Papers of the Linguistic Society of Belgium

17 | 2023

Complexity in Complementation: Understanding lifespan change in the verb complementation of individuals

Eleanor Smith, Peter Petré, Hubert Cuyckens, Lauren Fonteyn

Universiteit Antwerpen, KU Leuven, Universiteit Leiden

URL: https://sites.uclouvain.be/bkl-cbl/en/journals/papers-of-the-lsb/volume-17-2023/smith-petre-cuyckens-fonteyn/

Electronic reference

Smith, Eleanor, Petré, Peter, Cuyckens, Hubert & Fonteyn, Lauren. (2023). Complexity in Complementation: Understanding lifespan change in the verb complementation of individuals. In Marie Steffens & Thomas Hoelbeek (eds.), *Papers of the Linguistic Society of Belgium* [online] 17, 107-137. DOI: https://doi.org/10.61430/NOYS7030

Complexity in Complementation: Understanding lifespan change in the verb complementation of individuals

Eleanor Smith[1,2], Peter Petré[1], Hubert Cuyckens[2], Lauren Fonteyn[3]

[1] Universiteit Antwerpen, [2] KU Leuven, [3] Universiteit Leiden

Abstract: This paper combines an individual and period level approach to the study of long term variation in English syntax. It is hoped that by combining these views, a more complete picture of how individuals process, accommodate, and spread changes in language can be constructed to complement the general literature on population level language variation. The data used is taken from twelve individuals across two periods. It consists of >500,000 words per individual, with complementation clauses of the verb *remember* annotated for six variables. Multifactorial classification models are then employed to determine which language-internal factors an individual uses to condition the variation in their linguistic output, and to compare the relative importance of the constraints across individuals and periods. Results show that individuals prioritise partly idiosyncratic systems over larger semantic groupings, creating substantial degrees of inter-individual variation; we argue this correlates with continued long-term variation across periods. A drop in the degree of idiosyncrasy between the earlier and later period shows potential standardisation at play. Finally, independently from the well-known connection with social background from sociolinguistics, tentative evidence is found that a minority of individuals predict the next period's usage patterns, perhaps marking themselves as 'way-pavers'; these individuals show in-group similarities in lifespan usage patterns.

1 Introduction

Across many theoretical frameworks, linguistic variation is believed to be best studied at the level of speech communities. Some linguists have previously claimed that in studying language change the individual is "reduced below the level of linguistic significance" (Labov 2012: 265) . But this view is too narrowly focused on social identity, underplays individual differences (Tamminga et al. 2016), and leaves important questions unanswered: Does individual (cognitive) processing impact the spread of a variant at the community level, and, if so, how/why? How do individuals accommodate change in their understanding or use of the language? For answers, the behaviour of individuals must be studied.

This paper provides such a pilot study of twelve writers across two periods (born in the 1660s and the 1710s respectively). It investigates changes in their use of finite and non-finite complement clauses (CCs) with the complement taking predicate (CTP) *remember*. *Remember* was chosen as the focus for this study, as it is a frequent member of a semantic set which presents with the required variation. Examples of the alternation at hand are given in (1).

(1)
a) Construction from here referred to as 0-COMP.
"I **remember** I was very well pleaſed with the Device of one that I met with on the Tomb of a young Roman Lady , which had been made for her by her Mother."
(Addison 1705)

b) Construction from here referred to as THAT.
"I do not **remember** , that the Delusions of our Dreams used to be objected against the Evidence of Sense […]"
(Sherlock 1697)

c) Construction from here referred to as *TO*-INFINITIVE.
"I **remember** to have ſeen but Two that are the Figures of Actors"
(Addison 1705)

d) Construction from here referred to as ING.
"I **remember** making a remark perfectly ſimple , and perfectly true: [...]"
(Gibbon 1796)

In the type of variation shown, the less frequent non-finite variants, first attested in Middle English (Los 2005: 254-255, Fischer et al. 2017: 111), coexist with the more frequent finite variants, complementing the historical variationist focus on

full replacement of outgoing by incoming variants (e.g. Nevalainen et al. 2011 & Los 2005). With this analysis, we aim to add to Fonteyn & Nini's (2020:18) usage-based modelling of individual variation. As such, we anticipate notable differences between the behaviours of individuals within periods, as each writer has a unique set of inputs and social experiences from which, we postulate, they generalise constraints and possible variants.

The focus on syntactic variation was chosen with regards to previous research which suggests that syntactic change resides below the level of awareness (Labov 1993:4, Labov 2001:28). Although more recent studies such as Levon & Buchstaller (2015) call into question the strong formulation of Labov's (1993:4) hypothesis, evidence is still found that although listeners may be judging structural features such as syntactic constructions "they do so in a comparatively more complex fashion than for phonetic ones" (Levon and Buchstaller 2015:304). This points to less explicit awareness of syntactic variation, which may contribute to a lower influence of social evaluation, as the population finds such variation less salient than that at other levels of the language. This makes syntactic variation, such as variation between finite and non-finite CCs, an ideal case to study the role of cognitive representations in individual behaviour.

Given the previous work in the field, in this study we aim to test two hypotheses, these are given in (2).

(2)

a) Following Rohdenburg's (1996:15) complexity principle, finite constructions will be favoured in cognitively more complex contexts. This is because they may be easier to process as they contain overt information on subject, tense, aspect and modality (Van Driessche and Cuyckens 2019:72) .
b) The direction of the lifespan change on the individual level will follow the direction of the observed inter-period change.

In this paper, section 2 covers the details related to the data used in this analysis. Section 3 describes an overview of the methodology, and section 4 the results of the analysis. Section 5 provides an in depth discussion of the findings and the final section, 6, outlines the conclusions which can be drawn from this work.

2 Data

The data set used in this study consists of two periods of individuals, each containing 6 writers. The data for individuals in period1 is taken from the EMMA corpus (Petré et al. 2019), the data for individuals in period2 is taken

from the ECCO TCP corpus (Text creation partnership 2005-2010). These corpora were specifically chosen as they collate the full oeuvre of the individuals in their data, allowing for a detailed individual level diachronic analysis. To minimise extraneous variables, individuals across all periods were selected in line with a general profile, characteristics of which include being male, white, British, and based mainly in London during their writing career. The genre of texts was not restricted and as such the two periods do not show the same balance of genres, with period1 containing more religious material and period2 containing more fiction. Genre is not considered in this paper but will be taken into account in later analysis once a larger number of authors have been annotated. The ranges in birth dates and active career dates for each period are shown in Table 1.

Table 1: Birth date, Active career ranges and sources of data for periods 1 & 2.

	Birth Years	Active Careers	Data Source
Period 1	1635-1657 (22 years)	1659-1737 (78 years)	EMMA corpus
Period 2	1685-1735 (50 years)	1707-1802 (95 years)	ECCO TCP corpus

Across the 12 individuals, 612 instances of the CC variation with *remember* were found and annotated for 6 variables (for details on annotation see section 3.1). The distribution of these instances across the data set is outlined in Table 2.

Table 2: Distribution of CC instances with *remember* across the data set.

Period Number	Author	Count Total Words	Count Finite Instances	Count Non-finite instances	Count Total Instances
1	Joseph Addison	487,207	THAT: 7 0-COMP: 19	*TO*-INFINITIVE: 9	35
1	William Sherlock	2,076,365	THAT: 82 0-COMP: 15 How: 1	*TO*-INFINITIVE: 2	100

1	Edward Stillingfleet	2,974,637	THAT: 28 0-COMP: 22 How: 1	*TO*-INFINITIVE: 7	58
1	William Wake	1,143,686	THAT: 48 0-COMP: 11	*TO*-INFINITIVE: 1	60
1	Daniel Whitby	1,925,091	THAT: 31 HOW: 2	*TO*-INFINITIVE: 6	39
1	George Whitehead	1,284,629	THAT: 18 0-COMP: 13 HOW: 6	*TO*-INFINITIVE: 2	39
2	George Berkeley	483,503	THAT: 7 0-COMP: 9	*TO*-INFINITIVE: 11	27
2	Edmund Burke	1,237,897	THAT: 58 0-COMP: 3	*TO*-INFINITIVE: 11	72
2	George Colman	488,978	THAT: 16 0-COMP: 13	*TO*-INFINITIVE: 13 ING: 3	45
2	Richard Cumberland	1,119,378	THAT: 17 0-COMP: 38	*TO*-INFINITIVE: 10 ING: 2	67
2	Edward Gibbon	1,239,403	THAT: 20 0-COMP: 2	*TO*-INFINITIVE: 4 ING: 1	27
2	John Trusler	2,801,867	THAT: 23 0-COMP: 11	*TO*-INFINITIVE: 4 ING: 5	43
-	-	17,262,641	521	91	612

3 Method

The general methodology of the study consists in annotating the data, then running conditional inference tree and random forest analyses on the individual and period levels. The aim of this is to determine the importance of the annotated factors in conditioning the choice between different constructions for individuals and on an aggregate level. Lifespan change analysis was also conducted using 2x2 Fisher's exact tests to test the significance of any differences in CC use in the first vs second half of each individual's active career.

3.1 Annotation

All instances were manually annotated for six variables. Table 3 provides an overview.

Table 3: Overview of the labels used in the annotation process.

Label	Options
Length of CC (length_of_clause)	Number of tokens
Negation in the CC (cc-neg)	present/ not present
Voice of the CC (voice)	active/ passive
Meaning of CC (cc_means)	event/ state
Animacy of CC subject (anim)	animate/ inanimate
Co-reference of the subject across the main clause and CC (co-ref)	same/ different

These labels were selected in line with the findings of previous work such as Cuyckens, D'hoedt & Szmrecsanyi (2014) and Van Driessche and Cuyckens (2019). The selection of variables takes into account Rohdenburg's (1996:151) complexity principle that "in the case of more or less explicit grammatical options, the more explicit ones(s) will tend to be preferred in cognitively more complex environments"' (1996:151) by trying to measure complexity of the surrounding environment. This is inline with our hypothesis, given in (2), that finite constructions can be seen as easier to process, and as such may be favoured in cognitively complex environments.

At this point in the project it was decided not to include individual or period as factors of a single multivariable model. Instead separate analyses for each period and individual were conducted to allow for an in depth look into the patterns at play at all levels while still allowing for some comparison between groups. In future research, with larger amounts of data available, we plan to include both the individual and period number as explicit factors in more holistic statistical analyses in order to test the significance of the inter-period/individual differences observed in this paper.

3.2 Analysis

Several types of statistical analyses were run in this study. Firstly the conditional inference tree (CIT) algorithm 'Ctree' from the R package 'partykit' (Hothorn and Zeileis 2015) was used. CITs are a decision tree method based on recursive partitioning, which finds the optimum characteristics from the given labels to explain splits in the data (see Figure 1). The alpha level, or significance threshold, for the CITs in this study was set to 0.05.

The second statistical analysis used in this study was the conditional random forest (CRF) algorithm 'Cforest', also from the R package partykit. CRF is an ensemble learning method for classification and regression tasks, which operates by constructing a multitude of decision trees and taking the average of their predictions. We combine these methods as "CITs can be particularly useful for explanation and interpretation, whereas CRFs are usually better in prediction" (Levshina 2020:614) giving us a wider scope of study than with one of the methods alone. Figure 1 (taken from Levshina 2020:612) shows a simplified example of the kind of binary partitioning performed in CIT and CRF. When the data is split based on shape 80% of objects grouped with stars are white and 80% of objects grouped with circles are blue, meaning that splitting the data based on shape allows us to predict the colour correctly in 80% of cases (Levshina 2020:612. For further reading also see Gries 2019).

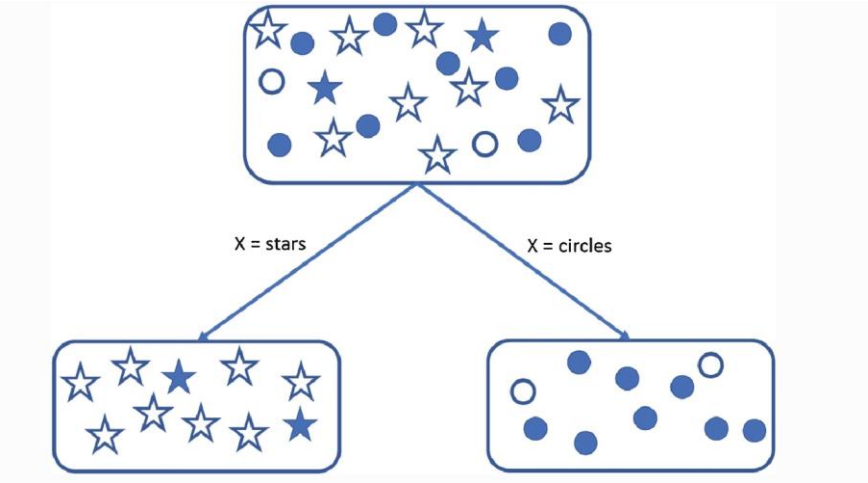

Figure 1: Simplified example of binary partitioning taken from Levshina 2020:612

One reason these analysis methods were specifically chosen is that they are robust in situations with limited amounts of data (Fonteyn & Nini 2020). CRFs were also chosen as they are particularly useful when predictors are highly intercorrelated (Levshina 2020:613) as is often the case with linguistic data.

For both the CIT and CRF analyses the six annotation labels shown in Table 3 were used as the possible factors. These are length of the CC, presence of negation in the CC, voice of the CC, meaning of the CC, animacy of the CC subject, and co-reference of the subject across the main clause and CC. A full overview of the distribution of each of the labels can be found in *appendix a* of this paper.

Alongside the statistical analysis of the data, a qualitative analysis concerning idiosyncratic constructions was also conducted on the individual level. Constructions are considered idiosyncratic when they fall outside of the general usage patterns of the period. This is operationalised as follows; no more than two authors may have the same behaviour for it to be considered idiosyncratic. Including this method adds depth to the individual level of the analysis, allowing for further consideration of individual motivation and personal cognitive and literary style.

4 Results

This section outlines the results of the different analyses. The results are first split by period, with 4.1 providing the details of the analysis of the period1 data

and 4.2 providing the information on the results of the analyses of the period2 data. The final section, 4.3, provides the results of the lifespan change analysis. This analysis covers all individuals from both periods and examines changes in their individual construction usages across the span of their active careers. As mentioned in (2), we hypothesised that the direction of the lifespan change on the individual level will follow the direction of the observed inter-period change.

4.1 Period 1

4.1.1 Aggregate

The CIT results for period1 at the aggregate level are shown in Figure 2. Here we can see that the only significant factor conditioning the choice between the use of a finite or non-finite construction is subject co-reference. 99.53% of instances with different subjects for the main clause and CC are finite constructions. Whitehead provides us with 1 instance of, a different subject non-finite construction, given in (3). When we look at instances with the same subject across both the main clause and CC the majority are also finite, but there is a small group of non-finites present all of which are *TO*-INFINITIVE.

(3)
"*Saving in its own nature and property , which we do affirm it to be ; but that he enlightens every man to Salvation ,* **I do not remember these to be our words** *, as T. D. lays them down [...]*"
(Whitehead 1669)

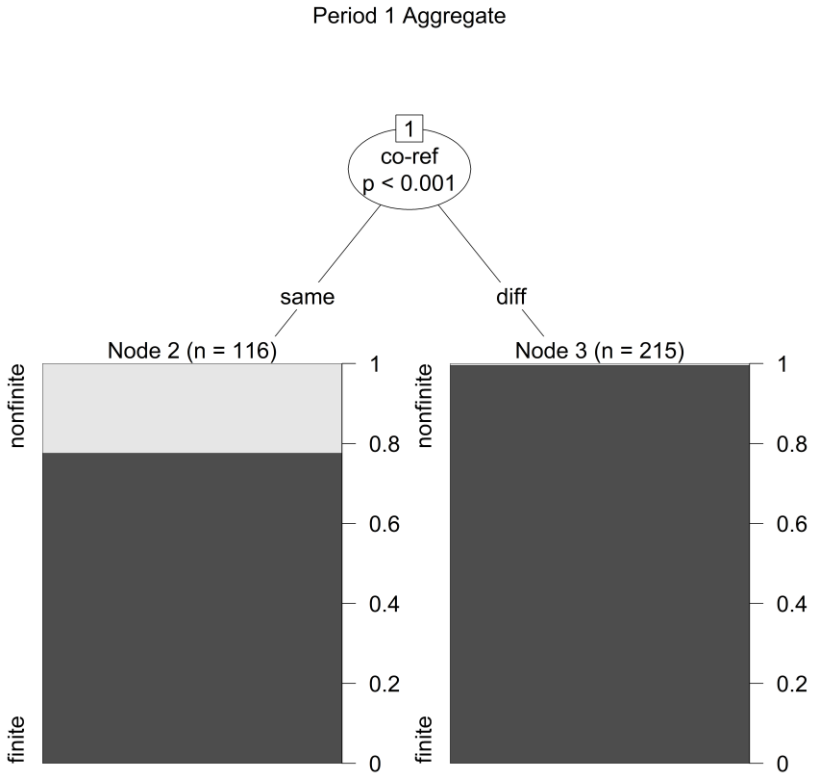

Figure 2: CIT results for the period1 aggregate level.

CRF results for period1 on the aggregate level (see Figure 3) show that co-reference of the subject is indeed the most important factor in the choice between finite and non-finite constructions. Although, 'co-ref' is still the only factor found to be significant by the Ctree in period1, we can clearly see that 'length_of_clause' seems to be more important than other factors in the CRF. It is also followed by 'voice' and 'cc_neg' which show some level of involvement.

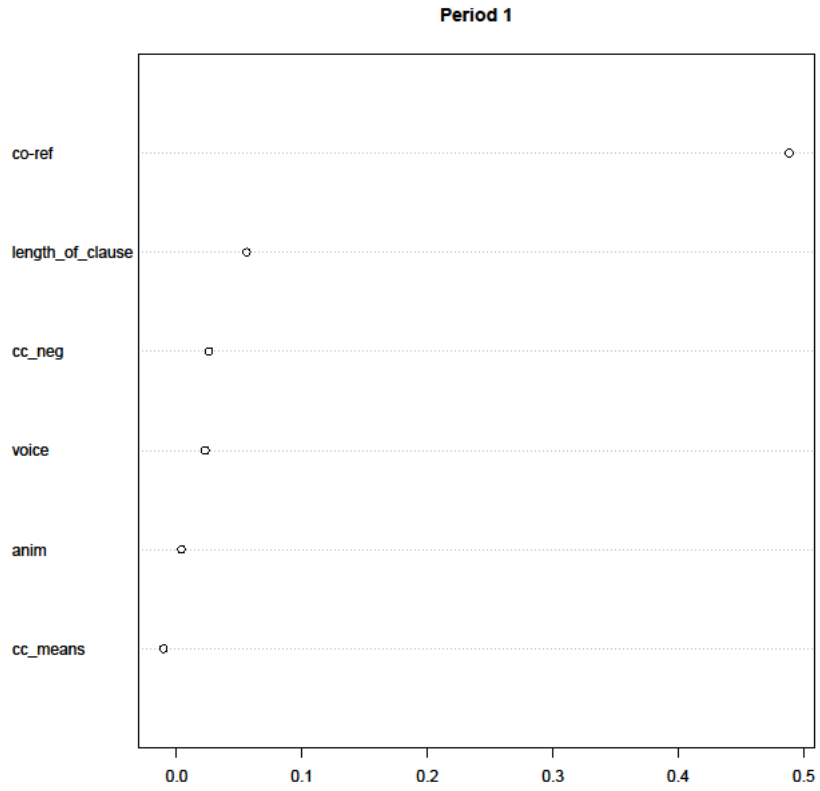

Figure 3: CRF results for the period1 aggregate level.

4.1.2 Individual

Of the 6 authors in period1, 3 show significant factors conditioning their construction choice. These individuals are Addison, Stillingfleet and Whitby. Figure 4 shows the individual CIT results generated from their data. The limited results on the individual level may be due to data scarcity issues. Future annotation of additional verbs will allow us to test this. The results of Figure 4 show that, as for the aggregate level, for all three individuals 'co-ref' is the only significant factor conditioning construction choice.

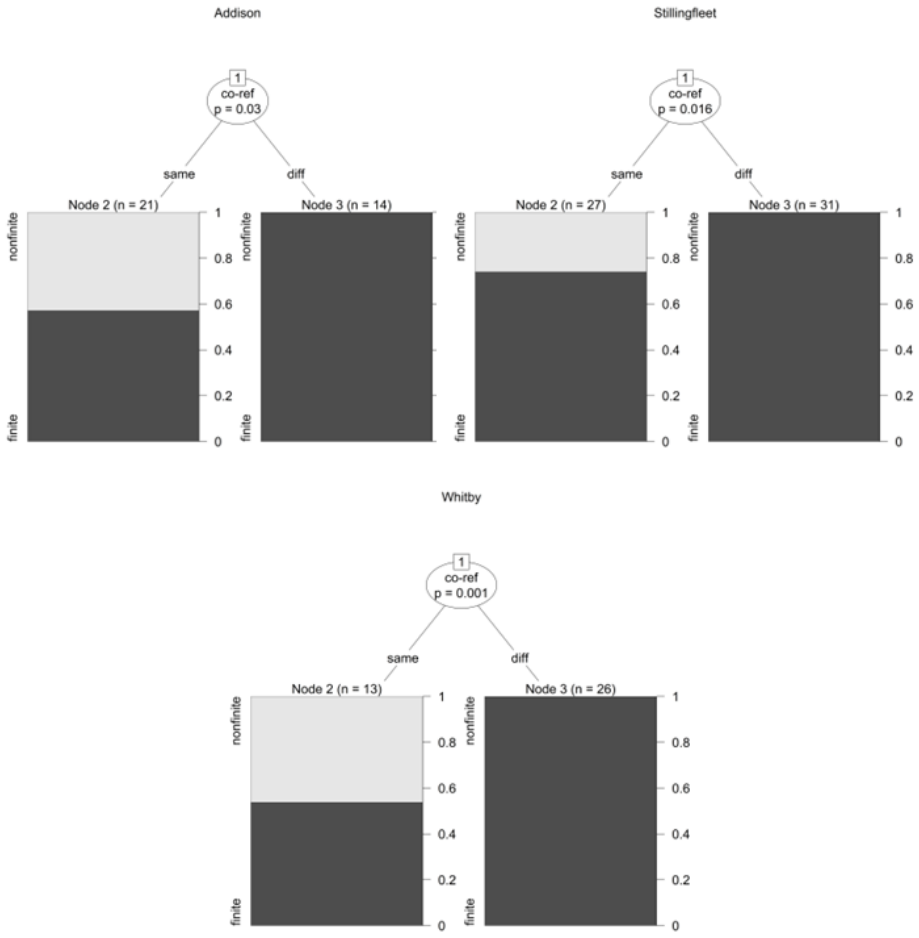

Figure 4:Individual CITs for period1 (Addison, Stillingfleet and Whitby)

Table 4: Split in individual behaviours of period1 into 2 groups comparing the minimally and maximally explicit constructions. Group1 is shown in red, group2 is shown in blue.

Author	Lifespan THAT use	% non-finite construction use	Idiosyncrasies present
Sherlock	Increase	2% *TO*-INFINITIVE= 2 ING= 0	None
Wake	Increase	1.7% *TO*-INFINITIVE= 1 ING= 0	Minimal
Whitehead	Stable	5.12% *TO*-INFINITIVE= 2 ING= 0	yes
Stillingfleet	Stable	12.06% *TO*-INFINITIVE= 7 ING= 0	yes
Whitby	Stable	15.39% *TO*-INFINITIVE= 6 ING= 0	yes
Addison	Stable	26% *TO*-INFINITIVE= 9 ING= 0	yes

When the constructions of all six individuals are examined, two groups can be distinguished, as shown in Table 4. Group1 (shown in red) is characterised by a lack of individual idiosyncrasies in constructions (see below), low usage of non-finite constructions and a trend towards increased THAT use over the individual's lifespan. Conversely, the behaviour of group2 (shown in blue) can be characterised by the presence of idiosyncrasies for each individual, a higher (in the cases of Whitby, and Addison, much higher) use rate for non-finite constructions and stable use of THAT over their lifespans.

Table 5 presents an overview of the individual idiosyncrasies of period1. Although this study focuses on verbal complementation, in Table 5 Stillingfleet is shown to have a nominal gerund CC which was deemed relevant as no other nominal or verbal gerund complements were found with *remember* in the data. It could be seen that this nominal construct is a forerunner for the verbal gerund structure found in period2 and beyond. As such, it is classified as an individual idiosyncrasy of Stillingfleet. Stillingfleet, Wake and Addison are all shown to have at least one instance each of the formulaic period2 style *TO*-INFINITIVE 'to

have X-ed' which is rare in the data of this period but plentiful in period2. Generally, then, higher use of idiosyncratic patterns (which might be considered as a higher level of syntactic creativity) appears to correlate with higher use of the incoming variant of *TO*-INFINITIVE. Whitehead shows a preference for constructions using finite HOW throughout his lifespan, even combining this with a THAT to create the novel HOW-THAT finite construction which is not found in any other individual. Whitby also provides us with a novel construction type in his use of 'NP+ *TO*-INFINITIVE', where the NP is (redundantly) coreferential with an argument of the infinitival clause. This use is not shown in the data of the other writers. An example with context is given in (4).

(4)
"*The Command for* **remembring the Seventh Day from the Creation to rest upon it from all manner of Work** *, was Ceremonial and not Moral [...]*" (Whitby 1688)

Table 5: Overview of the idiosyncrasies present across period 1.

Author(s)	Idiosyncrasy	Example
Stillingfleet	'*Remember*+nominal gerund'	*"he remembers* **the Christening of bells** *among them"*
Stillingfleet, Wake & Addison	Instance(s) of period 2 style *TO*-INFINITIVE	*"remember* **to have X-ed***"*
Whitehead	Preference for finite 'how'	*"Do you not remember* **how** *you came with your Drums and Fiddles?"*
Whitehead	Finite 'HOW-THAT' construction	*"you may Remember* **how that**, *in my Letters to you, I gave you Information[...]"*
Whitby	Innovative 'NP+ *TO*-INFINITIVE' construction	See example given in (4)

4.2 Period 2

4.2.1 Aggregate

Figure 5 shows the results of the CIT analysis on the period2 aggregate data. One difference between these results and those for period1 (Figure 2) is that, while subject co-reference is still the most significant factor conditioning the choice

between finite and non-finite constructions, it no longer gives rise to a near-perfect (99.53%) split. In the period2 results we can see that there is now a small group of differing subject non-finite constructions, something that was generally not observed within the constraints of period1. It is also interesting to see that in the period2 results a second branch now appears showing that a further significant split can be made within the same subject instances based on clause meaning. The significance of clause meaning is especially interesting due to its status as the least important factor for period1 as shown in Figure 3, thus highlighting a definite change in the importance of variables between the periods. The figure shows that instances describing a state tend to be finite constructions, and those describing an event tend to be non-finite constructions.

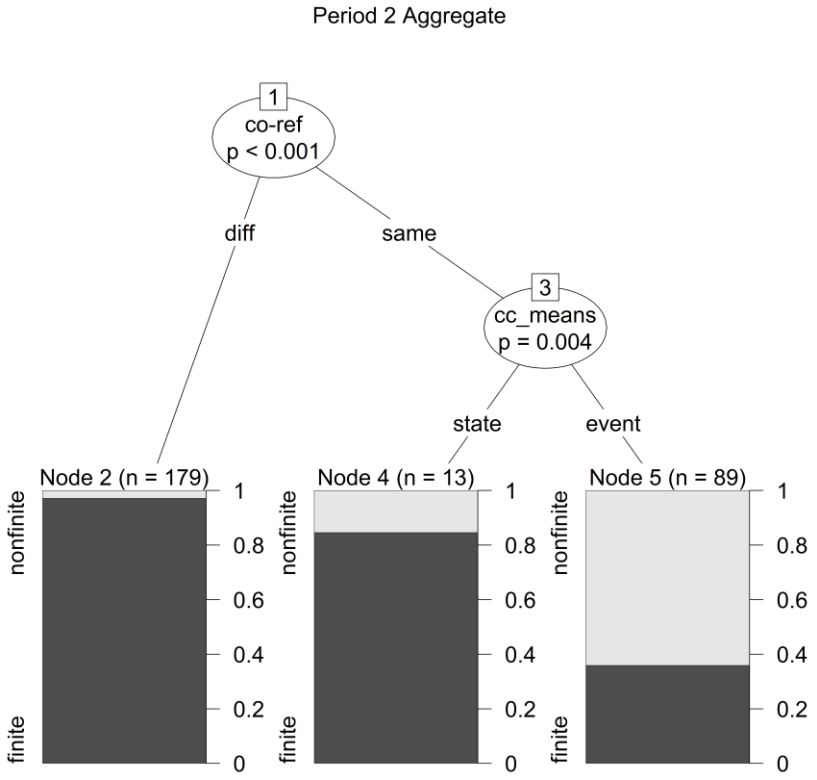

Figure 5: CIT results for the period 2 aggregate level.

CRF was also conducted on the period2 aggregate data; the results can be seen in Figure 6. Here, subject co-reference is still by far the most important factor conditioning the choice between finite and non-finite construction. Both CC meaning and presence of negation in the CC also appear to show a small effect. Compared to the results of period1 (Figure 3) voice and length of the CC no longer appear in this group of factors. As such it can be suggested that the appearance of clause meaning as a significant factor has displaced structural variables, signalling a switch to functional variables as being more significant.

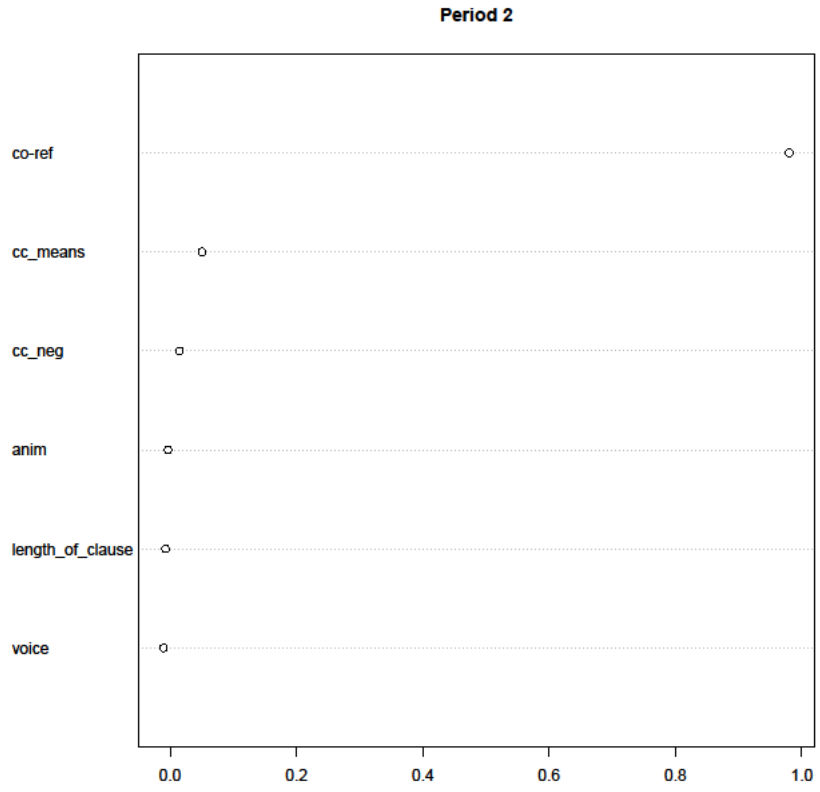

Figure 6: CRF results for the period 2 aggregate level.

4.2.2 Individual

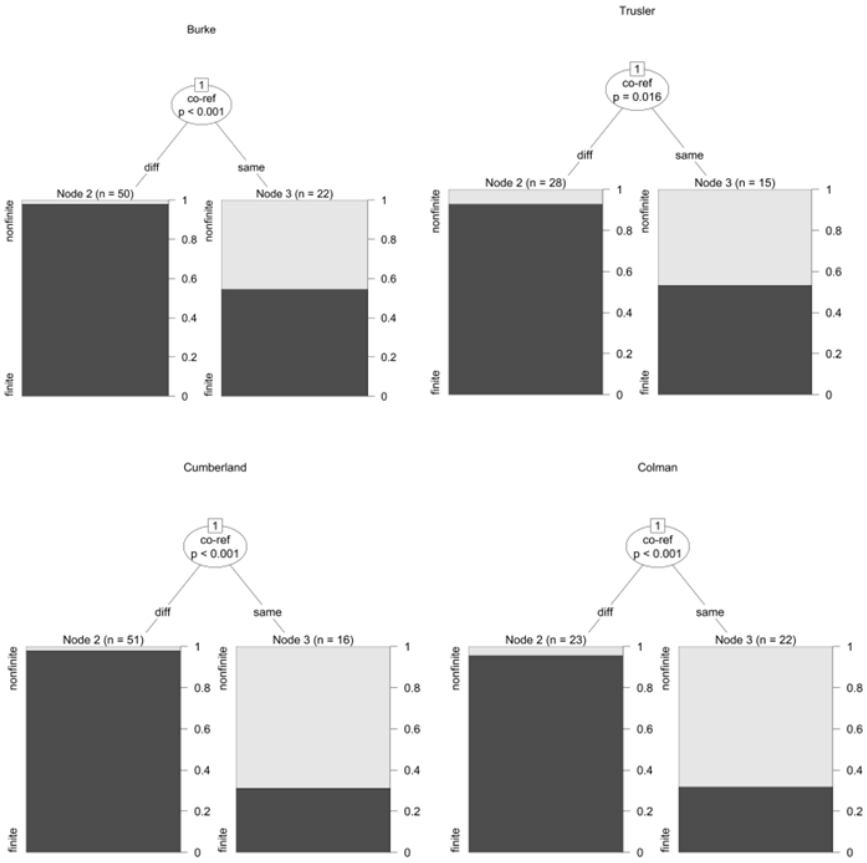

Figure 7: CITs for individuals in period2 which exhibit differing subject non-finite constructions. (Burke, Trusler, Cumberland and Colman)

In contrast to period1, all 6 individuals in period2 show statistically significant factors conditioning their construction choice. The CITs for period2 on the individual level are shown in Figure 7 and Figure 8. For all of the individuals, subject co-reference is found to be the only significant factor conditioning the choice between finite and non-finite constructions. Two of the individuals follow the pattern found in period1, with the vast majority of non-finite constructions having same subject co-reference (Figure 8), on the other hand, 4/6 individuals allow a higher rate of differing subject non-finite constructions (Figure 7), showing loosening of the strength of a constraint compared to period1.

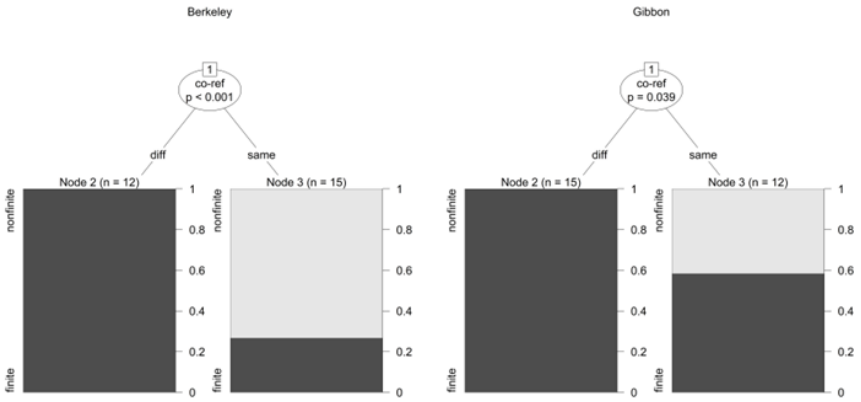

Figure 8: CITs for individuals in period2 which exhibit only finite differing subject co-reference instances. (Gibbon and Berkeley)

Table 6 shows an overview of the results for individual analysis in period2. The behaviour in this period is much more homogenous than period1, 5/6 show a general increase in THAT use across their lifespan mirroring the behaviour of the group1 individuals in period1 (Table 4). General non-finite usage is higher in this period than in period1, also showing a wider variety of uses. With regards to the innovations in non-finite use in this period, only 2/6 individuals don't make use of ING (Burke and Berkeley), only 2/6 don't generate any differing subject non-finite constructions (Gibbon and Berkeley), and Cumberland is the only individual to use *TO*-INFINITIVE outside of the 'to have x-ed' construction. The individual that appears to be most committed to the new constructions is Trusler, with 5 verbal gerunds, 2 of which have differing subjects from their main clauses, and 100% use of 'to-have x-ed'.

Table 6: An overview of the individual level results of period 2.

Author	Lifespan THAT use	% non-finite construction use	Differing subject co-reference non-finites
Burke	**Increase** 1st half: 78.26% 2nd half: 81.63%	15.2% *TO*-INFINITIVE = 11 (100% 'to have x-ed') ING = 0	*TO*-INFINITIVE x1
Cumberland	**Increase** 1st half: 14.81% 2nd half: 32.5%	17.9% *TO*-INFINITIVE =10 (60% 'to have x-ed') ING =2	ING x1
Gibbon	**Decrease** 1st half: 100% 2nd half: 58.82%	18.5% *TO*-INFINITIVE =4 (100% 'to have x-ed') ING =1	0
Trusler	**Increase** 1st half: 40.62% 2nd half: 90.9%	20.9% *TO*-INFINITIVE =4 (100% 'to have x-ed') ING =5	ING x2
Colman	**Increase** 1st half: 26.66% 2nd half: 53.33%	35.5% *TO*-INFINITIVE =13 (100% to have x-ed) ING =3	*TO*-INFINITIVE x1
Berkeley	**Increase** 1st half: 20% 2nd half: 42.85%	40.7% *TO*-INFINITIVE =11 (100% 'to have x-ed) ING =0	0

4.3 Lifespan Change

Using a 2x2 Fisher's exact test, it was found that 4 individuals showed statistically significant changes in their use of CC constructions over their active careers. Fisher's exact test was chosen due to the low sample size and presence of low counts for some constructions. The span of publishing dates for each individual was split in half, with the first and second half of the publishing career spans compared. Two individuals with significant results, Stillingfleet and Sherlock, are from period1 and the other two, Gibbon and Trusler, are from period2 of the data set.

Sherlock (p= 0.0443) and Trusler (p=0.005) show a statistically significant increase of THAT, whereas Gibbon shows a significant decrease (p=0.0261) of THAT. On the other hand, Stillingfleet (p=0.0104) shows a significant decrease of non-finites.

While the authors that show significant lifespan change present differing patterns, the behaviour of Stillingfleet and Trusler is interesting when compared to their overall data. These are the two authors that showed the most innovative behaviour of their respective periods, and yet all of their innovative use is situated in the first half of their careers. Trusler's significant increase of THAT late in his career shows a reversion to the most explicit variant, and Stillingfleet's significant decrease of non-finites in the second half of his career similarly shows a turn away from the least explicit variant. This lifespan change is against the direction of the observed intergenerational change.

5 Discussion

The findings presented in section 4 show several broad patterns. In general, there is higher heterogeneity in period1 shown in the range and frequency of idiosyncratic use, with the majority of idiosyncrasies shown in period1 not reappearing in period2. This perhaps indicates that the possible range of variation is 'standardised' somewhat in this time. One example of this is the homogeneous way in which *TO*-INFINITIVE is constructed in the second period (e.g. 'to have x-ed') as opposed to the many varying constructions found in the first period. However, it should also be considered that the genre differences between the two periods may have contributed to the change in variation in this case.

Another finding in the results is the increase from 8% non-finites to 23% between periods, brought about by the expansion of the *TO*-INFINITIVE and the appearance of ING in period2. This direction of inter-period change reflects the more general expansion of non-finite construction use and variation in this time period. Appearance of a functional second constraint, clause meaning, in period2 shows a move towards possible exaptation echoing de Cuypere (2008:15) as "a

neutral or non-functional structure becomes functional", the formal variation between construction types acquires its own functional interpretation.

The results shown in Table 4 on period1 show a possible correlation between lifespan stability of THAT, a higher degree of non-finite use and the presence of idiosyncrasies in group2. Considering Rohdenburg (1996), one possible shared motivation for these behaviours is a lower concern for explicitness and as such perhaps a lower concern for clarity or reader processing load. Group1 shows a preference for the most explicit choice, especially as they age, and avoids use of non-finite constructions when possible throughout their lifespans. The lack of individual idiosyncrasies in this group may also be seen as an attempt at clarity or explicitness, as novel forms could be seen as difficult for readers to process and understand (Tzuyin Lai et al. 2009). In comparison group2 does not seem to share these concerns. With respect to the complexity principle (Rohdenburg, 1996:151), the type of complement itself may also be ordered along a cline of complexity; non-finites would be more complex to process because of the lack of explicit subject and tense/aspect-marking, followed by 0-Comp, which is more explicit but lacks an explicit starting point of the CC, and finally THAT as the most explicit option. When all individuals in period1 are taken together, the correlation between high use of non-finites and less insistence on THAT over 0-COMP, shown in the groupings of Table 4, reaches statistical significance (pearson's correlation two-sided test: rho=0.95, N=5, df=3, t=5.5721, P=0.01141), again strengthening the idea that these two distinct combinations of behaviours show some clear difference in the strategies employed, perhaps due to different cognitive/ processing styles (Rayner and Riding 1997) or shared motivation such as lowering complexity (Rohdenburg 1996).

The results in section 4 show a possible 'way paver' in each period. When the idiosyncrasies of period1 (Table 5) are considered, the behaviour of Stillingfleet seems to lay the groundwork for what becomes the aggregate level patterns in period2. He is the only individual in this period who employs a precursor to the verbal gerund, uses 100% period2 style *TO*-INFINITIVE, and has the second highest percentage use of non-finites (13.2%) which also puts him in line with the general level of non-finite use found in period2. Looking at the individual behaviours in period2 (Table 6) Trusler seems the most committed to the new patterns appearing in this time period. Trusler shows the highest use of the new ING, 2/5 also have differing subject co-reference. The appearance of non-finites of any construction with differing subject co-reference was exceptional in period1 and becomes more entrenched in this period, so combining this new construction with previously disfavoured differing subject co-reference shows commitment and confidence in the newly emerging variants. The analysis of lifespan change shown in section 4.3 shows that our two 'way-

pavers' show significant changes in their lifespan usage patterns. Both Stillingfleet and Trusler's behaviour can be characterised as being full of early career innovation followed by retrograde lifespan change. This suggests that it may have been difficult for them to keep up this innovative behaviour. Whether this retrograde change in the second half of their careers was due to social pressure or change in self-alignment (Sankoff & Wagner 2006:214), which assumes some socio-cognitive accessibility of the variation for these individuals (Buchstaller 2016:221), or some other reason, is not possible to know at present, but we hope to shed more light on how general this behaviour is, and where it comes from, in future research.

6 Conclusions

Overall, there are several conclusions that can be drawn from this study. First, idiosyncrasies are more defining of individuals' usage than larger categories found at the period level. This adds to Fonteyn & Nini's (2020:18), usage-based modelling of individual variation as the idiosyncrasies shown in period1 are localised to specific individuals, with most not appearing in the data for period2 .

Another key finding shows how individual processing may impact the spread of a variant, as possible 'way-pavers' are found in both periods. Stillingfleet (period1) and Trusler (period2) are tentatively labelled as 'leaders of change'. The fact that these two individuals show shared lifespan usage patterns also raises questions about the general behaviour of individuals with this profile/status and how they accomplish their roles in terms of promoting language variation and change. The retrograde pattern of their usage in the second half of their careers prompts further research into why this kind of patterning may occur in this context.

Finally, we find that long term variation on the period level is facilitated by heterogeneous idiosyncrasies on the individual level. This is shown in the reflection of previous idiosyncrasies of some authors in period1 becoming the main aggregate patterns in period2. Without the heterogeneous behaviour of the individuals in geneation1 testing the constraints and generating novel constructions, innovation, variation and change would not occur in period2.

With regards to future work, we will further investigate the retrograde lifespan change exhibited by 'way-paver' individuals. The study will also be expanded to include 15 individuals per period, and a third period (born c.1810) will be added in order to test that the patterns found in this study hold when a second period jump is considered. Lastly, we plan to repeat and collate the analyses shown here for 6 CTPs in total to compare behaviours across different verbs and discover if there are larger patterns at play in variation of verbal complementation constructions at a more abstract level. Further in depth lifespan change analysis with multiple CTPs will also facilitate discussion of how

individuals accommodate change in their use of the language as population patterns change. Another point of future research would be to include the semantic senses of *remember*, as described in Cuyckens et al (2014). This would allow us to contrast differences in the patterns of variation between the senses, adding further depth to our analysis.

References:

Anthonissen,L. (2020). Special passives across the lifespan Cognitive and social mechanisms. Dissertation, Universiteit Antwerpen & R. Ludwig-Maximilians-Universität München.

Buchstaller, I. (2016). Investigating the Effect of Socio-Cognitive Salience and Speaker-Based Factors in Morpho-Syntactic Life-Span Change. *Journal of English Linguistics.* Vol. 44(3) pp.199–229. DOI: 10.1177/0075424216639645

Cuyckens, Hubert, Frauke D'hoedt & Benedikt Szmrecsanyi. (2014). Variability in verb complementation in Late Modern English: Finite vs. non-finite patterns. In Marianne Hundt (ed.), 182-203.Cambridge: Cambridge University Press.

de Cuypere, L. (2005). Exploring exaptation in language change. *Folia Linguistica Historica*, 39(Historica-vol-26-1-2), 13-26. https://doi.org/10.1515/flin.26.1-2.13

Fischer, O., De Smet, H., & Van der Wurff, W. (2017). The Verb Phrase. *In A Brief History of English Syntax* (pp. 106-132). Cambridge: Cambridge University Press. doi:10.1017/9781139049559.007

Fonteyn, L. and Nini, A. (2020). Individuality in syntactic variation: An investigation of the seventeenth-century gerund alternation . *Cognitive Linguistics* 31 (Issue 2), pp. 279-308.

Th Gries, Stefan. (2019). "On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement" *Corpus Linguistics and Linguistic Theory*, 16, no. 3, 2020, pp. 617-647. https://doi.org/10.1515/cllt-2018-0078

Hothorn T, Zeileis A (2015). "partykit: A Modular Toolkit for Recursive Partytioning in R." *Journal of Machine Learning Research*, 16, 3905-3909. https://jmlr.org/papers/v16/hothorn15a.html.

Labov, W. (2012). What is to be learned: The community as the focus of social cognition. *Review Cognitive Linguistics* 10.263−29

Labov, W. (1993). The unobservability of structure and its linguistic consequences. Paper presented at the 22nd New Ways in Analyzing Variation conference, University of Ottawa, October pp.22–25.

Labov, W. (2001). Principles of linguistic change. Vol. 2: Social Factors. Philadelphia: John Benjamins.

Levon, E., & Buchstaller, I. (2015). Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. *Language Variation and Change*, 27(3), pp.319-348. doi:10.1017/S0954394515000149

Levshina, N. (2020). Conditional Inference Trees and Random Forests. In: Paquot, M., Gries, S.T. (eds) *A Practical Handbook of Corpus Linguistics*. Springer, Cham. https://doi.org/10.1007/978-3-030-46216-1_25

Los, B. (2005). The Rise of the To-Infinitive. OUP Oxford. https://doi.org/10.1093/acprof:oso/9780199274765.001.0001

Nevalainen, T., H. Raumolin-Brunberg & H. Mannila. (2011). The diffusion of language change in real time. *Language Variation and Change* 23. pp.1–43

Petré,P. ,Anthonissen,L. ,Budts,S. ,Manjavacas,E. ,Silva,E. ,Standing,W. & Strik,O. (2019). Early Modern Multiloquent Authors (EMMA): Designing a large-scale corpus of individuals' languages. *ICAME Journal*,43(1) pp.83-122. https://doi.org/10.2478/icame-2019-0004

Rayner,S. Riding, R. (1997) Towards a Categorisation of Cognitive Styles and Learning Styles, *Educational Psychology*, 17:1-2, 5-27, DOI: 10.1080/0144341970170101

Rohdenburg, G. (1996) Cognitive complexity and increased grammatical explicitness in English. , Vol. 7 (Issue 2), pp. 149-182.

Sankoff, Gillian and Wagner, Suzanne Evans (2006) "Age-grading in retrograde movement: The inflected future in Montréal French," *University of Pennsylvania Working Papers in Linguistics:* Vol. 12: Iss. 2, Article 16.
Available at: https://repository.upenn.edu/pwpl/vol12/iss2/16

Tamminga, M. L, MacKenzie and D, Embick. (2016) The Dynamics of Variation in Inidividuals. *Linguistic Variation* Vol.16 Iss.2. pp.300-336. https://doi.org/10.1075/lv.16.2.06tam

Text creation partnership, Eighteenth Century collections online. 2005-2010. Available at https://quod.lib.umich.edu/e/ecco/.
Last accessed 02/08/2022.

Tzuyin Lai, V. Curran, T. Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research* Vol.1284, pp.145-155. https://doi.org/10.1016/j.brainres.2009.05.088.

Van Driessche, L. Cuyckens, H. (2019). Clausal verb complementation in varieties of English. *JournaLIPP*, 2019(7) pp.72-91.

Appendix:

Appendix a: An overview of the distribution of annotation labels per author.

Author	Period number	Average clause length	Co-reference distribution	CC subj Animacy distribution	CC meaning distribution	CC negation distribution	CC Voice distribution
Addison	1	19.71	*Same*: 60% *Diff*: 40%	*Anim*: 85.74% *Inan*: 14.26%	*Event*: 65.71% *State*: 34.29%	*Yes*: 42.85% *No*: 57.15%	*Active*: 88.57% *Passive*: 11.43%
Sherlock	1	19.64	*Same*: 33% *Diff*: 77%	*Anim*: 72% *Inan*: 28%	*Event*: 48% *State*: 52%	*Yes*: 16% *No*: 84%	*Active*: 90% *Passive*: 8%
Stillingfleet	1	20.94	*Same*: 46.55% *Diff*: 53.45%	*Anim*: 81.03% *Inan*: 18.97%	*Event*: 74.13% *State*: 25.87%	*Yes*: 20.68% *No*: 79.32%	*Active*: 89.65% *Passive*: 10.35%

Author	Period number	Average clause length	Co-reference distribution	CC subj Animacy distribution	CC meaning distribution	CC negation distribution	CC Voice distribution
Wake	1	22.25	*Same*: 36.66% *Diff*: 63.33%	*Anim*: 70% *Inan*: 30%	*Event*: 60% *State*: 40%	*Yes*: 8.33% *No*: 91.67%	*Active*: 93.33% *Passive*: 6.67%
Whitby	1	19.92	*Same*: 33.33% *Diff*: 66.67%	*Anim*: 74.35% *Inan*: 25.65%	*Event*: 66.66% *State*: 33.34%	*Yes*: 7.69% *No*: 92.31%	*Active*: 89.74% *Passive*: 10.26%
Whitehead	1	22.94	*Same*: 25.64% *Diff*: 74.36%	*Anim*: 79.48% *Inan*: 20.52%	*Event*: 76.92% *State*: 23.08%	*Yes*: 7.69% *No*: 92.31%	*Active*: 87.17% *Passive*: 12.83%

Author	Period number	Average clause length	Co-reference distribution	CC subj Animacy distribution	CC meaning distribution	CC negation distribution	CC Voice distribution
Berkeley	2	27.25	*Same*: 55.55% *Diff*: 44.45%	*Anim*: 81.48% *Inan*: 18.52%	*Event*: 70.37% *State*: 29.63%	*Yes*: 18.51% *No*: 81.49%	*Active*: 96.29% *Passive*: 3.71%
Burke	2	24.73	*Same*: 30.55% *Diff*: 69.45%	*Anim*: 69.45% *Inan*: 30.55%	*Event*: 79.16% *State*: 20.84%	*Yes*: 15.27% *No*: 84.73%	*Active*: 86.11% *Passive*: 13.89%
Colman	2	23.57	*Same*: 48.88% *Diff*: 51.12%	*Anim*: 71.11% *Inan*: 28.89%	*Event*: 71.11% *State*: 28.89%	*Yes*: 0% *No*: 100%	*Active*: 82.22% *Passive*: 17.78%

Author	Period number	Average clause length	Co-reference distribution	CC subj Animacy distribution	CC meaning distribution	CC negation distribution	CC Voice distribution
Cumberland	2	15.13	*Same*: 23.88% *Diff*: 76.12%	*Anim*: 73.14% *Inan*: 26.86%	*Event*: 53.74% *State*: 46.26%	*Yes*: 11.94% *No*: 88.06%	*Active*: 97.02% *Passive*: 2.98%
Gibbon	2	19.59	*Same*: 44.44% *Diff*: 55.56%	*Anim*: 78.78% *Inan*: 22.22%	*Event*: 70.38% *State*: 29.62%	*Yes*: 7.4% *No*: 92.6%	*Active*: 85.19% *Passive*: 14.81%
Trusler	2	22.69	*Same*: 34.88% *Diff*: 65.12%	*Anim*: 58.14% *Inan*: 41.86%	*Event*: 69.77% *State*: 30.23%	*Yes*: 27.9% *No*: 72.1%	*Active*: 93.03% *Passive*: 6.97%