Adjective amplification in English - determining factors of success in language change

Dr. Martin Schweinberger

slides available at

www.martinschweinberger.de

R code upon request

Research project

Acquisition, Variation, and Diachronic Change in and across English Amplifier Systems (Schweinberger fcb, 2021a, 2020b,c, fca, 2021b, 2020d,e,a).

(1) i believe they do a very good job (WSC#DGB009:0505:HA)

(2) that was really cheap (WSC#DPC123:1255:VV)

(3) that's so bad (WSC#DPC212:0105:TM)

(4) and a lot of them are moving into an area which <,> has been extremely difficult to get networked (WSC#DGZ064:0855:DA)

Phenomenon

Intensification is related to the semantic category of *degree* (degree adverbs) and ranges between very low intensity (downtoning) and very high (amplifiers) (Quirk et al. 1985: 589–590).

- Amplifiers (Tagliamonte 2008)
  - Boosters, e.g. *very*
  - (Maximizers, e.g. *completely*)
- Downtoners
  - Approximators, e.g. *almost*
  - Compromisers, e.g. *more or less*
  - Diminishers, e.g. *partly*
  - Minimizers, e.g. *hardly*

Why analyze adjective amplification?

- Amplification is major area of grammatical change (cf. Brinton and Arnovick 2006: 441)
- Amplification is crucial for the "social and emotional expression of speakers" (Ito and Tagliamonte 2003: 258)
  $\rightarrow$ interesting for studies of social identity construction and identity marking
- Amplification is a linguistic subsystem which allows precise circumscription of a variable context (Labov 1972, 1966: 49)

Amplification represents an ideal case for testing mechanisms underlying language change!

Previous Research

Previous Research

Amplification

- substantial amount of corpus-based research on intensification (e.g Aijmer 2011, 2018; Fuchs 2016, 2017; Núñez Pertejo and Palacios 2014; Palacios and Núñez Pertejo 2012)
  → but mostly either focused on individual intensifiers or without regard to the intensified adjectives
- associated with teenage talk and young(ish) (female) speakers
  (Bauer and Bauer 2002; D'Arcy 2015; Macaulay 2006; Tagliamonte 2006, 2008)
- recently amplifier-adjective bigrams have come more into focus (e.g. Schweinberger 2017; Wagner 2017a,b)

Previous Research

- Intensifying *really* replaces *very* (lexical replacement)
  (e.g. D'Arcy 2015; Ito and Tagliamonte 2003; Tagliamonte 2005, 2008)
- Previous study of intensification in NZE (D'Arcy 2015; Bauer and Bauer 2002)

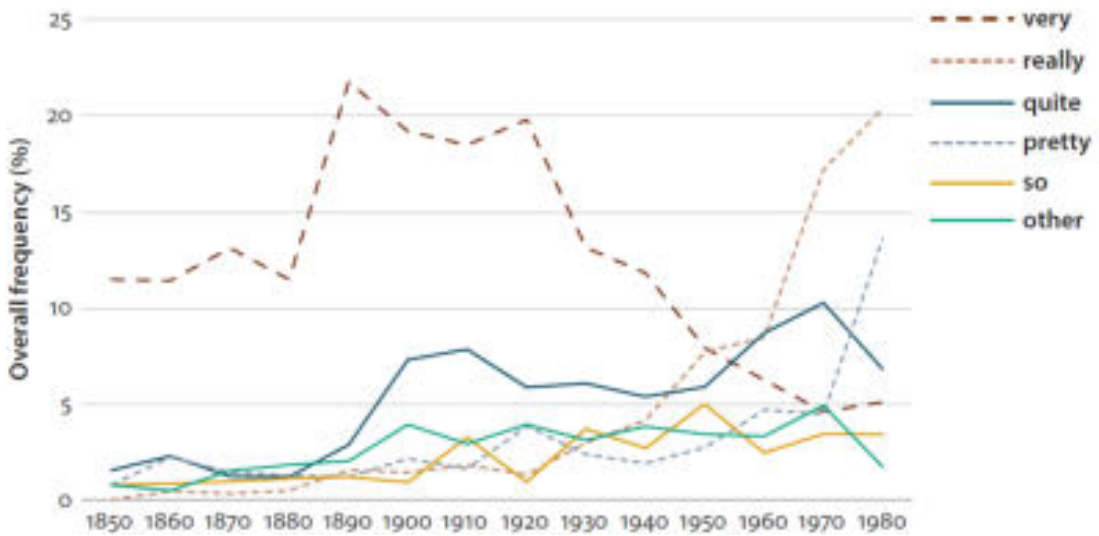


Figure 1: Adapted from D'Arcy (2015: 468)

Previous Research

"The dramatic expansion of *very* weakened its ability to amplify an adjectival head, necessitating a new form[. . .]: very lost its pragmatic strength and *really* was recruited in its place." D'Arcy (2015: 468)

- *very* (D'Arcy 2015: 480)
  Correlated with only adjective type (gradable vs non-gradable) and age of speakers among speakers born between 1932 and 1980.
- *really* (D'Arcy 2015: 481)
  Correlated with speaker age, syntactic function, and gender among speakers born between 1932 and 1980.

Research Question

Q

What sets *really* apart?

Why is *really* replacing *very* and not, e.g., *so*, *quite*, *pretty*?

Hypothesis 1 (Broadening)

Successful variants are more bleached

(Mair 2004: "delayed increase of discourse frequency" hypothesis)

Argument

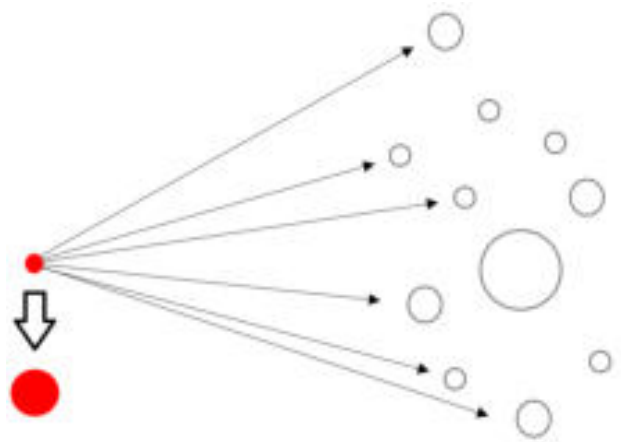
→ co-occurrence with many different adj. types

→ frequent use

→ deeper cognitive entrenchment

→ easier retrieval from memory

→ dominance within the amplifier system.



Prediction

Co-occurrence with many different adjective types

→ high lexical diversity

→ weak coll. attraction with specific adj. types

Hypothesis 2 (Specialization)

Successful variants associate with few but frequent adj. types

(Lorenz 2002: 144; Méndez-Naya 2003: 375; Tagliamonte and Roberts 2005: 285)

Argument

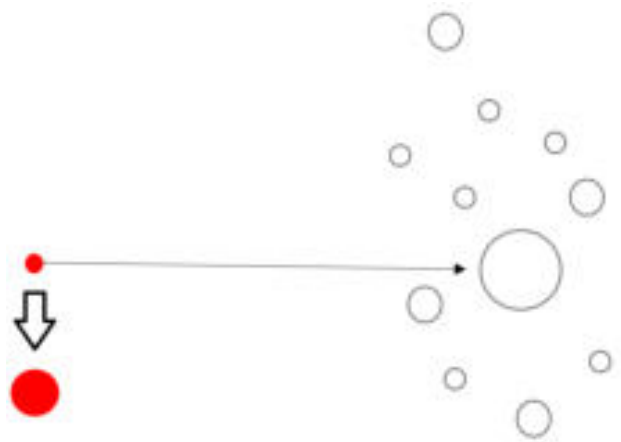
→ co-occurrence with high-freq. adj. types

→ frequent use

→ deeper cognitive entrenchment

→ easier retrieval from memory

→ dominance within the amplifier system.



Prediction

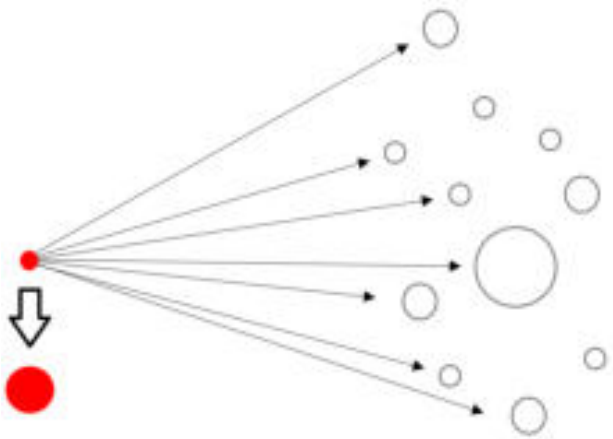
Co-occurrence with few high frequency adjectives

→ low lexical diversity

→ strong coll. attraction with high-freq. adj. types

Hypothesis 3 (Randomness)

We cannot predict which variants become successful based on their coll. profile.



Data and Methodology

Wellington Corpus of Spoken New Zealand English (WSC)

One-million-word corpus of transcribed English compiled between 1988 and 1994 (Peters fc)

- Formal Speech/Monologue 12
- Semi-formal Speech/Elicited Monologue 13
- Informal Speech/Dialogue 75

Accompanied by metadata and biodata of speaker (extremely interesting resource for variationist analyses)

Data Processing

- WSC
  (https://www.wgtn.ac.nz/lals/resources/corpora-default/corpora-wsc)
- Part–of-speech tagged (OpenNLP vis R) the
- Retrieved adjectives (PoS–tag JJ)
- Determined whether adjective were preceded by an amplifier (member of a predefined set of amplifiers)
- Implemented a Sentiment Analysis of adjective types (emotional vs non-emotional) using the Syuzhet library in R (Jockers 2017)

Data Processing

- Determined if the same amplifier type had occurred within a span of three adjective slots previously ($\rightarrow$ priming)
- Added token frequency of adjective type (Tagliamonte and Roberts 2005)
- Removed...
  - negated adjectives
  - comparative and superlative forms
  - adjectives that were not amplified by at least two different amplifier types
  - adjectives that were preceded by downtoners
  - strange forms (e.g. *much*)

Data Processing

- Added semantic classification of adjective types based on Dixon (1977) (cf. also D'Arcy 2015; Tagliamonte and Roberts 2005; Tagliamonte 2006, 2008)
- Manual cross–evaluation of automated classification
- Metadata (genre, audience size, conversation type: samesex, mixedsex) and speaker information(age, sex)

Data Processing

Implementation of a Semantic VSM (Levshina 2015) (data driven/usage-based exclusion of maximizers/definition of variable context)

	get	see	use	hear	eat
knife	31	16	69	0	2
cat	36	38	4	4	6
dog	66	58	9	34	28
boat	46	21	17	4	0
cup	59	6	5	1	1
pig	4	15	3	1	7
banana	7	2	2	0	12

Table 1: Example based on Evert (2018).

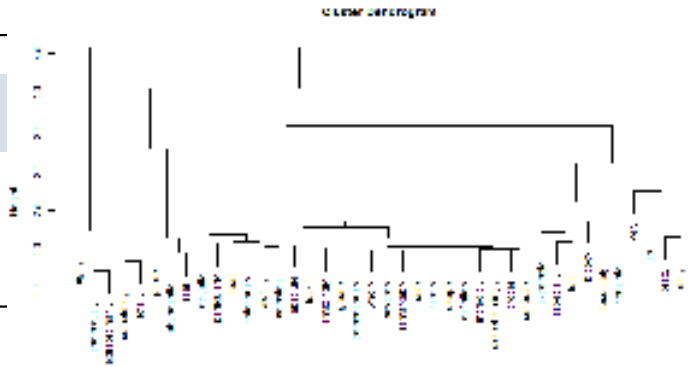


Figure 2 (right): Dendrogram showing the classification of adjective types based on their co-occurrence profile with adjectives.

Data Overview (WSC: spoken private dialogue)

Amplification	**N**	**%**	**Variants** (%)
∅ Amplification	9,570	84.8	
really	774	6.86	45.13
very	319	2.83	18.60
so	250	2.22	14.58
pretty	161	1.43	9.39
real	50	0.44	2.93
absolutely/bloody	19 (38)	0.17 (0.34)	1.11 (2.22)
totally	15	0.13	0.87
fucking	13	0.12	0.76
completely	11	0.10	0.64
others (<10)	84	0.74	4.90
Total	11,285 (1,715)	100 (15.2)	100

Table 2: Overview of amplifier frequencies and percentages in the final data set.

Data Overview

Age	**Sex**	**Speakers** (N)	**Adj.** (N)	***really*** (N)	***really*** (%)
16-19	Men	26	77	30	38.96
16-19	Women	47	224	125	55.80
20-29	Men	74	295	115	38.98
20-29	Women	124	573	351	61.26
30-39	Men	26	67	21	31.34
30-39	Women	35	136	70	51.47
40-49	Men	19	52	11	21.15
40-49	Women	34	100	27	27.00
50+	Men	22	71	9	12.68
50+	Women	34	118	13	11.02
Total		441	1,713	772	34.97

Table 3: Overview of adjective and *really* frequencies and percentages by age and gender in the final data set.

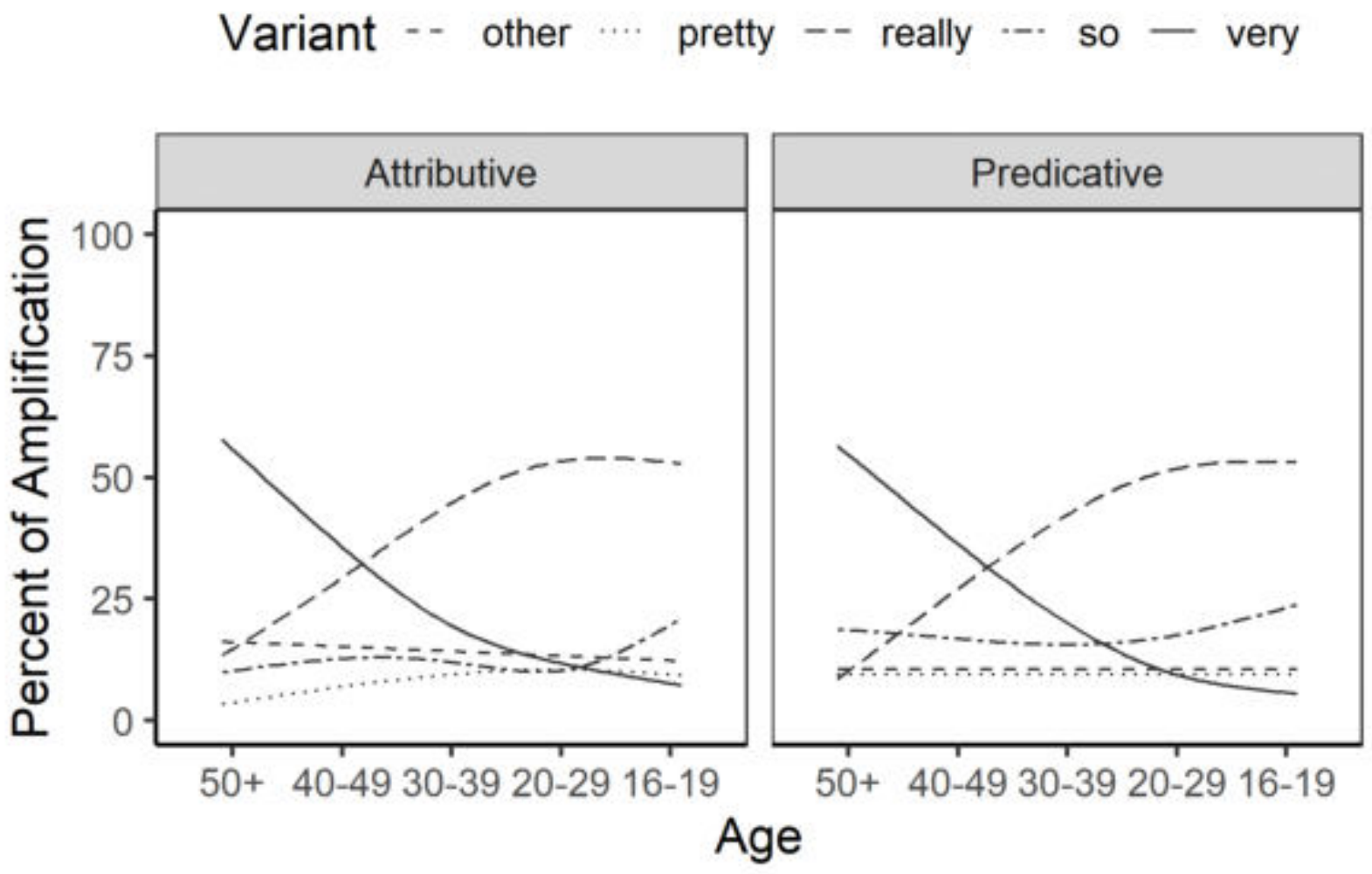


Figure 3: Percent of Amplifier Variants in Amplified Pre-Adjectival Slots across Syntactic Function.

Lexical Diversity

Hypothesis 1

- Co-occurrence with many different adjective types
  $\rightarrow$ high lexical diversity

Hypothesis 2

- Co-occurrence with few (but high frequency) adjective types
  $\rightarrow$ low lexical diversity

Lexical Diversity

$$LD = \frac{N_{Adj.\,Types}}{N_{Amp.\,Tokens}}$$

Example

Amplifier	Amp. Tokens	Adj. Types	Calculation	LD value
variant$_A$	10	1	$\frac{1}{10}$	.1
variant$_B$	10	5	$\frac{5}{10}$	.5
variant$_C$	10	10	$\frac{10}{10}$	1
very	67	12	$\frac{12}{67}$	.18
pretty	37	12	$\frac{12}{37}$	.32
bloody	2	1	$\frac{1}{2}$	.50

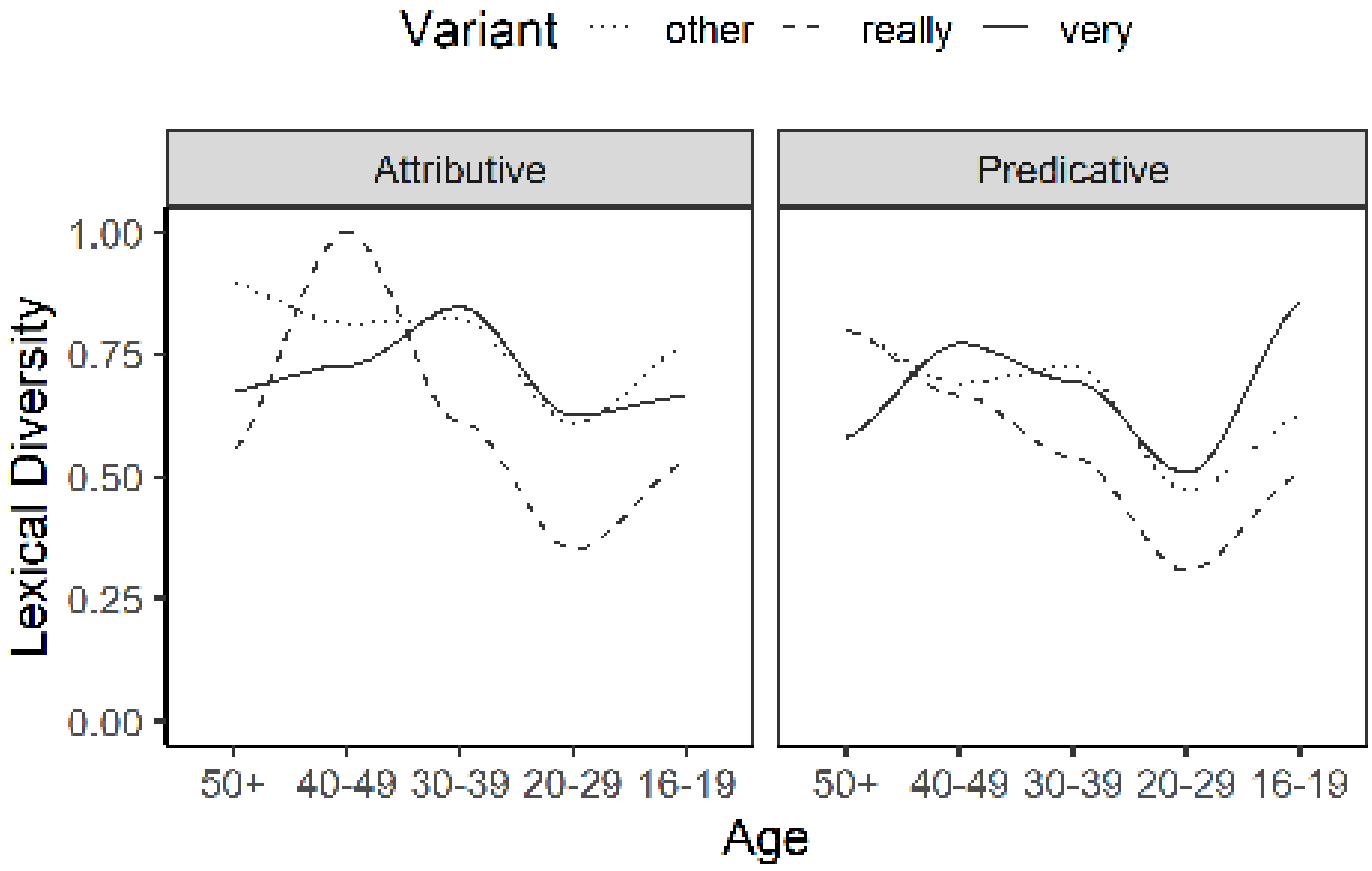


Figure 4: Lexical Diversity by Amp. Variant

Boruta Analysis

(Kursa et al. 2010)

Why use Boruta?

- Alternative to regressions that can handle small data sets

What is Boruta?

- Variable selection procedure
- Name derived from a demon in Slavic mythology who dwelled in pine forests
- Extension/improvement of random forests
- Hundreds of forests are grown $\rightarrow$ distribution of parameters rather than singe values (higher reliability)

Problems of Boruta

- Cannot handle multicollinearity(!)
- Does not model nested/grouped data structure

Boruta Analysis

(Kursa et al. 2010)

Procedure

1. Addition randomness: shuffling copies of all features (shadow features).
2. Training of a random forest classifier on the extended data
3. Application of a feature importance measure (Mean Decrease Accuracy)
4. Checking whether a real feature has a higher importance than the best shadow features at each iteration
5. Continuous removal of unimportant features (features that are less important than shadow features)

Variable Coding

Dependent Variable(s)				
really	nominal	yes/no occurrence of pre-adjectival *really*		
Independent Variable(s)				
Age	categorical	16-19 | 20-29 | 30-39 | 40-49 | 50+	extra	linguistic
AudienceSize	nominal	Dyad | MultipleInterlocutors	extra	linguistic
ConversationType	nominal	MixedSex | SameSex	extra	linguistic
Gender	nominal	Woman | Man	extra	linguistic
Priming	nominal	Primed | NotPrimed	extra	linguistic
Emotionality	nominal	positive | non-emotional | negative	intra	linguistic
Function	nominal	attributive | predicative	intra	linguistic
SemanticCategory	categorical	semantic category of adj.	intra	linguistic
Gradability	numeric	logged scaled probability of gradability	intra	linguistic
Adjective	categorical	260 adj. types	intra	linguistic
Frequemcy	numeric	logged + scaled frequency of adj. by age	intra	linguistic

Boruta Analysis

(Kursa et al. 2010)

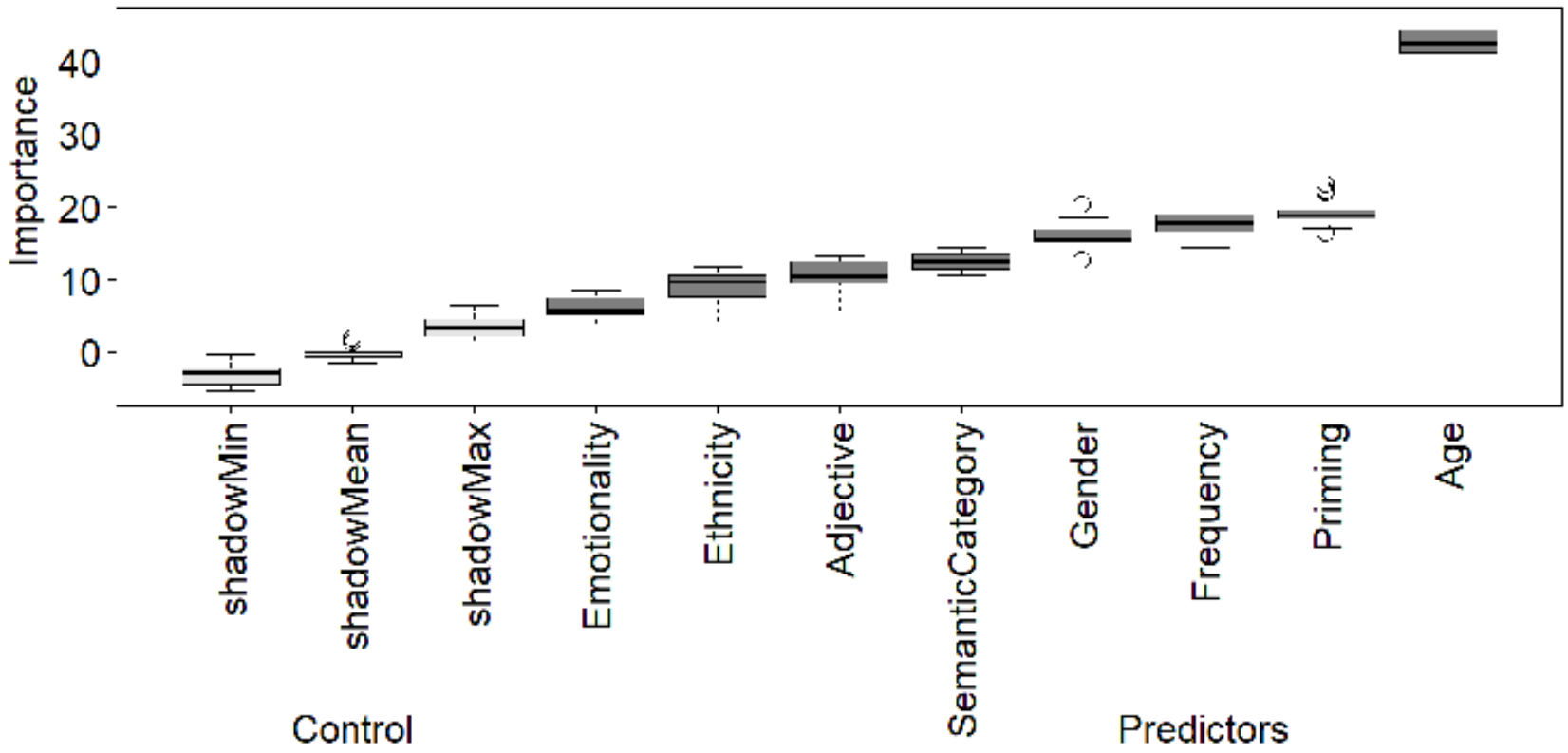


Figure 5: Predictor Strength based on Boruta Feature Selection

Mixed-Effects Binomial Logistic Regression

(Baayen 2008; Faraway 2016)

Why MEBLoR?

- Standard models for multivariate analyses
- Can handle nested/grouped data
- Can handle multicollinearity

What is MEBLoR?

- Evaluates the impact of various variables (and interactions) on dependent variable

Problems of MEBLoR

- Cannot handle small data sets (well)

Mixed-Effects Binomial Logistic Regression

(Baayen 2008; Faraway 2016)

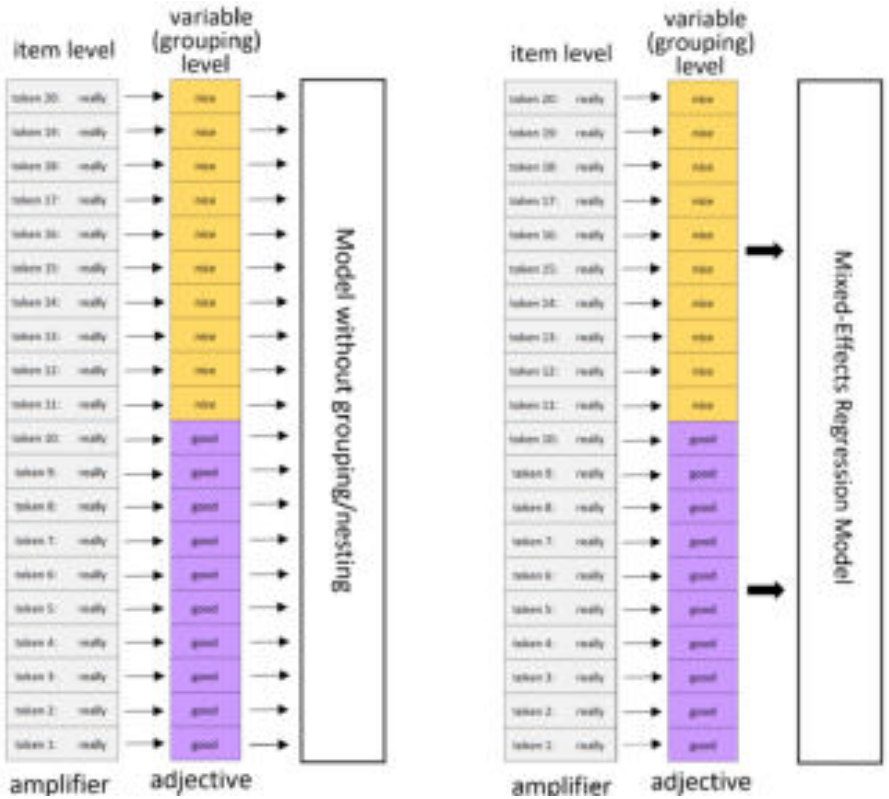


Figure 6: Difference between models without grouping/nesting and mixed-effects models (with grouping/nesting).

Mixed-Effects Binomial Logistic Regression

(Baayen 2008; Faraway 2016)

	Group(s)	Variance	Std. Dev.
Random Effect(s)	Adjective	0.3802	0.6166
Fixed Effect(s)	**OddsRatio**	**CIs**	**Significance (p)**
(Intercept)	0.22	0.13 - 0.37	<.001***
Age::20-29	1.09	0.81 – 1.47	0.552
Age::30-39	0.91	0.61 – 1.37	0.661
Age::40-49	0.37	0.23 – 0.60	<0.001***
Age::50+	0.12	0.07 – 0.21	<0.001***
Priming::Prime	2.58	1.92 – 3.48	<0.001***
Gender::Woman	2.20	1.73 – 2.80	<0.001***
Ethnicity::MAORI/PAKEHA	1.07	0.54 – 2.13	0.839
Ethnicity::OTHER	2.24	1.34 – 3.75	0.002**
Ethnicity::PAKEHA	1.76	1.17 – 2.65	0.007**
Model statistics			**Value**
Number of Groups			260
Number of cases in model			1707
Observed successes			770
Marginal/Conditional R^2			0.203 / 0.286
C			0.79
Somers' D_{xy}			0.59
Prediction accuracy			72.8% (54.9)
Model Likelihood Ratio Test		L.R. $\chi^2(3)$: 25.48	<.001***

Fixed-Effects

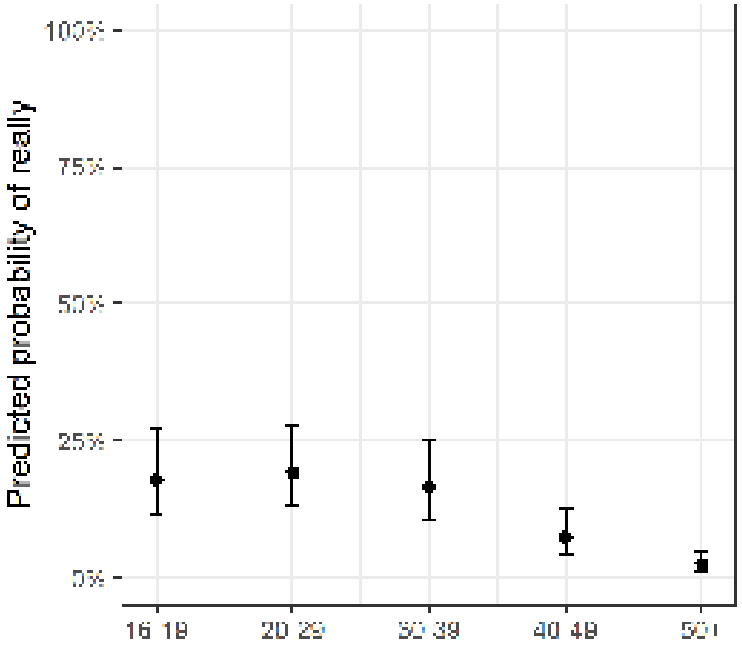


Figure 7: Probability of really by Age.

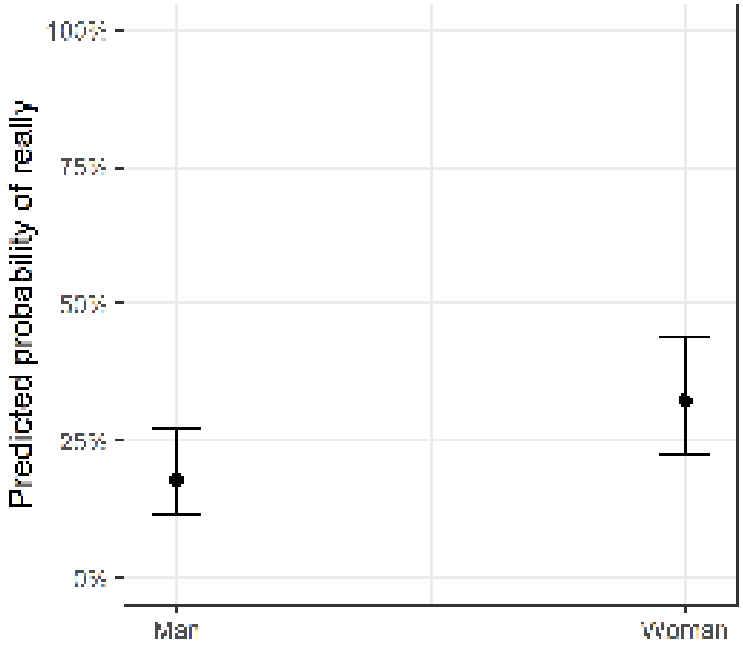
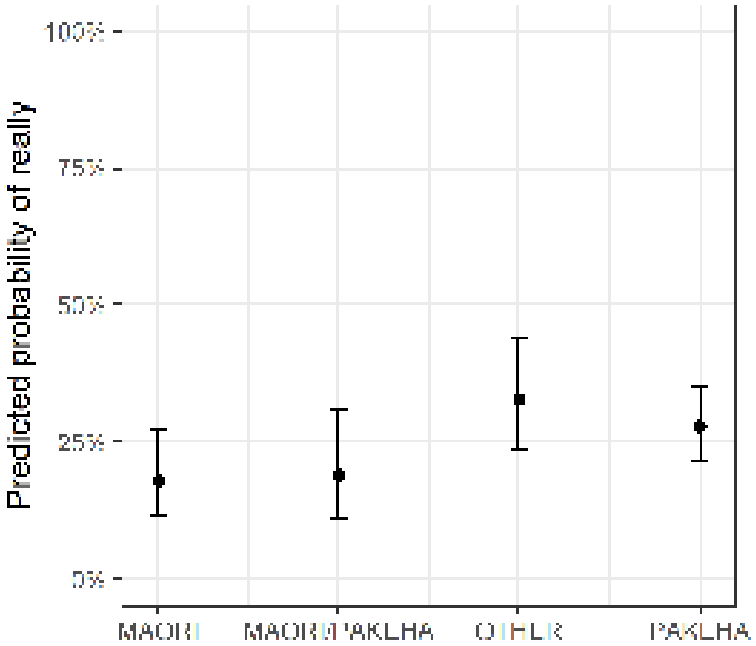


Figure 8: Probability of really by Gender.

Fixed-Effects



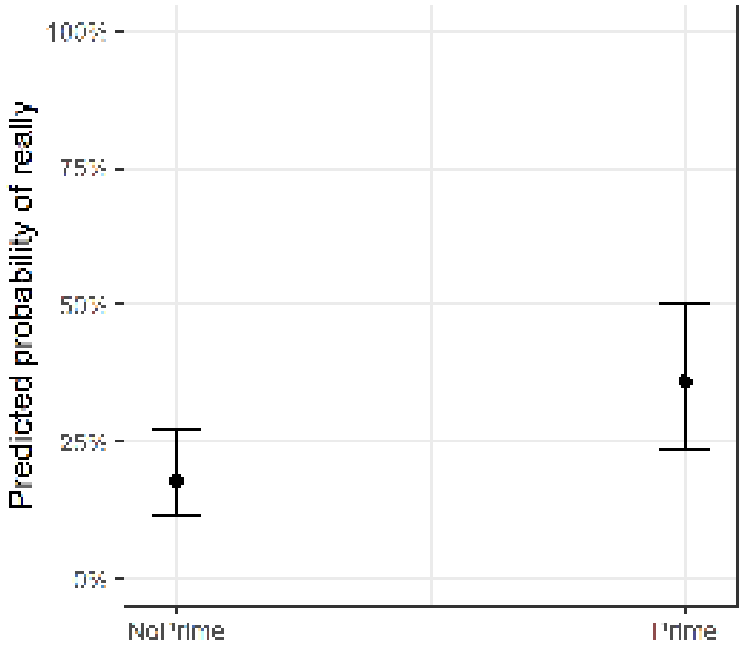


Figure 9: Probability of really by Ethnicity. Figure 10: Probability of really by Priming.

Random-Effects

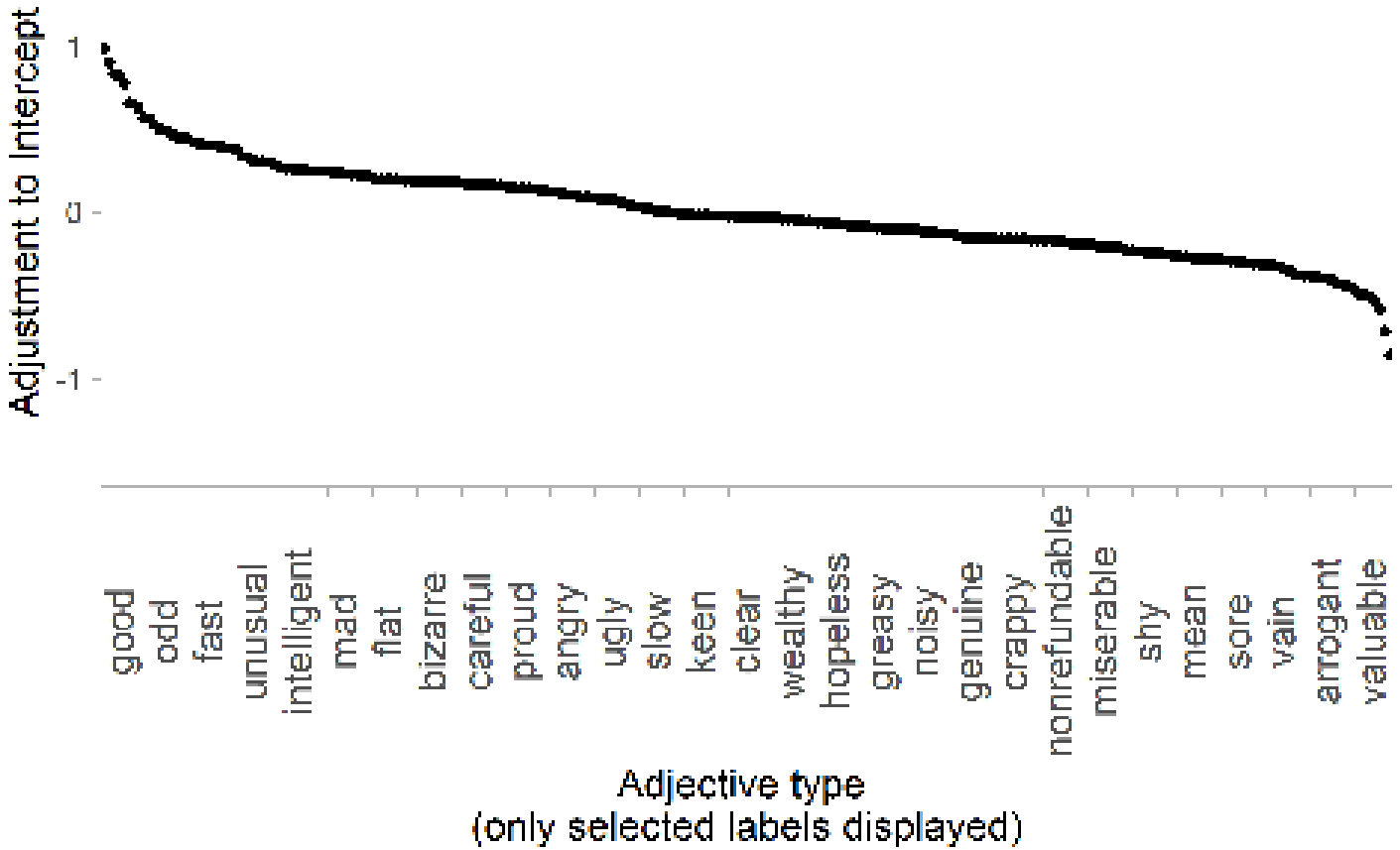


Figure 11: Adjustments to intercepts by Adjective.

Discussion & Outlook

Summary

Analysis suggests that really collocates with. . .

(a) few adjective types (low LD) but

(b) frequent adjectives (Boruta)

- MEBLoR: attraction *good* and *really* (highest adjustment)

	Lexical Diversity	**Coll. Strength (with *good*)**
H1 (bleaching/broadening)	high ✗	weak ✗
H2 (specialization)	low ✓	strong ✓

Discussion

- In the NZE amplifier system, the innovative variant (*really*) successfully replaced the dominant form (*very*) because *really* is favored by young speakers, women, and Pakehas and it collocates strongly with *good*.
- There are no signs that *really* experienced a broadening before taking over the system.
- Rather, the findings suggest that the broadening happens once a variant has already become dominant (substantiates Tagliamonte and Denis 2014)

Argument

1. The co-occurrence with HFAs lead to the innovative variant being used as a more expressive variant to amplify certain HFAs.
2. The frequency of the innovative form increased because it piggybacked on the frequency of the HFA.
3. Increase in use $\rightarrow$ more deeply entrenched.
4. Deeper entrenchment $\rightarrow$ increased ease of retrieval.
5. Higher ease of retrieval $\rightarrow$ advantage over rival variants.
6. Innovative variant broadens because it increasingly co-occurs with more adj. types.

Challenges

Challenges (Problems) of the present study

- LD scores are (heavily) skewed by frequency
  (frequent items are skewed towards low LD scores)
- Not (yet) analyzed how *really* differs from its direct competitors (*so*, *pretty*, *real*, etc.)
  (we checked what correlates with the use of *really*)
- Observational (correlation) not experimental (causation)
- Limited scope
  (only one phenomenon in one language)

Outlook

Compare *really* directly to other rival variants

- What distinguishes successful variants from non-successful variants? (frequency, length, social profile, etc.)

Could this be a universal mechanism?

- Test if the mechanisms can be shown to have worked in analogous changes in English
  3^{rd} p. sg. ind. morpheme: <eth> → <(e)s>
- Test if the mechanisms can be shown to have worked in analogous changes in languages other than English

Thank you so, really, very much!

Aijmer, K. (2011). Are you totally spy? a new intensifier in present-day american english. In S. Hancil (Ed.), *Marqueurs discursifs et subjectivité*, pp. 155–172. Rouen: Universités de Rouen and Havre.

Aijmer, K. (2018). That's well bad. some new intensifiers in spoken in british english. In V. Brezina, R. Love, and K. Aijmer (Eds.), *Corpus Approaches to Contemporary British English*, pp. 60–95. New York and London: Routledge.

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University press.

Bauer, L. and W. Bauer (2002). Adjective boosters in the english of young new zealanders. *Journal of English Linguistics 30*, 244–257.

Brinton, L. J. and L. K. Arnovick (2006). *The English Language: A Linguistic History*. Oxford: Oxford University Press.

D'Arcy, A. F. (2015). Stability, stasis and change - the longue duree of intensification. *Diachronica 32*(4), 449–493.

Dixon, R. M. W. (1977). Where have all the adjectives gone? *Studies in Language 1*, 19–80.

Evert, S. (2018). V14. lexikalische semantik. https://www.youtube.com/watch?v=kZRBR89Uwhc&t=4645s.

Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*, Volume 124. CRC Press.

Fuchs, R. (2016). Register variation in intensifier usage across asian englishes. In H. Pichler (Ed.), *Discourse-Pragmatic Variation and Change: Insights from English*, pp. 185–213. Cambridge: Cambridge University Press.

Fuchs, R. (2017). Do women (still) use more intensifiers than men? *International Journal of Corpus Linguistics 22*(3), 345–374.

Ito, R. and S. Tagliamonte (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in english intensifiers. *Language in Society 32*, 257–279.

Jockers, M. L. (2017). Syuzhet 1.0.4 now on cran. http://www.matthewjockers.net/2017/12/16/syuzhet-1-0-4/.

Kursa, M. B., W. R. Rudnicki, et al. (2010). Feature selection with the boruta package. *Journal of Statistical Software 36*(11), 1–13.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins Publishing Company.

Lorenz, G. (2002). Really worthwhile or not really significant? a corpus-based approach to the delexicalization and grammaticalization of intensifiers in modern english. In I. Wischer and G. Diewald (Eds.), *New Reflections on Grammaticalization*, pp. 143–161. Amsterdam: John Benjamins.

Macaulay, R. (2006). Pure grammaticalization: The development of a teenage intensifier. *Language Variation and Change 18*, 267–283.

Mair, C. (2004). Corpus linguistics and grammaticalisation theory. statistics, frequencies and beyond. In H. Lindquist and C. Mair (Eds.), *Corpus approaches to grammaticalization in English*, pp. 121–150. Amsterdam and Philadelphia: John Benjamins.

Méndez-Naya, B. (2003). On intensifiers and grammaticalization: The case of swithe. *English Studies 84*, 372–391.

Núñez Pertejo, P. and I. Palacios (2014). That's absolutely crap, totally rubbish. the use of intensifiers absolutely and totally in the spoken language of british adults and teenagers. *Functions of Language 21*(2), 210–237.

Palacios, I. and P. Núñez Pertejo (2012). He's absolutely massive. it's a super day. madonna, she is a wicked singer. youth language and intensification: A corpus-based study. *Text and Talk 32*(6), 773–796.

Peters, P. (fc.). The australian component of the international corpus of english (ice-aus).

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985). *A Comprehensive Grammar of the English Language*. London & New York: Longman.

Schweinberger, M. (2017). Using intensifier-adjective bi-grams to investigate mechanisms of change. Paper presented at ICAME38. Prague, 27/5/2017.

Schweinberger, M. (2020a). Analyzing change in the american english amplifier system in the fiction genre. In P. Rautionaho, A. Nurmi, and J. Klemola (Eds.), *Corpora and the Changing Society. Studies in the evolution of English, Part II*, pp. 223–249. John Benjamins.

Schweinberger, M. (2020b). A corpus-based analysis of adjective amplification among native speakers and learners of english. *International Journal of Learner Corpus Research 6*(2), 163–192.

Schweinberger, M. (2020c). How learner corpus-research can inform language learning and teaching: An analysis of adjective amplification among l1 and l2 english speakers. *Australian Review of Applied Linguistics 43*(2), 195–217.

Schweinberger, M. (2020d). Using intensifier-adjective collocations to determine mechanisms of change. In A. Cermakova and M. Mala (Eds.), *Variation in Time and Space: Observing the World through Corpora*, pp. 231–256. De Gruyter.

Schweinberger, M. (2020e). Using semantic vector space models to investigate lexical replacement – a corpus based study of ongoing changes in intensifier systems. In Y. Asahi (Ed.), *Proceedings of Methods XVI. Papers from the sixteenth international conference on Methods in Dialectology, 2017*, pp. 241–249. Peter Lang.

Schweinberger, M. (2021a). On historical developments in the irish english intensifier system. *Anglistik 32*(1), 139–158.

Schweinberger, M. (2021b). On the waning of forms – a corpus-based analysis of decline and loss in adjective amplification. In S. Kranich and T. Breban (Eds.), *Lost in change: Causes and processes in the loss of grammatical constructions and categories*, pp. ??–?? John Benjamins.

Schweinberger, M. (fca). Absolutely fantastic and really, really good – language variation and change in irish english. In C. Amador-Moreno and S. Lucek (Eds.), *Expanding the Landscapes of Irish English: Research in Honour of Jeffrey Kallen*, pp. ??–?? Taylor & Francis.

Schweinberger, M. (fcb). Ongoing change in the australian english amplifier system. *Australian Journal of Linguistics ?*(?), ??–??

Tagliamonte, S. (2005). So who? like how? just what?: Discourse markers in the conversations of young canadians. *Journal of Pragmatics 37*(11), 1896–1915.

Tagliamonte, S. (2006). "so cool, right?": Canadian english entering the 21st century. *The Canadian Journal of Linguistics/La revue canadienne de linguistique 51*(2), 309–331.

Tagliamonte, S. (2008). So different and pretty cool! recycling intensifiers in toronto, canada. *English Language and Linguistics 12*(2), 361–394.

Tagliamonte, S. and C. Roberts (2005). So weird; so cool; so innovative: The use of intensifiers in the television series friends. *American Speech 80*(3), 280–300.

Tagliamonte, S. A. and D. Denis (2014). Expanding the transmission/diffusion dichotomy: Evidence from canada. *Language 90*(1), 90–136.

Wagner, S. (2017a). Amplifier-adjective 2-grams world-wide: focus on pretty. Paper presneted at ICAME 37. Charles University Prague, 27/5/2017.

Wagner, S. (2017b). Totally new and pretty awesome: Amplifier–adjective bigrams in globwe. *Lingua 200*, 63–83.

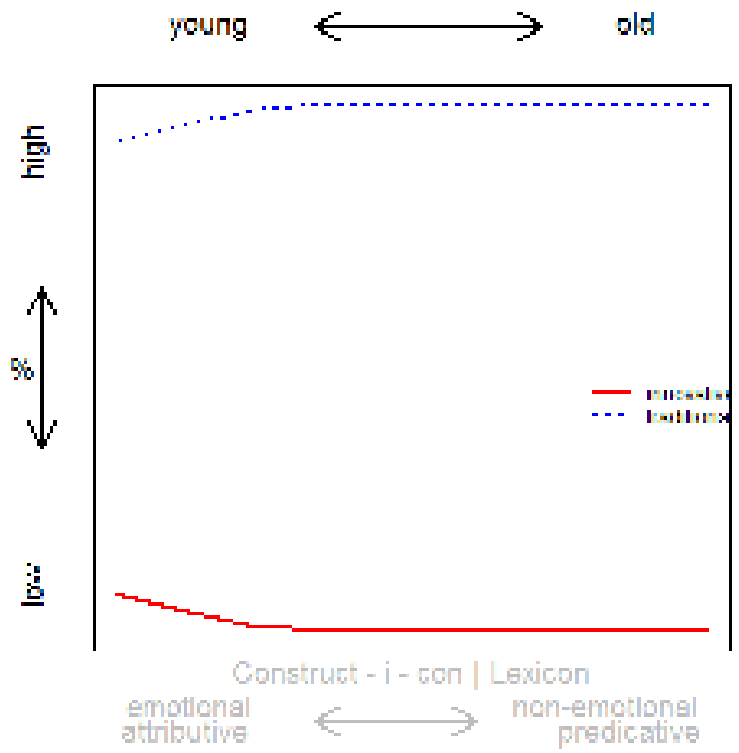
Diffusion of Innovations

young ⟷ old

upper

⟷

working



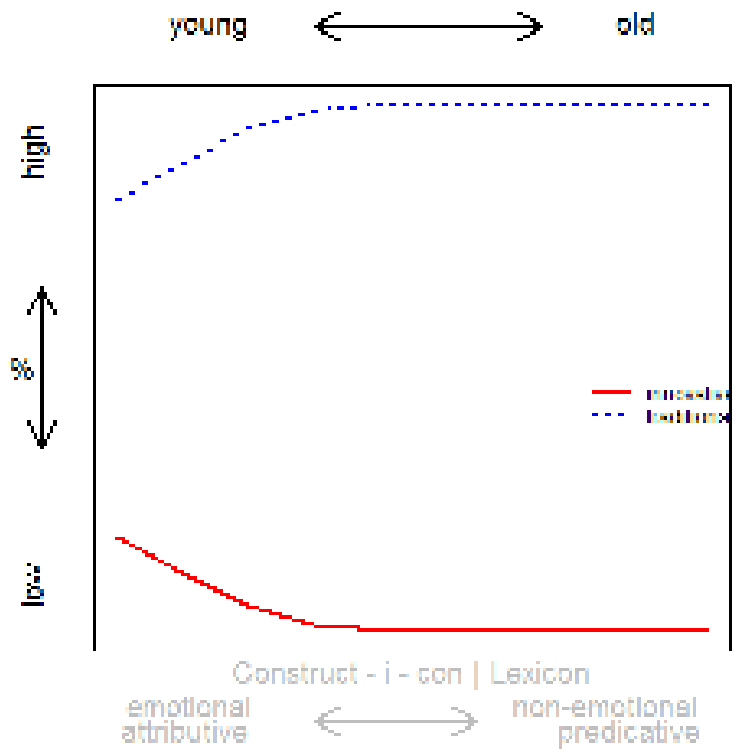
Diffusion of Innovations

young ⟷ old

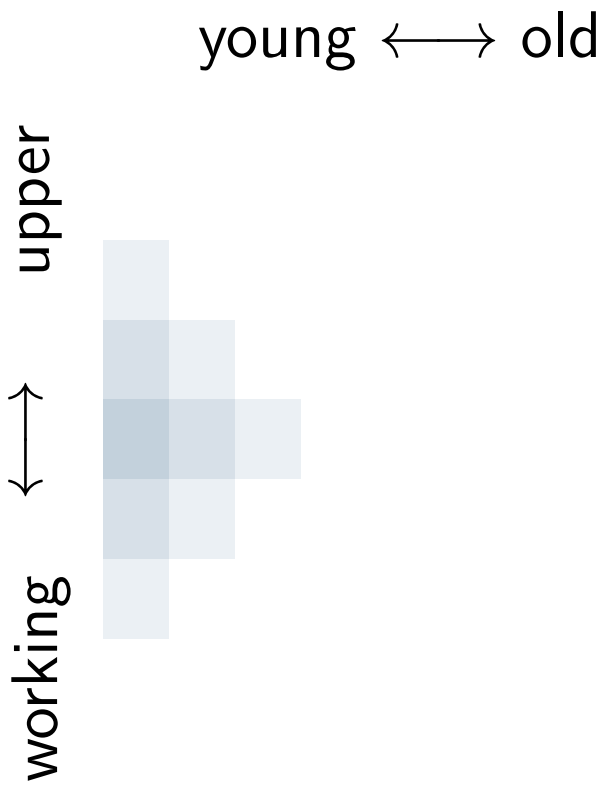
upper

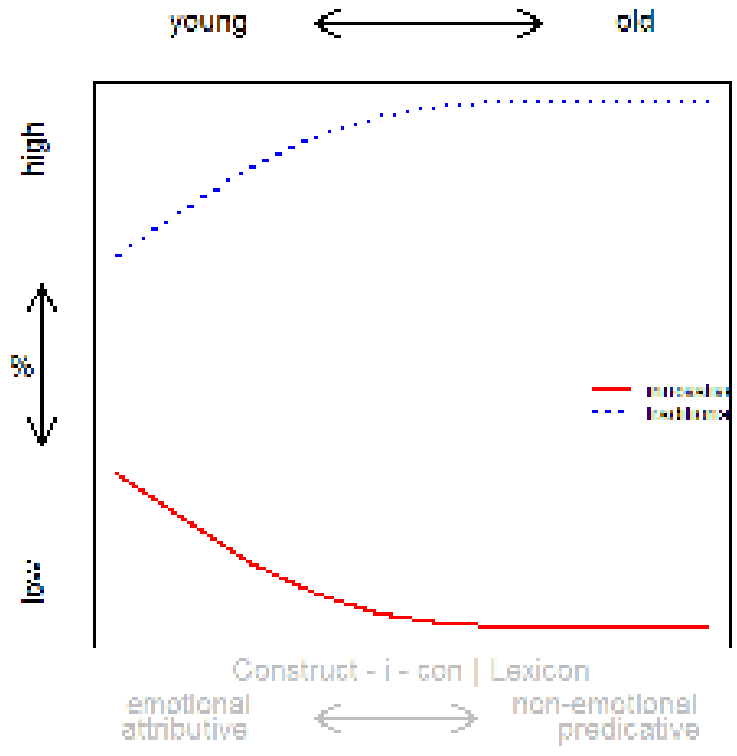
⟷

working

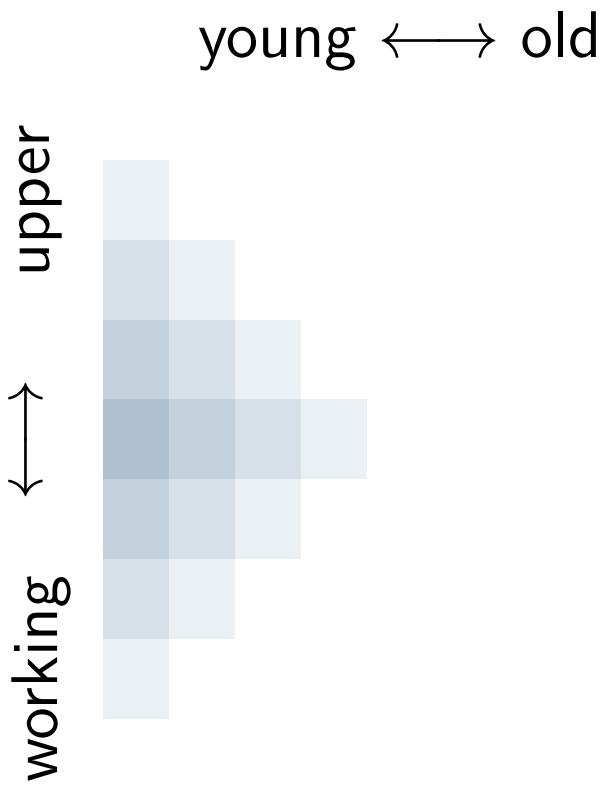


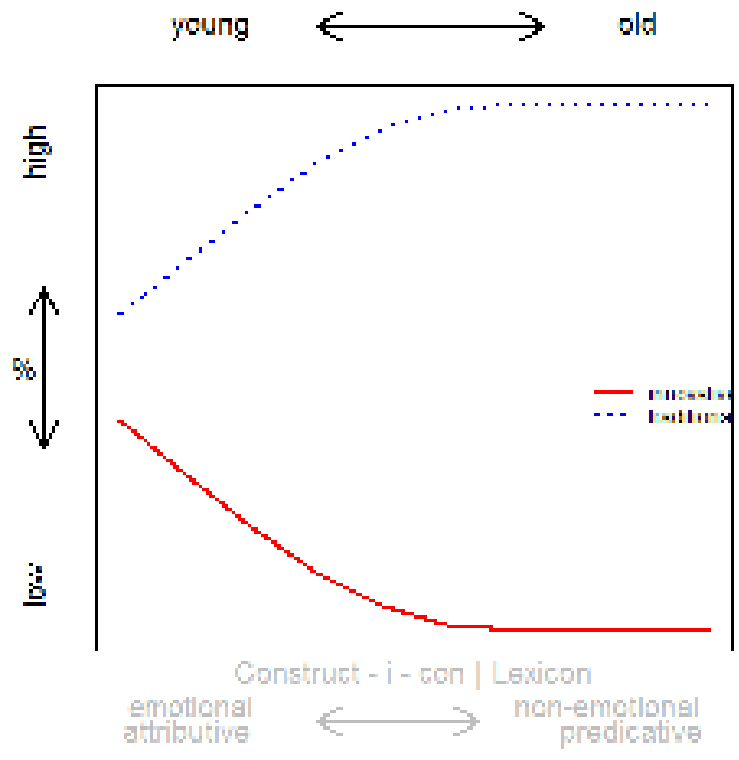
Diffusion of Innovations



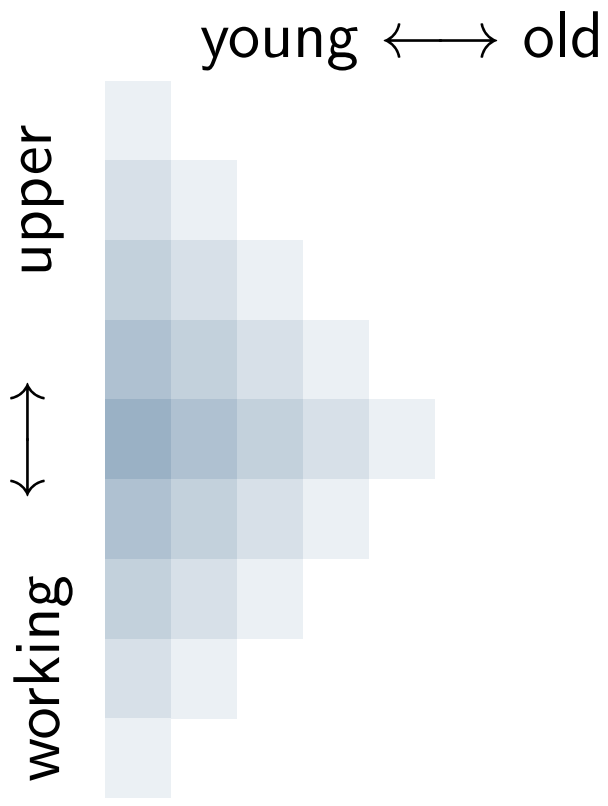


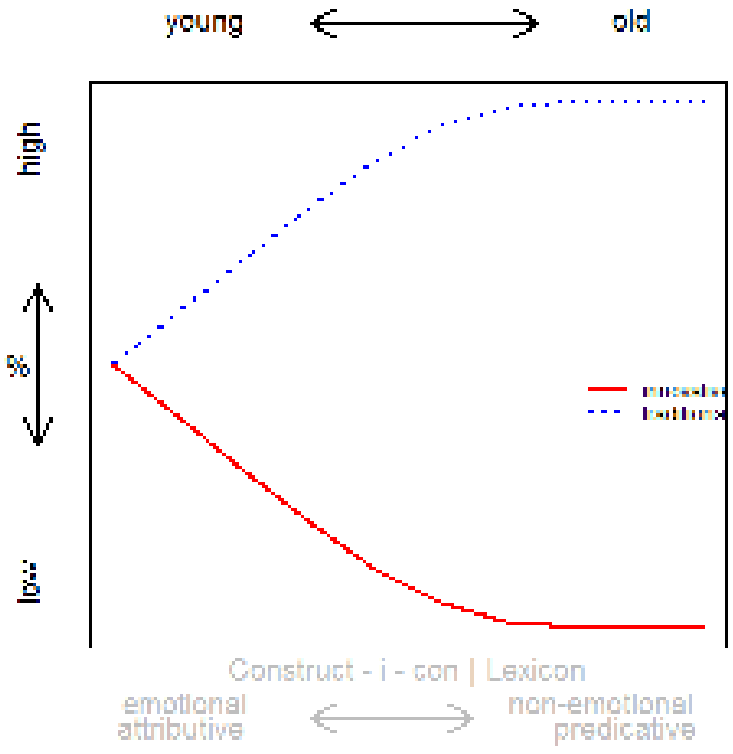
Diffusion of Innovations



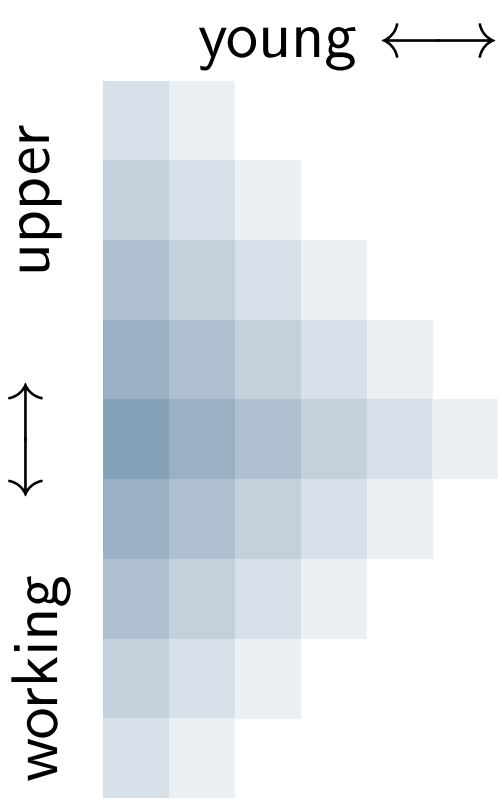


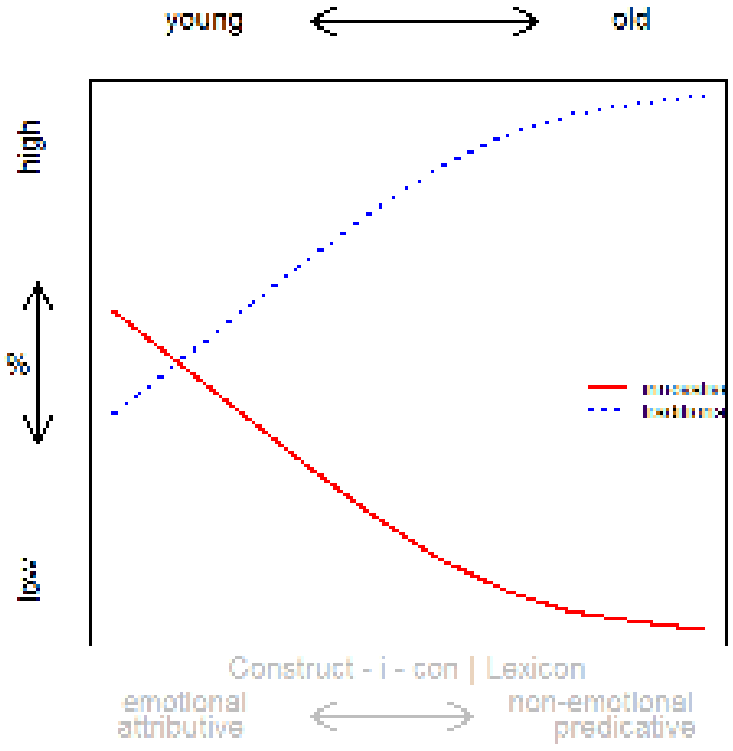
Diffusion of Innovations



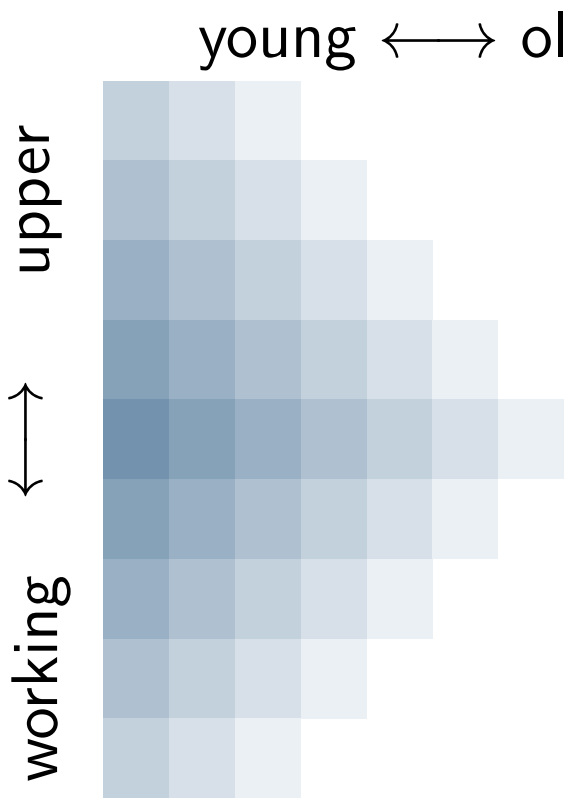


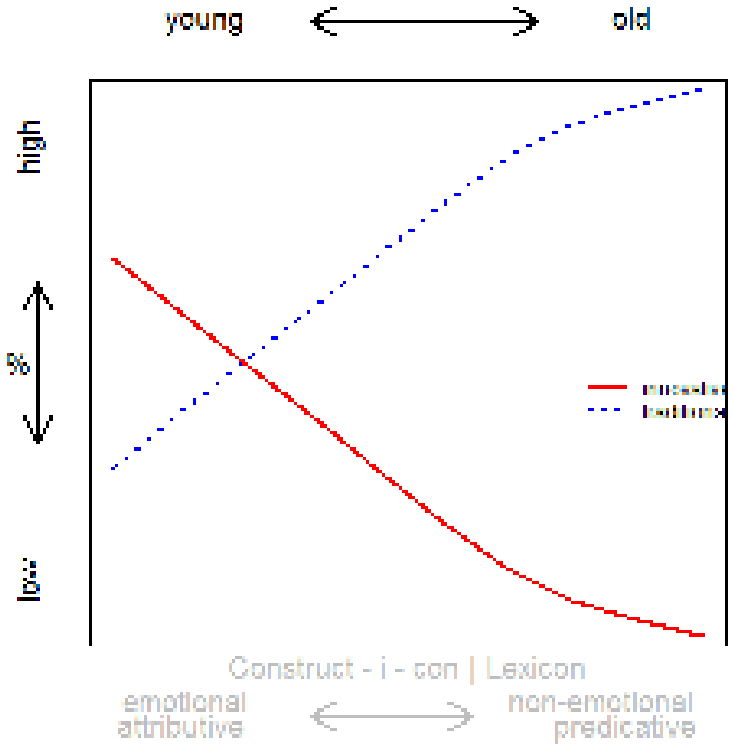
Diffusion of Innovations



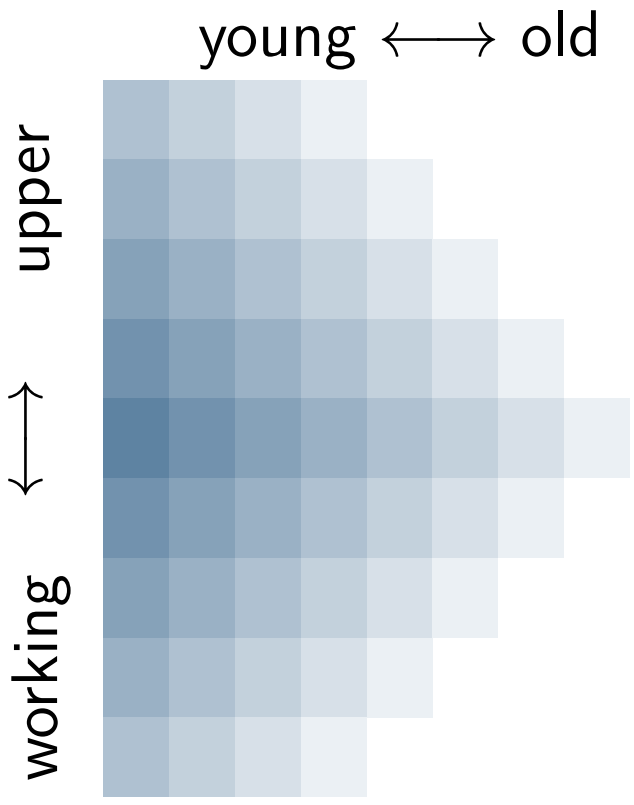


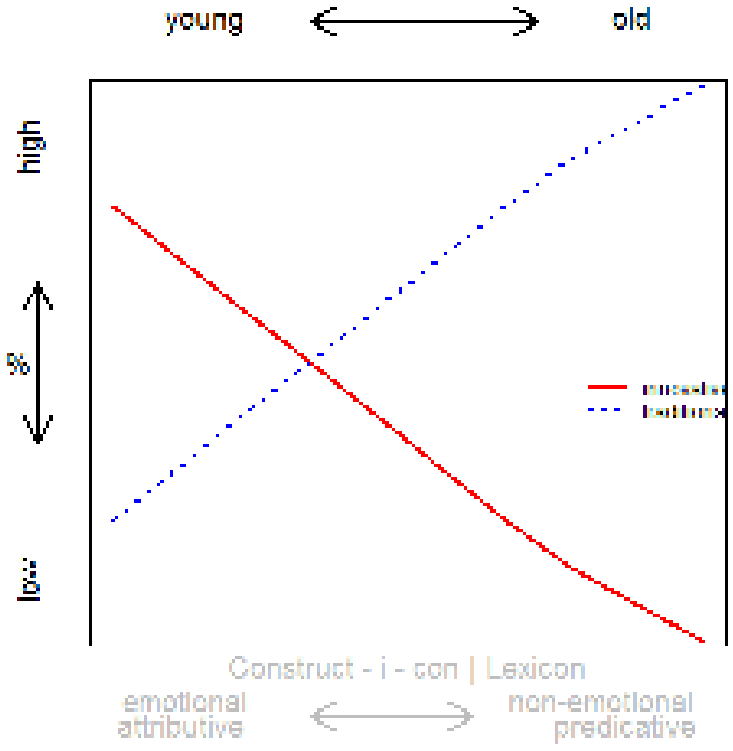
Diffusion of Innovations



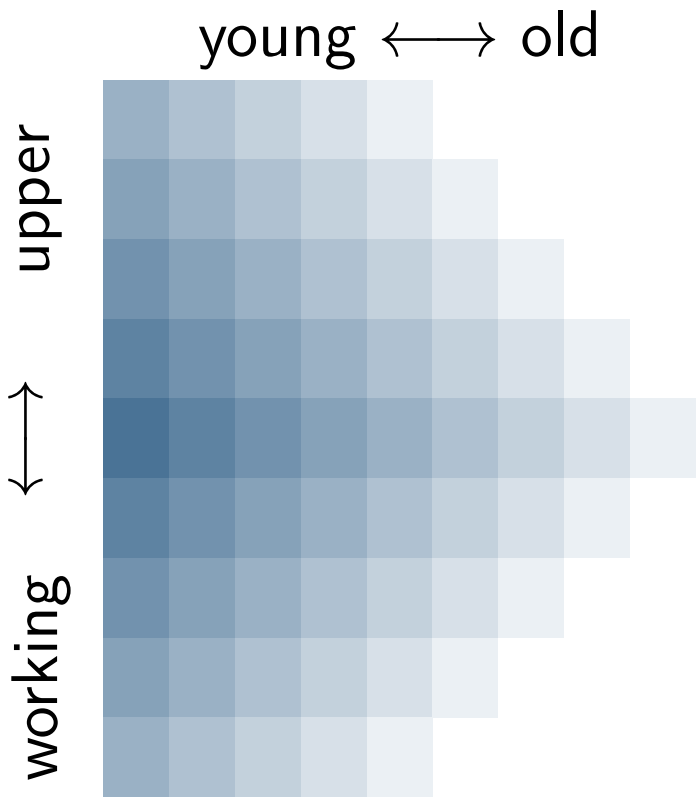


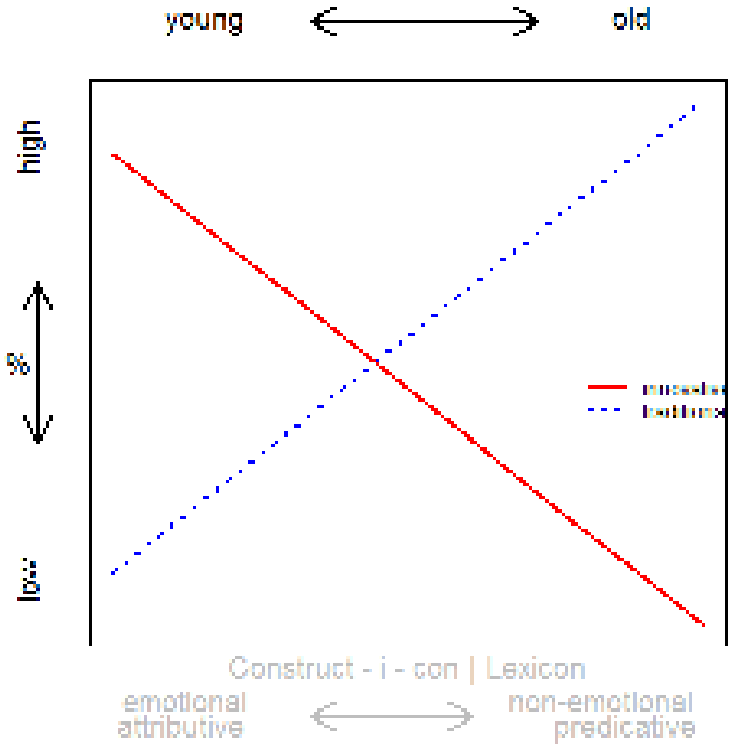
Diffusion of Innovations



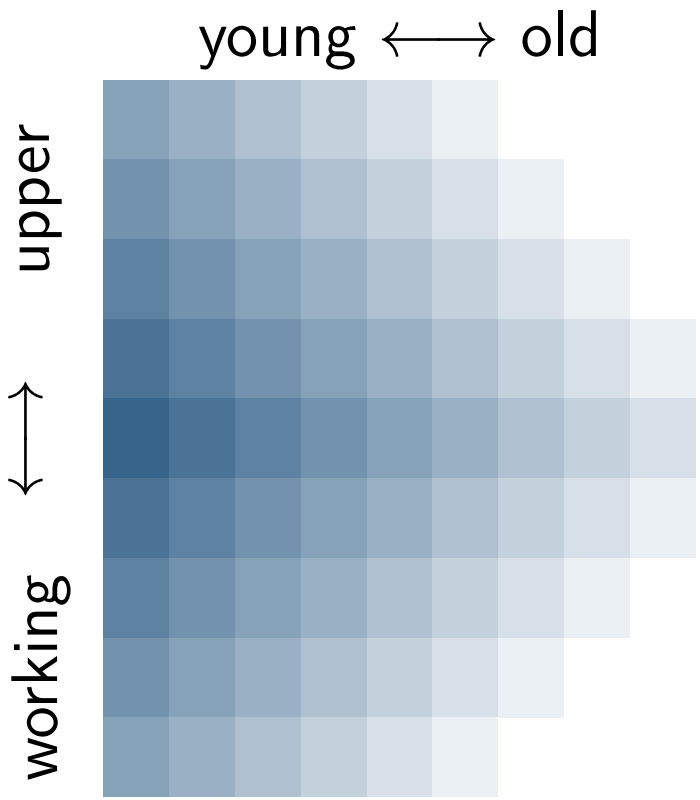


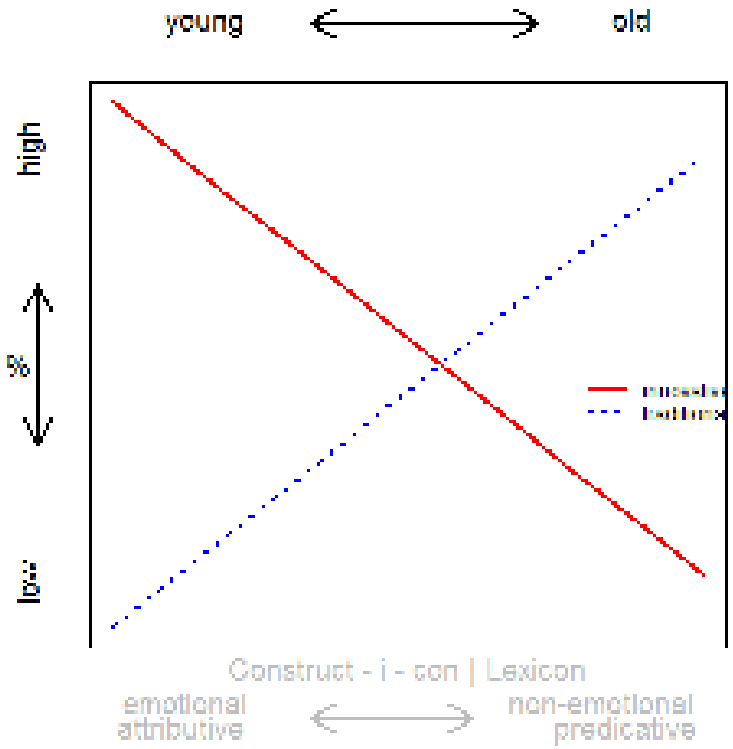
Diffusion of Innovations



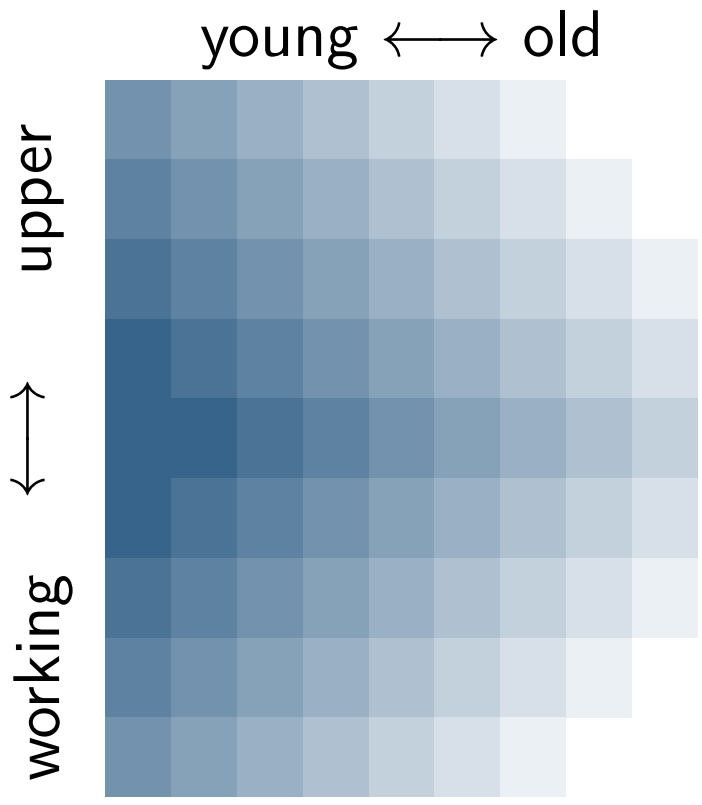


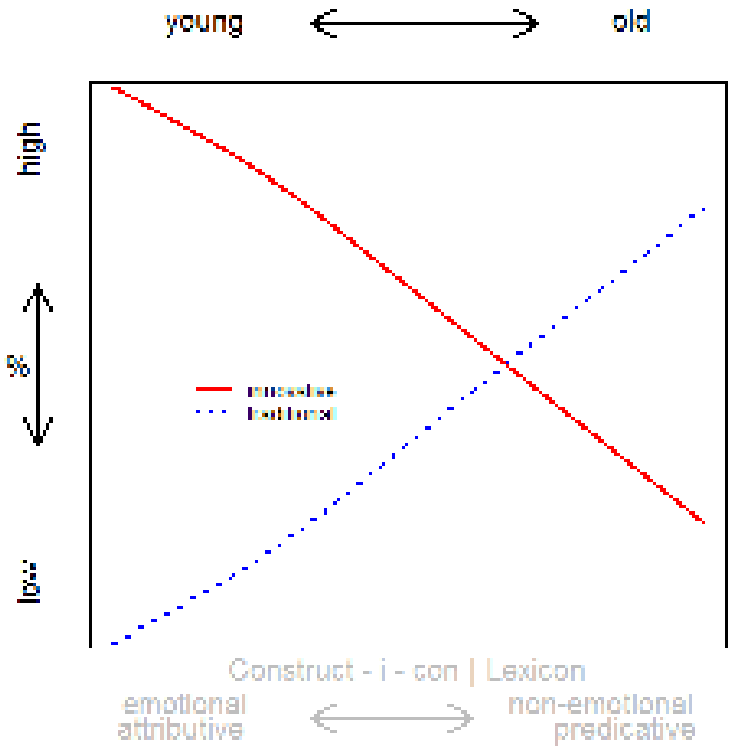
Diffusion of Innovations



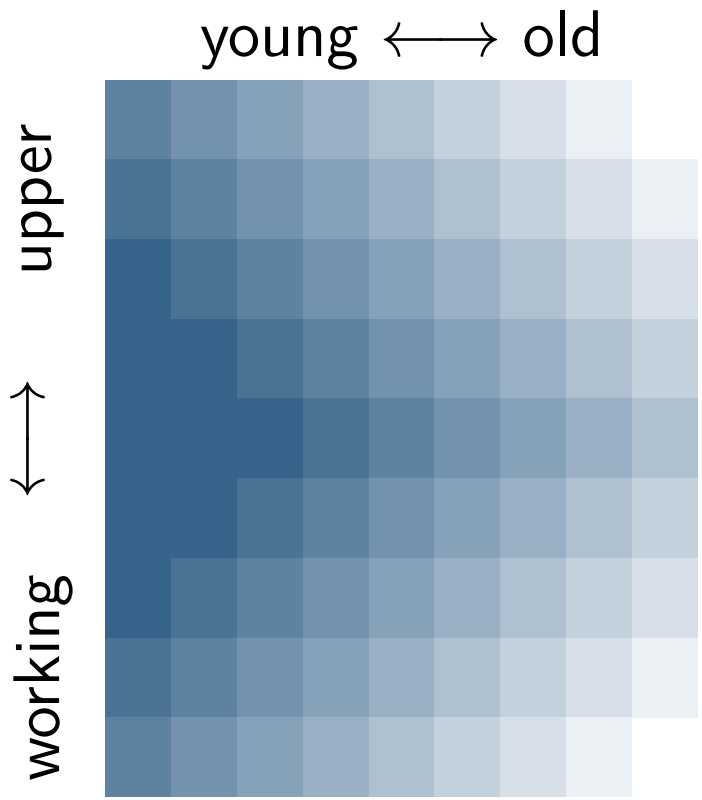


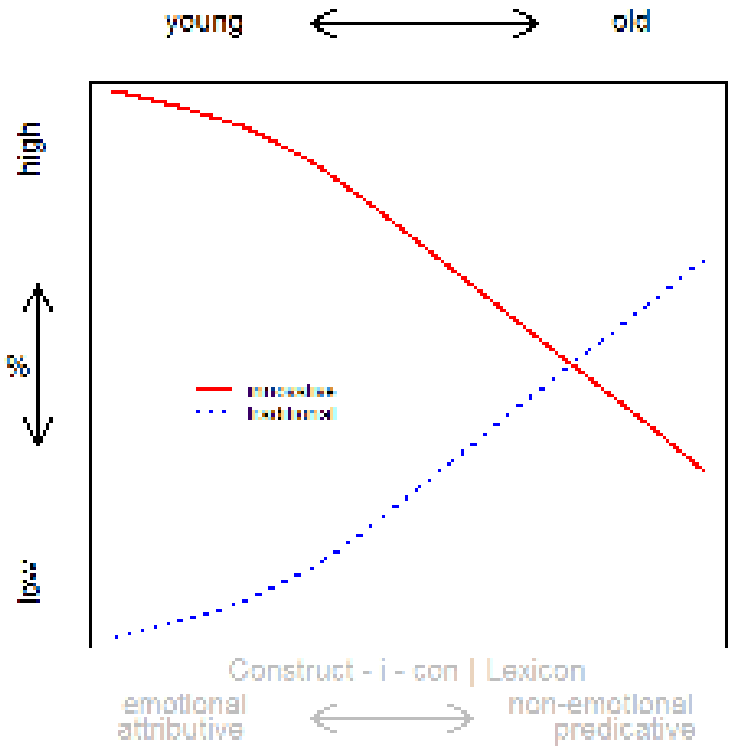
Diffusion of Innovations



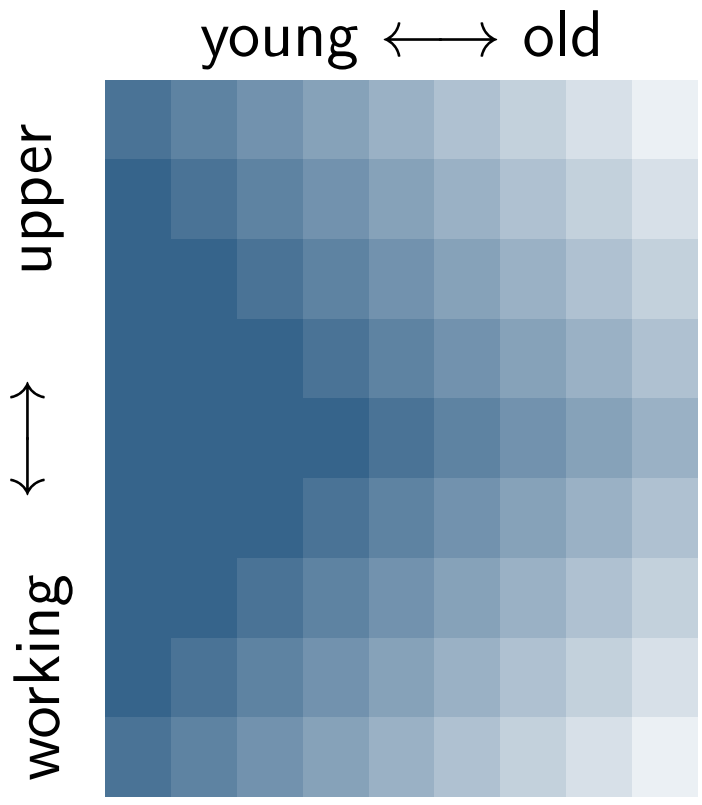


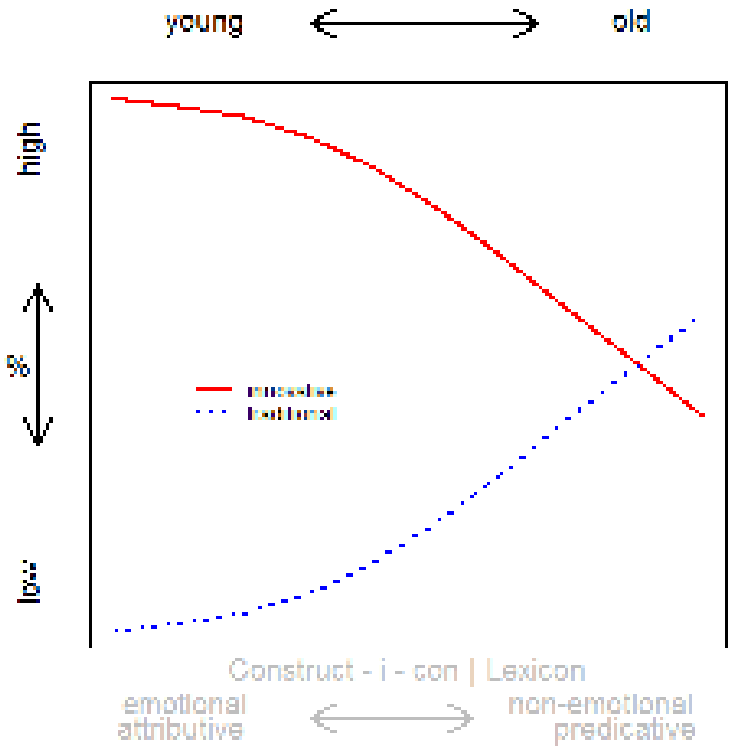
Diffusion of Innovations



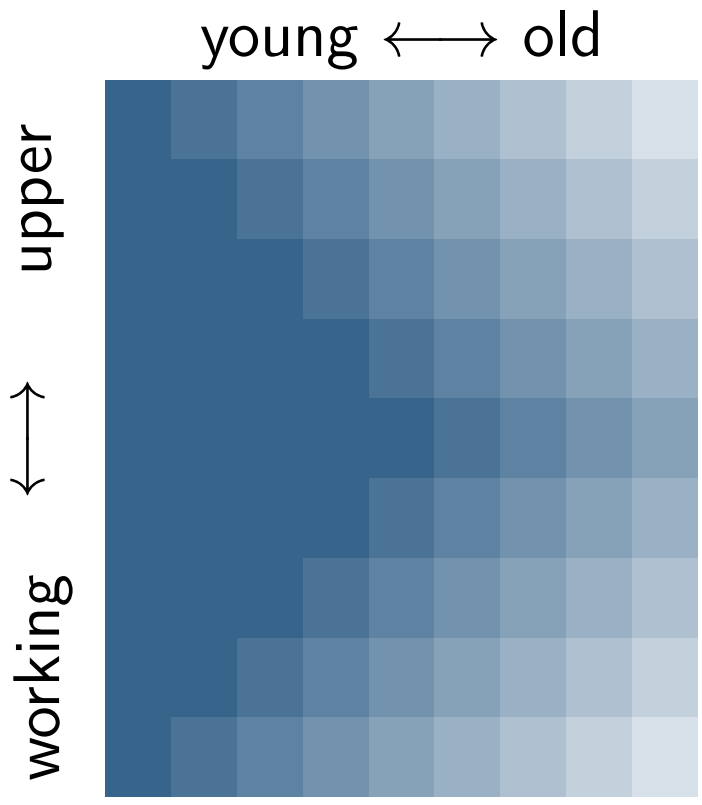


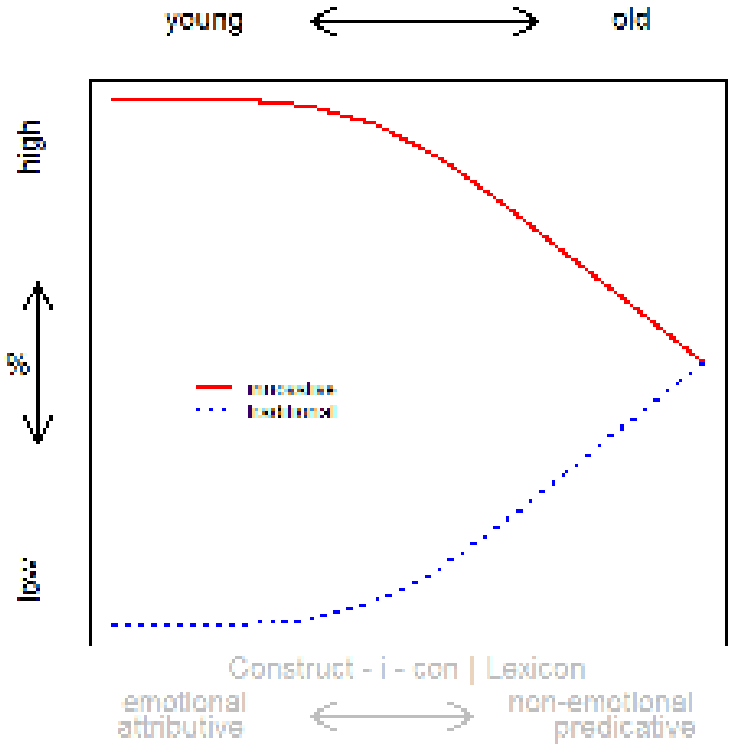
Diffusion of Innovations



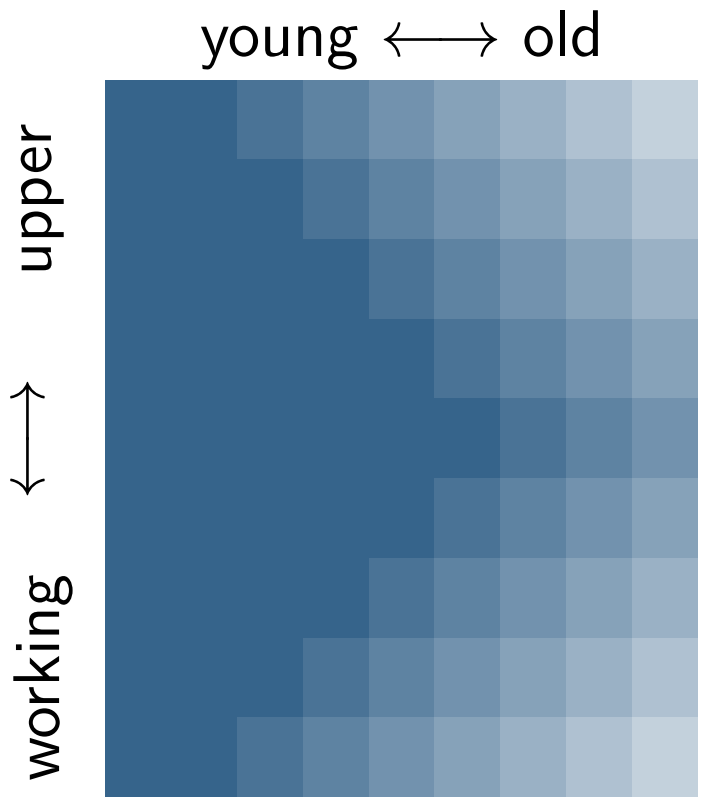


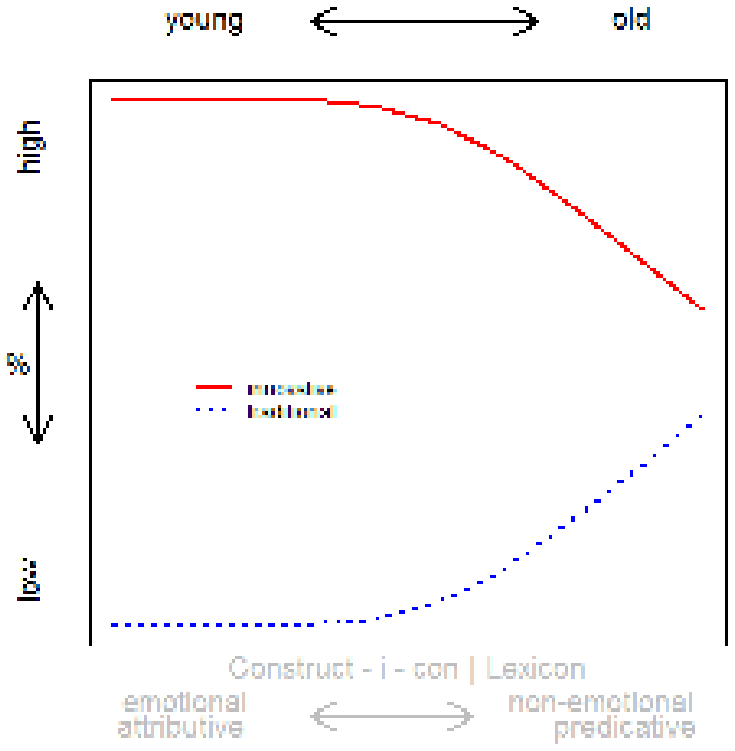
Diffusion of Innovations



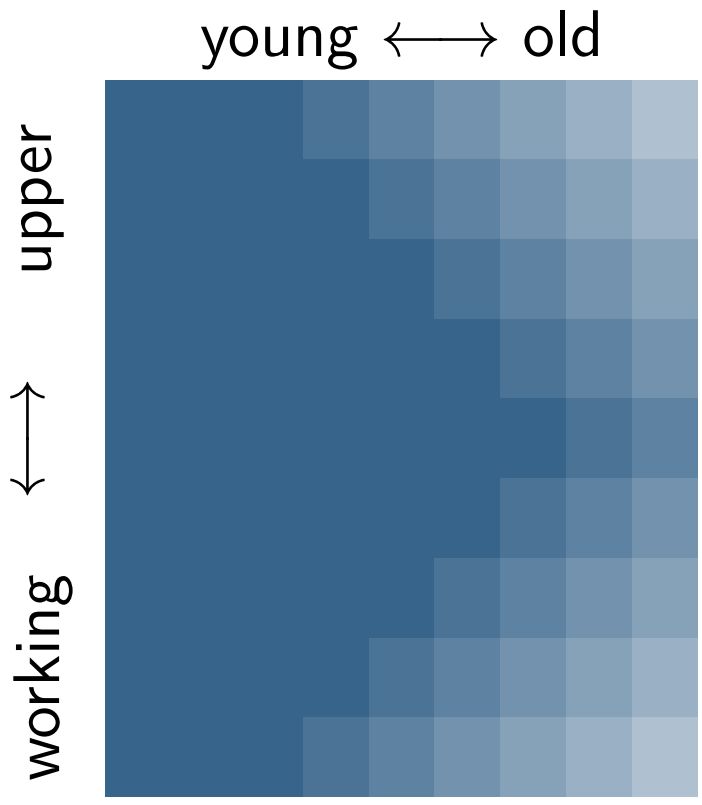


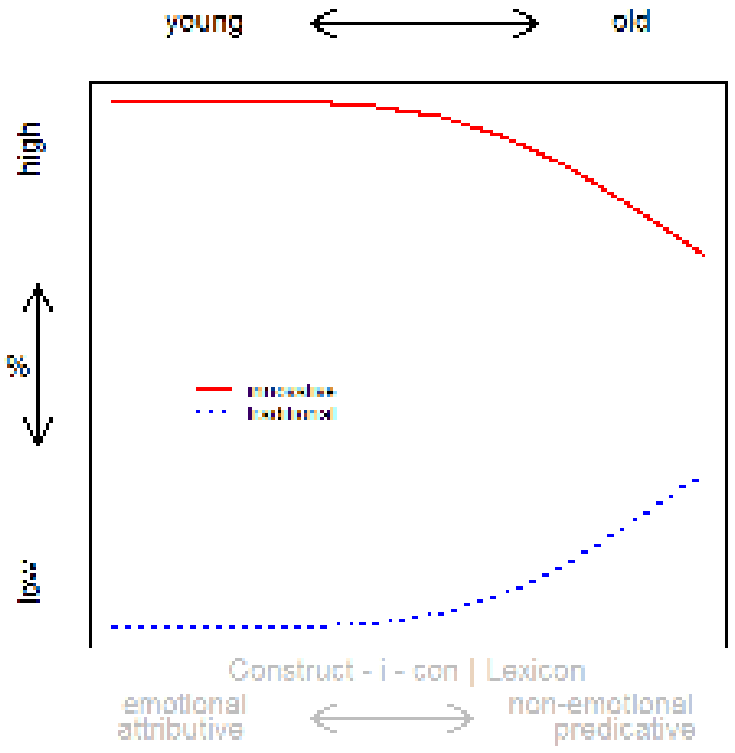
Diffusion of Innovations



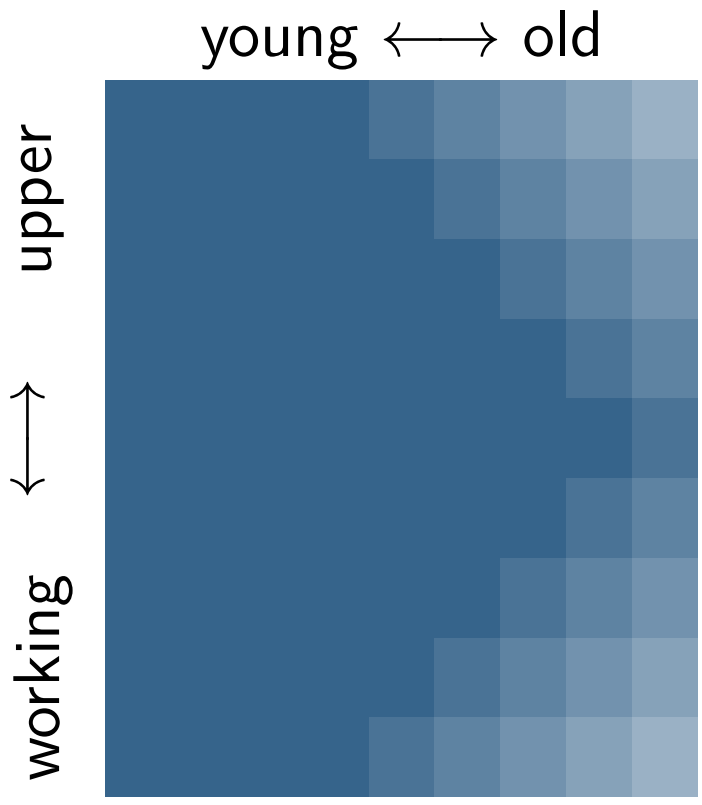


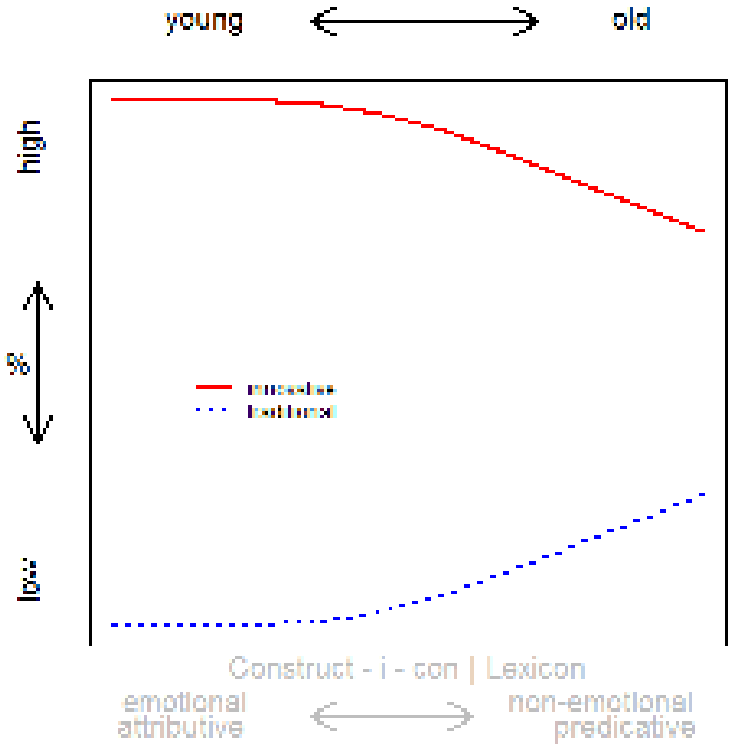
Diffusion of Innovations



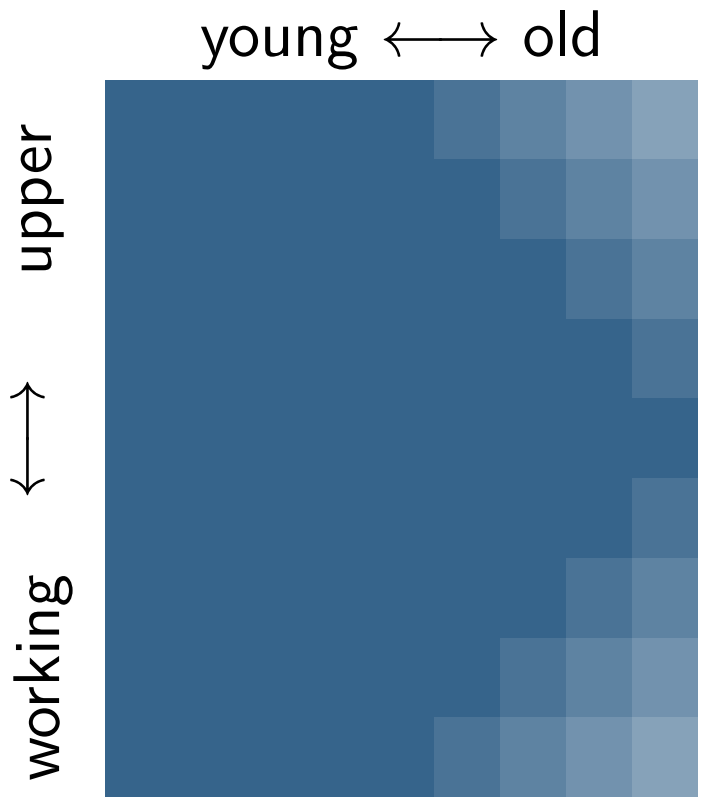


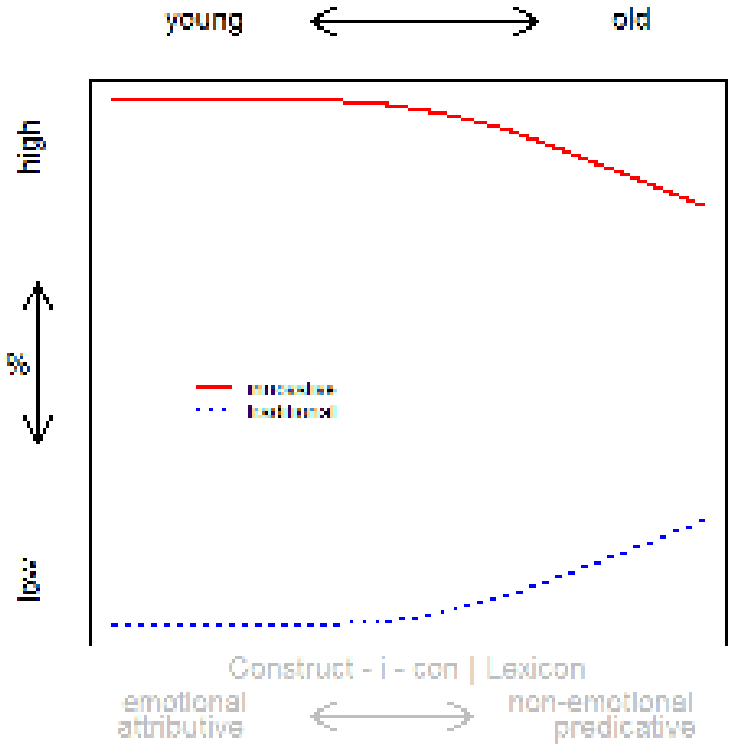
Diffusion of Innovations



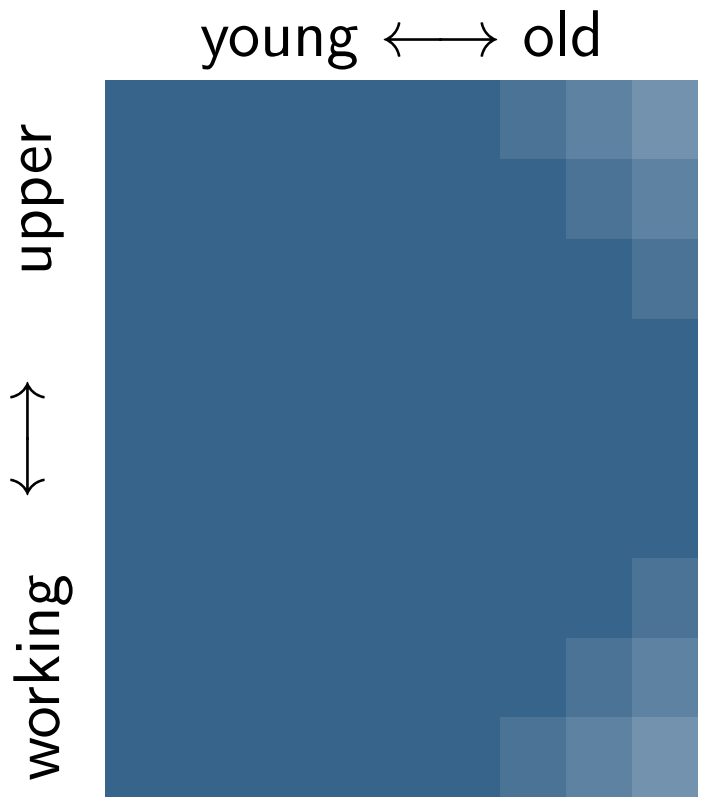


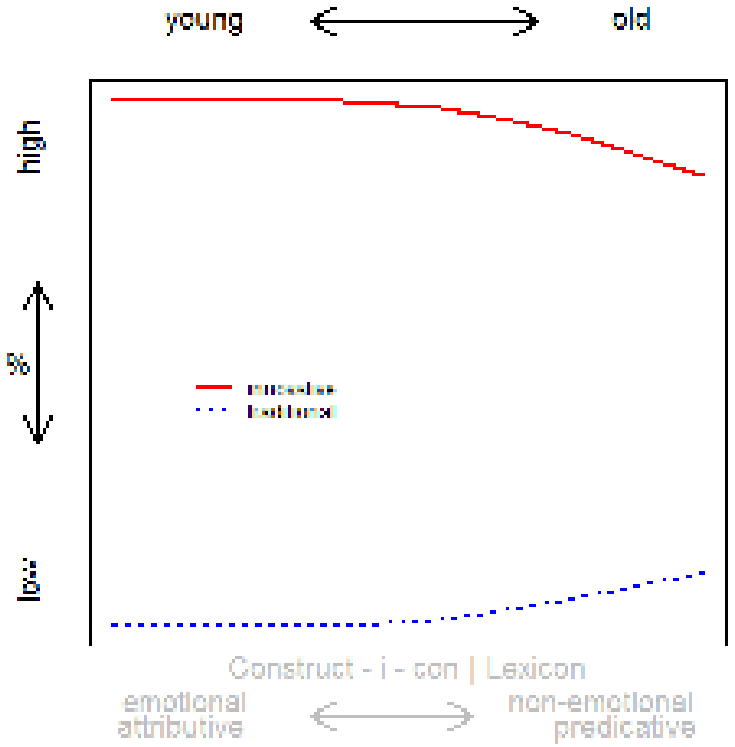
Diffusion of Innovations



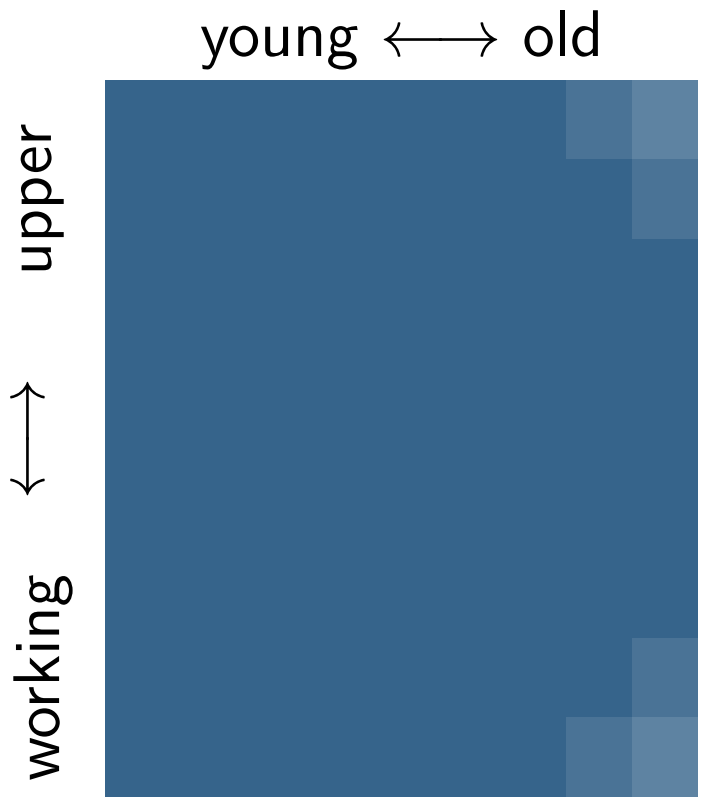


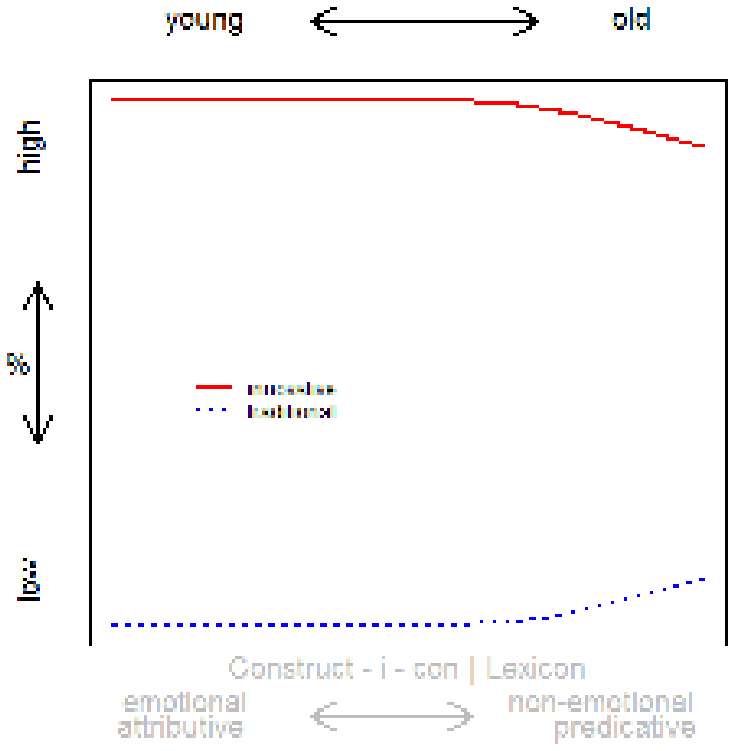
Diffusion of Innovations



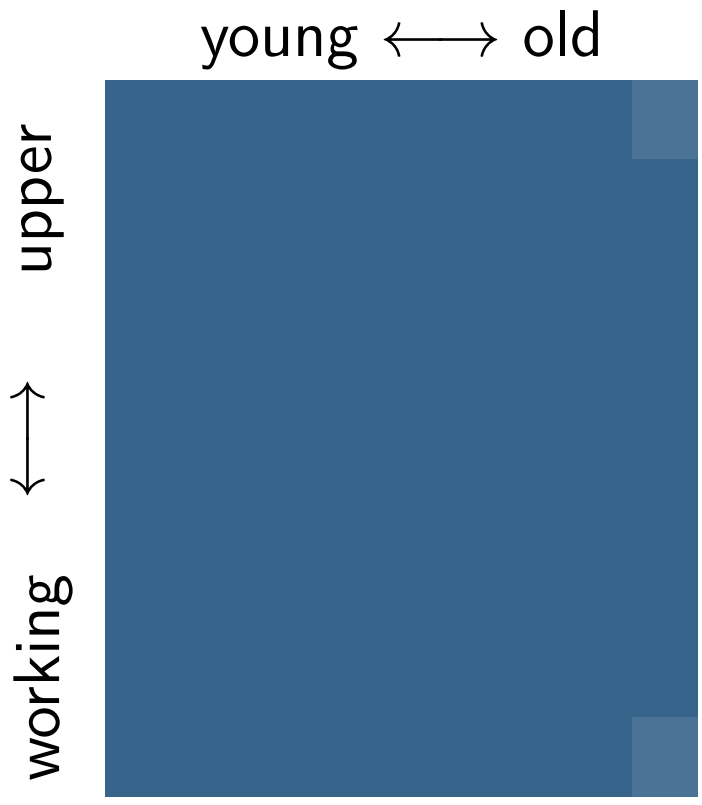


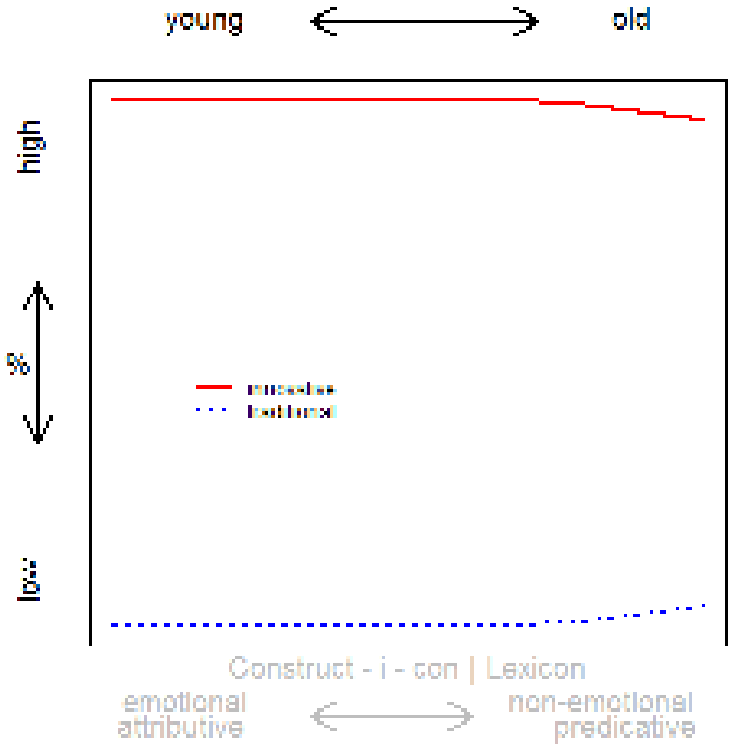
Diffusion of Innovations



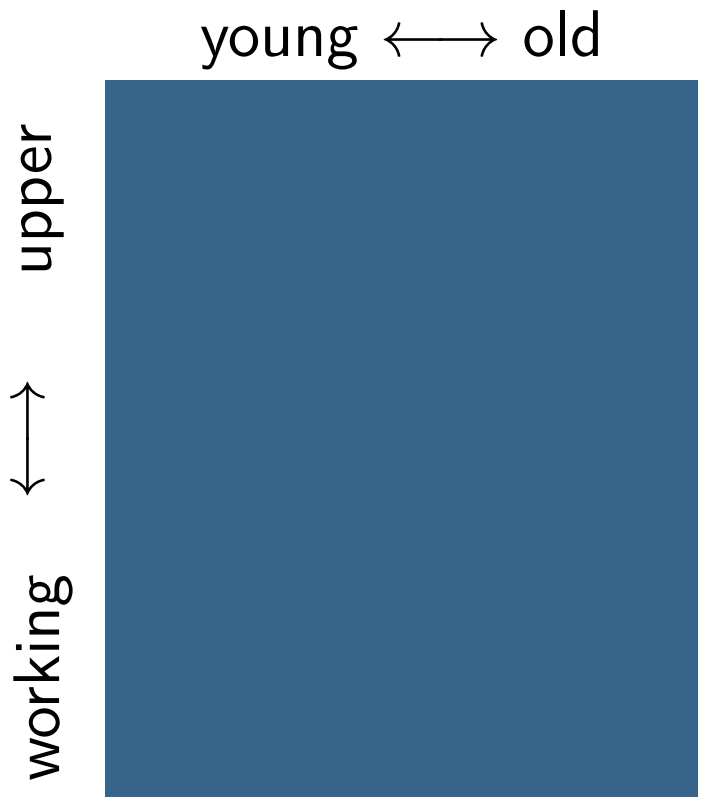


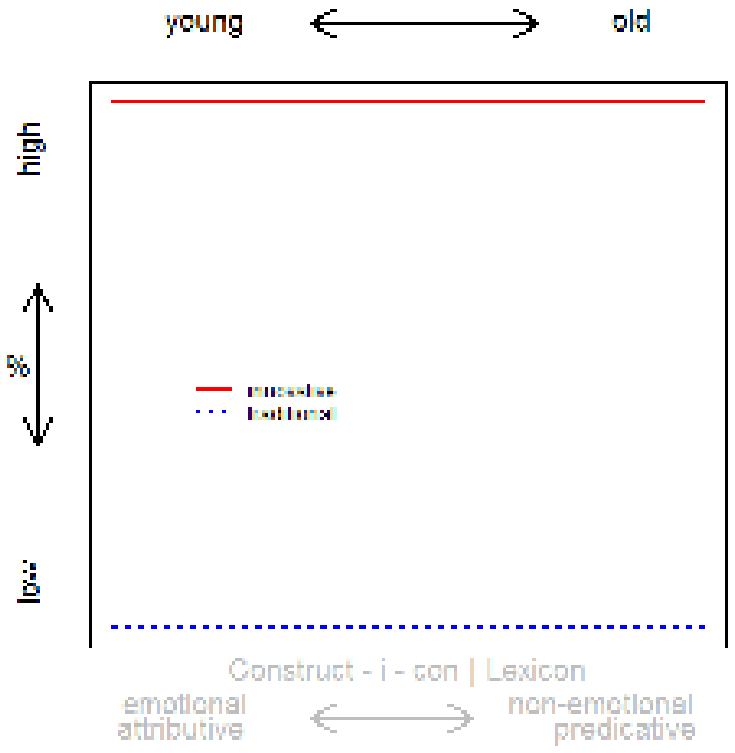
Diffusion of Innovations





Diffusion of Innovations





Adjective amplification in English - determining factors of success in language change

Dr. Martin Schweinberger

slides available at

www.martinschweinberger.de

R code upon request