Language change is a function of historical demography

Freek Van de Velde

University of Leuven

Quantitative Lexicology and Variational Linguistics – QLVL

Take-home message

1. Language change is a function of historical demography
2. Demographic upheaval leads to morphosyntactic change
3. Influx of L2 speakers makes languages more analytic
4. This can be seen across language families, but also within language families
5. This can be seen across languages, but also within languages

Internal-external divide fallacy

- Many linguists are wary of external explanations for language change (Lass, Kurylowicz, Ohala ... see e.g. Woods 2001: 974-975)

> "In view of the confusion and controversies surrounding causes of language change, it is not surprising that some reputable linguists have regarded the whole field as a disaster area, and opted out altogether." (Aitchison 1991: 106)

Reasons for the wariness

- If morphosyntax is too responsive to demographic change, this potentially undermines universalism, which reeks of crypto-racism and discredited romantic ideas (Herder, Humboldt, Schiller ...) about the deep connection between language and people

- There is no shortage of crackpot theories

- Results of research into the impact of demographic factors on language change are unclear:
  - Nettle (1999): smaller languages, faster change
  - Wichmann & Holman (2009): 1° No clear effects: "The test shows mainly negligible effects of population", 2° "... the exception being an apparently faster rate of change in the larger of two closely related variants."

Morphological types and demography

- Two big, related themes with a time-honoured scholarly interest:
  1. Morphological types
  2. Demographic correlations

Types of languages

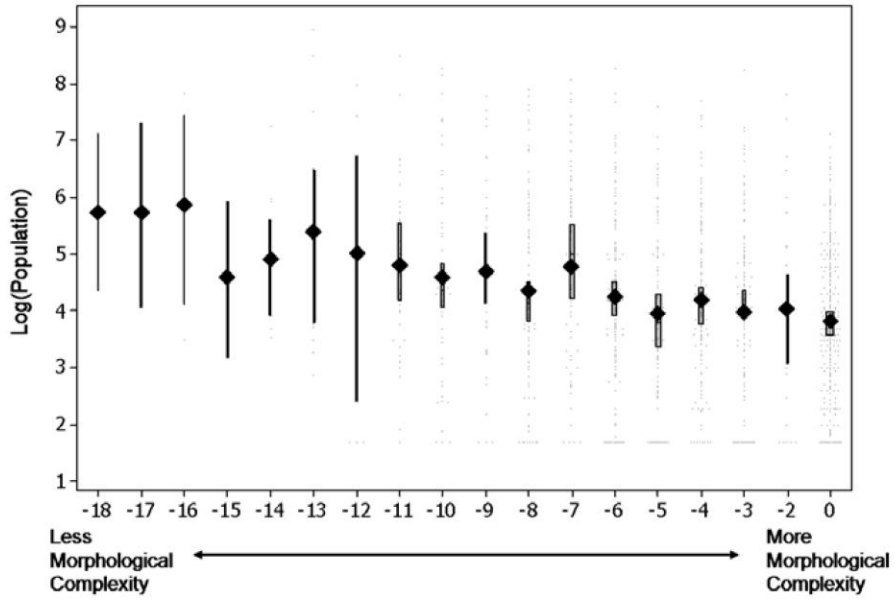
- Morphological types: isolating, agglutinative, fusional, polysynthetic, introflexive ... (disagreement and confusion, Bickel & Nichols 2013)
- Analytic vs. synthetic:

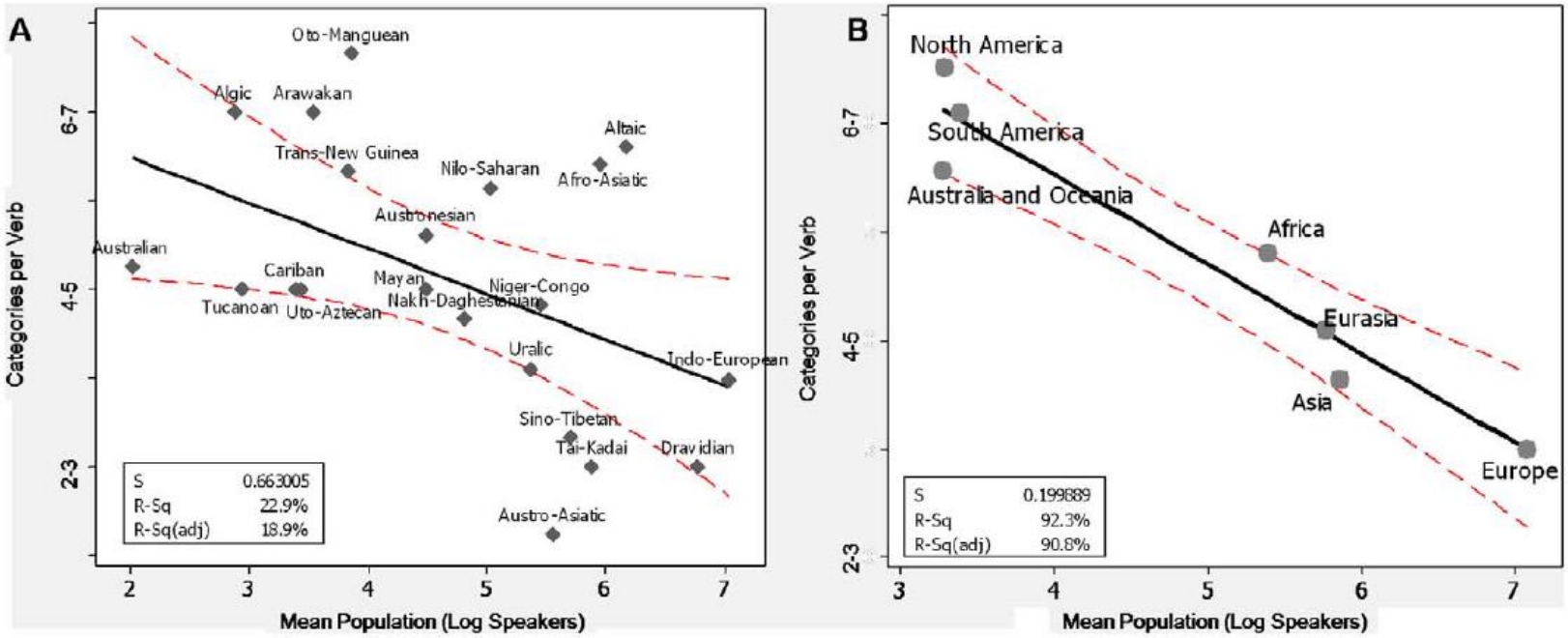
> "En Europe les langues dérivées du latin, et l'anglais, ont une grammaire tout analytique (...) synthétiques dans leur origine (...) elles penchent fortement vers les formes analytiques" (Von Schlegel 1846:161, cited in Szmrecsanyi 2012)

Demographic explanations for the analytic-synthetic difference

- Lupyan & Dale (2010)
  - Due to evolutionary pressures, languages adapt to their community (see also Christiansen and Chater 2008)

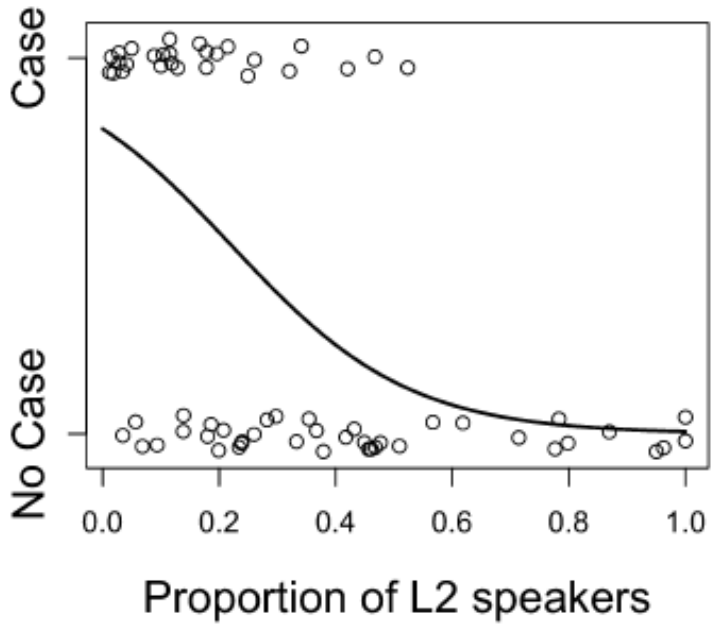
    ⇨ *Linguistic Niche Hypothesis*
  - **Esoteric languages**: morphologic complexity, redundancy, synthetic, favouring L1 acquisition ⇨ smaller languages
  - **Exoteric languages**: analytic-syntactic complexity, transparency, analytic, favouring L2 acquisition ⇨ patterns with bigger languages

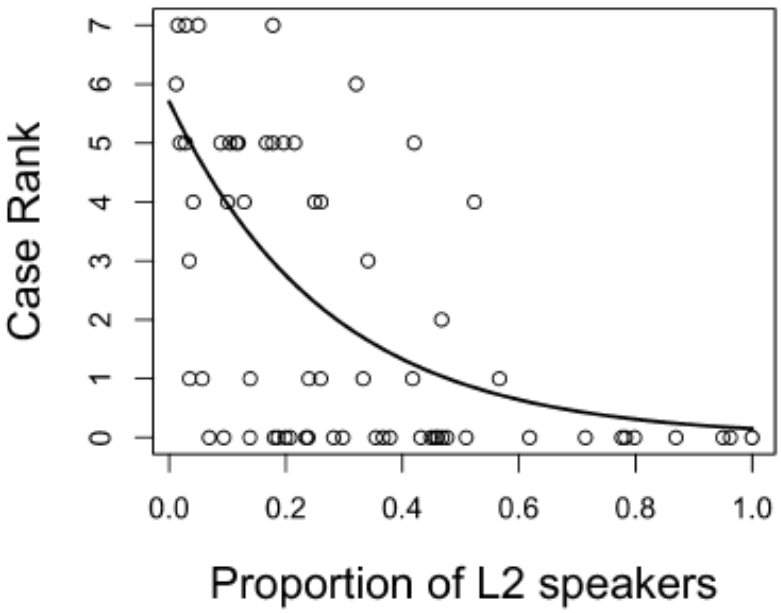

Log(Population)
9
8
7
6
5
4
3
2
1
-18 -17 -16 -15 -14 -13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 0
Less Morphological Complexity
More Morphological Complexity


A
Oto-Manguean
Algic
Arawakan
Altaic
Trans-New Guinea
Nilo-Saharan
Afro-Asiatic
Austronesian
Australian
Cariban
Mayan
Niger-Congo
Nakh-Daghestanian
Tucanoan
Uto-Aztecan
Uralic
Indo-European
Sino-Tibetan
Tai-Kadai
Dravidian
Austro-Asiatic
S 0.663005
R-Sq 22.9%
R-Sq(adj) 18.9%
Categories per Verb
Mean Population (Log Speakers)
B
North America
South America
Australia and Oceania
Africa
Eurasia
Asia
Europe
S 0.199889
R-Sq 92.3%
R-Sq(adj) 90.8%
Categories per Verb
Mean Population (Log Speakers)

Demographic explanations for the analytic-synthetic difference

- Bentz & Winter (2013)
  - Data:
    - 226 languages using the *SIL Ethnologue*, the *Rosetta project website* and the *UCLA Language Materials Project*; area and family information from *AUTOTYP database,* case information from *WALS*
    - Overlap: 66 languages (26 language families, 16 areas)
  - Operationalisation:
    - L2 speakers: adult L2 speakers as opposed to early bilinguals
    - Case: productive morphological inflections of nouns (loose definition: possessive clitic *-s* in English is counted as case)
  - Method:
    - Generalized linear mixed effects models: logistic regression (case vs. no-case), and negative binomial regression (count of case). Response variable: case; explanatory variable: proportion of L2 speakers
    - Throw in population count in the regression models to see whether it is a predictor on top of the L2 proportion. ⇨ It isn't.


Case
No Case
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of L2 speakers


Case Rank
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of L2 speakers

Demographic explanations for the analytic-synthetic difference

- Lupyan & Dale (2010), Bentz & Winter (2013): synchronic quantitative evidence for Von Schlegel's diachronic claim

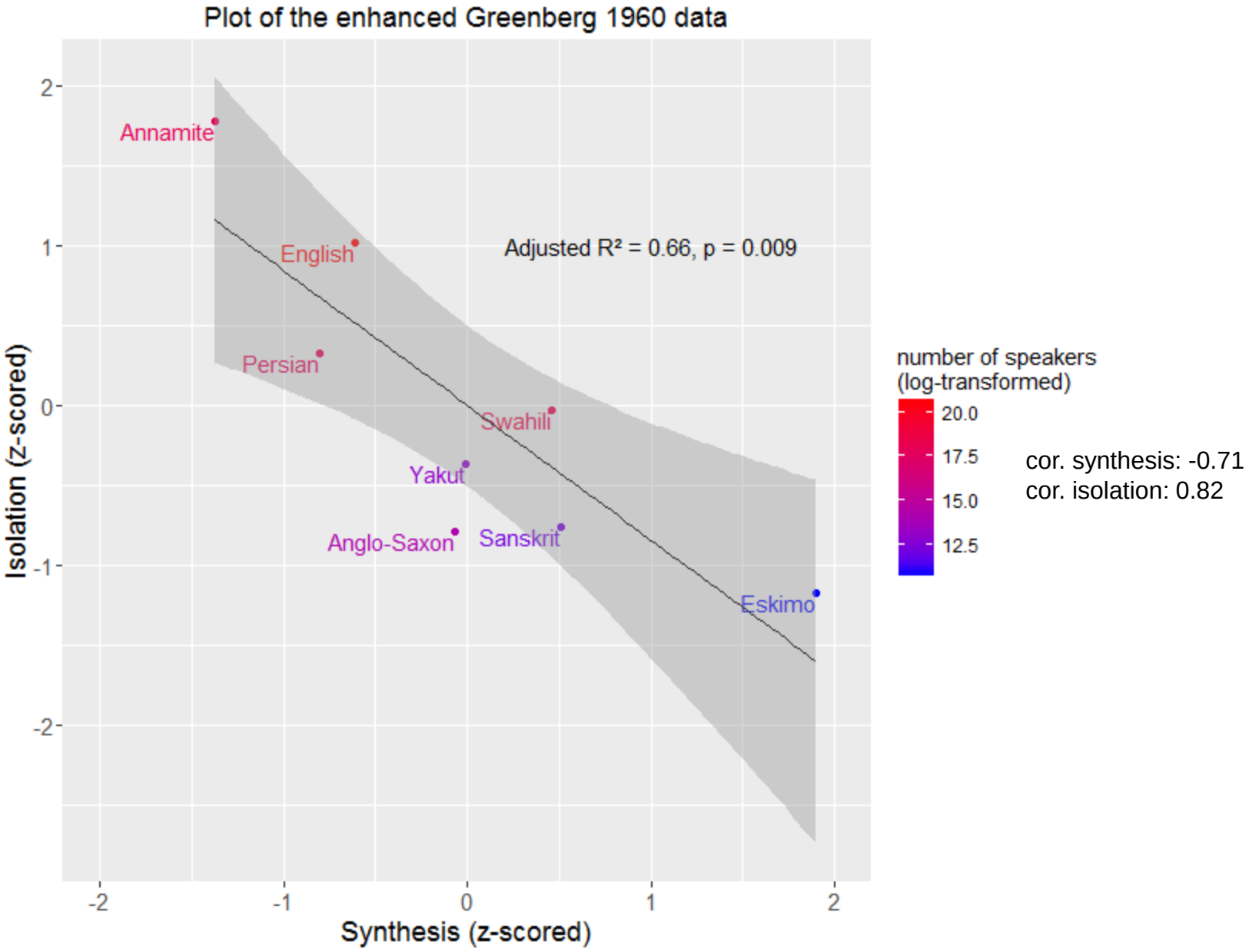
- What about diachronic quantitative evidence? (see Kusters 2003; Szmrecsanyi 2012)

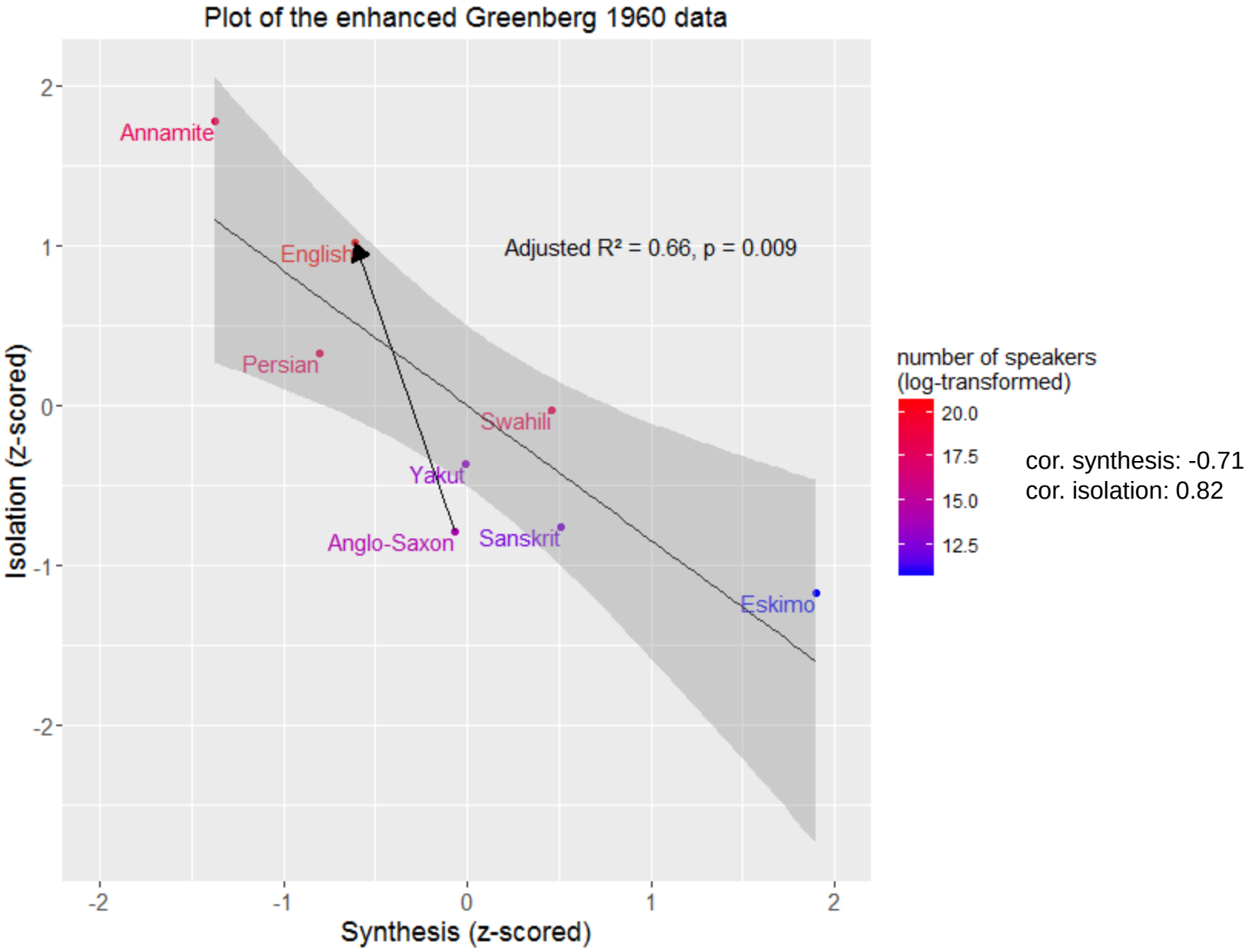
From synchrony to diachrony

Greenberg (1960)

- Index of synthesis (proportion of morphemes to words)
- Index of isolation (proportion of word order as a grammatical marker to the total number of nexus)
- Along with a number of other indices (Index of agglutination, Index of compounding, Index of inflection, Index of prefixation ...).
- Calculated on 100 word stretches of different languages (labour-intensive):

	Sanskrit	Anglo-Saxon	Persian	English	Yakut	Swahili	Annamite	Eskimo
Synthesis	2.59	2.12	1.52	1.68	2.17	2.55	1.06	3.72
Isolation	.16	.15	.52	.75	.29	.40	1.00	.02


Plot of the enhanced Greenberg 1960 data
Annamite
English
Persian
Swahili
Yakut
Anglo-Saxon
Sanskrit
Eskimo
Adjusted R² = 0.66, p = 0.009
Isolation (z-scored)
Synthesis (z-scored)
2
1
0
-1
-2
-2
-1
0
1
2
number of speakers
(log-transformed)
20.0
17.5
15.0
12.5
cor. synthesis: -0.71
cor. isolation: 0.82


Plot of the enhanced Greenberg 1960 data
Annamite
English
Persian
Swahili
Yakut
Anglo-Saxon
Sanskrit
Eskimo
Adjusted R² = 0.66, p = 0.009
Isolation (z-scored)
Synthesis (z-scored)
2
1
0
-1
-2
-2
-1
0
1
2
number of speakers
(log-transformed)
20.0
17.5
15.0
12.5
cor. synthesis: -0.71
cor. isolation: 0.82

Two case studies

- MORPHOLOGY: weak preterite in Germanic (*helped* < *hVlp*)
- SYNTAX: new future in Early Modern English (*going to*)

⇨ The two poster childs of grammatical language change (in English)

Case study: weak preterites

- Germanic languages have two morphological strategies for building preterites (not counting analytic perfects, *he has written a book*):

1. Strong inflection:
   - English *sing – sang*
   - Ablaut, based on Indo-European aspectual system (perfect > preterite)

PIE root *b^h_id^h-	e-grade (present)	o-grade (perfect)
Greek	*peíth-omai*	*$_2$pé-poith-a*
Gothic	*beid-an*	**baid-* (PGm ă < PIE ŏ)

2. Weak inflection
   - English *work – worked*
   - Dental suffix, based on a analytic formation [VERB + *d^heh_1-, *d^hoh_1- ('did')]

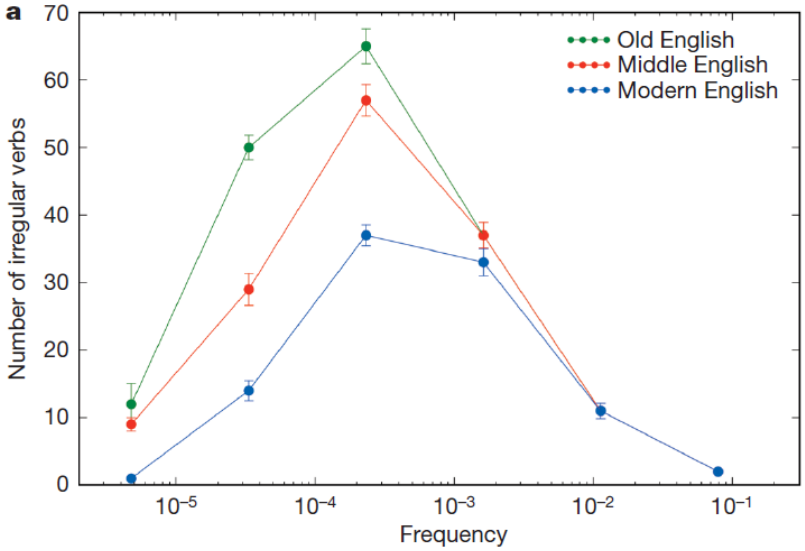
Gothic	*beid-an*	**baid-*
Dutch	*beid-en*	*beid-de*

Case study: weak preterites

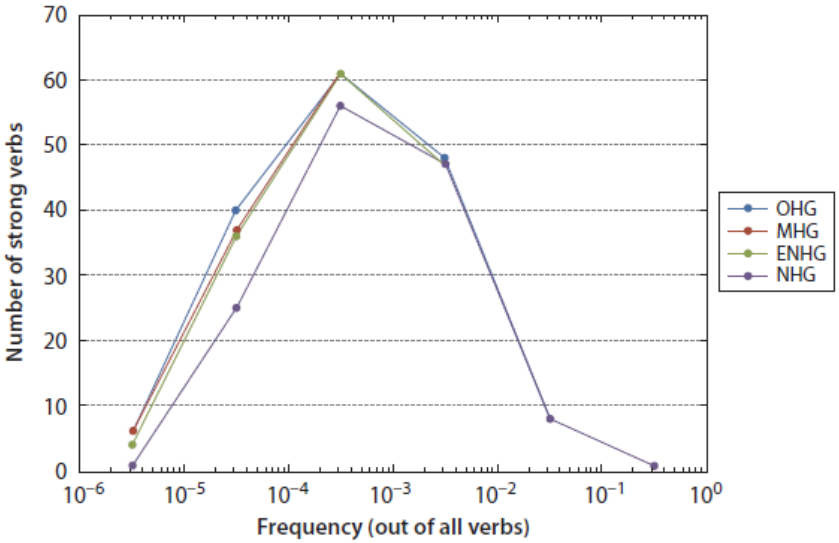
- Various changes occur:
  - irregularisation (Eng. *buy* – *bought*)
  - one strong ablaut class to another (Du. *heffen* – *hief* < *hoef* (Germ. *hob*, *hub*))
  - weak to strong (Du. *vragen* – *vroeg* (vs. Germ. *fragte*))
  - strong to weak (Eng. *carve* – *carved* < *cearf* (Du. *kerfde* < *karf*))

⇨ Long-term drift, over many centuries

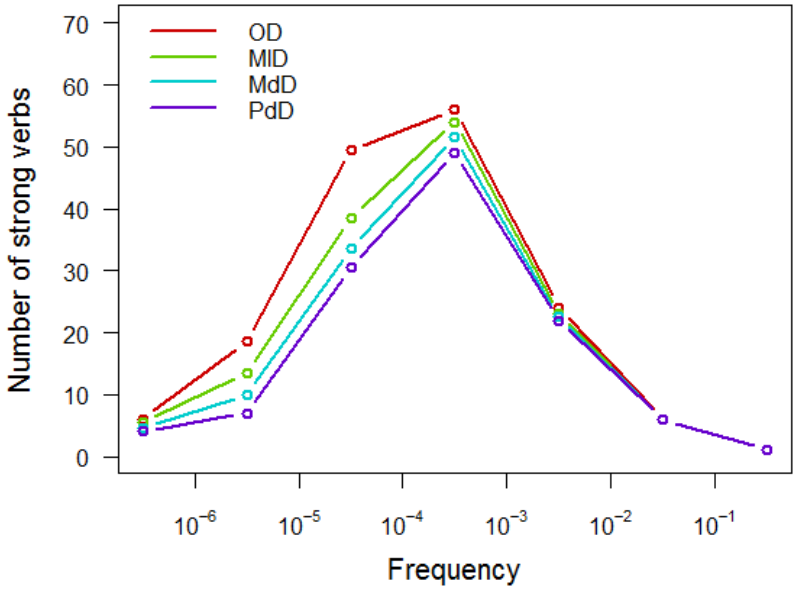
ENGLISH: Lieberman et al. 2007



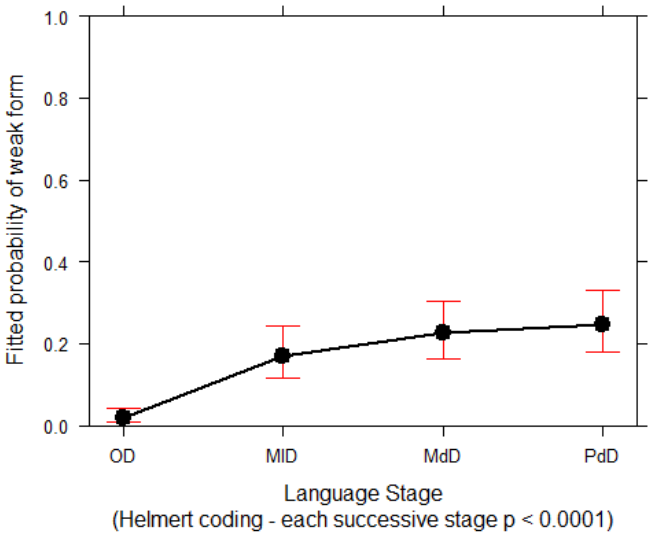
GERMAN: Carroll et al. 2012



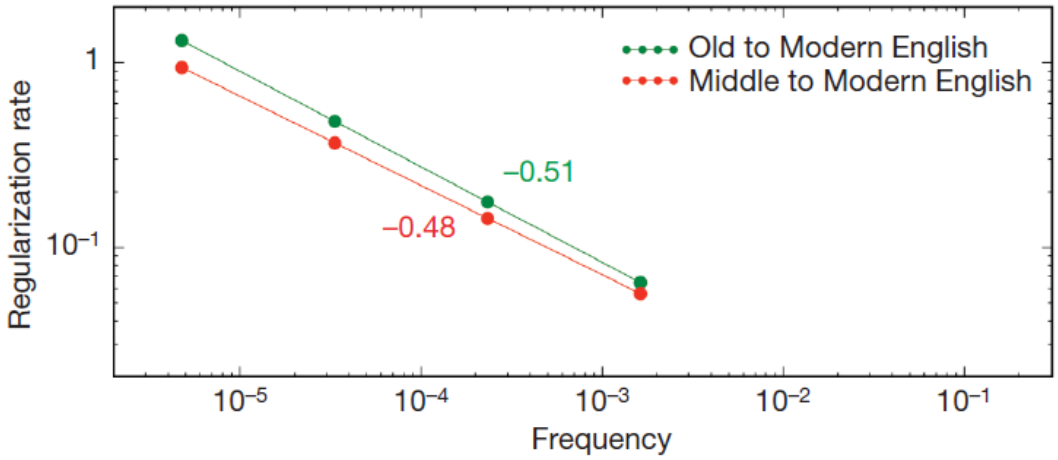
DUTCH (joint work with Isabeau De Smet)



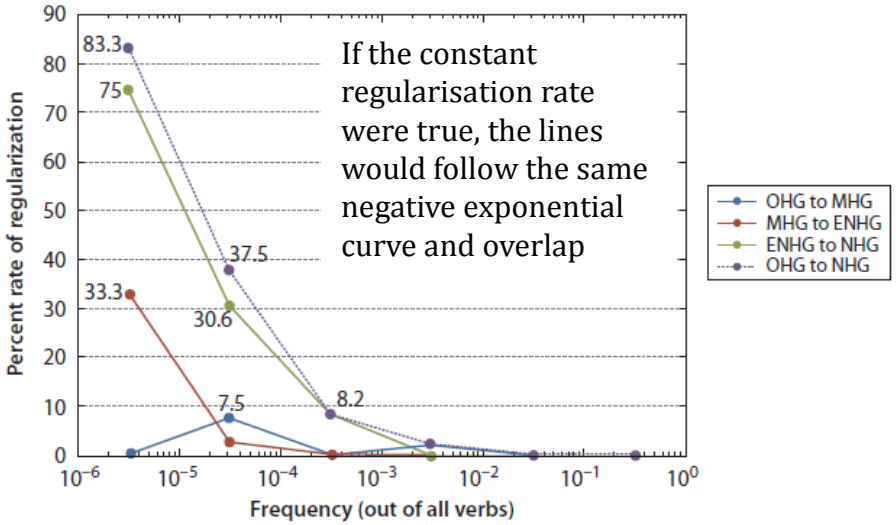
Partial effect plot
multiple logistic regression (incl. frequency)



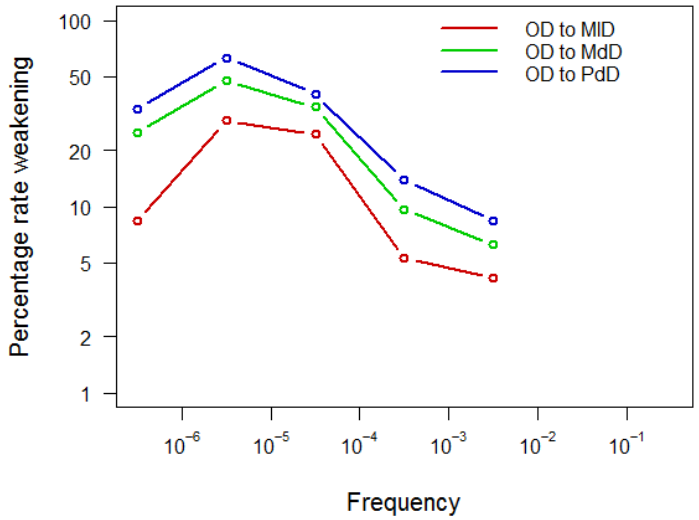
Lieberman et al. 2007: Constant rate of regularisation through time, only dependent on frequency



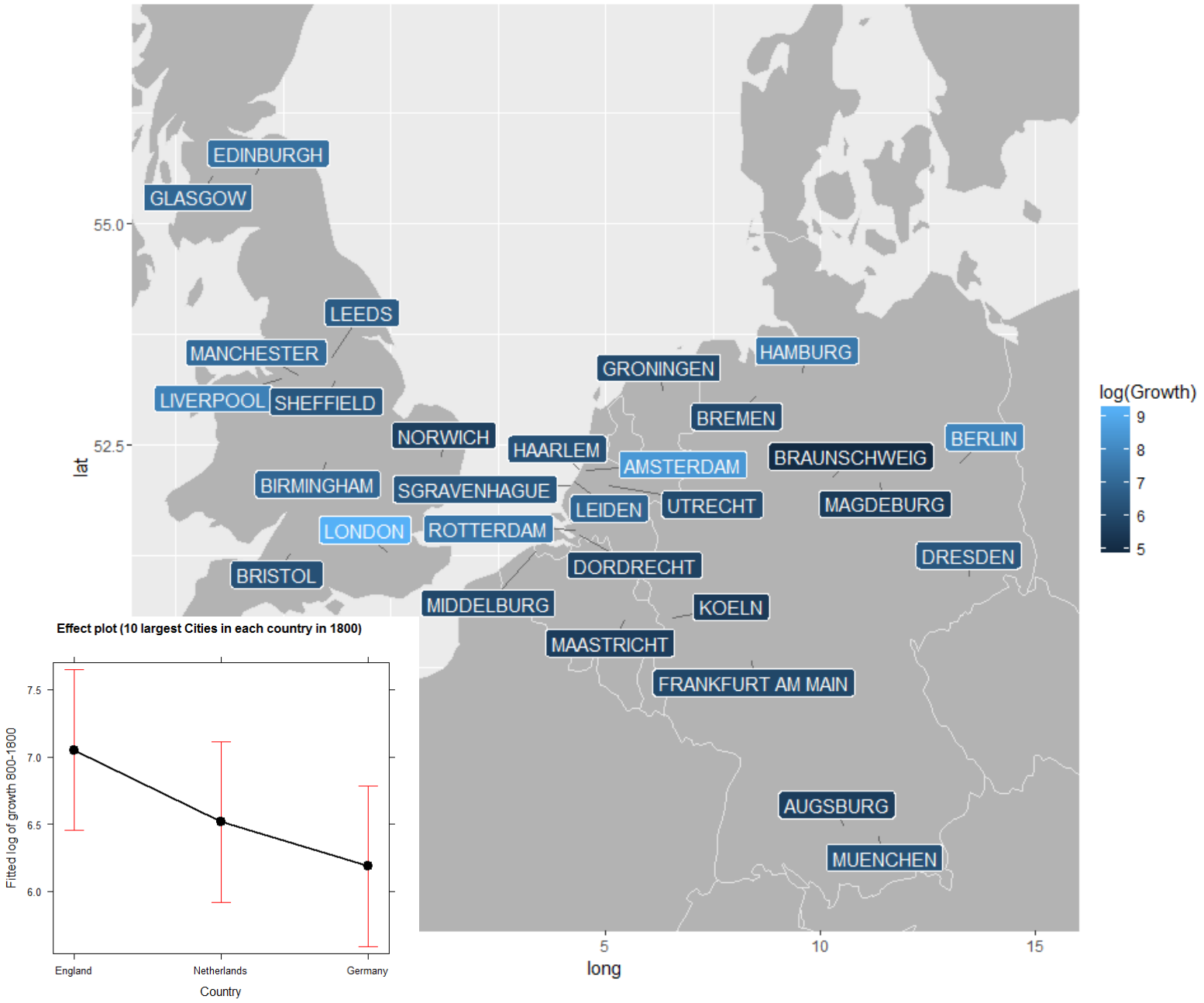
Carroll et al. 2012: Constant rate does not work for German

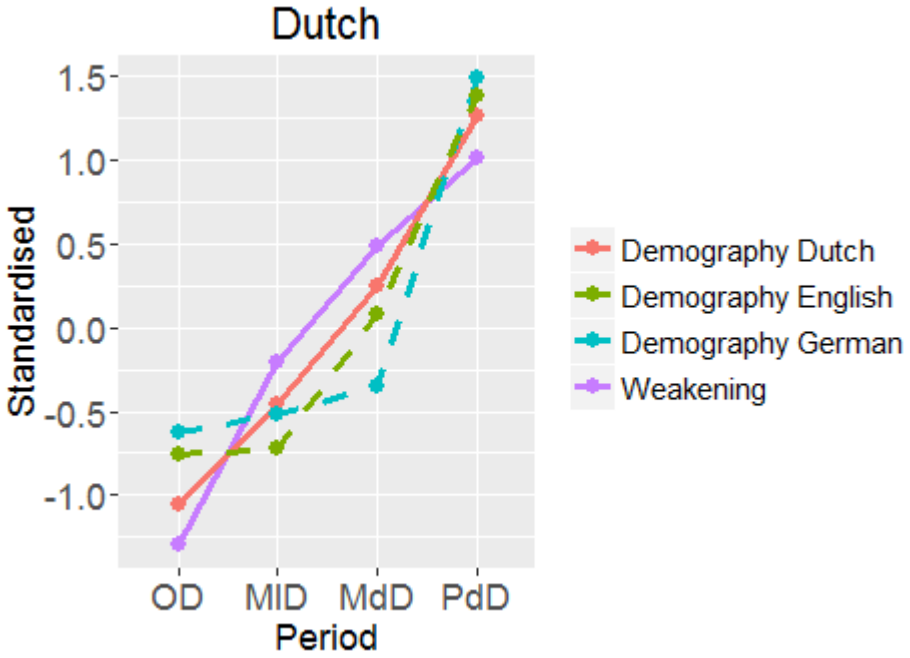


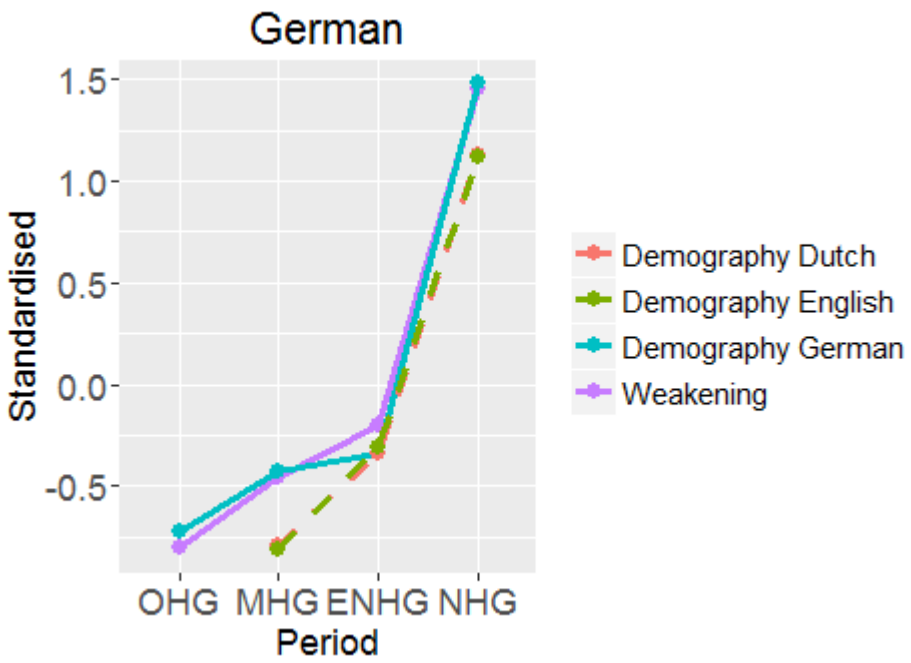
... neither for Dutch

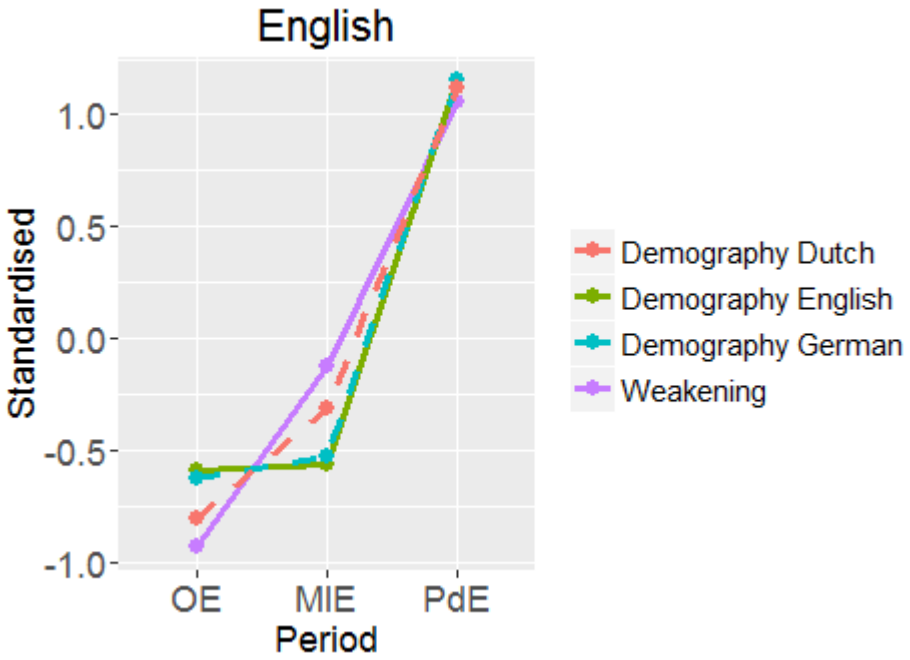



EDINBURGH
GLASGOW
LEEDS
MANCHESTER
LIVERPOOL
SHEFFIELD
NORWICH
BIRMINGHAM
LONDON
BRISTOL
SGRAVENHAGUE
ROTTERDAM
HAARLEM
AMSTERDAM
LEIDEN
UTRECHT
DORDRECHT
MIDDELBURG
MAASTRICHT
GRONINGEN
BREMEN
HAMBURG
BRAUNSCHWEIG
BERLIN
MAGDEBURG
DRESDEN
KOELN
FRANKFURT AM MAIN
AUGSBURG
MUENCHEN
55.0
52.5
50.0
47.5
lat
-5
0
5
10
15
long
log(Growth)
9
8
7
6
5


EDINBURGH
GLASGOW
LEEDS
MANCHESTER
LIVERPOOL
SHEFFIELD
NORWICH
BIRMINGHAM
LONDON
BRISTOL
SGRAVENHAGUE
ROTTERDAM
HAARLEM
AMSTERDAM
LEIDEN
UTRECHT
DORDRECHT
MIDDELBURG
MAASTRICHT
GRONINGEN
HAMBURG
BREMEN
BRAUNSCHWEIG
BERLIN
MAGDEBURG
DRESDEN
KOELN
FRANKFURT AM MAIN
AUGSBURG
MUENCHEN
55.0
52.5
lat
5
10
15
long
log(Growth)
9
8
7
6
5
Effect plot (10 largest Cities in each country in 1800)
Fitted log of growth 800-1800
7.5
7.0
6.5
6.0
England
Netherlands
Germany
Country



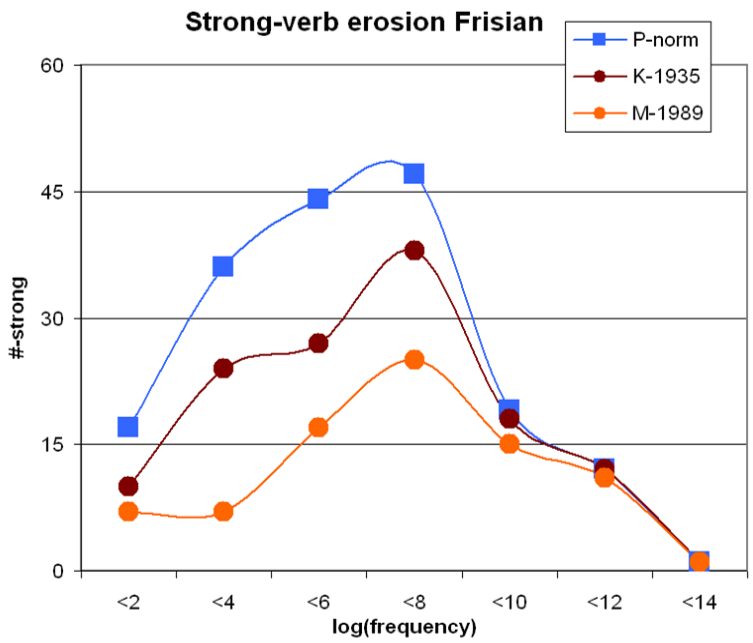


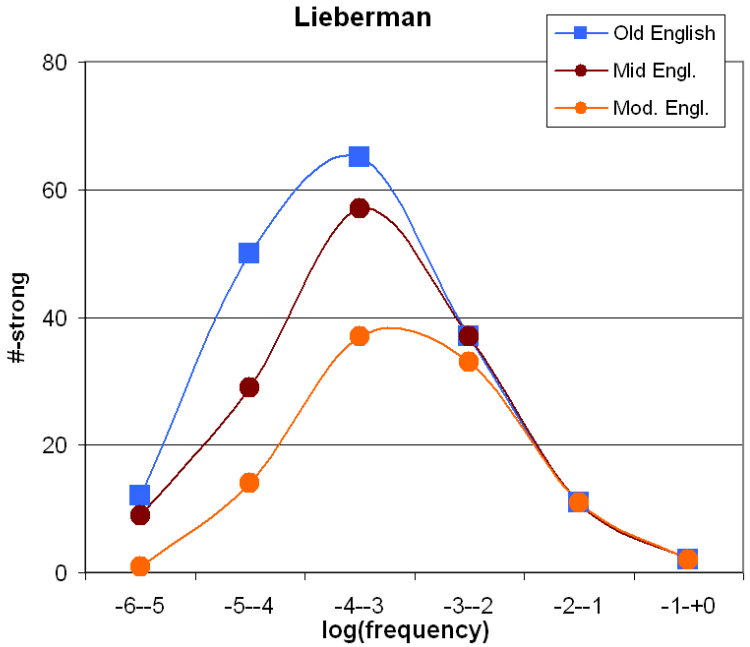


Average of largest city in each century covering the linguistic periods in each area

log(inh) ⇨ Weakening ⇩	English	Dutch	German
English	0.95 (n.s.)	0.98 (n.s.)	0.95 (n.s.)
Dutch	0.94 (n.s.)	**0.99****	0.82 (n.s.)
German	0.86 (n.s.)	0.97 (n.s.)	**0.99***

Time scales can be radically different

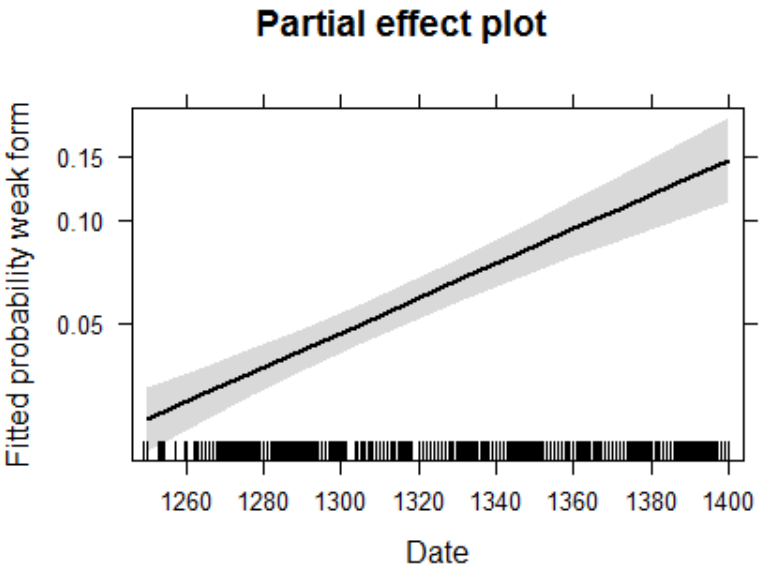


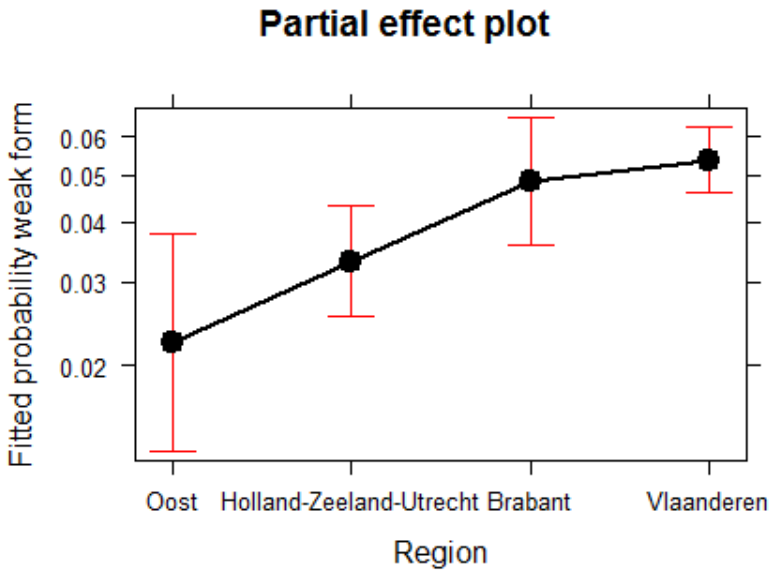


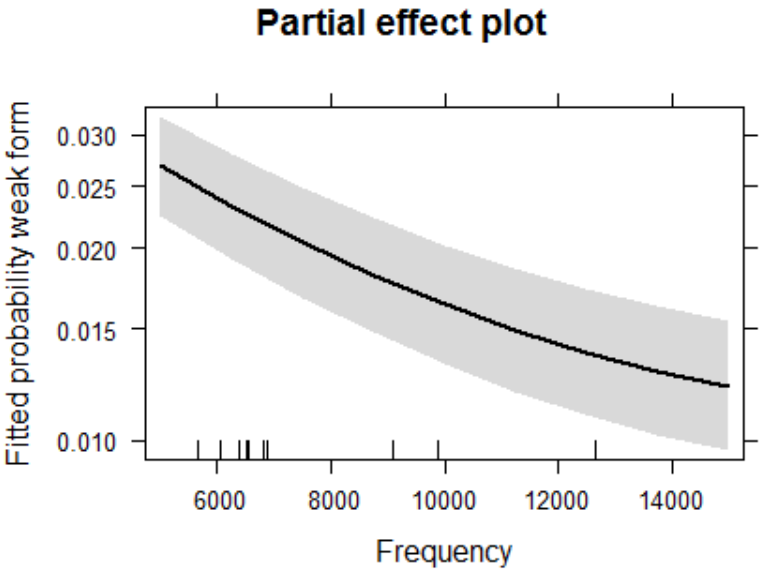
Graphs from: Versloot, Arjen (p.c.)

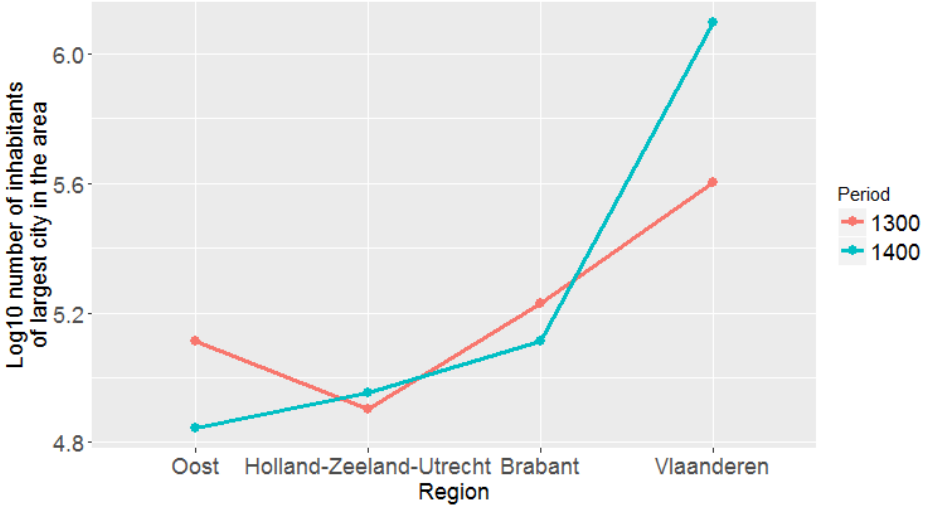
Frisian has recently been heavily influenced by Dutch

From grammars to corpora: 13th and 14th century Dutch

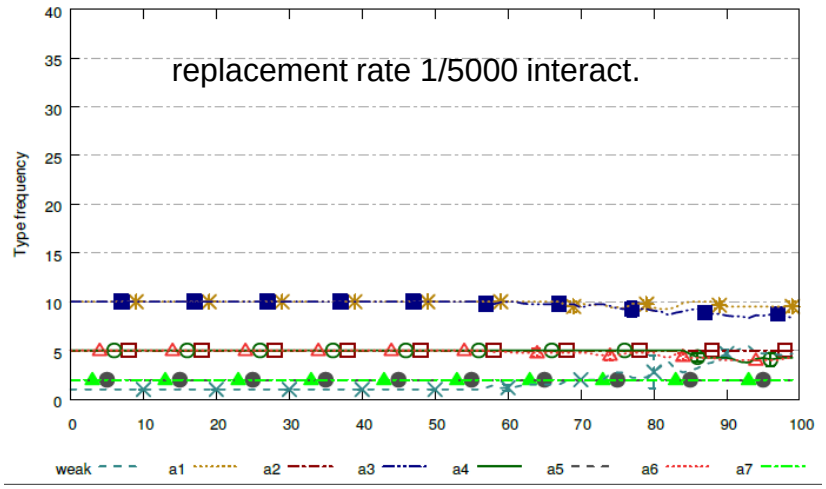


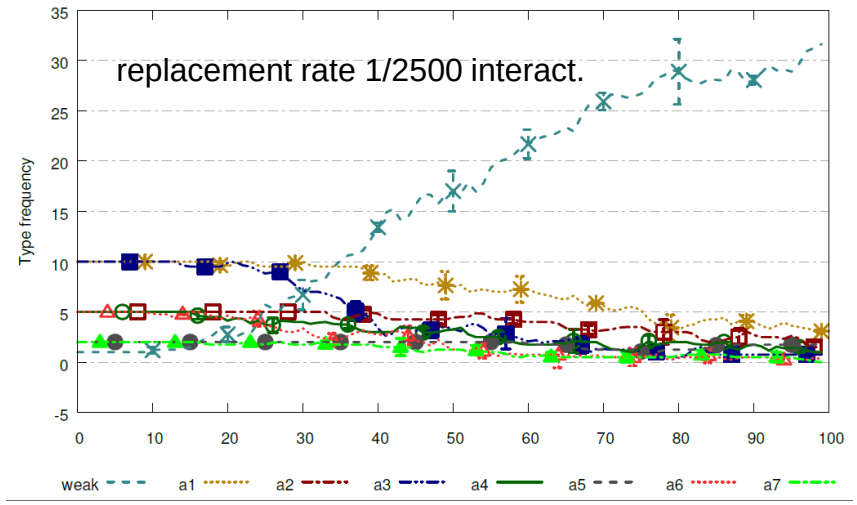


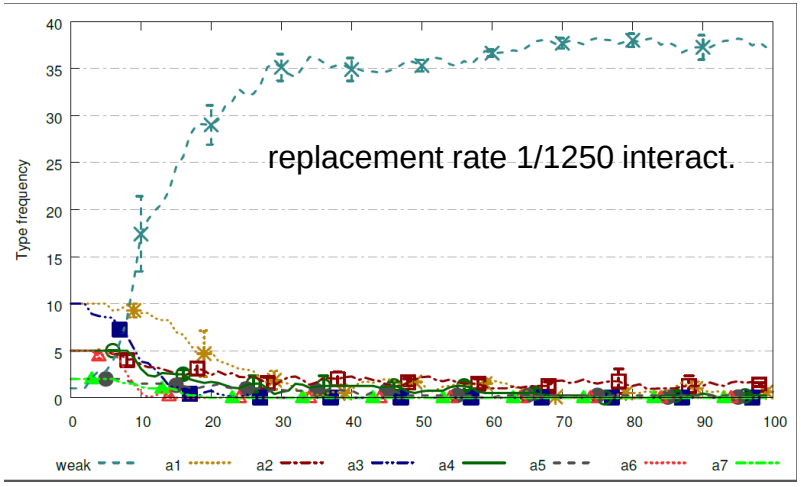


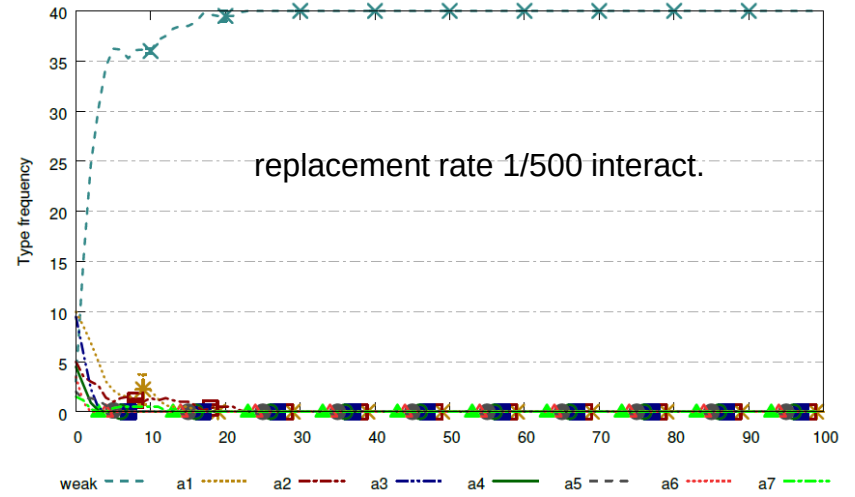


In silico simulation: agent-based models (Pijpops, Beuls & Van de Velde 2015)









Case study: *be going to*

Joint work with Peter Petré

- Analytic future, arose in Early Modern English
- [I am going] [to buy some chocolate] > [I am going to buy some chocolate]
- Source construction: [[*go*][allative motion]] + [[*be* V*ing*][imperfectivity/on-goingness]] + [[*to* Inf][purpose adjunct]]

  *I **am goyng** to the Pope, **to praie** him to place me in mariage.* (1566, *The palace of pleasure beautified* ... [EEBOCorp 1.0])
- The Rolls Royce of grammaticalisation studies. Previous work by Hilpert (2008), Traugott & Trousdale (2013), Disney (2009, ms.), Petré (2013a)

Methodology

- Corpus description
  - EEBOCorp 1.0 (Petré 2013b)
  - EEBO-database (eebo.chadwyck.com): English books printed 1473-1700.
  - Selection criteria:
    1. Prolific writers
    2. Constant register over time
    3. Writers are from roughly the same social status.
  - Resulting corpus: 50 million+ tokens, with individual author token counts ranging between *ca.* 300,000 and 14,000,000 words
  - Perl scripts for retrieving all instances of *going* (n = 10,000+), including variants
  - After semi-manual filtering a total of 1137 instances of *be going to* + INF remained

Methodology

- Analysis
  - Coding of formal + semantic features
  - commonly associated with the grammaticalisation of *be going to*.
  - Each datapoint is given a score per feature
  - summatively, provide an overall score (max. 10) for the level of grammaticalisation reached in a particular observation

1. ‘adjacency’, i.e. the linear contiguity of *go* and the *to*-inf part
   He's going (now) to see some fresher beauties.

2. ‘structural’ features (‘fronting’, ‘parenthetical use’, ‘coordination’ with existing aspectual auxiliaries)
   that barbarous action he was going to commit.

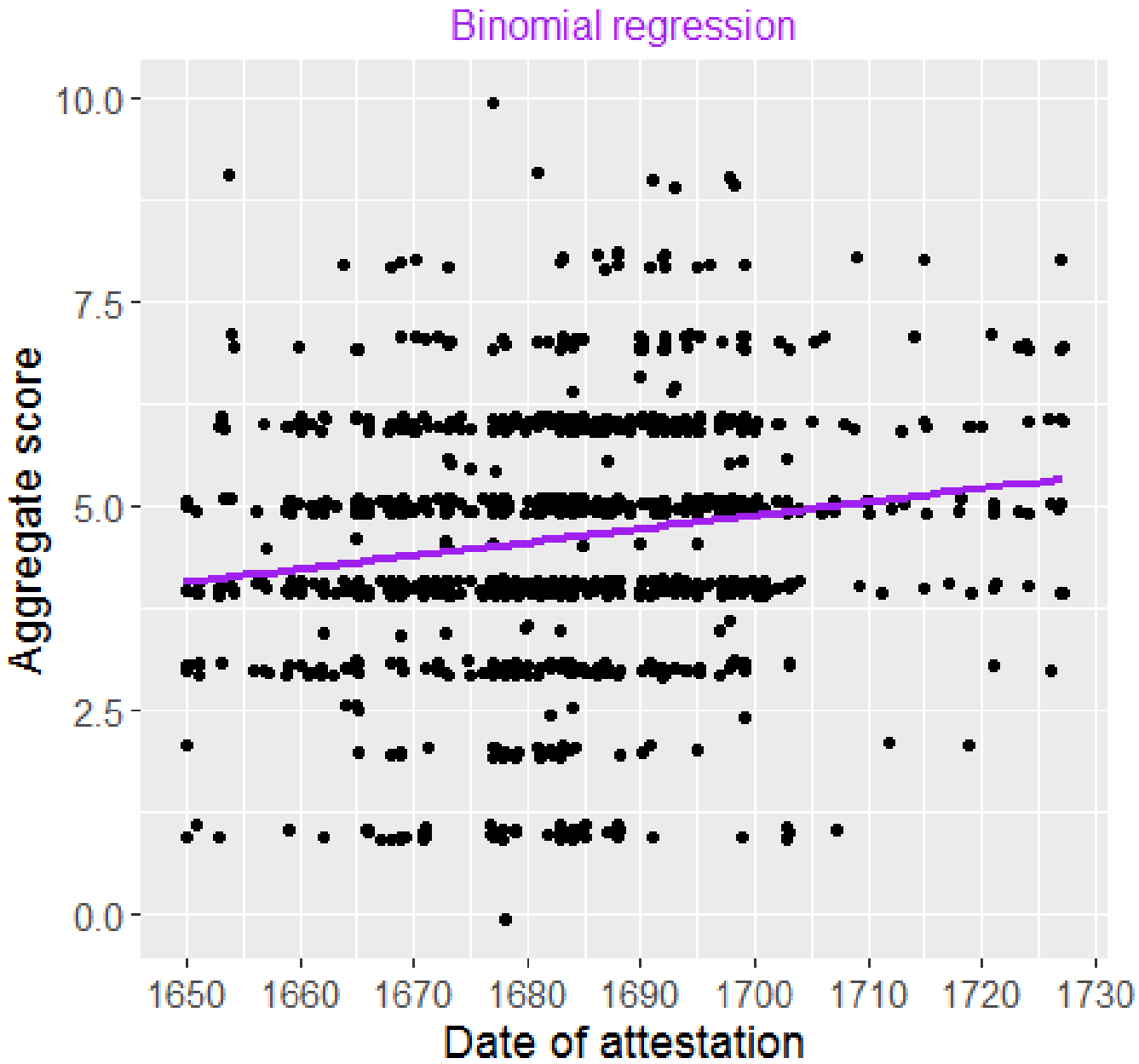
3. ‘goal’, i.e. the presence or absence of a goal
   Sir, I am just now going to a Lawyer (to aske his Councell).

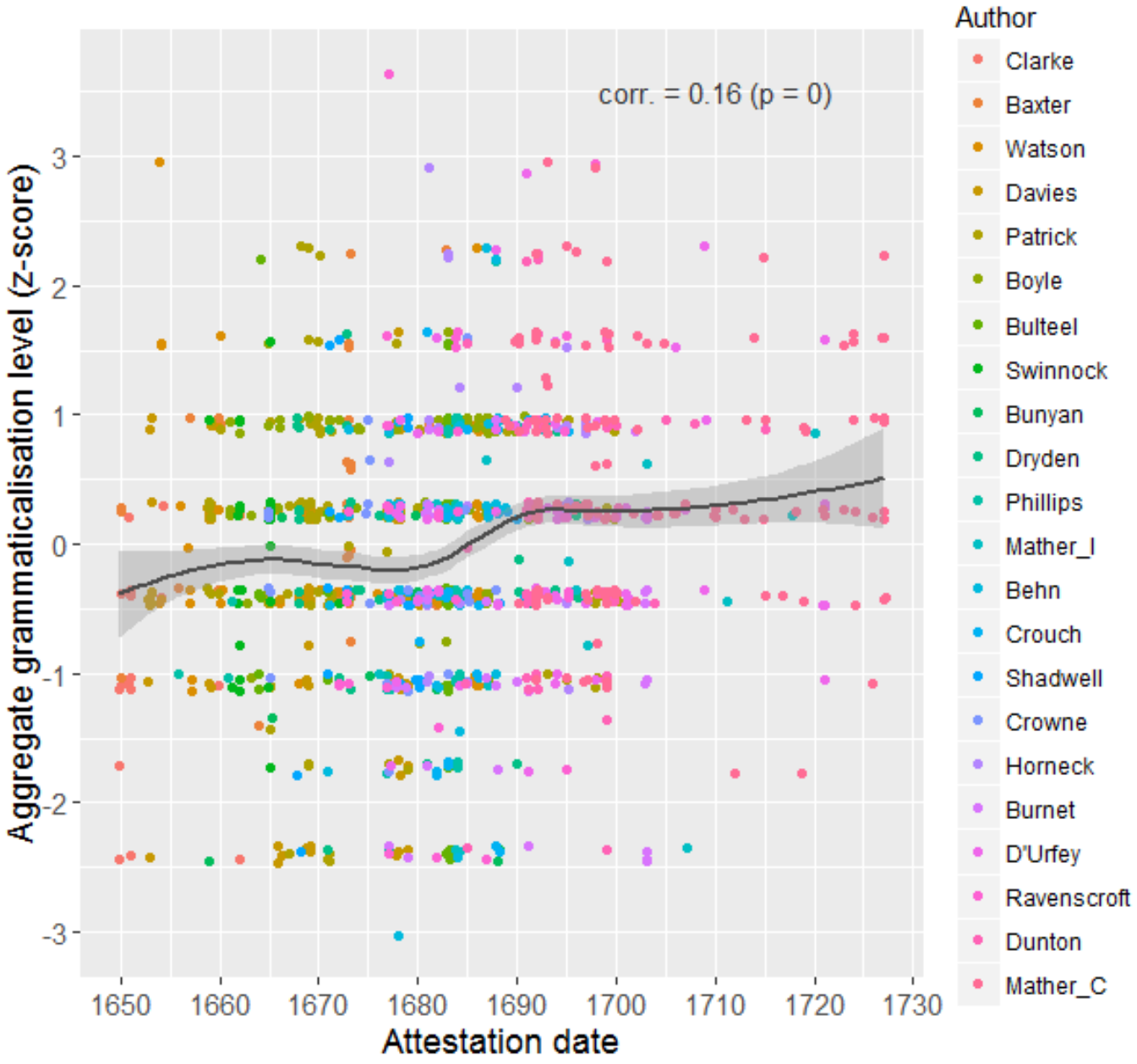
4. ‘voice’, i.e. whether *go* is followed by a passive *to*-inf
   Are not you going to be married?

5. ‘motion’, i.e. can *going* be interpreted as expressing spatial motion?
   Count de Saluces was going to be her lover.

6. ‘animacy’, i.e. whether the subject is animate or not.
   Examples which are now going to be Familiar to me.

7. ‘predictiveness’, i.e. is a prediction about absolute future involved.
   little before the Second Coming ... the Devil is going to be Dislodged of the Air, where his present Quarters are; God will ... cause him to fall.


Binomial regression
10.0
7.5
5.0
2.5
0.0
Aggregate score
1650
1660
1670
1680
1690
1700
1710
1720
1730
Date of attestation


corr. = 0.16 (p = 0)
Aggregate grammaticalisation level (Z-score)
3
2
1
0
-1
-2
-3
1650
1660
1670
1680
1690
1700
1710
1720
1730
Attestation date
Author
Clarke
Baxter
Watson
Davies
Patrick
Boyle
Bulteel
Swinnock
Bunyan
Dryden
Phillips
Mather_I
Behn
Crouch
Shadwell
Crowne
Horneck
Burnet
D'Urfey
Ravenscroft
Dunton
Mather_C

London: unusually high population turnover

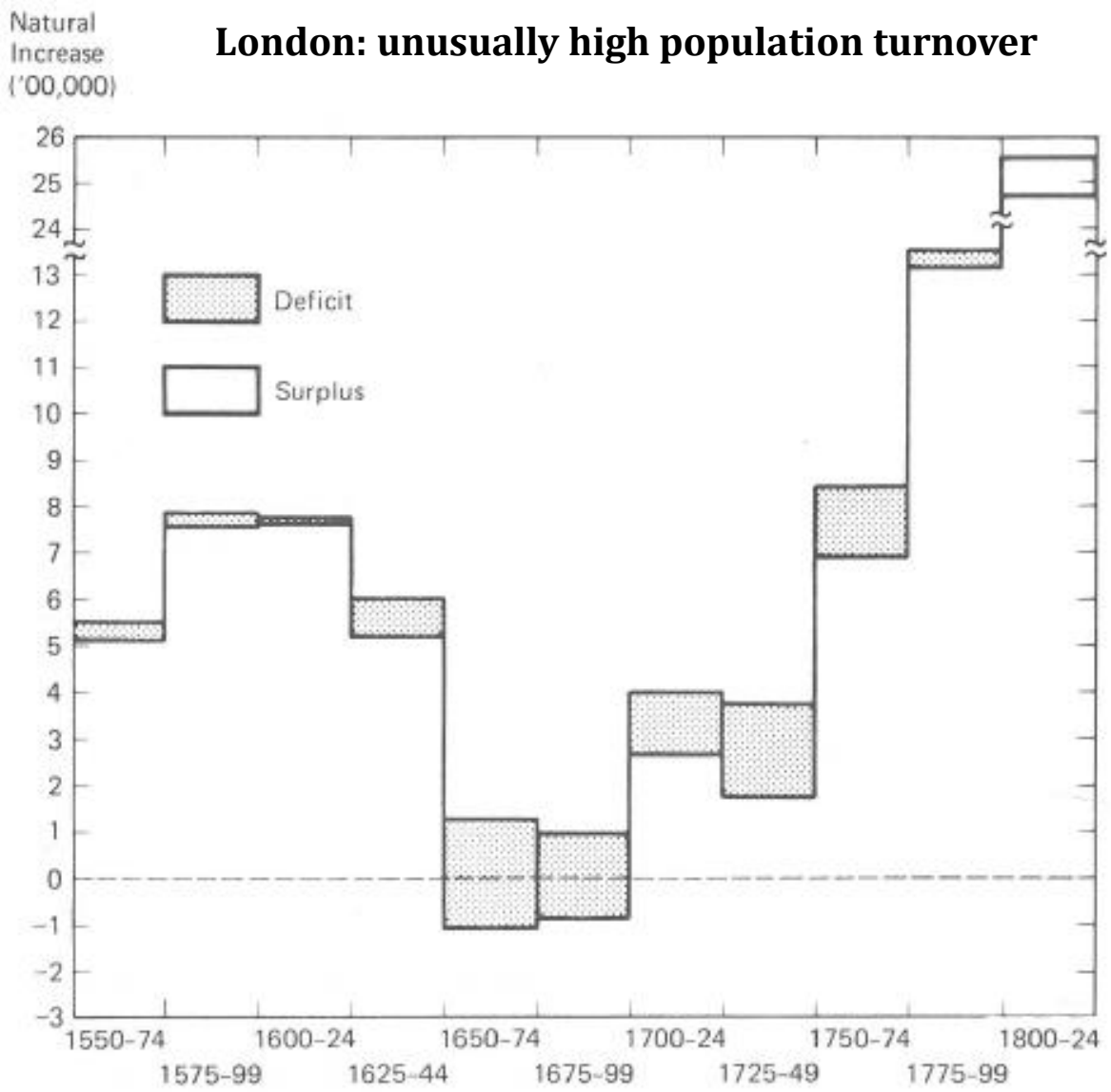


Figure 6.2: Natural increase in London and England by quarter century, 1550–1824

Wrigley & Schofield 1994: 166-169

"(...) London's particular demographic characteristics of massive immigration and high mortality."

"In the second half of the seventeenth century the extra-metropolitan surpluses fell precipitously while the London deficits rose so that the latter were almost exactly double the former, thereby converting the non-London surplus into a national deficit of almost the same magnitude. In the early eighteenth century the rest of the country bounced back to record more substantial surpluses, but the London deficits continued at a high level and reduced the surpluses in the first two quarters of the century by 33 and 55 per cent respectively."

London: unusually high population turnover

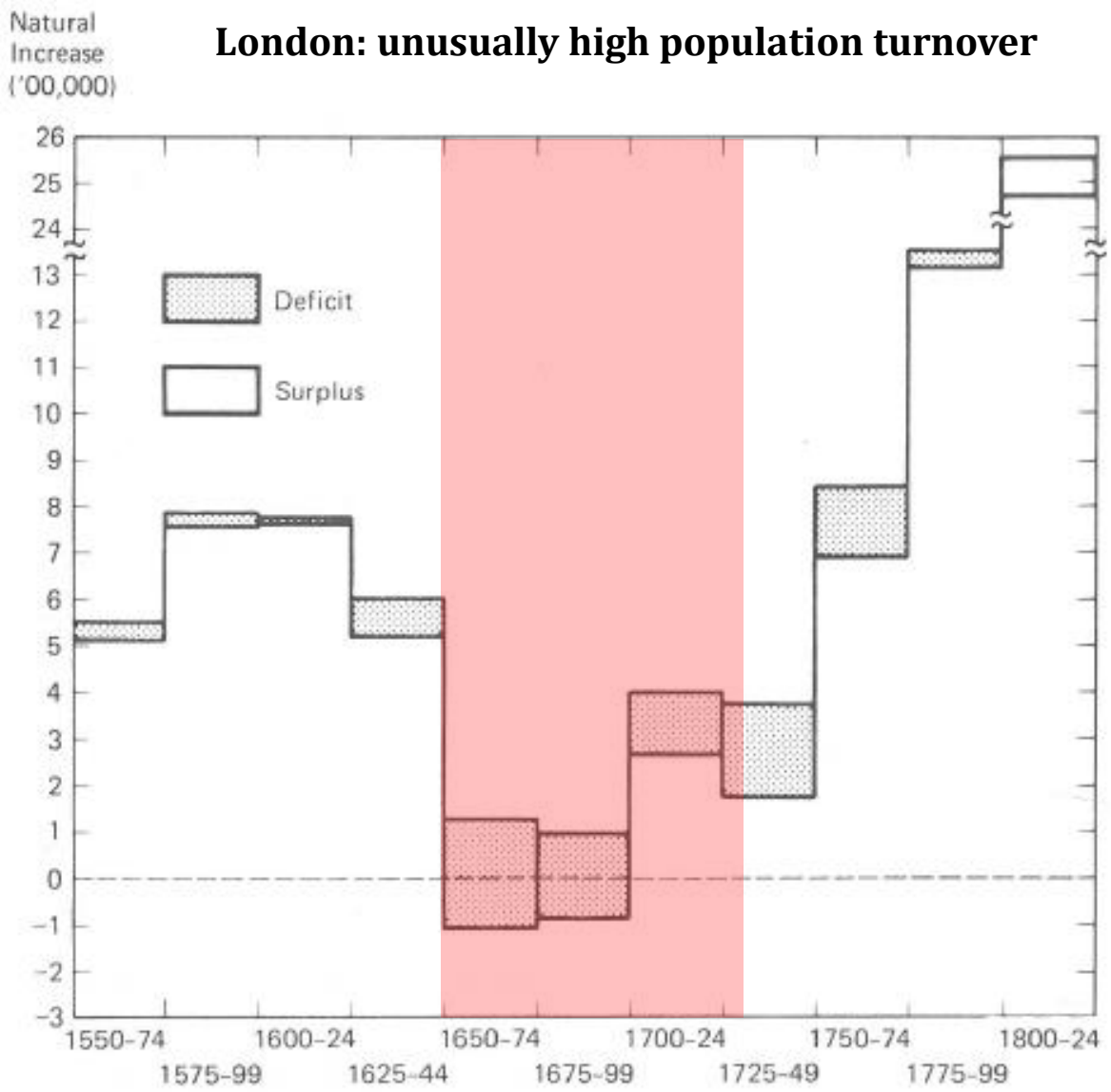
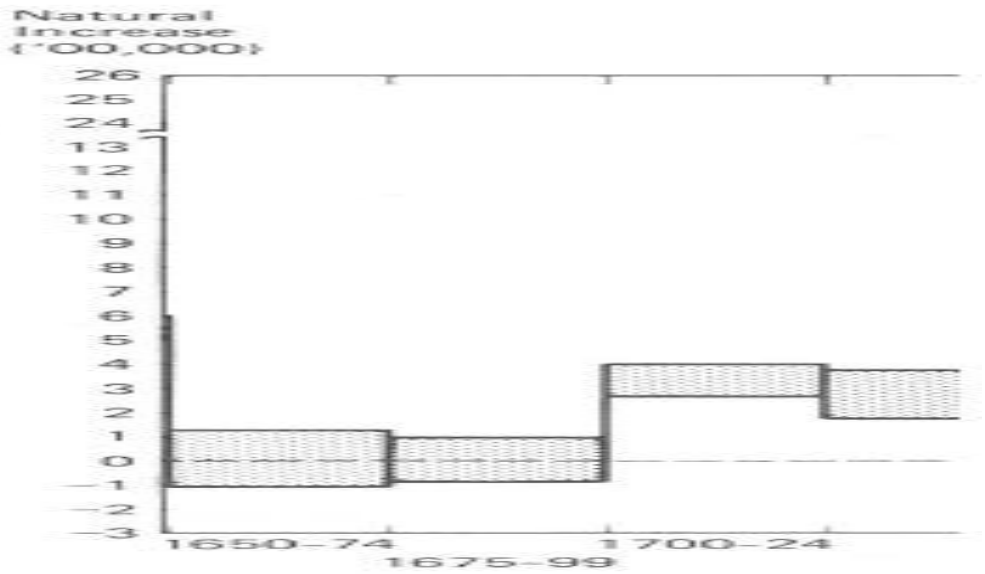


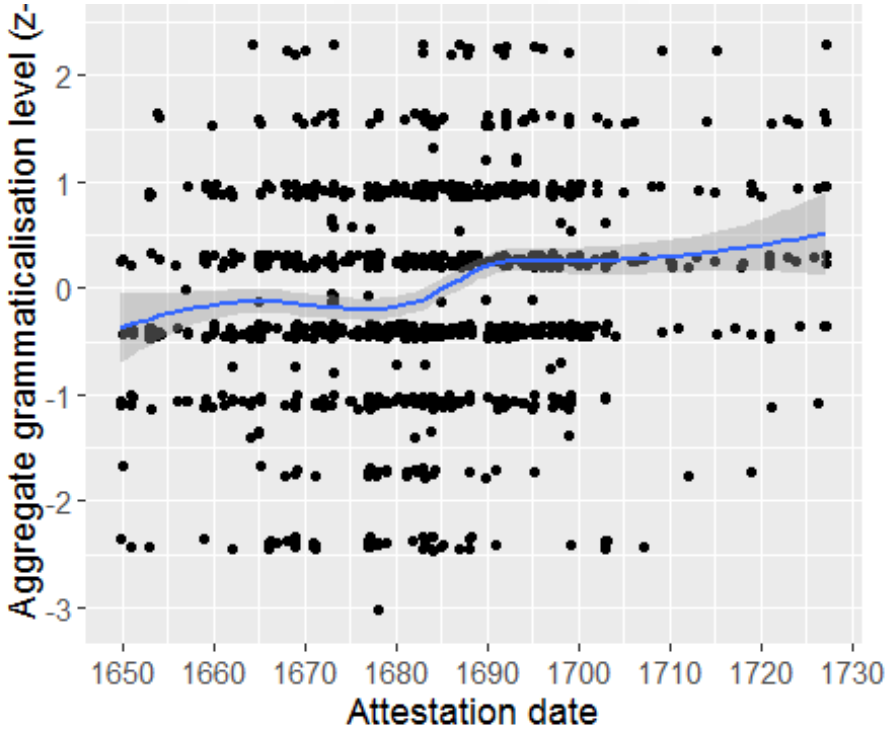
Figure 6.2: Natural increase in London and England by quarter century, 1550–1824

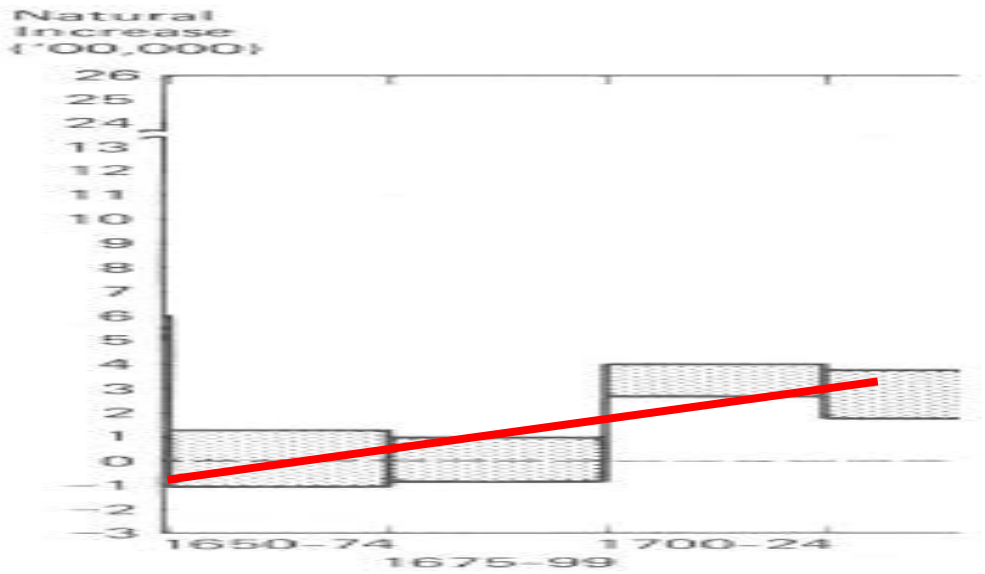
Wrigley & Schofield 1994: 166-169

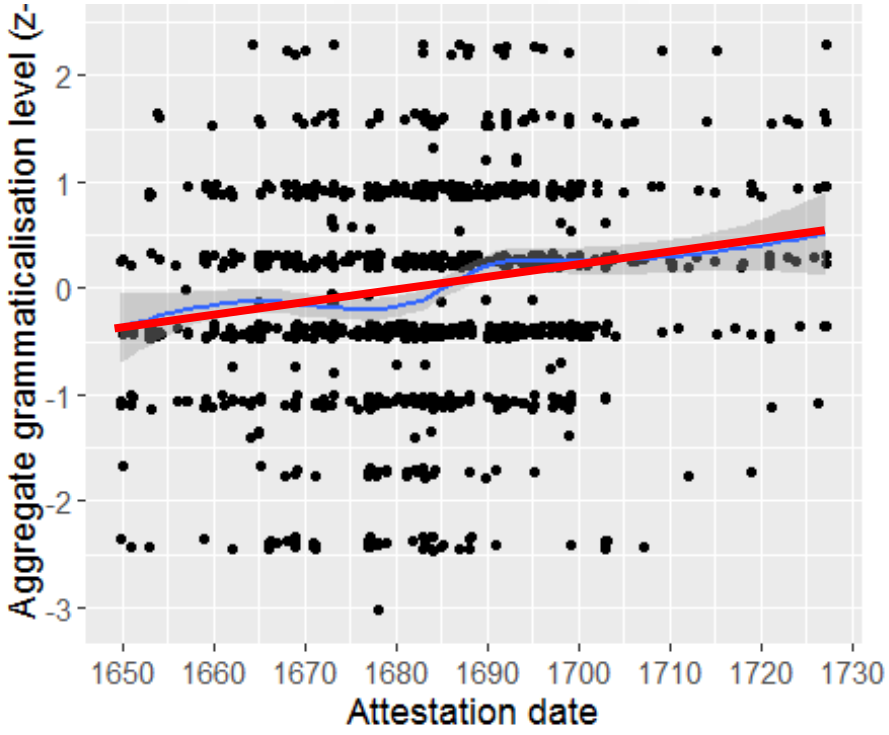
"(...) London's particular demographic characteristics of massive immigration and high mortality."

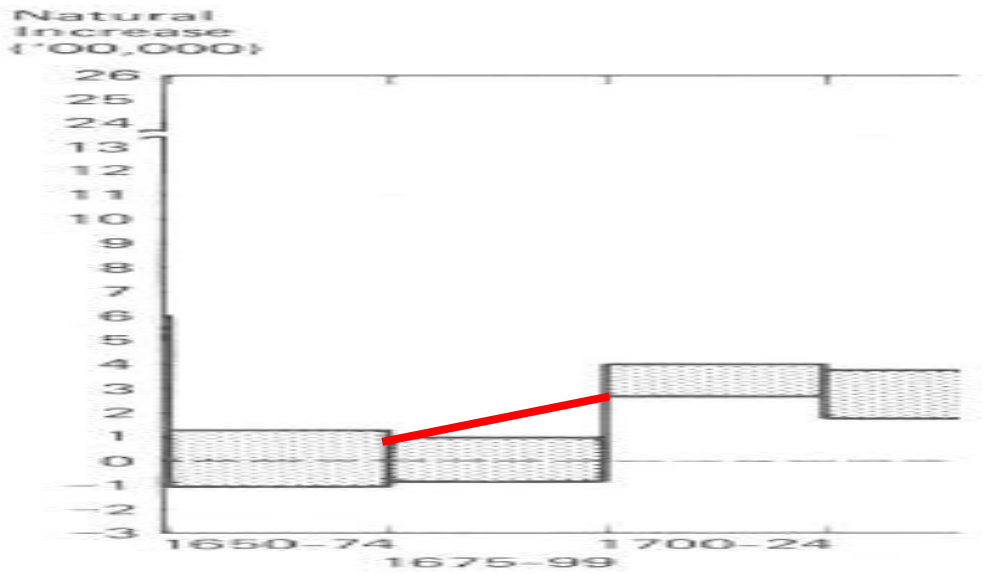
"In the second half of the seventeenth century the extra-metropolitan surpluses fell precipitously while the London deficits rose so that the latter were almost exactly double the former, thereby converting the non-London surplus into a national deficit of almost the same magnitude. In the early eighteenth century the rest of the country bounced back to record more substantial surpluses, but the London deficits continued at a high level and reduced the surpluses in the first two quarters of the century by 33 and 55 per cent respectively."

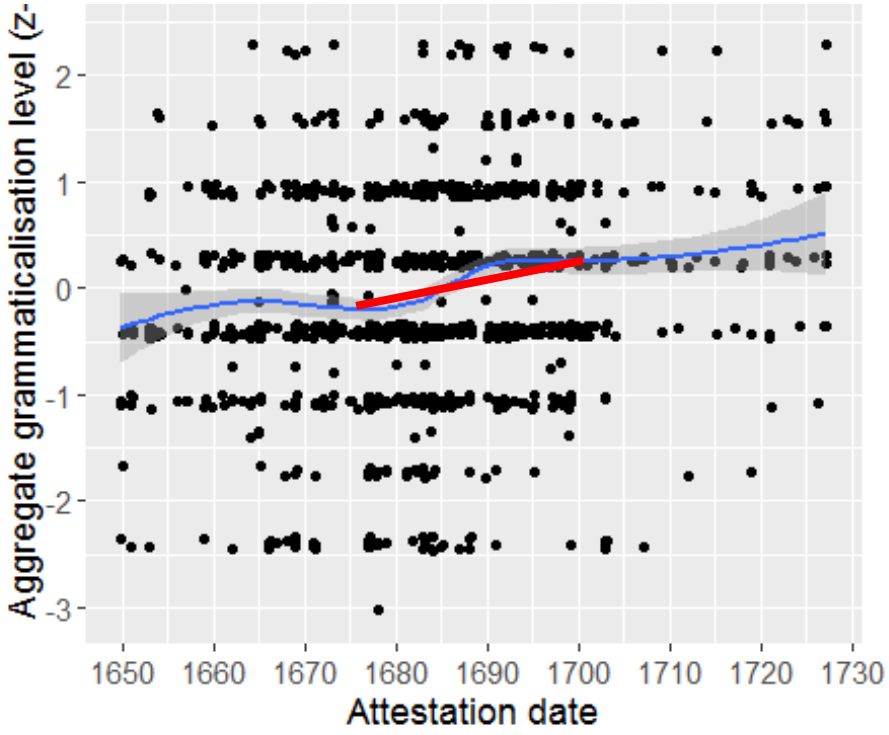

Natural
Increase
('00,000)
26
25
24
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
1650–74
1675–99
1700–24


Aggregate grammaticalisation level (z-
2
1
0
-1
-2
-3
1650
1660
1670
1680
1690
1700
1710
1720
1730
Attestation date

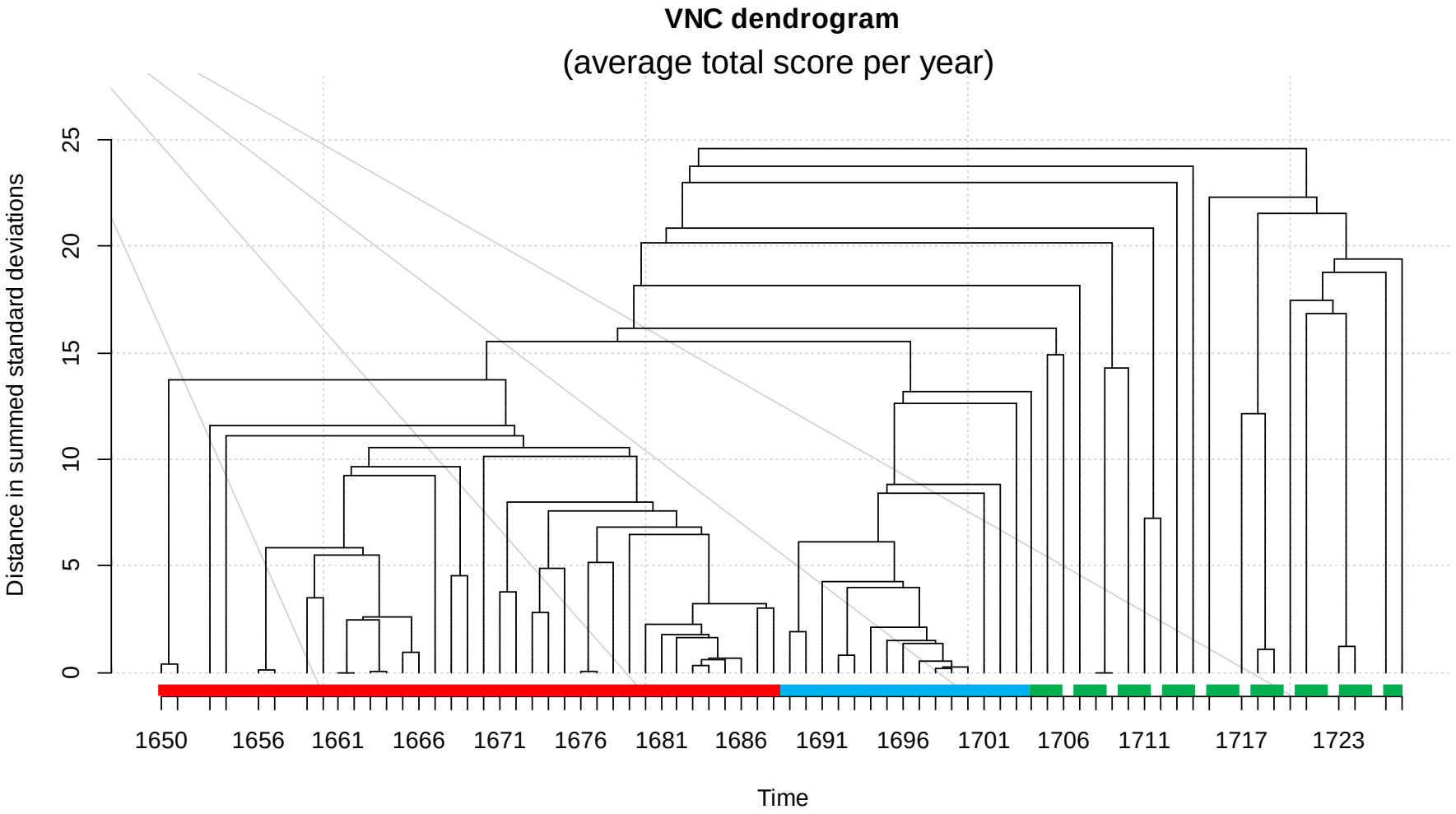

Natural
Increase
('00,000)
26
25
24
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
1650–74
1675–99
1700–24


Aggregate grammaticalisation level (z-
2
1
0
-1
-2
-3
1650
1660
1670
1680
1690
1700
1710
1720
1730
Attestation date


Natural
Increase
('00,000)
26
25
24
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
1650–74
1675–99
1700–24


Aggregate grammaticalisation level (z-
2
1
0
-1
-2
-3
1650
1660
1670
1680
1690
1700
1710
1720
1730
Attestation date

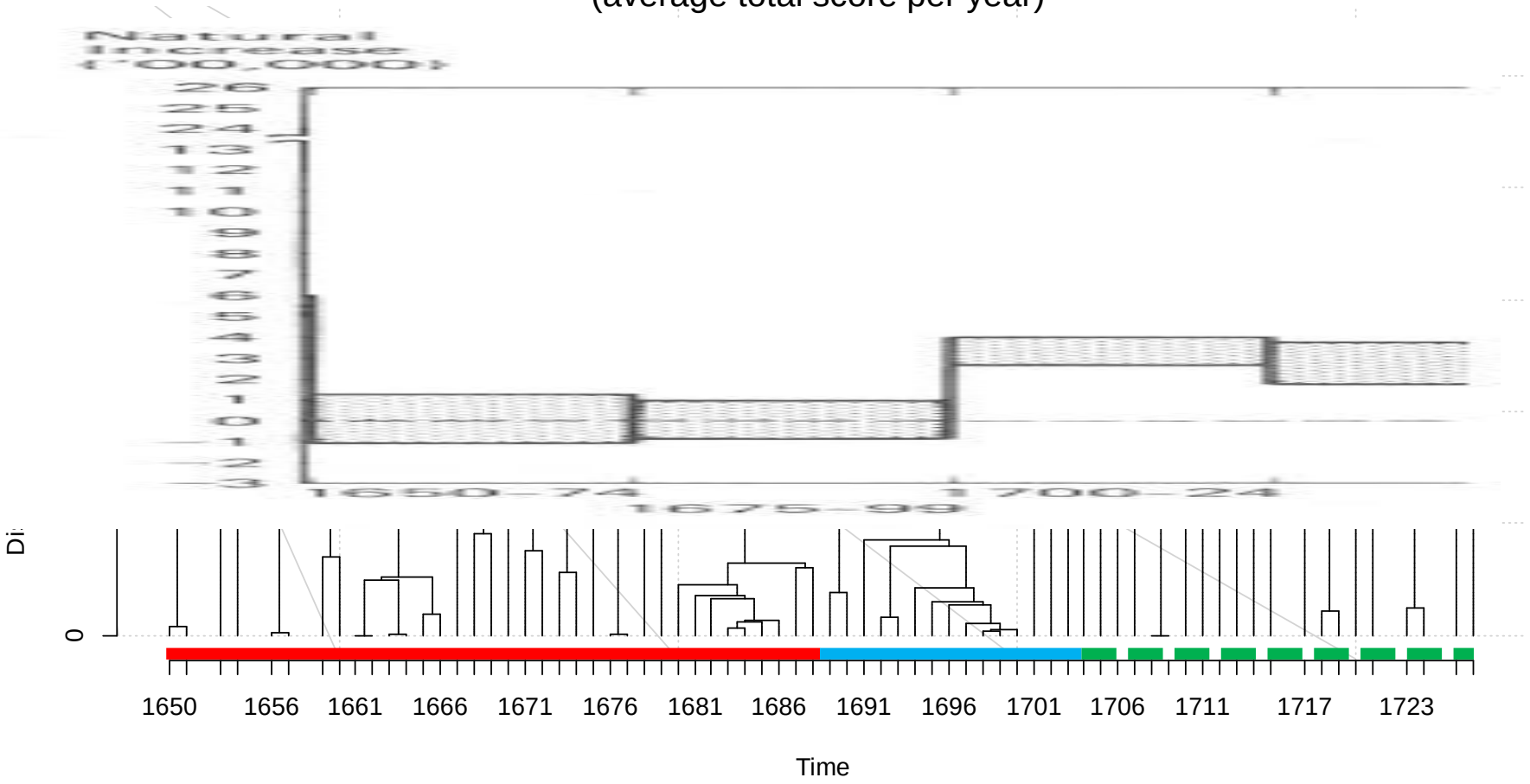
Variability-based neighbor clustering (Gries & Hilpert 2008)



Variability-based neighbor clustering (Gries & Hilpert 2008)

VNC dendrogram

(average total score per year)



Conclusions

- Language change is a function of historical demography
- between languages of different families (Kusters 2003)
- within languages of the same family (English-Dutch-German)
- within stages of one language (English-Dutch)

Acknowledgements

Katrien Beuls

Isabeau De Smet

Peter Petré

Dirk Pijpops

References

- Aitchison, J. 1991. *Language change: progress or decay*. 2nd edn. Cambridge: Cambridge University Press.
- Bentz, C. & B. Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3: 1-27.
- Bickel, B. & J. Nichols. 2013. 'Fusion of Selected Inflectional Formatives'. In: M. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/20, Accessed on 2015-06-11.)
- Carroll, R., R. Svare & J. Salmons. 2012. 'Quantifying the evolutionary dynamics of German verbs'. *Journal of Historical Linguistics* 2: 153-172.
- Christiansen, M.H. & N. Chater. 2008. 'Language as shaped by the brain'. *Behavioral and Brain Sciences* 31(5): 489-509.
- Disney, S. 2009. 'A domain matrix view of the uses and development of BE going to + infinitive'. *Linguistics & Language Teachings* 3: 25-44.
- Greenberg, J. 1960. 'A Quantitative Approach to the Morphological Typology of Language'. *International Journal of American Linguistics* 26(3): 178-194.
- Gries, S. & M. Hilpert. 2008. 'The identification of stages in diachronic data: variability-based neighbor clustering'. *Corpora* 3(1): 59-81.
- Hilpert, M. 2008. *Germanic future constructions: a usage-based approach to language change*. Amsterdam: Benjamins.
- Kusters, W. 2003. *Linguistic complexity: the influence of social change on verbal inflection*. Utrecht: LOT Dissertation Series.

References

- Lieberman, E., Michel, J.-B., J. Jackson, T. Tang, M.A. Nowak. 2007. 'Quantifying the evolutionary dynamics of language'. *Nature* 449: 713-716.
- Lupyan, G. & R. Dale. 2010. 'Language structure is partly determined by social structure'. *PLoS ONE* 5(1).
- Nettle, D. 1999. 'Is the rate of linguistic change constant?' *Lingua* 108: 119-136.
- Petré, P. 2013a. 'Analogy, frequency and system-dependent competition in grammaticalization: the show is (going) (about) to begin'. Forty-sixth annual meeting of the SLE (SLE 46), Split, 18-21 September.
- Petré, P. 2013b. *EEBOCorp 1.0*.
- Pijpops, D, K. Beuls & F. Van de Velde. 2015. 'The rise of the verbal weak inflection in Germanic. An agent-based model'. *Computational Linguistics in the Netherlands Journal* 5: 81-102.
- Szmrecsanyi, B. 2012. Analyticity and Syntheticity in the history of English. In: T. Nevalainen, E. Traugott (eds.), *The Oxford handbook of the history of English*. Oxford: Oxford University Press. 654-665.
- Traugott. E. & G. Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Wichmann, S. & E.W. Holman. 2009. 'Population Size and Rates of Language Change'. *Human Biology* 81: 259-274.
- Woods, N. 2001. 'Internal and external dimensions of language change: the great divide? Evidence from New Zealand English'. *Linguistics* 39(5): 973-1007.
- Wrigley, E.A. & R.S. Schofield. 1994. The population history of England 1541-1871. A reconstruction. Cambridge: Cambridge University Press.