On the waning of forms - a corpus-based analysis of losers in language change

Martin Schweinberger

slides available at

www.martinschweinberger.de

R code at

https://github.com/MartinSchweinberger/isle6_verynze/tree/master

THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

CREATE CHANGE

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)

This talk is an elaboration of Schweinberger (2021b).

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
- recently amplifier-adjective bigrams have come more into focus (e.g. Schweinberger 2017; Wagner 2017a,b)
- focus on incoming variants rather than receeding forms
  (notable exceptions are D'Arcy 2015; Schweinberger 2021b; Tagliamonte 2008)

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)

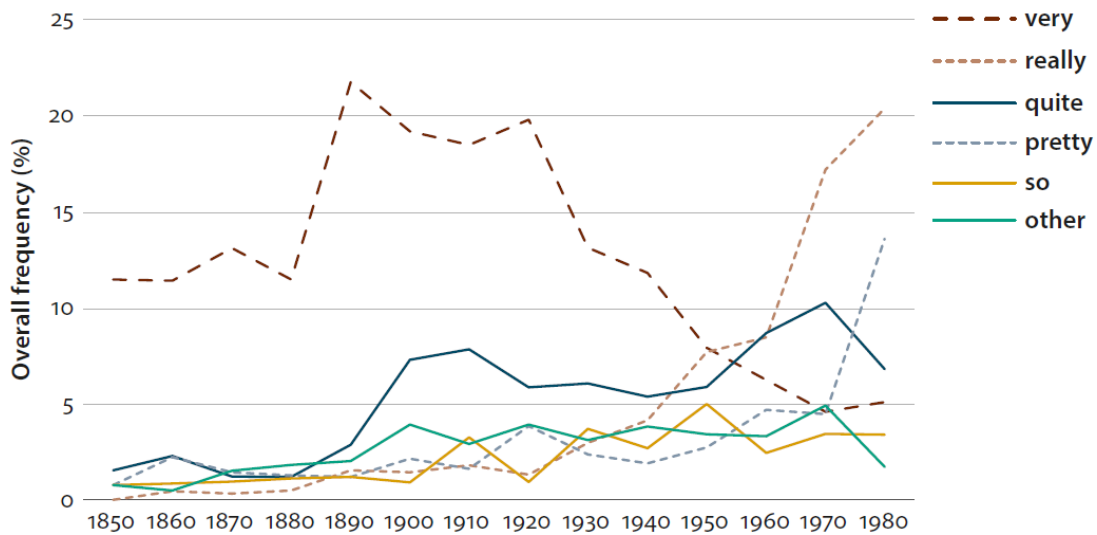


Figure 1: Adapted from D'Arcy (2015: 468)

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)

- *really* (D'Arcy 2015: 481)
  Correlated with speaker age, syntactic function, and gender among speakers born between 1932 and 1980.
  $\rightarrow$ increase accompanied by intra- and extra-linguistic stratification (ordered heterogeneity) (Weinreich et al. 1968: 100)
- *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.

Research Question(s)

Does the retreat of *very* mirror the trajectory of *really*?

Does loss also proceed in a systematic and ordered manner?

Data and Methodology

Wellington Corpus of Spoken New Zealand English (WSC)

One-million-word corpus of transcribed English compiled between 1988 and 1994 (Holmes et al. 1998)

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

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

Data Processing

Data processing in RStudio (Allaire 2012)

- WSC (Holmes et al. 1998)
- Part–of-speech tagged (Hornik 2019)
- Retrieved adjectives (PoS–tag JJ)
- Determined whether adjective were preceded by an amplifier (member of a predefined set of amplifiers)
- Auotomated spell-checking (Ooms 2020)
- Checked pos-tagging (Straka and Straková 2017)
- Implemented a Sentiment Analysis of adjective types (emotional vs non-emotional) (Silge and Robinson 2016)

Data Processing

- Calculated logged and scaled frequency of adjectives per age group
- Determined if the same amplifier type had occurred within a span of three adjective slots previously ($\rightarrow$ priming)
- Added gradability scores (likelihood of occurring in comparative contexts)
- Semantic classification of adjectives (Dixon 1977; D'Arcy 2015; Tagliamonte and Roberts 2005; Tagliamonte 2006, 2008)

Data Processing

- Removed...
  - negated adjectives
  - comparative and superlative forms
  - adjectives that were never amplified
  - adjectives that were preceded by downtoners
  - strange forms (e.g. *much*)
- Manual cross–evaluation of automated classification
- Metadata ((Text-)Type and speaker information(age, sex)

Variable Coding

Dependent Variable(s)				
very	nominal	yes/no occurrence of pre-adjectival *very*		
Independent Variable(s)				
Age	categorical	16-29 | 30-39 | 40+	extra	linguistic
Ethnicity	categorical	Maori | Other | Pakeha	extra	linguistic
Type	categorical	Formal | Private | Public	extra	linguistic
Gender	nominal	Woman | Man	extra	linguistic
Priming	nominal	Primed | NotPrimed	extra	linguistic
L1	nominal	English | Other	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	420 adj. types	intra	linguistic
Frequemcy	numeric	logged + scaled frequency of adj. by age	intra	linguistic

Statistical Analysis

Conditional Inference Tree (Hothorn et al. 2015)

- to check which variable levels can be collapsed

Boruta (Kursa et al. 2010)

- to check which variables to include in the regression modeling

Mixed-effect binomial logistic regression (Bates et al. 2014)

- to inspect the size and direction of effects

Results for very

Data Overview

Amplification	**N**	**%**	**Variants** (%)
∅ Amplification	15,229	86.66	
really	851	4.84	36.31
very	647	3.68	27.60
so	277	1.58	11.82
pretty	241	1.37	10.28
real	59	0.34	2.52
totally	29	0.17	1.24
absolutely	28	0.16	1.19
bloody	22	0.13	0.94
completely	19	0.11	0.81
extremely	18	0.10	0.77
particularly	14	0.08	0.60
fucking	13	0.07	0.55
incredibly	12	0.07	0.51
others (<10)	114	0.65	4.86
Total	17,573 (2,344)	100 (13.3)	100

Table 1: Overview of amplifier frequencies and percentages in the final data set.

Data Overview

Age	**Sex**	**Speakers** (N)	**Adj.** (N)	***very*** (N)	***very*** (%)
16-19	Man	39	116	22	18.97
16-19	Woman	50	224	14	6.25
20-29	Man	92	359	73	20.33
20-29	Woman	152	589	57	9.68
30-39	Man	50	123	38	30.89
30-39	Woman	54	197	57	28.93
40-49	Man	56	169	83	49.11
40-49	Woman	74	245	112	45.71
50-59	Man	35	104	61	58.65
50-59	Woman	33	104	68	65.38
60+	Man	16	58	37	63.79
60+	Woman	18	56	25	44.64
Total		669	2,344	647	36.86

Table 2: Overview of adjective and *really* frequencies and percentages by age and gender in the final data set.

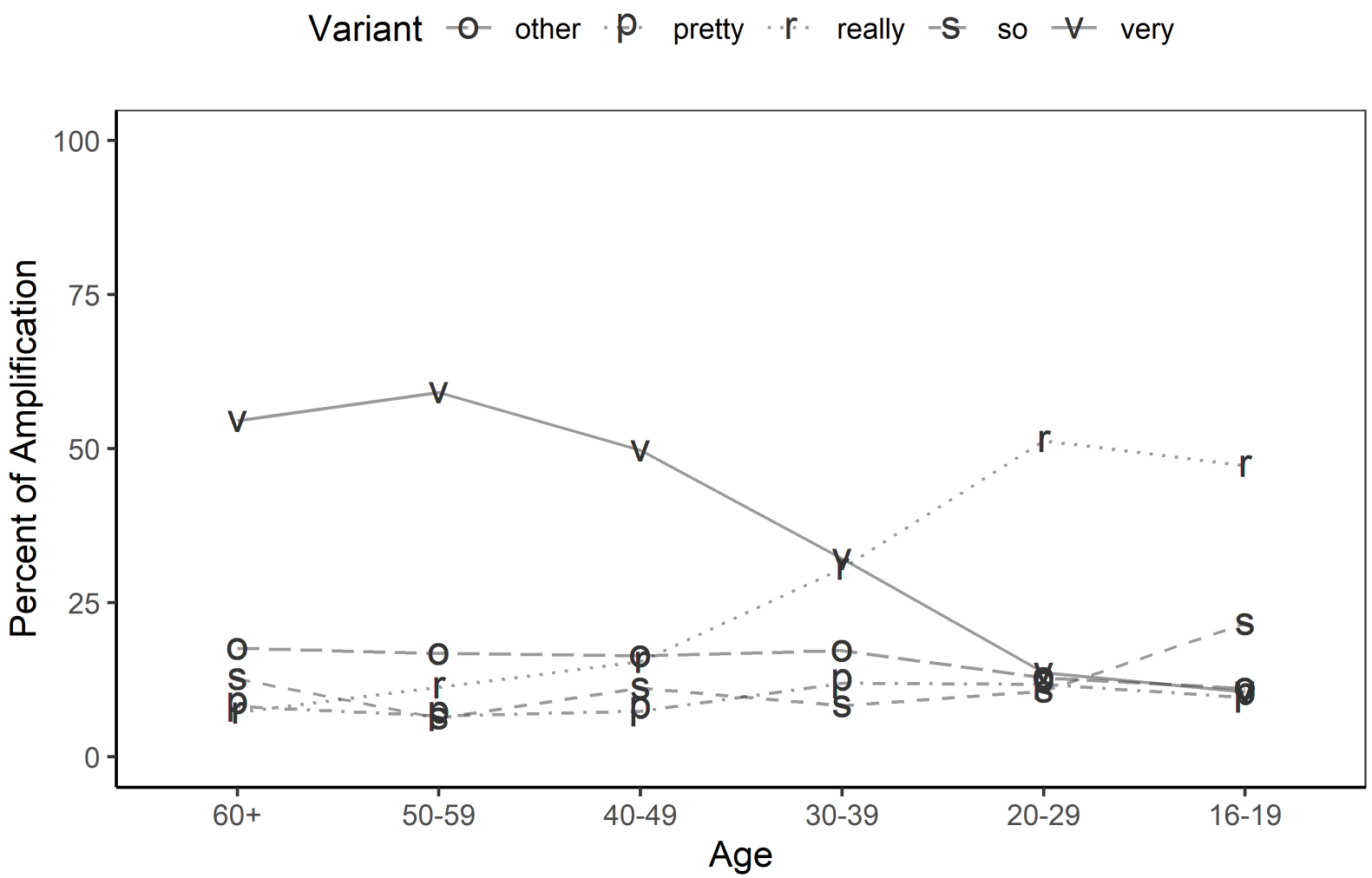
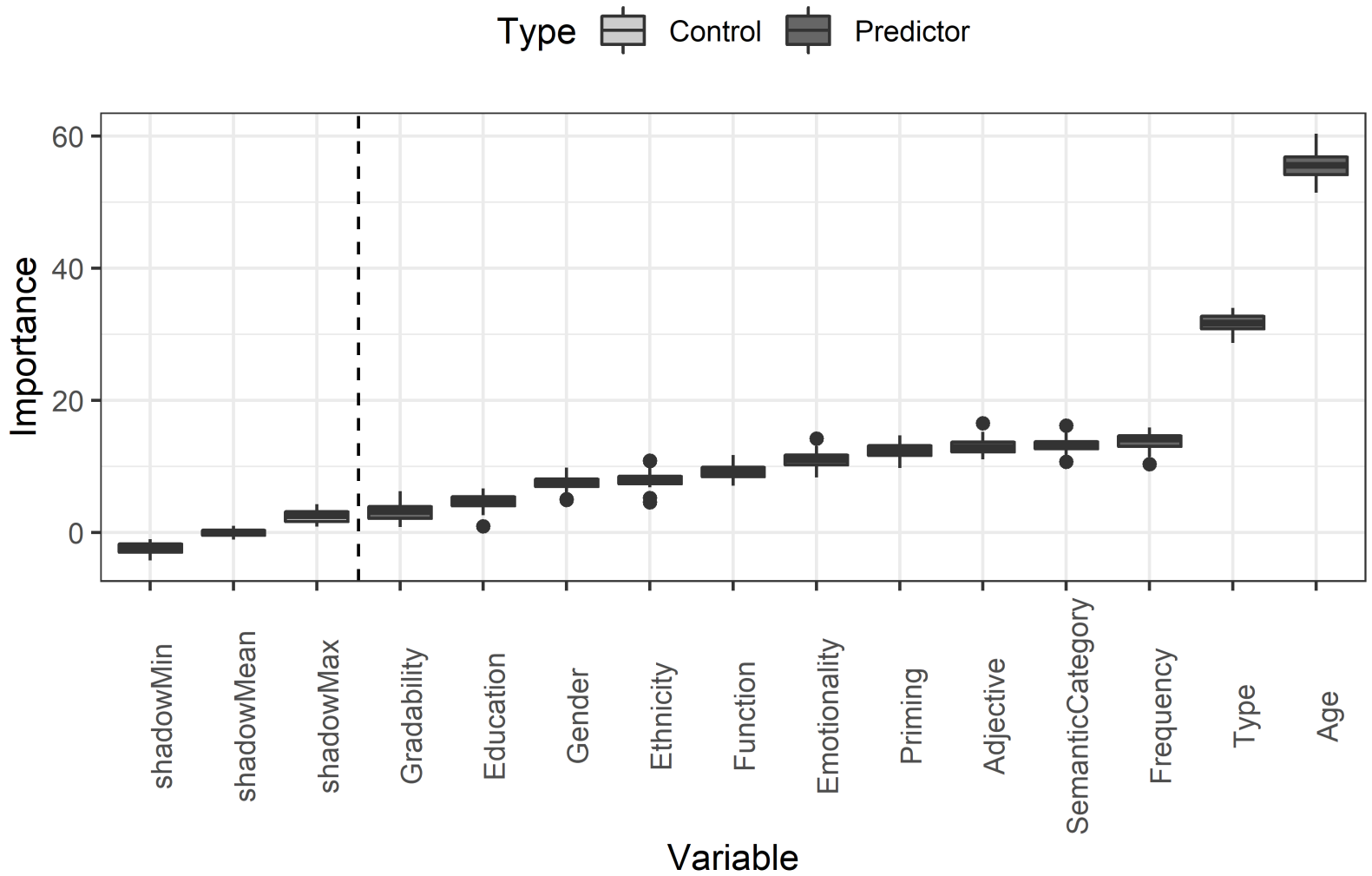


Figure 3: Percent of Amplifier Variants in Amplified Pre-Adjectival Slots across Syntactic Function.

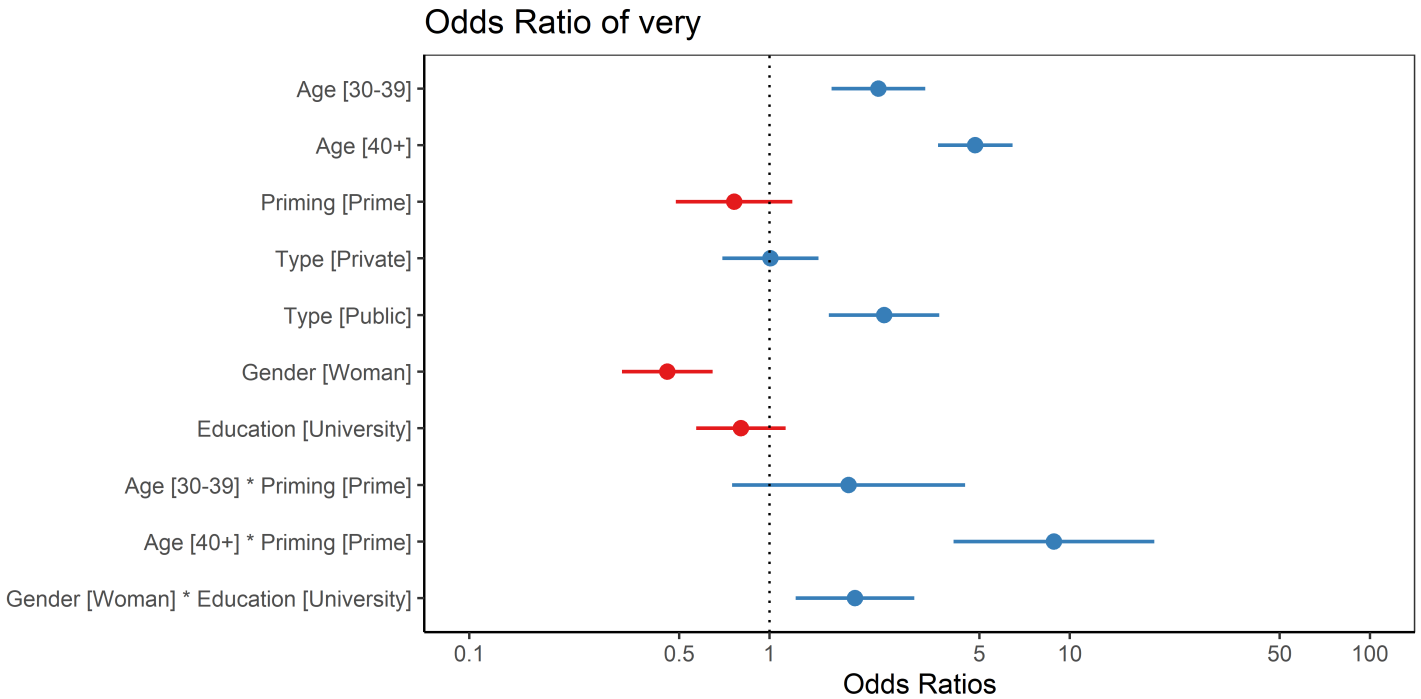

Type
Control
Predictor
Importance
60
40
20
0
shadowMin
shadowMean
shadowMax
Gradability
Education
Gender
Ethnicity
Function
Emotionality
Priming
Adjective
SemanticCategory
Frequency
Type
Age
Variable

Mixed-Effects Binomial Logistic Regression

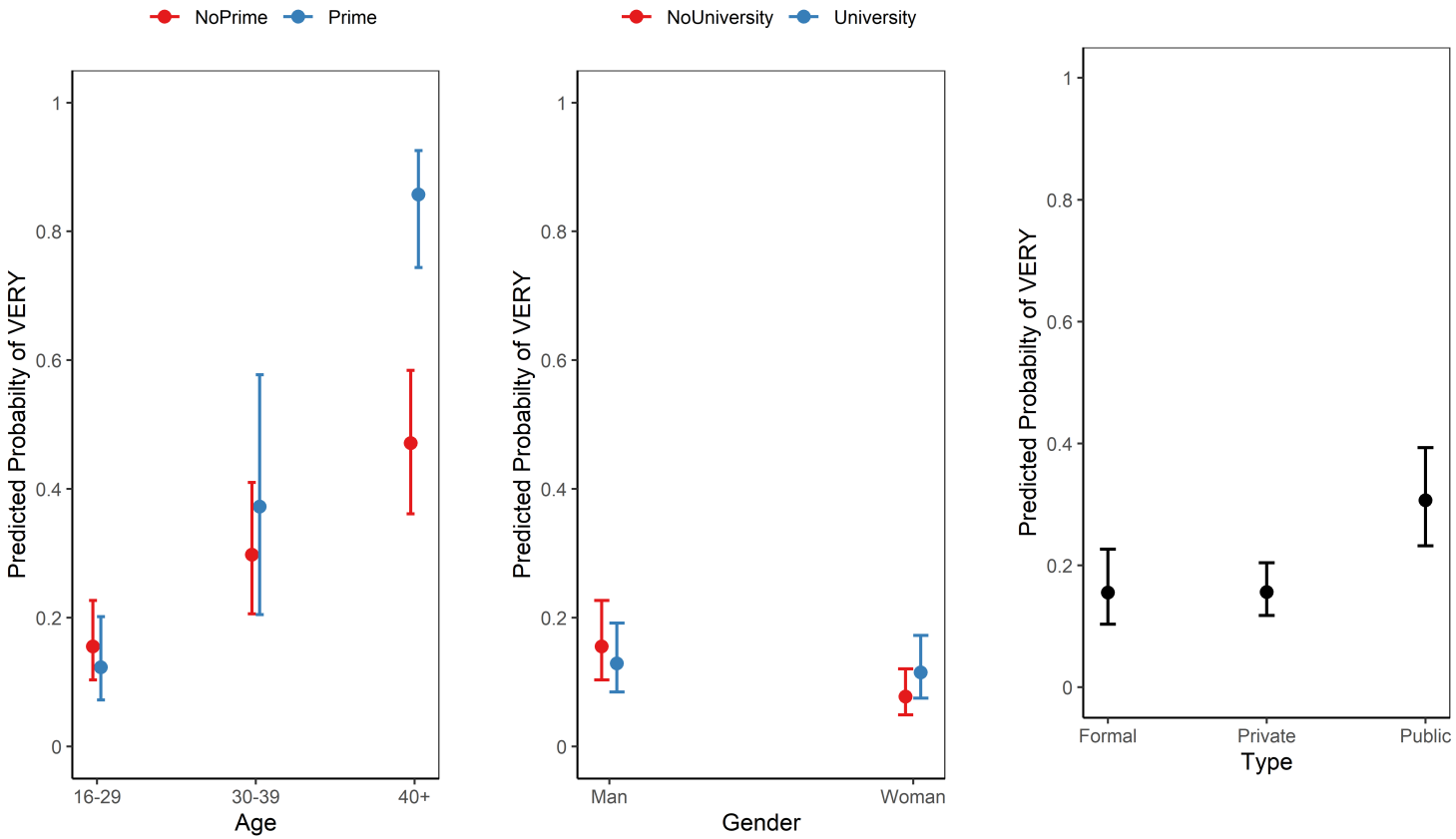
	Intercept-only Model	**Age-only Model**	**Final Minimal Model**
Predictors	Odds Ratios	Odds Ratios	Odds Ratios
(Intercept)	0.37 *** (0.30 – 0.44)	0.14 *** (0.11 – 0.17)	0.18 *** (0.12 – 0.29)
Age30-39		2.72 *** (1.98 – 3.73)	2.31 *** (1.61 – 3.31)
Age:40+		7.40 *** (5.79 – 9.45)	4.85 *** (3.64 – 6.45)
Priming:Prime			0.76 (0.49 – 1.19)
Type:Private			1.01 (0.70 – 1.46)
Type:Public			2.41 *** (1.58 – 3.67)
Gender:Woman			0.46 *** (0.32 – 0.65)
Education:University			0.80 (0.57 – 1.13)
Age30-39:PrimingPrime			1.84 (0.75 – 4.49)
Age:40+::Priming:Prime			8.85 *** (4.10 – 19.12)
Gender:Woman::Education:University			1.93 ** (1.22 – 3.04)
Random Effects			
$\sigma 2$	3.29	3.29	3.29
$\tau 00$	1.23 Adjective	0.82 Adjective	0.76 Adjective
ICC	0.27	0.20	0.19
N	420 Adjective	420 Adjective	420 Adjective
Observations	2344	2344	2344
Marginal/Conditional R^2	0.000 / 0.273	0.164 / 0.331	0.236 / 0.379
C / Somers D_{xy}	0.747 / 0.493	0.777 / 0.554	0.761 / 0.523
Accuracy (NIR)	0.724	0.8072 (0.724)	0.7969 (0.724)

$p < .05^{*}, < .01^{**}, < .001^{***}$ (Lüdecke 2021)

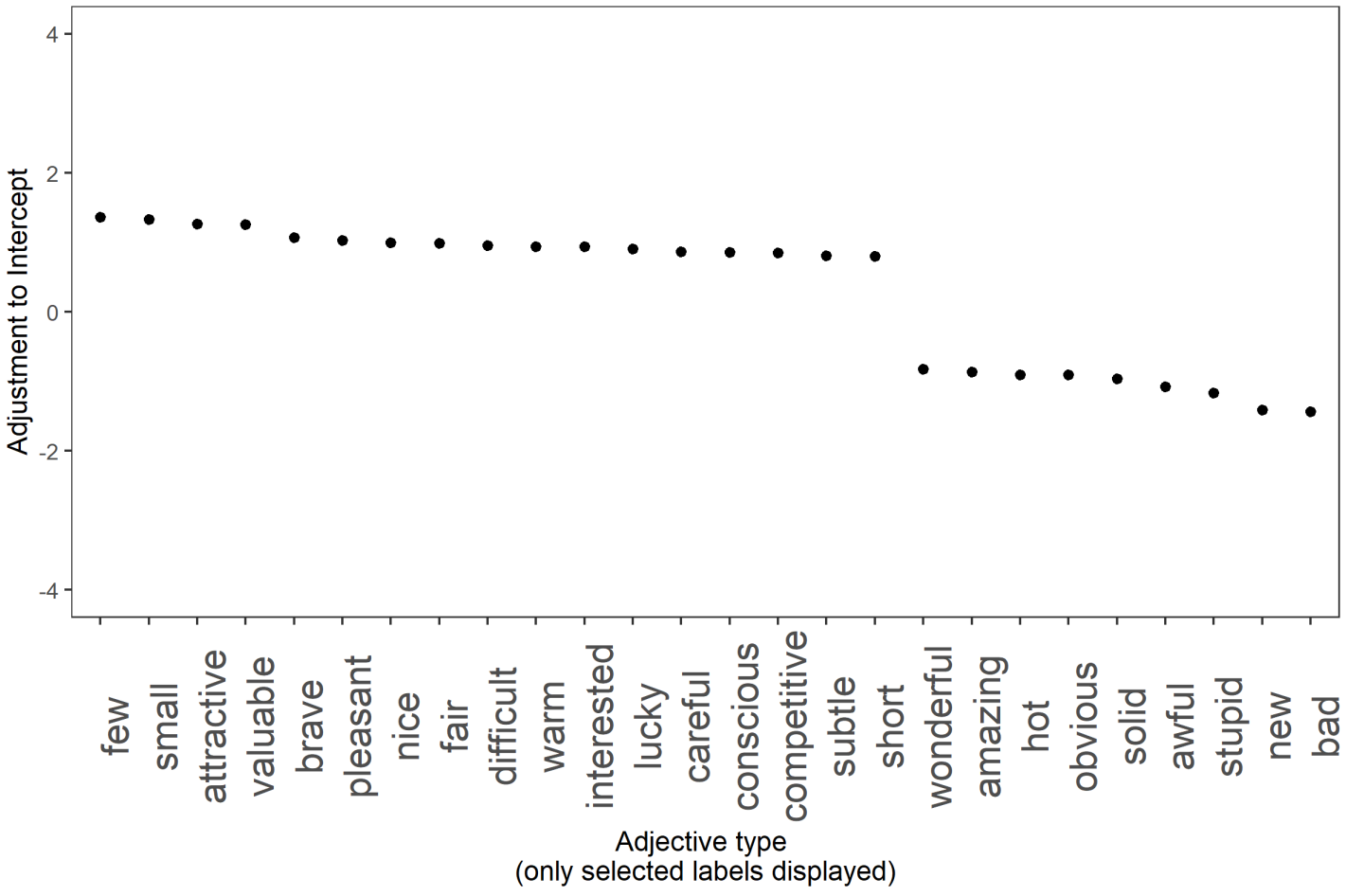
Random-Effects



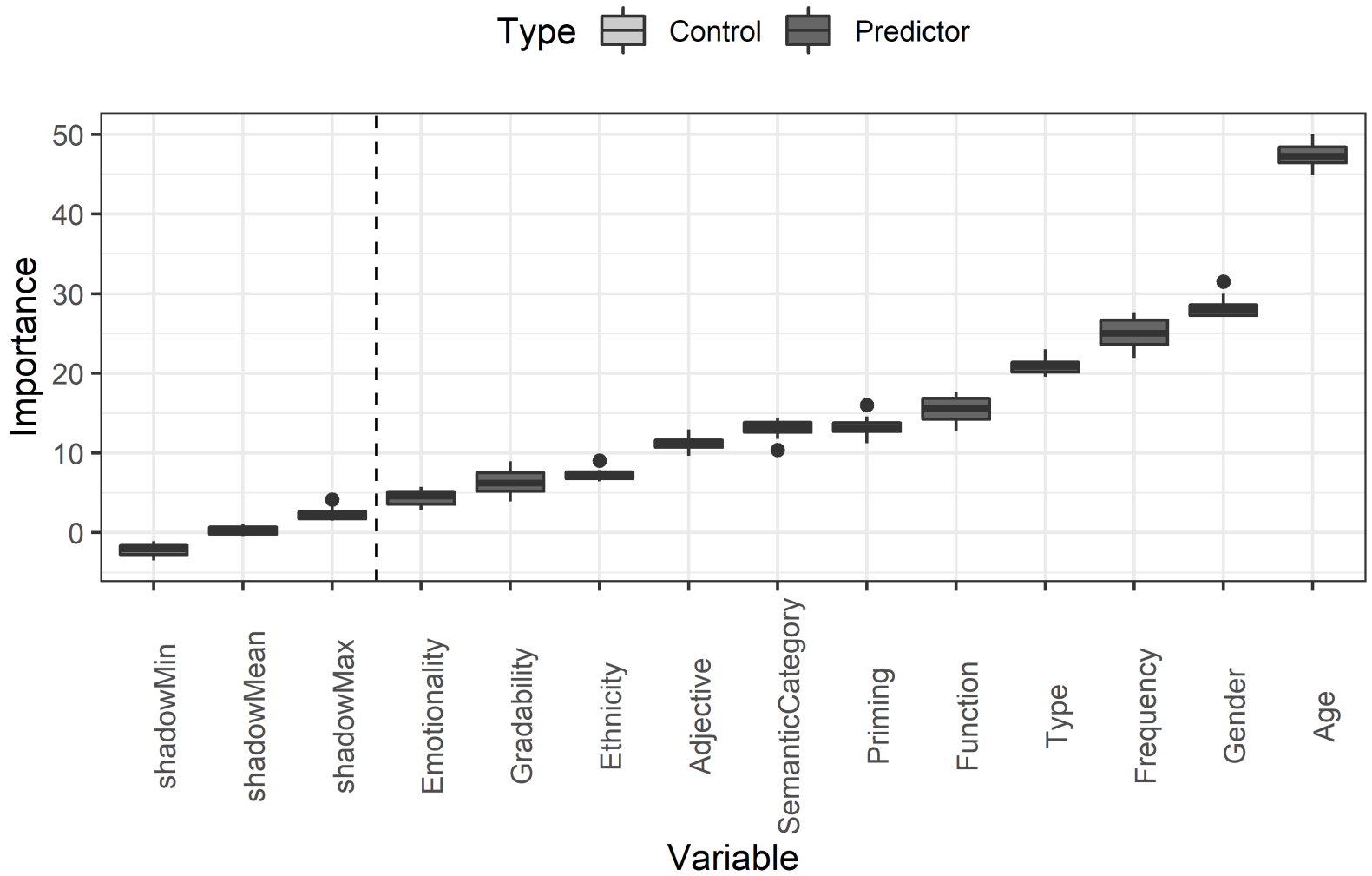
Fixed-Effects



Varying Intercepts

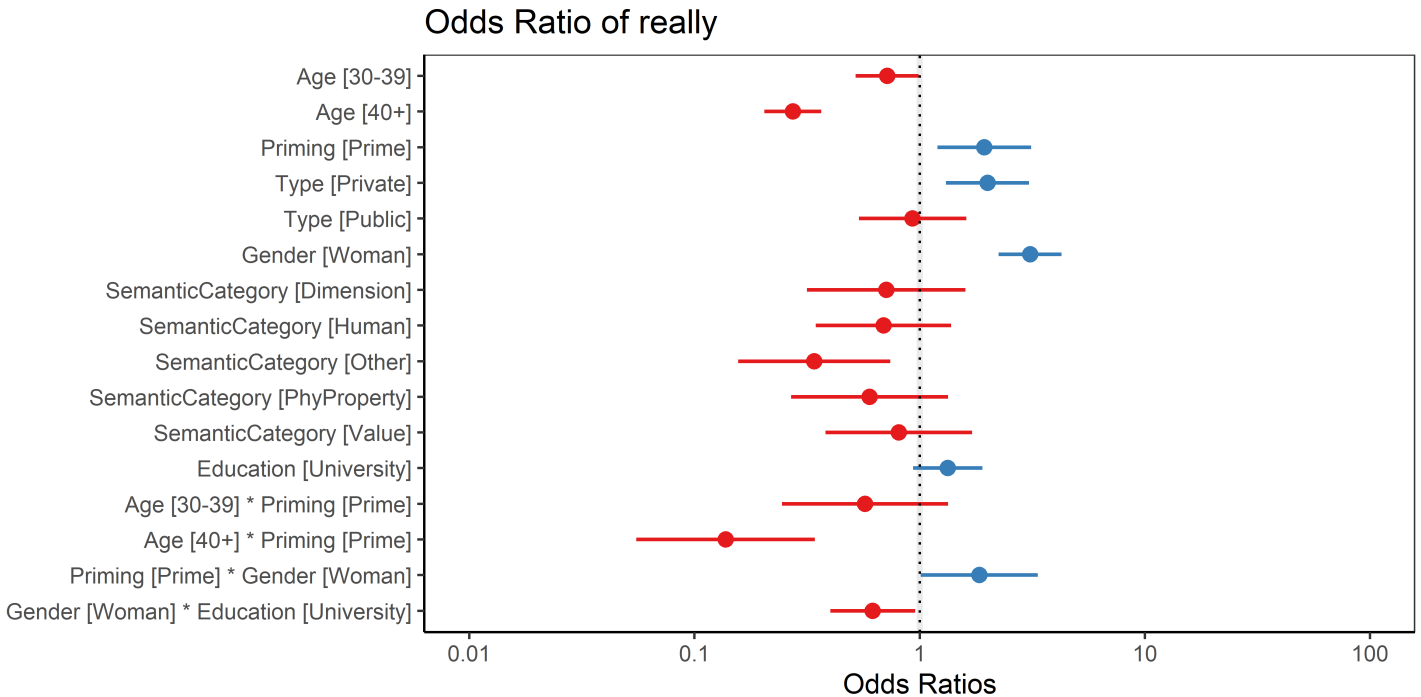


Results for really

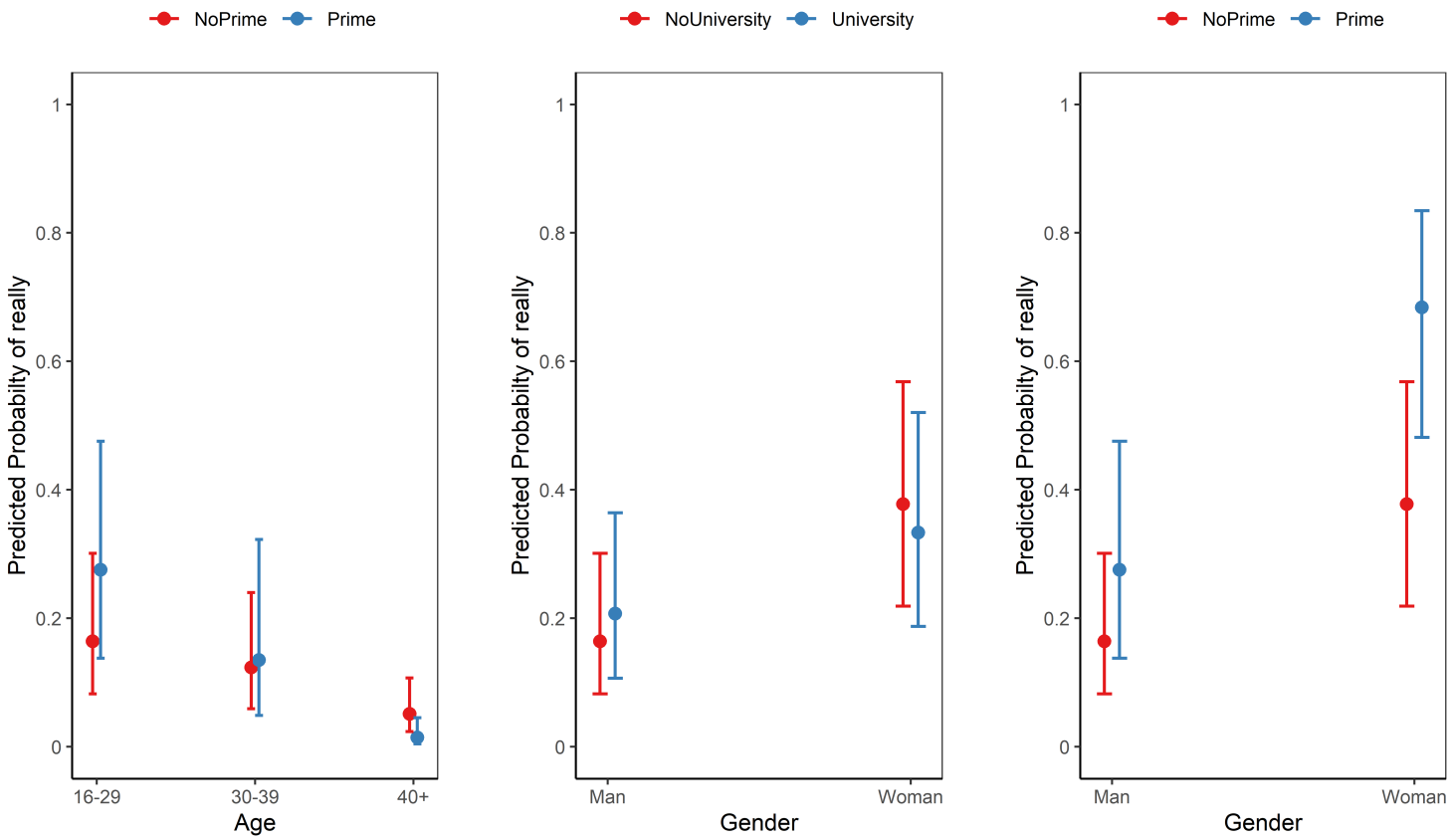

Type
Control
Predictor
Importance
50
40
30
20
10
0
shadowMin
shadowMean
shadowMax
Emotionality
Gradability
Ethnicity
Adjective
SemanticCategory
Priming
Function
Type
Frequency
Gender
Age
Variable

	Intercept-only Model	Age-only Model	Final Minimal Model
Predictors	Odds Ratios	Odds Ratios	Odds Ratios
(Intercept)	0.28*** (0.22 – 0.34)	0.55*** (0.45 – 0.68)	0.20*** (0.09 – 0.43)
Age30-39		0.56** (0.42 – 0.74)	0.72** (0.52 – 0.99)
Age [40+]		0.17*** (0.13 – 0.22)	0.27*** (0.20 – 0.37)
Priming:Prime			1.94*** (1.20 – 3.13)
Type:Private			2.00* (1.31 – 3.06)
Type:Public			0.93*** (0.54 – 1.61)
Gender:Woman			3.09* (2.24 – 4.27)
SemanticCat.:Dimension			0.71 (0.31 – 1.60)
SemanticCat.:Human			0.69 (0.35 – 1.38)
SemanticCat.:Other			0.34** (0.16 – 0.74)
SemanticCat.:PhyProperty			0.60 (0.27 – 1.34)
SemanticCat.:Value			0.81 (0.34 – 1.71)
Education:Univ.			1.33 (0.93 – 1.90)
Age:30-39::Priming:Prime			0.57 (0.24 – 1.34)
Age:40+::Priming:Prime			0.14*** (0.06 – 0.34)
Priming:Prime::Gender:Woman			1.84* (1.01 – 3.35)
Gender:Woman::Education:Univ.			0.62* (0.40 – 0.96)
Random Effects			
$\sigma 2$	3.29	3.29	3.29
$\tau 00$	0.85 Adjective	0.62 Adjective	0.50 Adjective
ICC	0.20	0.16	0.13
N	420 Adjective	420 Adjective	420 Adjective
Observations	2344	2344	2344
Marginal R^2, Conditional R^2	0.000, 0.205	0.138, 0.275	0.278, 0.373
C / D_{xy}	0.651 /0.301	0.707 / 0.415	0.732 / 0.466
Accuracy (NIR)	0.637	0.730 (0.637)	0.750 (.637)

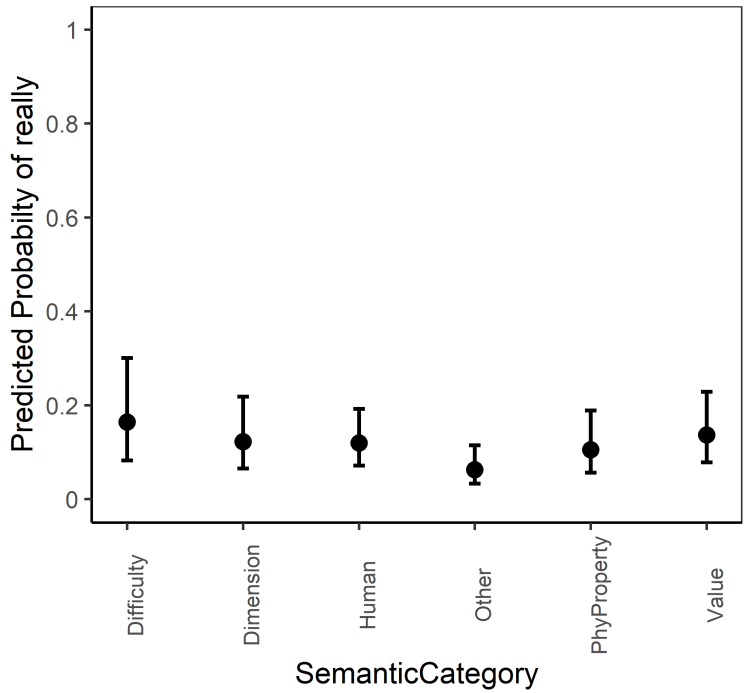
Random-Effects

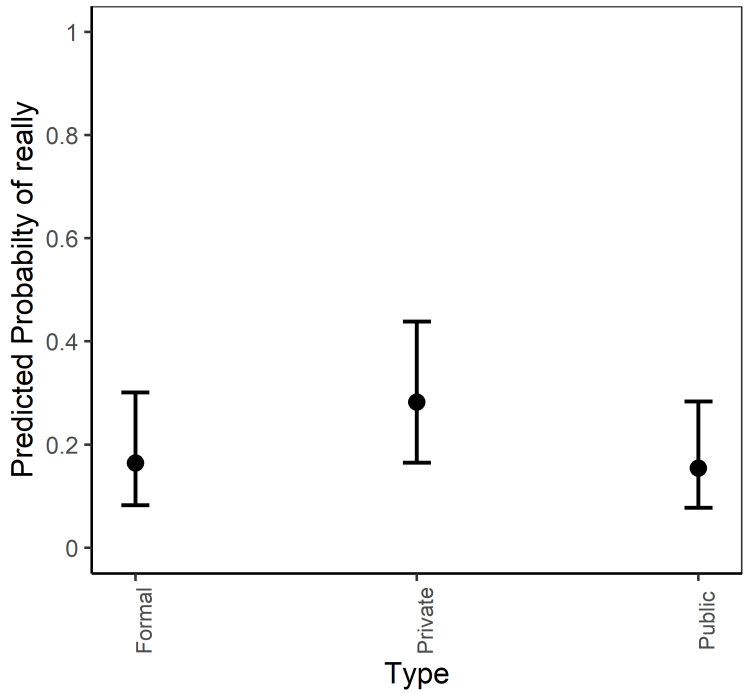


Fixed-Effects

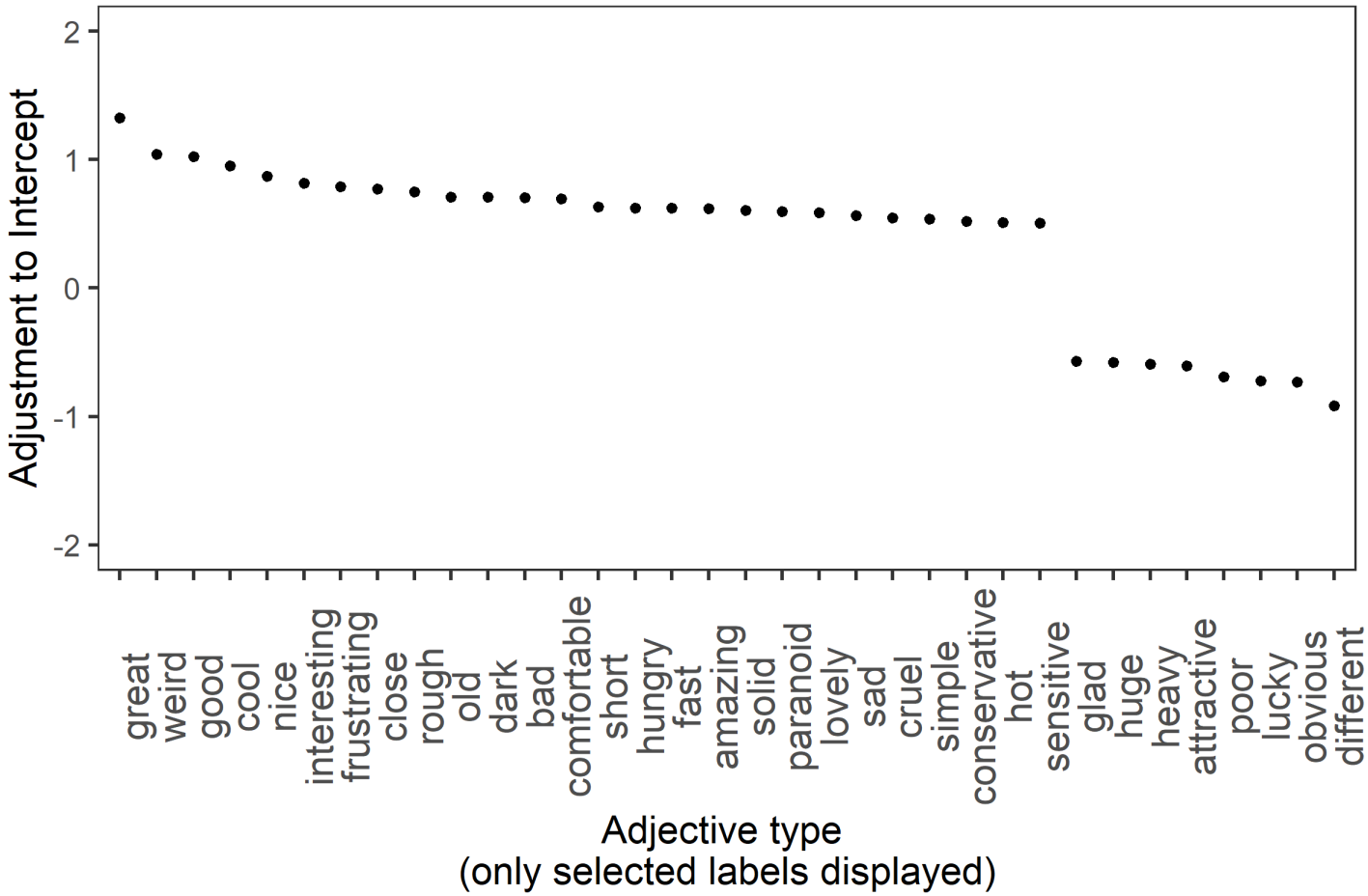


Fixed-Effects





Varying Intercepts



Discussion & Outlook

Summary

Analysis shows that *very* is

- used predominantly by older speakers and in public discourse
- rejected by younger speakers and in private as well as in formal contexts.

The use of *very* collocates with several extra-linguistic factors but the impact of these factors is limited.

Discussion

Waning of *very*: ***ruins of stratification***

- no intra-linguistic variability weak social stratification (the obvious exception is age).
- motivated by functional factors (lack of expressivity).
- selected adjectives prefer *very* as their default amplifier (see pos. adjustments to intercepts).
- Effect of genre (Type): while loosing ground to rival variants - *very* more stable in certain registers (public discourse) $\rightarrow$ lack of social meaning.

Waning not symmetric to waxing (accompanied by substantive social and linguistic stratification)

Things to be noted

A few words on effect size

- Discussion between Adam Kilgarriff and Stefan Gries (Kilgarriff 2005; Gries 2005): with large N everything becomes significant (check effect size!)

Checking effect size does not really help in the present case because when dealing with low probabilities, Odds Ratios can become large very easily: 1/100,000 to 10/100,000 would have and OddsRatio of 10! (although the effect is still super weak)

Outlook

Questions we may want to discuss...

How weak do effects have to be to be meaningless?

Do we have a problem with overpowered studies?

Cut-off for effect size? Rather not...

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.

Allaire, J. (2012). Rstudio: integrated development environment for r. *Boston, MA 770*, 394.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

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.

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.

Gries, S. T. (2005). Null-hypothesis significance testingof word frequencies: a follow-up on kilgarriff. *Corpus Linguistics and Linguistic Theory 1*(2), 277–294.

Holmes, J., B. Vine, and G. Johnson (1998). The wellington corpus of spoken new zealand english.

Hornik, K. (2019). opennlp: Apache opennlp tools interface.

Hothorn, T., K. Hornik, and A. Zeileis (2015). ctree: Conditional inference trees. *The comprehensive R archive network 8*.

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.

Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory 1*(2), 263–275.

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.

Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.7.

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.

Ooms, J. (2020). hunspell: High-performance stemmer, tokenizer, and spell checker.

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.

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 ?*(?), ??–??

Silge, J. and D. Robinson (2016). tidytext: Text mining and analysis using tidy data principles in r. *Journal of Open Source Software 1*(3), 37.

Straka, M. and J. Straková (2017). Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pp. 88–99.

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.

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 glowbe. *Lingua 200*, 63–83.

Weinreich, U., W. Labov, and M. I. Herzog (1968). Empirical foundations for a theory of language change. In W. P. Lehmann and Y. Malkiel (Eds.), *Directions for Historical Linguistics*, pp. 95–195. Austin, TX: University of Texas Press.

On the waning of forms - a corpus-based analysis of losers in language change

Martin Schweinberger

slides available at

www.martinschweinberger.de

R code at

https://github.com/MartinSchweinberger/isle6_verynze/tree/master

THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

CREATE CHANGE