

Chapter 6

Chance

[T]o my imagination it is far more satisfactory to look at such instincts as...consequences of one general law leading to the advancement of all organic beings – namely, multiply, vary, let the strongest live and the weakest die.

–Charles Darwin (1859)

In our analyses of historical change we have used mean dynamics that presuppose effectively infinite populations. This assumption allows us to smooth out chance occurrences and study the expected motion of what is arguably a stochastic process. However, as we noted in the previous chapter, this does not rule out the possibility that the changes we observe are actually due to chance. Here we relax the assumption of effectively infinite populations and determine its effect on the acquisition dynamics. While the acquisition dynamics predict stability, it is only a weak kind of stability stemming from the size of the population. Either of the transitions of the formal cycle could be the result of random sampling errors in a finite rather than an infinite population. This is particularly relevant for the second transition from *ne...not* to *not*, which we noted cannot be explained by either use or acquisition. In this chapter we investigate the possibility that the transitions of the

formal cycle are indeed due to chance.

First, we briefly discuss the conceptual role of drift in the history of population genetics. Whereas early approaches to population genetics from Darwin on have emphasized selection over drift, more recent work has developed tools for addressing both theoretical possibilities. Second, we demonstrate the dynamics of drift in finite populations. We show that change can indeed come about through drift. More importantly, we show that drift can lead to change that exhibits the qualitative trajectories so frequently observed in historical linguistics, and selection can lead to change that is decidedly not like what we observe historically. Third, we introduce a statistical method for testing whether we can reject the hypothesis of drift in favor of selection for a particular trajectory. We apply this test to both of the transitions of the formal cycle and show that while we can reject the possibility of drift in the transition from *ne* to *ne...not*, we cannot reject it in the transition from *ne...not* to *not*. Finally, we discuss these results in light of the previous chapters. Given that we can reject the role of drift in the first transition, this adds support for our model of the functional cycle. Given that we cannot reject the role of drift in the second transition, we can make sense of the varying times to completion of the transition across languages.

The main contributions of this chapter are twofold. First, we offer the first application of statistical methods for distinguishing between selection and drift in linguistic time series. This has important implications not just for the formal and functional cycles, but for linguistic change as a whole. The crucial fact is that the typical trajectories observed in linguistic change may arise from selection or drift. Simply put, we cannot distinguish between these two possibilities from simple visual examination. Rather, we need a means of testing competing hypothesis about the underlying cause of the change. Second, the application of these methods offers insight into the formal cycle, and the second transition in particular given that neither use nor acquisition offer an explanation. More broadly, it also offers constraints on the nature of stable variation.

6.1 Drift

The nineteenth century saw both the discovery of the mechanics of genetic inheritance by Gregor Mendel and the formulation of the principle of natural selection by Charles Darwin. Yet, it was not until the early twentieth century that these two notions were reconciled and put on a rigorous mathematical foundation by the work of Fisher, Sewall Wright, and Haldane in what has become known as the modern *evolutionary synthesis* (Huxley 1942). As the term suggests, these foundations offer a coherent and compelling view of observed changes in biological populations.

The role of random drift in explaining these changes, however, was taken to be minimal in comparison to selection. The balance between selection and drift was revisited most notably by Kimura (1968) in his *neutral theory* of evolution, which emphasized the role of drift rather than selection at the molecular level. While the proper emphasis on of each has been the subject of intense debate, more recent work in population genetics has taken the balance between these two forces as an empirical matter, developing theoretical tools for distinguishing the potential role of each. In what follows, we adopt this approach to understanding changes in a population over time.

In particular, our goal will be to test the role of drift in both of the transitions of the formal cycle. Our first step towards this approach is to demonstrate the need for statistical methods for distinguishing the quantitative signatures of selection versus drift in linguistic time series. We do so by simulating the dynamics of change in a finite population to demonstrate some counter-intuitive possibilities regarding drift versus selection.

6.2 Dynamics

Here we outline the dynamics of a simple model of change in a finite population. Using simulations of the *Moran model* (Moran 1958), we show that our intuitions about how drift

and selection look are not as useful as we might think. That is, we cannot simply look at the trajectory of a change and determine intuitively if it occurred due to selection or drift. This is true despite the fact that we often observe a similar qualitative *S*-shaped trajectory in the course of language change (Bailey 1973).

The *Moran model* describes the dynamics of selection and drift in a finite population. At each point in time one individual is chosen to reproduce and one individual is chosen to die, yielding a continuous-time Markov chain.¹ The probability that a variant is chosen to reproduce is proportional to its relative fitness. For example, consider a population composed of a particular number of two variants, $N = n_1 + n_2$. Let the selective advantage of the second variant be s , then the probabilities of each being selected to reproduce are given by the following.

$$p_1^{birth} = \frac{n_1}{n_1 + n_2(1 + s)} \quad (6.1)$$

$$p_2^{birth} = \frac{n_2(1 + s)}{n_1 + n_2(1 + s)} \quad (6.2)$$

For the case where the two variants are selectively neutral, $s = 0$, the two variants are chosen directly proportional to their respective frequencies in the population. Where $s > 0$, the second variant is selected for, so the second variant is slightly more likely to be chosen. This makes sense, if one variant has a selective advantage, then we would expect it to be more likely to reproduce.

While the probability of being chosen to reproduce is proportion to the relative fitness of the two variants, the probability of being chosen to die is directly proportional to just the

¹While the Moran model assumes continuously overlapping generations as we did in the previous chapter regarding syntactic acquisition, the *Wright-Fisher model* can be taken as the discrete-time analogue where generations do not overlap and are sampled all at once. The same general point holds for the Wright-Fisher process as well though: we cannot determine the selective advantage of particular variants from the trajectory they create.

prevalence of the two variants in the population.

$$p_1^{death} = \frac{n_1}{n_1 + n_2} \quad (6.3)$$

$$p_2^{death} = \frac{n_2}{n_1 + n_2} \quad (6.4)$$

Given that a single individual is chosen for birth and death at each point in time, then the number of any variant either increases, decreases, or stays the same. For example, we can keep track of the number of the second variant in the population., n_2 . This number will go up by one if an individual of the second variant is chosen to reproduce and an individual of the first variant is chosen to die. If the selections are the opposite, where an individual of the first variant is chosen to reproduce and an individual of the second variant is chosen to die, then the number of the second variant will decrease by one in the population. If the type of both individuals selected is the same, then there will not be any change in the number of either variants. The probability of all these outcomes is determined by the probability of selection for birth and death that we listed above.

In fact, we can specify the probability of each of these changes from one point in time to the next. In particular, we can show the probability of a particular change in the number of the second variant in the population. We list the probability that n_2 will increase, decrease, or stay the same.

$$p_{n_2, n_2+1} = p_2^{birth} p_1^{death} = \frac{n_2(1+s)}{n_1 + n_2(1+s)} \frac{n_1}{n_1 + n_2} \quad (6.5)$$

$$p_{n_2, n_2-1} = p_1^{birth} p_2^{death} = \frac{n_1}{n_1 + n_2(1+s)} \frac{n_2}{n_1 + n_2} \quad (6.6)$$

$$p_{n_2, n_2} = 1 - p_{n_2, n_2+1} - p_{n_2, n_2-1} \quad (6.7)$$

There are two important points where the population does not change at any subsequent points in time. Namely, if $n_2 = 0$ or $n_2 = N$, then the population will not change at any point moving forward. These are referred to as the *absorbing states* of the process, whereas all other states are *transient*. This follows from the fact that at the absorbing states the sampling probabilities for the two variants are either one or zero.

So, neutral selection in a finite population provides an interesting contrast to the acquisition dynamics we presented in the previous chapter. Indeed, if we loosen the assumption of an infinite population, then the acquisition dynamics makes a similarly strong prediction of no stable variation between grammars that are totally mutually incompatible. In other words, if acquisition is the only force acting on a language, then a single syntactic position will only ever be expressed by a single variant. We return to the implications of this prediction for observed stable variation below.

Now that we have specified the dynamics of the model, we can simulate trajectories of a population of size N for selection coefficient s . Before doing so though, it is useful to consider what our expectations would be about a population evolving under neutral versus positive selection. Intuitively, we would expect selection to exhibit a clear trajectory, where the variant that is being selected against is quickly driven from the population. In contrast, we would expect neutral selection to be a more random kind of change, not necessarily tending in direction or the other. That is, we would not expect any clear trajectory favoring one variant over the other. It is useful to dwell on these expectations a bit before considering the simulation results. We want a clear baseline to compare the results to.

When we do compare individual trajectories of the model for cases where $s = 0$ versus $s > 0$, we sometimes have the exact opposite results, as can be seen in Figures 6.1 and 6.2.

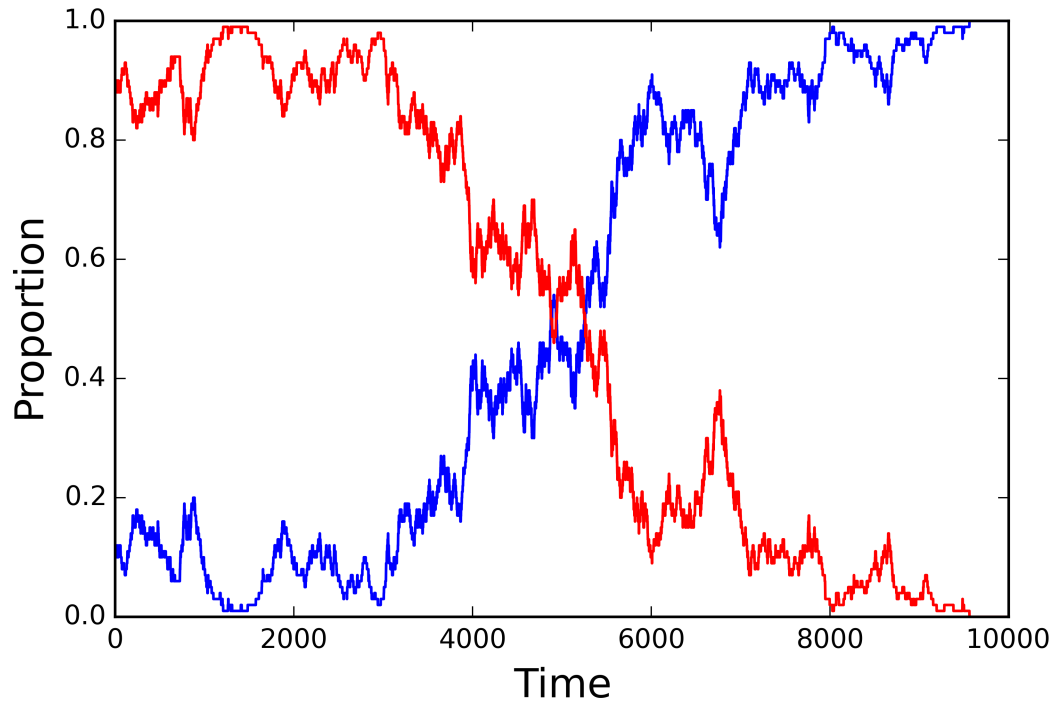


Figure 6.1: Proportion of forms over time in simulation of Moran model for $N = 100$ and $s = 0$

That is, in Figure 6.1 we show the trajectory of a population under neutral selection $s = 0$ that exhibits the characteristic *S*-shaped curve observed in language change. Likewise, in Figure 6.2 we show the trajectory of a population under positive selection $s = .1$ that exhibits a rather different kind of growth. The fundamental fact is that the underlying cause of a particular change cannot be read off its form, any particular *S*-shaped curve may be due to drift or selection.

Now, one objection to this point might be that the individual trajectories presented in Figures 6.1 and 6.2 are atypical. More often than not, selection and drift will match our expectations. This is certainly true, these trajectories were chosen particularly to emphasize the potentially counterintuitive result of selection and drift in finite populations. So, we might be tempted to conclude that the probability that a given change is due to drift given

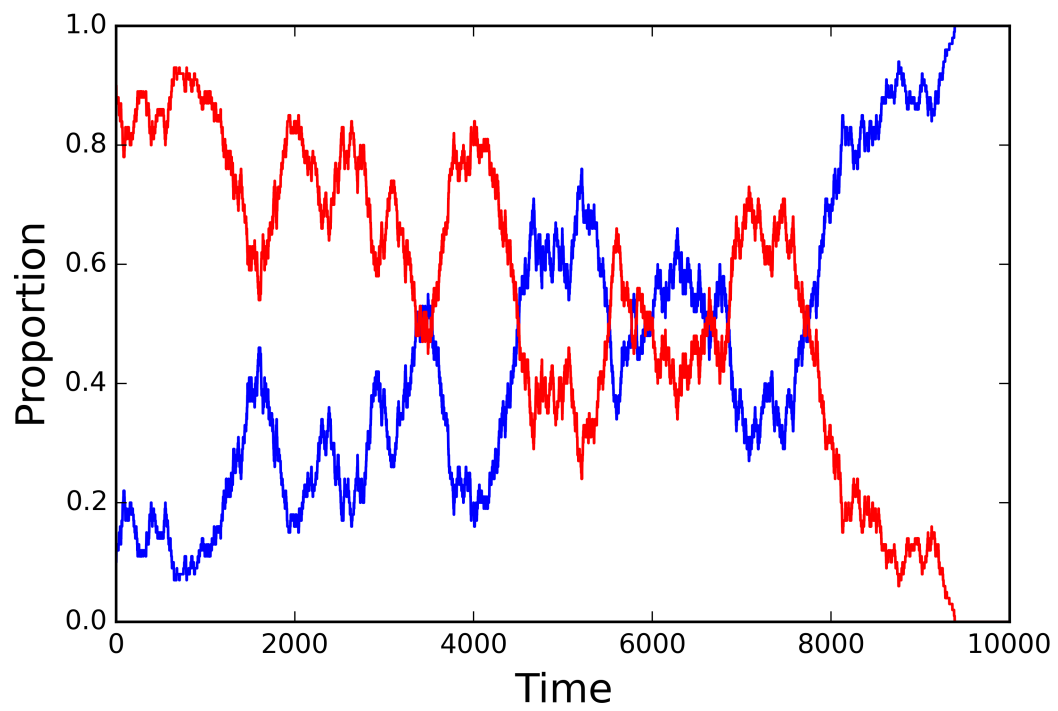


Figure 6.2: Proportion of forms over time in simulation of Moran model for $N = 100$ and $s = .1$

that it is *S*-shaped is actually fairly small.

However, this objection rests on several assumptions. First, it assumes that we have some prior expectation over drift versus selection. But, we do not have sufficiently developed causal models of linguistic change that would warrant any particular prior. In fact, in particular cases, we have very strong theoretical reasons for expecting drift, as was shown in the previous chapter regarding the transitions of the formal cycle. Second, the size of the population is not a parameter we know beforehand. That is, we can never be sure from the data itself what the probability of a particular curve should be under drift. Third, and perhaps most importantly, it is not clear what actually counts as being *S*-shaped or not. While we may have intuitions about what does or does not count, these would need to be clarified.

If all of these assumptions cannot be clearly articulated and justified, then it is reasonable to assume that we simply cannot know the cause of a particular change given its shape. If this is indeed the case, then we need some means of testing different hypotheses about the underlying causes of change. For the formal cycle, we want a means of testing for whether or not each of the transitions is due to drift or selection.

6.3 Modeling

If drift and selection can yield counter-intuitive trajectories in finite populations, then we need some means of distinguishing the two in linguistic time series. In what follows, we discuss the *fitness increment test* described by Feder et al. (2014) as a means for testing the hypotheses of drift versus selection. First, we begin by noting the motivation for the test, as well as some of details regarding its application. Second, we apply the test to the two transitions of the formal cycle.

The fundamental comparison to be made in distinguishing between drift and selection

is between two models. The first model assumes that there is no selection $s = 0$ and determines the population size N that would best explain the data. The second model assumes that there is selection $s > 0$ determines the selection coefficient and population size N that would best explain the data. Given that the first model can be taken as a special case of the second, the two can be compared using a likelihood ratio test. In this case, drift is our null hypothesis and selection is our alternative hypothesis.

However, Feder et al. (2014) note that doing so is not without complications. First, even using standard approximations to the Moran process (Kimura 1955a,b, Ewens 2012), calculating the likelihoods necessary to compare the two models is computationally intensive. Second, even if these values were simple to obtain, the relevant test statistic is not χ^2 distributed as is often assumed (cf. Wilks 1938). In fact, using the χ^2 distribution systematically underestimates the false positive rate, meaning that if we assume the test-statistic is χ^2 distributed, then we will reject the null hypothesis of drift more often than we would like to when it is true.

To address these problems Feder et al. (2014) propose an approximation to the Moran process that consists of the combination of a deterministic logistic process and a Gaussian noise process.² Under this approximation, the changes in the number of different forms over time have certain properties. Suppose we have measurements from several points in time of the two variants. Let p_i be the proportion of the second variant in the population at time t_i . We are interested in how these proportions change over time, so will look at the differences in proportion between different points of time, $p_i - p_{i-1}$. In particular, we rescale these *fitness increments* in the following manner.

$$Y_i = \frac{p_i - p_{i-1}}{\sqrt{2p_{i-1}(1 - p_{i-1})(t_i - t_{i-1})}} \quad (6.8)$$

²See Feder et al. (2014:521-522) for the mathematical details. Note that while this approximation technically only holds for the Moran process, in practice it works well for the Wright-Fisher process as well.

For the Gaussian approximation, under the null hypothesis of drift these scaled fitness increments are independent and approximately normally distributed around zero with a variance inversely proportional to the size of the population. In contrast, under the alternative hypothesis the increments are independent and approximately normally distributed, but with a non-zero mean and a different variance. For our purposes, we want to know whether the mean of the scaled fitness increments is greater than zero. That is, we want to know if the incoming variant has some selective advantage. This *fitness increment test* can be accomplished using a one-tailed *t-test*. So, under the Gaussian approximation, all we have to do to do is rescale the fitness increments and test if their mean is greater than zero.

Returning to the formal cycle, we are interested in determining whether we have sufficient evidence in favor of selection to reject the null hypothesis of drift in either of the two transitions. That is, we want to know not just whether a model with additional parameters fits the data better, but if it fits the data sufficiently better for us to reject the null hypothesis of drift. In the case of the formal cycle, we want to know whether we can reject the role of drift in the transitions from from *ne* to *ne...not* and from *ne...not* to *not*.

The overall trajectory of these transitions in Middle English is shown in Figure 6.3. However, as we noted in the previous chapter, we are really interested in two independent changes at different locations in the negative phrase. In what follows, we will treat the two transitions independently. The data relevant to the first transition is shown in Figure 6.4, where the contrast between *ne* versus *ne...not* and *not* indicates potential competition for what expresses the specifier of the negative phrase. The data relevant to the second transition are shown in Figure 6.5, where the contrast between *not* versus *ne* and *ne...not* indicates potential competition for what expresses the head of the negative phrase.

For each of the transitions we need to decide on how to group the data together to calculate the fitness increments and perform the fitness increment test. There are two general considerations in doing so. The first consideration is that we need to make sure the assump-

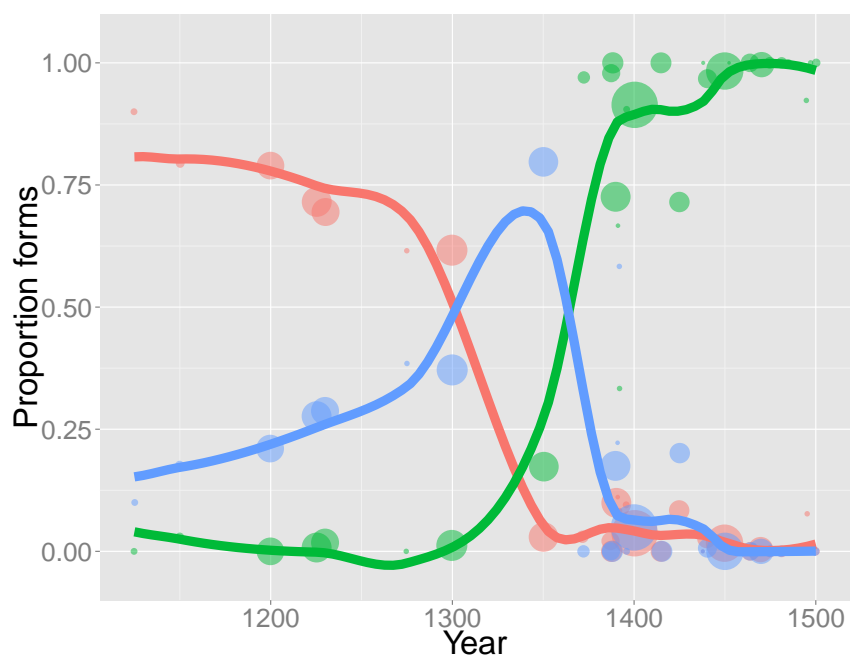


Figure 6.3: Proportion of forms of negation in Negative Declaratives

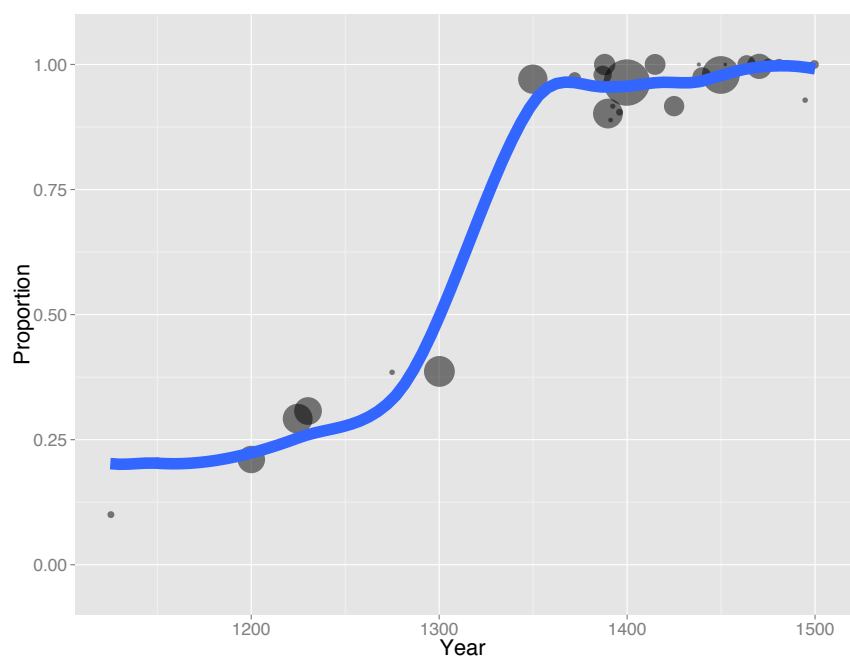


Figure 6.4: Proportion of *ne...not* and *not* versus *ne* over time

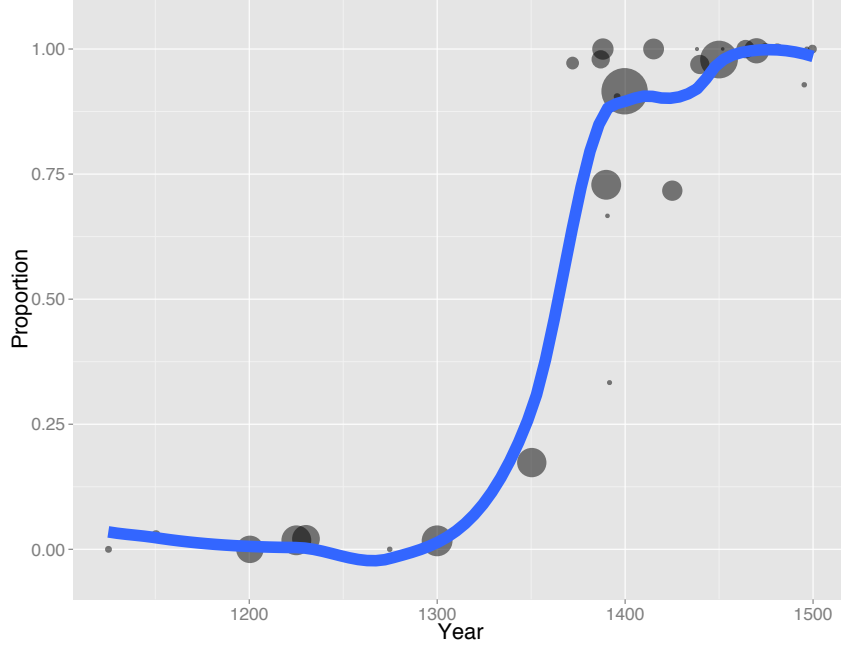


Figure 6.5: Proportion of *not* versus *ne* and *ne...not* over time

tions of the underlying Gaussian approximation are met. In particular, the approximation cannot deal with absorption events before the last bin. If such an absorption event occurred, then all subsequent bins should be the same. So, we need to make sure there are no bins before the last one where only one variant is present. This basically serves as a limit on how finely we can bin the data.

The second consideration is that we want the test to have as much statistical power as possible, allowing us to reject the null hypothesis when it is indeed false. In practice, the fitness increment test shows reasonable power when the increments include around a thousand data points. To guarantee that bins have approximately equal numbers of data points we bin by quantiles. This allows for the width of bins to vary, and we treat the midpoint of each bin as the time of the sample measurement. Note that the scaling of the fitness increments takes these time differences into account as well.

So, for each transition we do three things. First, we bin the data by quantiles and perform the fitness increment test on the resulting bins. In what follows we use the finest

partition of the data that give us almost one thousand data points per bin but meets the assumptions of the Gaussian approximation. Namely, there are no absorption events and the fitness increments are normally distributed. Second, we perform the fitness increment test on the rescaled fitness increments to determine if we can reject the null hypothesis of drift. Third, if we can reject the null hypothesis then we numerically estimate the mostly likely selection coefficient and population size under the alternative hypothesis of selection; if we cannot reject the null hypothesis of drift then we numerically estimate the most likely population size without selection.

For the first transition we bin the data into six bins and confirm that there are no absorption events prior to the last bin and that the rescaled fitness increments are approximately normal according to the *Shapiro-Wilk test* ($W = , p = 0.2050$). Given that these conditions are met, we perform the fitness increment test and note that we can reject the null hypothesis of drift ($\bar{Y} = 0.0278, t(5) = 2.6394, p = 0.0288$). The selection coefficient and population size that best explain the data under the alternative hypothesis of selection can be numerically inferred as $\hat{s} = 0.01913$ and $\hat{N} = 15900$.³ While the selection coefficient from fitting the acquisition dynamics to the first transition are not directly comparable, it is interesting to note that in a finite population the selection coefficient is about five times smaller. This is a potentially important point to keep in mind when fitting deterministic mean dynamics to data.

For the second transition we bin the data into five bins and confirm that there are no absorption events prior to the last bin and that the rescaled fitness increments are approximately normal according to the *Shapiro-Wilk test* ($W = , p = 0.1300$). Given that these conditions are met, we perform the fitness increment test and note that we cannot reject the

³Many thanks to Mitchell Johnson for help installing and understanding the code, which can be found at <https://github.com/mnewberry/tsinfer>. Technically speaking, the population parameter reported here conflates another nuisance parameter regarding the time scale of the change. We leave interpreting the implications of the population parameter for future research.

null hypothesis of drift ($\bar{Y} = 0.0624$, $t(5) = 1.7021$, $p = 0.0820$). The population size under the null hypothesis of drift that best explains the data can be inferred numerically as $\hat{N} = 506$. We cannot reject the null hypothesis, so it makes sense that the population would have to be fairly small to account for the rapid change in the second transition.

So, we can reject the null hypothesis of drift in the first transition, but not in the case of the second transition. This first result makes sense. given the model of the functional cycle presented in Chapter 4. That is, the first transition of the formal cycle cannot be explained by random drift in syntactic acquisition, but it can be explained as an instance of the functional cycle due to pragmatic pressures. This second result interesting given that the second transition is the more dramatic of the changes, however, there two important things to note.

First, we have not shown that the second transition is due to drift. Rather, we have simply shown that we cannot reject the possibility. This could be due to the fact that the second transition is actually due to drift, or that we have simply failed to reject the null hypothesis even though it is not true. In this regard, we should note that the fitness increment test actually loses the power to detect selection under certain circumstances when selection is particularly strong (Feder et al. 2014:Figure 2 and 514-515). However, the power of the test depends on the number of data points per bin and the length of the time series in relation to the size of the population in complicated way. Determining whether the failure to reject the null hypothesis for the second transition is due to this is something we leave for future research.⁴ For now though, given that we have no other alternative explanations to put forward, we lose nothing by simply noting that the second transition is consistent with random drift in syntactic acquisition.

Second, if the failure to reject the null hypothesis is indeed because the second tran-

⁴Although, these results do not change if we alter the number of bins that we divided the data into. We can never reject the null hypothesis. See Appendix D for the details.

sition is actually due to drift, then this offers an interesting explanation for the differing time courses across languages. For example, while the second transition quickly follows the first in the history of Middle English Wallage (2008) as well as Middle High German Jäger (2008). However, even in Middle Low German, where the second transition does go to completion we cannot reject drift using the data cited in Breitbarth (2009:109) ($\bar{Y} = 0.02644$, $t(3) = 0.8011$, $p = 0.2408$). Moreover, the second transition can take several hundred years, as is the case in the history of French (Martineau and Mougeon 2003) and Dutch (Burridge 1993). Indeed, some Flemish dialects still retain the embracing form (van der Auwera and Neuckermans 1997, Zeijlstra 2004). The fact that languages differ so widely in the amount of time spent between the first and the second transition has often been taken as a puzzling. However, when the second transition is viewed as a stochastic process, the varying amount of time makes perfect sense.

Now, claiming that the second transition in all of these languages may be due to drift obscures quite a bit of linguistic detail. But, it offers both theoretical and empirical leverage. For example, if we can rule out drift quantitatively in the second transition in a particular language, then we have a compelling reason to dive into theoretical analysis of the change in question. Ultimately, such an analysis should yield explanations of the observed change similar in form to our model of the functional cycle in Chapter 4 or the acquisition dynamics we presented in Chapter 5. From the other direction, if we have compelling theoretical reasons to believe that the second transition in a language happened for a particular reason, and we can build a model that explains its trajectory, but we cannot reject the null hypothesis of drift, then this shows us some of the limitations of the quantitative methods we have applied here.

At worst then, claiming that the second transition of the formal cycle may be due to drift is both empirically plausible and scientifically useful. Absent a better explanation, it makes sense of what has often been taken as a puzzling fact. More importantly, it clarifies

what would be needed to provide a better explanation.

Summary

In this chapter we applied statistical methods developed in population genetics to test the hypothesis of selection and drift in both transitions of the formal cycle in Middle English. We found that we could reject the null hypothesis of drift in the first but not the second transition. The result for the first transition makes sense. In fact, given the explanation of the formal cycle we offered in Chapter 4, we would be surprised if we could not reject drift in the first transition of the formal cycle. The result for the second transition also makes sense. The model we presented in Chapter 4 does not offer an explanation of the second transition. Moreover, given the syntactic structures posited to underly the formal cycle, the acquisition dynamics presented in Chapter 5 do not offer an explanation either. If neither acquisition nor use can explain the transition, then it might just be due to random drift.

Beyond the formal cycle, these results have important implications for stable linguistic variation in general. The acquisition dynamics allow for stable variation only when grammars are perfectly mutually incompatible. That is, when $s = 0$, the two grammars are perfectly balanced and will not change over time. However, if we relax the assumption of an effectively infinite population that underlies this mean dynamics, we know that one or the other variant will win it. This follows from the fact that the absorbing states of the stochastic process occur where only one or the other variant is used. So, if acquisition is the only force acting on forms over time, then we will not observe linguistic variation at a particular level. Namely, there can never be variation between how a particular syntactic location is expressed.

The fact that we do observe stable variation suggests that there are other forces acting over the distribution of forms in the linguistic environment. Indeed, in some cases stable

variation has been observed over centuries, as is the case for the apical and velar variants of (ING) (Labov 1994): *hunting* versus *huntin'*. If not for some countervailing force, this variation should have arguably been extinguished, if these variants differ only in how they express some progressive aspectual head in the sense of *distributed morphology* (Embick and Noyer 2007).

Broadly speaking, it would seem that the most likely candidate for a countervailing force to the monomorphic effect of syntactic acquisition is meaning, which we take to include semantic, pragmatic, and sociostylistic information. That is, the stability of (ING) arguably stems from the fact that it signals style: *hunting* is formal whereas *huntin'* is not. So, while two variants may compete with each other to express the same syntactic position, they might stably coexist if they find complementary informational niches. This is exactly what we see with the specialization of the two (ING) variants to particular styles. As a slogan then, we might say: no stable variation without information. Note that this does not mean that stable informational variation is inherently stable. However, it would seem that it is more likely that two meanings might specialize rather than pool together. Of course, this is an empirical matter. One potential line of research to investigate such developments would be to employ word-embeddings to construct measures of similarity between the meanings of different variants over time.

Bibliography

- van der Auwera, Johan, and Annemie Neuckermans. 1997. Jespersen's cycle and the interaction of predicate and quantifier negation. In *Dialectology meets typology. Dialect grammar from a cross-linguistic perspective*, ed. Bernd Kortmann, 453–478. Mouton de Gruyter.
- Bailey, Charles-James N. 1973. *Variation and linguistic theory*. Center for Applied Linguistics.
- Breitbarth, Anne. 2009. A hybrid approach to Jespersen's cycle in West Germanic. *The Journal of Comparative Germanic Linguistics* 12(2):81–114.
- Burridge, Kate. 1993. *Syntactic change in Germanic: aspects of language change in Germanic with particular reference to Middle Dutch*, vol. 89. John Benjamins Publishing.
- Darwin, Charles. 1859. On the origin of the species by natural selection.
- Embick, David, and Rolf Noyer. 2007. Distributed morphology and the syntax/morphology interface. *The Oxford handbook of linguistic interfaces* 289–324.
- Ewens, Warren J. 2012. *Mathematical population genetics 1: Theoretical introduction*, vol. 27. Springer Science & Business Media.
- Feder, Alison F, Sergey Kryazhimskiy, and Joshua B Plotkin. 2014. Identifying signatures of selection in genetic time series. *Genetics* 196(2):509–522.

- Huxley, Julian. 1942. *Evolution. the modern synthesis*. George Alien & Unwin Ltd.
- Jäger, Agnes. 2008. *History of german negation*, vol. 118. John Benjamins Publishing.
- Kimura, Motoo. 1955a. Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences of the United States of America* 41(3):144.
- . 1955b. Stochastic processes and distribution of gene frequencies under natural selection. In *Cold spring harbor symposia on quantitative biology*, vol. 20, 33–53. Cold Spring Harbor Laboratory Press.
- . 1968. Evolutionary rate at the molecular level. *Nature* 217(5129):624–626.
- Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.
- Martineau, France, and Raymond Mougeon. 2003. A sociolinguistic study of the origins of ne deletion in european and quebec french. *Language* 79(1):118–152.
- Moran, Patrick Alfred Pierce. 1958. Random processes in genetics. In *Mathematical proceedings of the cambridge philosophical society*, vol. 54, 60–71. Cambridge Univ Press.
- Wallage, Phillip. 2008. Jespersen's cycle in Middle English: parametric variation and grammatical competition. *Lingua* 118(5):643–674.
- Wilks, Samuel S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* 9(1):60–62.
- Zeijlstra, Hedde. 2004. Sentential negation and negative concord. Ph.D. thesis, Netherlands Graduate School of Linguistics.