Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Global diffusion and local implementation – the discourse particle LIKE around the world

Martin Schweinberger

Universität Hamburg
martin.schweinberger@uni-hamburg.de

Research question

- What happens when an innovation diffuses through a speech community…
  - Does it spread in a predictable way or does it spread in a rather chaotic fashion?
- Well, there are models which describe exactly that…

  for instance, the models developed in the Labovian school of variationist sociolinguistics

Relevance and research question

- Our current models have the advantages that they are
  - Based on many studies (highly stable)
  - High predictive and explanatory power
- But these models also have shortcomings, e.g.
  - Based mostly on studies of AmE and EngE
  - Focus on phonological changes
  - Neglect of language contact and multilingualism
  - (Overemphasizing generational change, the apparent time construct and face-to-face contact)

Relevance and research question

- However, nowadays innovations are available globally due to increased language and dialect contact

- Sociolinguists are of course aware of that and have begun to look at globalization from a variatonist perspective (e.g. Meyerhoff & Niedzielski 2003; Buchstaller 2008; Buchstaller & D'Arcy 2009)

- Implications for the standard model of language change from below?

  in other words…

  How appropriate is the Labovian paradigm, i.e. the standard model, in cases of…

  – dialect contact and multilingualism
  – lexical change
  – culturally diverse settings?

Labov's standard model of language change from below

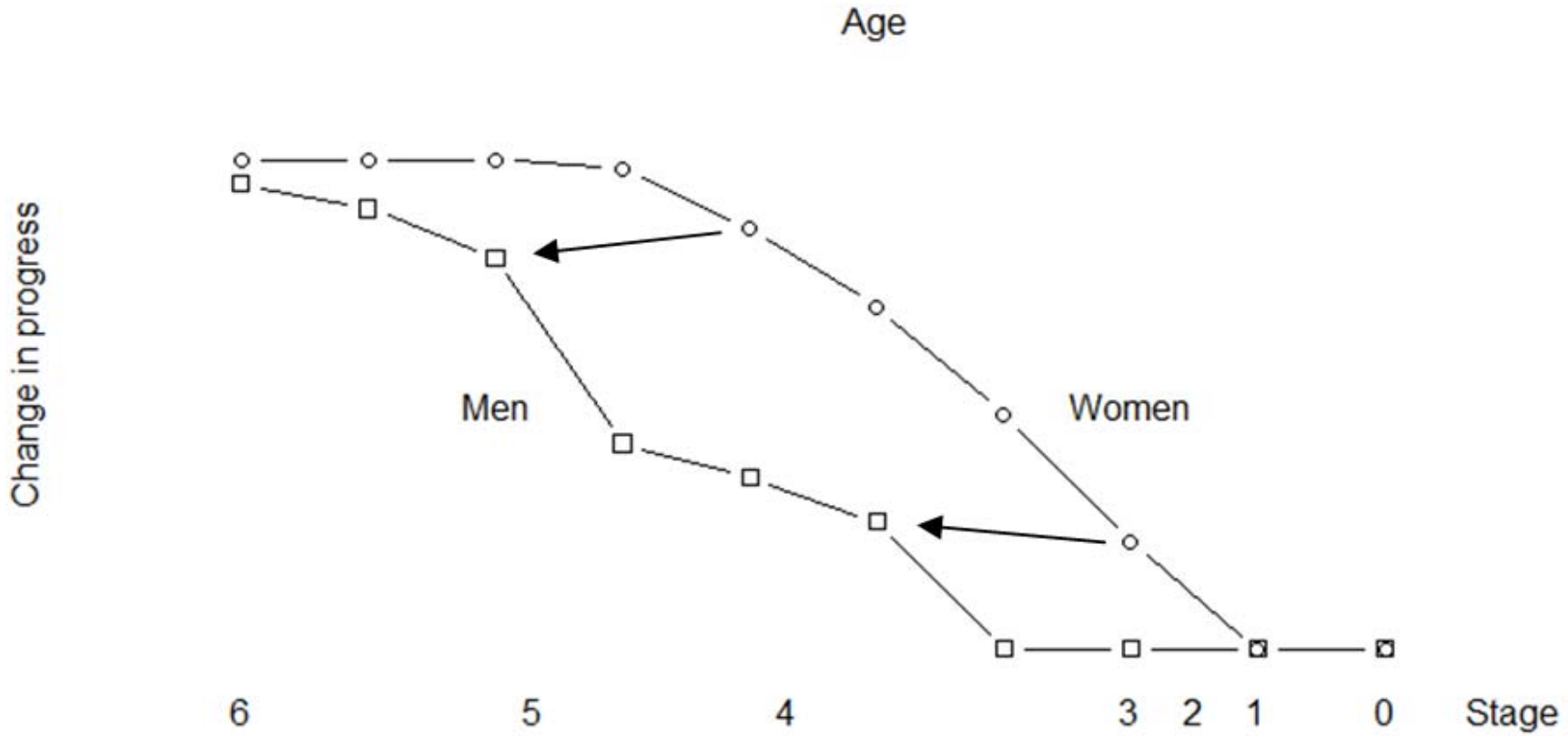


Figure 1: Six-stage model of gender relations in linguistic change from below (Labov 1994: 65)

The impact of gender

- Most of the linguistic changes which have been studied in the 2nd half of the 20th century show a high degree of social stratification and gender differenciation (Labov 1994, 2002).

- "[G]irls must rely more on symbolic manifestations of social membership than boys" (Eckert 1989: 265; cf. also Eckert 1999, 2003, 2004)

  - Female adolescents are expected to show a preference for linguistic means to indicate group membership

  - Male adolescents are expected to express group membership less than females and through non-linguistic means .

- Labov's (1994, 2002, 2010) model and Eckert's (1989, 2003. 2004) analyses both predict that innovations are adopted first by female adolescents.

The discourse marker LIKE

(1) Clause-medial LIKE

a. Should I make *like* tartar sauce or something really decadent? (Santa Barbara Corpus: sbc003$Marilyn)

b. Cos he just won a place to *like* <,> Canterbury Cathedral Choir School. (ICE-Canada: S1A-051$A)

- Properties
  - Modifies element to its right (rightward scope)
  - Hedges or focuses lower level constructions (phrases and words, not clauses and sentences)
  - Globally available innovation (occurs in almost all regional varieties)

Data editing & processing

- *International Corpus of English* (ICE)
  - Distinct regional components
    - Santa Barbara Corpus of Colloquial American English
    - ICE Canada
    - ICE Ireland
    - ICE New Zealand
  - Matching standardized design
  - Most informal register (S1A): face-to-face conversation, telephone calls (highest frequency of non-standard and discourse features)

Data editing & processing

- *SPEAKER-BASED ICE*
  - Allows speaker-based analyses (information about a speaker‘s age, gender, etc.)
  - For each speaker a PERL skript created a file that contained only the utterances of that speaker
  - Extraction of the word counts for each speaker
  - Using the word counts, it was possible to calculate the per-1,000-word frequencies of clause-medial LIKE for each speaker

    Making use of the speaker information provided by the ICE teams the normalized frequencies of clause-medial LIKE were correlated with extra-linguistic factors (speaker age, gender, occupation, current place of residence, education level, etc)

Data editing & processing

- *SPEAKER-BASED ICE*
  - Allows speaker-based analyses (information about a speaker's age, gender, etc.)
  - For each speaker a PERL skript created a file that contained only the utterances of that speaker
  - Extraction of the word counts for each speaker using PERL
  - Using the word counts, it was possible to calculate the per-1,000-word frequencies of clause-medial LIKE for each speaker

Making use of the speaker information provided by the ICE teams the normalized frequencies of clause-medial LIKE were correlated with extra-linguistic factors (speaker age, gender, occupation, current place of residence, education level, etc)

Data cleaning, editing & processing

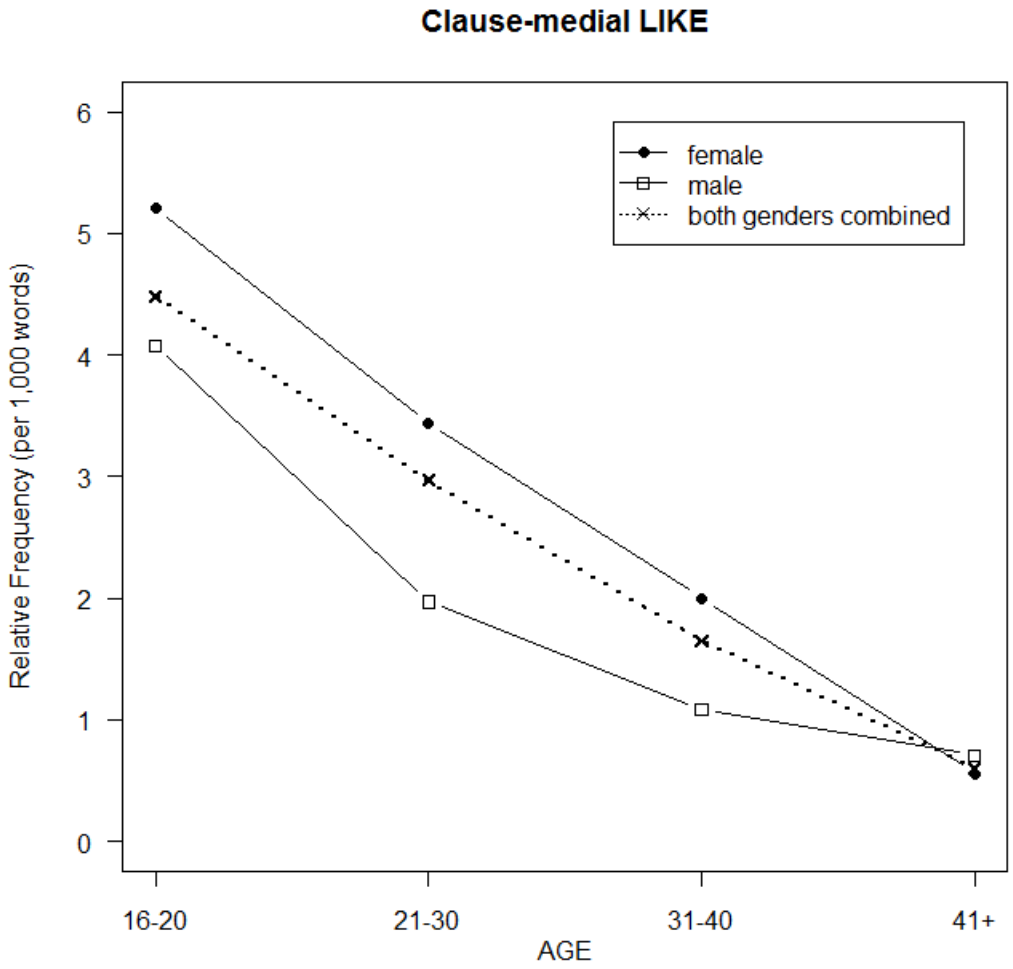
Variety (ICE component)	**Words** (SUM)	**Speaker** (N)	**INI** (N)	**MED** (N)	**FIN** (N)	**NON** (N)	**NA** (N)	**ALL** (N)
Canada	194,574	244	368	381	26	112	13	900
Santa Barbara C.	246,258	163	220	390	1	234	15	860
Ireland	189,787	309	249	237	318	118	14	936
New Zealand	229,193	227	209	183	20	115	2	529
SUM	**859,812**	**943**	**1,046**	**1,191**	**365**	**579**	**44**	**3,225**

Table 3: Overview of the data used for the present analysis

Results

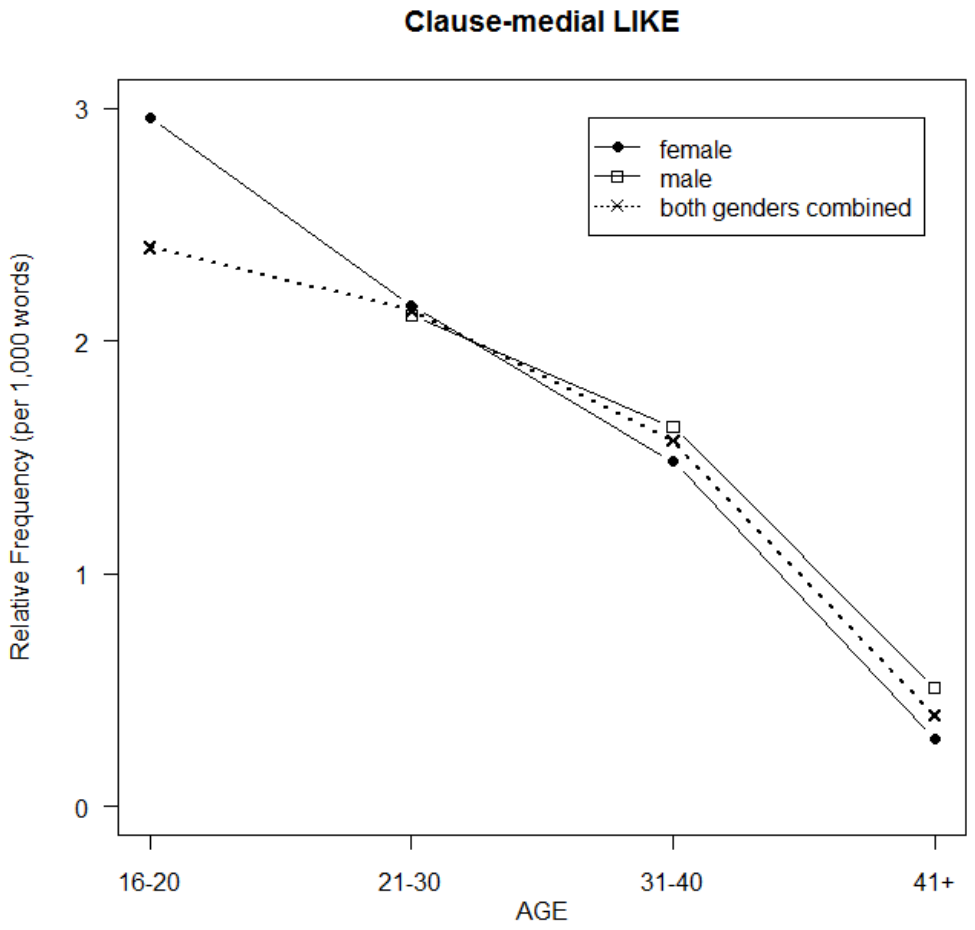
Results

Canadian English



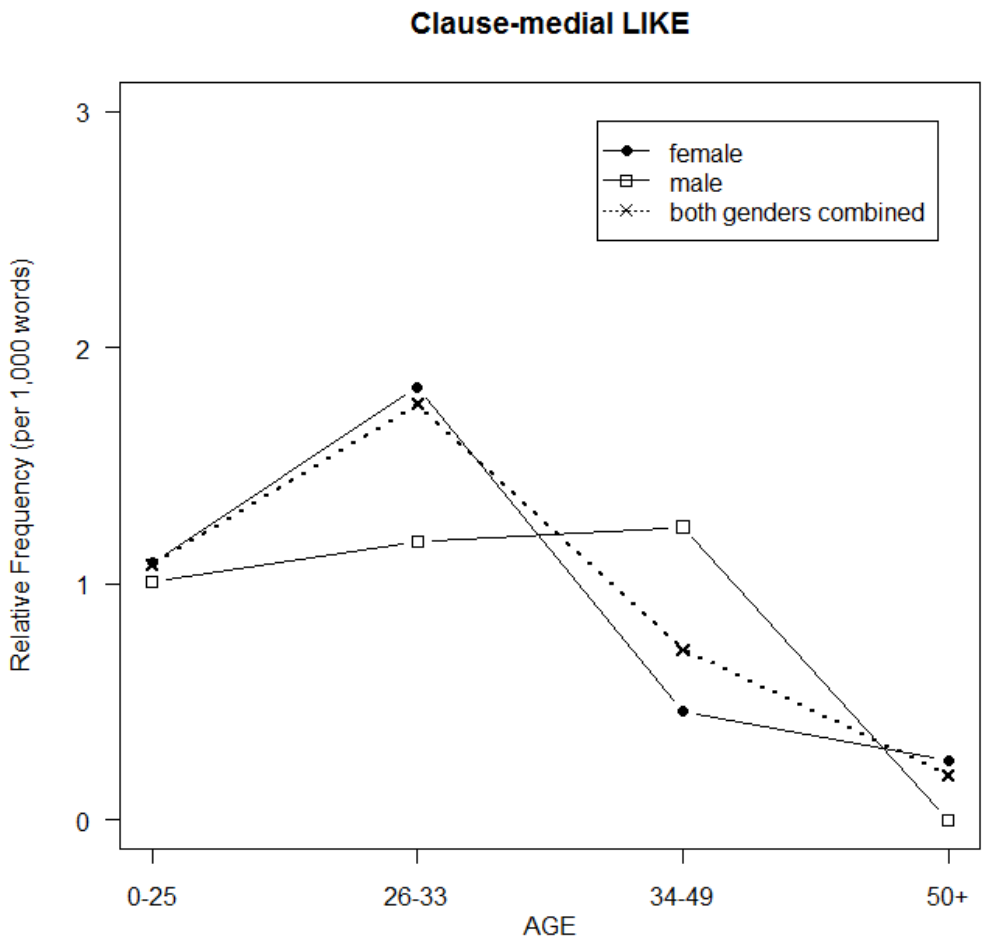
Results

American English



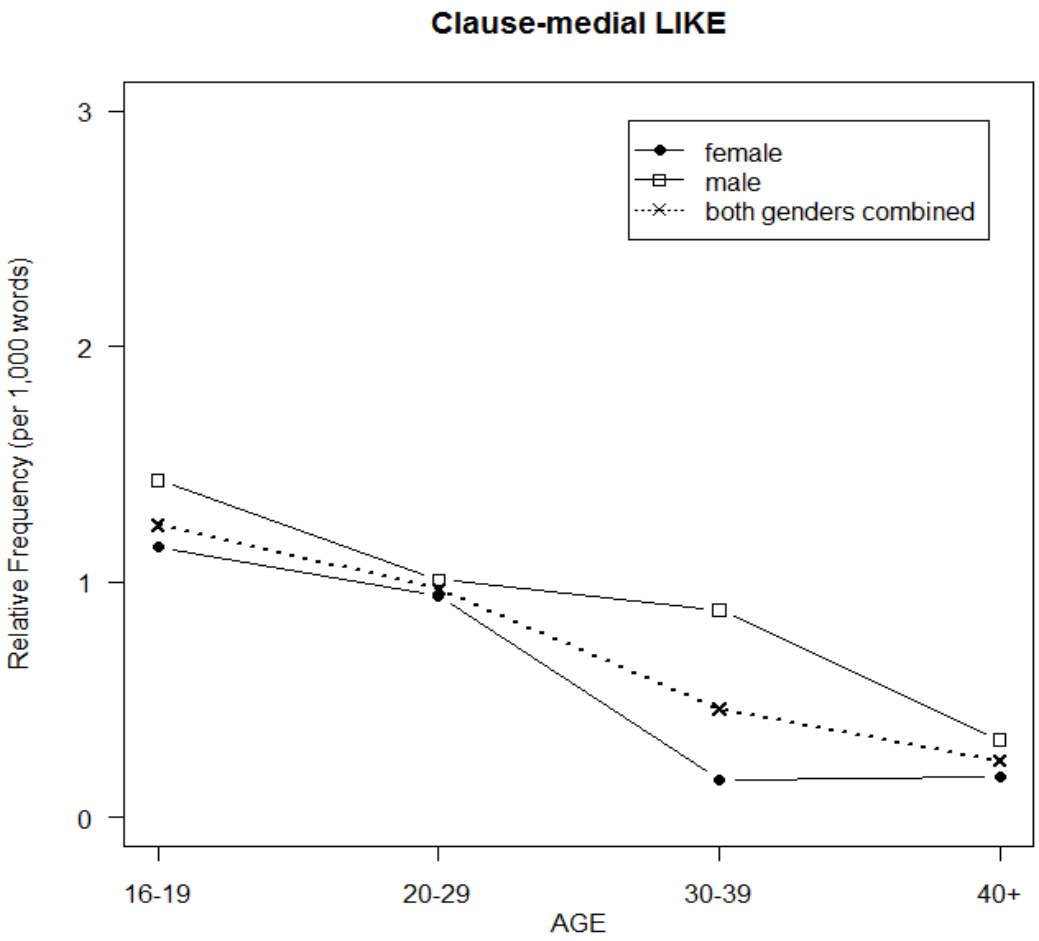
Results

Irish English



Results

New Zealand English



Summary & Discussion

- Supra-locally stable patterns
  - Monotonic recess with age

    The results confirm that “the association of *like* with younger speakers seems to hold across the English-speaking world” (D’Arcy 2007: 391).
- Variety-specific patterns
  - Degree and direction of gender differences

    “These trends show that sex differences […] are developmental, and are learned. They do not appear to be endemic to the features themselves, but are created in the speech community, within the peer group” (Tagliamonte 2005: 1912-1913).

Summary & Discussion

- Phonological change (transmission)
  - generational change (slow)
  - Distinct social stratification and gender differentiation
  - High quality of face-to-face contact required: Media are negligible with respect to transmission (Labov 2001: 2001: 228-229, 362-363, 385)
- Lexical change (diffusion)
  - (partial) communal change (rapid)
  - Less social stratification and gender differentiation
  - No high quality of face-to-face contact required
  - Transmission via mass-media (Muhr 2003)

Summary & Discussion

Outlook

- Expansion of the use of speaker-based analyses using the ICE 2.0
  - include more varieties
    - upcoming ICE components of ESL varieties
    - maybe use ICLE components (EFL varieties)
  - expand the pool of innovations
    e.g. quotative complementizer BE LIKE, innovative intensifiers (totally, etc.), innovative general extenders (and stuff, and shit)
  - Sociolinguistics from a global perspective
  - Evaluating the stability of systematic patterns underlying the diffusion of innovations

Thank you very much for LIKE your attention

References

Baayen, Harald. 2008. *Analyzing linguistic variation – a practical introduction using R*. Cambridge: Cambridge University Press.

Buchstaller, Isabelle. 2008. The localization of global linguistic variants. *English World-Wide* 291: 15–44.

Buchstaller, Isabelle & Alexandra D'Arcy. 2009. Localized globalization: A multi-local, multivariate investigation of quotative be like. *Journal of Sociolinguistics* 13.3: 291–331.

Cameron, A.C. & P.K. Trivedi. 1998. Regression analysis of count data. Cambridge: Cambridge University Press.

Chambers, Jack K. 22003. *Sociolinguistic Theory: Linguistic Variation and Its Social Significance.* Oxford: Blackwell.

D'Arcy, Alexandra. 2007. LIKE and language ideology: disentangling the fact from fiction. American Speech 82: 386–419.

Downes, William. 1998. *Language and Society*. Cambridge&NewYork: Cambridge University Press.

Du Bois, John W., Chafe, Wallace L., Meyer, Charles, and Thompson, Sandra A. 2000. *Santa Barbara corpus of spoken American English, Part 1.* Philadelphia: Linguistic Data Consortium. ISBN 1-58563-164-7.

Du Bois, John W., Chafe, Wallace L., Meyer, Charles, Thompson, Sandra A., and Martey, Nii. 2003. *Santa Barbara corpus of spoken American English, Part 2.* Philadelphia: Linguistic Data Consortium. ISBN 1-58563-272-4.

References

Du Bois, John W., and Englebretson, Robert. 2004. *Santa Barbara corpus of spoken American English, Part 3.* Philadelphia: Linguistic Data Consortium. ISBN 1-58563-308-9.

Du Bois, John W., and Englebretson, Robert. 2005. *Santa Barbara corpus of spoken American English, Part 4.* Philadelphia: Linguistic Data Consortium. ISBN: 158563-348-8.

Eckert, Penelope. 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1: 245-267.

Eckert, Penelope. 2004. The good woman. In Mary Bucholtz (ed.), *Language and woman's place: Text and commentaries*, 165–170. New York: Oxford University Press.

Eckert, Penelope & Sally McConnell-Ginet. 2003. *Language and Gender*. New York: Cambridge University Press.

Eckert, Penelope. 2003. Language and gender in adolescence. In Janet Holmes & Miriam Meyerhoff (eds.), *Handbook of Language and Gender*. Oxford: Blackwell.

Eckert, Penelope & Sally McConnell-Ginet. 1999. New generalizations and explanations in language and gender research. *Language in Society* 28(2). 185–202.

References

ICE website. International Corpus of English homepage: www.ucl.ac.uk/english-usage/ice/index.htm.

ICE-Ireland 1.2.1: The International Corpus of English: The Irish Component. 2008. CD-ROM.

Labov, William. 1994. *Principles of Linguistic Change. Vol. 1, Internal Factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of Linguistic Change. Vol. 2, Social Factors*. Oxford: Blackwell.

Labov, William. 2002. "Driving forces in linguistic change". In: *Proceedings of the 2002 International Conference on Korean Linguistics.*

Meyerhoff, Miriam & Nancy Niedzielski. 2003. The globalization of vernacular variation. *Journal of Sociolinguistics* 7: 534-555.

Muhr, Rudolf. 2003. Language change via satellite: The influence of German television broadcasting on Austrian German. *Journal of Historical Pragmatics* 4.1: 103-128.

Tagliamonte, Sali. 2005. So who? Like how? Just what? Discourse markers in the conversation of young Canadians. In: *Journal of Pragmatics* 37: 1896-1915.

Tagliamonte, Sali A. & Alexandra D'Arcy. 2009. Peaks Beyond Phonology: Adolescence, Incrementation, and Language Change. Language 85.1: 58-108.

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Global diffusion and local implementation – the discourse particle LIKE around the world

Martin Schweinberger

Universität Hamburg
martin.schweinberger@uni-hamburg.de

Appendix

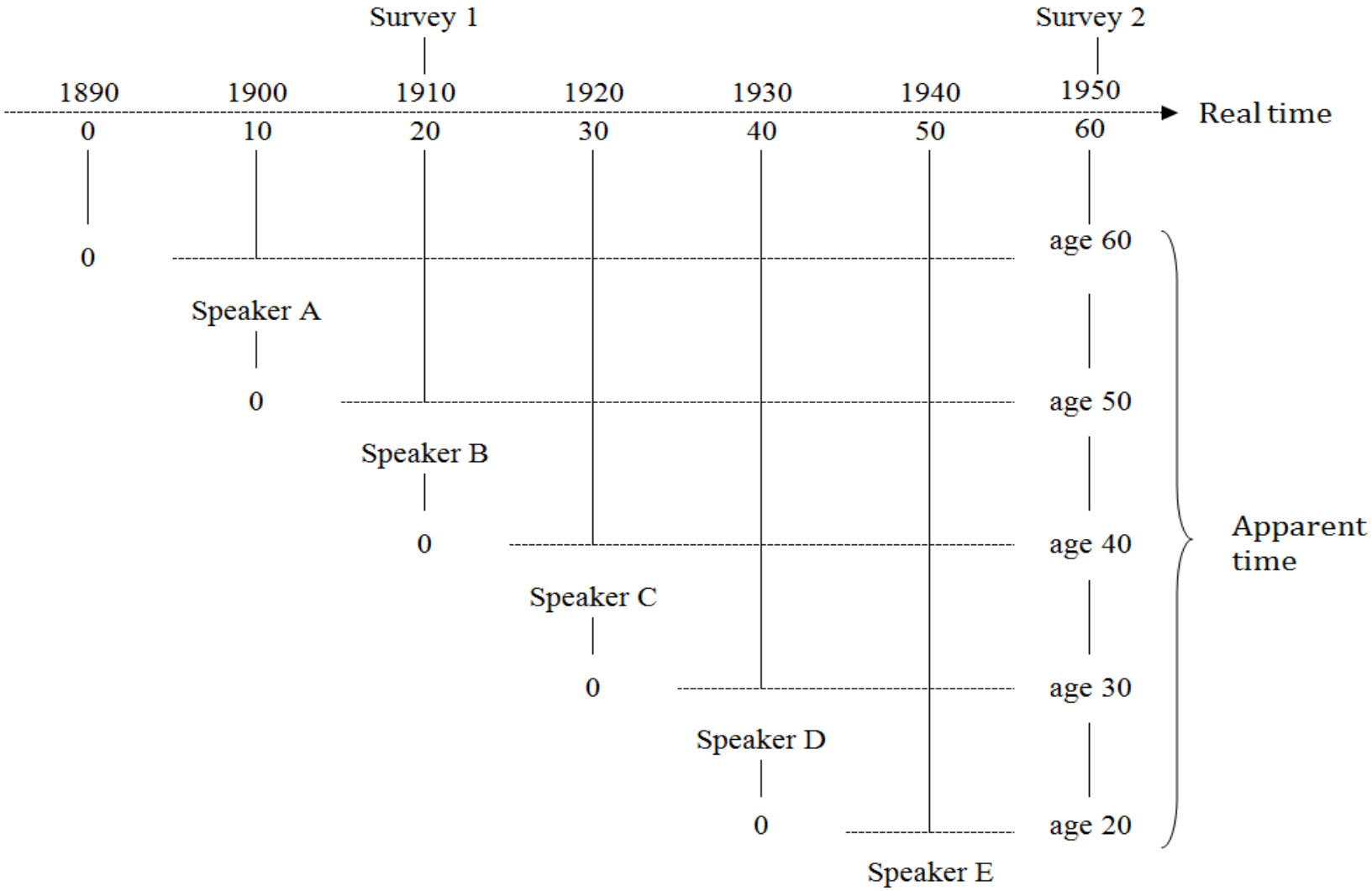


Figure 2: Real and apparent time in language change (Downes 1998: 238)

Types of change (Labov 1994: 84)

- *Age-grading*
  Individuals change their linguistic behaviour throughout their lifetimes, but the community as a whole does not change.

- *Generational change*
  "Individual speakers enter the community with a characteristic frequency for a particular variable, maintained throughout their lives; but regular increases in the values adopted by individuals, often incremented by generations, lead to linguistic change for the community."

- *Communal change*
  "In communal change all members of the community alter their frequencies together or acquire new forms simultaneously."

Multivariate statistics

- Multivariate regression model (Poisson Regression)
  - Based on a probability distribution which describes the occurrence of discrete events in a given interval (cf. Baayen 2008:296)
  - Used for modeling rare events, i.e. count data
    - For example, volcano outbreaks per 100 years, instances of cancer in a village within one year, or LIKE per 1000 words, .
  - glm package in R
    `family=quasipoisson` to compensate for moderate overdispersion, i.e. λ>1 (λ=probability of event*intervall size (number of draws) or variance of errors/mean>1); link=log
  - Relatively strict model requirements!

Multivariate statistics

- Multivariate regression model (Poisson Regression)
  - Based on a probability distribution which describes the occurrence of discrete events in a given interval (cf. Baayen 2008:296)
  - Used for modeling rare events, i.e. count data
    - For example, volcano outbreaks per 100 years, instances of cancer in a village within one year, or LIKE per 1000 words, .
  - glm package in R
    `family=quasipoisson` to compensate for moderate overdispersion, i.e. λ>1 (λ=probability of event*intervall size (number of draws) or variance of errors/mean>1); link=log
  - Relatively strict model requirements!

Multivariate statistics

- Multivariate regression model (Poisson Regression)
- Dependent Variable
  - Clause-medial LIKE per 1,000 words (counts)
- Independent Variables
  - Age (nominal: age group 1, 2, 3, or 4; 1 = dummy)
  - Sex/Gender (nominal: m/f)
  - PAI (priming, accommodation, Idiosyncratic overuse; numeric) (to save-guard against over-estimating extra-linguistic variables; ratio: instances of all instances of LIKE in a given conversation divided by the token count of that conversation)

Additional real-time analyses

- Non-parametric t-tests (t-score and significance level)
- Dependent Variable
  - Use of clause-medial LIKE per 1,000 words (frequencies) by a certain age group
- Independent Variables
  - Date of data compilation
    - X (1990-1994)
    - Y (2001-2005)

Results

Canadian English

Table 2: Results of the multivariate regression for clause-medial LIKE in CanE.

MED	Estimate (coefficient)	Std. Error	z value	Pr(>|z|)
(Intercept)	0.52	0.256	2.04	<.05*
SEX: Male	-0.00	0.192	-0.01	.98
A2	-0.36	0.268	-1.36	.17
A3	-0.32	0.195	-1.67	<.1
A4	-1.50	0.275	-5.44	<.001***
PAI	0.04	0.006	7.28	<.001***
SEX: Male*A3	-0.74	0.377	-1.93	<.05*

Table 3: LIKE use in CanE with respect to AGE and the date of data compilation.

	A1 (16-20)	A2 (21-30)	A3 (31-40)	A4 (41+)
ALL	n.s.	N.A.	n.s.	n.s.
INI	n.s.	N.A.	n.s.	n.s.
MED	n.s.	N.A.	n.s.	n.s.
FIN	n.s.	N.A.	n.s.	n.s.
NON	-1.607.	N.A.	n.s.	n.s.

Results

American English

Table 4: Results of the multivariate regression for clause-medial LIKE in AmE.

MED	Estimate (coefficient)	Std. Error	z value	Pr(>|z|)
(Intercept)	0.32	0.342	0.95	.33
A2	0.15	0.379	0.40	.68
A3	-0.03	0.437	-0.08	.93
A4	-1.45	0.416	-3.48	<.001***
PAI	0.01	0.005	2.90	<.01**

Irish English

Table 5: Results of the multivariate regression for clause-medial LIKE in IrE.

MED	Estimate (coefficient)	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.35	0.230	2.30	<0.05*
A2	0.06	0.289	0.230	.81
A3	-0.58	0.488	-1.19	.23
A4	-1.19	0.441	-2.71	<.01**
PAI	0.04	0.009	5.25	<.001***
REG:South	-0.50	0.270	-1.88	<.1
A4*SEX: Male	-15.09	0.636	-23.71	<.001***

Table 6: LIKE in IrE use with respect to AGE and the date of data compilation.

	A1 (0-25)	A2 (26-33)	A3 (34-49)	A4 (50+)
ALL	-1.36·	-3.13**	n.s.	n.s.
INI	-1.60·	n.s.	n.s.	n.s.
MED	-1.48·	-3.00**	-1.50·	n.s.
FIN	1.39·	-3.22**	n.s.	n.s.
NON	-2.29*	n.s.	n.s.	n.s.

Results

New Zealand English

Table 7: Results of the multivariate regression for clause-medial LIKE in NZE.

MED	Estimate (coefficient)	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.13	0.433	-2.61	<.01**
SEX: Male	0.85	0.327	2.90	<.01**
A2	-0.08	0.402	2.16	<.05*
A3	-0.28	0.487	-0.59	.55
A4	-0.96	0.564	-1.70	<.1
PAI	0.16	0.023	7.02	<.001***
ADC	-073	0.277	-2.65	<.01**
SML	-0.26	0.247	-1.06	.28
SEX:Male:A2	-1.24	0.418	-2.97	<.01**