

On the probability and direction of morphosyntactic lifespan change

Lauren Fonteyn^{1*} and Peter Petré²

¹Centre for Linguistics, Leiden University, Leiden and ²Department of Linguistics, University of Antwerp

*Corresponding author. E-mail: l.fonteyn@hum.leidenuniv.nl

Abstract

The aim of this study is to further contribute to the ongoing debate regarding the nature of “morphosyntactic lifespan change,” defined here as observable shifts in the grammatical choices individuals make between competing morphosyntactic structures. Through a quantitative case study of competition between two types of *ing*-nominals in seventeenth-century English, in which we factor in the grammatical contexts in which the variant structures can be used, we show that individuals vary in the extent to which they participate in the contextual diffusion of a new structure. We furthermore show that there is interindividual variability with respect to whether and what kind of lifespan change—frequency, constraint, and inventory change—is attested and highlight different patterns of intraindividual change: progressive, retrograde, and “mixed.”

The question of whether individual language users modify their internalized grammatical system during adulthood is one that has long divided linguists of different theoretical persuasions. On the one hand, usage-based models of language tend to consider language as an inherently “adaptive” system that is in a constant state of change, both at the individual and the community level (Diessel, 2019). In such models, the linguistic behavior of individuals is considered to be based on the rules they have inducted from “the frequency of the different expressions and constructions encountered by the speaker” during the past social interactions they have had with other individuals (Barlow, 2013:444; also see Beckner, Blythe, Bybee, Christiansen, Croft, Ellis, Holland, Ke, Larsen-Freeman, & Schoenemann, 2009:14). Importantly, these utterance production rules inducted from language use are, essentially, never entirely fixed and can be affected by every new instance of language use the individual encounters (Bybee, 2006). Thus, the idea that an individual’s internalized grammar is susceptible to change lies at the core of the utterance-based (Croft, 2000:45–55) or usage-based model, in which the “current and past interactions [of individual language users] together feed forward into [their] future behavior” (Beckner et al., 2009:2).

On the other hand, there are various theoretical models of language where grammar, at least at the level of the individual speaker, is conceived as a relatively fixed

system. Such models are often (but not exclusively) characterized by a more modular or componential view of language, wherein the linguistic knowledge of individuals comprises a lexicon and a separate rule system (or “grammar”) that generates the syntactic patterns in which lexical items can be inserted. With respect to the possibility of language change across the individual’s lifespan then, it is generally held that modifications in the lexicon are not restricted by age (Kerswill, 1996; Meisel, Elsig, & Rinke, 2013:26). By contrast, more “systemic changes” that affect the rule system that generates grammatical sentences, “only take place in the process of the transmission and incrementation of a change, that is, during childhood” (Meisel et al., 2013:37), when the grammatical system is imperfectly transmitted from parent to child (for recent proponents, see Anderson [2016], Lightfoot [2019]). Systemic (e.g., morphosyntactic) change through time is thus conceptualized as “merely a function of changes in the population, with older speakers becoming inactive and dying, and younger speakers continually entering the community” (Denison, 2002:61). Such models of language change, which have been proposed since at least the late nineteenth century (López-Couso, 2017:339), conceptualize grammatical change as being child-based (or, at least, child-centered) rather than utterance-based.

At the same time, the two types of models are not always at odds. Both the utterance-based model and current versions of the child-based model would, for instance, accept that it is possible for adults to participate in lexical or so-called “surface structure changes,” such as the replacement of the Early Modern English third-person suffix *-(e)th* by *-(e)s* (Raumolin-Brunberg, 2005), or cases of competition between different pronominal forms (e.g., Raumolin-Brunberg, 2009). Furthermore, there is also consensus on the idea that, in cases of competition between mutually exclusive structural properties that relate to different grammars generating them (e.g., presence or absence of *do*-support in different clause types [e.g., Raumolin-Brunberg, 2009]), adults freely (and frequently) change the rate by which they use either the old or the new system. In a wide range of studies, it has been shown that “adult frequencies of linguistic forms are labile” (Tagliamonte & D’Arcy, 2007:213; in reference to Labov, 2001:447), empirically confirming that the language of adults, including the frequency with which they employ one morphosyntactic “rule” over another (or others), is not fixed in quantitative terms (Bergs, 2005; Buchstaller, 2015, 2016; Nahkola & Saanilahti, 2004; Neels, 2020; Raumolin-Brunberg, 2005, 2009; Raumolin-Brunberg & Nurmi, 2011; Sharma, Bresnan, & Deo, 2008).

Having established that the grammar of adults can and does change in quantitative terms, the focus of the discussion on adult lifespan change has subsequently shifted to the question of whether adult grammar can also be altered in “qualitative terms.” Notably, it has proven much more difficult to attest instances of what is termed “core” grammatical change, which would involve a more or less radical change “from one invariable and uniform grammatical system to another” (Meisel et al., 2013:30) in adulthood (also see the comments in Raumolin-Brunberg [2005:47], [2009:192]). Within the variationist tradition, such radical categorical changes have, to our knowledge, not yet been documented. However, some studies have reported on partial categorical or “inventory” changes (Buchstaller, Krause-Lerche, & Mechler, 2021), where a novel form is introduced or is used in a novel way (or an older form or usage ceases to be used). Petré and Van de Velde (2018) and

Anthonissen and Petré (2019) found that novel, grammaticalized uses of infinitive *be going to* are (to a limited extent) adopted in adulthood by speakers who previously only had the lexical use of infinitive *be going to* in their repertoire. Furthermore, adults may adopt uses that are more advanced on the grammaticalization scale than those in their repertoire, which may signal that usage constraints of grammatical patterns change qualitatively across the lifespan. A recent example includes the constraint change and inventory change in, among other variables, individuals' systems of quotation reported in Buchstaller et al.'s (2021) panel study. Such findings seem to call for amendment of the conclusions by Tagliamonte and D'Arcy (2007:210), who observe changes in the constraints on the use of *be like* between generations, but not within generations.

The aim of this paper is to further contribute to the ongoing debate on morphosyntactic lifespan change, defined here as observable intraindividual shifts in the grammatical choices individuals make between functionally equivalent (or competing) structures. We focus on an ongoing change involving two types of *ing*-nominals in seventeenth-century English, using statistical model comparison.

Our first hypothesis is defined simply to corroborate earlier findings on frequency-based change, which state that when a competition-driven syntactic change is taking place in a linguistic community, individuals in that community will change the frequency or rate by which they use the competing syntactic structures across their lifespan (Raumolin-Brunberg, 2005, 2009). Furthermore, based on earlier studies, we hypothesize that these changes in frequency will predominantly proceed in the direction of the ongoing change in the rest of the community (Anthonissen, 2020:325; Baxter & Croft, 2016:156; Raumolin-Brunberg, 2009:170).

A second hypothesis is formulated to assess how gradual contextual diffusion, through which syntactic changes often proceed at the community level, recurs in the linguistic behavior of individuals. Population-level studies have amply demonstrated that innovative grammatical structures tend to diffuse gradually from one grammatical context to another (e.g., Kroch, 1989; Rutten & van der Wal, 2014; Smirnova, 2015; Tagliamonte & Denis, 2016; Traugott & Trousdale, 2013). Innovative variants tend to emerge in specific contexts and only after they become sufficiently frequent or entrenched in those contexts do they "spark off the innovations that constitute the following stage" (De Smet, 2016:100). Studies that consider such stepwise patterns of diffusion at the level of the individual are relatively rare, but it has tentatively been argued that the "innovation-entrenchment-further innovation" trajectory observable in population-level language also exists "in patterns of variation across individual speakers" (De Smet, 2016:100). By examining the frequency by which individuals adopt innovative variants in different grammatical contexts, we can help determine the extent to which this hypothesis holds true and whether the rate with which the variants occur in these contexts is comparable across individuals.

A third hypothesis we assess is that adults can participate in the diffusion of the new variant to new grammatical contexts and reach a more advanced stage of the development occurring at the community level (for similar endeavors, see Anthonissen & Petré, 2019; Buchstaller et al., 2021; De Smet, 2016; Mackenzie, 2019; Tagliamonte & D'Arcy, 2007:213). In investigating this hypothesis with real-time data, we provide some tentative (and relatively rare, see Mackenzie [2019:17])

support for the idea that the grammar of adults may be susceptible to qualitative life-span change if qualitative lifespan change is defined as changes in the grammatical contexts or constraints under which competing morphosyntactic constructions are used by each individual across their lifespan.

Data and method

Early Modern English ing-nominals

The case study chosen to address the issues at hand is the diachronically unstable variation between two types of English *ing*-nominals, illustrated in (1) and (2):

- (1) Idolatry consists in *giving of that worship which is due to God*, to that which is not God. (EMMA, Daniel Whitby, 1674)
- (2) ... the greatest part of Leviticus is employ'd in *giving Laws concerning Sacrifices* (EMMA, Daniel Whitby, 1697)

The two types of *ing*-nominals in (1) and (2) exemplify a well-known case of morphosyntactic variation in the Early Modern English period, whereby two different syntactic structures can be used largely interchangeably: in both examples, a derivational suffix *-ing* is attached to a base verb (i.e., *give*) to create a deverbal noun phrase that behaves much like other, more prototypical English noun phrases. The main difference between the examples in (1) and (2) appears to be that, in example (1), the object of *giving* occurs as an *of*-phrase, whereas the structure in (2) does not require an *of*-phrase. In terms of their contextual distribution in (Early) Modern English, both types of *ing*-nominal frequently functioned as prepositional complements but also occurred as the direct object or subject of a clause (De Smet, 2008; Fanego, 2004; Houston, 1989; Nevalainen, Raumolin-Brunberg, & Manilla, 2011).

The distributional overlap between *ing*-OF and *ing*-Ø in Modern English can be explained by the fact that the two variants are historically related: the innovative variant, *ing*-Ø, most likely developed from *ing*-OF in the Middle English period. As shown by Fanego (2004), the first instances of *ing*-Ø were attested around 1250. The subsequent diffusion of the progressive variant *ing*-Ø proceeded according to what Fanego (2004:38) has termed a “hierarchy of relative nominality.” First, *ing*-Ø became a well-established alternative to *ing*-OF in “bare” contexts, where the *ing*-form was not preceded by a determiner. Such contexts are illustrated in (3):

[no determiner: bare]

- (3) a. The soul can't dye, cannot therefore the Man dye? If not, there is no such thing, as *killing of Men*, or mortal Men. (1673, PW3)
- b. It was a cruel mercy which Tamberlane shewed to three hundred Lepers in *killing them* to rid them out of their misery (1662, SG2)

It was only after *ing*-Ø was established in determinerless contexts that it further diffused to contexts with a determiner. Over the course of the Early Modern and Late Modern period, the second stage of diffusion involved a spread to contexts with a determiner that is a possessive, as illustrated in (4).

[possessive pronoun: POSS]

- (4) a. But I perswade my selfe, it cannot well be used in the defence of **his killing of the Dragon**: (1631, Peter Heylyn)
 b. then I must believe that **his killing my Father** was no murder, and that they died wrongfully who were Executed for having a hand in his Death (1680, Roger L'Estrange)

Finally, in the third stage of diffusion, *ing-Ø* spread further to contexts where the determiner “is an article, usually definite”—illustrated in (5) (Fanego, 2004:38).¹ At this point, *ing-Ø* also started appearing following demonstratives—as illustrated in (6a)—and indefinite articles—illustrated in (6b)—albeit to a lesser extent.

[definite article: *the*]

- (5) a. how long thou hast lived to little purpose, yea, to **the killing of thy soul for ever** (1659, George Swinnock)
 b. it was given out, that **the killing an Usurper**, was always esteemed a commendable Action (1679, Gilbert Burnet)

[demonstrative, indefinite article, quantifier: other]

- (6) a. Ninthly, Protest against **that most terrible and odious shedding of the innocent blood of those 3 forementioned** (1640, William Prynne)
 b. thou wert thereby kept from **a further shedding the blood of thy soul**. (1659, George Swinnock)

Specifying our hypotheses to the diachronically unstable variation between *ing-OF* and *ing-Ø* in the seventeenth century, we can reformulate them as follows:

- Hypothesis 1: The proportion of *ing-OF* and *ing-Ø* in the linguistic output of individuals changes across their lifespan. These developments can, but need not be, in the direction of the community development (that is, an increase of *ing-Ø* at the expense of *ing-OF*).
- Hypothesis 2: There is a determiner constraint hierarchy (bare contexts > possessives > (definite) articles). This hierarchy attested at the community level can also be witnessed in the linguistic output of individuals.
- Hypothesis 3: Changing usage proportions of *ing-Ø* and *ing-OF* can occur and vary in extent and direction in different grammatical contexts within and between individuals. Additionally, determiner constraints (and accompanying usage of *ing-OF* and *ing-Ø*) can, but need not, be stable across an individual's lifespan.

Early Modern Multiloquent Authors (EMMA) corpus

The data was collected from the corpus of *Early Modern Multiloquent Authors* (EMMA), the compilation of which has been detailed by Petré, Anthonissen, Budts, Manjavacas, Silva, Standing, and Strik (2019). For the present study, it is important to highlight the following:

- The corpus contains the works of fifty authors born between 1599 and 1686, whose *oeuvres* comprise at least 500,000 words and are relatively evenly distributed across a long career (average = thirty-eight years). The authors are grouped by birth date in five separate “generations.”
- The authors included in EMMA constitute a relatively homogeneous group, with the majority (thirty-seven authors) being members of London society who spent a considerable time of their active life in London. The remaining thirteen authors are not strongly connected to London society (neither socially nor geographically). The corpus is enriched with metainformation about text type and the authors’ place of birth, social circles, and geographical mobility.

Rather than querying the entire corpus (87,126,198 words) for strings of more than two letters followed by *ing* (to exclude *sing*, *king*, etc.) and manually filtering tokens that did not represent instances of deverbal nominalizations (e.g., *evening*, *something*, etc.), we restricted our sample in the following ways:

- To control for genre variation, we restricted querying to texts labeled as descriptive or argumentative prose and letters.
- Our sample is restricted to texts produced by authors from the first three generations in the corpus (born between 1599 and 1644). As a result, the resulting dataset spans across approximately one hundred years (the “oldest” tokens dating from 1626 and the most recent ones from 1721), during which the progressive variant *ing-Ø* was steadily on the rise. With respect to the stages of diffusion described by Fanego (2004), *ing-Ø* had already been widely adopted in bare contexts. Following Nevalainen et al.’s (2011:7) classification of Labov’s (1994) five stages of linguistic change, the development in bare contexts was virtually completed (representing over 85% of all tokens). In possessive contexts, the stage of development can be classified as “mid-range” to “nearing completion” (between 35% and 85%), whereas *ing-Ø* was “new and vigorous” (15%-35%) in contexts with a (definite) article.

In total, twenty-one authors are included in our sample.² The total number of observations for each author is listed in Table 1.³

For each of these individuals, the total number of *ing*-nominals they produced in their prose and letter writing was collected. This totals up to 16,633 tokens, of which 11,862 are examples of *ing-Ø* (69%) and 4,767 are examples of *ing-OF* (31%).

Statistical analysis

We fitted a number of different logistic regression models, ranging from a basic model with no independent variables to more complex multilevel models. In the multilevel models, the following variables are included: the author, the age of the author

Table 1. Absolute and relative frequencies of *ing*-OF and *ing*-Ø per author

Generation 1 (1599–1616)	Generation 2 (1621–1631)	Generation 3 (1635–1644)
Heylyn, Peter (1599–1662) OF: 344 (46%) Ø: 401 (53%)	Boyle, Roger (1621–1679) OF: 79 (29%) Ø: 192 (71%)	Whitby, Daniel (1638–1726) OF: 506 (26%) Ø: 1426 (74%)
Prynne, William (1600–1669) OF: 496 (47%) Ø: 558 (53%)	Pierce, Thomas (1622–1691) OF: 91 (24%) Ø: 295 (76%)	Mather, Increase (1639–1723) OF: 201 (24%) Ø: 639 (76%)
Fuller, Thomas (1607–1661) OF: 172 (39%) Ø: 271 (61%)	Fox, George (1624–1691) OF: 213 (36%) Ø: 381 (64%)	Crouch, Nathaniel (1640–1725) OF: 197 (20%) Ø: 806 (80%)
Milton, John (1608–1674) OF: 235 (41%) Ø: 339 (59%)	Boyle, Robert (1627–1691) OF: 79 (13%) Ø: 515 (87%)	Behn, Aphra (1640–1689) OF: 18 (7%) Ø: 256 (93%)
Taylor, Jeremy (1613–1667) OF: 102 (21%) Ø: 380 (79%)	Swinnock, George (1627–1673) OF: 55 (17%) Ø: 262 (83%)	Burnet, Gilbert (1643–1715) OF: 509 (21%) Ø: 1948 (79%)
L'Estrange, Roger (1616–1704) OF: 460 (39%) Ø: 728 (61%)	Bunyan, John (1628–1688) OF: 401 (52%) Ø: 374 (48%)	Penn, William (1644–1718) OF: 571 (24%) Ø: 1804 (76%)
	Flavell, John (1630–1691) OF: 58 (26%) Ø: 165 (74%)	
	Tillotson, John (1630–1694) OF: 82 (32%) Ø: 176 (68%)	
	Dryden, John (1631–1700) OF: 90 (22%) Ø: 320 (78%)	

at the time a variant was produced (in order to model lifespan developments), and the determiner that preceded the variant (to model effects of grammatical conditioning). The independent variable “determiner” comprises four levels: bare contexts, possessive contexts, definite article contexts, and other contexts (including demonstratives and indefinite articles). Models will be fitted with varying “author” intercepts or varying “author” slopes. The final, most complex model includes an interaction effect between “age” and “determiner” with varying author slopes.

Bayesian regression models

A common problem for logistic mixed-effect models, which, unlike linear models, involve a binary dependent variable, is that there is a fairly high risk of (quasi-)

Table 2. Overlap and quasi-separation in the data set

	JOHN DRYDEN	
	<i>ing-OF</i>	<i>ing-Ø</i>
BARE	11 (96.3%)	284 (3.7%)
POSSESSIVE	0 (0%)	10 (100%)
THE	73 (86.9%)	11 (13.1%)
OTHER	3 (60%)	2 (40%)

separation problems, which occur when “one predictor completely or almost completely separates the binary response in the observed data” (Kimball, Shantz, Eager, & Roy, 2019:231). Ideally, when logistic regression is used to model variation, there is at least partial overlap between the dependent variables with respect to all levels of the independent variables. Overlap is demonstrated by the dataset for John Dryden in Table 2, where the levels “bare,” “the,” and “other” include observations for both of the dependent variables. With the level “possessive,” however, John Dryden’s data exhibits separation: there are ten *ing*-nominals preceded by a possessive, which are exclusively of the *ing-Ø* type. In John Dryden’s data, the predictor level “possessive” thus perfectly predicts the dependent variable.

Such (quasi-)separation could be due to the low number of observations for some conditioning factors or because one of the individuals in the dataset truly perfectly separates the dependent variable under particular conditions. Regardless, (quasi-)separation presents an issue for the way in which mixed-effect models (such as *glmer*) compute the probability of the observed data, as it can cause the estimates for the fixed and random effects to become extremely—even infinitely—large.⁴ In the present case, the problems caused by quasi-separation ultimately led to a failure to converge for *glmer*.⁵

The present study relies on a Bayesian implementation of binomial logistic regression fitted with *brm* (R [R Core Team, 2021]; package: *brms* [Bürkner, 2017, 2018]).⁶ The function underlying the models is identical to that of *glmer* in that the logistic function of the proportion of the binary dependent variable is represented as a linear function of fixed and random effects. However, the way in which the results are computed differs slightly. Bayesian models do not start from the assumption that all possible estimate values of the effects are equally likely. Instead, Bayesian computation allows the analyst to impose expectations (and constraints) on estimate values in the form of priors. Such priors assume the shape of a distribution (normal/Gaussian, Cauchy, etc.) around a mean with a certain standard deviation. They are used to state, for instance, that estimate values for the effects are *likely* near a given mean and very *unlikely* to reach the extreme values that may cause failure to converge.

In the present case, the priors for all models are “weakly informative priors”: the expectations that are set still take a wide range of possible outcomes into account but

exclude highly unlikely, extreme values that cause convergence issues. The (varying) intercept priors and priors for the fixed effects were set as distributed around a mean of zero (translating into the starting assumption that they have no effect on the choice between *ing*-OF and *ing*-Ø), with a standard deviation between one and five.⁷

Model comparison

To demonstrate the procedure, we will start by formulating the basic hypothesis that there is a considerable amount of variation between authors in the extent to which they use *ing*-Ø. To assess whether this hypothesis holds true, we fit two models. First, we fit a simple intercept model ($\text{ing} \sim 1$), which estimates the probability of the dependent variable at the population level without taking any conditions (or effects) into account. Such a model assumes that all authors exhibit identical behavior and use the innovative variant at the same rate. Second, we fit a varying intercept model, which estimates the population-level rate or probability of the dependent variable while taking differences between authors into account (thus “correcting” the population estimate for differences between individuals).

The output of Bayesian regression models should be understood as follows: they do not report a single, most likely coefficient combined with a *p*-value, but an estimate coefficient *mean*, an error rate, and a credible interval within which the estimate is predicted to fall. In Table 3, we report the mean estimate of the intercept (to be understood as the log odds of the innovative variant for the entire population) for both the simple intercept and varying intercept models. The simple intercept model estimates the intercept mean at 0.91, reporting a 95% credible interval (or CI) between 0.88 and 0.95. The varying intercept model, described in the second column, lists a population-level effect as well as a group-level effect (i.e., varying author intercept). The main difference with the simple intercept model is that the varying intercept model yields a 95% credible interval that covers a much broader range (between 0.67 and 1.24).⁸ Because the varying intercept model allows for intercept corrections by individual author, the increased range of the credible interval indicates substantial interindividual variation in terms of how often *ing*-Ø is used.⁹

The next step is to assess whether the fitted varying intercept model is better at generalizing to any new, out-of-sample data than the fitted simple intercept model. To this end, we can perform a model evaluation by means of Leave-One-Out cross-validation, or by examining their Widely Applicable Information Criterion (WAIC; Gelman, Hwang, & Vehtari, 2014; Watanabe, 2010).¹⁰ Smaller WAIC values are considered superior, but it is not immediately clear *how much* smaller the WAIC values should be. As such, we also calculate the WAIC weights, which take a value between zero and one and should be understood as the posterior probability that a particular model makes better out-of-sample predictions.¹¹

In Figure 1, the WAIC value of the varying intercept model is estimated at 19,052, which is 875 units smaller than that of the simple intercept model, which is estimated at 19,928. Figure 1 also demonstrates that the difference between the WAIC scores remains substantial, even when the standard error range of the WAIC estimates (represented by the width of the lines running through the dots) is taken into account. This suggests that the varying intercept model, which corrects estimates for

Table 3. Model comparison of simple intercept model, predicting the posterior probability of the dependent variable at the population level, and varying author intercept model predicting the posterior probability of the dependent variable and correcting the population estimate for differences between individuals (group level effects). Number of observations = 16629, chains = 4, iterations = 4000, warmup = 1000, thin = 1

Simple intercept model					Varying intercept model				
ing ~ 1					ing ~ 1 + (1 author)				
Population-level estimates					Population-level estimates				
Estimate	Est error	95% CI lower	95% CI upper	Rhat	Estimate	Est error	95% CI lower	95% CI upper	Rhat
0.91	0.02	0.88	0.95	1.00	0.96	0.15	0.67	1.24	1.00
					Group-level estimates				
					~ author (21 levels)				
					standard deviation Intercept:				
					Estimate	Est error	95% CI lower	95% CI upper	Rhat
					0.66	0.12	0.47	0.96	1.00

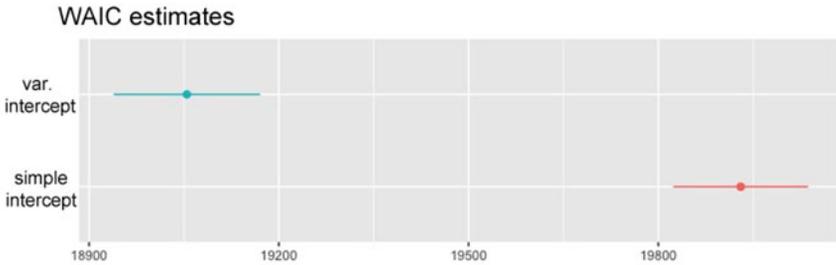


Figure 1. WAIC estimates and standard error of simple intercept model and varying author intercept model.

interindividual variation, exhibits a better fit for the data, confirming our basic hypothesis of there being nontrivial differences between authors.¹² It is this model comparison procedure that has been applied to all models presented in the analysis below.

Model overview

The assessment of the three hypotheses hinges on the comparison of twenty-two models in total. A full overview of each model can be found in Table 4.¹³ The models can be divided into four groups: base models, age-only models, determiner-only models, and full models.

As the first hypothesis stated that individuals change the extent to which they use the conservative (*ing*-OF) and innovative variant (*ing*-Ø) during their lifespan, we need to establish that a model including an independent variable “age” (e.g., $\text{ing} \sim \text{age}$) provides a better fit for the data than the base model ($\text{ing} \sim 1$).¹⁴ Additionally, we wish to determine whether there is interindividual variation in the data with respect to these changing usage rates. We can model interindividual variation in two different ways. The first way is to simply add age as a fixed effect to the varying intercept model ($\text{ing} \sim \text{age} + [1|\text{author}]$). In such a model, we can correct the population-level intercept for interindividual variation (meaning that the proportion of variant usage can differ among individuals), but we still assume that the direction in which any potential lifespan change proceeds is identical across all individuals. In other words: while the population-level *intercept* is corrected for interindividual variation, the *slope* is not. Alternatively, then, we could introduce interindividual variation as a “varying slopes” model ($\text{ing} \sim \text{age} + \text{age}|\text{author}$), which does allow for individual differences in terms of the direction of the change. The difference in how *ing*-nominal usage across the lifespan of individuals is modeled by a varying intercept model (left, by means of parallel lines) and varying slopes model (right, by means of nonparallel lines) is visualized in Figure 2.

The second hypothesis relates to the effect of the grammatical context in which the *ing*-nominal is used. As such, we include three “determiner only” models in which the main effect “determiner” is included, either in isolation or in relation to the behavior of individual authors with respect to these determiners ($\text{ing} \sim \text{det}$; $\text{ing} \sim \text{det} + [1|\text{author}]$; $\text{ing} \sim \text{det} + [\text{det}|\text{author}]$).

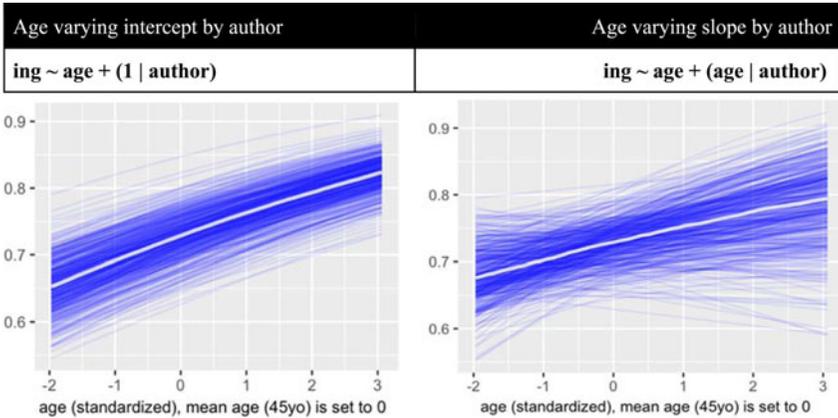


Figure 2. Effect plots of varying intercept model and varying slope model ($n = 500$).

Finally, the third hypothesis relates to the combined effect of age and grammatical context. Besides the models where either age or determiner are considered as the only fixed effects, we also consider a number of “full” models, in which a range of different feature combinations are tested. The most basic version of the full model includes “age” and “determiner” as fixed effects, without including any random effects ($\text{ing} \sim \text{age_sd} + \text{det}$). A slightly more complex model also includes the interaction between “age” and “determiner” as a fixed effect ($\text{ing} \sim \text{age_sd} * \text{det}$). We also included versions of these models where the intercept is corrected for variation between authors ($\text{ing} \sim \text{age_sd} + \text{det} + [1|\text{author}]$; $\text{ing} \sim \text{age_sd} * \text{det} + [1|\text{author}]$). Finally, full model versions were fitted with varying author slopes for age ($\text{ing} \sim \text{age_sd} + \text{det} + [\text{age_sd}|\text{author}]$; $\text{ing} \sim \text{age_sd} * \text{det} + [\text{age_sd}|\text{author}]$), determiner ($\text{ing} \sim \text{age_sd} + \text{det} + [\text{det}|\text{author}]$; $\text{ing} \sim \text{age_sd} * \text{det} + [\text{det}|\text{author}]$), the interaction of age and determiner ($\text{ing} \sim \text{age_sd} + \text{det} + [\text{age_sd}:\text{det}|\text{author}]$; $\text{ing} \sim \text{age_sd} * \text{det} + [\text{age_sd}:\text{det}|\text{author}]$), and a combination of age, determiner, and their interaction ($\text{ing} \sim \text{age_sd} + \text{det} + [\text{age_sd}*\text{det}|\text{author}]$; $\text{ing} \sim \text{age_sd} * \text{det} + [\text{age_sd}*\text{det}|\text{author}]$).

Results

Model comparison

Table 4 lists the specifications for all models. The models are ranked according to their WAIC estimates (and accompanying weights). The estimated effective number of parameters per model is also listed (p_{WAIC}). To further discuss the effect of age and determiner type across authors, we report the results of the model with the highest WAIC weight ($\text{ing} \sim \text{age_sd} + \text{det} + [\text{age_sd}*\text{det}|\text{author}]$) in the next section (“Selected model”).

The two models with the lowest WAIC estimates are the most complex models, which include age, determiner, and their interaction as random effects (with varying author slopes). The model comparison supports the proposal that both age and grammatical context are meaningful parameters and that there is substantial variation between individuals with respect to those parameters. Particularly, the inclusion of

Table 4. WAIC and WAIC weight (rounded to 0.000001) for all models, ranked by WAIC score (ascending)

Type	Parameters	p_WAIC (SE)	WAIC	SE WAIC	weight
full	age_sd + det + age_sd*det author	91.8 (2.5)	11810.6	156.9	0.737985
full	age_sd * det + age_sd*det author	94.8 (2.6)	11812.7	156.9	0.261966
full	age_sd + det + (age_sd+det author)	73.3 (1.8)	11830.0	156.7	0
full	age_sd * det + (age_sd+det author)	76.0 (1.9)	11834.8	156.8	0
full	age_sd + det + det author	58.9 (1.5)	11932.9	157.0	0
full	age_sd * det + det author	58.9 (1.6)	11938.3	157.0	0
det	det + det author	57.0 (1.5)	12003.0	157.6	0
full	age_sd + det + age_sd:det author	63.9 (1.7)	12039.0	158.3	0
full	age_sd * det + age_sd:det author	66.0 (1.8)	12040.5	158.4	0
full	age_sd * det + age_sd author	38.3 (0.8)	12095.2	158.2	0
full	age_sd + det + age_sd author	35.4 (0.7)	12096.9	158.1	0
full	age_sd + det + 1 author	22.7 (0.4)	12205.4	158.7	0
full	age_sd * det + 1 author	25.4 (0.5)	12209.4	158.7	0
det	det + 1 author	21.6 (0.4)	12272.6	159.3	0
full	age_sd + det	4.9 (0.1)	13053.7	158.4	0
full	age_sd * det	7.7 (0.2)	13056.7	158.5	0
det	det	4.0 (0.1)	13067.1	158.5	0
age	age_sd + age_sd author	37.3 (0.3)	18856.5	118.0	0
age	age_sd + 1 author	21.6 (0.3)	18985.2	116.3	0
base	1 + 1 author	20.4 (0.3)	19052.4	115.8	0
age	age_sd	2.0 (0.0)	19920.3	106.4	0
base	1	1.0 (0.0)	19928.3	106.3	0

grammatical context appears to have a substantial positive effect on the estimated prediction error of the model, as evident from the large difference between the WAIC values of the model that solely factors in determiner as a fixed effect (13067.1 [standard error: 158.5]) from the most complex age-only model (18856.5 [standard error: 118]).¹⁵

Selected model

Table 5 summarizes the selected model's parameters. At the population level, the presence of a determiner has a strong negative effect: *ing-Ø* is less likely to

Table 5. For each parameter, we provide the population-level mean estimate coefficient (Estimate), the standard deviation of the posterior distribution of the estimate (Est.Error), and the 95% Credible intervals (CI lower bound and CI upper bound). Rhat helps evaluate the estimate the posterior distribution of the parameter (Rhat = 1.0 indicates convergence). The covariance matrix of group-level effects can be found in the online repository

Family: bernoulli Links: mu = logit Formula: gerund ~ det + age_sd + (det * age_sd author) Number of observations: 16629 Draws: 4 chains, each with iter = 4000; warmup = 1000; thin = 1; total post-warmup draws = 12000					
	Estimate	Est.Error	CI lower bound	CI upper bound	Rhat
<i>Population-level effects</i>					
<i>intercept</i>	2.55	0.16	0.61	1.23	1.00
<i>determiner</i>	<i>reference level: bare</i>				
• poss	-1.36	0.17	-1.69	-1.02	1.00
• the	-3.78	0.24	-4.22	-3.29	1.00
• other	-3.35	0.23	-3.77	-2.86	1.00
<i>age</i>	0.20	0.10	0.01	0.40	1.00
<i>Group-level effects: ~author (Number of levels: 21)</i>					
<i>sd(Intercept)</i>	0.87	0.16	0.61	1.23	1.00
<i>sd(detposs)</i>	0.53	0.18	0.24	0.93	1.00
<i>sd(detthe)</i>	1.00	0.18	0.71	1.42	1.00
<i>sd(detother)</i>	0.77	0.27	0.31	1.37	1.00
<i>sd(age_sd)</i>	0.40	0.09	0.26	0.60	1.00
<i>sd(detposs:age_sd)</i>	0.71	0.24	0.33	1.26	1.00
<i>sd(detthe:age_sd)</i>	0.20	0.12	0.01	0.48	1.00
<i>sd(detother:age_sd)</i>	0.23	0.17	0.01	0.65	1.00

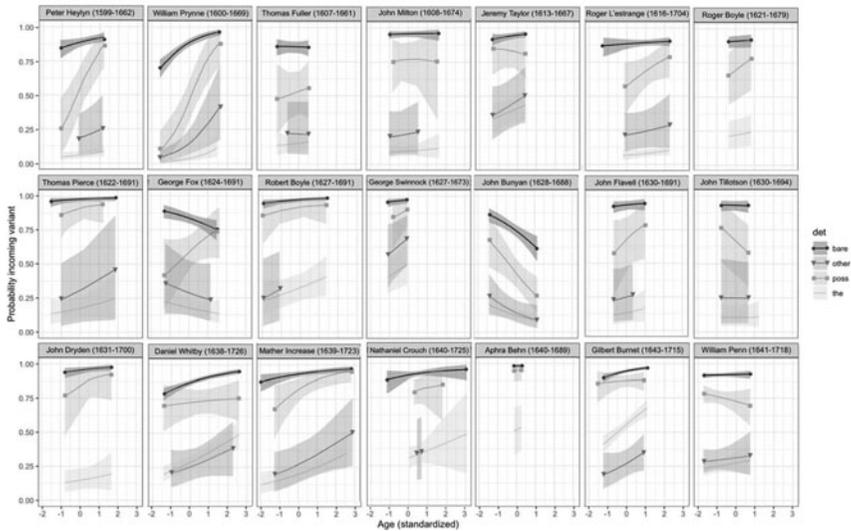


Figure 3. Effect plots of interaction effect: age*det|author.

occur following a possessive (mean coefficient estimate: -1.36 [95% credible interval: -1.69 ; -1.02]), a definite article (-3.78 [CI: -4.22 ; -3.29]), or any other determiner type (-3.35 [CI: -3.77 ; -2.86]). Age appears to have a small, positive effect (0.20 [CI: 0.01 – 0.40]). These effects are on the log-odds scale, which means the positive effect estimate of age can be interpreted as the odds of *ing-Ø* increasing by 20% as the age of the authors increases.

Interindividual variation between authors can be examined by considering the estimated standard deviations at the group level. With the exception of the interaction between age and definite article and “other” contexts, the mean standard deviation estimates as well as their lower bound are clearly away from zero, which indicates the relevance of including the random intercept and random slope terms.

To better understand these figures, it is helpful to scrutinize the interaction between age and grammatical context per author, as visualized in Figure 3. For each graph in Figure 3, the y-axis represents the posterior probability of the incoming variant, *ing-Ø*. The x-axis represents a standardized representation of age, whereby the mean, which is set to zero, should be understood as approximately forty-five years old (the average age at which tokens were produced). From these figures, it is apparent why grammatical context has a substantial effect on the explanatory power of the model: despite interauthor variation with respect to the coefficient estimates per grammatical context, the probability of *ing-Ø* is highest when the *ing*-nominal is not preceded by a determiner (line marked with dots), followed by possessive contexts (line marked with squares), and lowest when the *ing*-nominal is preceded by a definite article (unmarked line) or other type of determiner (e.g., indefinite article, demonstrative, etc.; marked with triangles) for the majority of each author’s life, mimicking the determiner constraint hierarchy observed at the population level.

The figure also helps clarify why the effect of the inclusion of age as a predictor variable appears to be more modest: demonstrable age effects are not present for every individual in the sample. This is either because there is insufficient data to factor in age for some individuals (e.g., Aphra Behn, George Swinnock), or because usage rates appear to have remained largely stable across their lifespan (e.g., Thomas Pierce, William Penn). Furthermore, there appears to be quite some variation between individuals with respect to the interaction between age and grammatical contexts, particularly in the case of possessive determiners (standard deviation: 0.71 [CI: 0.33-1.26]).

Finally, the patterns in [Figure 3](#) may also help identify individuals that seem to alter their grammatical preferences across their lifespan under certain grammatical conditions. An overview of the observed changes is presented in [Table 6](#), which may help indicate how typical certain trajectory types may be (Sankoff, 2019:223). Note that we treated the observed patterns very conservatively: despite there being more apparent patterns of change in the raw data, we retained the default assumption of stability if the 95% credible interval ranges presented in [Figure 3](#) allow for horizontal (nonascending or nondescending) trend lines.¹⁶ In total, eight individuals were flagged as altering their usage rates of *ing*-OF and *ing*-Ø. Of these eight individuals, there are five cases where the raw data suggest the possibility of a constraint change. For some individuals, such as Gilbert Burnet, the constraint change is a partial one: Burnet exhibited mixed usage but turned to categorical usage of *ing*-Ø in determinerless contexts later in life. While it is difficult to statistically verify categoricity, the rightmost (i.e., latest) mean posterior probability estimates of the relevant categories for these individuals closely approaches one.

A full constraint change may be attested for two individuals: the preferences of Peter Heylyn and William Prynne (the first and second author listed in [Figure 3](#)), born in 1599 and 1600 respectively, appear to switch radically from *ing*-OF to *ing*-Ø in possessive contexts. Given the low raw frequencies (twenty-nine tokens in total), and the possibility of more superficial, stylistic explanations, we should consider Heylyn's patterns with a healthy dose of uncertainty. Still, we could tentatively suggest the following: In his early prose and letters (before age thirty-seven), Heylyn exhibited variable use of the conservative (19%) and progressive variant (81%) in determinerless contexts. In contexts with a nominal determiner, however, Heylyn's early life grammar seemed to be largely or perhaps even entirely conservative: in the context of a possessive preceding the *ing*-form, the direct object of the *ing*-form required a periphrastic realization with *of*—conforming to the internal structure of a noun phrase. Later in life, Heylyn's writings start to allow progressive, mixed forms, and ultimately Heylyn develops a clear, near-categorical preference for the progressive form following possessives after the age of sixty. A similar picture emerges from the raw usage frequencies of William Prynne, who appeared to be categorically conservative in possessive contexts before the age of thirty-six but fully turned toward the progressive variant after the age of fifty-six.¹⁷ Thus, both Heylyn and Prynne may have categorically switched to the progressive variant when a possessive is involved around similar ages (sixty and fifty-six respectively), and times (in 1659 and 1664).

Given these results, the question arises whether the observed trends can be understood as a full constraint change, where these individuals categorically switch from

Table 6. Observed changes and metadata per author (excluding Aphra Behn and George Swinnock)

Author	Frequency change?	direction	Constraint change?	location(s)	family	university
Heylyn, Peter (1599–1662)	yes (poss)	com trend	maybe (poss; partial)	Oxfordshire, Hampshire	upper class	yes
Prynne, William (1600–1669)	yes (bare, poss & the)	com trend	maybe (poss; full)	Somerset, London	upper class	yes
Fuller, Thomas (1607–1661)	no	stable	no	<i>Various</i>	upper class	yes
Milton, John (1608–1674)	no	stable	no	London	upper class	yes
Taylor, Jeremy (1613–1667)	no	stable	no	London, Wales and Ireland	upper class	yes
L'Estrange, Roger (1616–1704)	no	stable	no	London & exile (Scotland, Europe)	upper class	yes
Boyle, Roger (1621–1679)	no	stable	no	London, Ireland	upper class	yes
Pierce, Thomas (1622–1691)	no	stable	no	Oxford and Salisbury	upper class	yes
Fox, George (1624–1691)	yes (bare, poss)	mixed	no	<i>Various</i>	craftsmen (middle)	no
Boyle, Robert (1627–1691)	no	stable	no	Ireland > London	upper class	yes
Bunyan, John (1628–1688)	yes (bare & poss)	retrograde	no	Bedford & imprisonment	craftsmen (middle)	no
Flavell, John (1630–1691)	no	stable	no	Devon	upper class	yes
Tillotson, John (1630–1694)	no	stable	no	North > London	upper class	yes
Dryden, John (1631–1700)	yes (bare)	com trend	no	Midlands > London	upper class	yes
Whitby, Daniel (1638–1726)	yes (bare & the)	com trend	maybe (bare; partial)	North > South	upper class	yes
Mather, Increase (1639–1723)	yes	com trend	maybe (bare & poss; partial)	New England > England	upper class	yes
Crouch, Nathaniel (1640–1725)	no	stable	no	London	craftsmen (middle)	no
Burnet, Gilbert (1643–1715)	yes (bare & the)	com trend	maybe (bare; partial)	Scotland > England	upper class	yes
Penn, William (1644–1718)	no	stable	no	<i>Various</i>	upper class	yes

exclusively using a conservative variant to exclusively using the progressive variant under a given grammatical condition.¹⁸ An interesting question with respect to such possible (partial or full) constraint changes is whether an author's apparent adoption of a new system (under a particular constraint) also involves a rejection of the old system, which would be even stronger evidence for a qualitative change. However, unless we find metalinguistic comments in their works (which we have not so far), this will remain an open question.¹⁹ With respect to direction, it appears that all (possible) constraint changes are in accordance with the larger population-level development (loss of *ing*-OF in bare and possessive contexts).

Discussion: inter- and intraindividual variation and change

What emerges from the results is that, while neither full nor partial inventory changes (where an individual ceases to use *ing*-OF and/or adopts *ing*-Ø) occur, we do observe a fair number of frequency and (to a lesser extent) constraint-based age effects. Additionally, the vast majority of observable quantitative shifts across all grammatical contexts appear to be in the direction of the larger population-level development, showing an increase of *ing*-Ø.

There are, however, two “retrograde” changers (Wagner & Sankoff, 2011). The first is John Bunyan, who appears to become incrementally more conservative than his peers (Age coefficient estimate: -0.54 [95% CI: -0.78 -0.32]). The second is George Fox, who similarly turns to making more conservative choices (Age coefficient estimate: -0.35 [95% CI: -0.58 -0.13]). These results corroborate earlier findings on interindividual variation in linguistic lifespan change trajectories: the frequency with which individuals use variants need not but can change, and when it does, it commonly is, but need not be, in line with the population-level development (e.g., progressive as well as retrograde change attested in panel studies by Wagner & Sankoff [2011] for inflected futures in Montreal French; Buchstaller [2015] for quotative *be like* in Tyneside English; Standing [2021:360-72] for cleft constructions in Early Modern English; also see Anthonissen [2020:247]). What is discussed somewhat less commonly than interindividual variation in lifespan trajectories is that there may also be *intraindividual* variation in this respect: for many authors where frequency-based change is attested, this change may be restricted to only a few grammatical contexts. Furthermore, an interesting example in our data set is George Fox, who becomes more conservative if the *ing*-nominal is not preceded by a determiner, while his choices following possessives likely concur with the population-level pattern.²⁰

In other words, while it may be uncommon, individuals need not consistently develop toward or away from the community trend in every respect (also see Buchstaller et al., 2021). Such intraindividual variation thus may not only be attestable across the sets of variants an individual employs (Bergs, 2005:255-56), but also across the grammatical contexts that condition a single variable set.

Confronted with these figures, the question arises whether labile linguistic behavior can somehow be linked to the contextual metainformation included in the corpus. With respect to stable speakers and speakers who adapted (slightly) in the direction of the community trend, there appear to be no striking differences in terms of social or

geographical (in)stability. By way of illustration, Tillotson and Dryden were born one year apart, belonged both to the upper class, and both went to Cambridge University in the early 1650s. After their studies, they both moved to London around the same time (1656 and 1657 respectively). Virtually all of their published work was written while they lived there. Yet, while Dryden shows a frequency (and possibly even a constraint) shift along the lines of the community trend, Tillotson is stable throughout his life. More generally, this could either mean that the differences can be explained by some unidentified factor or that some speakers are simply more likely to model their usage rates to the perceived community trend later in life (due to different “cognitive learning styles,” see for example Riding & Rayner, 1998).

Regarding the two individuals who become more conservative, George Fox and John Bunyan, there are some striking similarities between them that may go some way in explaining their retrograde behavior (see Sankoff [2018] and [2019] for an overview of studies attesting retrograde change). First, after a relatively uneventful youth spent in small countryside communities, both experienced a more diversified social life.²¹ Fox spent about a year in London at the age of nineteen, whereas Bunyan enlisted in the Parliamentary army between the age of sixteen and nineteen. This may have led them to accommodate to language use in more urban settings and establish the grammar shown in their first published work. However, some years later they were both cut off from society, being repeatedly imprisoned for their dissenting religious views and actions. Fox was in prison at least ten times (between 1649 and 1675), adding up to a total of more than six and a half years, or about 15% of his writing career. Bunyan was in prison uninterrupted (though he was allowed outside occasionally) for twelve years from 1660 until 1672, and another half year in 1676 (almost 40% of his writing career). While early modern prisons were rather “porous” and brimming with short-term inmates and visitors, normal social life was nevertheless sincerely disrupted, and prisoners were confined to their own cells in poor conditions for much of their time (Murray [2009:147, 152]). When in such relative isolation, one may fall back, due to the lack of sustained input from a more constructive social network structure, on entrenched language patterns such as they knew from youth (see the conservative behavior of the English poet Browning living in Italy [Arnaud, 1998:132-33]). Fox and Bunyan’s inmates, most of them from a lower class origin, may have been an additional conservative influence (see also the linguistic stagnation in prison documented in Baugh [1996:408-9], where this is interpreted as a case of covert prestige). The lower classes were typically less mobile and therefore less influenced by London life, which played a central role in many linguistic changes at the time, including the shift from *ing*-OF to *ing*-Ø (Nevalainen, 2015:288).²²

Besides Fox and Bunyan, there is one other speaker, William Prynne, who was incarcerated for about ten years or one fifth of his writing career. He was transported to the island of Jersey, where he was kept “incommunicado,” not even allowed a pen and paper (Cressy, 2018:741). His isolation was therefore more severe than in either Fox’s or Bunyan’s case. This period of seclusion may have slowed down his accommodation to ongoing change, similar to Fox and Bunyan. However, while he was imprisoned first around the same age (thirty-three) as Bunyan’s (thirty-two), the majority of his time in prison (seven out of ten years) occurred in the first half of

Table 7. WAIC estimates and WAIC weight (rounded to 0.0000001) for additional models, ranked by WAIC score (ascending)

Type	Parameters	p_WAIC (SE)	WAIC	SE WAIC	weight
full	age_sd + det + age_sd*det author + age_sd incarcerated	92.7 (2.5)	11809.8	156.9	0.4069601
full	age_sd + det + age_sd*det author	91.8 (2.5)	11810.6	156.9	0.3309124
full	age_sd * det + age_sd*det author + age_sd university	90.3 (2.4)	11810.7	157.0	0.2621275

his career (between 1633-1640). After 1640, Prynne remained productive and mostly free until 1668 and was generally well-connected, even becoming MP in 1648. While reconnecting to the speech community again, Prynne caught up on the ongoing changes, which might partly explain his more radical shift in preferences (in line with the community trend) in possessive contexts (also see Petré et al. [2019:110] on a similar shift in the language of Margaret Cavendish after her return from exile). Prynne's political success and general well-connectedness after his seclusion distinguishes him from Fox and Bunyan.

Finally, together with Nathaniel Crouch, Fox and Bunyan are the only authors in our sample who did not grow up in an upper-class family, instead being the sons of middle class artisans. They are also the only ones that did not attend university. Yet, what sets Fox and Bunyan apart from Crouch is that they were by far the least connected to “society” in its sense of the aggregate of fashionable, wealthy, or otherwise prominent people. Nathaniel Crouch, as a successful London-based publisher, was in the middle of buzzing London city life. By contrast, Fox spent most of his time preaching throughout the country, experiencing relatively few periods of stable participation in large-scale weak-tied urban communities such as London. Bunyan, in turn, was a poor man for most of his life and, as a result, bound to the countryside of Bedfordshire. It is interesting to note that Bunyan, who is arguably the author with the smallest social network, becomes more conservative across the board, whereas Fox may pick up at least on the innovation in possessive contexts. Yet overall, the similarity of their life paths (extended period of isolation, class background, lack of university attendance) seem to provide a plausible group of explanations for their shared retrograde behavior.

To tentatively explore whether these effects could help explain interauthor variation with respect to age effects, the selected model ($\text{ing} \sim \text{age_sd} + \text{det} + \text{age_sd} * \text{det} | \text{author}$) could be further extended with the parameters “incarcerated” (levels: yes/no, indicating whether a given author has spent more than five years imprisoned) and “university” (levels: yes/no, indicating whether the author has attended university).²³ These factors should be understood as groupings that introduce slope variance with respect to the factor age. As such, they are modeled as group-level effects with varying age slopes. The resulting models were submitted to a WAIC model comparison, presented in Table 7.²⁴

While the highest weight is assigned to the model including incarceration, it does not bring about a substantial reduction in the WAIC estimate. Considering the fact that the scores are near identical for each model, it remains undetermined whether the extended incarceration model indeed yields a better fit. We do, however, invite readers to extend our analysis to a new sample of authors that is more nicely balanced with respect to isolation versus connectedness, middle versus upper class, and/or (lack of) university education, or to explore alternative social dimensions by adapting our analysis scripts for use on the openly released data.

Conclusion

In recent years, a growing number of studies have indicated that an individual's grammar is not necessarily a fixed system and may be subject to change even after what is traditionally understood as the critical acquisition period. This has been amply shown for gradual shifts in pronunciation and morphophonology (e.g., Bowie & Yaeger-Dror, 2015; Harrington, Palethorpe, & Watson, 2000; MacKenzie, 2017; Nahkola & Saanilahti, 2004; Sankoff & Blondeau, 2007). Recent studies have moreover shown that it also applies to morphosyntactic constructions, which may be adopted (Anthonissen & Petré, 2019; Petré & Anthonissen, 2020; Petré & Van de Velde, 2018), significantly expanded (Anthonissen, 2020; Neels, 2020), or changed (Buchstaller, 2015, 2016; Standing & Petré, 2021) beyond the critical period. Similarly, lifespan change has also been attested in contexts where a choice between morphosyntactic variants needs to be made (e.g., Raumolin-Brunberg, 2005, 2009; Raumolin-Brunberg & Nurmi, 2011). While several of these changes may be regarded as stochastic shifts in preference, some changes do appear to result in a rejection of one of the variants, which could ultimately point to qualitative change, and stochastic shift also cannot explain cases where a language user starts to adopt grammatical innovations later in life. What may set such changes apart from grammatical change rooted in child language acquisition is that changes later in life may involve more (top-down) explicit learning (similar to second-language learning, see Anthonissen & Petré, 2019:3; Dąbrowska, 2015:661), but the resulting shifts in representation may arguably be just as "qualitative."

Our study corroborates earlier findings that individuals can change the rate by which they use the competing morphosyntactic structures across their lifespan (see, among others, Buchstaller, 2015; Raumolin-Brunberg, 2005, 2009; Raumolin-Brunberg & Nurmi, 2011; Wagner & Sankoff, 2011), and, for the majority of individuals, this change proceeds in the direction of the ongoing change in the rest of the community. Yet, in line with Baxter and Croft's (2016) suggestion that individual speakers may differ in their inclination to adjust their behavior to the community (also see Anthonissen, 2020), we also found nontrivial interindividual variation regarding the extent to which these usage rates change, as well as the direction in which they change (with retrograde changers such as John Bunyan opposing the general, community-directed patterns of change, which may partly be the result of their particular life trajectories).

We additionally established that there is intraindividual variation in the slopes and direction of change under different grammatical conditions. As such, we further extend Bergs' (2005:255) study on intraindividual variation, which demonstrates that

individuals need not be consistently progressive or conservative in all aspects of their grammar. While Bergs considered the choices of individuals with respect to multiple morphosyntactic variables without factoring in grammatical constraints, we suggest that, even with respect to a single variable (*ing-OF* versus *ing-Ø*), individuals can also become more progressive over time with respect to some grammatical contexts while remaining stable or even becoming more conservative in others. That the rate of change attested need not be constant across grammatical contexts in the linguistic output of individuals (as it is assumed to be at the community level, see Kroch's [1989] Constant Rate Hypothesis) does, however, call for further empirical inquiry.

Finally, we also probed into the question of whether individuals may also “participate” in qualitative change by reaching a more advanced stage of the community grammar during their lifespan. In a recent study with panel data, Buchstaller et al. (2021:32-3) suggested that “evidence of inventory change is rare and difficult to come by,” and “speakers are less likely to change the constraints that govern the use of a variable than they are to change proportional use of a variant” (also see MacKenzie, 2019). Finding evidence of full (and even partial) inventory changes in real-time corpus data may be equally difficult or even impossible, particularly with protracted changes such as the one examined in the present study (Nevalainen et al., 2011). Yet, when taking a constraint-based definition of qualitative change (which is facilitated by large-scale individual-oriented corpora such as EMMA) and identifying an appropriate time window (e.g., when the population-level patterns these individuals model themselves toward are sufficiently unstable to allow for more radical adjustments), we may in fact attest cases where individuals allow novel variants to diffuse to (and/or take over in) new grammatical contexts, albeit indeed to a lesser extent than frequency-based changes.

Notes

1. Note that structures such as *the killing an usurper* are no longer current in Present-day English (see Fanego, 2004). Furthermore, the extent to which there is distributional overlap between *ing-Ø* and *ing-OF* has been reduced substantially after the eighteenth century. See Maekelberghe (2017) for a discussion of whether the variants *ing-OF* and *ing-Ø* can still be considered a morphosyntactic alternation pair in Present-day English.
2. The higher number of authors in the second generation is due to the fact that, overall, the number of relevant tokens per author is smaller than in the first and third generation. Besides Aphra Behn, the only other woman in the corpus did not produce any relevant tokens for the selected genres.
3. The full annotated data set and R code to replicate the analysis is accessible in an online repository: https://github.com/LFonteyn/LVC_lifespan.
4. Issues related to separation in response classes have previously been described as “knockouts” (Rousseau & Sankoff, 1978; Sankoff & Labov, 1979) and “empty cells” (Tagliamonte & Baayen, 2012:24).
5. A description of why frequentist models struggle with estimating maximum likelihood values when there is (quasi-)separation can be found in Kimball et al. (2019). On convergence errors more generally, see Bates, Kliegl, Vasishth, and Baayen (2018).
6. For a very detailed tutorial on using the *brms* package, see Nalborczyk, Batailler, Løevenbruck, Vilain, and Bürkner (2019).
7. More specifically, the priors for the fixed effects were distributed around a mean of zero, with a standard deviation of one. The intercept priors were also set as normally distributed, with a mean at zero and a standard deviation of five. The varying author intercepts were modeled using a Half-Cauchy (long-tail distribution) with zero mean and a standard deviation of two.

8. The simple intercept model estimates the posterior predictive probability of the incoming variant at 0.713 [95% CI: 0.706-0.721], while the varying intercept model puts the estimate at 0.723 [95% CI: 0.661-0.776].
9. The extent of interindividual variation can also be expressed when examining the “group-level” standard deviation estimate (0.66), for which we can state with 95% confidence that this value lies between 0.47 and 0.94. The closer this value is to zero, the smaller the relevance of interindividual variation.
10. Vehtari, Gelman, and Gabry (2017) pointed out that WAIC estimates may be inaccurate with weak priors or influential observations. They recommend using Approximate Leave-One-Out cross-validation with Pareto-smoothed importance sampling (PSIS-LOO). Repeating the entire model comparison with PSIS-LOO rather than WAIC yielded similar estimates. Results of PSIS-LOO can be found and compared in the R notebook included in the online repository.
11. WAIC weights were obtained with `model_weights` (Bürkner, 2017, 2018). Further model evaluation involved examining the models’ trace plots and Rhat statistic. If the Rhat exceeds one, the algorithm has not yet converged. In such cases, more iterations are needed (possibly along with more informative priors).
12. The WAIC weight of the varying intercept model, which closely approximates one (as opposed to the WAIC weight of the simple intercept model, which approximates zero), corroborates that there are nontrivial differences between authors. Note that the weights are distributed between the models that are included in the comparison only. Neither the simple nor the varying intercept base model illustrated here will receive any weight when more complex models are included in the comparison.
13. All coefficient tables can be found at https://github.com/LFonteyn/LVC_lifespan/blob/main/Notebook_models.pdf
14. Apart from modeling lifespan trends by looking at the author’s age of production, we can also examine age effects by modeling the author’s date of birth. In EMMA, authors have been grouped into generations based on their date of birth. Adding the author’s generation category as an additional variable increases the WAIC estimate slightly (by approximately 0.2). The WAIC weight of a model that only includes age is estimated at 0.384, which is slightly higher than that of a model with age and generation added as fixed effects (0.323) and a model that includes the interaction between age and generation (0.292).
15. It should be noted that the WAIC mean estimates of the six models with the lowest WAIC estimates are very similar (see Figure A in the online repository). This indicates some uncertainty in the model comparison. Ultimately, the WAIC weight estimation seems to be in favor of the model where the interaction between age and determiner is not included as a population-level effect.
16. Crucially, failure to attest linguistic change is not the same as confirming stability—claims that involve the independence of time/age effects “are not testable and may only be assumed true” (see Paolillo, 2011:271).
17. The only late instance of the conservative variant (in 1664) is found in what turns out to be a reprint of a text first printed in 1655 and therefore should have been excluded.
18. For both Heylyn and Prynne, the lower bound of the leftmost (i.e., earliest) posterior probability estimate’s 95% credible interval approaches zero, and the upper bound of the rightmost (i.e., latest) estimate approaches one. Exact values can be consulted in the Appendix.
19. Additionally, it would be interesting to consider whether the rejection of the old system was inspired by prescriptivist comments. As pointed out by Fanego (2004), eighteenth and nineteenth century prescriptivists advised against the combination of definite articles with *ing*-Ø, possibly halting the diffusion of the innovative variant in those contexts. The developments described here predate such prescriptivist comments.
20. To be considered a retrograde trend, the lower bound of the credible interval of the leftmost (i.e., earliest) probability estimate of bare *ing*-forms must be higher than the upper bound of the rightmost (i.e., latest) probability estimate. Conversely, to be considered a progressive trend, the upper bound of the leftmost estimate of possessive *ing*-forms must be lower than the lower bound of the rightmost estimate. For bare forms, a retrograde trend can be established at the 95% credible interval (0.84-0.81) as well as the 89% credible interval (0.85-0.79). For possessive forms, the 95% CI still includes a retrograde trend (0.64-0.57), but these values should be considered unlikely. At the 89%, the bounds suggest a positive trend (0.58-0.63).
21. Drayton-in-the-clay (today known as Fenny Drayton) and Elstow respectively. Biographical data was taken from the EMMA corpus metadata (Pétré et al., 2019).
22. One other explanatory factor that has been identified with respect to retrograde change is *ascent* upon the social ladder, which has been found to correlate with more conservative behavior in Early Modern

England (particularly Nevalainen & Raumolin-Brunberg, 2016:143). While it is true that Fox and Bunyan both gradually became influencers in their own ways, the first as the leader of the Quaker movement, the latter as an influential preacher, there is no evidence that their social status went up among the establishment, nor that they had aspirations in that direction. We therefore suggest that isolation from the speech community where the change is taking place is a more likely factor and in line with what is known about the role of social networks (Sharma & Dodsworth, 2020).

23. We cannot distinguish the influence of class from the influence of university attendance, as the subset of authors who did not attend university (George Fox, John Bunyan and Nathaniel Crouch) is the same as the subset of authors with a middle-class background.

24. PSIS-LOO yielded similar estimates and weights.

References

- Anderson, Stephen R. (2016). Synchronic versus diachronic explanation and the nature of the language faculty. *Annual Review of Linguistics* 2:11–31.
- Anthonissen, Lynn. (2020). *Special Passives across the Lifespan: Cognitive and Social Mechanisms*. Doctoral dissertation, University of Antwerp and Ludwig-Maximilians-Universität München.
- Anthonissen, Lynn & Peter Petré. (2019). Grammaticalization and the linguistic individual: new avenues in lifespan research. *Linguistics Vanguard* 5:1–12.
- Arnaud, René. (1998). The development of the progressive in 19th century English: A quantitative survey. *Language Variation and Change* 10:123–52.
- Barlow, Michael. (2013). Individual differences and usage-based grammar. *International Journal of Corpus Linguistics* 18:443–78.
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. (2018). Parsimonious Mixed Models. <https://arxiv.org/pdf/1506.04967.pdf>
- Baugh, John. (1996). Dimensions of a theory of econolinguistics. In G. Guy, F. Crawford, D. Schiffrin & J. Baugh (Eds.), *Towards a social science of language, vol. I*. Amsterdam: John Benjamins. 397–419.
- Baxter, Gareth & William Croft. (2016). Modeling language change across the lifespan: Individual trajectories in community change. *Language Variation and Change* 28:129–73.
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. (2009). Language is a complex adaptive system. *Language Learning* 59:1–26.
- Bergs, Alexander. (2005). *Social networks and historical socio-linguistics: studies in morphosyntactic variation in the Paston Letters (1421–1503)*. Berlin: Mouton De Gruyter.
- Bowie, David & Malcah Yaeger-Dror. (2015). Phonological change in real time. In P. Honeybone & J. Salmons (Eds.), *The Oxford handbook of historical phonology*. Oxford: Oxford University Press. 603–18.
- Buchstaller, Isabelle. (2015). Exploring linguistic malleability across the life span: Age-specific patterns in quotative use. *Language in Society* 44:457–96.
- Buchstaller, Isabelle. (2016). Investigating the effect of socio-cognitive salience and speaker-based factors in morpho-syntactic life-span change. *Journal of English Linguistics* 44:199–229.
- Buchstaller, Isabelle, Anne Krause-Lerche & Johanna Mechler. (2021). Exploring the effect of linguistic architecture and heuristic method in panel analysis. In K. Beaman & I. Buchstaller (Eds.), *Language variation and language change across the lifespan: Theoretical and Empirical Perspectives from Panel Studies*. New York: Routledge. 185–208.
- Bürkner, Paul-Christian. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80(1):1–28.
- Bürkner, Paul-Christian. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 10(1):395–411.
- Bybee, Joan L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language* 82: 711–33.
- Cressy, David. (2018). Puritan Martyrs in Island Prisons. *Journal of British Studies* 57:736–754.
- Croft, William. (2000). *Explaining Language Change: An Evolutionary Approach*, 2nd edn. Oxford: Oxford University Press.

- Dąbrowska, Ewa. (2015). Individual differences in grammatical knowledge. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics*. Berlin: De Gruyter Mouton. 649–67.
- Denison, David. (2002). Log(istic) and simplistic S-curves. In R. Hickey (Ed.), *Motives for Language Change*. Cambridge: Cambridge University Press. 54–70.
- De Smet, Hendrik. (2008). Functional motivations in the development of nominal and verbal gerunds in Middle and Early Modern English. *English Language and Linguistics* 12:55–102.
- De Smet, Hendrik. (2016). How gradual change progresses: The interaction between convention and innovation. *Language Variation and Change* 28:83–102.
- Diessel, Holger. (2019). *The Grammar Network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.
- Fanego, Teresa. (2004). On reanalysis and actualization in syntactic change: The rise and development of English verbal gerunds. *Diachronica* 21:5–55.
- Gelman, Andrew, Hwang, Jessica & Aki Vethari. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24:997–1016.
- Harrington, Jonathan, Sallyanne Palethorpe & Catherine I. Watson. (2000). Does the Queen speak the Queen's English? *Nature* 408:927–28.
- Houston, Ann. (1989). The English gerund: syntactic change and discourse function. In R. Fasold & D. Schiffrin (Eds.), *Language change and variation*. Amsterdam: John Benjamins. 173–96.
- Kerswill, Paul. (1996). Children, adolescents, and language change. *Language Variation and Change* 8:177–202.
- Kimball, Amelia E., Kailen Shantz, Christopher Eager & Joseph Roy. (2019). Confronting Quasi-Separation in Logistic Mixed Effects for Linguistic Data: A Bayesian Approach. *Journal of Quantitative Linguistics* 26:231–55.
- Kroch, Anthony. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.
- Labov, William. (1994). *Principles of linguistic change: Internal factors*. Vol 1. Oxford: Blackwell.
- Labov, William. (2001). *Principles of Linguistic Change: Social Factors*. Vol. 2. Malden and Oxford: Wiley-Blackwell.
- Lightfoot, David. (2019). Chomsky's I-languages: Rethinking catastrophic changes. *Acta Linguistica Academica* 66:349–69.
- López-Couso, María José. (2017). Transferring insights from child language acquisition to diachronic change (and vice versa). In M. Hundt, S. Mollin & S. E. Pfenninger (Eds.), *The changing English language: Psycholinguistic perspectives*. Cambridge: Cambridge University Press. 332–47.
- MacKenzie, Laurel. (2017). Frequency effects over the lifespan: A case study of Attenborough's r's. *Linguistics Vanguard* 3:1–12.
- MacKenzie, Laurel. (2019). Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation. *Glossa* 4:1–23.
- Maelkelberghe, Charlotte. (2017). *The use of nominal and verbal gerunds in Present-day English*. A multifunctional comparative analysis. Doctoral dissertation, KU Leuven.
- Meisel, Jurgen, Martin Elsig & Esther Rinke. (2013). *Language acquisition and change: A morphosyntactic perspective*. Edinburgh: Edinburgh University Press.
- Murray, Molly. (2009). Measured Sentences: Forming Literature in the Early Modern Prison. *Huntington Library Quarterly* 72:147–67.
- Nahkola, Kari & Marja Saanilahti. (2004). Mapping language changes in real time: A panel study on Finnish. *Language Variation and Change* 16:75–92.
- Nalborczyk, Ladislav, Cédric Batailler, Hélène Løvenbruck, Anne Vilain & Paul-Christian Bürkner. (2019). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research* 62:1225–42.
- Neels, Jakob. (2020). Lifespan change in grammaticalisation as frequency-sensitive automation: William Faulkner and the let alone construction. *Cognitive Linguistics* 31:339–65.
- Nevalainen, Terttu, Helena Ramoulin-Brunberg & Heikki Manilla. (2011). The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change* 23:1–43.
- Nevalainen, Terttu. (2015). Social networks and language change in Tudor and Stuart London-only connect? *English Language and Linguistics* 19:269–92.

- Nevalainen, Terttu & Helena Raumolin-Brunberg. (2016). *Historical sociolinguistics. Language change in Tudor and Stuart England*, 2nd edn. London & New York: Routledge.
- Paolillo, John. (2011). Independence claims in linguistics. *Language Variation and Change* 23:257–74.
- Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique Manjavacas, Emma-Louise Silva, William Standing & Odile A.O. Strik. (2019). Early Modern Multiloquent Authors (EMMA): Designing a large-scale corpus of individuals' languages. *ICAME Journal* 43:83–122.
- Petré, Peter & Lynn Anthonissen. (2020). Individuality in complex systems: A constructionist approach. *Cognitive Linguistics* 31:185–212.
- Petré, Peter & Freek Van de Velde. (2018). The real-time dynamics of the individual and the community in grammaticalization. *Language* 94:867–901.
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Raumolin-Brunberg, Helena. (2005). Language change in adulthood: Historical letters as evidence. *European Journal of English Studies* 9:37–51.
- Raumolin-Brunberg, Helena. (2009). Lifespan changes in the language of three early modern gentlemen. In A. Nurmi, M. Nevala & M. Palander-Collin (Eds.), *The Language of Daily Life in England (1400–1800)*. Amsterdam: John Benjamins. 165–96.
- Raumolin-Brunberg, Helena & Arja Nurmi. (2011). Grammaticalization and language change in the individual. In H. Narrog & B. Heine (Eds.), *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press. 251–62.
- Riding, Richard & Stephen Rayner. (1998). *Cognitive styles and learning strategies: Understanding style differences in learning and behavior*. London: Fulton.
- Rousseau, Pascale & David Sankoff. (1978). A solution to the problem of grouping speakers. In D. Sankoff (Ed.), *Linguistic variation: Models and methods*, 97–117. New York: Academic Press.
- Rutten, Gijsbert & Marijke van der Wal. (2014). Social and constructional diffusion: Relative clauses in seventeenth- and eighteenth-century Dutch. In R. Boogaert, T. Colleman & G. Rutten (Eds.), *Extending the scope of construction grammar*, 181–206. Berlin: De Gruyter.
- Sankoff, David & William Labov. (1979). On the uses of variable rules. *Language in Society* 8(2–3): 189–222.
- Sankoff, Gillian & Hélène Blondeau. (2007). Language change across the lifespan: /r/ in Montreal French. *Language* 83:560–88.
- Sankoff, Gillian. (2018). Language change across the lifespan. *Annual Review of Linguistics* 4:297–316.
- Sankoff, Gillian. (2019). Language change across the lifespan: three trajectory types. *Language* 95(2): 197–229.
- Sharma, Devyani, Joan Bresnan, & Ashwini Deo. (2008). Variation and change in the individual: Evidence from the Survey of English Dialects. In R. Cooper & R. Kempson (Eds.), *Language in flux: Dialogue coordination, language variation, change and evolution*. London: College Publications. 265–321.
- Sharma, Devyani & Robin Dodsworth. (2020). Language variation and social networks. *Annual Review of Linguistics* 6:341–61.
- Smirnova, Elena. (2015). Constructionalization and constructional change: the role of context in the development of constructions. In J. Barddal, E. Smirnova, L. Sommerer & S. Gildea (Eds.), *Diachronic Construction Grammar*. Amsterdam: John Benjamins. 81–106.
- Standing, William. (2021). *It-cleft change across the lifespan in Early Modern Authors*. Doctoral dissertation, University of Antwerp.
- Standing, William & Peter Petré. (2021). Exploiting convention: Lifespan change and generational incrementation in the development of cleft constructions. In K. Beaman & I. Buchstaller (Eds.), *Language variation and language change across the lifespan: Theoretical and Empirical Perspectives from Panel Studies*. New York: Routledge. 141–63.
- Tagliamonte, Sali A. & Alexandra D'Arcy. (2007). Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change* 19:199–217.
- Tagliamonte, Sali A. & Harald Baayen. (2012). Models, forests and trees of York English: *was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2):135–78.
- Tagliamonte, Sali A. & Derek, Denis. (2016). The changing FUTURE: competition, specialization and reorganization in the contemporary English future temporal reference system. *English Language and Linguistics* 22:403–30.

- Traugott, Elisabeth C. & Graeme Trousdale. (2013). *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Vehtari, Aki, Gelman, Andrew & Jonah Gabry. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27:1413–32.
- Wagner, Suzanne Evans, & Gillian Sankoff. (2011). Age grading in the Montréal French inflected future. *Language Variation and Change* 23:275–313.
- Watanabe, Sumio. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research* 11:3571–94.

Cite this article: Fonteyn L, Petré P (2022). On the probability and direction of morphosyntactic lifespan change. *Language Variation and Change* 34, 79–105. <https://doi.org/10.1017/S0954394522000011>