

GENETIC DRIFT EXPLAINS SAPIR'S DRIFT IN SEMANTIC CHANGE

IGOR YANOVICH

*Carnegie Mellon University / Universität Tübingen
Pittsburgh, PA, USA / Tübingen, Germany
igor.yanovich@uni-tuebingen.de*

The linguistic notion of “Sapir’s drift” refers to the phenomenon when genetically related languages, long after their separation, undergo the same or very similar linguistic changes. Such “drift” may seem almost magical: given that language change is generally a random process, why would separate linguistic varieties exhibit the same change? There exist possible explanations demystifying Sapir’s drift, including (Joseph, 2012) who argues that if the sister languages all possessed the same variation in a given construction, that variation can serve as the basis for parallel changes long after the languages separate. Here, I propose another, complementary, explanation for Sapir’s drift for changes involving grammatical or semantic reanalysis. The new account is based on evolutionary modeling in the finite-population setting. All finite populations show the effect called *genetic drift* (unrelated to Sapir’s drift) that delays the effect of forces pushing the language in a particular direction. For reanalysis-based changes, this means that even when reanalysis of individual utterances could already occur in the proto-language, the full takeover by the new form may under the right conditions happen many centuries later, in the proto-language’s descendants. Given the introduced model, it would have been surprising if Sapir’s drift never arose, and not the fact that it does.

1. Phenomenon of interest: **motan* modals in the Germanic

Though my argument is general, it is best illustrated with some specific linguistic puzzle. I use for that purpose parallel semantic changes in **motan* modal verbs in Germanic languages.^a Today, this common Germanic word turned into English *must*, Dutch *moeten* and German *müssen*, all of which are necessity modals, roughly synonymous with “it is necessary that...”. But in the early historically attested Germanic languages, **motan* had a very different meaning. The traditional story is that it was then a possibility modal, roughly synonymous with “it’s possible that...” (see (Ono, 1958) for Old English, (Lühr, 1997) for Old High German, (Tellier, 1962) for Old Saxon, a.o.) Meanwhile, (Yanovich, 2015) and to some extent (Standop, 1957) argued instead that **motan* had a special, neither necessity nor possibility, meaning. We do not assume here Standop’s or Yanovich’s specific semantic analyses, but do assume with them that early **motan* must have meant

^aThe word is written with a star because “motan” is the reconstructed form of the verb’s infinitive that was never observed in actual texts. Finite forms of the verb are actually attested.

something special: if it were a regular possibility modal, we would expect a similar development to necessity to happen sometimes to other possibility modals in other linguistic families. But such changes have not been reported.

Here is why the semantic change of **motan* is a Sapir's-drift puzzle. In the early Germanic texts from the 8-9-10th centuries, which already exhibit very different languages (cf. 1), **motan* still was not a necessity modal. In English, it turned completely into such only by the 16th century; in Dutch and High German, even later. Thus the same change eventually occurred to **motan* in those related languages, but it only happened after many centuries of independent development. This is a case of Sapir's drift. Informally, there must have been something about the early **motan* which made it predestined for such a change; however, its effect was delayed for a long time.

(1) **9th century translations of *Pater Noster* into Germanic languages:**

fater unser, du pist in himilum, kauuihit si namo din	<i>Old High German</i>
ure fæder, þe art on heofonum, sy þin nama gehalgod	<i>Old English</i>
fadar ūsa <...>, thu bist an them hōhon himila rīkea, geuuīhid sī thīn namo	<i>Old Saxon</i>
‘Our Father, who is in {heaven (OHG, OE) / heaven’s high kingdom (OSax)}, let your name be sacred’	

We will solve the puzzle of **motan* using explicit evolutionary modeling of semantic change through reanalysis. The crucial assumption deriving differing onsets of change in the different Germanic languages is that the stability of **motan*’s old meaning was due to the forces of *genetic drift*, always observed in finite populations. The role of genetic drift is to delay the progression of an already predestined change. (Sapir’s drift and genetic drift are unrelated notions, creating an unfortunate homonymy.)

2. Framework for the evolutionary modeling of semantic change

The items subject to evolution in our model are meaning-form pairs. We will refer to **motan* with the old meaning as A, and to **motan* with the new necessity meaning as B.^b A and B serve the role of alleles of the same gene in biological evolution. At a given discrete stage of evolution, our speaker community produces *n* instances of **motan*, each of them A or B. We look only at this pool of **motan*

^bThis is an oversimplification, though it is unlikely to do harm in our case. In reality, the number of meanings and sub-meanings that **motan* could have is much greater than two (see e.g. (Yanovich, 2015)). In particular, the meaning that seems to have replaced the old meaning was specifically that of *circumstantial* necessity, as opposed to other types of necessity (see (Lühr, 1997) for Old High German, (Yanovich, 2015) for Old and Middle English). To keep modeling simple, we abstract away from the complexities of multiple meanings.

utterances, not at the grammars of individual speakers. (A remark on how to model some grammar effects is made on p. 8.) For simplicity, we assume that n is kept constant across stages. All utterances at the same stage are treated equally, so a given population state may be described simply as a vector $\vec{x} = (x_A, x_B)$, where x_A is the number of As.

An utterance *reproduces*, i.e. gets a descendant at the next stage, through being used as the model for a new utterance. In the normal case, reproduction will be faithful — e.g., an old instance of A will give rise to a new A. But As and Bs are subject to semantic reanalysis, or *mutation*: when a speaker uses some past A-utterance as the model, she may have misanalyzed it as a B-utterance, or vice versa. In this case, an old A may give rise to a new B, or the other way round. The mechanism of semantic reanalysis is by definition necessary for a meaning change to take place, (Eckardt, 2006) a.o.: for A-meaning-form pairs to turn into B-pairs, some speakers must mistake some As for Bs. Furthermore, if As may be mistaken for Bs in some suggestive contexts, some Bs will usually also sometimes be mistaken for As. But the rates of A-to-B and B-to-A reanalysis (in symbols, $r_{A \rightarrow B}$ and $r_{B \rightarrow A}$) are generally not equal.

Finally, types of form-meaning pairs may be associated with different *fitnesses* $f_A(\bar{x}) \neq f_B(\bar{x})$. Fitness is by definition the measure of reproductive success: some types may serve as the model more frequently, and fitnesses capture that. We do not have to take a stand on what causes differences in fitness, but one possible interpretation is as follows: fitness stems from better communicative efficiency. For example, it may be that given the rest of the linguistic system, it is extremely useful to utter some B tokens. In this case, old B tokens, when available, would be used as a model more frequently, i.e. have higher fitness. The mechanism causing utterances with higher fitness to reproduce more frequently is called *selection*.

The introduced framework conceptualizes semantic change as a familiar and reasonably well-studied type of finite-population evolutionary process guided by selection, mutation and genetic drift. Using it, we can apply well-known analytical machinery to puzzles in semantic change, including Sapir's drift of **motan*.

3. When drift is stronger than mutation, Sapir's drift may ensue

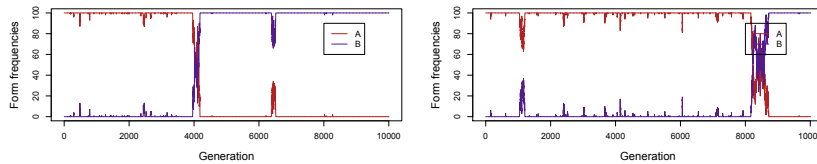
The old meaning of **motan* was relatively stable, as it persisted for many centuries in several independent Germanic languages. Such stability may come from three sources in our model: (i) selection, based on fitness (e.g. communicative efficiency), significantly favored A-**motan* over the new B-**motan*; (ii) B \rightarrow A semantic reanalysis occurred much more frequently than A \rightarrow B reanalysis; and (iii) it was B that was favored by mutation and selection, but due to the strength of genetic drift in a relatively small population of utterances, the initial all-A state persisted for a long time before crossing over to all-B.

Reasons (i) and (ii) do not work well for **motan*: under them, we need to stipulate that a great change happened either to the relative usefulness, (i), or to the

reanalysis potential of **motan*, (ii), to explain the subsequent A→B development in later Germanic languages. We do not know, of course, that such drastic changes did not happen. But presumably such changes would be random events. If so, they would be very unlikely to occur independently in long-separated languages. Thus if we explain **motan*'s drift under (i) or (ii), and thus through drastic changes, we then need to explain why those changes were coordinated. In other words, we just kick the explanatory can down the road.

But reason (iii) is different. It effectively says that the A→B change was lurking under the surface in the common Germanic because the possibility for A→B reanalysis of *individual utterances* was already there. But under particular parameters of the process, the eventual takeover by B can be inevitable yet delayed, possibly for a very long time. More precisely, when genetic drift is sufficiently strong relative to mutation and selection, it will take a long time for sporadic A→B reanalysis to cause a systemic change, even though on the level of individual utterances it will always create some B-noise. Example trajectories for such delayed B-takeover are given in 2.

(2) **Sample trajectories for delayed B-takeover** ($r_{A \rightarrow B} = 10^{-4}$, $r_{B \rightarrow A} = 10^{-5}$)



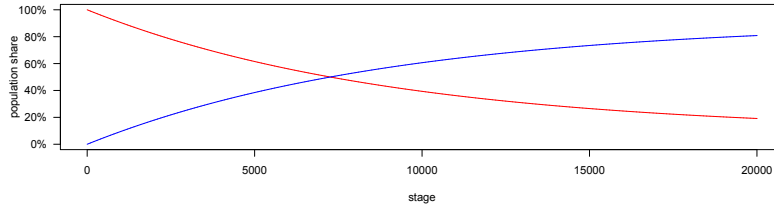
What is *genetic drift* anyway? Imagine an infinite population of A utterances that have a 1% chance of being reanalyzed as B, i.e. $r_{A \rightarrow B} = 0.01$. For simplicity, assume equal fitnesses: $f_A = f_B$. The share of B will then be steadily increasing with each stage, at least until a certain point where back-reanalysis will become strong enough to stop the B-growth.^c As the population is infinite, there will always be A-individuals mutating into B.

Now imagine the same process happening to a population of 10 A-utterances. Most likely, the next stage's population will not feature any Bs at all. Even if it will (which will happen with $p=0.1$), at the next stage the new B can still die out without offspring simply because reproduction is a random process. It is mathematically necessary that in a finite population, uniform states will be more favored. The smaller the population, the more bias towards uniformity there is. The mathematical force causing that is called genetic drift. Importantly, genetic drift by itself does not care about the actual states: A is as good as B for it. Bias for A-to-B shifts is created by asymmetry in mutation, selection or both. Genetic drift, in

^cFor example, with $r_{A \rightarrow B} = 10^{-4}$, $r_{B \rightarrow A} = 10^{-5}$, that point is when the share of Bs reaches $\frac{1}{11}$. It is easy to check that $\frac{1}{11} r_{B \rightarrow A} = \frac{10}{11} r_{A \rightarrow B}$, i.e. the inflow and outflow become equal.

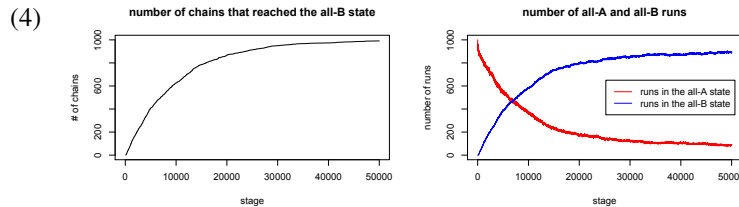
its turn, may delay the effects of that bias.

(3) **Infinite-population trajectories for $r_{A \rightarrow B} = 10^{-4}$, $r_{B \rightarrow A} = 10^{-5}$**



It is instructive to compare the drift-dominated trajectories in 2 with the trajectories we would observe in an infinite population, without stochasticity, 3. In the stochastic case 2, the process stays close to the all-A state, and the jumps between the states are rapid. In the deterministic, infinite-population case 3, the share of Bs starts to gradually rise from the very start. No bounces back into all-A occur, the system just moves gradually and steadily towards the equilibrium point.

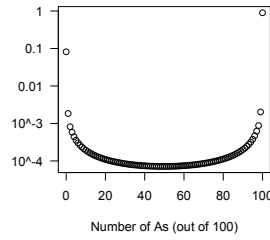
Graphs in 4 show how the same drift-dominated process as in 2 behaves statistically, on the scale of 1000 simulations starting from identical conditions. The left graph shows how many of the runs (*y* axis), all starting in an all-A state, passed through an all-B state by a given stage (*x* axis). To make this more concrete, imagine that the change in **motan*'s meaning proceeded exactly according to the laws of our evolutionary process. English, Dutch and High German would each correspond to a randomly placed horizontal line on the plot. Where that line would cross the graph would determine the timing of a complete switch away from the non-necessity meaning. It is easy to see that the switch would be inevitable in the long run, but would likely happen at quite different times for the three languages.



The right plot in 4 shows the number of runs, out of the same 1000, which happen to be in the all-A (red) or the all-B (blue) state at a given stage. We can note that initially, the share of runs in the all-B state rises sharply, but then starts to level off. The share of runs in the all-A state does not go to 0 either. In fact, both of them tend towards specific *equilibrium values*. As both A-to-B and B-to-A reanalysis is possible, our evolutionary process will not only reach all-B starting from all-A, but will also inevitably return back to all-A after having been in all-B.

The long-term behavior of our process can be studied analytically by using its stationary distribution S . As long as transitions between all states are possible and there is a state which can exactly reproduce itself at the next stage (as all our population states are)^d, by the Perron-Frobenius theorem there is a unique vector S that describes how much time the process will spend in each state in the limit of eternity. Computing S for large n is computationally prohibitive, but for small n , even a personal computer can do the work fast enough.^e The stationary distribution for the same population size and reanalysis rates as we used above are given in 5. The shares of runs in the all-A and all-B states will in the limit tend towards the stationary distribution's shares of all-A and all-B. In our examples, the red and blue lines on the right graph of 4 tend towards 8.2% and 90.0%. The rest 1.8% of the time the process is in the intermediate states.

(5) The states' time shares for $r_{A \rightarrow B} = 10^{-4}$, $r_{B \rightarrow A} = 10^{-5}$, $f_A = f_B$ (log scale).



For cases of Sapir's drift, such long-term behavior is not a good approximation. Examples with constant switching back and forth between the archaic and innovative forms are not generally found in historical-linguistic literature. But this does not undermine the usefulness of our analysis for two reasons. First, in realistic linguistic cases, the system would often switch from a B-dominated state to a further innovative meaning-form type C. If the outflow from all-B states to C, D, and so forth down the reanalysis chain, is strong enough, the process will rarely have time to return to all-A.

Second, cases of constant and rapid back-and-forth reanalysis were not something that historical linguists looked for. The standard thinking about language

^dOur evolutionary process is a Markov chain. The conditions described in the text are irreducibility and aperiodicity.

^eBy Perron-Frobenius, S is the probability left eigenvector of the transition matrix T of the process. Rows of T correspond to the current population state x , columns, to the next population state y . The entries t_{ij} of T state the probability of getting from state x_i to state y_j . Let i and j in t_{ij} refer to the number of A-utterances out of n in the current and next states respectively, and let's call individuals with an A-parent A-offspring. Then:

$$t_{ij} = \sum_{0 \leq k \leq n} p(k \text{ A-offspring in } y \mid x_i) \sum_{0 \leq l \leq j} f_{bin}(l; k, 1 - r_{A \rightarrow B}) f_{bin}(j - l; n - k, r_{B \rightarrow A})$$

where $f_{bin}(a; b, p)$ is the probability of a successes after b trials with probability p of success.

change does not feature a scenario when there is a constant flux between two states (as opposed to cases of stable variation when two forms coexist for a long time). So when linguists observe form B to be frequent in year 1500, and then again in year 1600, the default assumption is that the later uses are direct descendants of the earlier ones. In our evolutionary model, this is not necessarily so. If the process jumps frequently between all-A and all-B, then there could have been a complete absence of Bs for some decades between 1500 and 1600; the new abundance of Bs at 1600 may be unrelated by descent to the old abundance. What enabled both cases of abundance in such a scenario would be the general potential for reanalysis from A to B in the language. Summing up, one reason why we do not find descriptions of erratic back-and-forth trajectories in the historical literature may be the fact that such trajectories have not been conceptualized before as possible for language change.

Let's conclude our analysis. There are three main cases for the behavior of our evolutionary process. (i) When asymmetrical reanalysis rates or asymmetrical fitness strongly dominate genetic drift, we see different languages starting from the same condition following roughly the same time course, with a little noise due to the stochastic nature of the process. This would be "regular", non-delayed linguistic development. (ii) As genetic drift becomes stronger, the time course of a change caused by the same asymmetry will become very different. The change will eventually happen in any language with the same starting conditions, but it may occur at quite different times (recall the distribution on the left plot in 4). Most of the time the modeled language will be in a uniform or almost-uniform state (see 2 for two examples, and 5 for the equilibrium shares of the uniform states). If the time-scale of our observations is comparable with the time scale of the sharply rising part of the curve in the left graph in 4, we will see languages undergoing the same A-to-B change with different timing. This is the condition for Sapir's drift. (iii) Finally, if drift dominates, but the time scale of our observations is such that we can only see the near-equilibrium portion of the curve on the left in 4 rather than its sharply rising portion, we will see rapid shifts between A and B happening from time to time, followed by periods of relative stability. Whether there exists actual changes that follow this pattern is an open empirical question.

4. Conclusion: Towards generalizing the model

We have seen above that a very simple model of language change predicts the existence of Sapir's drift. What we have *not* seen is whether our simple model can be an adequate approximation for actual language change. First, speaker's grammars are conspicuously absent from analysis. Yet we know that however usage-based language evolution may be, utterances are produced by speakers who may have their idiosyncratic constraints as to what to say and how to understand what they hear. Another conspicuous oversimplification is that our model above treats all linguistic utterances as equally accessible for all speakers (who are implicit, but

present in the model). A more realistic model would involve several subpopulations of utterances, corresponding to different speakers or tightly knit speech communities, with some exchange of utterances between them. However, both preliminary theoretical analysis and computer simulations suggest that accounting for those more complex conditions would not make Sapir's drift disappear.

Second, Sapir's drift is predicted to ensue only for particular sets of values of the evolutionary parameters in the model, including the size of the utterance population n , reanalysis rates $r_{A \rightarrow B}$ and $r_{B \rightarrow A}$, fitnesses of forms f_A and f_B , number of stages passed t , and also the mode of reproduction for utterances (i.e. the probability distribution over the number of offspring, see (Der, Epstein, & Plotkin, 2011)). I have not, however, presented here any mapping from real-world linguistic observations to parameters of the model. Thus we know that the model predicts Sapir's drift, but we do not know if it predicts it for reasonable parameter ranges.

One reason I have refrained from providing any specific mapping is that lifting the model to more realistic cases has crucial consequences for the parameter ranges that create Sapir's drift situations. For example, suppose we include into the model the following effect of individual speaker grammars: reanalyzing A as B is easier if the speaker can already produce some Bs than if they have to innovate B without any prior direct evidence of its existence. In this case, the process will become much more strongly biased towards uniform states than by genetic drift alone. Thus very different predictions about the number of stages needed for Sapir's drift's emergence will be made by the simpler and more complex models.

That said, there is hope for getting reasonable empirical estimates for some of the evolutionary parameters. For example, if we equate one evolving pool of utterances with the memory pool of one speaker, we can then use the simple model as it stands to estimate, using both experimental and single-speaker-corpus techniques, the effective size of the utterance pool for real-life speakers (i.e. n), the effective rates of reanalysis of given constructions (i.e. different rs) and/or their relative fitnesses (i.e. fs), and the scale of t . We can then lift those results into a more complex model by looking at the linguistic output of whole speech communities as recorded in corpora and other types of language surveys.

In other words, estimation of the parameters of the model will not be an easy task, but it does not appear *a priori* impossible. If this line of research bears fruit, we will get a set of explicit models of language change that can be used for inferring crucial evolutionary characteristics of diachronic processes, and in effect for probabilistically predicting when and which language change can take place.

Acknowledgements

I am grateful to Adrian Brasoveanu, Gerhard Jäger and audiences at CMU and UCSC for useful discussions, and to the anonymous reviewers for EVOLANG for their valuable comments which made the text better. I also gratefully acknowledge the support of the Alexander-von-Humboldt foundation and of CMU Philosophy.

References

- Der, R., Epstein, C. L., & Plotkin, J. B. (2011). Generalized population models and the nature of genetic drift. *Theoretical Population Biology*, 80, 80-99.
- Eckardt, R. (2006). *Meaning change in grammaticalization. an enquiry into semantic reanalysis*. Oxford: Oxford University Press.
- Joseph, B. D. (2012). A variationist solution to apparent copying across related languages. In *Copies vs. cognates in bound morphology*. Brill.
- Lühr, R. (1997). Zur Semantik der althochdeutschen Modalverben. In G. Fritz & T. Gloning (Eds.), *Untersuchungen zur semantischen entwicklungsgeschichte der modalverben im deutschen* (pp. 159–176). de Gruyter.
- Ono, S. (1958). Some notes on the auxiliary *motan. *Anglica*, 3(3), 64-80.
- Standop, E. (1957). *Syntax und Semantik der modalen Hilfsverben im Altenglischen* magan, motan, sculan, willan. Bochum-Langendreer: Pöppinghaus.
- Tellier, A. (1962). *Les verbes perfecto-présents et les auxiliaires de mode en anglais ancien: (viii^e s. - xvi^e s.)*. Paris: C. Klincksieck.
- Yanovich, I. (2015). *Old English *motan, variable-force modality, and the pre-supposition of inevitable actualization*. (to appear in *Language*)