Using Adverb Placement to Identify the Changing Position of American Possessive *Have*

Richard Zimmermann

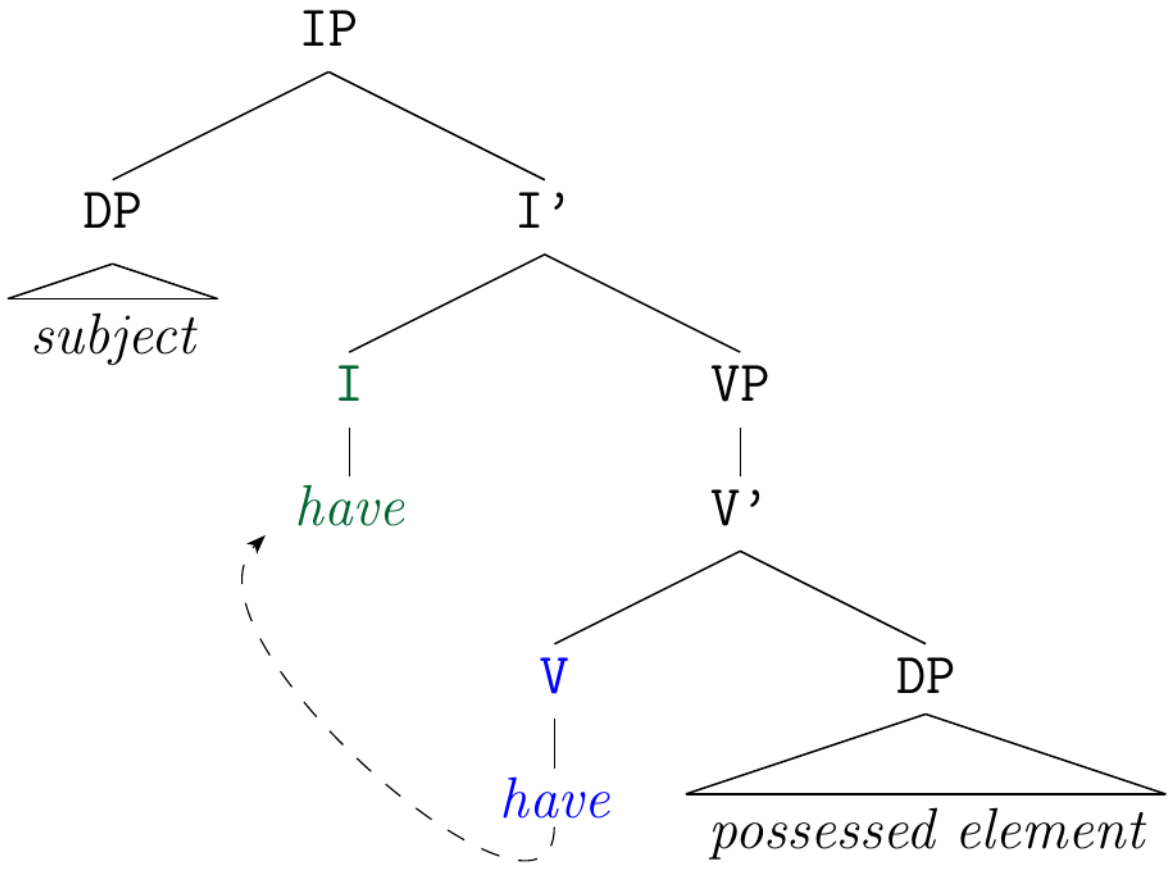
University of Geneva
SHES 2016

11 June 2016

Introduction

Late Modern American English Possessive *Have*

- loss of V-to-I (negation, inversion, VP-ellipsis)



Effects of the Change

- VP-adjoined adverbs are one way to diagnose the change

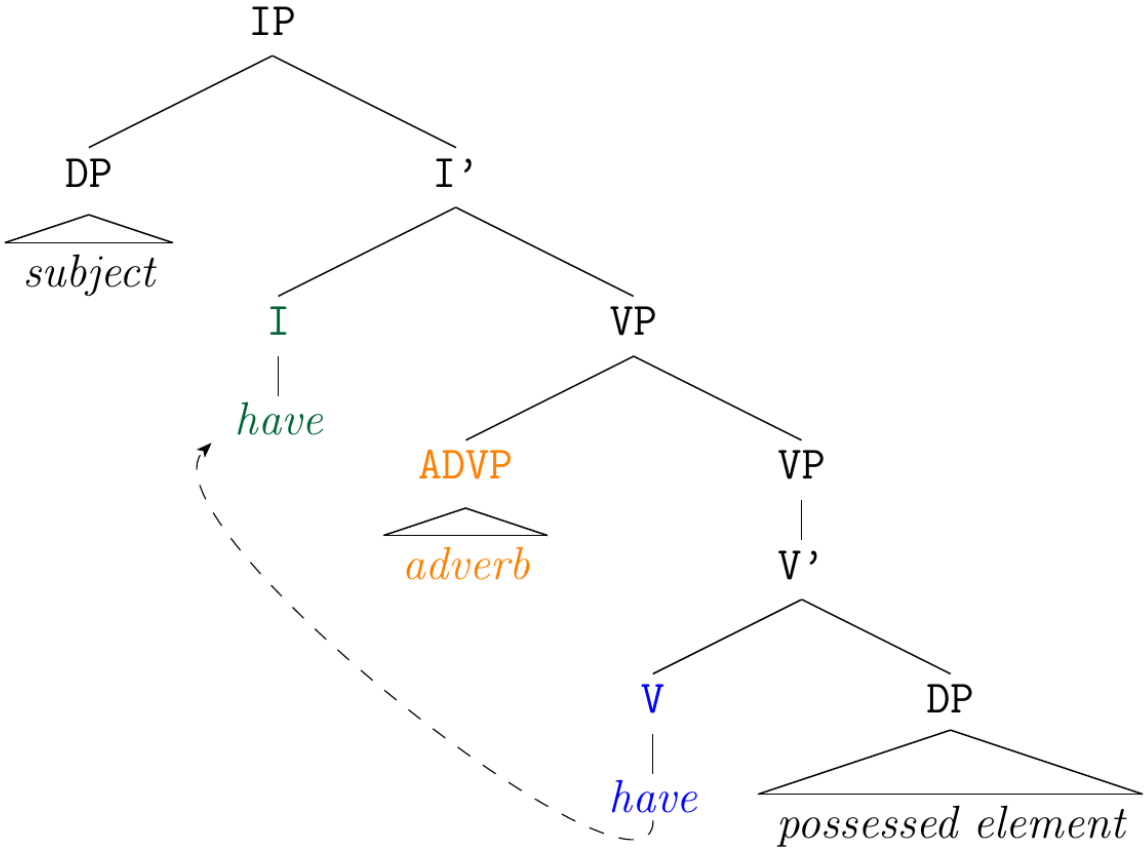


Illustration - temporal adverb *sometimes*

(1) a. *conservative variant: 'have'-adverb*
Phoenician temples **had** sometimes **adjuncts**, as cathedrals have their chapter-houses
HistoryPhoenicia, 1889

b. *innovative variant: adverb-'have'*
High school football rivalries in Texas sometimes **have a quality of tribal warfare**.
Time, 1971

Illustration - temporal adverb *always*

(2) a. *conservative variant: 'have'-adverb*
The preacher **had** always **a belief in witchcraft**
RachelDyerANorth, 1828

b. *innovative variant: adverb-'have'*
My father always **had** **a serene philosophy**
Atlantic, 1946

Illustration - epistemic adverb *necessarily*

(3) a. *conservative variant: 'have'-adverb*
A period of war, moreover, **has** necessarily **a tendency to strengthen the governmental action**.
USDemRev, 1838

b. *innovative variant: adverb-'have'*
The government, in any argument over facts, necessarily **has the last word**, since it can claim to have more facts.
Nation, 1965

Objective of this Talk

- assemble a large-scale quantitative dataset to measure the change in adverb placement

Objective of this Talk

- assemble a large-scale quantitative dataset to measure the change in adverb placement
- investigate the effect of high adverb adjunction

Objective of this Talk

- assemble a large-scale quantitative dataset to measure the change in adverb placement
- investigate the effect of high adverb adjunction
- taking this effect into consideration, compare the change in adverb placement to other contexts, such as negation and inversion

Measuring the Change

COHA

- *Corpus of Historical American English* (Davies 2010)

COHA

- *Corpus of Historical American English* (Davies 2010)
- 385m words; high register, formal, written, standard AE

COHA

- *Corpus of Historical American English* (Davies 2010)
- 385m words; high register, formal, written, standard AE
- string-based & POS-tagged

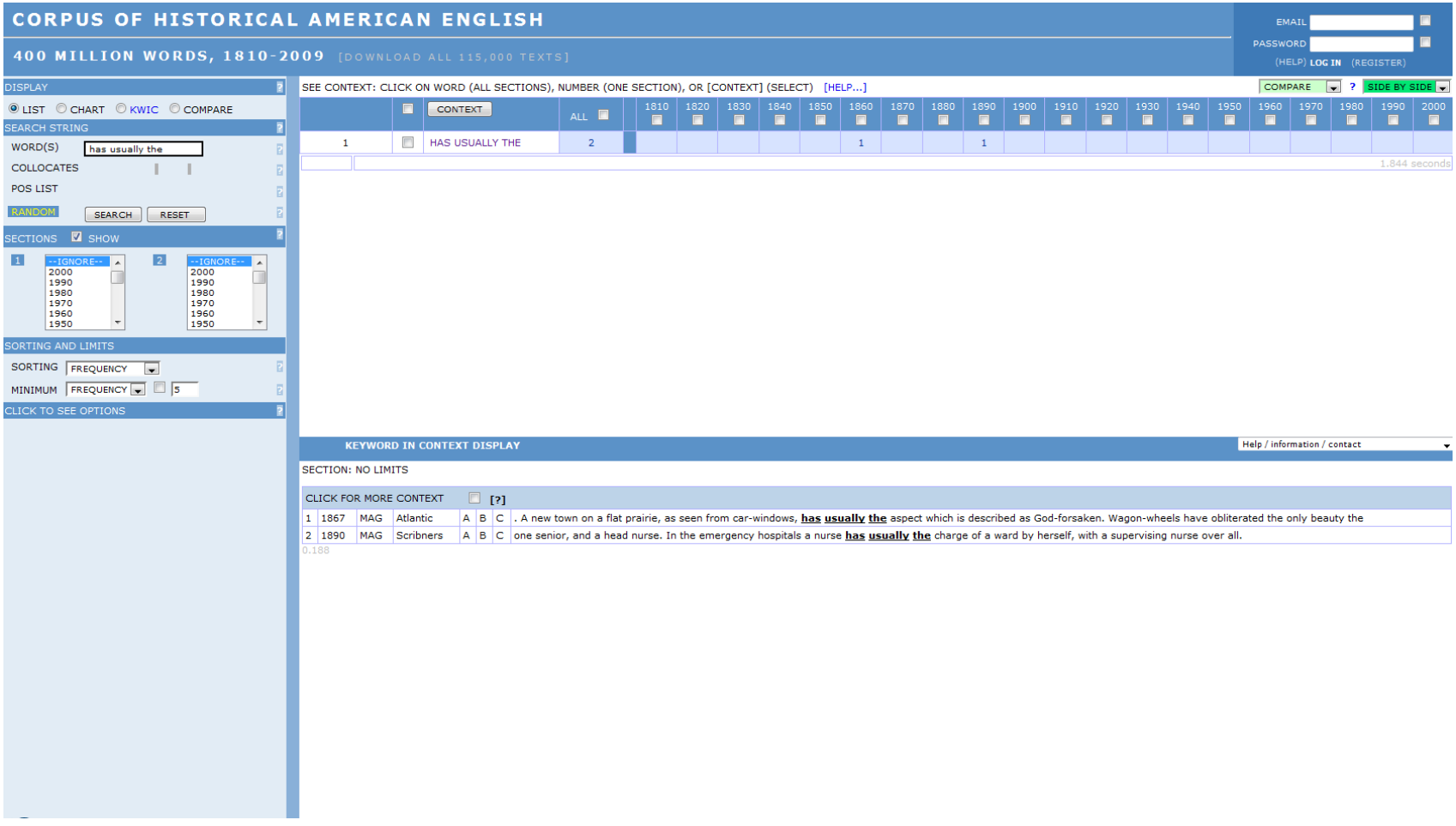
COHA

- *Corpus of Historical American English* (Davies 2010)
- 385m words; high register, formal, written, standard AE
- string-based & POS-tagged
- iVs: year (1810-2009), genre (FIC, NF, MAG, NEWS)

COHA

- *Corpus of Historical American English* (Davies 2010)
- 385m words; high register, formal, written, standard AE
- string-based & POS-tagged
- iVs: year (1810-2009), genre (FIC, NF, MAG, NEWS)
- problems:
  faulty OCR, search term restrictions, inaccurate POS-tagging, duplicate texts, British texts

COHA interface



Principles of search scripting

- Guided searches
- Symmetric searches
- Precision over Recall

		token found by search query	
		yes	no
token should have been found by search query	yes		*recall error*
	no	*precision error*	

- Large data

Queries

- dependent variable: '*have ADV*' vs. '*ADV have*' - 2
- independent variables:
  different inflections + object element
  - `have|has|had` - 3
  - object element was requierd to increase the likelihood of possessive *have* - 10
    (the, a, demonstrative, weak quantifiers, possessives etc.)
- examples of a search query string:

```
had usually the
really have this|these|that|those
has sometimes [MC*]
```

- 2*3*10 = 60 search queries per adverb

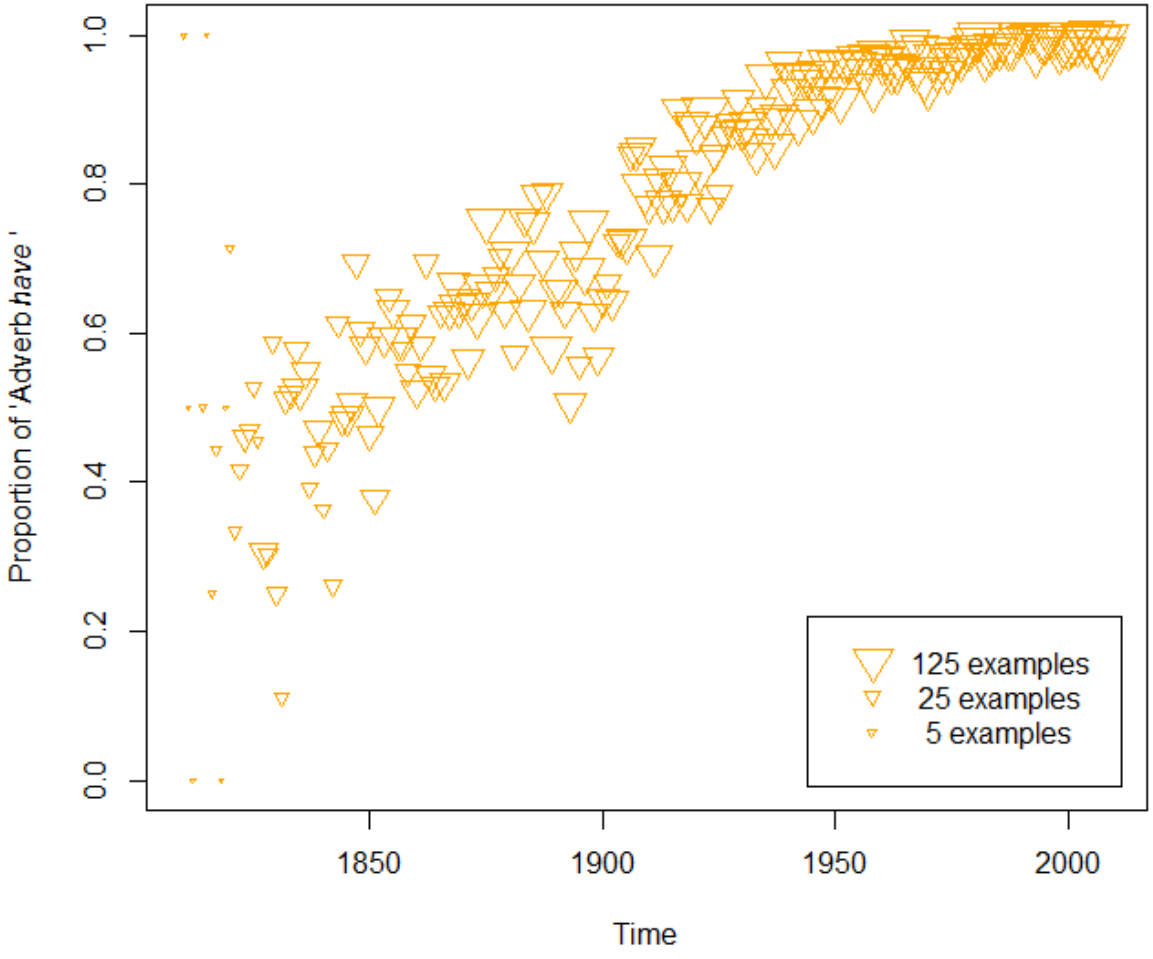
Manual correction

- precision errors were corrected by hand
- e.g., the search query `have also every|each|all|most|least|both|neither` can lead to false hits:
  - non-finite *have*
    But to **have also each** *of the premisses true is not merely to have syllogized*
  - targeted object is no an object
    My pensioners, the poorest, the oldest, the youngest, **have also every** *one received something that may be useful*
  - ... (causative *have*, perfect *have*, different category etc.)

Adverbs

- 26 temporal and epistemic adverbs
  actually, already, always, apparently, certainly, definitely, frequently, indeed, necessarily, never, nevertheless, no longer, now, obviously, of course, often, possibly, probably, rarely, really, seldom, sometimes, suddenly, surely, then, usually
- total of 18,704 examples

Development of adverb-*have* order, n=18,704



Logistic regression model with 'Time' as only predictor

- Intercept and coefficient for 'Time' variable

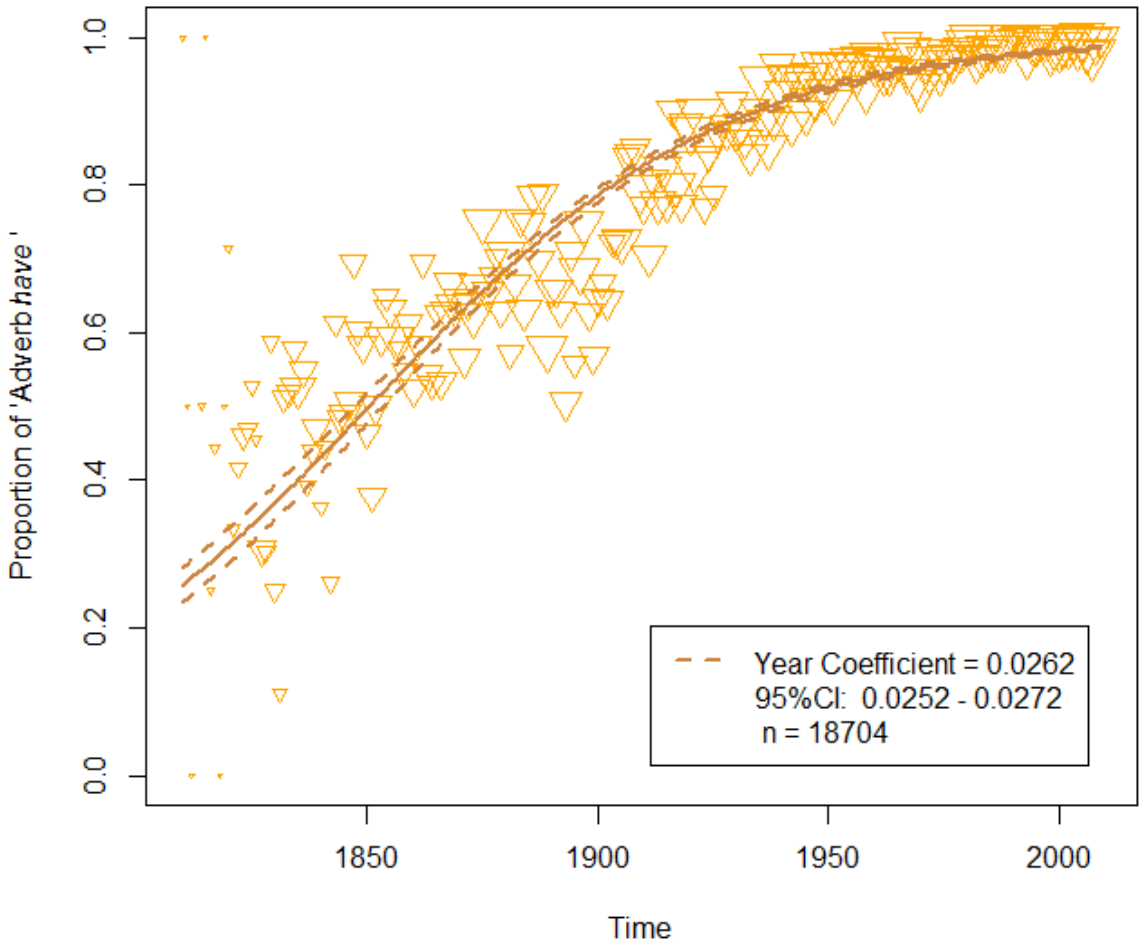
`formula = Adverb-Have ~ Time`

	Estimate	Std.Error	z-value	p
Intercept	-48.53	0.9531	-50.91	$<0.001^{***}$
Time	0.02622	0.0005031	52.13	$<0.001^{***}$

- 95% confidence intervals for 'Time' coefficient

	2.5%	97.5%
Time	0.0252	0.0272

Plot of Model (dark orange line)



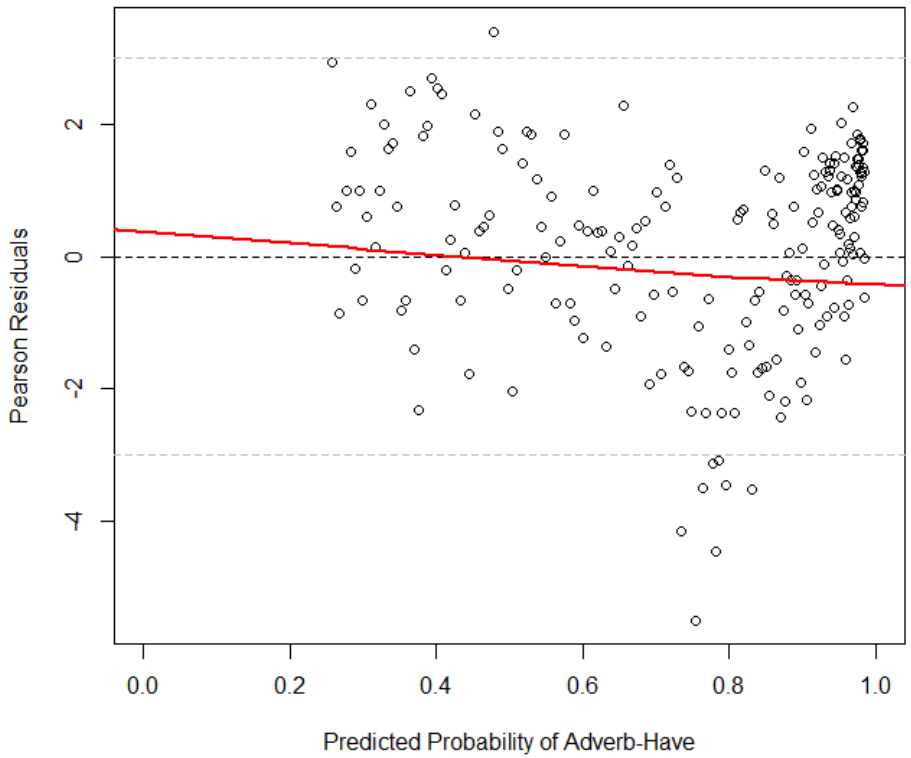
Model Fit

```
Null deviance:     4236.33 on 198 df
Residual deviance: 467.5   on 197 df
```

- Significance of the overall model:
  huge reduction in deviance; rejection of null hypothesis that the model is not better than chance at predicting the outcome; model is a significant fit to the data χ^2=3768.82, df = 1, p<0.001***
- Pseudo R^2: 'Time' predicts the outcome in a very good way
  Hosmer and Lemeshow R^2 = 0.89
- Low Predictive Accuracy: Model does not classify considerably more examples correctly than null model
  correct: 83.25%, baseline: 83.04%

Residual Plot

- residuals fall within a horizontal band; constant variance
- most values within +/- 3
- cluster of outliers between c. 1900-1950



Summary of Measurement

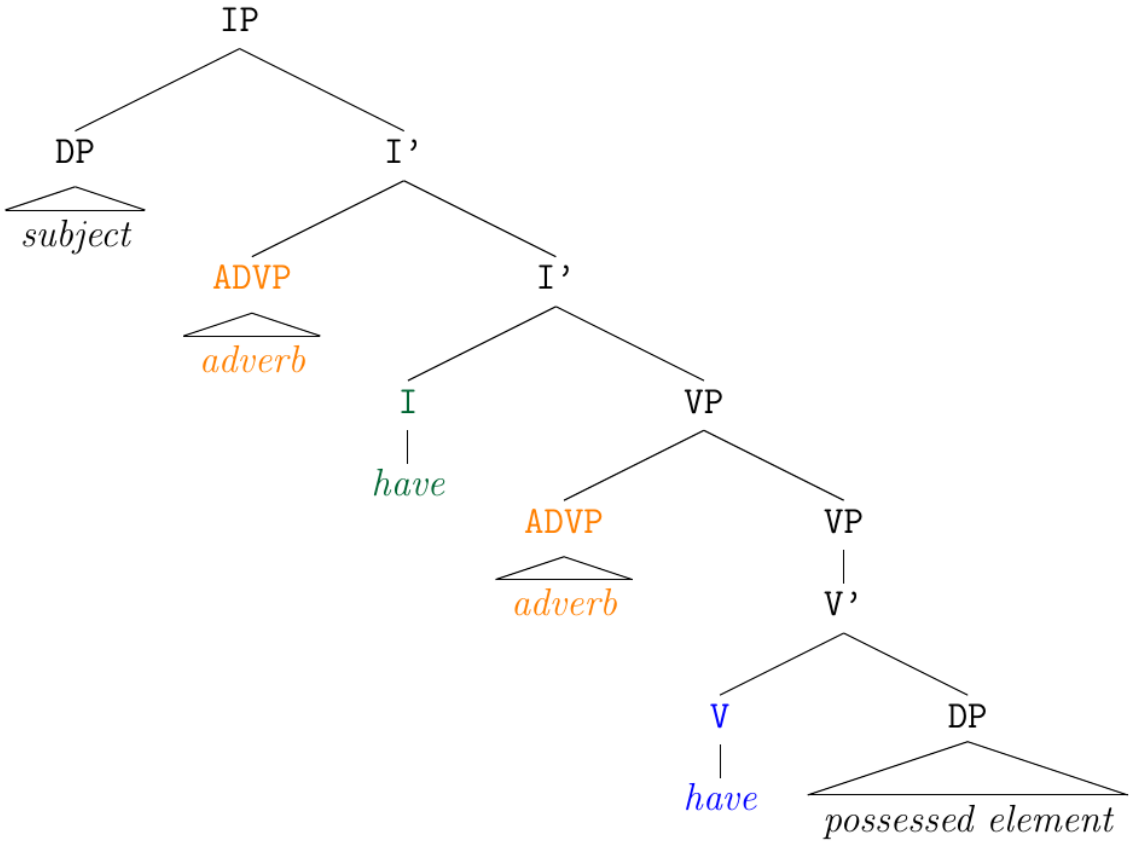
- the loss of V-to-I with possessive *have* surfaces as the loss of post-verbal adverbs
- there is in fact a clear change in relative order of *have* + adverb as predicted
- but: the change seems unreasonably show (it would take c. 400 years for the change to go to completion)
- skewed data; seemingly many more innovative than conservative *have*s; low predictivity
- a cluster of unexplained outliers in the mid 20th century

$\rightarrow$ confounding factor obscuring the true rate of change?

Effect of High Adjunction

Adverb Positions

- assume that adverbs can also occur before conservative *have*



Conservative *have* and high adverbs

(4) *presence of low and high adverb*
The poor Indians **now have often reason** to rejoice
NaturalistInNicaragua 1847

Conservative *have* and high adverbs

(5) *direct negation forces conservative* have + *high adverb:*
We **certainly have not the** solidarity and endurance of the Jews
Atlantic 1927

Conservative *have* and high adverbs

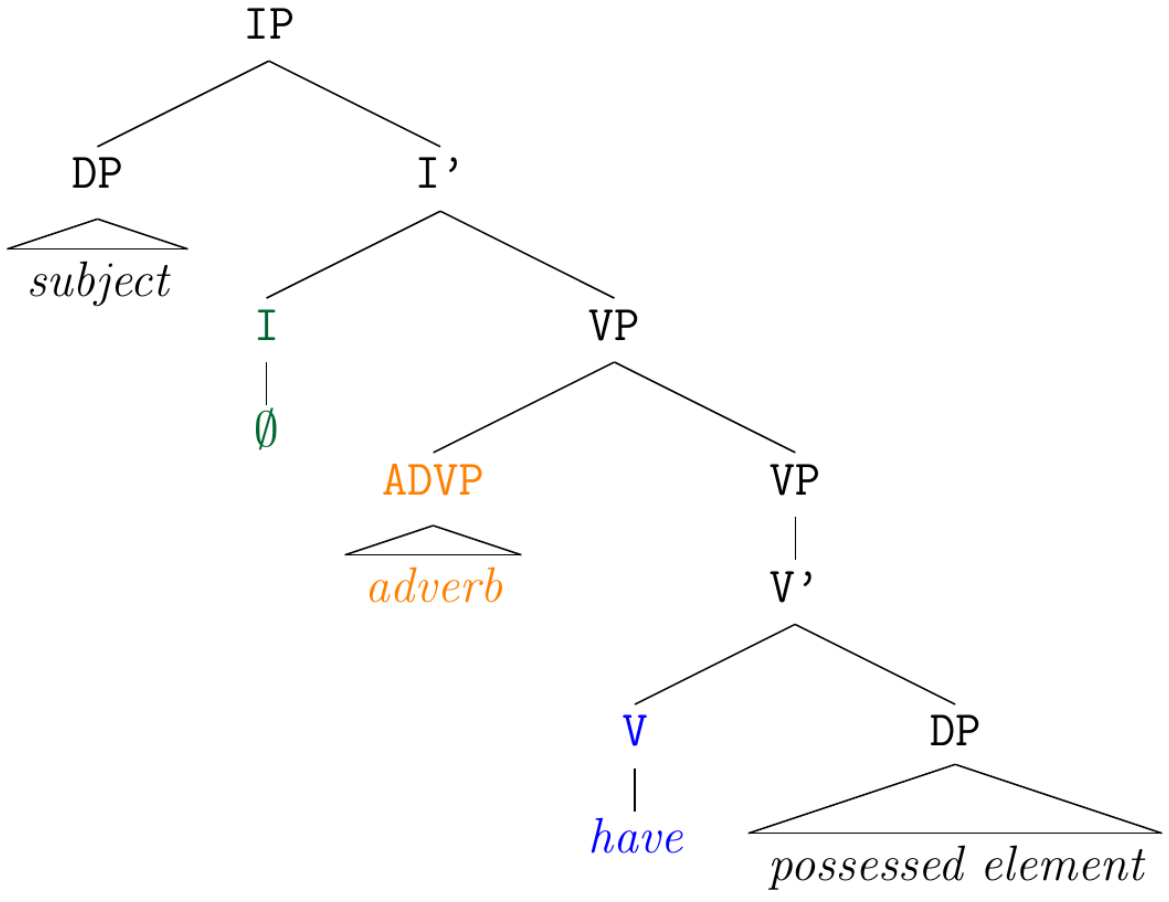
(6) *floating quantifier forces conservative* have + *high adverb:*
they **still had each** **a** strong following
Atlantic 1893

Conservative *have* and high adverbs

(7) *parallel adverb+modal suggests conservative* have + *high adverb:*
a woman **never can** obtain rank by merit, therefore **never has reason** to be proud of it.
LoversVows 1814

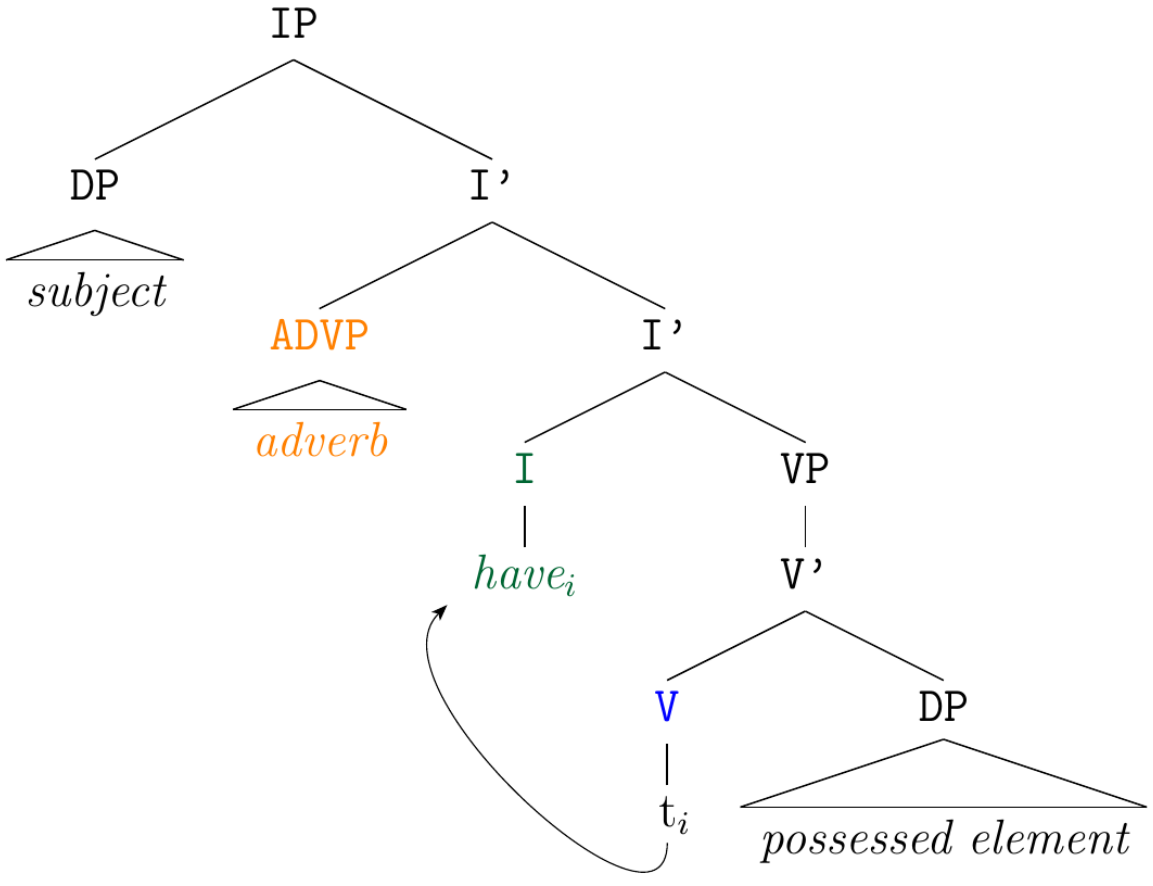
Structural ambiguity

- adverb-*have* with innovative variant

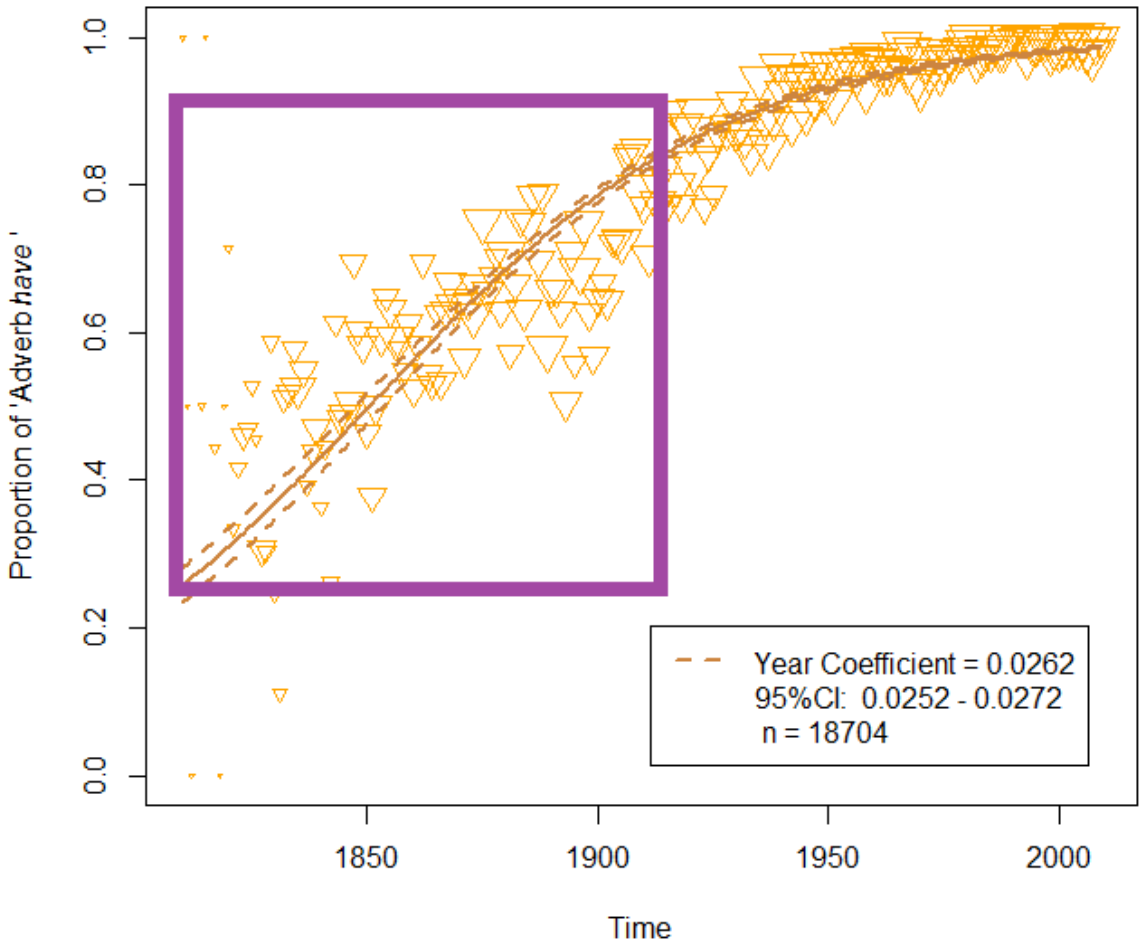


Structural ambiguity

- adverb-*have* with conservative variant + high adverb



Result of structural ambiguity



Correlation between *have* and other auxiliaries

- Hypothesis: If a substantial number of adverb-*have* sentences involve the conservative variant of possessive *have*,
  then, the propensity of an adverb to be adjoined to I' should correlate with the rate of change it measures.

How to determine probability of high adjunction?

- idea: instantiate I-position by another auxiliary
- compare pre-auxiliary and post-auxiliary position

Example of High Adjunction measurement - *never*

aux	[pp*]|[nn*] never aux	[pp*]|[nn*] aux never	% before aux
may	42	1,414	2.9
might	9	1,044	0.9
can	1,480	5,617	20.9
could	2,687	6,809	28.3
shall	614	4,155	12.9
should	580	3,387	14.6
will	1,322	8,151	14.0
would	2,008	11,643	14.7
must	12	945	1.3
is	139	2,923	4.5
are	94	1,674	5.3
was	1,824	7,921	18.7
were	324	2,459	11.6
TOTAL	11,135	58,142	16.1

Low Frequency Adverbs

- point estimates of rate of change for low-frequency items are very imprecise
- only use adverbs with more than 500 examples

Low Frequency Adverbs

adverb	lower	higher	range	*n*
possibly	-0.005	0.037	0.043	26
necessarily	0.011	0.057	0.046	39
definitely	0.023	0.142	0.119	53
nevertheless	0.014	0.036	0.022	107
frequently	0.012	0.038	0.026	108
of course	-0.006	0.009	0.015	109
surely	0.011	0.033	0.022	112
suddenly	0.036	0.098	0.061	138
rarely	0.011	0.034	0.023	146
indeed	0.012	0.026	0.015	162
obviously	0.025	0.051	0.025	181
seldom	0.027	0.057	0.031	196
apparently	0.034	0.058	0.024	284
actually	0.022	0.043	0.021	311
sometimes	0.014	0.028	0.014	318
then	0.021	0.030	0.010	549
usually	0.029	0.045	0.017	576
often	0.023	0.034	0.011	691
probably	0.019	0.028	0.010	752
certainly	0.018	0.029	0.011	773
really	0.020	0.029	0.009	876
no longer	0.037	0.047	0.010	1087
already	0.035	0.043	0.009	1732
always	0.019	0.024	0.005	2399
now	0.035	0.040	0.005	3381
never	0.016	0.024	0.008	3596

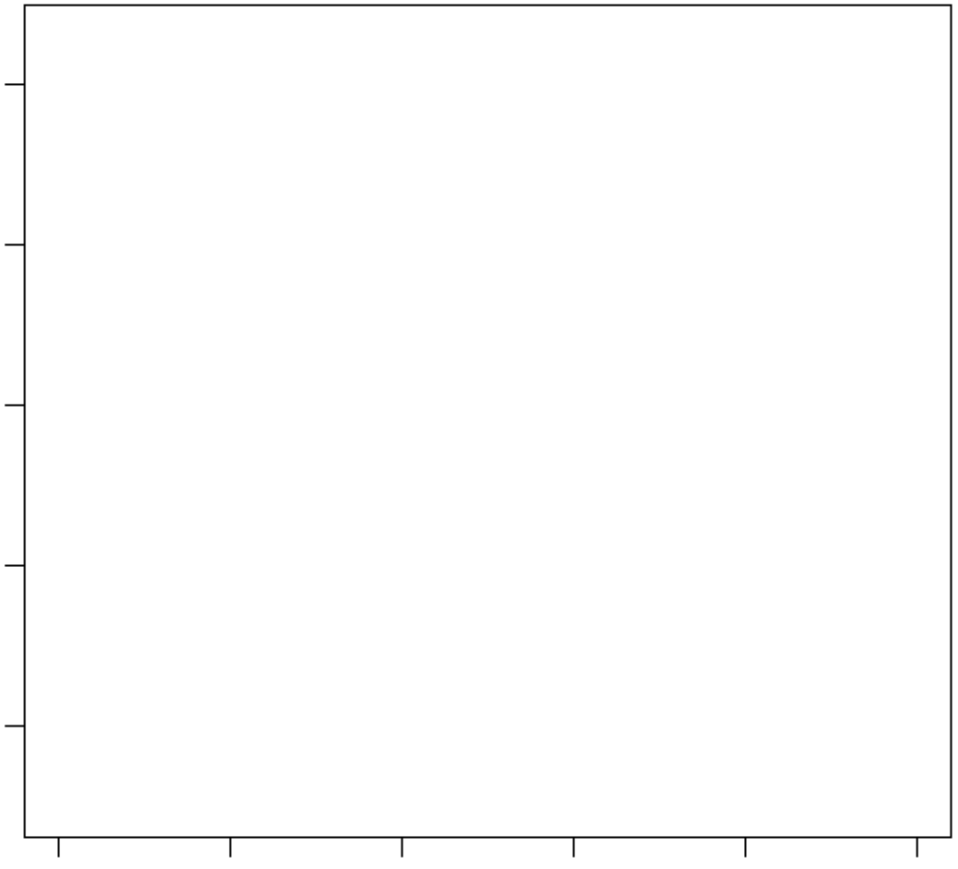
Low Frequency Adverbs

adverb	lower	higher	range	*n*
possibly	-0.005	0.037	0.043	26
necessarily	0.011	0.057	0.046	39
definitely	0.023	0.142	0.119	53
nevertheless	0.014	0.036	0.022	107
frequently	0.012	0.038	0.026	108
of course	-0.006	0.009	0.015	109
surely	0.011	0.033	0.022	112
suddenly	0.036	0.098	0.061	138
rarely	0.011	0.034	0.023	146
indeed	0.012	0.026	0.015	162
obviously	0.025	0.051	0.025	181
seldom	0.027	0.057	0.031	196
apparently	0.034	0.058	0.024	284
actually	0.022	0.043	0.021	311
sometimes	0.014	0.028	0.014	318
then	0.021	0.030	0.010	549
usually	0.029	0.045	0.017	576
often	0.023	0.034	0.011	691
probably	0.019	0.028	0.010	752
certainly	0.018	0.029	0.011	773
really	0.020	0.029	0.009	876
no longer	0.037	0.047	0.010	1087
already	0.035	0.043	0.009	1732
always	0.019	0.024	0.005	2399
now	0.035	0.040	0.005	3381
never	0.016	0.024	0.008	3596

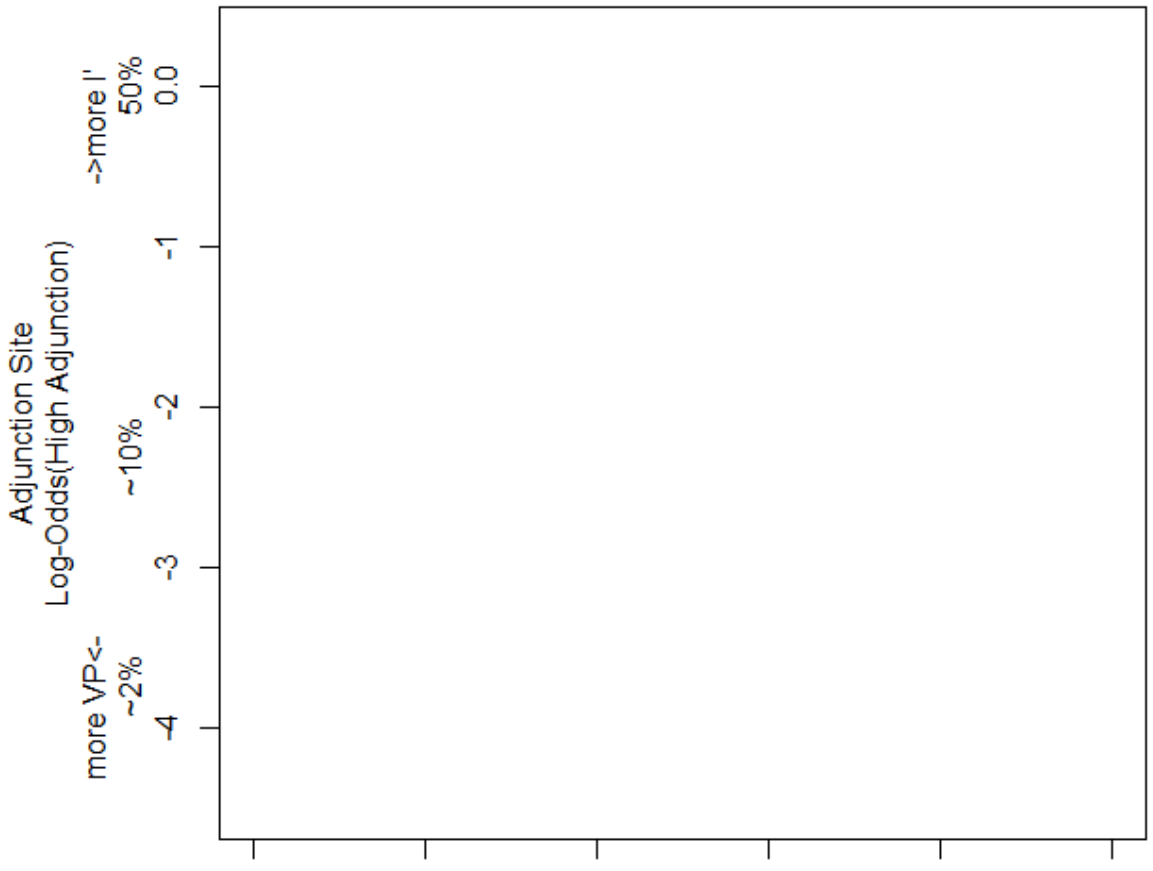
Low Frequency Adverbs

- point estimates of rate of change for low-frequency items are very imprecise
- only use adverbs with more than 500 examples
- 11/26 adverbs item 16,412 / 18,407 ($\sim$ 90%)

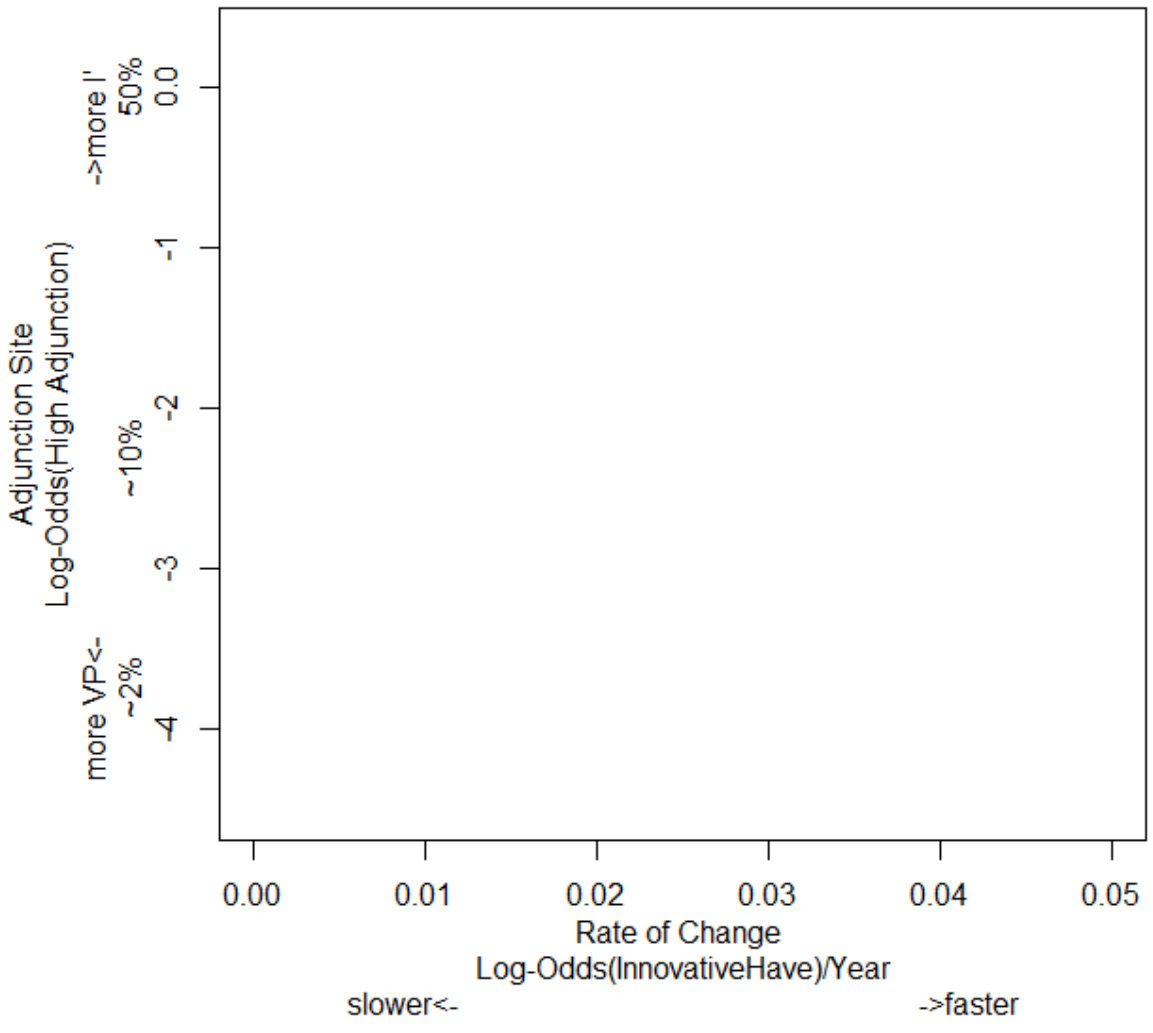
Correlation high adjunction $\sim$ rate of change



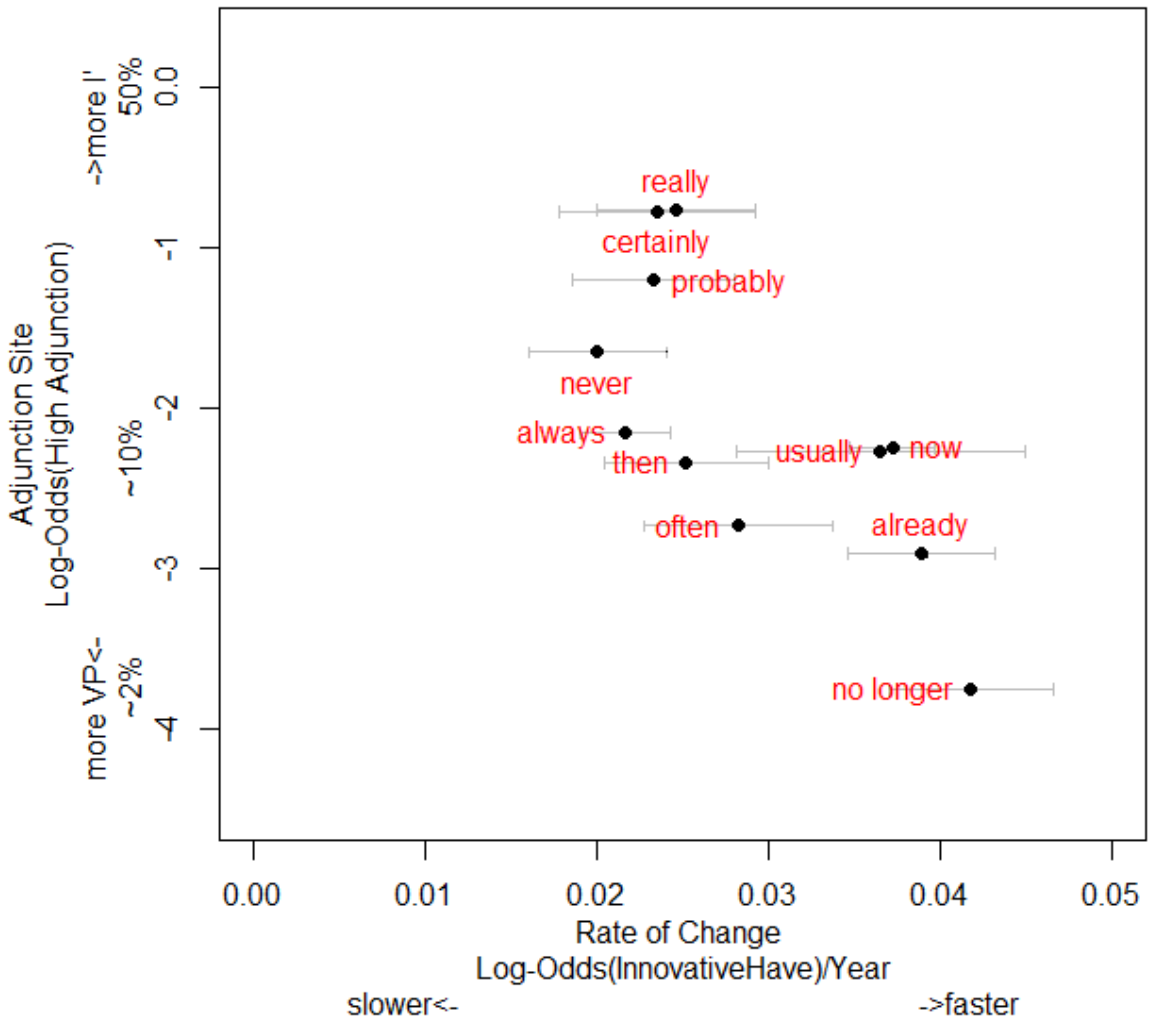
Correlation high adjunction ∼ rate of change



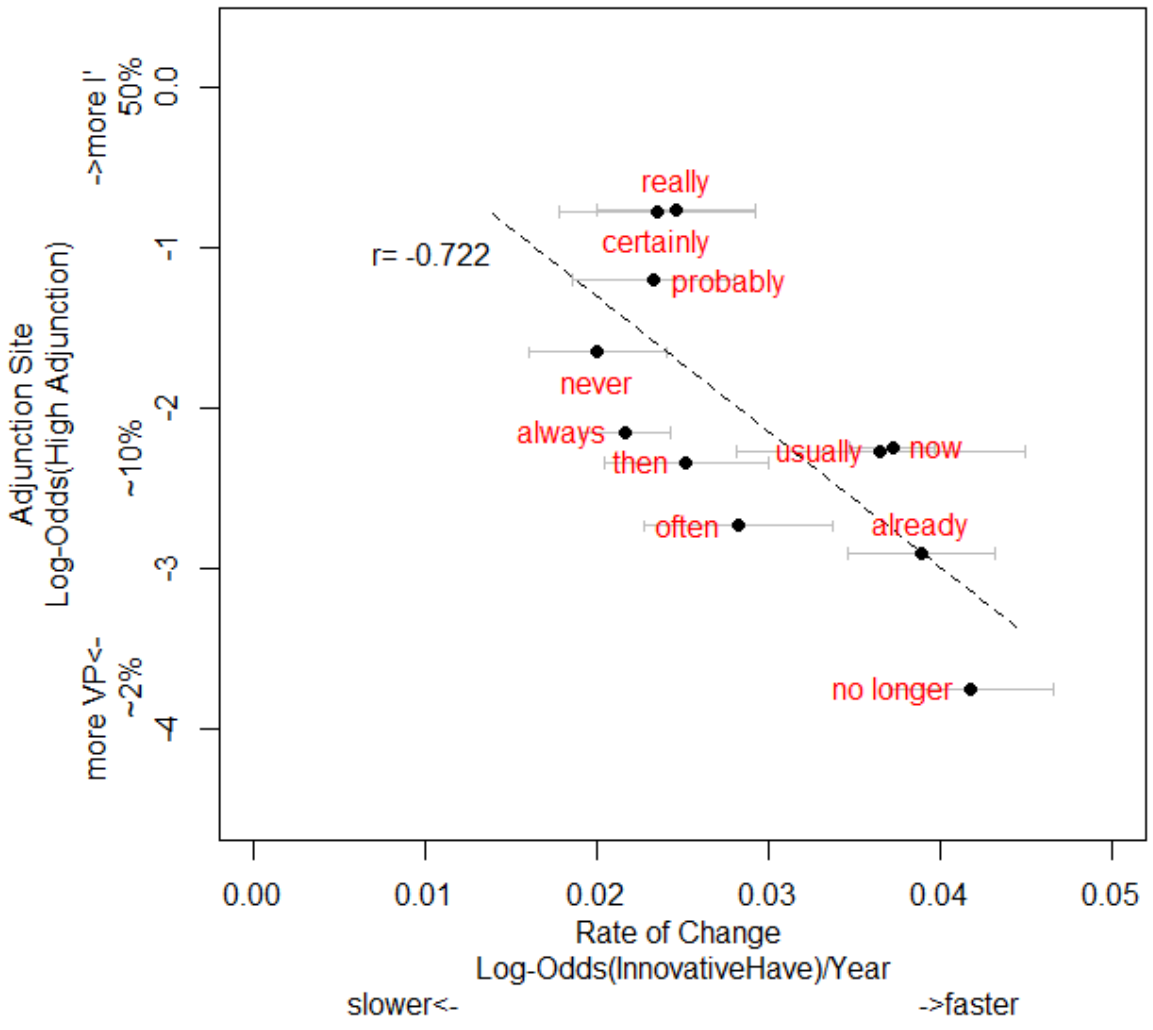
Correlation high adjunction ∼ rate of change



Correlation high adjunction ∼ rate of change



Correlation high adjunction ∼ rate of change



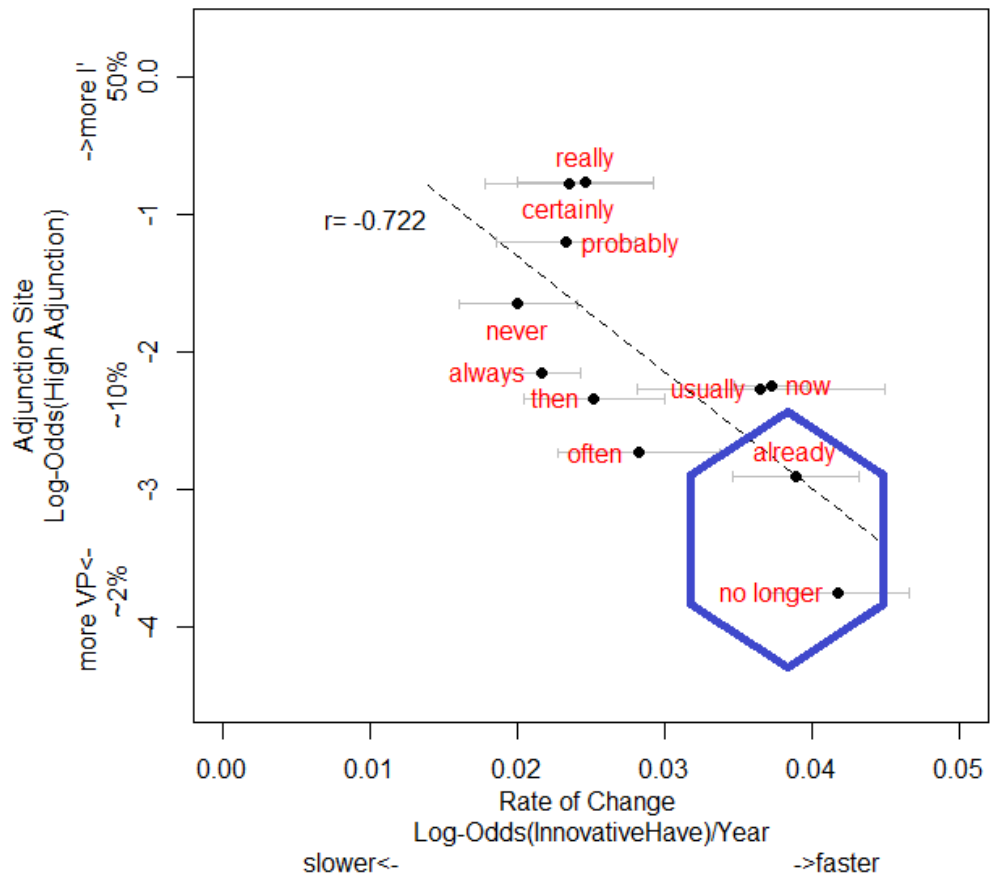
Summary of High Adjunction

- There is in fact a strong correlation between the (apparent) rate of change in adverb placement with possessive *have* and the propensity of an adverb to adjoin high.
- Hence, the actual rate of the loss of V-to-I movement is likely to be higher than the initial measurement suggested

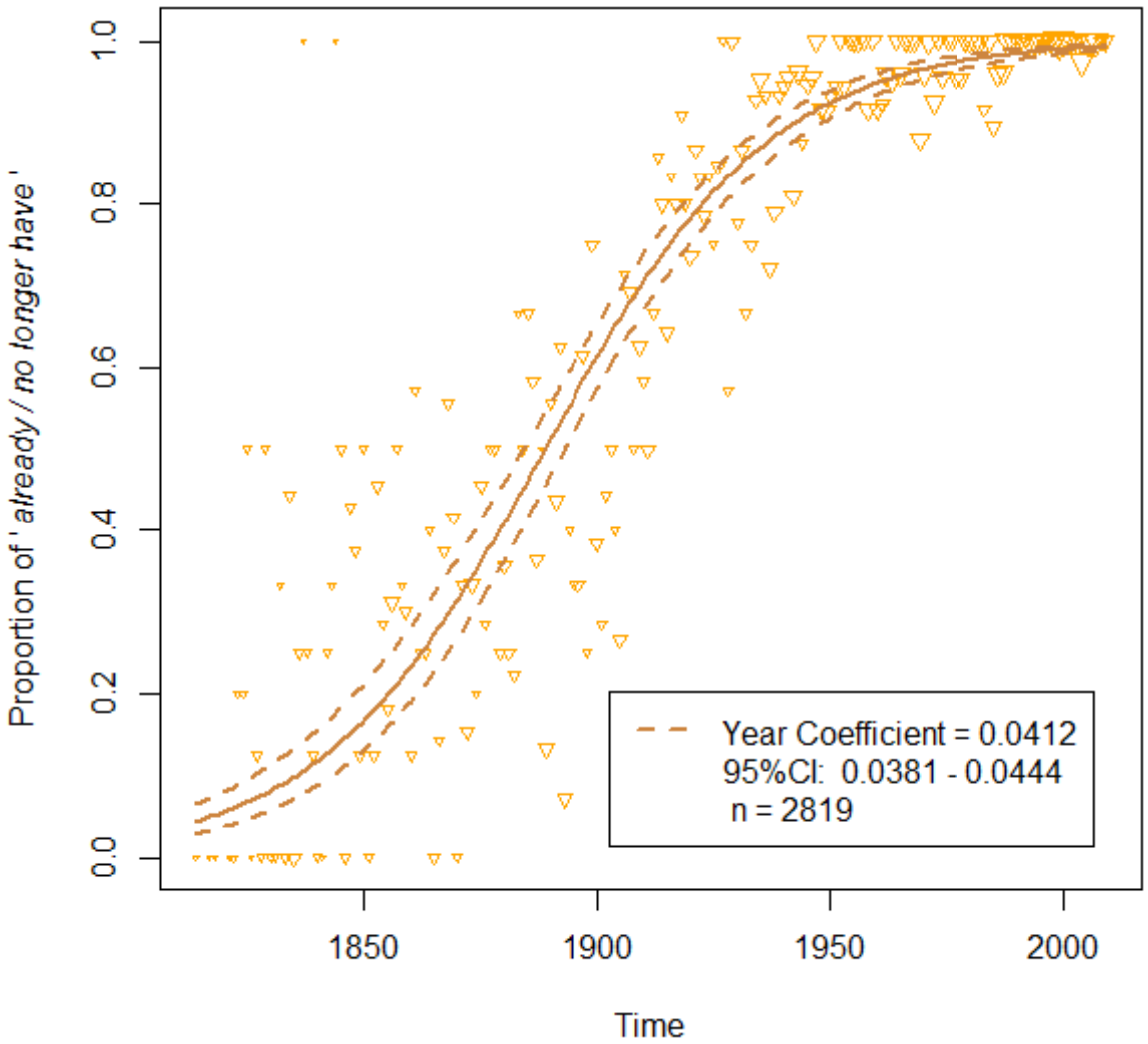
Comparison with other Contexts

Repeat measurement with VP-adverbs

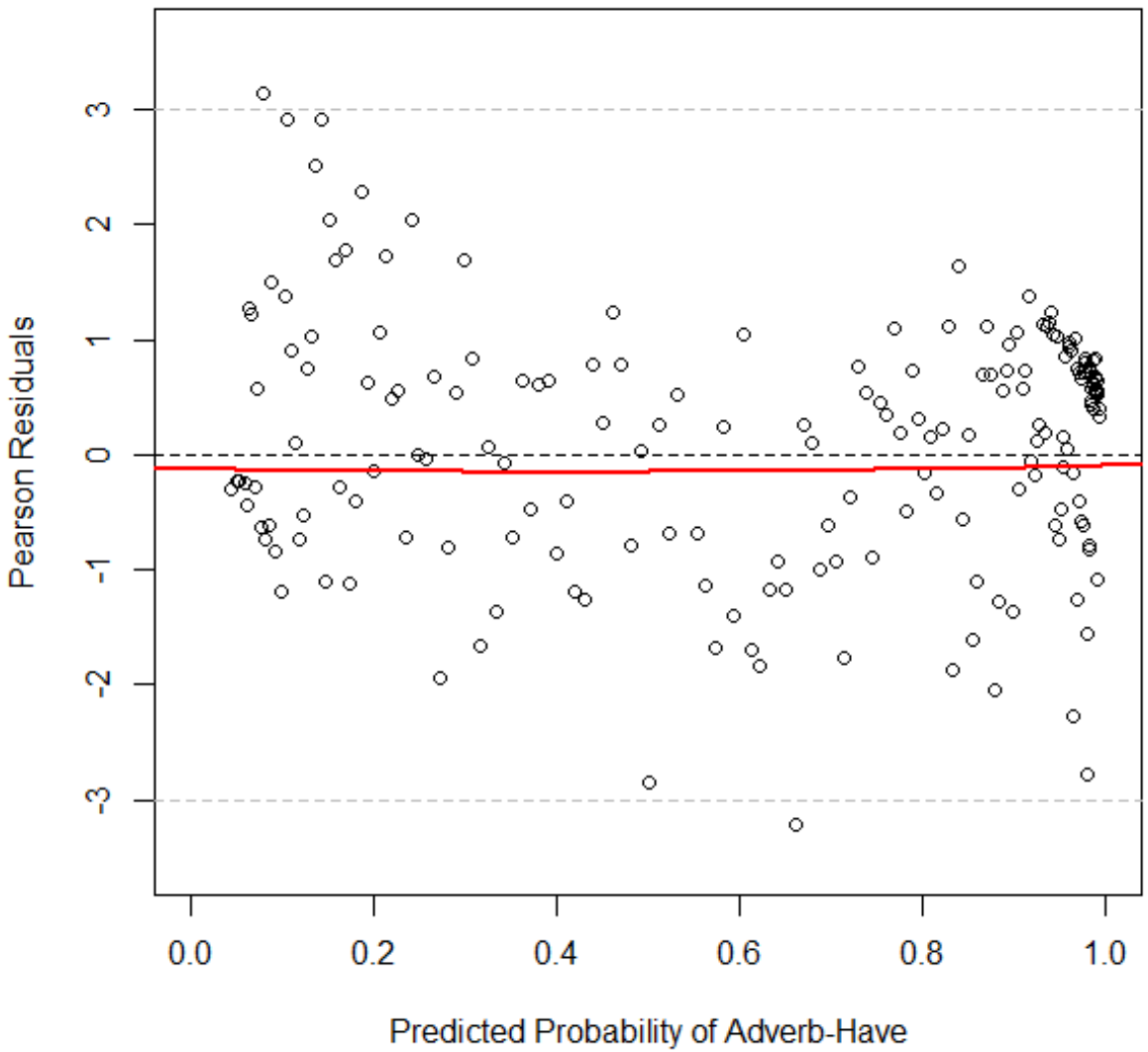
- *already* (High Adj. 5.2%), *no longer* (High Adj. 2.3%)



Development of adverb-*have* order revised



Residual plot for *already/no longer*



Model improvement

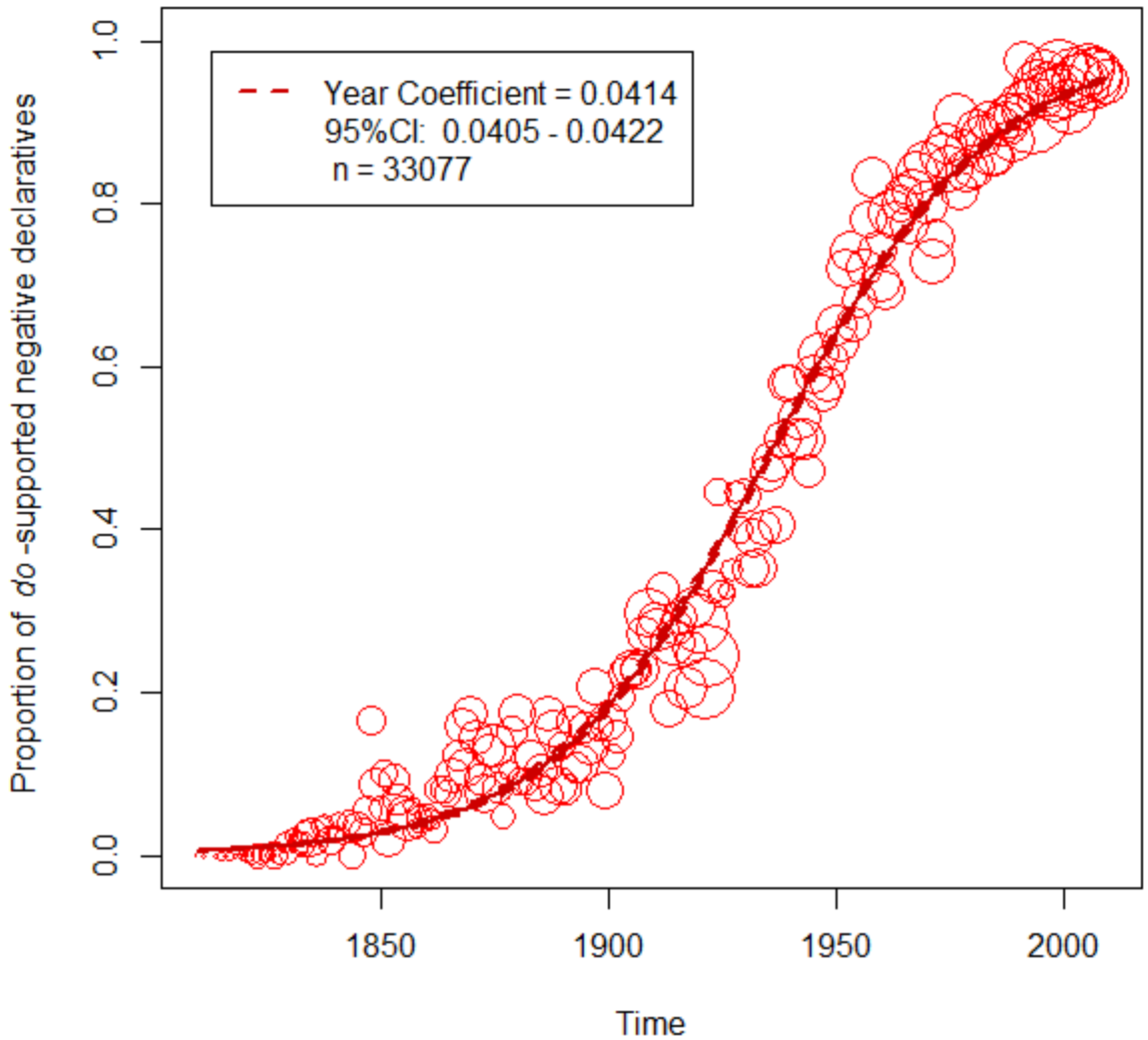
- rate of change now appears more realistic
  ($\sim$ 0.04 log-odds/year)
- outliers largely disappear
- predictivity improves considerably
  `correct:  86.73%, baseline:  79.50%`

Context: Negation

(8) a. *conservative variant: direct negation*
There, now, add the salt and pepper fixings, and the king himself **hasn't a slicker supper**.
A Romance of the Mohawk, 1840

b. *innovative variant:* do*-support*
The farming community of 900 people **doesn't have a single fast-food restaurant**.
Weight loss x 2, 2005

Development of negation context

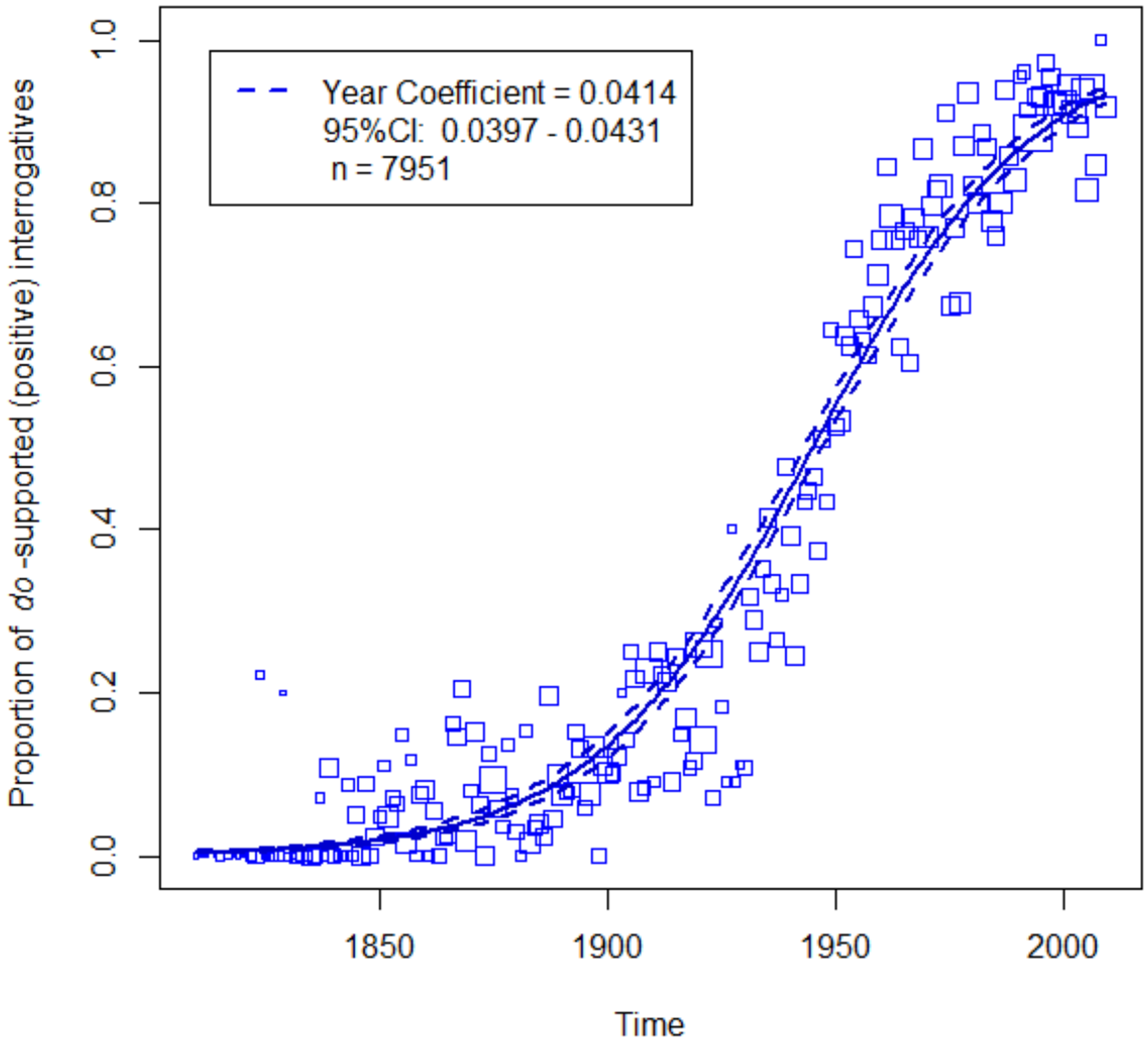


Context: Inversion

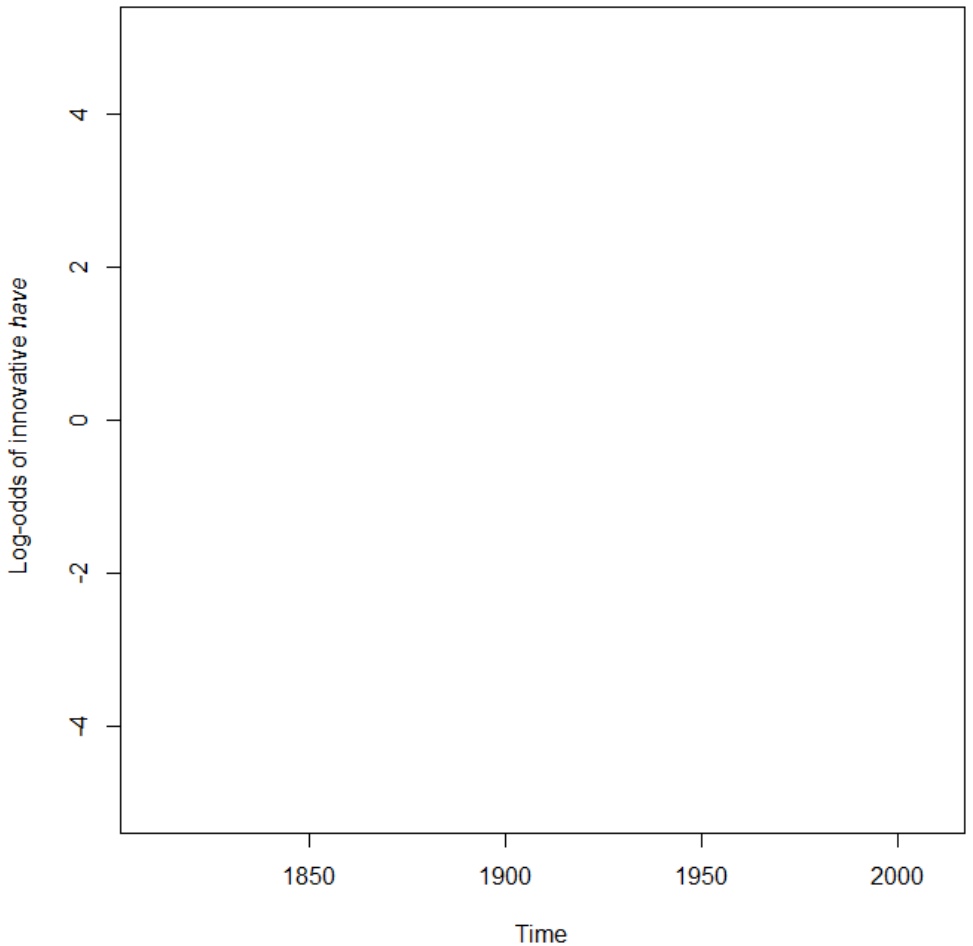
(9) a. *conservative variant: inversion with* have
Pray, how much money **have you**?
DukesPrizeStory, 1848

b. *innovative variant: inversion with* do-*support*
FEMALE INSPECTOR #2: "And how much money **do you have** ?"
Mov:MariaFullGrace, 2004

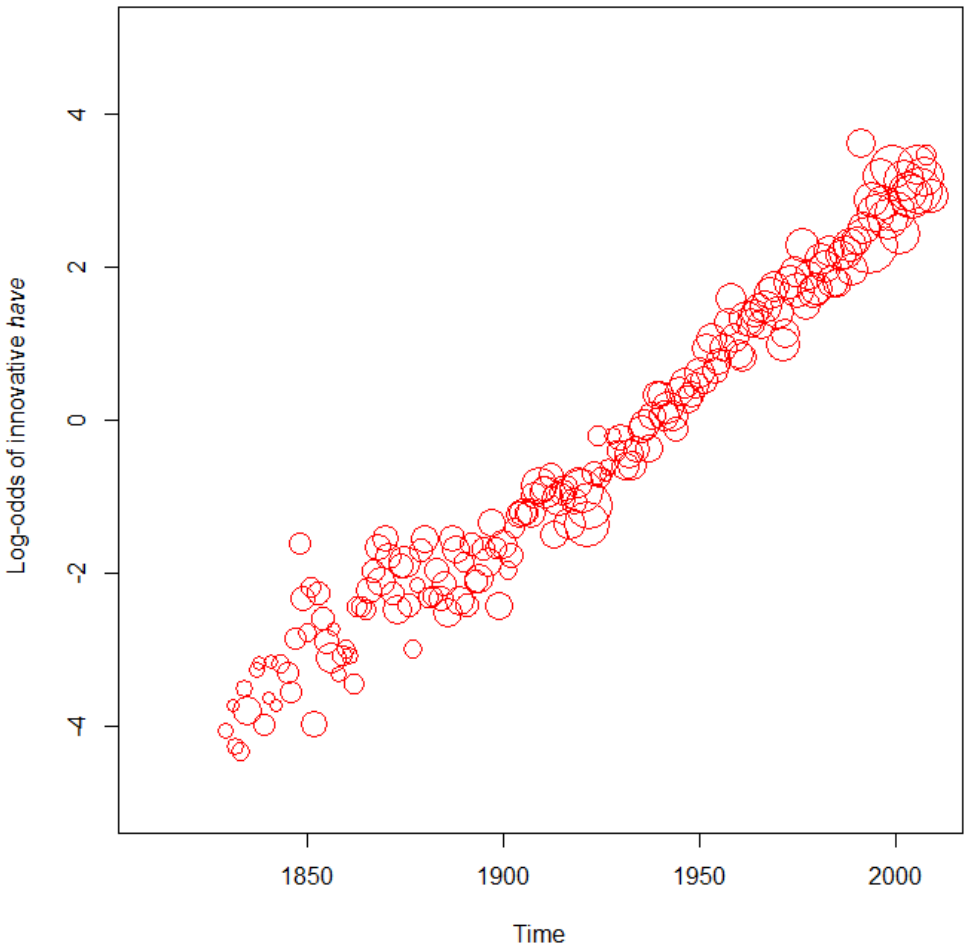
Development of inversion context



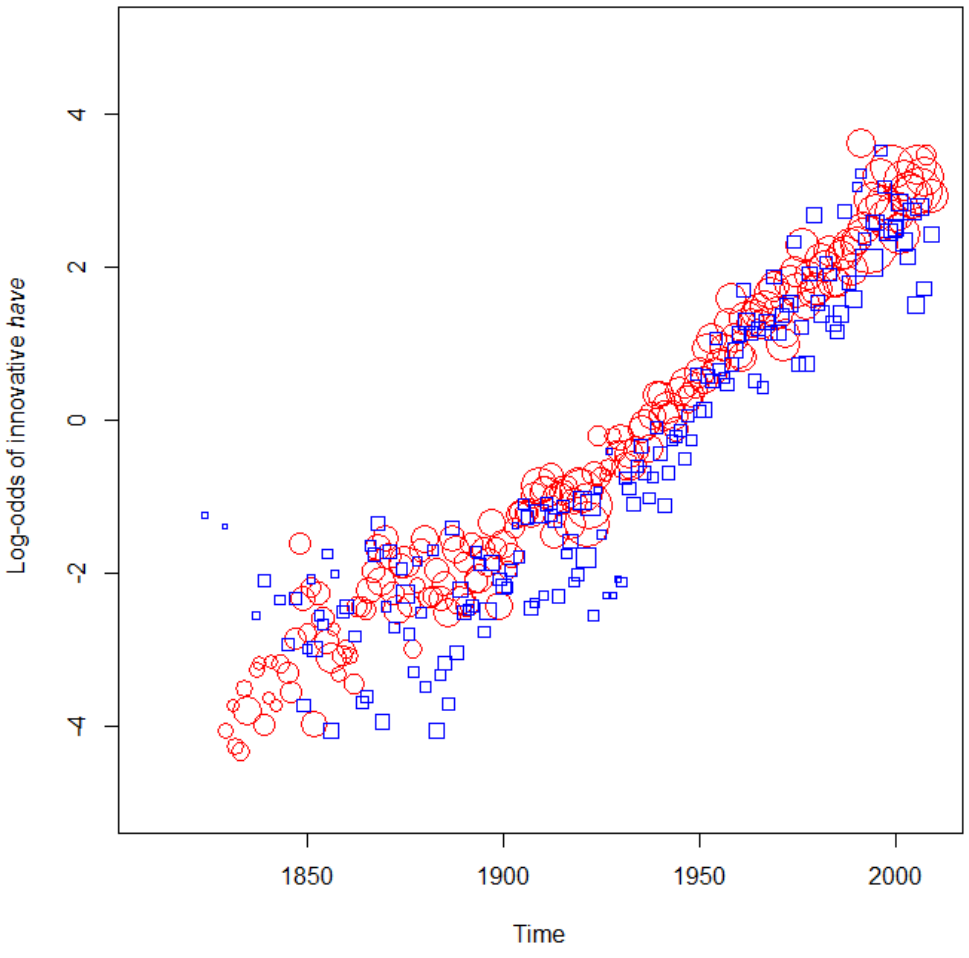
Comparison of different contexts



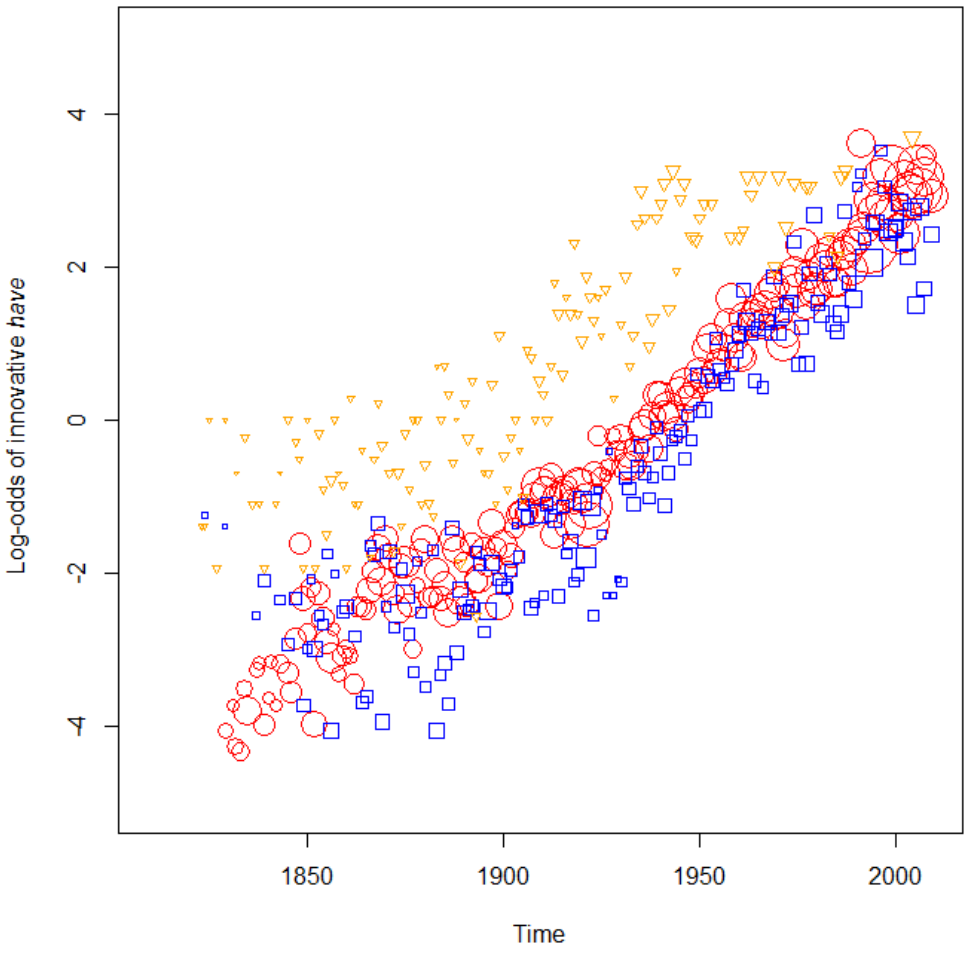
Comparison of different contexts



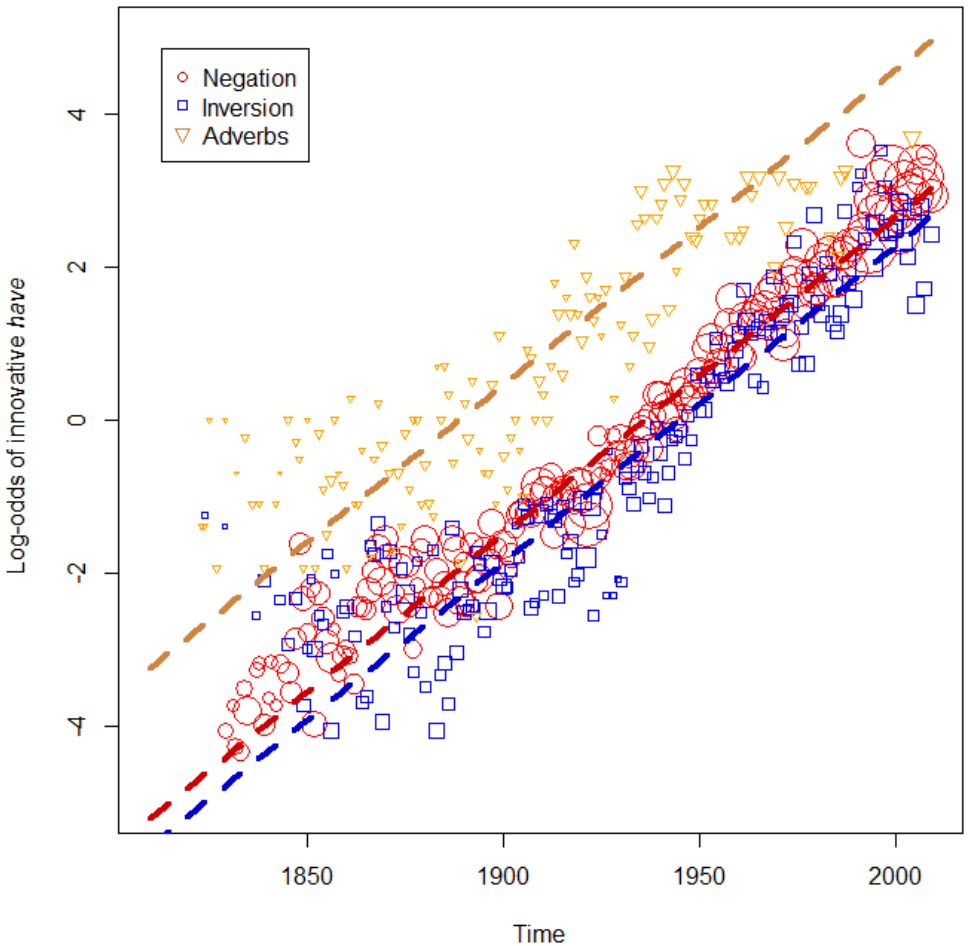
Comparison of different contexts



Comparison of different contexts



Comparison of different contexts



Test for significance of interaction effect

```
Model 1:
  InnovativeHave ~ Year + Context
Model 2:
  InnovativeHave ~ Year + Context + Year*Context
```

Model	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	584	1214.1			
2	582	1214.1	2	0.014959	0.9925

Table 1: Analysis of Deviance table for combined model

- interaction term does not significantly reduce deviance
- there is no good reason to assume that the rates of change are significantly different between the different contexts

Summary of Context Comparison

- when the effect of high adjunction is taken into account, the adverb change proceeds at the same rate of change as the negation and inversion changes with possessive have

Support for the Constant Rate Hypothesis

Support for the Constant Rate Hypothesis

- Possessive *have* loses V-to-I movement in 19^{th} - 21^{st} century American English.

Support for the Constant Rate Hypothesis

- Possessive *have* loses V-to-I movement in 19^{th} - 21^{st} century American English.
- This analysis entails that there should be a change in the relative order of adverbs and possessive *have*.

Support for the Constant Rate Hypothesis

- Possessive *have* loses V-to-I movement in 19^{th} - 21^{st} century American English.
- This analysis entails that there should be a change in the relative order of adverbs and possessive *have*.
- It is in fact true that such a change took place in late Modern American English.

Support for the Constant Rate Hypothesis

- Possessive *have* loses V-to-I movement in 19^{th} - 21^{st} century American English.
- This analysis entails that there should be a change in the relative order of adverbs and possessive *have*.
- It is in fact true that such a change took place in late Modern American English.
- The possibility of high adverb adjunction reduces the quality of the adverb diagnostic.
  $\rightarrow$ strong correlation between the frequency of adverbs in pre-auxiliary position and their apparent rate of change.

Support for the Constant Rate Hypothesis

- Possessive *have* loses V-to-I movement in 19^{th} - 21^{st} century American English.
- This analysis entails that there should be a change in the relative order of adverbs and possessive *have*.
- It is in fact true that such a change took place in late Modern American English.
- The possibility of high adverb adjunction reduces the quality of the adverb diagnostic.
  $\rightarrow$ strong correlation between the frequency of adverbs in pre-auxiliary position and their apparent rate of change.
- The adverbs *no longer* and *already* have the highest probability of adjoining to VP and are therefore the best diagnostics.

Support for the Constant Rate Hypothesis

- Possessive *have* loses V-to-I movement in 19^{th} - 21^{st} century American English.
- This analysis entails that there should be a change in the relative order of adverbs and possessive *have*.
- It is in fact true that such a change took place in late Modern American English.
- The possibility of high adverb adjunction reduces the quality of the adverb diagnostic.
  $\rightarrow$ strong correlation between the frequency of adverbs in pre-auxiliary position and their apparent rate of change.
- The adverbs *no longer* and *already* have the highest probability of adjoining to VP and are therefore the best diagnostics.
- When these adverbs are used, the rate of change appears to be identical to those of negation and inversion.

Thank you for your attention!

> "... when one grammatical option replaces another with which it is in competition across a set of linguistic contexts, the rate of replacement, **properly measured**, is the same in all of them." (Kroch 1989)

Appendix

Floating quantifiers - *each, all, both*

(10) a. Fichte, Schelling, Hegel, **had each** his own system, though they have been called transcendentalists.
NewEngYaleRev 1843

b. His mother, Fraulein Schlote and Miss Letitia Lamb **each had** her own accent and intonation
LastPuritan 1936

(11) a. the red limit and the violet limit **have both** the same luminous intensity
TreatiseOnForces 1845

b. Arcturus and Capella **both have** the same magnitude
ExploringDistant 1956

- total of 595 examples