

Fluid Construction Grammar Meeting, 8-13 August 2014

Longitudinal data on individual grammaticalisation

Freek Van de Velde

Peter Petré

University of Leuven

Research Foundation Flanders – FWO

Grammaticalisation

- Grammaticalisation: the recruitment of lexical material to fabricate new grammar
 - auxiliaries from main verbs:
(PIE **uert-* 'turn around' > OE *weorðan*,
Germ. *werden*)
 - condensation of small clause fragments
(*het mag (ge)schieden* > *misschien*)
(*it may be* > *maybe*)
 - conjunctions from word groups:
(*þa hwile þe* > *while*)



Supporting stakes are made of beams (i.e. of trees)

Grammaticalization: regular or erratic?

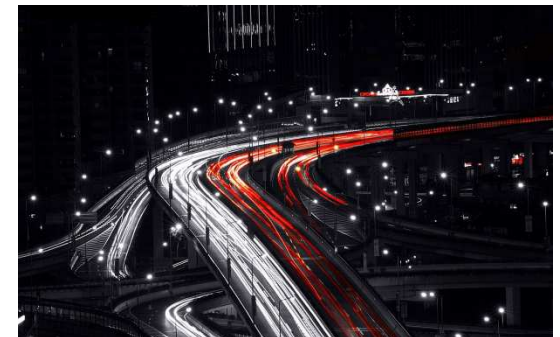
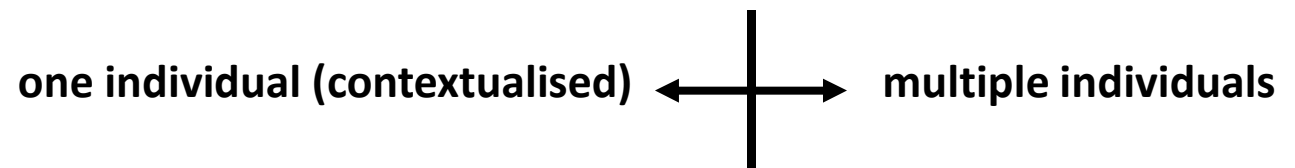
- Regularity:
 - Across languages: Heine & Kuteva 2002
 - Within languages: Van Gelderen 2011
- Variability:
 - Co-determined by structural environment (Mithun 1991:160; Fischer 1997:265; 2007, ch.4; Hilpert 2008; Verstraete 2008; De Vogelaer 2010; Coleman & De Clerck 2011; Ghesquière & Van de Velde 2011; De Smet & Van de Velde 2013; Petré 2014, among many others).
 - Individual differences?

Individual vs. aggregate level

"Unfortunately, of course, how to define language change remains a controversial issue. The basic division is between two schools of thought. On the one hand there are those who, following Chomsky and earlier writers such as Paul (1880), view the competence of the individual speaker (I language) as the primary object of study in linguistics. On the other hand there are those who stress the social dimension of language and focus their research on the "orderly heterogeneity" (Weinreich, Labov and Herzog 1968: 100) of usage by different social groups within a speech community. The latter school naturally views language change as minimally a change common to all members of a particular subgroup of a speech community – anything less is merely individual variation, not change." (Lucas 2014)

Individual vs. aggregate level

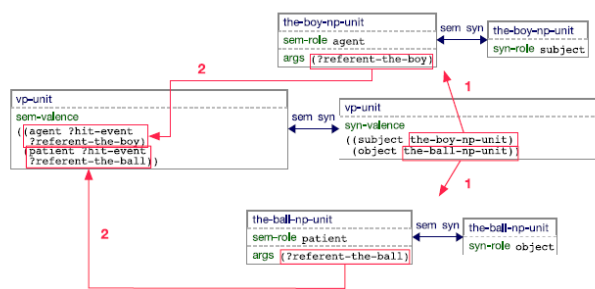
"Unfortunately, of course, how to define language change remains a controversial issue. The basic division is between two schools of thought. On the one hand there are those who, following Chomsky and earlier writers such as Paul (1880), view the competence of the individual speaker (I language) as the primary object of study in linguistics. On the other hand there are those who stress the social dimension of language and focus their research on the "orderly heterogeneity" (Weinreich, Labov and Herzog 1968: 100) of usage by different social groups within a speech community. The latter school naturally views language change as minimally a change common to all members of a particular subgroup of a speech community – anything less is merely individual variation, not change." (Lucas 2014)



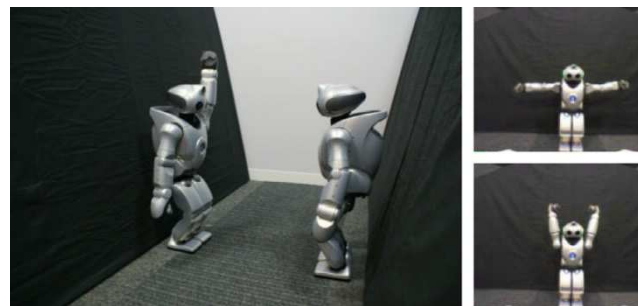
Individual vs. aggregate level

"Unfortunately, of course, how to define language change remains a controversial issue. The basic division is between two schools of thought. On the one hand there are those who, following Chomsky and earlier writers such as Paul (1880), view the competence of the individual speaker (I language) as the primary object of study in linguistics. On the other hand there are those who stress the social dimension of language and focus their research on the "orderly heterogeneity" (Weinreich, Labov and Herzog 1968: 100) of usage by different social groups within a speech community. The latter school naturally views language change as minimally a change common to all members of a particular subgroup of a speech community – anything less is merely individual variation, not change." (Lucas 2014)

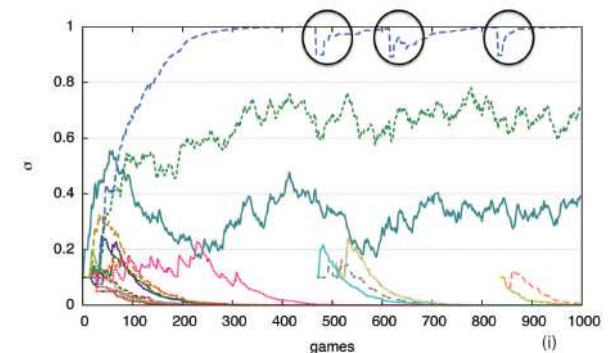
one individual (contextualised) ← → multiple individuals



(From: Van Trijp 2014)



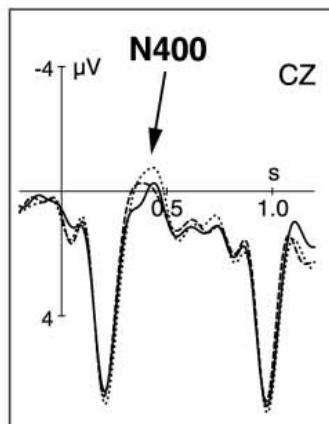
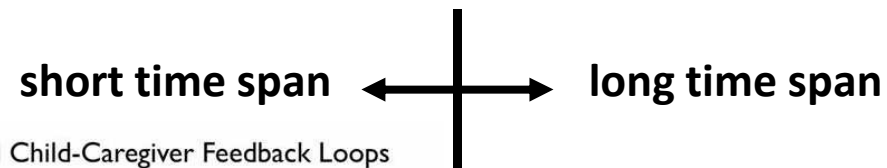
(From: Steels 2011)



(From: Beuls & Steels 2013)

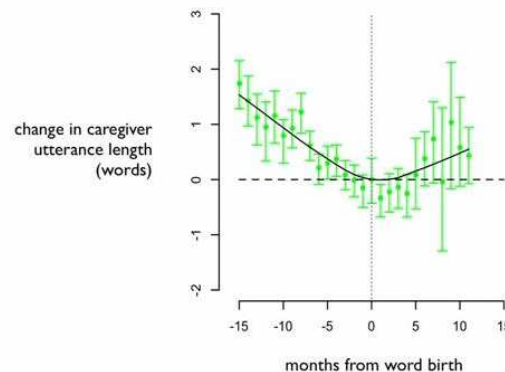
Individual vs. aggregate level

"Unfortunately, of course, how to define language change remains a controversial issue. The basic division is between two schools of thought. On the one hand there are those who, following Chomsky and earlier writers such as Paul (1880), view the competence of the individual speaker (I language) as the primary object of study in linguistics. On the other hand there are those who stress the social dimension of language and focus their research on the "orderly heterogeneity" (Weinreich, Labov and Herzog 1968: 100) of usage by different social groups within a speech community. The latter school naturally views language change as minimally a change common to all members of a particular subgroup of a speech community – anything less is merely individual variation, not change." (Lucas 2014)

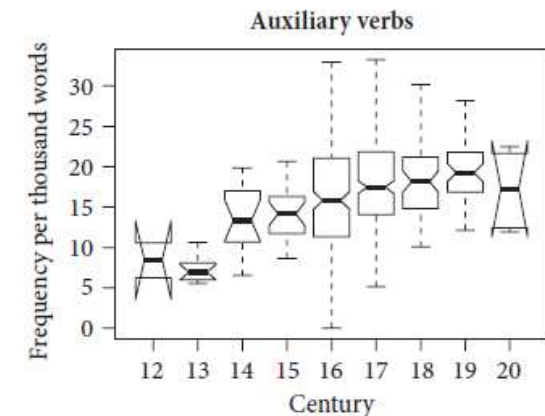


(From: Wang et al. 2009)

Evidence for Finely Tuned Child-Caregiver Feedback Loops



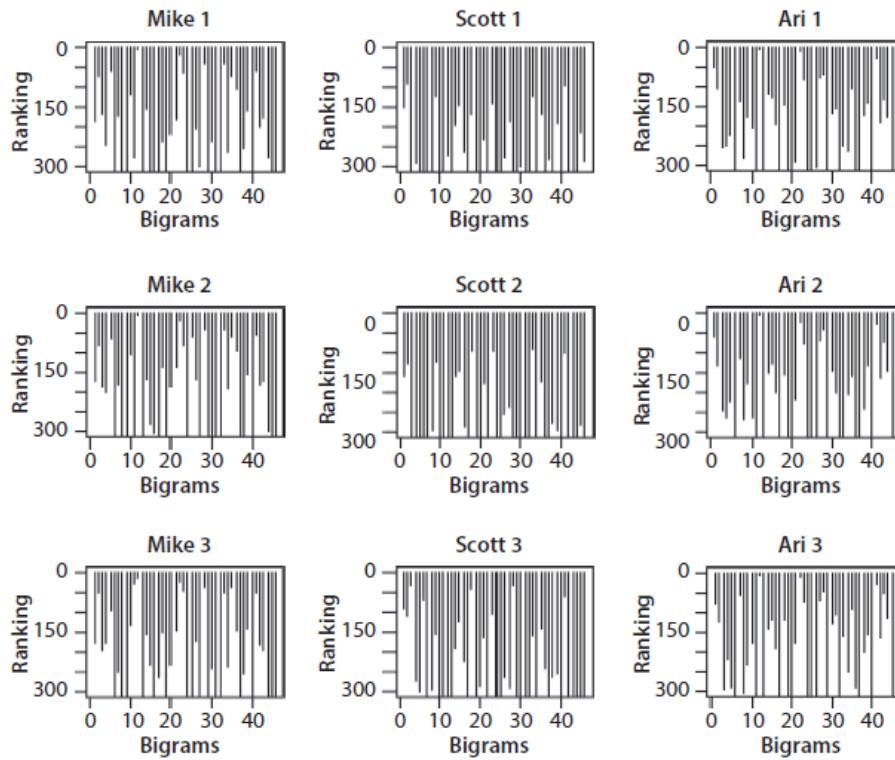
(From: Roy 2011)



(From: Szmrecszanyi 2012)

Individual vs. aggregate level

synchronic / short time span:



(From: Barlow 2013)

diachronic / non-longitudinal:

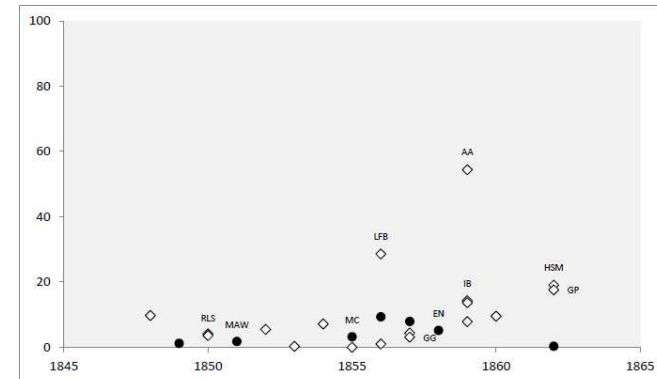
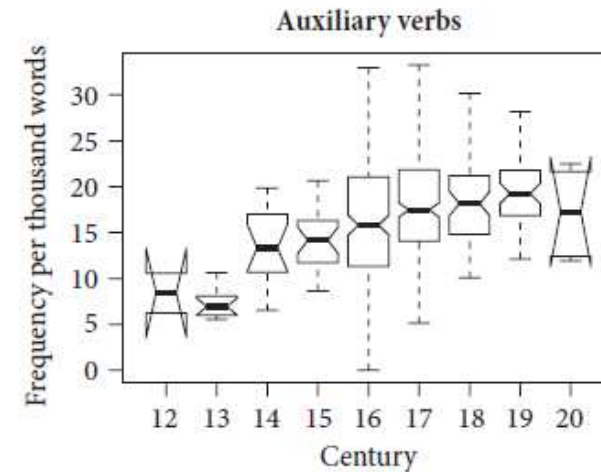


Figure 6. Percentage of *-ing*-clauses relative to *to*-infinitives following *begin* for 25 male (white diamonds) and female (black circles) authors in CEN, arranged by author's year of birth.

(From: De Smet, ms.)

no individual differences:



(From: Szmrecszanyi 2012)

diachronic, semi-longitudinal:

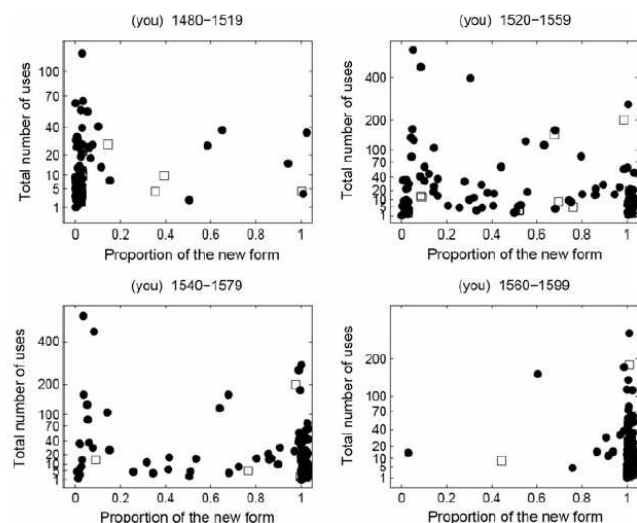


FIGURE 7. Diffusion of *you* in the subject function, 1480–1600 (note the overlapping period, 1540–1579).

(From: Nevalainen et al. 2011)

diachronic, semi-longitudinal, small scale:

‘old’ letters + ‘new’ letters

	Holland vocalized <i>ou/ au</i>	Brabant vocalized <i>ou/ au</i>	East NLS/N. Ger. non-vocalized <i>old</i>	Other auld
Magdaleine, 1592-1604		✓	✓	
Magdaleine, 1607-1610		✓		
Philip, 1597-1609		1597-1606, <i>ou & au</i> forms		1607-1616
Philip, 1642-1652		✓ <i>ou</i> forms only		

(From: Hendriks 2013)

- Research on individual differences in
 - language attainment experiments (see Dąbrowska 2012)
 - Synchronic corpus studies (Barlow 2013)
 - Historical corpus studies (Nevalainen et al. 2011; De Smet, ms.)
 - Diachronic (longitudinal) corpus studies (Bergs 2005; Raumolin-Brunberg 2009; Hendriks 2013)
- What is lacking?
 - Diachronic (longitudinal) LARGE-SCALE corpus studies

Be going to · grammaticalization

- [I am going] [to buy some chocolate] > [I am going to buy some chocolate]
- Source construction: [[*go*][allative motion]] + [[*be Ving*][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])

- Previous work by Hilpert (2008), Traugott (2011, 2012), 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. Sufficient material for first and second halves of writer's careers
 2. Constant register over time
 3. Writers are from roughly the same social status.
 - Resulting corpus: 31 million tokens, with individual author token counts ranging between *ca.* 300,000 and 10,000,000 words
 - Perl scripts for retrieving all instances of *going* (n = 5821), including variants
 - After semi-manual filtering a total of 1591 instances of *be going* remained

Methodology

- Data coding and analysis
 - Several formal and semantic features that are commonly associated with the grammaticalization of *be going to*. Summatively, we use these diagnostics to assess the level of grammaticalization reached in a particular individuals, which serves as the dependent variable in our inquiry.

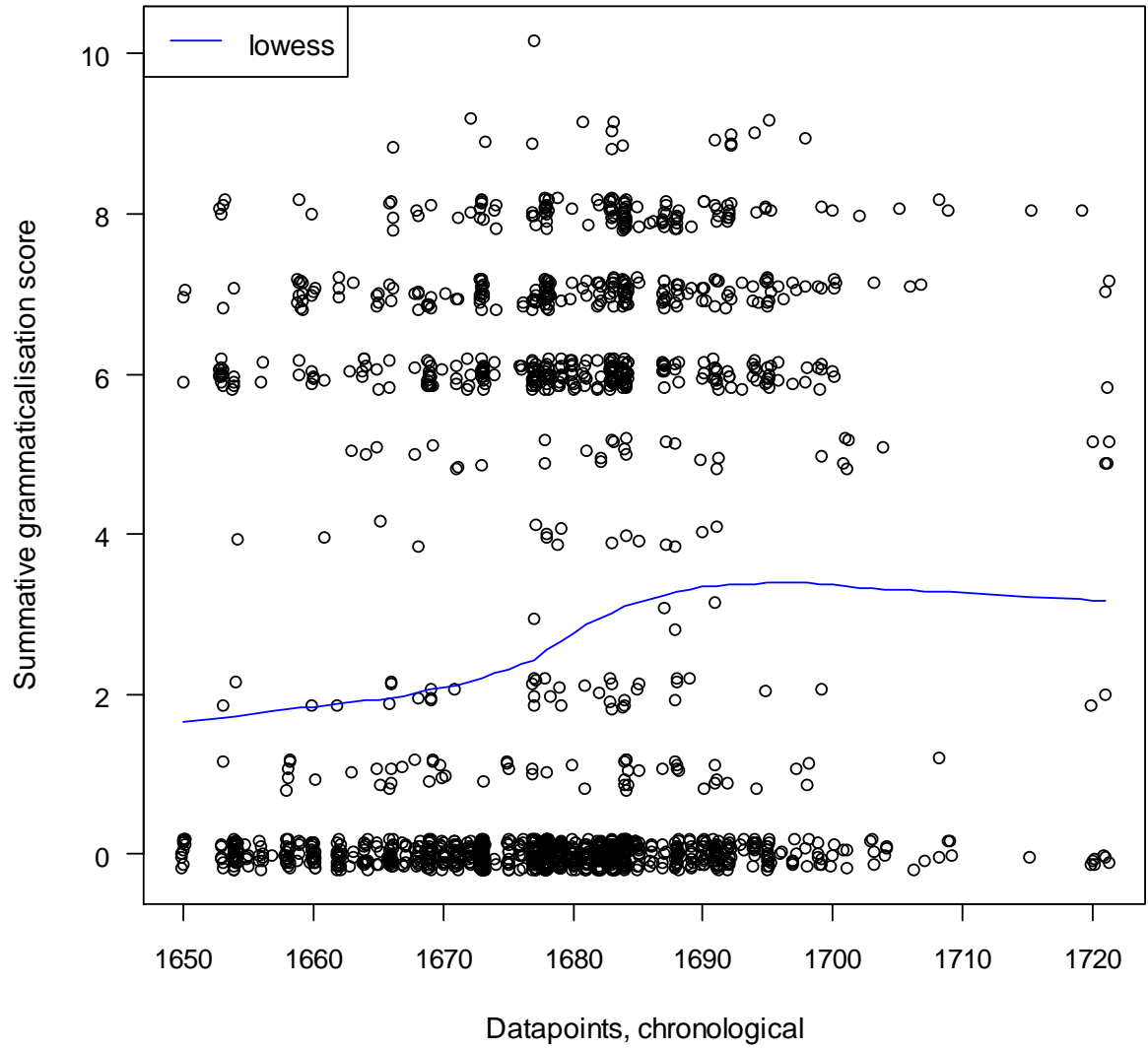
Methodology

- Data coding and analysis
 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.

Methodology

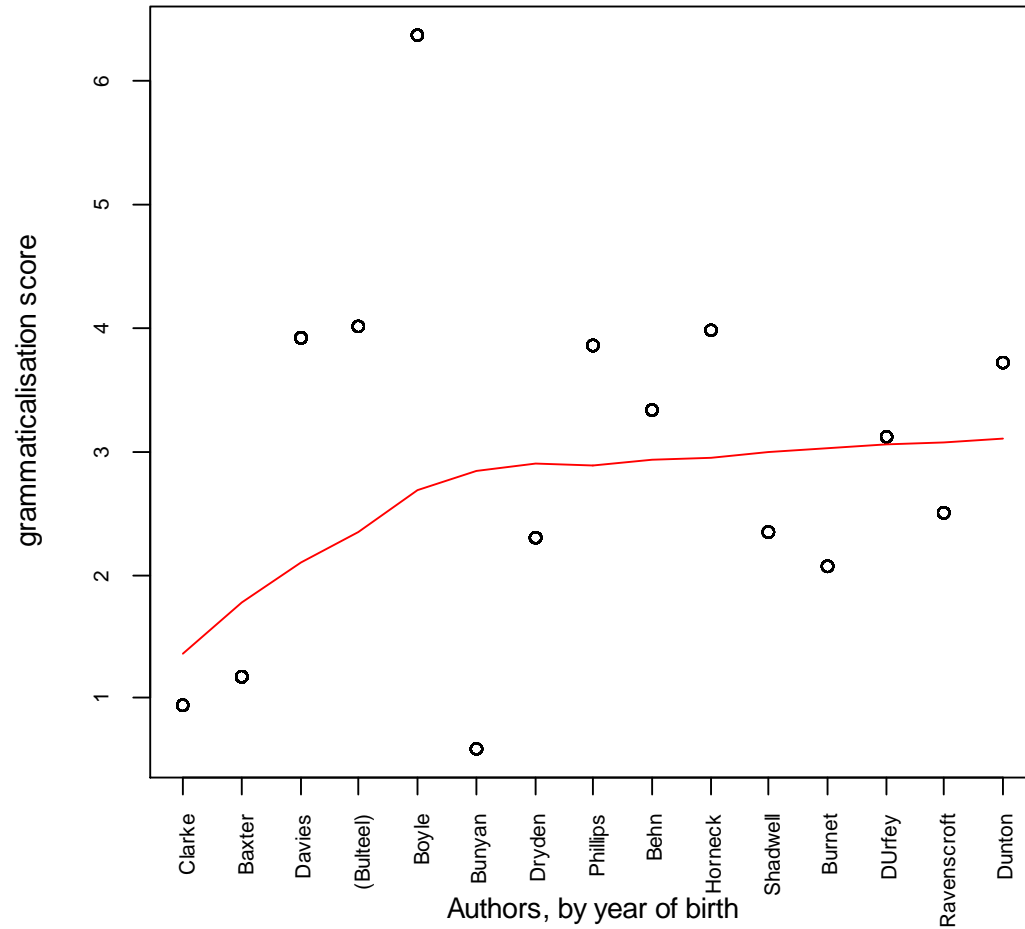
- Data coding and analysis
 - For each of the authors, we divided the collected data in half, to arrive at two categories 'earlier work' and 'later work', in order to check whether differences occurred through the years.

Aggregate grammaticalisation score (jittered)



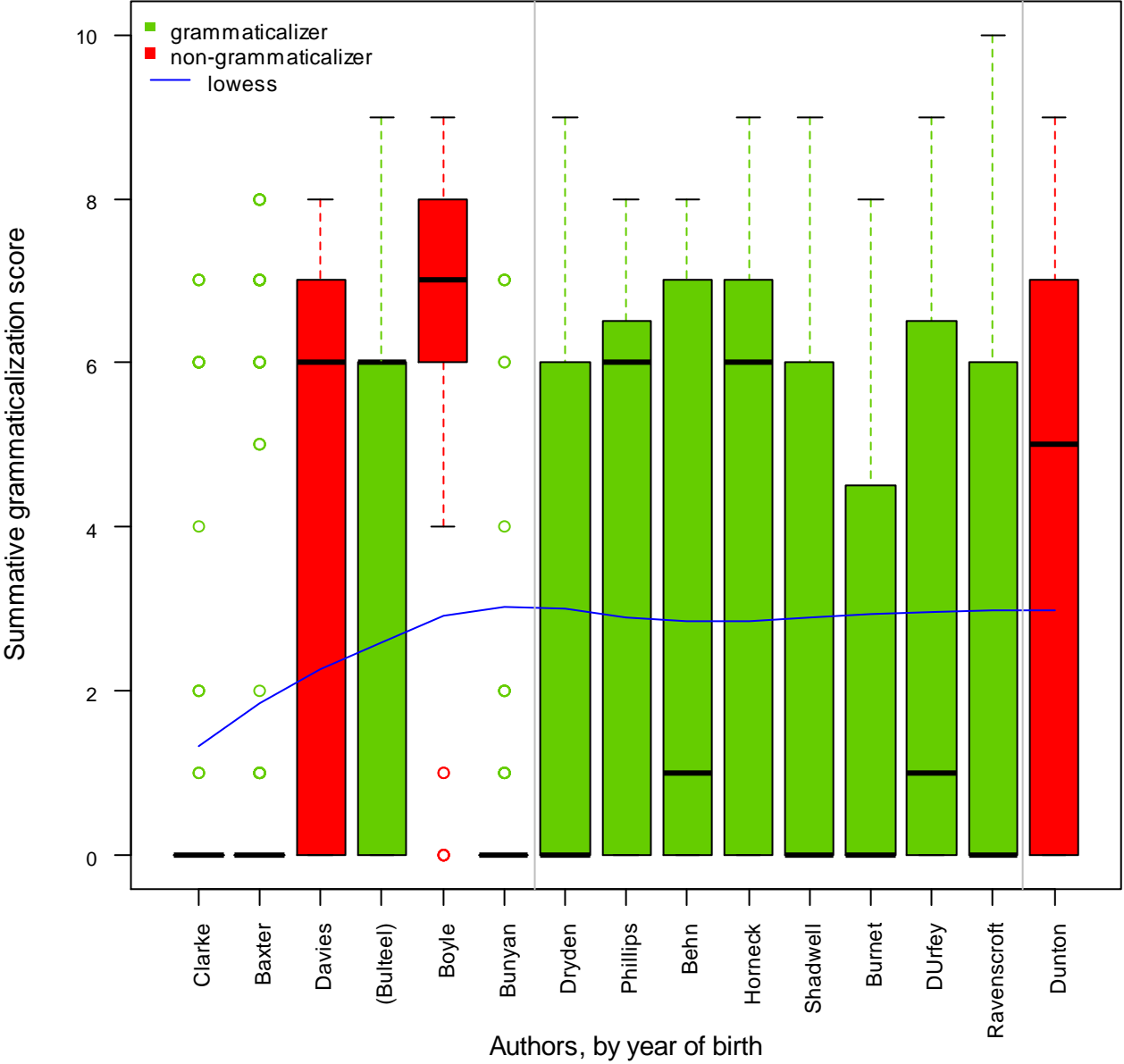
Aggregate grammaticalisation, with lowess regression line
(Correlation: Kendall tau = 0.126, $p < 0.0001$)

Findings

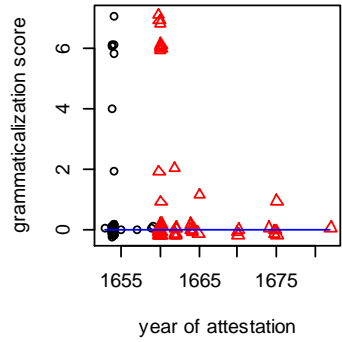


Author aggregate scores, with lowess regression line
(Correlation: Kendall tau = 0.221, $p < 0.0001$)

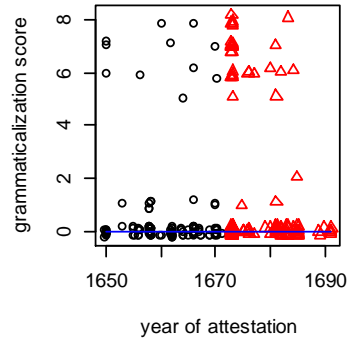
Summative grammaticalization Score



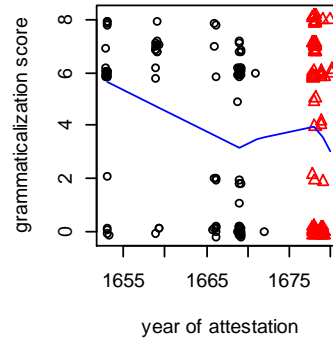
Samuel Clarke



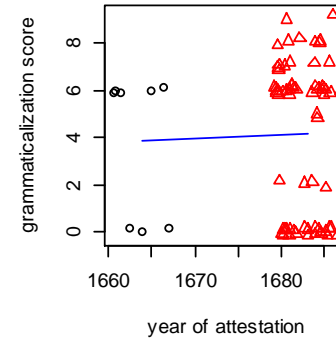
Richard Baxter



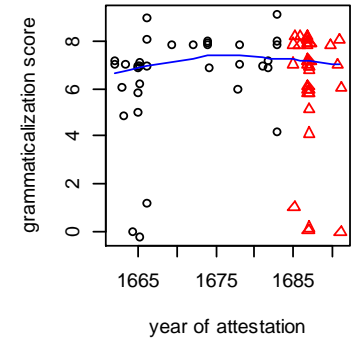
John Davies



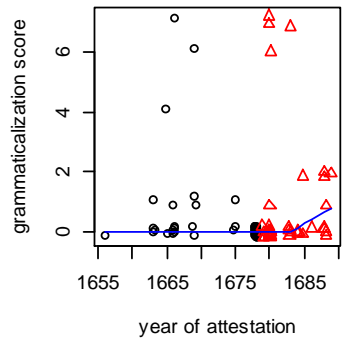
(John Bulteel)



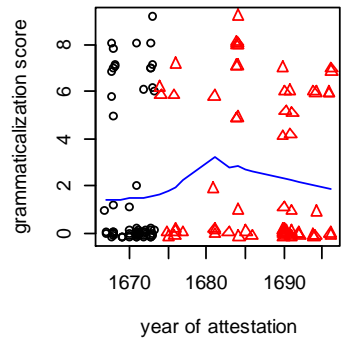
Robert Boyle



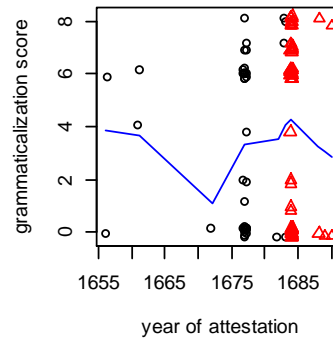
John Bunyan



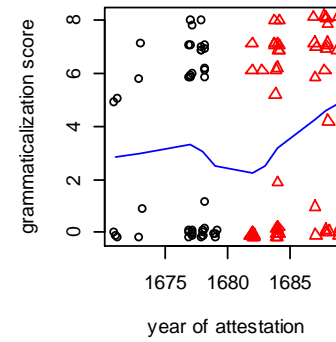
John Dryden



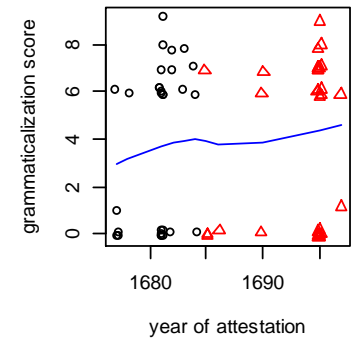
John Phillips



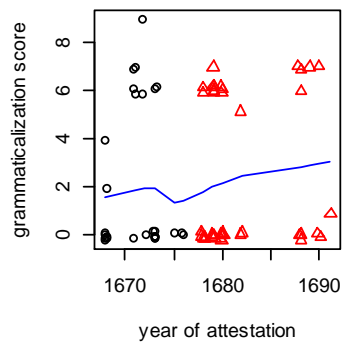
Aphra Behn



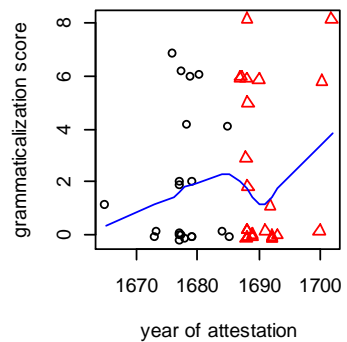
Anthony Horneck



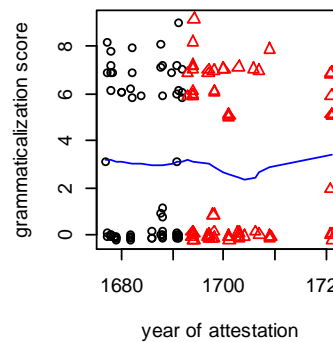
Thomas Shadwell



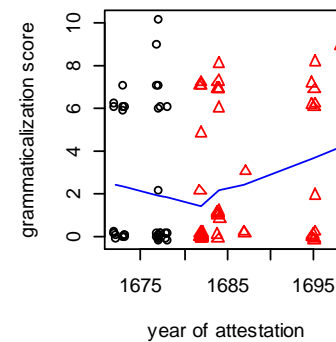
Gilbert Burnet



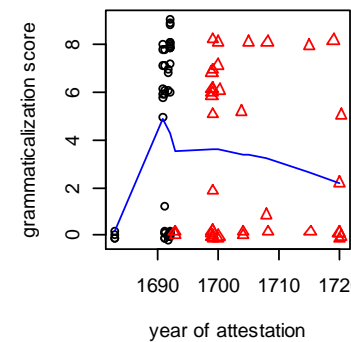
Thomas D'Urfey



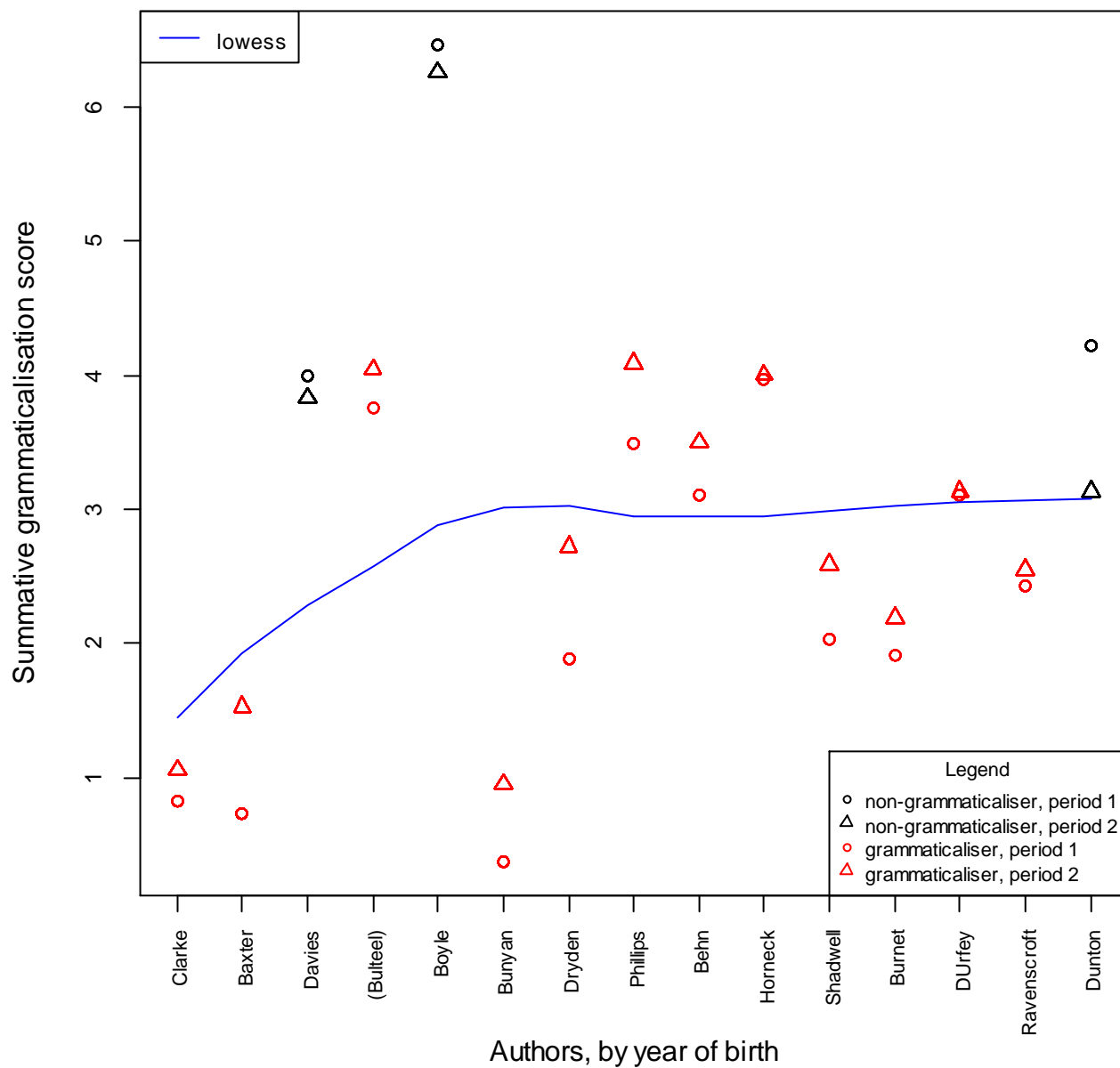
Edward Ravenscroft



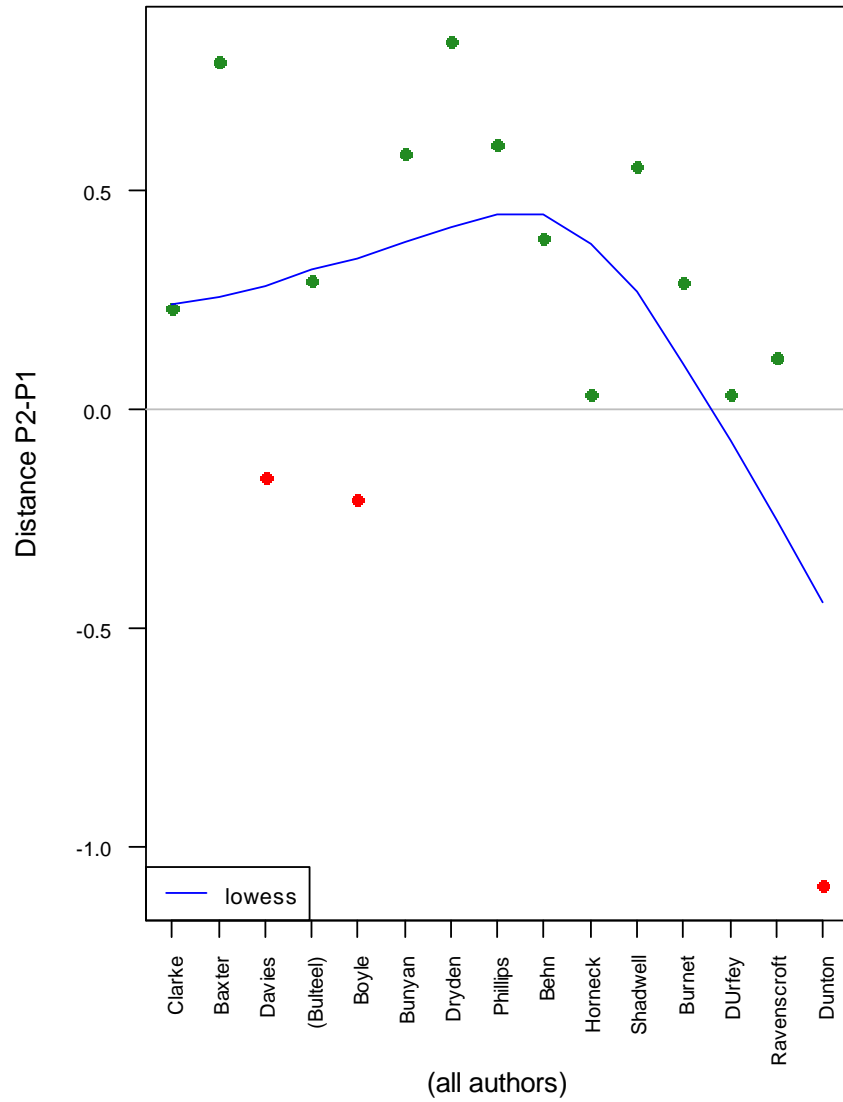
John Dunton



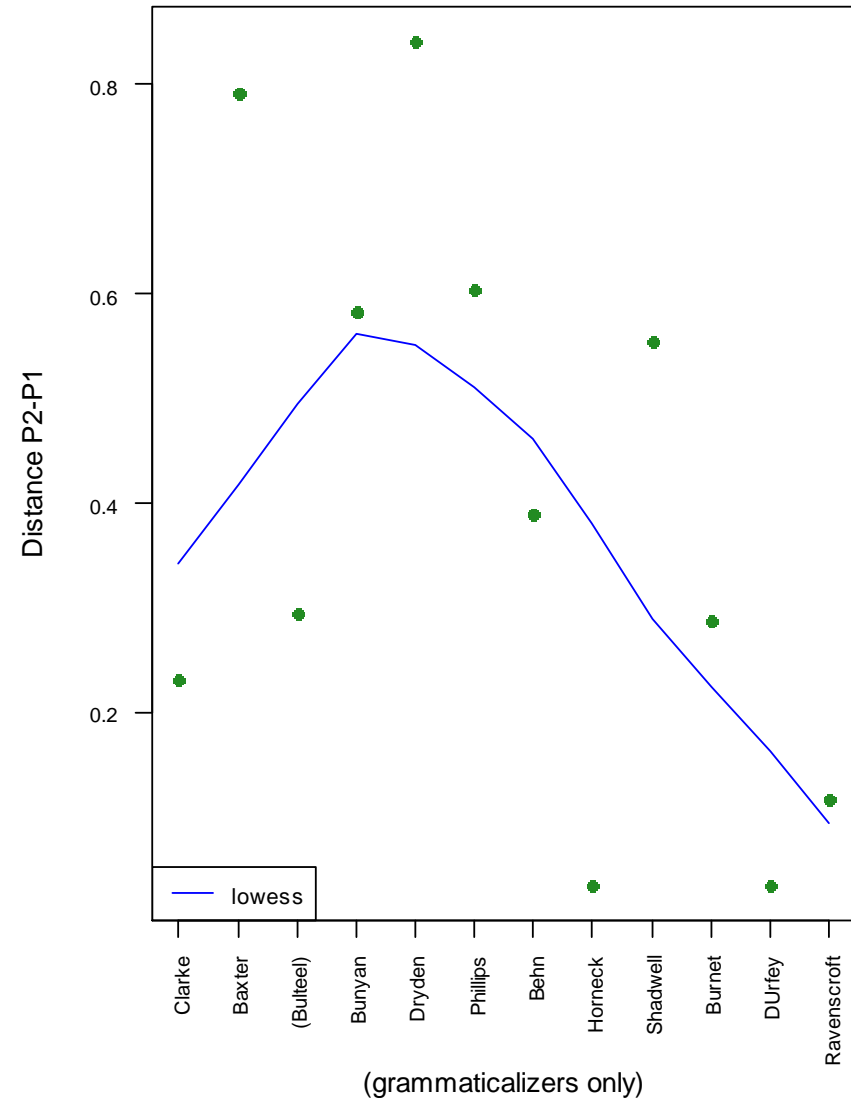
Summative grammaticalisation score



individual grammaticalization rate

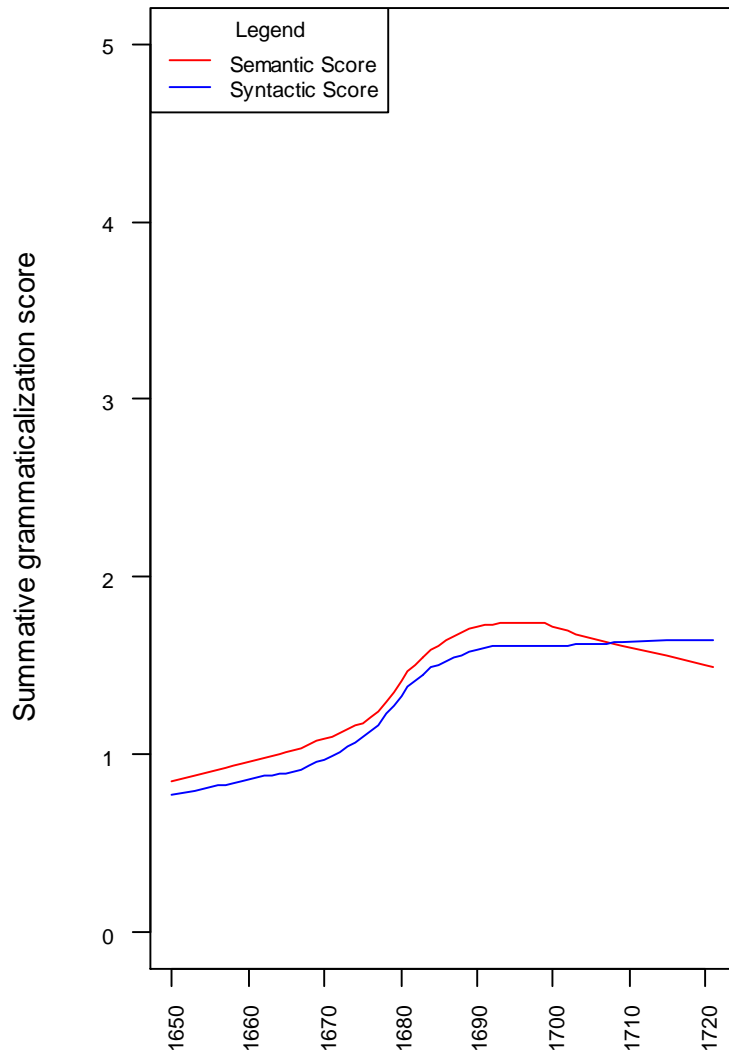


individual grammaticalization rate



Grammaticalisation rate follows a bell-shaped pattern

Syntactic and semantic change (lowess)

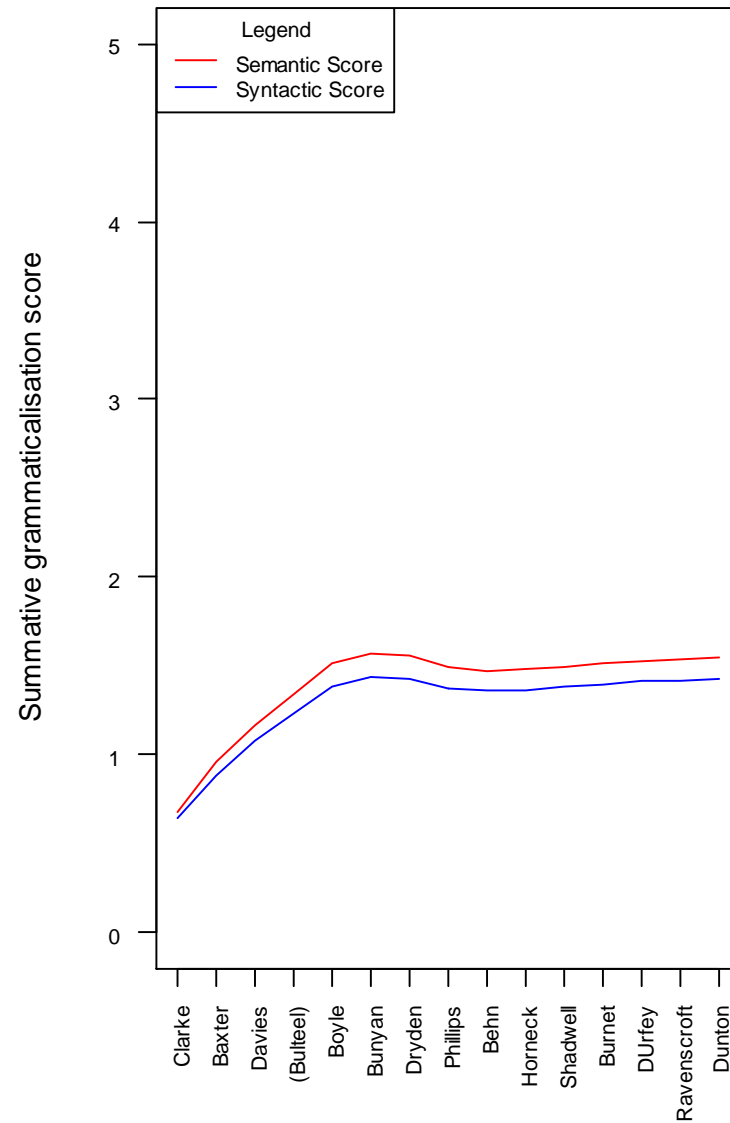


Datapoints, chronologically

Semantics: Kendall tau = 0.12, $p < 0.0001$

Syntax: Kendall tau = 0.13, $p < 0.0001$

Syntactic and semantic change (lowess)

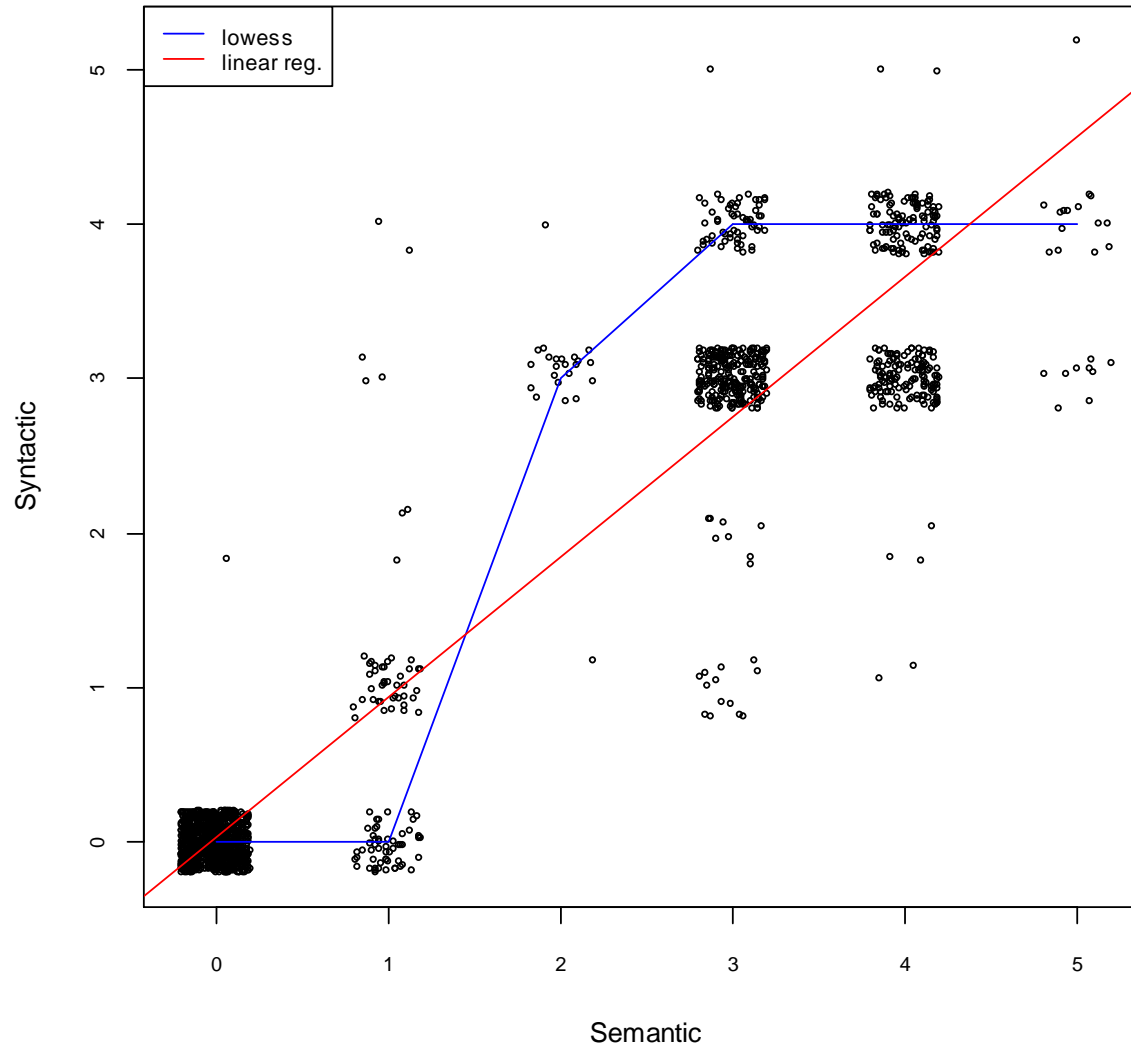


Authors, by year of birth

Semantics: Kendall tau = 0.12, $p < 0.0001$

Syntax: Kendall tau = 0.11, $p < 0.0001$

Semantic and Syntactic grammaticalization score



- Linear regression: Adjusted R-squared: 0.90 , $p < 0.0001$. Semantics and syntax co-evolve
- Lowess: syntax lags behind semantics (Heine et al. 1991:213; Givón 1991:123; Haspelmath 1999:1062; Francis & Yuasa 2008; Traugott 2008:17), **then takes over**

Comments welcome

freek.vandavelde@arts.kuleuven.be
wwwling.arts.kuleuven.be/qlvl/freek.htm