A syntactic change with lots of data: The rise of *do*-support with possessive 'have' in American English

Richard Zimmermann
Université de Genéve

Manchester LEL research seminar
November 3, 2015

1 Introduction

- this study is essentially a test of the Constant Rate Hypothesis, the idea that a linguistic change spreads at the same speed in all environments in which it is used
- a new test of the Constant Rate Hypothesis is expedient because very large historical corpora now allow for statistical analyses with unprecedented precision
- my test case concerns changes in the realization of possessive *have* in late Modern to Present-Day American English.

1.1 Outline

- introduction to the Constant Rate Hypothesis (section 2)
- the changing status of possessive *have* in American English (section 3)
- data source (section 4)
- data analysis (section 5)
- evaluation of the Constant Rate Hypothesis (section 6)
- conclusion (section 7)

1.2 Focus

- focus is on empirical side of research: data collection, data analysis, statistics, hypothesis testing
- in contrast, formal syntax is greatly reduced; lots of simplifications; common denominator approach so everybody understands the basic assumptions

2 The Constant Rate Hypothesis

2.1 Overview

- the Constant Rate Hypothesis was first introduced by Kroch (Kroch 1989)
- in essence it states the following:

  > Linguistic changes proceed at the same rate of change in all contexts. Contexts are linguistically meaningful divisions, like different constructions, different clause types, different semantic classes etc.

- Constant Rate Effects have frequently been observed in actual linguistic changes

2.2 Explanation of Constant Rate Effects

2.2.1 Grammar Model

- A grammar model G consists of a lexicon L and a set of rules R.
  - L is a set of lexical items with detailed, syntactically relevant information
  - examples: lexical items may contain information on features, subcategorization, possible argument alternations, constructional templates.
  - information of lexical items is assumed to form a list (but could be hierarchically ordered)
  - rules in R describe permissible combinations of lexical items as a generative procedure
  - application results in representations, most importantly here: hierarchical phrase structures
  - rules are assumed to have varying complexity (R could be very impoverished if most combinatorial information is stored in lexical items)
- G models the competence of an individual i, G_i

2.2.2 Usage

- in everyday life, speakers use G to produce finite output
  - output is assumed to simply consist of strings (e.g. as found in texts)
  - output can and does have performance errors
  - but output is also at least to some degree a reflection of the underlying generative possibilities of G
- output forms a corpus
  - a corpus is the sum of outputs from a number of individuals (possibly 1),
  - a diachronic corpus is a corpus in which every output is systematically annotated for a temporal independent variable
- the output of a grammar is complex
  - the specific nature (content) of the output of a grammar depends on numerous situational and psychological factors that are so complicated that the output cannot be predicted precisely
  - one can of course find statistical tendencies when factors related to processing constraints, priming, pragmatic contexts or sociolinguistics are investigated
  - some researchers (especially in sociolinguistics) conclude from these statistical regularities that grammar itself must include a probabilistic component - I don't think that such a model is plausible
  - I will assume that all "soft constraints" are a result of performance and not represented in G

2.2.3 Change

- change is regarded as the addition of a lexical item into L or of a rule into R
- increased use of a new element follows S-shaped curve
  - after a new item becomes available in G_i, it may be used more frequently in its output, and "spread" to other individuals G_j ... G_n as well
  - hence the new element will initially be used exponentially more frequently
  - it will then approach an upper limit (namely full penetration in the population and in "speech situations") and its rise will slow down
  - in other words, spread of a form in a diachronic corpus C follows logistic growth, and thus an S-shaped curve.
  - logistic growth has frequently been hypothesized for linguistic changes (Weinreich et al. 1968; Altmann 1983; Kroch 1989)

2.2.4 Constant Rate Effects

- once a new lexical item or rule has entered the language system, all relevant aspects of the grammar model G have access to it
- the new linguistic form and the environments in which it is used may be independent from each other
- under this assumption, Constant Rate Effects become expected

2.2.5 Competition

- a subset of linguistic changes proceed via competition
- a new lexical item or rule may form a set with another lexical item or rule
  - the innovative and conservative variant are mutually exclusive
  - the innovative and conservative variant are functionally very similar
- in variationist sociolinguistics, the same concept is called the "linguistic variable"
- competing linguistic forms are easier to investigate statistically because we can assume that at one point in time the distribution of the competing forms follow the binomial distribution with a parameter p representing the probability of using the innovative form at that time; logistic regression models build on this assumption

2.3 The ideal change to investigate the Constant Rate Hypothesis

- The Constant Rate Hypothesis has been investigated before, but the case studies were often defective
- In order to advance our understanding of Constant Rate Effects, a case study should include a linguistic change with the following "ideal" properties:
  - the hypothesized change should in fact be a single change; there should be no other obvious interferences, such as other relevant changes, subsequent reanalysis etc.
  - the change should be competitive; the statistical analysis will be more powerful
  - the change should only have two competing forms; an innovative and a conservative variant
  - the competing variants should be easy to identify
  - the change should be complete; it should spread from 0% to 100% of use
  - there should be large amounts of data available for the change; this makes linguistic changes observed in medieval texts bad contestants for test cases of the Constant Rate Hypothesis
  - the different linguistic contexts considered must not be used simply because they are convenient; instead they must be *deduced* from a theory of a grammar or a grammar fragment
  - there should be no obvious prescriptive pressures against the innovative form in writing

3 Possessive *have* in American English

3.1 Overview

- definition of "possessive *have*:"
  - *have* + possessed DP
  - contrast with perfect auxiliary *have* (*have done*), causative *have* (*have someone do something*), modal *have* (*have to do something*)
  - both concrete (*have a car*) and abstract possession (*have an idea*)
- the usage of possessive *have* in certain syntactic contexts changed substantially between the 19^{th} and 21^{st} centuries in written American English.
- illustration:

(1) *conservative usage of possessive 'have'*
Have we not botanical gardens? **We have**, indeed, and much good they should do ...
Sir Samuel White Baker (1855) *Eight Years' Wanderings in Ceylon*

(2) *innovative usage of possessive 'have'*
"**Don't you have** a spare key?" "**I do**." He went into the house and returned with a large iron key ...
Frederick Ramsay (2008) *Stranger Room*

1. in (1) *not* negates the proposition, but in (2) negation requires do-support
2. in (1) *have* and subject invert, but in (2) *have* does not invert with the subject and instead *do*-support is used again
3. in (1) an elliptical answer is formed with *have* directly, but in (2) *do* is used

- the characteristics in (2) represent the innovative patterns that are becoming generalized

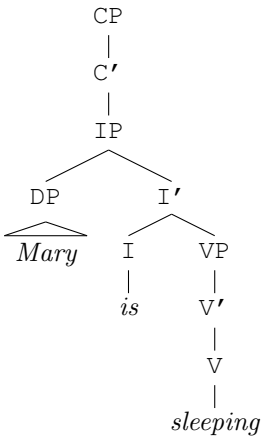
3.2 A simple model of Modern English clause structure

- standard account of Modern English, as frequently taught in introductory classes, found in syntac text books etc.
- rudimentary, not all details included
- my implementation is representational, no movement

3.2.1 Basics

- CP, IP, VP hierarchy of projections
- auxiliaries belong to category I (passive / progressive *be*, perfect *have* and modals like *can*, *might*, *would*) are of category I
- finite and non-finite lexical main verbs are of category V

(3)

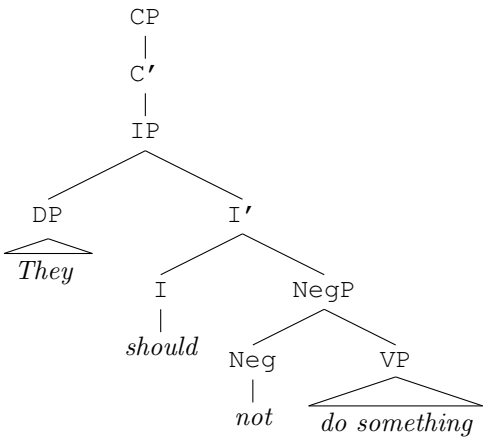


3.2.2 Negation

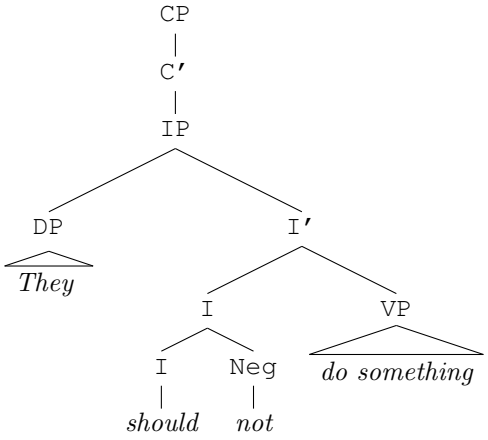
- negation modeled as `NegP` above `VP` (4a.) or as head-adjunction to `I` (4b.)

(4)

a.



b.

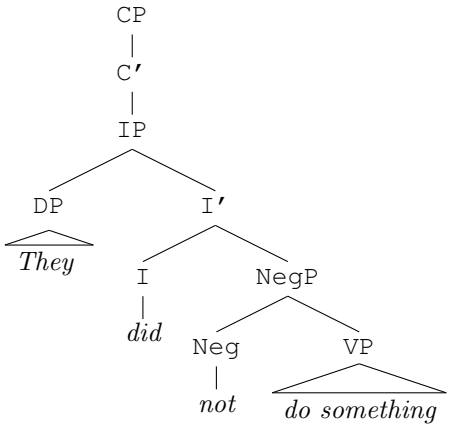


- without an element of category `I`, *do*-support is required
- this follows directly from top-down constraints: there are no phrase structures rules that allow the introduction of negation without `I`

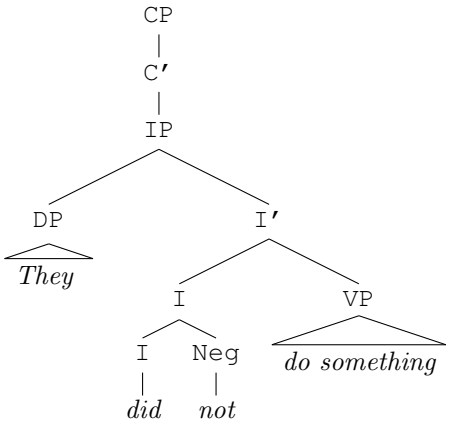
(5) `I′ → I NegP`
`I → I Neg`
but no rule like:
`I′ → NegP`
etc.

(6)

a.



b.



3.2.3 Inversion

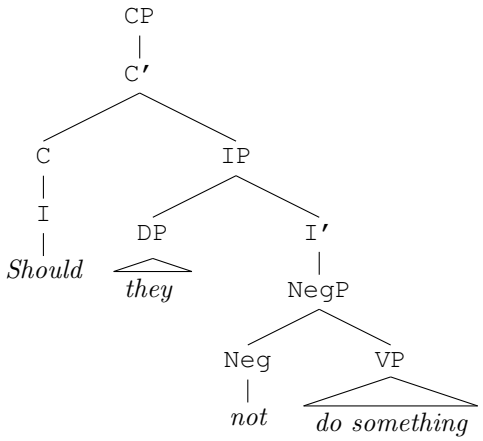
- inversion is modeled as insertion of an `I` element under `C`

  (7) `C` → `I`

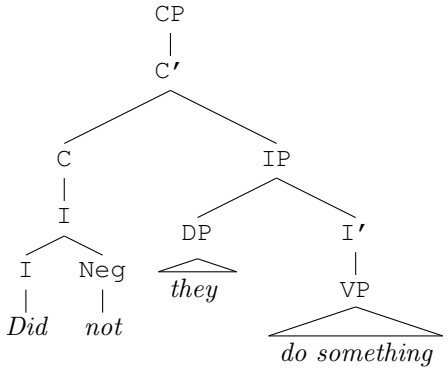
- inversion is licensed by an interrogative feature, or some other feature on an initial constituent
- if an `I` element is not present, *do*-support is required
- interaction with negation: `NegP` can remain low, or a complex `I+Neg` cluster can be placed under `C`

(8)

a.



b.

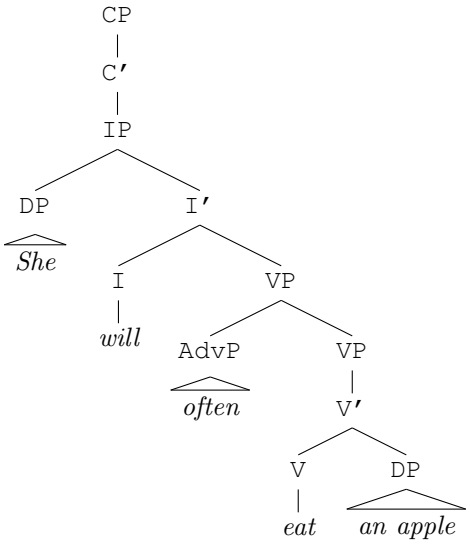


3.2.4 Adjunction

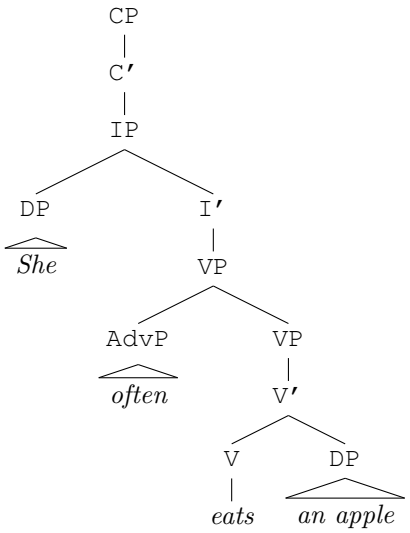
- adverbs can be adjoined to `VP`
- hence, adverbs will appear between auxiliaries and main verbs, and before finite main verbs

(9)

a.



b.



- however, an alternative position for adverbs is possible and in fact frequent - they can be placed before auxiliaries
- some natural examples from the BNC:

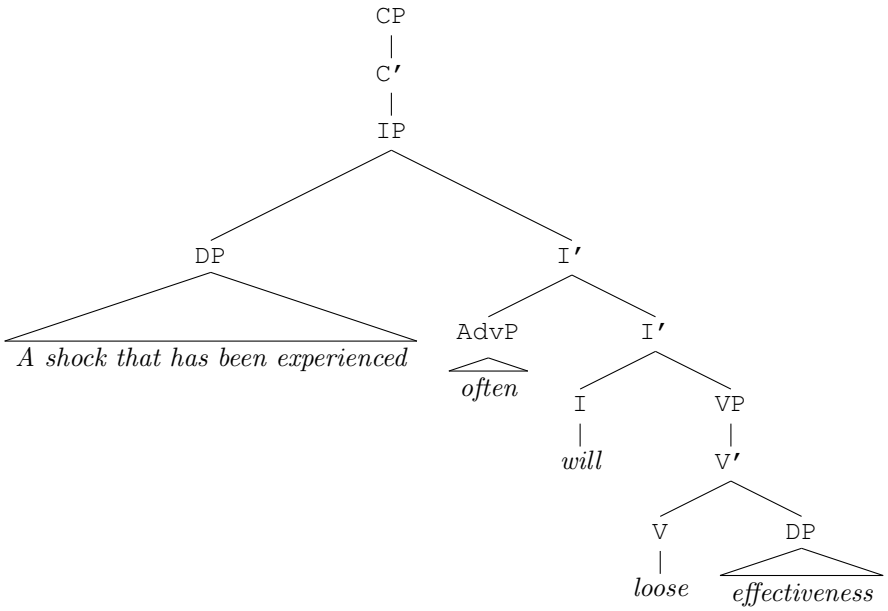
(10) a. Apparently a shock that has been experienced **often will** lose effectiveness as a reinforcer.
BNC, APH W_ac_soc_science

b. I found that the story they tell **sometimes has** changed dramatically
BNC, F71 S_speech_unscripted

c. if he was not prepared to take a risk with his money, he **most certainly was** not prepared to take a risk which might lose him the one person he valued most.
BNC, FPK W_fict_prose

- I model high adverbs as adjunction to `I'`

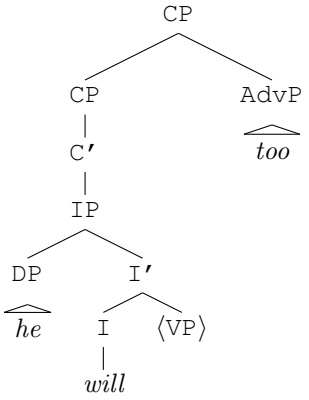
(11)



3.2.5 Ellipsis

- ellipsis phenomena target the `VP`
- requires the presence of an element in `I`
- the missing string in the `VP` is recovered on pragmatic and syntactic principles

(12) Close your eyes and ...



3.3 The diachrony of possessive *have*

3.3.1 The change

- within the model of Modern English clause structure presented, the central change that possessive *have* undergoes can be expressed like this:

Possessive *have* changes from an auxiliary of category `I` to a lexical main verb of category `V`

- in derivational frameworks, this can be modeled as the loss of verb movement

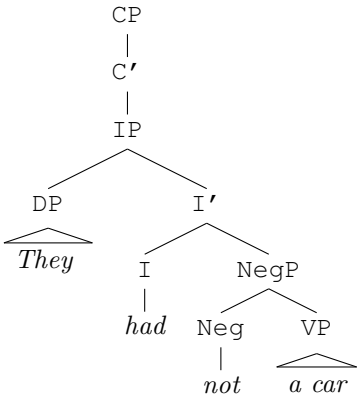
3.3.2 Predictions

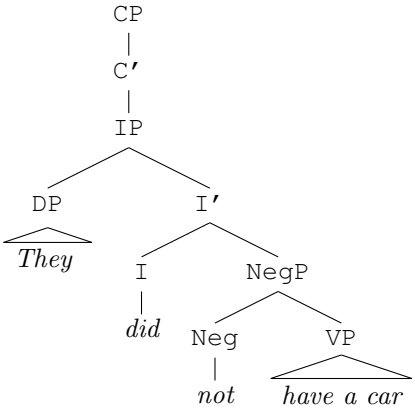
- the theory that *have* changes from `I` to `V` predicts that its realization changes in four domains:

1. Negation

(13)

a. conservative variant of *have* → b. innovative variant of *have*

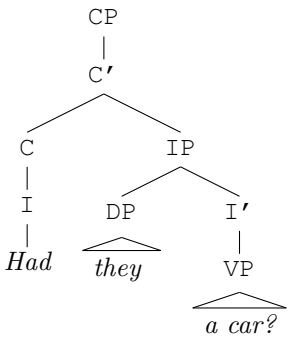


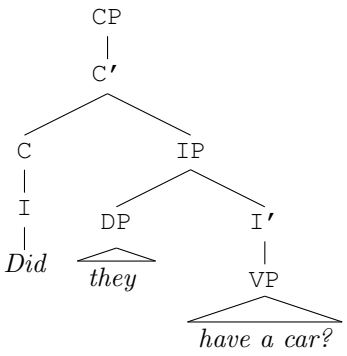


2. Inversion

(14)

a. conservative variant of *have* → b. innovative variant of *have*





3. Relative order of possessive *have* and VP-diagnostics

(15)

a. conservative variant of *have* → b. innovative variant of *have*

4. Ellipsis

(16)

a. conservative variant of *have* → b. innovative variant of *have*

I had a car and ... I had a car and ...

3.4 Evaluation

- we have here an ideal case to test the Constant Rate Hypothesis:
  - only one small competitive change occurs in an otherwise relatively stable system, i.e. the category change of possessive *have*; easy to identify
  - the grammar theory is relatively well-understood / standard / uncontroversial
    no ad-hoc assumptions about the change; the theory follows from independent observations regarding differences between auxiliaries and lexical verbs
    it is relatively uncontroversial how / that the grammar generates the different surface patterns; we can deduce which constructions to look for from the theory
  - we are dealing with a recent change; massive amounts of data from 19$^{\text{th}}$ to 21$^{\text{st}}$ century are available

> **The Constant Rate Hypothesis**
> The linguistic change of possessive *have* from an auxiliary to a main verb should proceed at the same rate of change in negation, inversion, adjunction and ellipsis contexts.

- we just need to collect lots of data for the different contexts and check if they actually change at the same rate

4 Methodology

4.1 Data source

- data was taken from the *Corpus of Historical American English* (COHA) (Davies 2010)
- c. 385m words
- high register, formal, written, standard American English
- part of speech tagged (unknown accuracy) but not parsed
- results automatically returns three independent variables:
  - every token is indexed for a year; sometimes year of publication, sometimes of second edition, but relatively reliable
  - every token is associated with a text name; quite a lot of duplicates (e.g. different editions)
  - every token is indexed for a generic genre (news, magazine, non-fiction, fiction)
- corpus can be searched with search query strings, e.g. "`has he the`"

4.2 Principles of search scripting

- **Guided searches**
  For every search query, a specific construction should be targeted. Search queries should not be random. For example, the search `has not the` might specifically target negative declaratives like *She has not the slightest idea* but not questions like *Has not the citizen a right*
- **Symmetric searches**
  Every pattern should have an equal chance to be found with the innovative and the conservative form. Hence, every search must be conducted twice with only a minimal difference between the conservative and innovative forms. For example, if there is a search query such as `has not the` there must also be a search query `does not have the` to find negative possessive have with and without *do*-support.
- **Precision over Recall**
  In order to assure data accuracy that is high enough for valid statistical inference, the result of every search query should be evaluated manually. The data set should have precision approaching 100%.

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

- **Large data**
  As many search queries as possible should be carried out. In order to increase certainty of the statistical estimates, large data sets must be used.

5 Data collection and analysis

5.1 Sentential negation

5.1.1 Logic

- sentential negation requires an overt element in `I`
- prediction:
  - the order '*have* - *not*' should exist as long as *have* can be inserted under `I`
  - but *do*-support should become regularized as *have* is increasingly `V`

5.1.2 Negation - Search queries

- dependent variable: '*have not*' vs. '*do not have*' - 2
- independent variables:
  searches for different inflections + negation with an without contraction + object element
  - searches for `have|has|had`, `do|does|did` - 3
  - negation with and without contraction, *not* and *n't* - 2
  - object element was requierd to increase the likelihood of possessive *have* - 10

targeted object	variable name	search term
definite description	the	`the`
indefinite nominal	a	`a`
indefinite nominal before vowel	an	`an`
quantified expression that cannot occur in existential construction	quantstrong	`every|each|all|most|least|both|neither`
quantified expression that can occur in existential construction	quantweak	`no|none|few|little|fewer|less|some|several|various|any|enough|more|many|much|plenty|lots`
cardinal number nominal	num	`[MC*]`
demonstrative nominals	dem	`this|these|that|those`
bare singulars	bare sg.	`[*nn1*]`
bare plurals	bare pl.	`[*nn2*]`
possessive adjective nominal	poss	`[app*]`

- example of a search query string:

  `had n't every|each|all|most|least|both|neither`
  vs.
  `did n't have every|each|all|most|least|both|neither`

- 2*3*2*10 = 120 search queries
- manual correction of precision errors

(17) 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

- total of 31,593 examples

5.1.3 Examples

(18) conservative variant: *have not*

a. There, now, add the salt and pepper fixings, and the king himself **hasn't** a slicker supper.
GreyslaerARomance 1840

b. ... whether the General Government have or **have not** a right to lay out roads and canals
NorthAmRev 1831

c. He found it difficult to even sit on the bed and he **hadn't** the strength to take off his clothes.
DeadlyIntent 2009

(19) innovative variant: *do not have*

a. The farming community of 900 people **doesn't have** a single fast-food restaurant
Prevention 2005

b. ... in order to determine if they do or **do not have** a conscious or unconscious prejudice.
Time 1964

c. it must be recollected that the plants in the middle **do not have** the chance to obtain so much air as the outside.
AmericanFruitGarden 1839

5.1.4 Diachronic development

- *do*-supported negation increases in an s-shaped curve

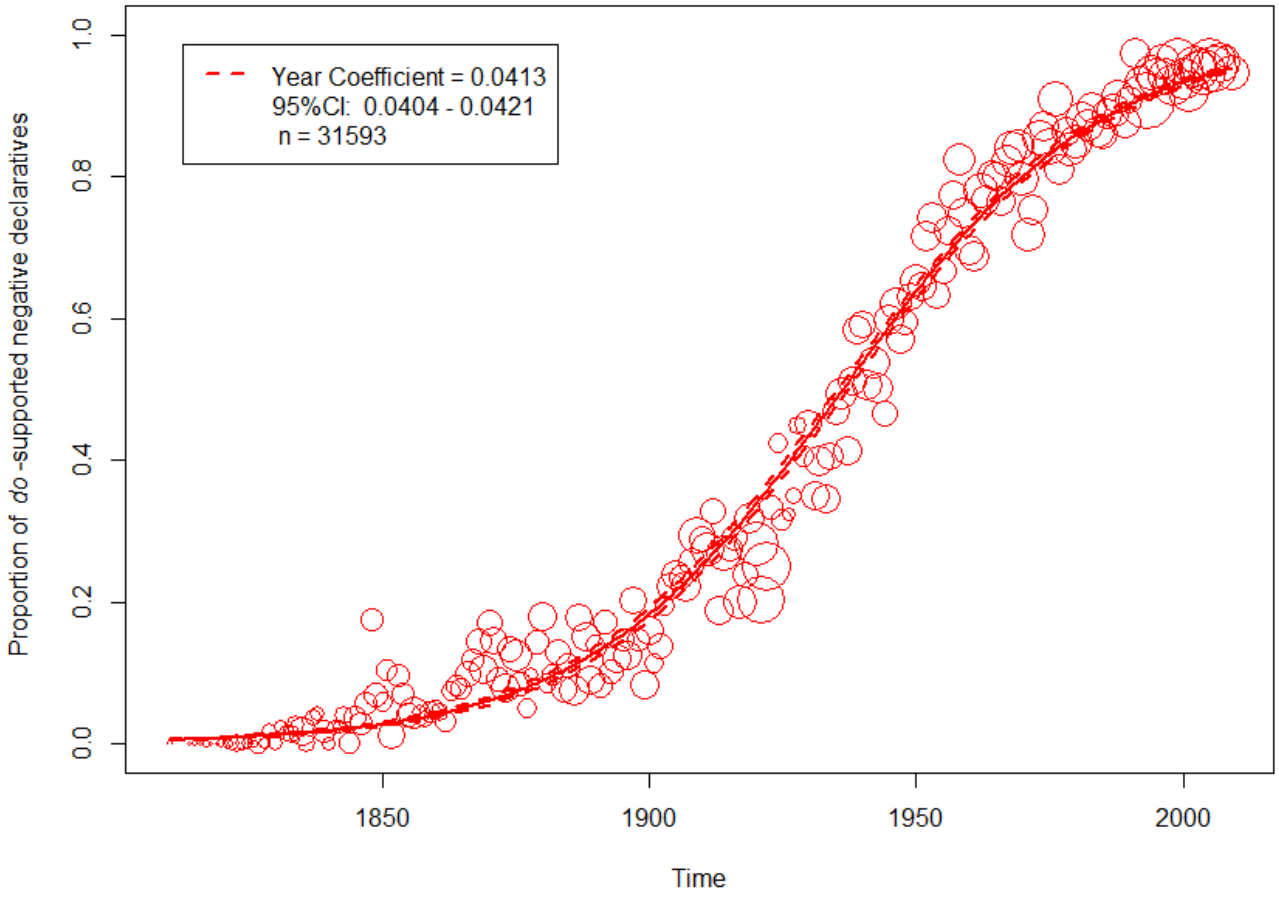


Figure 1: The development of *do*-support with possessive *have* in negative declaratives

– sufficient material for 198 data points, 1810-2009 except 1812 and 1813

– size of points is proportional to number of examples; standard point character = 25 examples

5.1.5 Model evaluation

- the logistic regression model is presented below.

```
Call:
glm(formula = cbind(DONOTHAVE, HAVENOT) ~ Year, family = binomial,
    data = neggraph)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.7096  -0.7924   0.1795   1.0514   6.5681

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.989e+01  8.420e-01  -94.88   <2e-16 ***
Year         4.126e-02  4.339e-04   95.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 18111.00  on 197  degrees of freedom
Residual deviance:   555.08  on 196  degrees of freedom
AIC: 1377.3

Number of Fisher Scoring iterations: 4
```

- most important information:
  - 'Year' coefficient = increase of 0.04126 log-odds(DONOTHAVE) / year
    (at this rate of change, it would take 223 years for a new form to spread from 1 to 99%)
  - Standard Error = 0.0004339, hence 95%-CI = [0.0404 – 0.0421]
- Given its simplicity, the model is an extraordinarily good fit to the data
  - 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=17555.92, df = 1, p<0.001***
  - Pseudo R^2:
    'Year' predicts the outcome in an excellent way
    Hosmer and Lemeshow R^2 = 0.969
  - Predictive Accuracy:
    Model classifies considerably more examples correctly than null model with intercept only
    correct: 82.4%, baseline: 51.7%

- Residual plot:
  - residual plot checks for assumption of constant variance and outliers
  - values should fall within a horizontal band between +/- 3 residuals

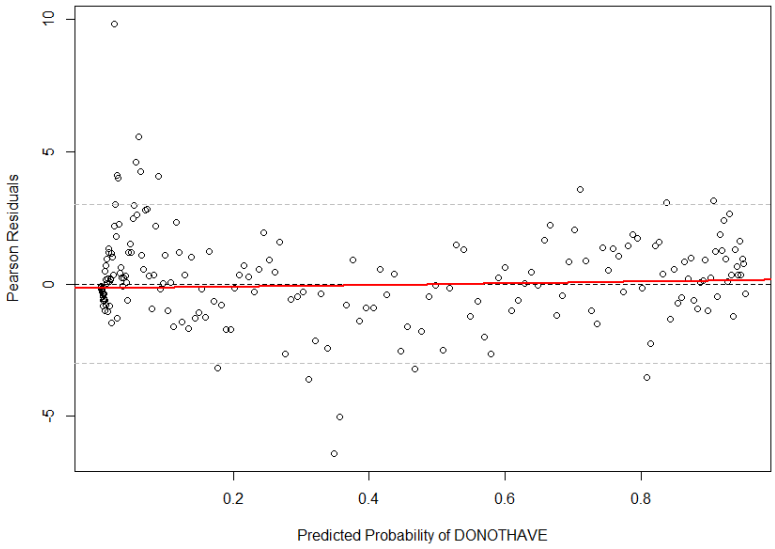


Figure 2: Residual plot of negation data

  - overall relatively constant variance, few outliers
  - but there is a cluster of outliers in early period indicating more instances of DO NOT HAVE than model predicts

- Discussion:
  - the outiers are probably due to register fluctuations: *do*-support is at first regarded as colloquial, (and perhaps direct negation later as archaic)
  - examples:

(20) a. Dea'! I don't see why they **don't have an elevata**
OakOpenings 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

  - 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
  - the outliers disappear if only non-fictional texts (annotated as NEWS, MAG or NF in corpus) are considered

5.2 Subject-auxiliary inversion

5.2.1 Direct questions

5.2.1.1 Logic

- subject auxiliary inversion in direct questions requires placing `I` before the subject
- prediction:
  - the order '*have*-subject' should exist as long as *have* can be inserted under `I`
  - but *do*-support should become regularized as *have* is increasingly `V`

5.2.1.2 Questions - Search queries

- dependent variable: '*have*-subject' vs. '*do*-subject-*have*' - 2
- independent variables:
  searches for different inflections + question type + polarity + object element + subject length
  - searches for `have|has|had`, `do|does|did` - 3
  - different question types - 6
    heuristics were used to find (i) polar (yes/no) questions:
    - `.|?|;|!|:` → initial punctuation mark
    - `and|but|or` → initial conjunction

    and (ii) *wh*-questions:
    - `[rrq*]|[pnq*]` → initial *wh*-adverb (*where, when, why, how*, also: *whom*)
    - `what|which` → initial *wh*-pronouns for object questions
    - `what|which *` → initial *wh*-adjective + a word for object questions
    - `how many|much *` → initial *wh*-quantifier + a word for object questions
  - questions can be positive (*do you have a car*), or negative with low negation (*do you not have a car*) or negative with high negation (*don't you have a car*) with or without contraction - 4
  - object element was requierd to increase the likelihood of possessive *have*; (the, a, an, quantstrong, quantweak, num, dem, bare sg., bare pl., poss) - 10
  - subject placeholders with a length between 1 and 5 words - 5
    - `[d*]|[pn1*]|[pp*]` → 1 word subject, personal pronouns, demonstratives, indefinites
    - `[at*]|[d*]|[app*] [n*]` → 2 word subject with head noun
    - `[at*]|[d*]|[app*] * [n*]` → 3 word subject with head noun
    - `[at*]|[d*]|[app*] * * [n*]` → 4 word subject with head noun
    - `[at*]|[d*]|[app*] * * * [n*]` → 5 word subject with head noun
- example of a search query string:

  `[rrq*]|[pnq*] do not [at*]|[d*]|[app*] [n*] have that|those|this|these`
  (to find for example: *why do not my friends have those characteristics*)
  vs.
  `[rrq*]|[pnq*] have not [at*]|[d*]|[app*] [n*] that|those|this|these`
  (to find for example: *why have not my friends those characteristics*)

- 2*3*6*4*10*5 = 7200 expected search queries
  however, due to length restrictions, omission of some bare sg. and pl. and other simplifications, only 2190 queries were actually carried out

- manual correction of precision errors

(21) a. *verb-first conditionals*
. Had she the daring then, she wouldn't be standing here
TamedByYourDesire 2002

b. *subject + object is really complex subject (many reasons)*
. Have some more milk, Master Hal.
ReelsSpindles 1900

c. But **how have all these** changes affected this visible image of Truth?
NorthAmRev 1850

d. *causative have*
Why don't you have your dad pay you a wage
BeyondBedroomWall 1975

- total of 4,419 examples

5.2.1.3 Examples

(22) conservative variant: *have*-subject

a. But **have you** no sense of obligation to be all that nature intended you to be
Scribners 1873

b. **Has the patriot or the philanthropist** a plan of beneficence?
NewEngYaleRev 1847

c. **Haven't you** some children?
FreaksFortuneThe 1854

d. And whom **have I** the honor of addressing?
BibleInSpainJourneys 1843

(23) innovative variant: *do*-subject-*have*

a. But **do we have** a legal or moral obligation to fight on in Korea
NYT-Reg 1951

b. **Does the age of the host have** an influence on susceptibility and the character of the disease?
Host-parasiteRelations 1927

c. **Don't you have** a girl?
Mov:AngelFace 1952

d. And whom **do I have** the pleasure of addressing?
RoadTamazunchale 1878

5.2.1.4 Diachronic development

- *do*-supported inversion in questions also increases in an s-shaped curve

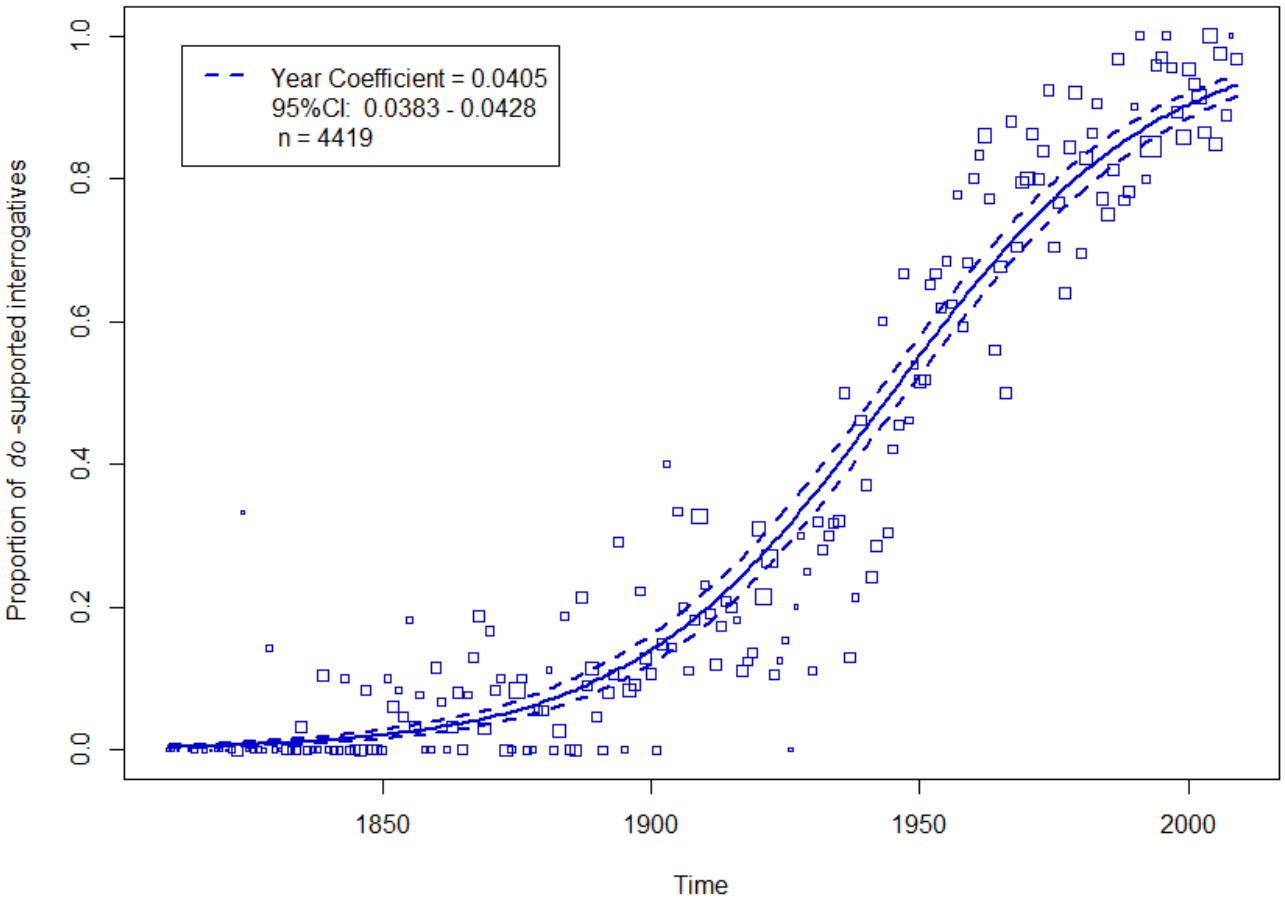


Figure 3: The development of *do*-support with possessive *have* in direct questions

  - again sufficient material for 198 data points, 1810-2009 except 1813 and 1816
  - size of points is proportional to number of examples; standard point character = 25 examples

5.2.1.5 Model evaluation

- as before, overall excellent fit to data
  - -2LL test: χ^2=2437.03, df = 1, p<0.001***
  - Pseudo R^2: Hosmer and Lemeshow R^2 = 0.894
  - Predictive Accuracy: correct: 83.9%, baseline: 41.2%
- early outliers; possible also related to register fluctuation

(24) a. I spec dey'll soon come. But **didn't we have a good time las' night** in Buffalo?
Escape 1861

b. why **don't he have** the witnesses examined apart?
ClintonBradshaw 1835

5.2.2 Other inversion

5.2.2.1 Logic

- exaclty as before, but now inversion is triggered outside of interrogative contexts

5.2.2.2 Other inversion - Search queries

- dependent variable: '*have*-subject' vs. '*do*-subject-*have*' - 2
- independent variables:
  searches for different inflections + different initial constituents triggering inversion
  - searches for `have|has|had`, `do|does|did` - 3
  - different negative and restrictive initial constituents - 8
    - *nor*
    - *not only*
    - *neither*
    - *no* + adverb
    - preposition + *no* + noun
    - *only* + adverb
    - *seldom*
    - *rarely*
    - *so much|many|little|few* + noun
  - use of collocation search option: for innovative variant, `have` should occur up to 9 words to the right
- example of a search query string:

```
rarely have
vs.
rarely do
COLLOCATES: have 0 9
```

- manual correction of precision errors

(25) *perfect*
Not only had I lost the books, but I had lost the man himself
Embarrassments 1896

- 2*3*8 = 48 search queries
- total of 1,023 examples

5.2.2.3 Examples

(26) conservative variant: *have*-subject
  a. and in no country **has he** more power for evil than in the federal courts of the United States.
  TrialTheodoreParker 1855
  b. Not only **had I** no right to take a step which might wound her, but ...
  GeorgeBalcombe 1836
  c. Seldom **have two ages** the same fashion in their pretexts and the same modes of mischief
  NatlReview 1990

(27) innovative variant: *do*-subject-*have*
  a. In no state **do minority groups have** statewide power.
  Time 1967
  b. not only **did I have** no idea where my clothes might be but ...
  Friday 1982
  c. Seldom **does a landscape architect have** the privilege of creating the same estate twice.
  SatEvePost 1940

5.2.2.4 Diachronic development

- once again *do*-supported increases in an s-shaped curve

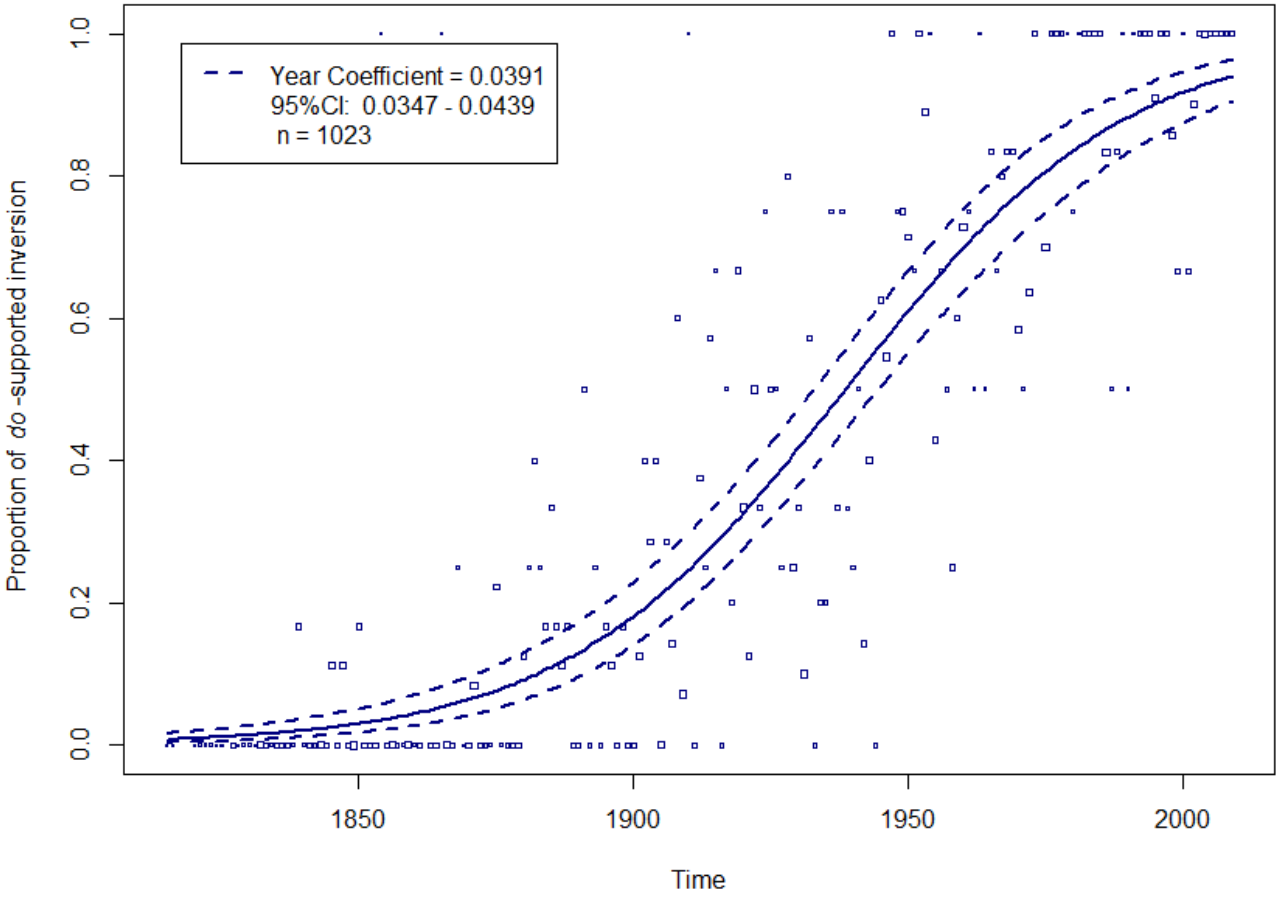


Figure 4: The development of *do*-support with possessive *have* in other inversion contexts

  - sufficient material for 189 data points, 1810-2009, no data 1810-14, 1817-19, 1826, 1862, 1974
  - size of points is proportional to number of examples; standard point character = 25 examples

5.2.2.5 Model evaluation

- overall very good fit to data
  - -2LL test: χ^2=524.43, df = 1, p<0.001***
  - Pseudo R^2: Hosmer and Lemeshow R^2 = 0.714
  - Predictive Accuracy: correct: 81.2%, baseline: 38.7%
- two outliers in early period; otherwise no outliers

5.3 Relative order of possessive *have* and VP-diagnostics

5.3.1 Adverbs

5.3.1.1 Logic

- adverbs are commonly adjoined to VP
- prediction:
  - the order '*have*-adverb' should exist as long as *have* can be inserted under I
  - but the order 'adverb-*have*' should become the only option once *have* is V

5.3.1.2 Adverbs - Search queries

- dependent variable: '*have*-adverb' vs. 'adverb-*have*' - 2
- independent variables:
  searches for *have* in three inflections + object element + an adverb
  - searches for have|has|had - 3
  - object element was requierd to increase the likelihood of possessive *have* (the, a, an, quantstrong, quantweak, num, dem, bare sg., bare pl., poss) - 10
  - 9 temporal adverbs (never|seldom|rarely|sometimes|often|usually|frequently|ever|always), 4 epistemic adverbs (probably|possibly|certainly|necessarily) - 13
- only positive polarity
- example of a search query string:
  - have often the vs. often have the
- 2*3*10*13 = 780 search queries
- manual correction of precision errors
- total of 10,636 examples

5.3.1.3 The problem of high adjunction

- adverbs can occur before auxiliaries, adjoined high to I′
- there is good reason to believe that this is also true for the conservative variant of possessive *have*
  - adverbs can occur before and after possessive *have*

    (28) The poor Indians **now have often reason** to rejoice
    NaturalistInNicaragua 1847
  - coordination structures of auxiliaries and possessive *have* with high adverbs suggests parallelism

    (29) a woman **never can** obtain rank by merit, therefore **never has** reason to be proud of it.
    LoversVows 1814
  - negation or inversion may reveal the conservative nature of *have* in conjunction with a high adverb

    (30) Perish the monster! I **have not** – **never had** father, or home
    RomanTraitorVol12 1846

→ the order 'adverb - *have*' does not necessarily indicate the innovative variant of possessive *have* - it can also be generated by high adjunction of the adverb

The prediction regarding the rate of change of possessive *have* as measured by adverb placement is slightly different from the other contexts: The rate of change should appear slower than in the other contexts.

5.3.1.4 Examples

(31) necessarily conservative variant: *have* - adverb

a. he had constant opportunities to cultivate oratory, for which he **had always** a great passion NorthAmRev 1817

b. The word is good English, and we **have certainly** a right to use it as often as we have occasion NewEngMag1832

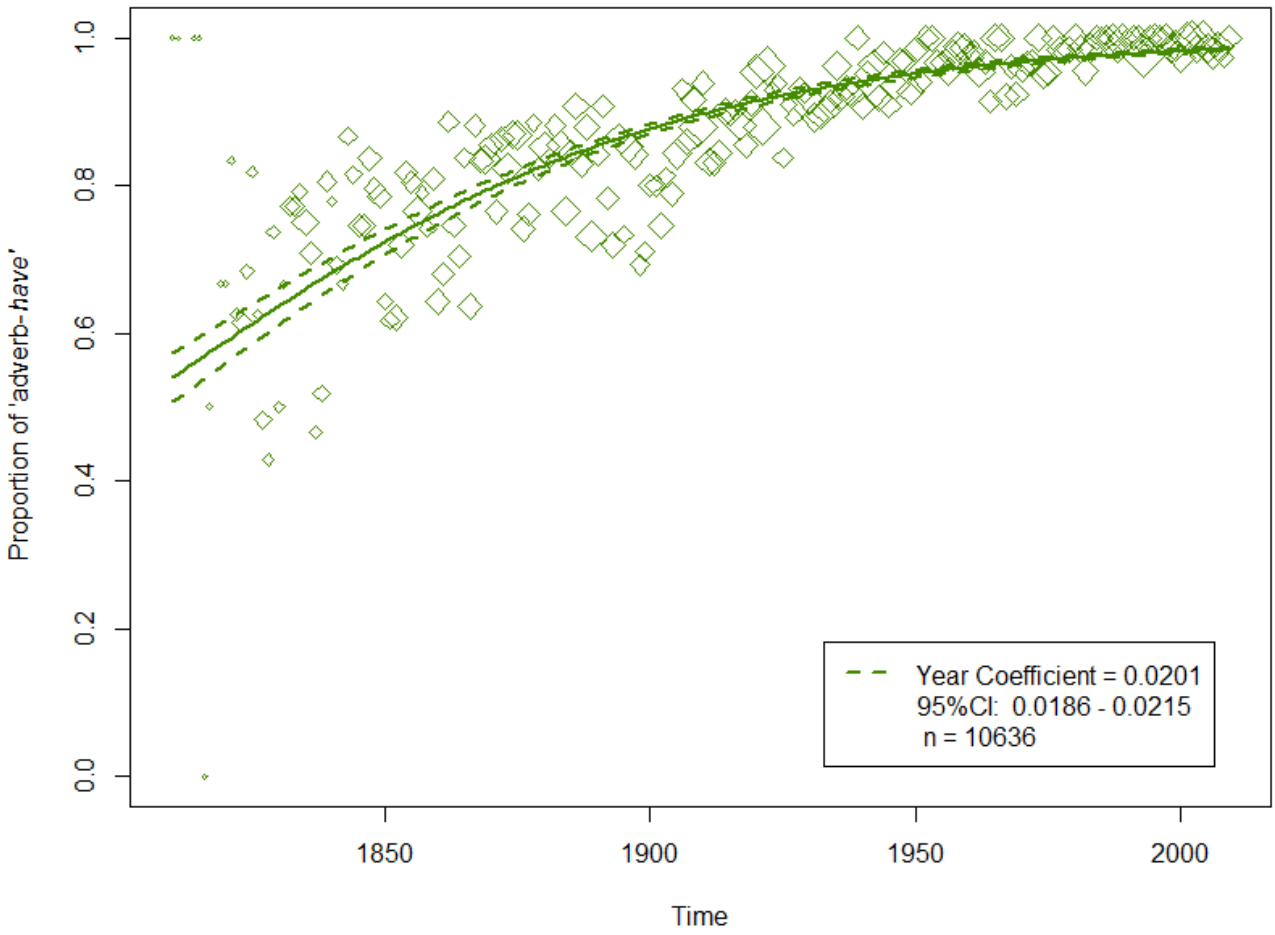
(32) potentially innovative variant: adverb - *have*

a. When a young girl, I **always had** a great desire to see London and Paris; HomeMission 1853

b. You and Jane **certainly have** a right to go to church HeFellInLoveWith1886

5.3.1.5 Diachronic development

- as expected, the order '*have* - adverb' becomes ungrammatical
- also as expected, the overall rate of 'adverb - *have*' is so high that the rate of change appears considerably slower than in the previous contexts



- very good model fit
- no outliers except in the very first years, where there is little data
- substantial random effect of adverb; different adverbs have different propensities to be adjoined high
  - e.g. time coefficients point estimates of *never* = 0.01749, *certainly* = 0.02336, *usually* = 0.03651

5.3.2 *no longer*

5.3.2.1 Logic

- in order to measure the rise of innovative possessive *have* adequately, we need an adjunct that virtually always occurs between an auxiliary and a main verb and never elsewhere, i.e. an adjunct that necessarily marks the `I-VP` boundary
- I investigated over 100 adjuncts or other interveners, but they all have fatal problems
  - adverbs are too flexible (*perhaps, maybe, essentially, even, particularly, indeed ...*)
  - similarly, parenthetical clauses are too flexible (*... , I think, ..., ... she thought, ...*)
  - heavy phrasal adjuncts (superlative, `PP`s like *at present*, adjunct `NP`s like *no doubt ...*) can occur in various places with an intonational break; in fact as a tendency, the heavier the worse
  - temporal measures work, but are too rare with possessive *have* (*daily, weekly, annualy, ...*)
- even true for
- only one singe item worked: *no longer*
  - *no longer* is relatively frequent - I found 1,162 examples
  - *no longer* does not readily occur before auxiliaries
    - native speakers dislike high adjunction of *no longer* (judgments vary; some speakers find high adjunction ungrammatical, some strongly disprefer high adjunction, and some see semantic differences regarding permanence, metaphorical readings or focus)
    - corpus study (COCA); on average only c. 7.5% of all instances of *no longer* are placed before an auxiliary

	(pro)noun *no longer* aux	(pro)noun aux *no longer*	% before aux
aux=*may*	5	182	0.03
aux=*might*	0	37	0.00
aux=*can*	151	1,648	0.08
aux=*could*	78	1,494	0.05
aux=*shall*	1	15	0.07
aux=*should*	5	113	0.04
aux=*will*	59	631	0.09
aux=*would*	39	600	0.06
aux=*are*	97	1,262	0.07
aux=*is*	154	1,845	0.08

5.3.2.2 Examples

(33) conservative variant: *have* - *no longer*

a. she **had no longer** a desire to live
RachelDyerANorth 1828

b. France **has no longer** any regard for the rights of Prussia or the Confederation
NYT-Reg 1860

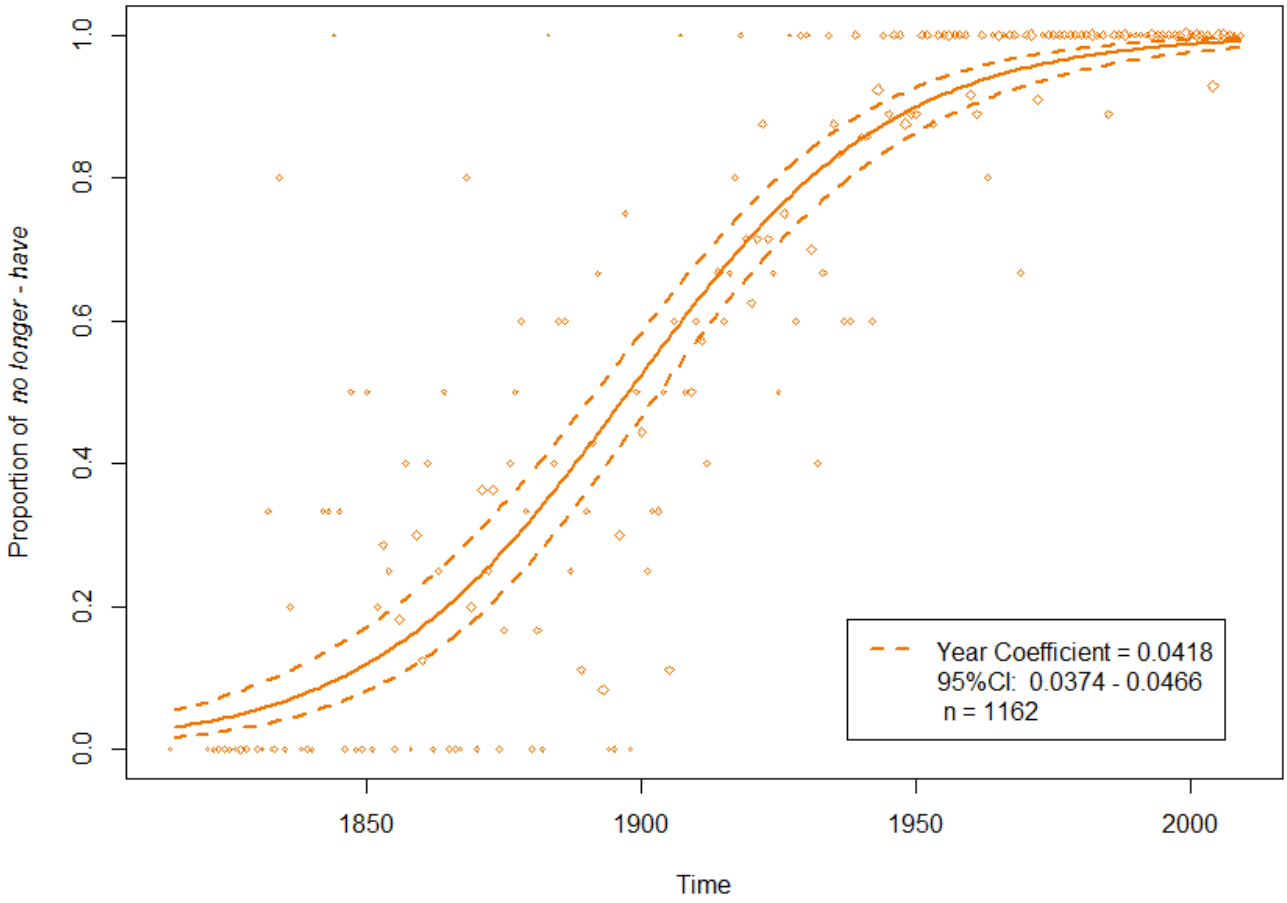
(34) innovative variant: adverb - *no longer* - *have*

a. he **no longer had** the desire to join them
NomadsNorth 1919

b. Great Britain **no longer has** the strongest fleet
Atlantic 1921

5.3.2.3 Diachronic development

- the order *no longer* - *have* increases in an s-shaped curve



  - sufficient material for 185 data points, 1810-2009, no data 1810-13, 1815-20, 1829, 1837, 1841, 1888, 1913
  - size of points is proportional to number of examples; standard point character = 25 examples

- overall very good fit to data
  - -2LL test: χ^2=635.74, df = 1, p<0.001***
  - Pseudo R^2: Hosmer and Lemeshow R^2 = 0.761
  - Predictive Accuracy: correct: 84.7%, baseline: 69.1%

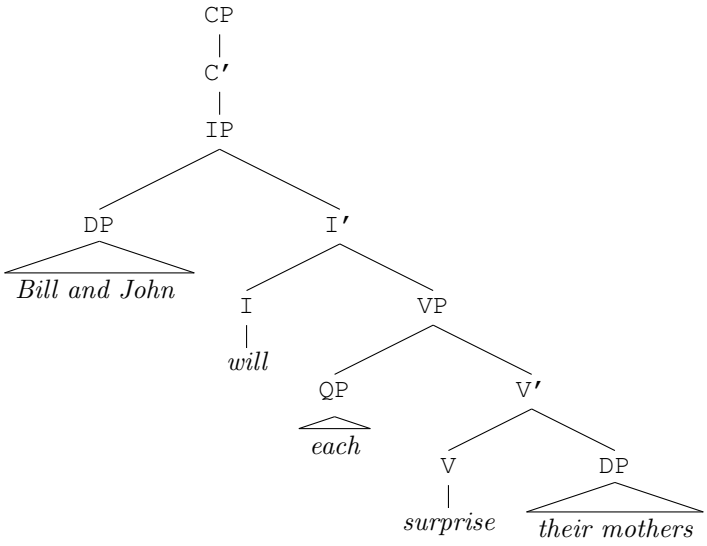
- one outlier in early period; otherwise no outliers

5.3.3 Floating quantifiers

5.3.3.1 Logic

- in English, the quantifiers *each*, *both* and *all* can float.
- analysed as element in Spec,VP

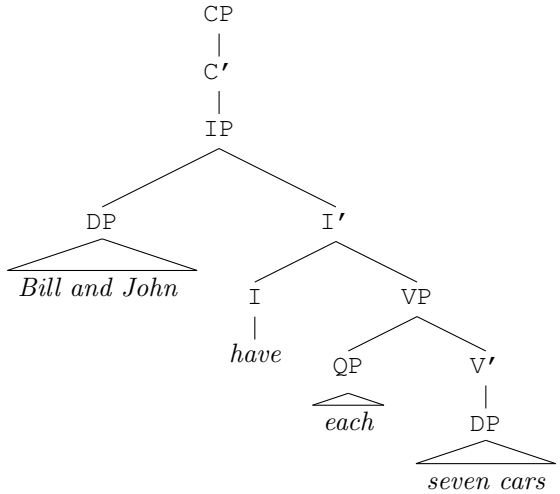
(35)



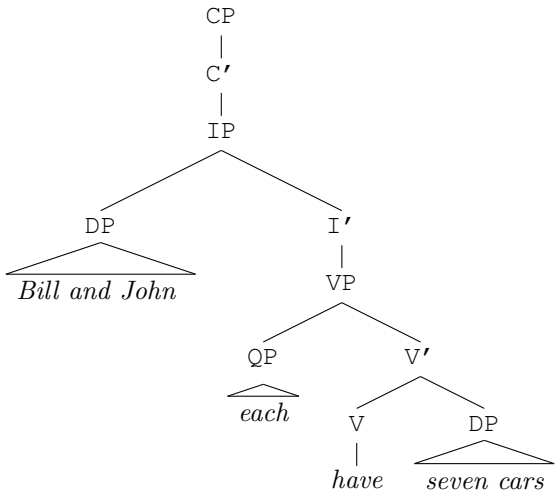
– the order '*have*-quantifier' should exist as long as *have* can be inserted under I
– but the order 'quantifier-*have*' should become the preferred option once *have* is V

(36)

a. conservative variant



b. innovative variant



5.3.3.2 Search queries

- lots of strict requirements on searches are necessary to be able to use floating quantifiers
  - *each* assumed to be unproblematic - all instances used
  - *both* and *all* require plural subjects; hence 3rd person singular inflection not searched for
  - *both* and *all* also require non-pronominal subjects

    (37) a. They will all do something.
    b. The car mechanics will all do something.

    (38) a. They all will do something.
    b. ?? The car mechanics all will do something.

    (39) a. I like them all.
    b. ?? I like the car mechanics all.

  - *all* is ambiguous with strong determiners (*they have all the cars*); required to occur with *same*

    (40) The first settlers in the interior have all the same story to tell
    NorthAmRev 1861

  - *both* must not occur with a coordinated object; I deleted phrases like *they had both X and Y*

    (41) but love and law have both their uncertainties
    NormanLeslieATale 1835

- small sample size, a total of only 595 examples

5.3.3.3 Examples

(42) conservative variant: *have* - Q
  a. Fichte, Schelling, Hegel, **had each** his own system, though they have been called transcendentalists.
  NewEngYaleRev 1843
  b. the discoveries in science, new courses of crops in agriculture, the extension of roads and canals, **have all** a tendency to increase the wealth of the country
  ShortHistoryPaper-money 1833
  c. the red limit and the violet limit **have both** the same luminous intensity
  TreatiseOnForces 1845

(43) innovative variant: Q - *have*
  a. His mother, Fraulein Schlote and Miss Letitia Lamb **each had** her own accent and intonation; but they had much the same vocabulary
  LastPuritan 1936
  b. American Negroes, Puerto Ricans, Indians and Mexicans **all have** an exceedingly direct stake in the Administration's posture in Vietnam.
  NYT-Let 1967
  c. Arcturus and Capella **both have** the same magnitude, 0.2.
  ExploringDistant 1956

5.3.3.4 Diachronic development

- Because of small sample size, s-shaped development may not be as obvious as in the other contexts

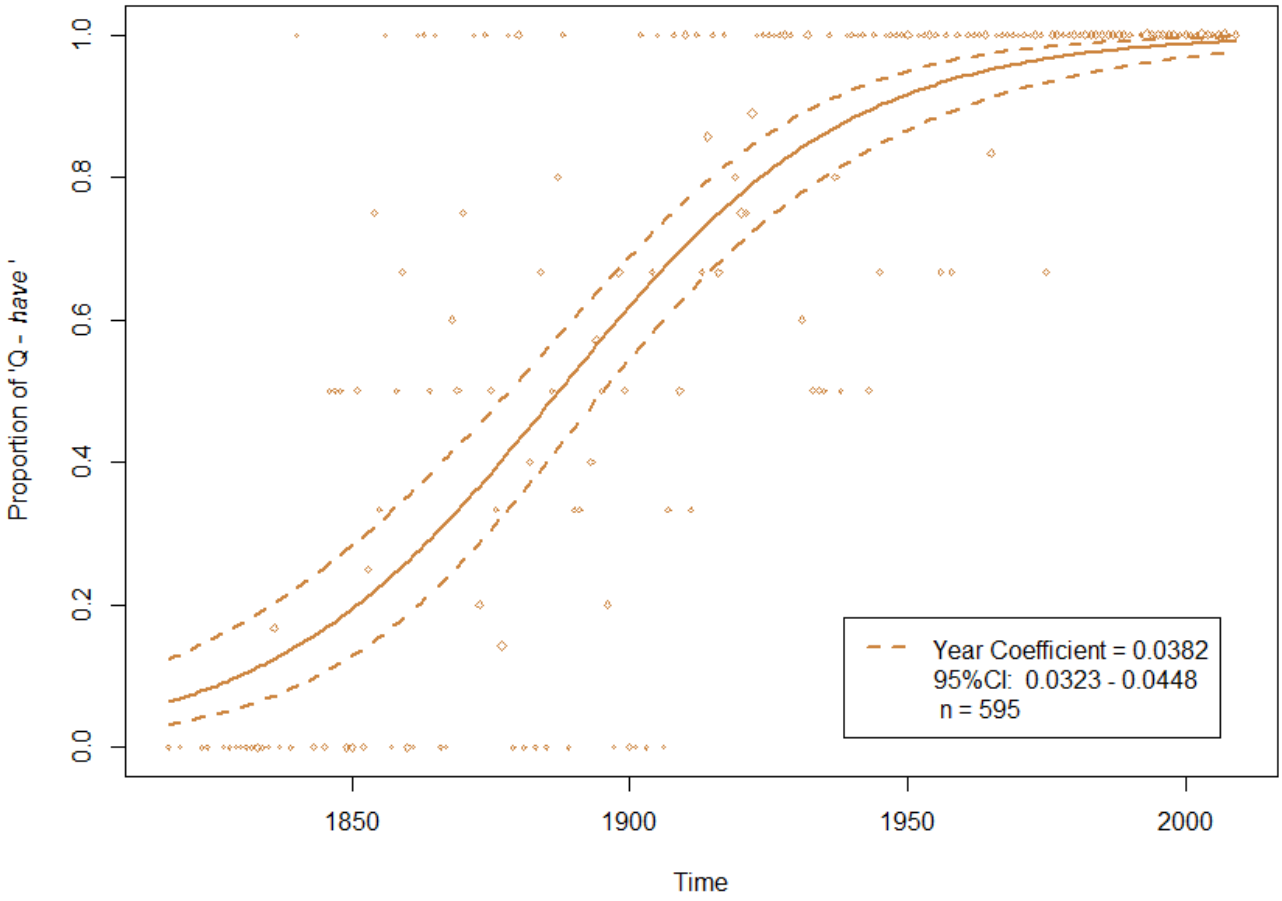


Figure 5: The development of the relative order of possessive *have* and floating quantifiers

  - 136 data points between 1810-2009
  - size of points is proportional to number of examples; standard point character = 25 examples
  - note considerable uncertainty for the rate of change

- despite small sample size, overall good fit to data
  - -2LL test: χ^2=283.37, df = 1, p<0.001***
  - Pseudo R^2: Hosmer and Lemeshow R^2 = 0.604
  - Predictive Accuracy: correct: 80.3%, baseline: 69.6%

- there are no outliers

5.4 Ellipsis phenomena

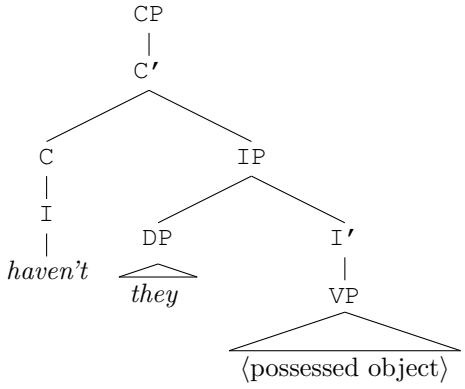
5.4.1 Negative tags

5.4.1.1 Logic

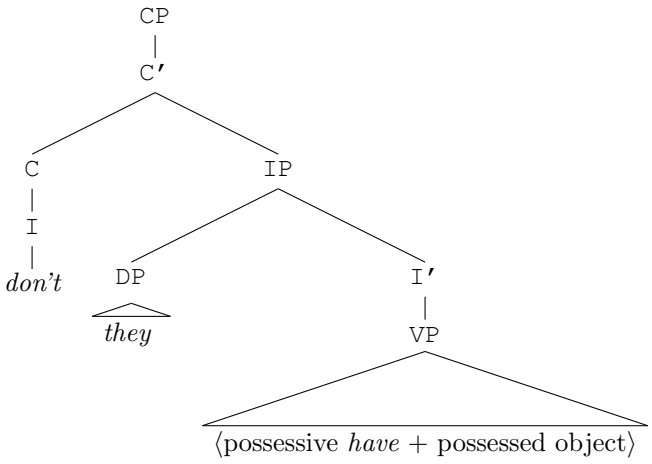
- ellipsis targets the VP
- prediction:
  - tag questions of the form '*have*-⟨VP⟩' should exist as long as *have* can be inserted under I
  - but tag questions of the form '*do*-⟨VP⟩' should become regularized as *have* is increasingly V

(44)

a. conservative negative tag



b. innovative negative tag



5.4.1.2 Negative tags - Search queries

- dependent variable: '*haven't*-subject?' vs. '*don't*-subject?' - 2
- independent variables:
  searches for different inflections + contraction negation
  - searches for `have|has|had`, `do|does|did` - 3
  - negation with and without contraction - 2
- use of collocate search option; a form of *have* should occur up to 9 words to the left
- only consider personal pronouns
- punctuation sign as heuristic for clause boundary
- example of a search query string:

```
have n't [pp*] ?|.|;|,|!|-|/|:
COLLOCATES have|has|had 9 0
vs.
do n't [pp*] ?|.|;|,|!|-|/|:
COLLOCATES have|has|had 9 0
```

- 2*3*2 = 12 search queries

- manual correction of precision errors

(45) a. *perfect have*
You **have** settled down, **haven't you**?
Play:ComicArtist 1928

b. *embedded possessive have*
I think we **have** our title song, **don't you**?
SongcatcherBallad 2002

c. *non-finite have*
you must **have** a kitchen somewhere, **haven't you**?
Play:DaddysGoneA 1921

d. *causative have*
And you **had** Pierre kill Powell, **didn't you**?
Mignon 1962

e. *modal have*
The train **has** to have water,**doesn't it**, Sergeant?
HawkeLawFastGun 2006

- total of 512 examples

5.4.1.3 Examples

(46) conservative variant: *have not*-subject

a. Destiny **has** brilliant spokes in her wheel, **hasn't she**?
Morgesons 1862

b. You **had** a daughter too, sir, **had you not**?
ItalianFather 1810

(47) innovative variant: *do not*-subject

a. Mary still **has** some of those candy bars, **doesn't she**?
INeverPromised 1964

b. You **had** such an interview, **did you not**?
StrangeDisappearance 1880

5.4.1.4 Diachronic development

- steady increase in *do*-support, but because of small sample s-shaped curve is not obvious

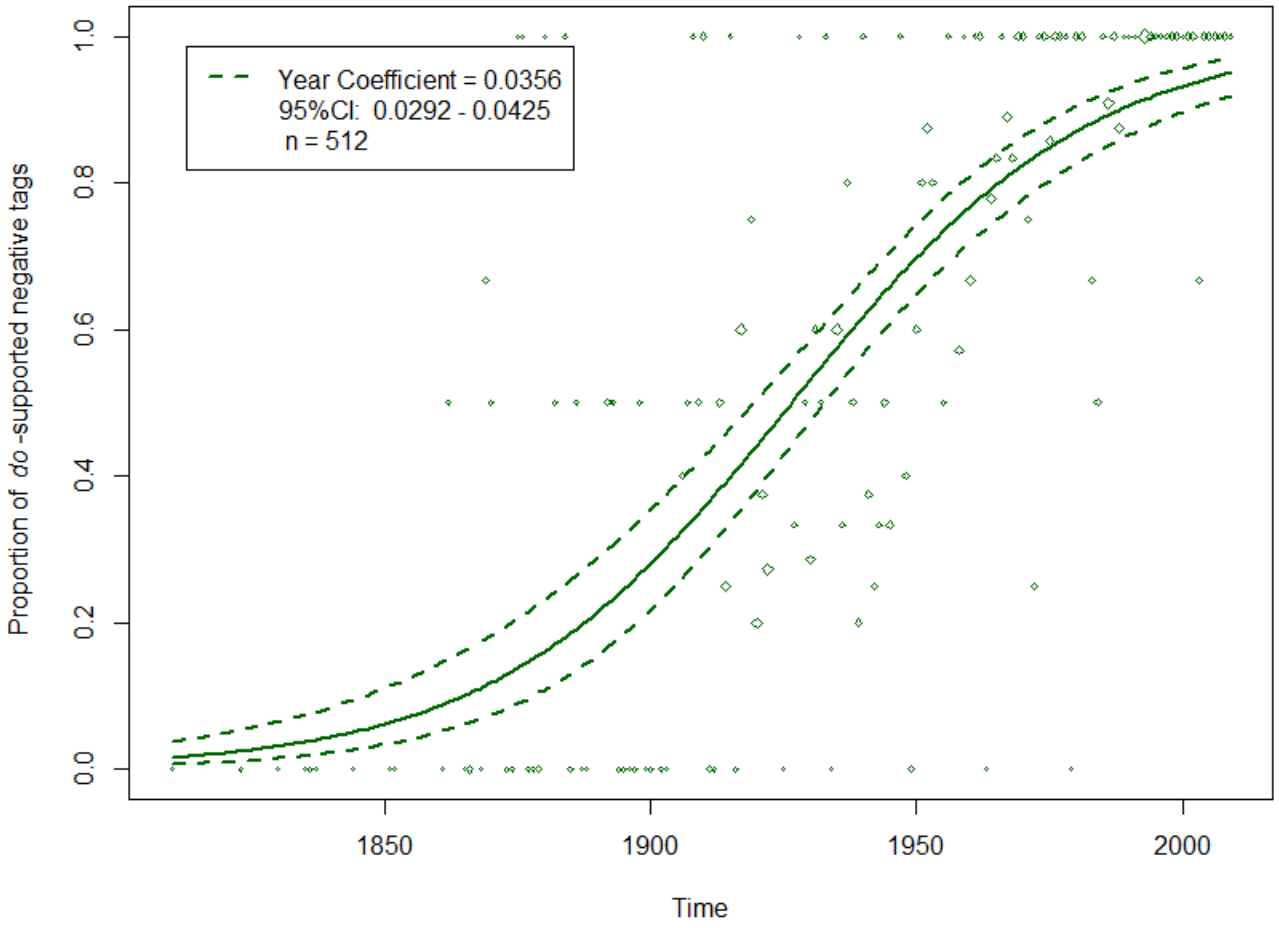


Figure 6: The development of *do*-support with possessive *have* innegative tags

  - again sufficient material for 198 data points, 1810-2009 except 1813 and 1816
  - size of points is proportional to number of examples; standard point character = 25 examples
  - small sample size leads to considerable uncertainty

5.4.1.5 Model evaluation

- overall acceptable fit to data
  - -2LL test: χ^2=165.83, df = 1, p<0.001***
  - Pseudo R^2: Hosmer and Lemeshow R^2 = 0.477
  - Predictive Accuracy: correct: 74.4%, baseline: 63.2%
- there are no outliers

5.4.2 Other ellipsis

- there exist a large number of other ellipsis contexts

(48) *positive tags*

a. Miss Penelope Perry, you **have** no insuperable objections to yielding, **have you**?
ClintonBradshaw 1835

b. Certainly you **have** no objections to my sitting next you, **do you**?
CryChildren 1952

(49) *positive ellipsis e.g. in comparative clauses*

a. and he **has** more power over them than **I have**.
WildernessBraddocks 1823

b. they **have** so many more obstacles than **I did**.
Atlanta 2007

(50) *negative ellipsis, e.g. in affirmative main clauses*

a. She dearly wished that she **had** a tongue, to explain. But **she hadn't**.
DepthsGlory 1985

b. [She] only wished she **had** another choice. Unfortunately, **she didn't**.
ApocalypseTroll 2000

- unfortunately, I haven't finished the manual correction of the precision errors
- it is hoped that increasing the sample size will result in a more precise point estimate of the rate of change than the context of negative tags alone can provide

6 Hypothesis Testing

- The rate of change from conservative possessive *have* of category `I` to innovative *have* of category `V` should be identical in all contexts
- Are the slopes of the curves presented in the last section really identical?

6.1 Graphical evaluation

- the following graph shows all the data from the preceding section combined
  - if the Constant Rate Hypothesis is correct, the curves should be parallel to each other

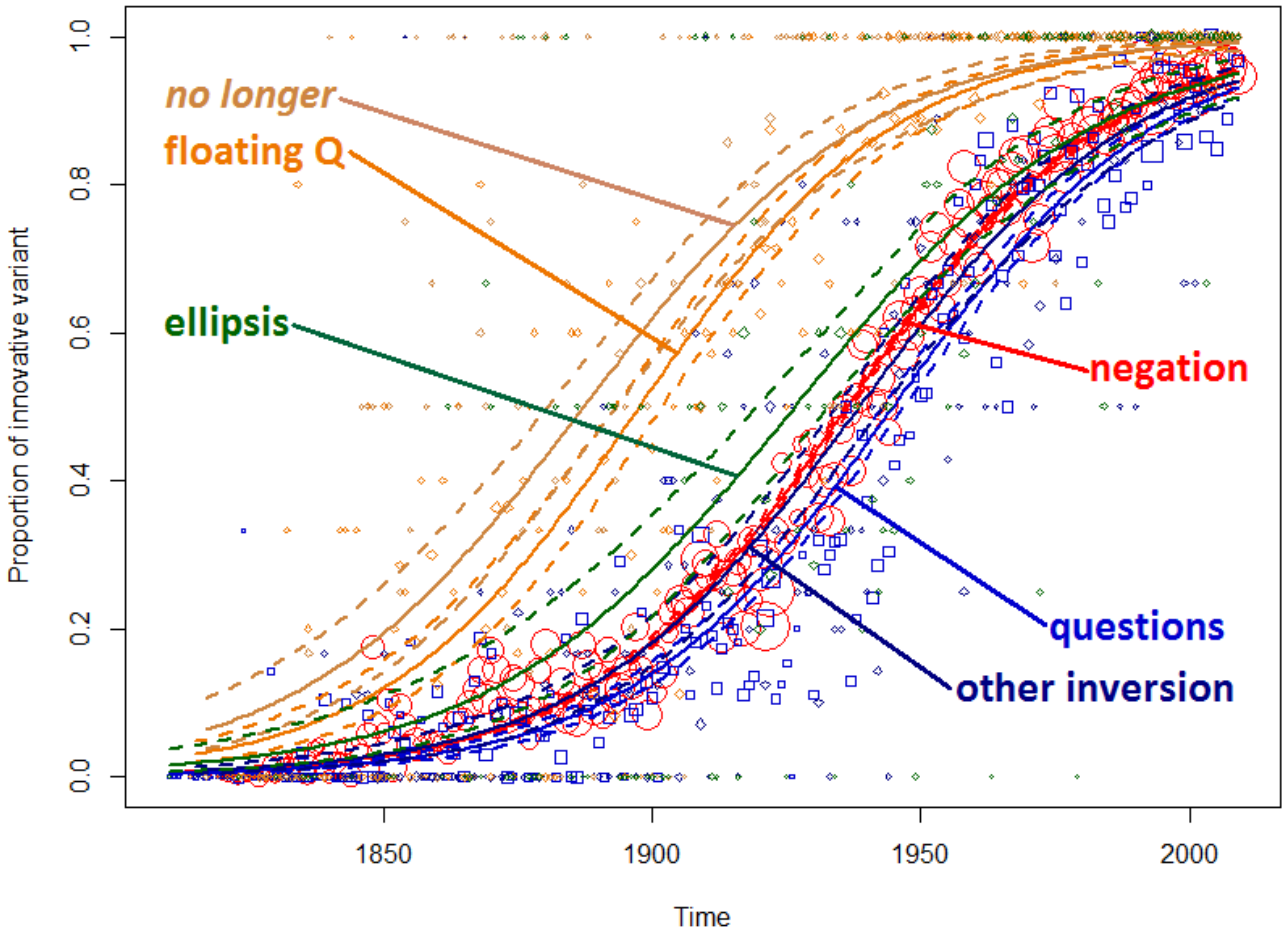


Figure 7: The development of innovative possessive *have* in six syntactic contexts

  - the curves look quite parallel to each other
  - the `VP` diagnostic contexts show a considerably overall probability of innovative possessive *have* than the other contexts
  - however, it is hard to assess if the curves are really parallel to each other because of their s-shaped form

- Logit plot
  - it is possible to "flatten out" the curves by plotting their logit transform

    (51) $p = \frac{e^{a+\beta*year}}{1+e^{a+\beta*year}}$

    (52) $ln(\frac{p}{1-p}) = a + \beta * year$ (logit transform)

  - the following graph shows only the point estimates (no CIs) for the year coefficient in the six contexts under investigation

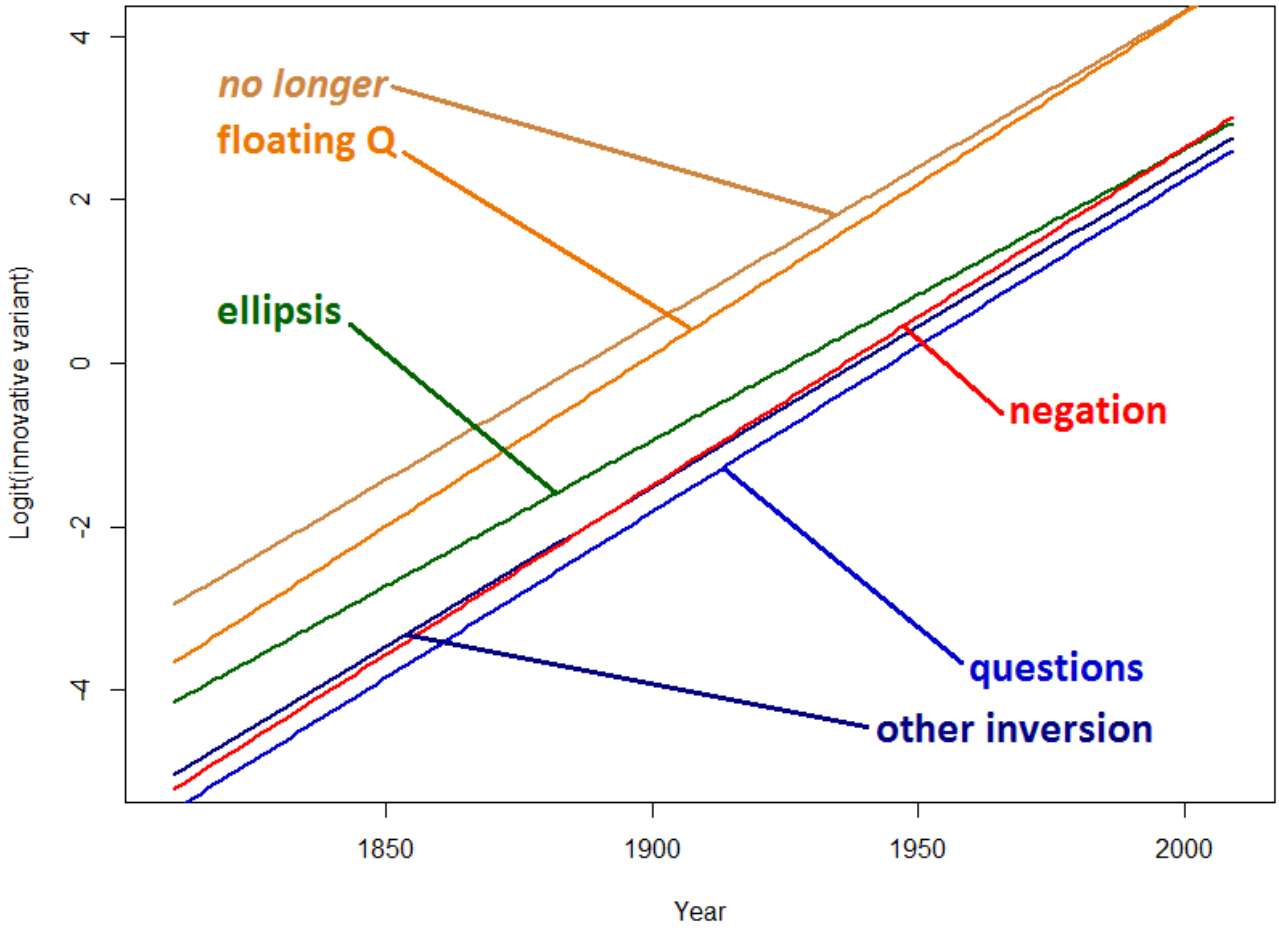


Figure 8: The development of the logit of innovative possessive *have* in six syntactic contexts

  - the lines look quite parallel to each other
  - only floating quantifiers and ellipsis (negative tag questions) show a slightly different slope - these are exactly the contexts with the least amount of data

→ inspection of the graphs suggest that the rates of change can plausibly be assumed to be identical

6.2 Test for significance of interaction effect

- the following is a common rationale for investigating the Constant Rate Hypothesis:
  - it is not possible to test for identity of two test statistics
  - however, it is possible to test if the six Year coefficients are significantly different from each other
  - fit a combined logistic regression model that includes 'Context' as an additional variable (here: six values for the six contexts investigated)
  - also include interaction term between 'Year' and 'Context'
  - essentially this interaction term will allow the regression lines to have different slopes for every context
  - if the six contexts develop independently of each other, the interaction effect may be significant
  - if the six contexts develop in the same manner, the addition of the interaction effect to a model with only main effects must not significantly improve the model fit
  - hence: if the interaction term is not significant, this lends credence to the Constant Rate Hypothesis (even though absence of evidence is not evidence of absence)
- Table 1 presents an Analysis of Deviance table comparing a model with to a model without the interaction effect.

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

Model	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	1075	1625.6			
2	1070	1621.2	5	4.3679	0.4978

Table 1: Analysis of Deviance table for combined model

  - as expected, the interaction term is not significant
  - there is no good reason to assume that the rates of change are significantly different for the six contexts
  - in fact, there is no evidence that any context develops innovative possessive *have* at a significantly different rate from any other context, as shown in Table 2 below

	first context included in the model					
second contexts included in the model	NEG	Q	OthInv	NoLong	FloatQ	NegTag
NEG	-	-	-	-	-	-
Q	0.55	-	-	-	-	-
OthInv	0.37	0.59	-	-	-	-
NoLong	0.81	0.62	0.42	-	-	-
FloatQ	0.36	0.50	0.82	0.37	-	-
NegTag	0.11	0.18	0.40	0.14	0.57	-

Table 2: p-values for model improvement with interaction effect in context-by-context comparisons

→ model comparison also supports the hypothesis that the contexts change at identical rates

7 Conclusion

- this study presented one of the must substantial tests of the Constant Rate Hypothesis to date
- the evaluation was based on one of the largest data sets of a syntactic change ever collected
- as a result, the statistical evaluation is far more precise than previous investigations

  → the results strongly support the validity of the hypothesis

  "A theory is tested not merely by applying it, or by trying it out, but by applying it to very special cases - cases for which it yields results different from those we should have expected without that theory, or in the light of other theories. In other words we try to select for our tests those crucial cases in which we should expect the theory to fail if it is not true. [...] It is an attempt to refute it; and if it does not succeed in refuting the theory in question - if, rather, the theory is successful with its unexpected prediction - then we say that it is corroborated by the experiment. **It is the better corroborated the less expected, or the less probable, the result of the experiment has been.**"
  (Karl Raimund Popper,*Conjectures and Refutations* 1963: 150)

- the Constant Rate Hypothesis survived a very hash falsification attempt - it would very easily have been possible for the slopes of the different contexts to be different
- the Constant Rate Hypothesis must be regarded as one of the best corroborated findings in all of quantitative historical linguistics

References

Altmann, G. (1983), Das piotrowski-gesetz und seine verallgemeinerungen, *in* 'Exakte Sprachwandelforschung. Theoretische Beitrge, 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.

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.