Cognitive indigenization effects in the English dative alternation

Melanie Röthlisberger

KU Leuven
Quantitative Lexicology and Variational Linguistics

Workshop: Probabilistic variation across dialects and varieties, April 4-5, Leuven

Introduction

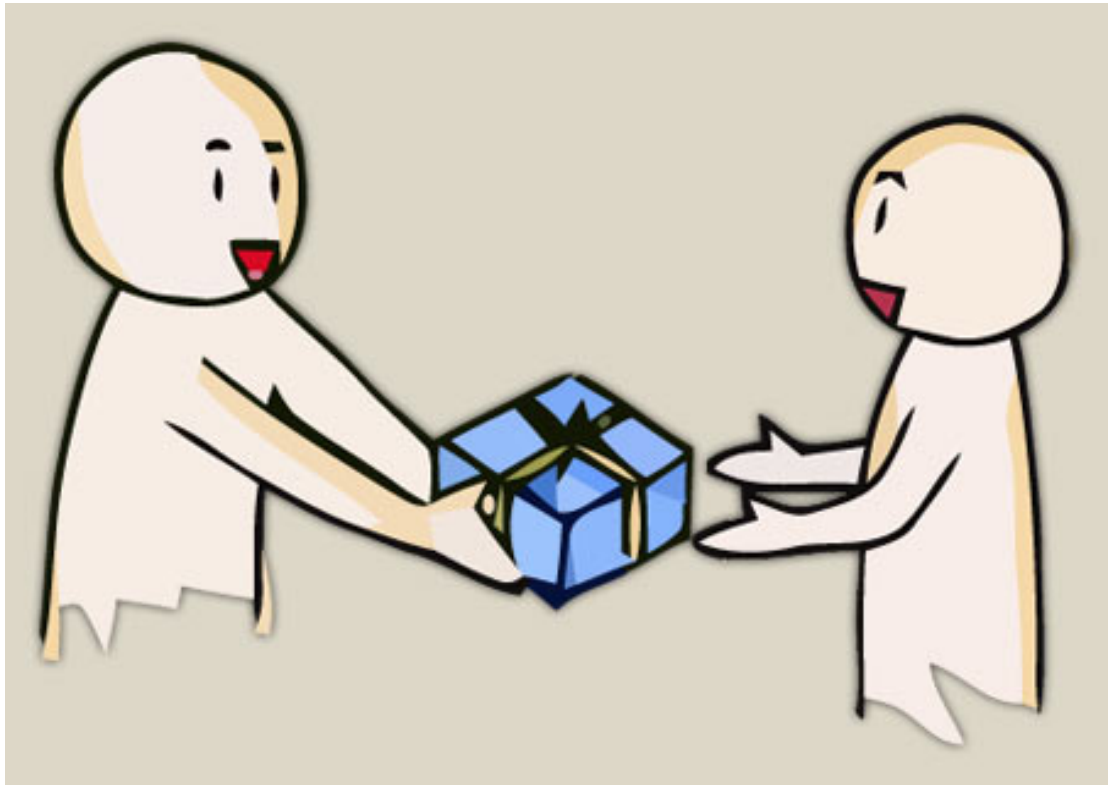
Cognitive indigenization effects
in
the English dative alternation

Cognitive indigenization

- nativization/indigenization = "the emergence of locally characteristic linguistic patterns" (Schneider 2007: 6)
- = indigenization on the level of underlying stochastic patterns that are shaped by language-internal (cognitive) factors (e.g. end-weight)

Cognitive indigenization effects in the English dative alternation

The dative alternation



The dative alternation

(1) ditransitive dative

He gives [Mary]$_{recipient}$ [a present]$_{theme}$

(2) prepositional dative

He gives [a present]$_{theme}$ to [Mary]$_{recipient}$

→ "alternate ways of saying 'the same' thing" (Labov 1972: 188)

Research questions

- What is the extent to which varieties of English share a stable probabilistic grammar?
- Are some factors more amenable to regional differences than others?

today

1. setting the frameworks
2. data & methods
3. analysis & results
4. discussion
5. outlook
6. unresolved issues

setting the frameworks

theoretical frameworks

- Probabilistic Grammar framework
  - grammar is gradient and probabilistic
  - constraint-based accounts
  - probabilistic indigenization

theoretical frameworks

- Probabilistic Grammar framework
  - grammar is gradient and probabilistic
  - constraint-based accounts
  - probabilistic indigenization
- Connection to: Cognitive sociolinguistics
  - cognitive factors and sociocultural factors both constrain linguistic variation, language planning, production, and comprehension

theoretical frameworks

- Probabilistic Grammar framework
  - grammar is gradient and probabilistic
  - constraint-based accounts
  - probabilistic indigenization
- Connection to: Cognitive sociolinguistics
  - cognitive factors and sociocultural factors both constrain linguistic variation, language planning, production, and comprehension
- Connection to: Psycholinguistic explanations
  - linguistic experience and statistical properties of the input shape language form

previous research

- statistical tendencies and processing principles underlying the dative alternation are shared across varieties

previous research

- statistical tendencies and processing principles underlying the dative alternation are shared across varieties
- stability in probabilistic grammars
  - 'easy' comes first $\rightarrow$ congruent effect
  - easy = animate, definite, pronominal, short

previous research

- statistical tendencies and processing principles underlying the dative alternation are shared across varieties
- stability in probabilistic grammars
  - 'easy' comes first $\rightarrow$ congruent effect
  - easy = animate, definite, pronominal, short
- variability (indigenization) in probabilistic grammars
  - recipient animacy: NZE vs. AmE
  - end-weight: AmE vs. AusE

(e.g. Bresnan and Hay 2008; Bresnan and Ford 2010)

data & methods

the corpus

- International Corpus of English (ICE) - series
- 60% spoken (transcriptions), 40% written texts
- 1m words per subcorpus
- 500 texts, 2,000 words per text
- 12 different registers, same corpus structure

the data

- British E, Canadian E, Indian E, Singapore E, Irish E, New Zealand E, Hong Kong E, Jamaican E, Philippines E

methods

(e.g. Bresnan et al. 2007)

1. extract dative tokens using verb list
2. define choice context (incl. pronouns), leave out, e.g.:
   - fixed and idiomatic expressions (e.g. *bring it to the boil*)
   - spatial goals (e.g. *send their daughter to school*)
   - beneficiaries (e.g. *We get them uh typed photo copies*)

N=8,549

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)
- syntactic complexity: postmodified = 'complex', no postmodification = 'simple'

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)
- syntactic complexity: postmodified = 'complex', no postmodification = 'simple'
- pronominality: 'pronoun' vs. 'non-pronoun'

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)
- syntactic complexity: postmodified = 'complex', no postmodification = 'simple'
- pronominality: 'pronoun' vs. 'non-pronoun'
- discourse givenness: occurrence in 100 words of preceding discourse/pronoun = 'given', otherwise = 'new'

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)
- syntactic complexity: postmodified = 'complex', no postmodification = 'simple'
- pronominality: 'pronoun' vs. 'non-pronoun'
- discourse givenness: occurrence in 100 words of preceding discourse/pronoun = 'given', otherwise = 'new'
- definiteness: existential reading in *There is/are* ...

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)
- syntactic complexity: postmodified = 'complex', no postmodification = 'simple'
- pronominality: 'pronoun' vs. 'non-pronoun'
- discourse givenness: occurrence in 100 words of preceding discourse/pronoun = 'given', otherwise = 'new'
- definiteness: existential reading in *There is/are* ...
- person of recipient: 1st & 2nd = 'local', otherwise = 'non-local'

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)
- syntactic complexity: postmodified = 'complex', no postmodification = 'simple'
- pronominality: 'pronoun' vs. 'non-pronoun'
- discourse givenness: occurrence in 100 words of preceding discourse/pronoun = 'given', otherwise = 'new'
- definiteness: existential reading in *There is/are* ...
- person of recipient: 1st & 2nd = 'local', otherwise = 'non-local'
- animacy of recipient: human/animal = 'animate'

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)
- syntactic complexity: postmodified = 'complex', no postmodification = 'simple'
- pronominality: 'pronoun' vs. 'non-pronoun'
- discourse givenness: occurrence in 100 words of preceding discourse/pronoun = 'given', otherwise = 'new'
- definiteness: existential reading in *There is/are* ...
- person of recipient: 1st & 2nd = 'local', otherwise = 'non-local'
- animacy of recipient: human/animal = 'animate'
- concreteness of theme: perceivable by 5 senses = 'concrete'

explanatory factors

- length (end-weight): measured as weight ratio = ln(# of characters in recipient / # of characters in theme)
- syntactic complexity: postmodified = 'complex', no postmodification = 'simple'
- pronominality: 'pronoun' vs. 'non-pronoun'
- discourse givenness: occurrence in 100 words of preceding discourse/pronoun = 'given', otherwise = 'new'
- definiteness: existential reading in *There is/are* ...
- person of recipient: 1st & 2nd = 'local', otherwise = 'non-local'
- animacy of recipient: human/animal = 'animate'
- concreteness of theme: perceivable by 5 senses = 'concrete'
- verb sense: t, f, p, c, a

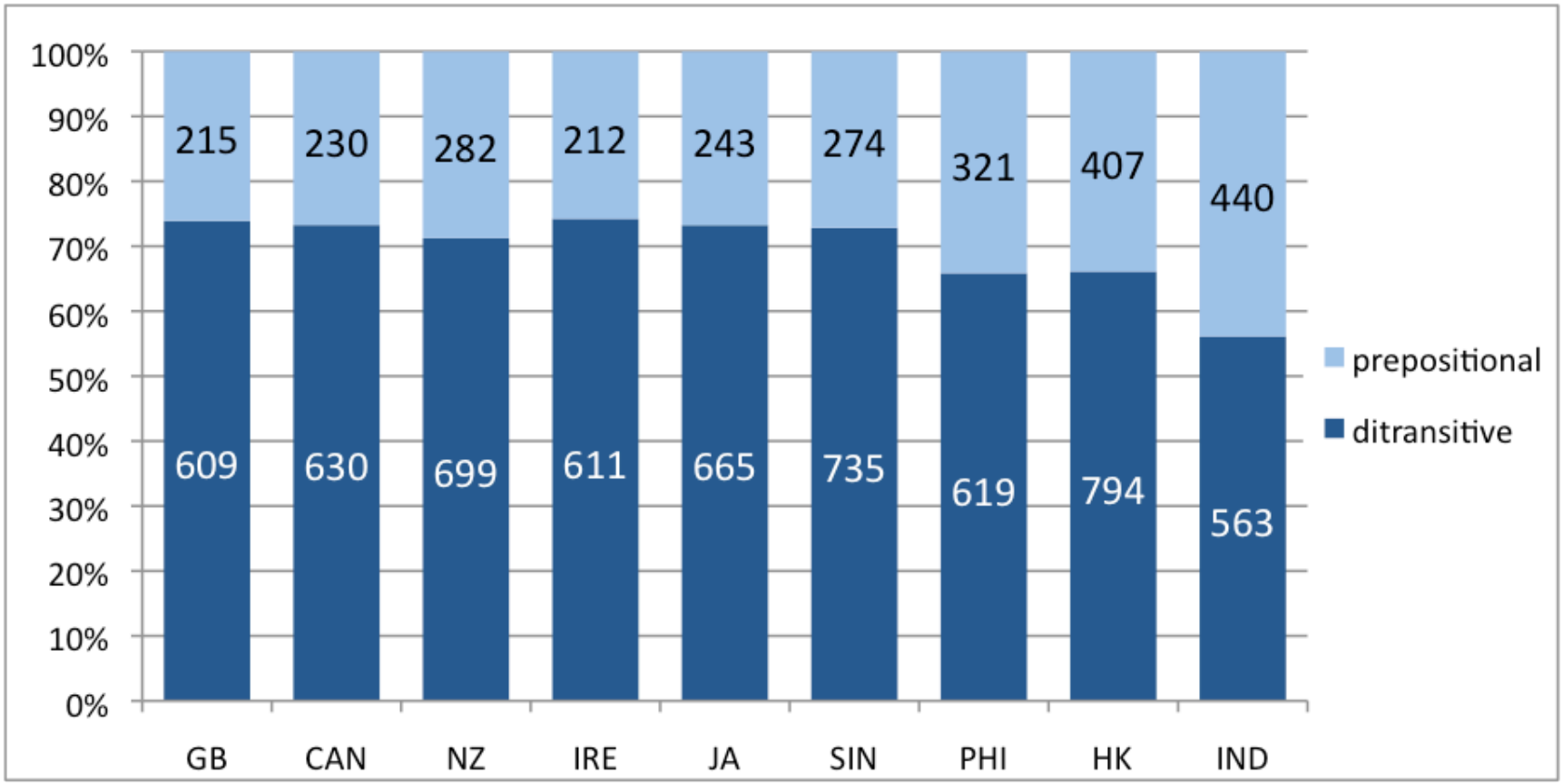
explanatory factors

- verb sense
  - transfer: *I pay you ten dollars*
  - future transfer: *They award him a silver medal*
  - prevention: *I'll charge you some money*
  - communication: *I owe you an apology*
  - abstract: *Can you please pay attention to the graph*

explanatory factors

- variety: BrE, CanE, SinE, etc.
- register
- corpus metadata: e.g. FileID, text category, etc.

variety

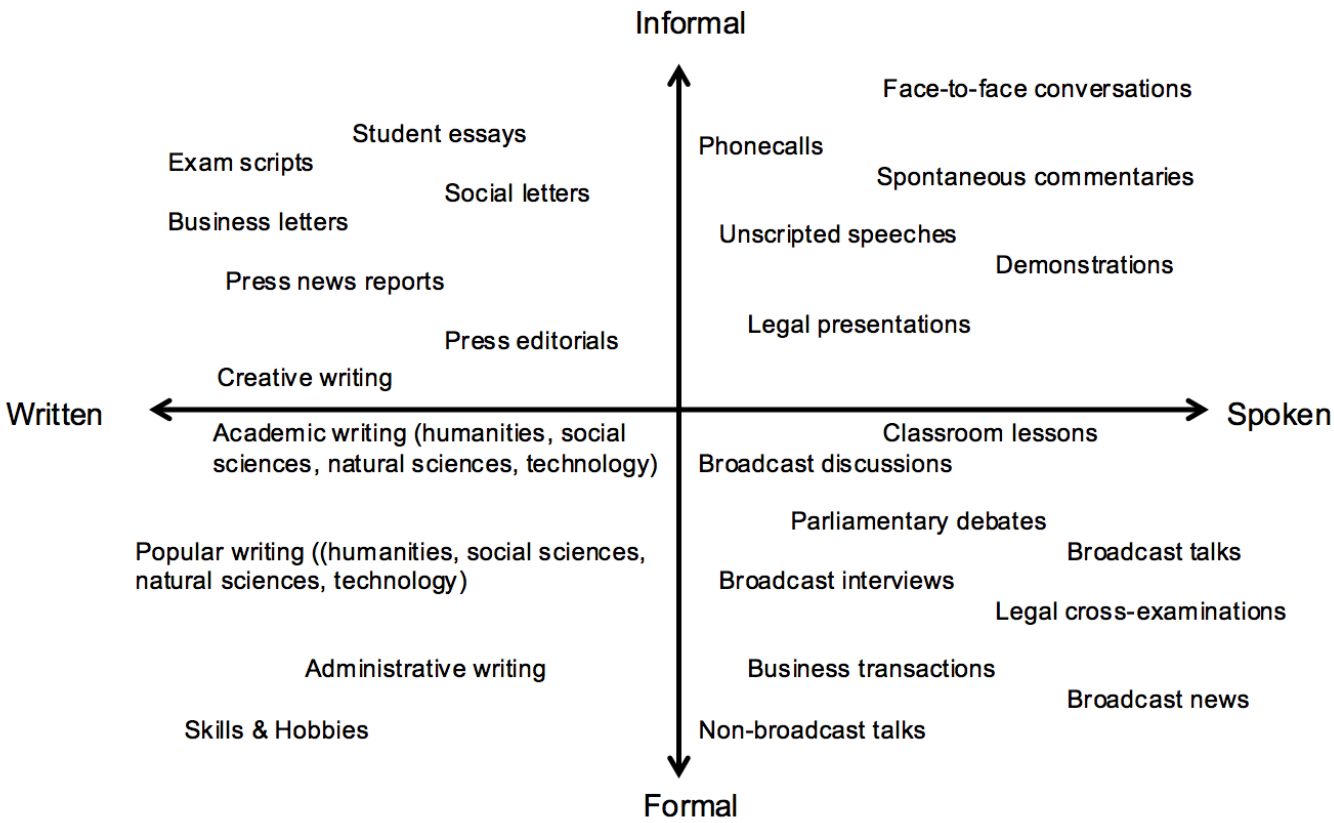


Dative proportions across all nine ICE corpora, N=8549

register coding ICE

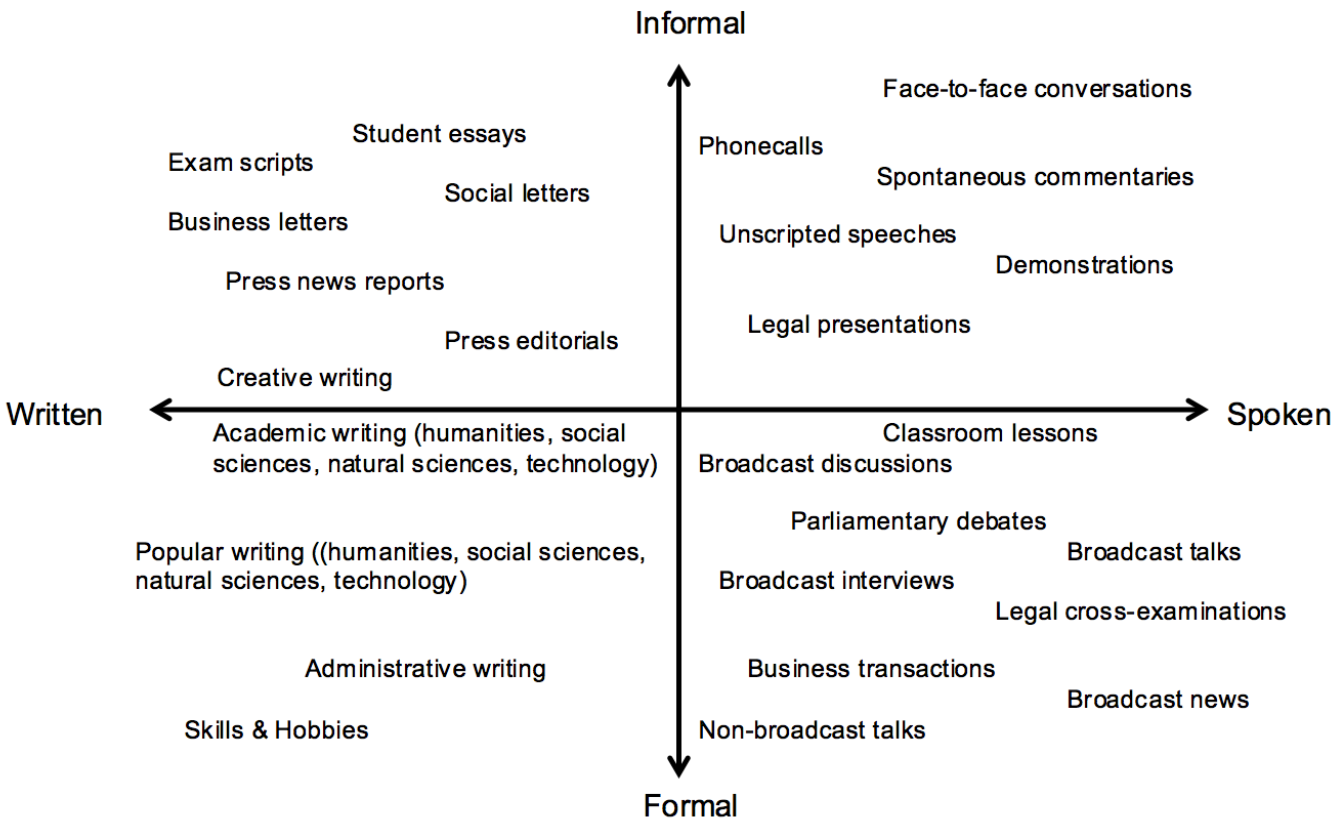
SPOKEN 300		Dialogues	180	Private	100	Face-to-face conversations	90	s1a
						Phonecalls	10	
				Public	80	Classroom lessons	20	s1b
						Broadcast Discussions	20	
						Broadcast Interviews	10	
						Parliamentary Debates	10	
						Legal cross-examinations	10	
						Business Transactions	10	
		Monologues	120	Unscripted	70	Spontaneous commentaries	20	s2a
						Unscripted Speeches	30	
						Demonstrations	10	
						Legal Presentations	10	
				Scripted	50	Broadcast News	20	s2b
						Broadcast Talks	20	
						Non-broadcast Talks	10	
WRITTEN 200	**200**	Non-printed	50	Student Writing	20	Student Essays	10	w1a
						Exam Scripts	10	
				Letters	30	Social Letters	15	w1b
						Business Letters	15	
		Printed	150	Academic writing	40	Humanities	10	w2a
						Social Sciences	10	
						Natural Sciences	10	
						Technology	10	
				Popular Writing	40	Humanities	10	w2b
						Social Sciences	10	
						Natural Sciences	10	
						Technology	10	
				Reportage	20	Press news reports	20	w2c
				Instructional writing	20	Administrative Writing	10	w2d
						Skills/Hobbies	10	
				Persuasive writing	10	Press editorials	10	w2e
				Creative writing	20	Novels & short stories	20	w2f

register coding in this study



(Koch and Oesterreicher 1985)

register coding in this study



(Koch and Oesterreicher 1985)

4 levels → **SpokInf, SpokForm, WritInf, WritForm**

analysis & results

analysis

- mixed-effects logistic regression
- deviation coding for VARIETY and REGISTER: compare every level to the mean of ALL levels
- predicted outcome: prepositional dative
- `glmer()` function in Rs `lme4` package

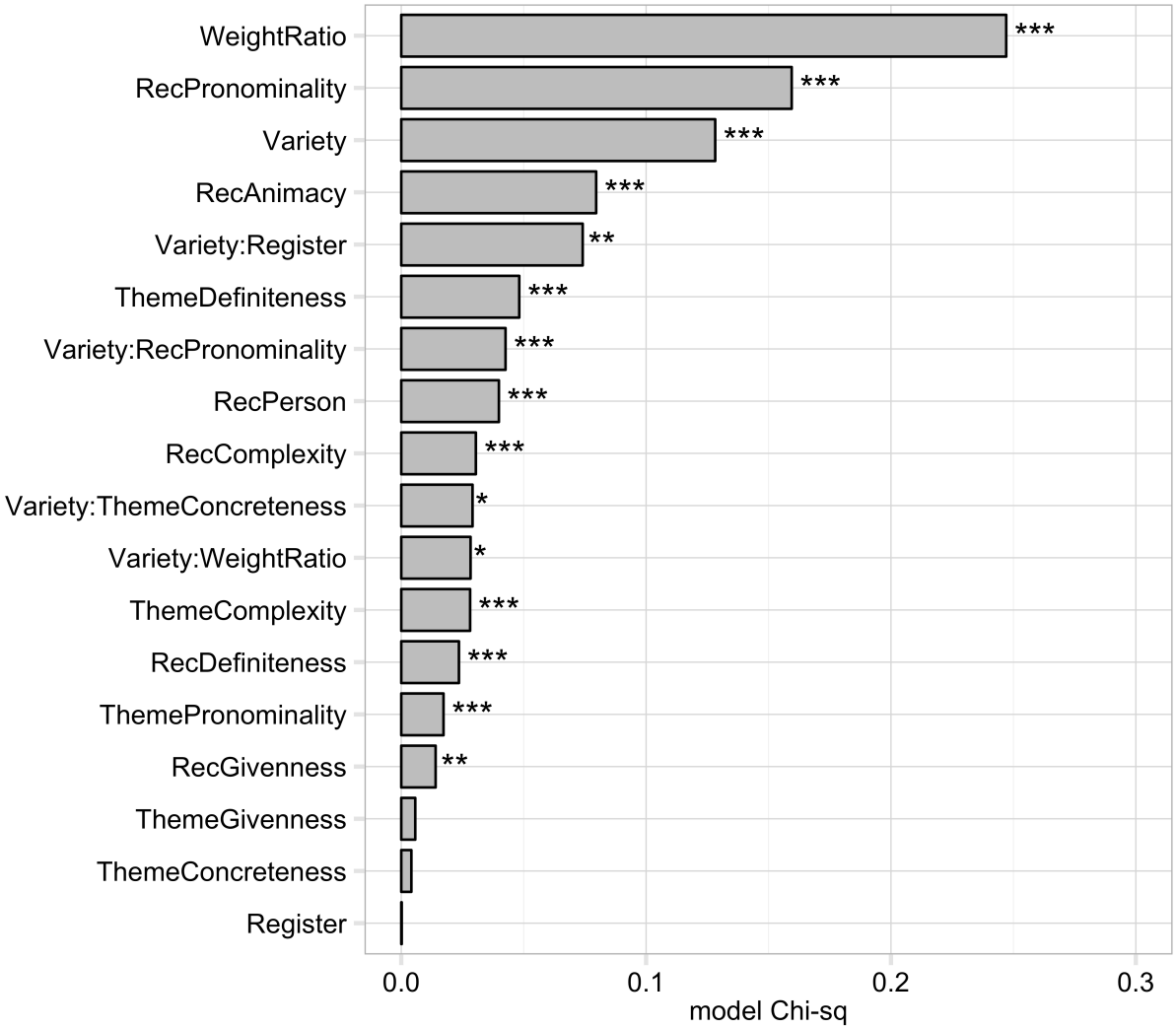
  (Bates, Maechler, and Bolker Bates et al.; Harrell 2001)
- random effects include
  - verb lemma and verb sense
  - corpus structure
  - recipient and theme head lemmas

dative model

$$\text{Response} = \{\text{ditransitive, prepositional}\}$$

$$\begin{aligned}\text{Response} \sim\ & (1|\text{VerbLemma}/\text{VerbSense}) \\ & + (1|\text{ThemeHead}) \\ & + (1|\text{CorpusStructure}) \\ & + \text{RecComplexity} \\ & + \text{RecGivenness} \\ & + \text{ThemeComplexity} \\ & + \text{RecPerson} \\ & + \text{RecDefiniteness} \\ & + \text{ThemePron} \\ & + \text{RecAnimacy} \\ & + \text{ThemeGivenness} \\ & + \text{ThemeDefiniteness} \\ & + \text{Variety} * \\ & (\text{Register} + \text{RecPron} + \text{ThemeConcreteness} + \text{WeightRatio})\end{aligned}$$

importance of predictors



Predicted outcome: PD; *C*-value: 0.98; Accuracy: 93.6% (baseline: 69 %)

results

- What is the extent to which varieties of English share a stable probabilistic grammar?
- Are some factors more amenable to regional differences than others?

main effects

Predictor	b	SE	p
(intercept)	2.525	0.405	<0.001
RECIPIENT COMPLEXITY			
simple $\Rightarrow$ complex	0.898	0.204	<0.001
THEME COMPLEXITY			
simple $\Rightarrow$ complex	-0.692	0.164	<0.001
RECIPIENT PERSON			
local $\Rightarrow$ non-local	0.882	0.175	<0.001
RECIPIENT ACCESSIBILITY			
given $\Rightarrow$ new	0.388	0.130	<0.01
RECIPIENT ANIMACY			
animate $\Rightarrow$ inanimate	0.994	0.140	<0.001
THEME PRONOMINALITY			
non-pronoun $\Rightarrow$ pronoun	1.552	0.468	<0.001
RECIPIENT PRONOMINALITY			
pronoun $\Rightarrow$ non-pronoun	1.945	0.191	<0.001
RECIPIENT DEFINITENESS			
definite $\Rightarrow$ indefinite	0.556	0.144	<0.001
THEME DEFINITENESS			
indefinite $\Rightarrow$ definite	0.696	0.126	<0.001
WEIGHT RATIO (rec/theme)	2.950	0.230	<0.001
VARIETY			
all $\Rightarrow$ CanE	-1.586	0.365	<0.001
all $\Rightarrow$ IndE	0.919	0.256	<0.001

main effects

- all predictors influence the choice of construction as predicted:
  - given >new
  - animate >inanimate
  - definite >indefinite
  - pron >non-pron
  - short >long

  recipient >theme $\rightarrow$ **ditransitive**

  theme >recipient $\rightarrow$ **prepositional**

interactions

Predictor	b	SE	p
VARIETY : RECIPIENT PRONOMINALITY			
CanE + non-pronoun	0.902	0.402	0.025
IndE + non-pronoun	1.108	0.353	0.002
JamE + non-pronoun	-1.253	0.402	0.002
VARIETY : WEIGHT			
IndE	-1.080	0.452	0.017
JamE	1.960	0.606	0.001
VARIETY : THEME CONCRETENESS			
CanE + concrete	1.250	0.397	0.002
VARIETY : REGISTER			
IrE + SpokForm	0.692	0.278	0.013
IrE + SpokInf	-0.604	0.287	0.035
HKE + SpokInf	0.679	0.244	0.005
HKE + WrittenForm	-0.912	0.293	0.002
HKE + WrittenInf	0.566	0.220	0.010
JamE + SpokInf	-0.703	0.312	0.024
JamE + WrittenForm	0.873	0.433	0.044
NZE + WrittenForm	0.673	0.295	0.023

cross-varietal differences

Table: Cross-varietal differences in effect size; - indicates decreased effect size, + indicates increased effect size

Variety	WeightRatio	RecPron	ThemeConcreteness
CanE	=	+	+
IndE	-	+	=
JamE	+	-	=

discussion

discussion

- general processes of language production and comprehension
  - . . . shape distributional patterns in speakers‘ experience
  - . . . which gives rise to subtle variation in the probabilistic effects of different linguistic features

discussion

- MacDonald (2013): Easy First, Plan Reuse

discussion

- MacDonald (2013): Easy First, Plan Reuse
  - consistent interplay between principles creates statistical regularities in language usage

discussion

- MacDonald (2013): Easy First, Plan Reuse
  - consistent interplay between principles creates statistical regularities in language usage
  - Easy First: creates stability in effect direction

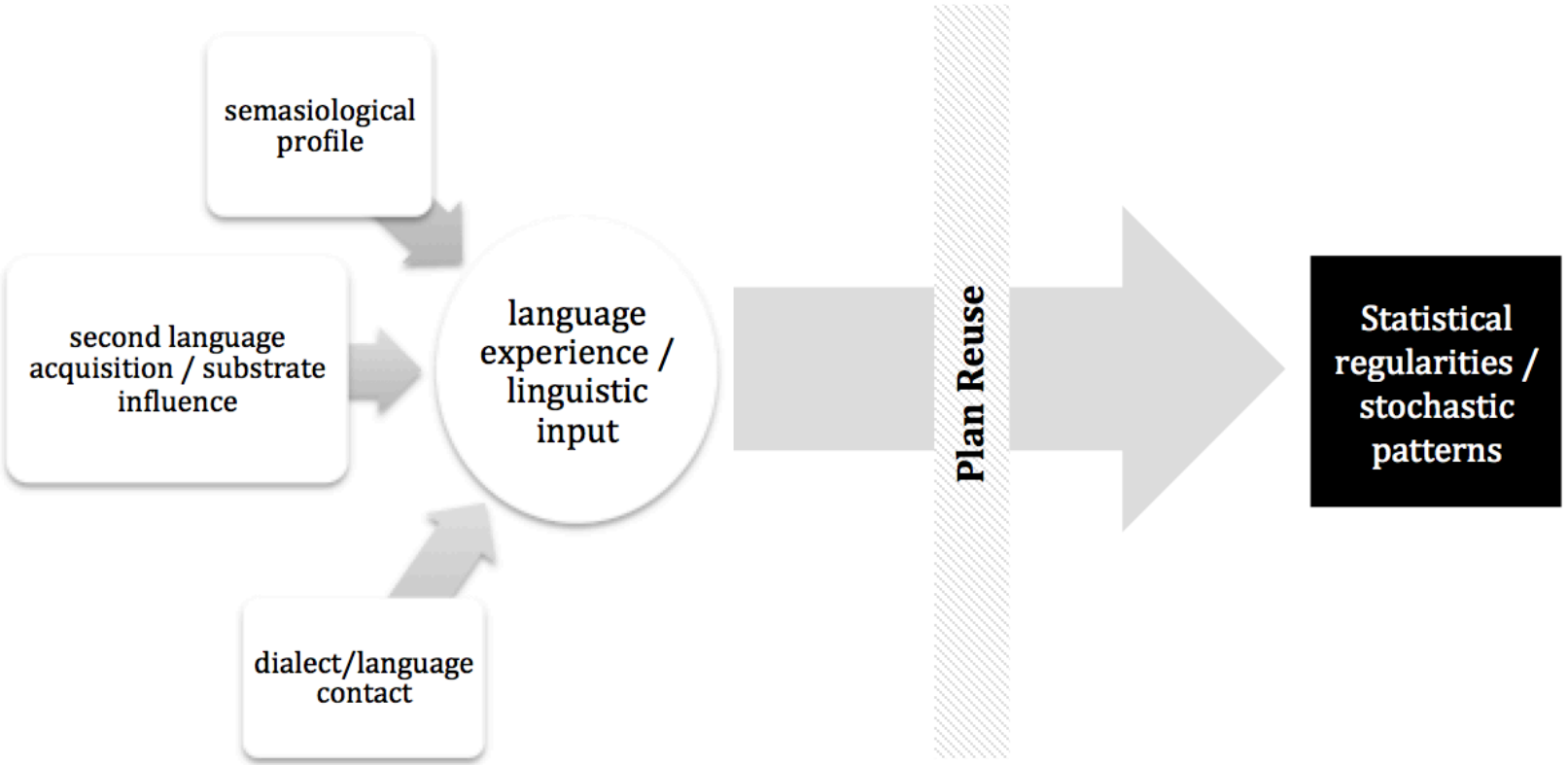
discussion

- MacDonald (2013): Easy First, Plan Reuse
  - consistent interplay between principles creates statistical regularities in language usage
  - Easy First: creates stability in effect direction
  - Plan Reuse: constantly reinforces the regularization of linguistic input $\rightarrow$ strengthens diverging statistical patterns of use

discussion

- MacDonald (2013): Easy First, Plan Reuse
  - consistent interplay between principles creates statistical regularities in language usage
  - Easy First: creates stability in effect direction
  - Plan Reuse: constantly reinforces the regularization of linguistic input $\rightarrow$ strengthens diverging statistical patterns of use
  - changes in lexis-syntax associations can result in diverging statistical regularities since the strength of effects that modulate these statistical regularities change as well

discussion


semasiological profile
second language acquisition / substrate influence
dialect/language contact
language experience / linguistic input
Plan Reuse
Statistical regularities / stochastic patterns

language and dialect contact

- emergence of localized linguistic structure with new lexical items in syntactic constructions

language and dialect contact

- emergence of localized linguistic structure with new lexical items in syntactic constructions
  - generalizing beyond the input

language and dialect contact

- emergence of localized linguistic structure with new lexical items in syntactic constructions
  - generalizing beyond the input
  - changes in abstract rules

second language acquisition

second language acquisition

- overuse of more transparent option (PD)
  $\rightarrow$ changes in the strength of specific cues as variants are used by L2 speakers in contexts where L1 speakers would not

second language acquisition

- overuse of more transparent option (PD)
  $\rightarrow$ changes in the strength of specific cues as variants are used by L2 speakers in contexts where L1 speakers would not
- transfer of cue strength from L1 (MacWhinney 1997)

constructional / semantic changes

- due to “normal” language usage

constructional / semantic changes

- due to “normal” language usage
- semasiological profile of variant might differ cross-variational

constructional / semantic changes

- due to “normal” language usage
- semasiological profile of variant might differ cross-variational
- 1st lang acq.: DO associated with certain lexical items

constructional / semantic changes

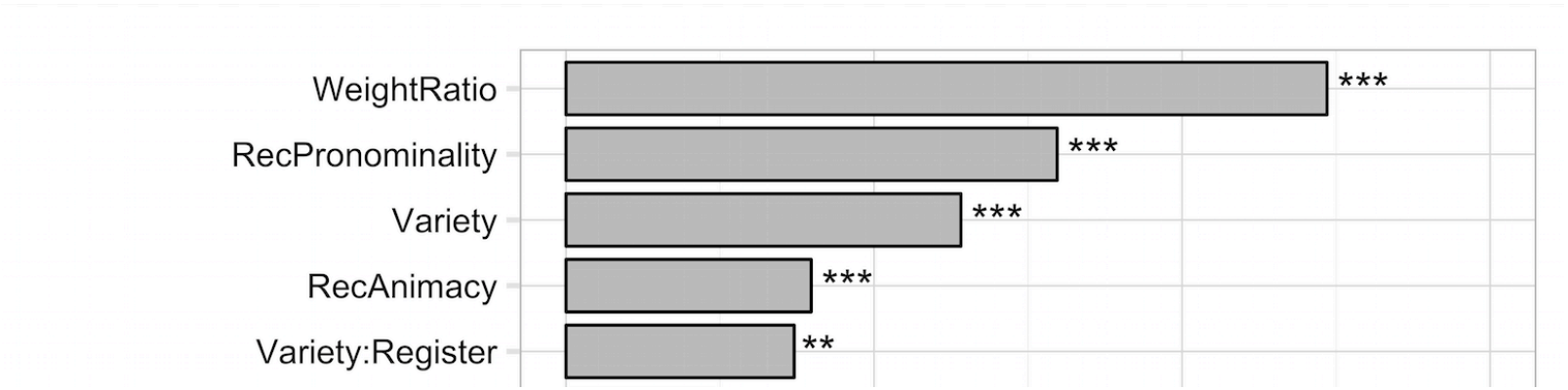
- due to “normal” language usage
- semasiological profile of variant might differ cross-variational
- 1st lang acq.: DO associated with certain lexical items
- 2nd lang acq.: DO is associated with certain lexical items

why length and RecPron?

- most amenable to probabilistic indigenization = length and recipient pronominality

why length and RecPron?

- most amenable to probabilistic indigenization = length and recipient pronominality
- most influential predictors = high cue validity



how does the study fit in with previous research?

how does the study fit in with previous research?

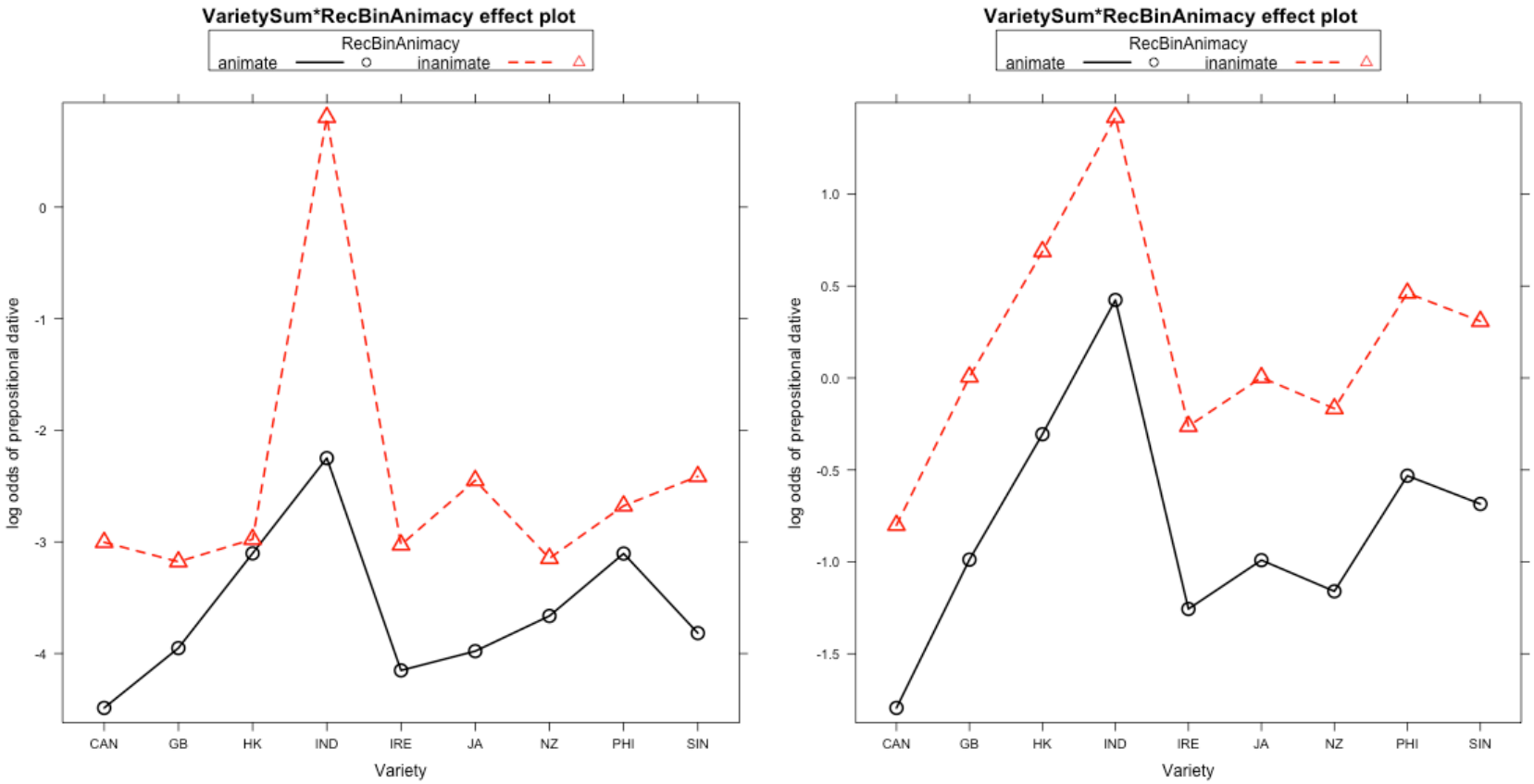
What about recipient animacy?

how does the study fit in with previous research?

Investigating the effect of recipient animacy:

- restrict dataset to *give*
- follow procedure in Bresnan and Hay 2008 in selection of predictors
- et voilà: $\rightarrow$ recipient animacy is a significant factor!

how does the study fit in with previous research?



(left: GIVE model; right: all verbs)

conclusion

- some effects seem to be sensitive to the lexical items that are used as syntactic constituents: verb-specific sensitivities vary across varieties

conclusion

- some effects seem to be sensitive to the lexical items that are used as syntactic constituents: verb-specific sensitivities vary across varieties
- variability in stochastic patterns emerges when speakers are exposed to diverging grammars (intergenerational and lifespan changes)

conclusion

- some effects seem to be sensitive to the lexical items that are used as syntactic constituents: verb-specific sensitivities vary across varieties
- variability in stochastic patterns emerges when speakers are exposed to diverging grammars (intergenerational and lifespan changes)
- . . . due to natural variation in the frequencies of specific lexical items, features and/or syntactic structures

conclusion

- some effects seem to be sensitive to the lexical items that are used as syntactic constituents: verb-specific sensitivities vary across varieties
- variability in stochastic patterns emerges when speakers are exposed to diverging grammars (intergenerational and lifespan changes)
- . . . due to natural variation in the frequencies of specific lexical items, features and/or syntactic structures
- diverging patterns of usage are constantly reinforced by Plan Reuse

conclusion

- some effects seem to be sensitive to the lexical items that are used as syntactic constituents: verb-specific sensitivities vary across varieties
- variability in stochastic patterns emerges when speakers are exposed to diverging grammars (intergenerational and lifespan changes)
- . . . due to natural variation in the frequencies of specific lexical items, features and/or syntactic structures
- diverging patterns of usage are constantly reinforced by Plan Reuse
- combining social as well as cognitive aspects is fruitful in order to more fully understand mechanisms of language production and comprehension

outlook

outlook

- focus on social constraint (Toronto)..
- ..and other syntactic alternations (Toronto)
- extend annotation (persistence)
- extend corpus material to include web-based language (GloWbE)
- separate analysis without pronouns?

unresolved issues

unresolved issues

1. Does cognitive indigenization also take place in other aspects of grammar (apart from syntax)?
2. The granularity of syntactic structure: to which extent is grammar tied to microCxs or specific lexical items?
3. How does the fact that L2 speakers are learners of English help us interpret the results?
4. How do substrate languages / creoles influence the effect that we observe?

Thank you!

melanie.rothlisberger@kuleuven.be

http://wwwling.arts.kuleuven.be/qlvl/ProbGrammarEnglish.html

references I

Bates, D., M. Maechler, and B. M. Bolker. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Bresnan, J., A. Cueni, T. Nikitina, and H. Baayen (2007). Predicting the Dative Alternation. pp. 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J. and M. Ford (2010). Predicting Syntax: Processing dative constructions in American and Australian Varieties of English. *Language 86*(1), 168–213.

Bresnan, J. and J. Hay (2008, feb). Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua 118*(2), 245–259.

Harrell, F. E. J. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.

Koch, P. and W. Oesterreicher (1985). Sprache der Nähe - Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In *Romanistisches Jahrbuch 36*, pp. 15–43. Berlin/New York: Mouton de Gruyter.

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

MacDonald, M. C. (2013, jan). How language production shapes language form and comprehension. *Frontiers in psychology 4*(April), 226.