

The Development of do-Support with Negative Possessive *have* in the History of American English

Richard Zimmermann
University of Geneva

NWAV 43

25 October 2014

Outline

From HAVE NOT to DO NOT HAVE

Variation in negative possessive *have*

Category change, not lexical variation

Data Collection

Testing of Hypotheses

Hypothesis I: S-shaped trajectory

Hypothesis I: Results

Hypothesis II: Constant Rate Effect

Hypothesis II: Results

Conclusion

The phenomenon under investigation

- ▶ *have* + NP, not auxiliary *have*

The phenomenon under investigation

- ▶ *have* + NP, not auxiliary *have*
- ▶ concrete (*have a car*) and abstract possession (*have an idea*)

The phenomenon under investigation

- ▶ *have* + NP, not auxiliary *have*
- ▶ concrete (*have a car*) and abstract possession (*have an idea*)
- ▶ variation: negation of possessive *have* can be
 - (i) direct (*She hasn't the faintest idea*)
 - (ii) or with *do*-support (*She doesn't have a clue*)

The phenomenon under investigation

- ▶ *have* + NP, not auxiliary *have*
- ▶ concrete (*have a car*) and abstract possession (*have an idea*)
- ▶ variation: negation of possessive *have* can be
 - (i) direct (*She hasn't the faintest idea*)
 - (ii) or with *do*-support (*She doesn't have a clue*)
- ▶ diachronic variation: *do*-support gradually becomes the dominant form during the 19th and 20th centuries

Illustration

- (1) 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

Analysis

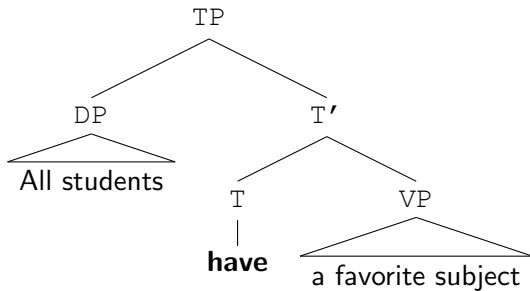
- ▶ the change does not simply involve a lexical change from *have not* to *do not have*

Analysis

- ▶ the change does not simply involve a lexical change from *have not* to *do not have*
- ▶ instead finite, possessive *have* changes category from T to V

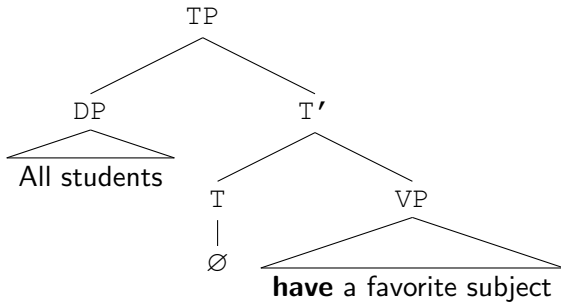
Analysis

(2)



Analysis

(3)

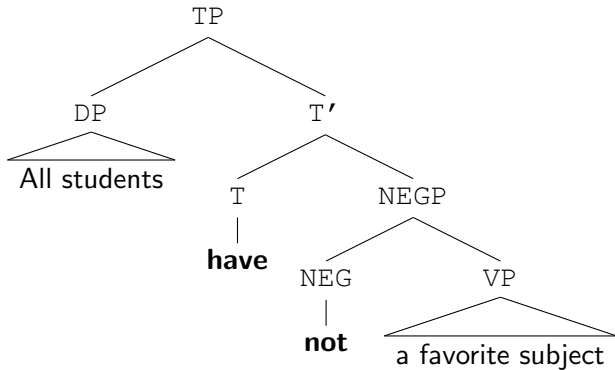


Analysis

- ▶ the change does not simply involve a lexical change from *have not* to *do not have*
- ▶ instead finite, possessive *have* changes category from T to V
- ▶ when *have* is V, negation requires *do*-support

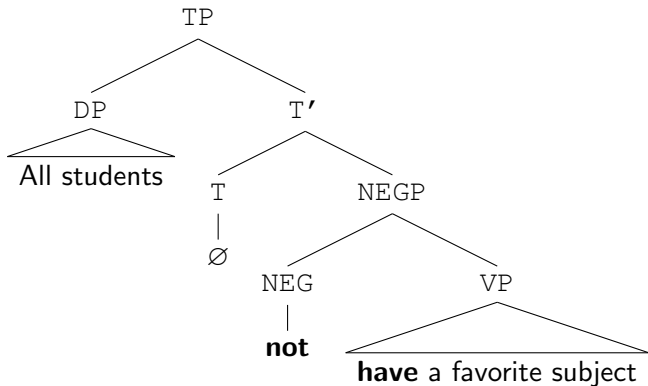
Analysis

(4)



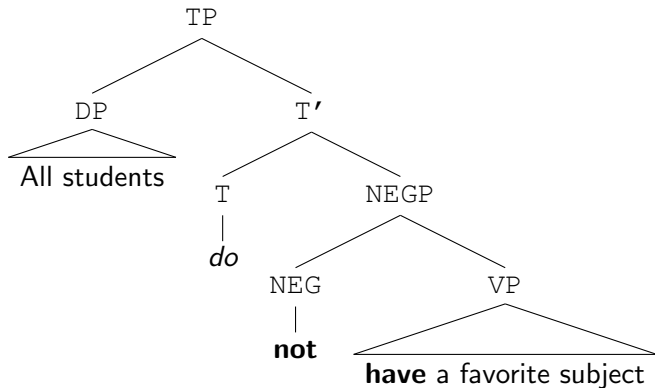
Analysis

(5)



Analysis

(5)



Analysis

- ▶ the change does not simply involve a lexical change from *have not* to *do not have*
- ▶ instead finite, possessive *have* changes category from T to V
- ▶ when *have* is V, negation requires *do*-support
- ▶ evidence:
 - (i) T Coordination, (ii) VP Ellipsis, (iii) Tag Questions

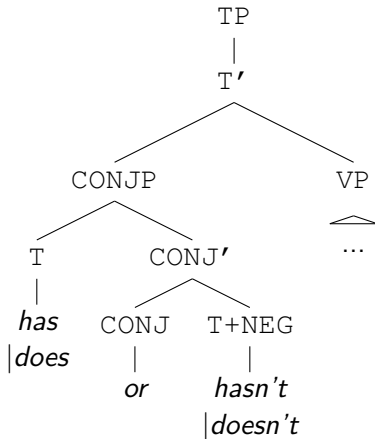
T Coordination

- (6) a. These lands are possessed by proprietors who either **have or have not** the capital necessary to cultivate them. *SystemEconomical*, 1878
- b. ... to set aside preconceptions as to whether religion **has or has not** any value; *IntroductionStudy*, 1908

- (7) a. ... insulting a prospective juror in order to determine if they **do or do not have** a conscious or unconscious prejudice. *Time*, 1964
- b. How are we to test whether a quality **does or does not have** parts? *MetaphysicsNatural*, 1966

T Coordination

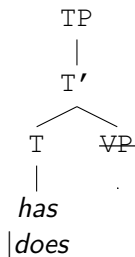
(8)



VP Ellipsis

- (9) a. ... if Quid hasn't a right to the property, **who has**?
FudgeDoingsBeing, 1855
- b. If they do not have a better idea, **who does**?
Time, 1979

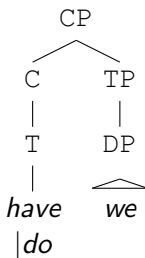
(10)



Tag Questions

- (11) a. But where is the sheet iron to come from? We haven't
any here – **have we?** *AdriftInIceFields*, 1877
- b. Natives! We don't have any agreements with them, **do**
we? *GreyBeginning*, 1984

(12)



Data collection

Data collection

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

Data collection

- ▶ *Corpus of Historical American English* (Davies 2010)
- ▶ 400m words

Data collection

- ▶ *Corpus of Historical American English* (Davies 2010)
- ▶ 400m words
- ▶ high register, formal, written, standard American English

Data collection

- ▶ *Corpus of Historical American English* (Davies 2010)
- ▶ 400m words
- ▶ high register, formal, written, standard American English
- ▶ collection of all non-interrogative sentences with negation and possessive *have*

Data collection

- ▶ *Corpus of Historical American English* (Davies 2010)
- ▶ 400m words
- ▶ high register, formal, written, standard American English
- ▶ collection of all non-interrogative sentences with negation and possessive *have*
- ▶ 120 string-based queries, such as "have not the", "hasn't any", "don't have a", "did not have his" etc.

COHA Interface

CORPUS OF HISTORICAL AMERICAN ENGLISH

400 MILLION WORDS, 1810-2009

EMAIL

PASSWORD

(HELP) LOG IN (REGISTER)

DISPLAY LIST CHART KWIC COMPARE

SEARCH STRING
 WORD(S)

COLLOCATES

POS LIST

SECTIONS SHOW

1 --IGNORE-- 2 --IGNORE--

2000 2000
 1990 1990
 1980 1980
 1970 1970
 1960 1960
 1950 1950

SORTING AND LIMITS

SORTING FREQUENCY

MINIMUM FREQUENCY

CLICK TO SEE OPTIONS

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...]

COMPARE
 SIDE BY SIDE

	CONTEXT	ALL	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
1	DID NOT HAVE THE	722				4	6	4	20	11	23	31	40	42	48	59	71	75	63	78	6	

4.281 seconds

KEYWORD IN CONTEXT DISPLAY [Help / information / contact](#)

SECTION: NO LIMITS

CLICK FOR MORE CONTEXT [?]

1	1863	MAG	Atlantic	A	B	C	he gained experiences which might have extinguished his mind, but which, as they did not have
2	1866	FIC	Outpost	A	B	C	Damon." Now we can watch the porpoises at play." But they did not have that sight to interest
3	1870	FIC	RudderGrange	A	B	C	river road, whence I always had the earliest view of my establishment, I did not have that view
4	1870	FIC	LostInFog	A	B	C	vapors, which surrounded them on all sides, it is true, but yet did not have that dampness which
5	1880	FIC	JollyFellowship	A	B	C	I waded ashore. The boy who was standing by me was Rectus. He did not have that name then
6	1880	NEWS	NYT-Ed	A	B	C	ranks arising out of local causes merely. At all events, the Maine election did not have that intensity
7	1885	MAG	NewEngYaleRev	A	B	C	was abundantly qualified to take the leading place in his department of study. He did not have !
8	1888	FIC	StrangeManuscript	A	B	C	death, of course, and poverty, too, very strongly; but I did not have that eager and energetic
9	1890	MAG	Atlantic	A	B	C	that end received his earnest support. Nevertheless, in the Convention of 1787 he did not have
10	1893	NEWS	NYT-Ed	A	B	C	" without the help of the American Minister and the United States marines. They did not have !

Manual clean-Up

- ▶ search queries yield several examples that are not relevant
- ▶ removal of irrelevant examples manually
- ▶ (manual evaluation almost finished)

(13) a. *questions*

Has not his sire With impious step invaded all our
temples? *AlexisCzarewitz, 1812*

b. *verb-first conditionals*

and he might have fallen, **had not the** bishop
stretched out his hand *JourneyInOther, 1894*

c. *intervening element*

you **have not much** longer to bear with my humours
YankeyInEngland, 1815

The final data set

The final data set

- ▶ resulting data set consists of $N = 32,145$ examples

The final data set

- ▶ resulting data set consists of $N = 32,145$ examples
- ▶ one of the largest datasets of a syntactic change

The final data set

- ▶ resulting data set consists of $N = 32,145$ examples
- ▶ one of the largest datasets of a syntactic change
- ▶ sufficient material to have a data point for every single year, 1810-2009

From HAVE NOT to DO NOT HAVE
Data Collection
Testing of Hypotheses
Conclusion

Hypothesis I: S-shaped trajectory

Hypothesis I: Results

Hypothesis II: Constant Rate Effect

Hypothesis II: Results

Hypothesis I

Hypothesis I

- ▶ The increase in the DO NOT HAVE variant follows an S-shaped curve.

Logistic Growth Model

- ▶ A phenomenon or entity spreads slowly at first, then quickly and finally more slowly again as it approaches an upper limit (e.g. full penetration in the population).

Logistic Growth Model

- ▶ A phenomenon or entity spreads slowly at first, then quickly and finally more slowly again as it approaches an upper limit (e.g. full penetration in the population).
- ▶ In other words, growth follows an S-shaped curve.

Logistic Growth Model

- ▶ A phenomenon or entity spreads slowly at first, then quickly and finally more slowly again as it approaches an upper limit (e.g. full penetration in the population).
- ▶ In other words, growth follows an S-shaped curve.
- ▶ The word "logistic" has no particular meaning in this context. It is used as a contrasting term to "exponential" growth. However, one common function to model logistic growth is in fact the logistic function.

Logistic Growth Model

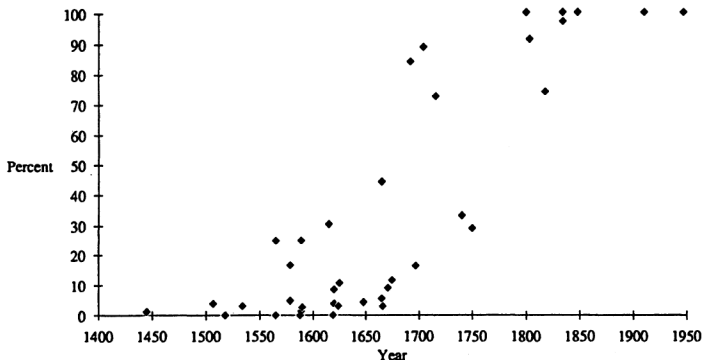
- ▶ A phenomenon or entity spreads slowly at first, then quickly and finally more slowly again as it approaches an upper limit (e.g. full penetration in the population).
- ▶ In other words, growth follows an S-shaped curve.
- ▶ The word "logistic" has no particular meaning in this context. It is used as a contrasting term to "exponential" growth. However, one common function to model logistic growth is in fact the logistic function.
- ▶ Used in a wide range of empirical phenomena, such as spread of a new technology, saturation of sales in a market, population growth, spread of a disease etc.

Logistic Growth Model

- ▶ A phenomenon or entity spreads slowly at first, then quickly and finally more slowly again as it approaches an upper limit (e.g. full penetration in the population).
- ▶ In other words, growth follows an S-shaped curve.
- ▶ The word "logistic" has no particular meaning in this context. It is used as a contrasting term to "exponential" growth. However, one common function to model logistic growth is in fact the logistic function.
- ▶ Used in a wide range of empirical phenomena, such as spread of a new technology, saturation of sales in a market, population growth, spread of a disease etc.
- ▶ Logistic growth has frequently been hypothesized for linguistic changes (Weinreich et al. 1968; Altmann 1983; Kroch 1989)

Example of linguistic application

- ▶ In practice, S-curves often imposed rather than observed



Percentage of INFL-medial subordinate clauses in Yiddish texts.

(From: Santorini 1992, 617)

From HAVE NOT to DO NOT HAVE
Data Collection
Testing of Hypotheses
Conclusion

Hypothesis I: S-shaped trajectory

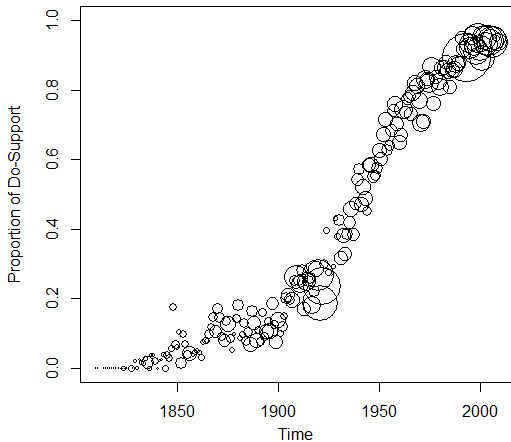
Hypothesis I: Results

Hypothesis II: Constant Rate Effect

Hypothesis II: Results

Results

Plot of *do*-support against 'Time'



Simple model with 'Time' as only predictor variable

- ▶ Intercept and coefficient for 'Time' variable

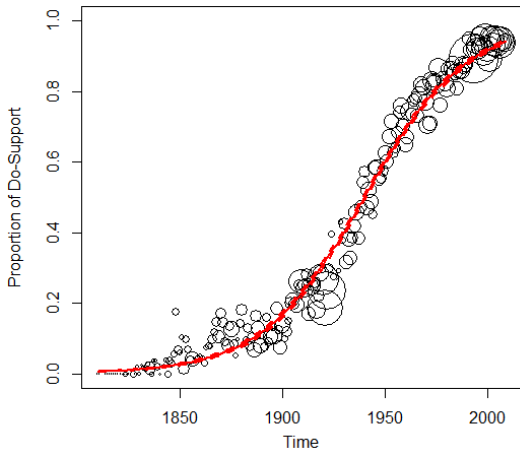
```
formula = DoSupport ~ Time
```

	Estimate	Std.Error	z-value	<i>p</i>
Intercept	-77.89	0.8068	-96.53	<0.001***
Time	0.04015	0.0004153	96.68	<0.001***

- ▶ 95% confidence intervals for 'Time' coefficient

	2.5%	97.5%
Time	0.03934	0.04097

Plot of Model (red line)



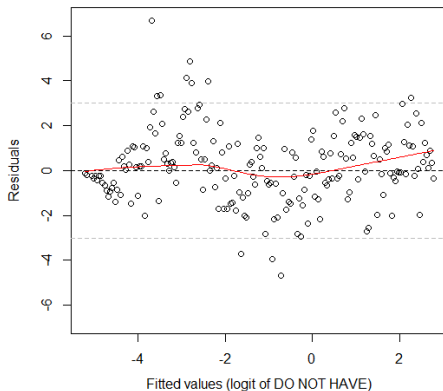
Model Fit

Null deviance: 18027.50 on 197 df
Residual deviance: 568.09 on 196 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
 $\chi^2=17459.42$, $df = 1$, $p<0.001^{***}$
- ▶ Pseudo R^2 : 'Time' predicts the outcome in an excellent way
Hosmer and Lemeshow $R^2 = 0.968$
- ▶ Predictive Accuracy: Model classifies considerably more examples correctly than null model with intercept only
correct: 81.62%, baseline: 49.43%

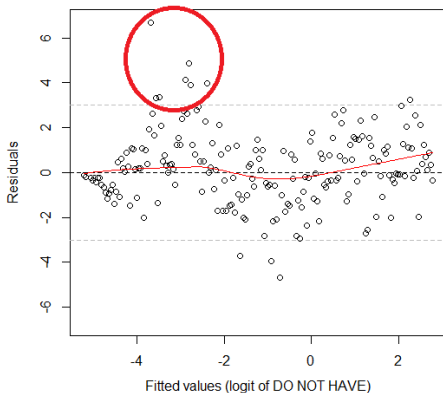
Residual Plot

- ▶ residuals fall within a horizontal band; constant variance
- ▶ most values within ± 3 ; few outliers



Residual Plot

- ▶ cluster of outliers in early period
- ▶ more instances of DO NOT HAVE than model predicts



Some examples

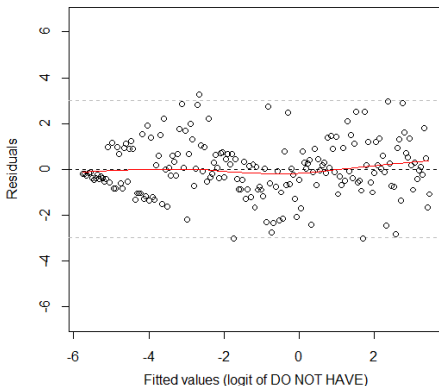
- (14) a. Dea'! I don't see why they **don't have an elevata**
Oak Openings, 1848
- b. I've seed picters of this place before, but I **didn't have no idee** it was so handsum, or that it was sich a grate curiosity.
Major Jones's Sketches of Travel, 1848
- c. ef we **don't have a supply o' water**, we're likely to perish
Ella Barnwell, 1853
- d. "Wal, wal," said the captain, " I **didn't have much hopes**; it's jest as I feared."
Lost in the Fog, 1870

Early outliers are a genre effect

- ▶ COHA usually represents a high register, but fictional texts may include direct speech with more vernacular features
- ▶ it seems that the use of DO NOT HAVE sometimes does not reflect an author's grammar but a caricature of a character's language usage
- ▶ the fact that *do*-support is often more frequent in the portrayal of lower classes, uneducated speakers or marginal groups may indicate that the change spreads from below

Residual Plot for Non-Fictional Texts Only

- ▶ if only texts annotated as NEWS, MAG or NF in corpus are considered, effect of outliers in early periods disappears



Summary

- ▶ Spread of DO NOT HAVE does indeed follow an S-shaped curve.
- ▶ The S-shaped pattern is not just assumed as a plausible model but in fact clearly visible in the data

From HAVE NOT to DO NOT HAVE
Data Collection
Testing of Hypotheses
Conclusion

Hypothesis I: S-shaped trajectory

Hypothesis I: Results

Hypothesis II: Constant Rate Effect

Hypothesis II: Results

Hypothesis II

Hypothesis II

- ▶ If the data is split up into linguistically meaningful contexts, the increase in the DO NOT HAVE variant proceeds at the same rate in all contexts. In other words, the change displays Constant Rate Effects (Kroch 1989).
- ▶ The context considered here as an example is the difference between strong and weak determiners.

Strong Determiners

- ▶ definite articles, demonstratives, universal quantifiers, possessives

(15) There is *the / *that / *every / *his child in the garden.

(16) a. Perhaps he **has not the means** – he looks poor
PoorLodger, 1811

b. I **don't have that advantage**.
BundleLetters, 1908

c. If the novel **has not all these objects** in view, ...
Atlantic, 1885

d. But he **does not have his facts** right
Atlantic, 2005

Weak Determiners

- ▶ indefinite articles, numerals, existential quantifiers, bare nouns

(17) There is / are ^{ok}a / ^{ok}seven / ^{ok}some / ^{ok}∅ child /
children in the garden.

- (18) a. If war **has not a natural tendency to harden the heart**, ... *WarInconsistent*, 1815
- b. But this lady, a lady of birth and some position,
certainly **did not have two hundred acres** under her
hands *ComingFriars*, 1889
- c. I know that I **have not any talent for battle pieces**
Scribners, 1892
- d. We **do not have ∅ flower gardens** indoors
NewRepublic, 1929

From HAVE NOT to DO NOT HAVE
Data Collection
Testing of Hypotheses
Conclusion

Hypothesis I: S-shaped trajectory

Hypothesis I: Results

Hypothesis II: Constant Rate Effect

Hypothesis II: Results

The Constant Rate Effect

The Constant Rate Effect

- ▶ Constant Rate Effects can be detected with logistic regression.
 1. The interaction term between 'Time' and 'Context' is a test for parallelism and should not be significant.
 2. Alternatively, individual logistic regressions fitted to each context should return the same coefficient for 'Time' ("slope" of the change).

The Constant Rate Effect

- ▶ Constant Rate Effects can be detected with logistic regression.
 1. The interaction term between 'Time' and 'Context' is a test for parallelism and should not be significant.
 2. Alternatively, individual logistic regressions fitted to each context should return the same coefficient for 'Time' ("slope" of the change).
- ▶ Theoretical importance: Falsify the hypothesis that a change spreads faster in preferred contexts than in dis-preferred contexts.

The Constant Rate Effect

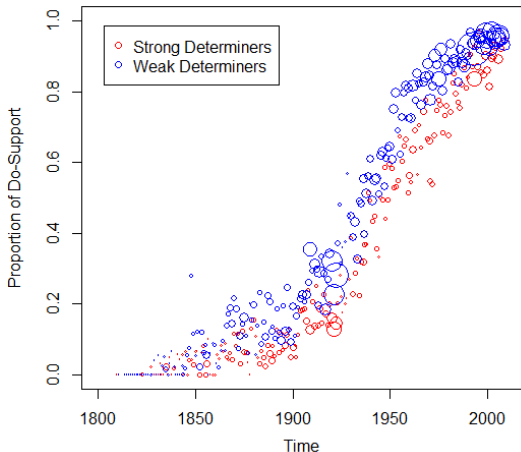
- ▶ Constant Rate Effects can be detected with logistic regression.
 1. The interaction term between 'Time' and 'Context' is a test for parallelism and should not be significant.
 2. Alternatively, individual logistic regressions fitted to each context should return the same coefficient for 'Time' ("slope" of the change).
- ▶ Theoretical importance: Falsify the hypothesis that a change spreads faster in preferred contexts than in dis-preferred contexts.
- ▶ Used to link potentially independent changes to one underlying cause.

From HAVE NOT to DO NOT HAVE
Data Collection
Testing of Hypotheses
Conclusion

Hypothesis I: S-shaped trajectory
Hypothesis I: Results
Hypothesis II: Constant Rate Effect
Hypothesis II: Results

Results

Plot of *do*-support against 'Time' by 'Determiner Type'



Model with 'Time' and 'DetType' as predictor variables

- ▶ Intercept and coefficients for 'Time' and 'DetType' variables

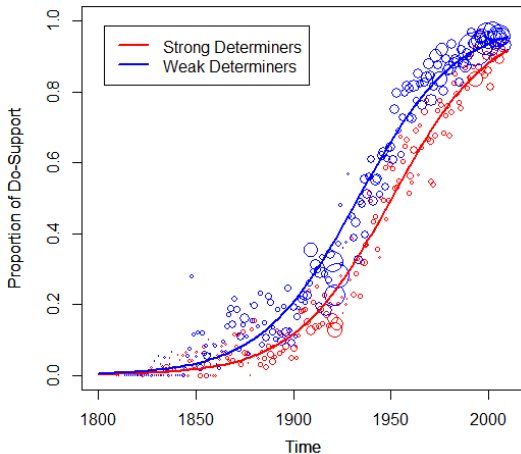
formula = DoSupport ~ Time + DetType

	Estimate	Std.Error	z-value	p
Intercept	-78.46	0.8168	-95.85	<0.001***
Time	0.04023	0.0004203	95.71	<0.001***
DetTypeWEAK	0.6825	0.03101	22.01	<0.001***

- ▶ 95% confidence intervals for 'Time' and 'DetType'

	2.5%	97.5%
Time	0.03941	0.04106
DetTypeWEAK	0.62183	0.74338

Plot of Model (red line, blue line)



Comparison between models

Analysis of Deviance Table

Model 1: DepVar ~ Time

Model 2: DepVar ~ Time + DetType

Model	Resid.Df	Resid.Dev	Df	Deviance	<i>p</i>
1	33230	28606			
2	33229	28113	1	492.67	<0.001***

- ▶ New model is significantly better than first model

From HAVE NOT to DO NOT HAVE
Data Collection
Testing of Hypotheses
Conclusion

Hypothesis I: S-shaped trajectory
Hypothesis I: Results
Hypothesis II: Constant Rate Effect
Hypothesis II: Results

Using the model to predict the future

Using the model to predict the future

- ▶ the model can be used to predict when HAVE NOT becomes ungrammatical

Using the model to predict the future

- ▶ the model can be used to predict when HAVE NOT becomes ungrammatical
- ▶ current evaluations of HAVE NOT: “weird”, “archaic”, “British”, “formal”, but rarely “bad”, “wrong” etc.

Using the model to predict the future

- ▶ the model can be used to predict when HAVE NOT becomes ungrammatical
- ▶ current evaluations of HAVE NOT: “weird”, “archaic”, “British”, “formal”, but rarely “bad”, “wrong” etc.
- ▶ model:
$$\text{DoSupport} = 0.04 * \text{Time} + 0.68 * \text{DetType}$$

Using the model to predict the future

- ▶ the model can be used to predict when HAVE NOT becomes ungrammatical
- ▶ current evaluations of HAVE NOT: “weird”, “archaic”, “British”, “formal”, but rarely “bad”, “wrong” etc.
- ▶ model:

$$\text{DoSupport} = 0.04 * \text{Time} + 0.68 * \text{DetType}$$

$$\text{DoSupport} = 0.04 * 2065 + 0.68 * 0 = 99.02\%$$

presence of DO NOT HAVE with strong determiners
predicted to cross 99% threshold in 2065

Using the model to predict the future

- ▶ the model can be used to predict when HAVE NOT becomes ungrammatical
- ▶ current evaluations of HAVE NOT: “weird”, “archaic”, “British”, “formal”, but rarely “bad”, “wrong” etc.
- ▶ model:

$$\text{DoSupport} = 0.04 * \text{Time} + 0.68 * \text{DetType}$$

$\text{DoSupport} = 0.04 * 2065 + 0.68 * 0 = 99.02\%$
presence of DO NOT HAVE with strong determiners
predicted to cross 99% threshold in 2065

$\text{DoSupport} = 0.04 * 2048 + 0.68 * 1 = 99.02\%$
presence of DO NOT HAVE with weak determiners
predicted to cross 99% threshold in 2048

Addition of Interaction Term

Analysis of Deviance Table

Model 1: DepVar ~ Time + DetType

Model 2: DepVar ~ Time + DetType
+ Time:DetType

Model	Resid.Df	Resid.Dev	Df	Deviance	<i>p</i>
1	33229	28113			
2	33228	28113	1	0.0785	0.7793

- ▶ Addition of interaction term `Time:DetType` to model with `DetType` does not significantly reduce deviance
- ▶ confirmation of presence of Constant Rate Effect in data

Fitting individual models

► strong determiners

```
formula = DoSupport ~ Time
```

	Estimate	Std.Error	z-value	p
Intercept	-78.21	1.339	-58.41	<0.001***
Time	0.0401	0.0006876	58.31	<0.001***

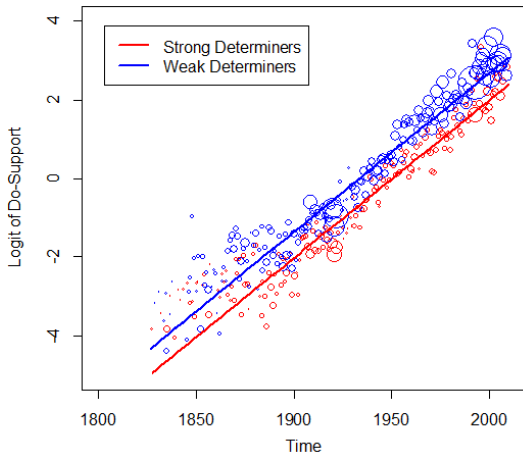
95% CIs: [0.03876 - 0.041457]

► weak determiners

	Estimate	Std.Error	z-value	p
Intercept	-77.99	1.031	-75.65	<0.001***
Time	0.0403	0.0005315	75.90	<0.001***

95% CIs: [0.03931 - 0.0413893]

Logits of DO NOT HAVE for strong and weak determiners



Summary

- ▶ the spread of *do*-support with negative possessive *have* displays a Constant Rate Effect with respect to strong and weak determiners
 - ▶ interaction term between time and determiner type is not significant in a model with determiner type as a main effect
 - ▶ slopes for strong and weak determiners in individual logistic regressions are extremely similar
- ▶ does not confirm that *all* changes display Constant Rate Effects, but adds another instance of the effect to a large body of examples

Relevance of this study

Relevance of this study

- ▶ This study makes use of one of the largest datasets for syntactic changes ever collected (30k examples).

Relevance of this study

- ▶ This study makes use of one of the largest datasets for syntactic changes ever collected (30k examples).
- ▶ It suggests that the logistic growth assumption often postulated for syntactic changes is in fact appropriate.

Relevance of this study

- ▶ This study makes use of one of the largest datasets for syntactic changes ever collected (30k examples).
- ▶ It suggests that the logistic growth assumption often postulated for syntactic changes is in fact appropriate.
- ▶ It also offers further support for the Constant Rate Hypothesis.

Thank you for your attention!

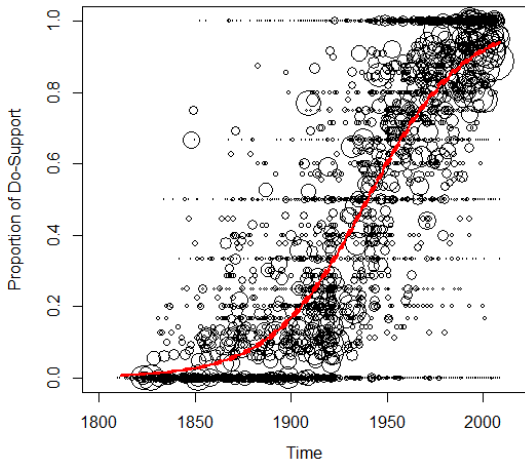


- Altmann, G. (1983), Das piotrowski-gesetz und seine verallgemeinerungen, in 'Exakte Sprachwandelforschung. Theoretische Beiträge, statistische Analysen und Arbeitsberichte', Göttinger Schriften zur Sprach- und Literaturwissenschaft. Bd. 2, Herodot, Göttingen, pp. 54–90.
- Davies, M. (2010), *The Corpus of Historical American English: 400 million words, 1810-2009*, <http://corpus.byu.edu/coha/> (Accessed 10 September 2013).
- Kroch, A. (1989), 'Reflexes of grammar in patterns of language change', *Journal of Language Variation and Change* **1.3**, 199–244.
- Santorini, B. (1992), 'Variation and change in yiddish subordinate clause word order', *Natural Language & Linguistic Theory* **10.4**, 595–640.

Weinreich, U., Labov, W. and Herzog, M. (1968), Empirical foundations for a theory of language change, *in* 'Directions for historical linguistics: a symposium', University of Texas Press, Austin, pp. 95–198.

Appendix

Development of *do*-support by individual text



Residuals for model with random effect

