Mathematical Models of Word Order Change in Middle English

W. Garrett Mitchener

August, 2004

Web site: `http://www.math.duke.edu/~wgm`


```
Id: Handout.tex,v 1.10 2004/08/31 15:26:54 wgm Exp
Id: Common.tex,v 1.11 2004/08/20 12:53:06 wgm Exp
```

1 Introduction

Explanations of language change frequently focus on answering these questions:

Explanations of syntactic change & reanalysis

- Three Hard Questions:
  - Description of language before
  - Description of ambiguous, unambiguous, reanalyzed forms
  - Description of language after
- Three Very Hard Questions:
  - Why did the change occur once feasible?
  - Why did it happen at the time it did?
  - Why didn't some other potential change happen?

The science of historical linguistics has advanced to the point that the first three items are quite approachable. Meetings and journals are full of excellent descriptions of languages before and after a change, and reasonable candidates for forms that are ambiguous and get reanalyzed, and forms that appear and disappear as one grammar yields to another.

However, the last three questions are always sticking points. (Ex: At DiGS 8 at Yale in June 2004, these questions came up repeatedly after many presentations.) Furthermore, the available answers are often speculative, untestable, or otherwise incomplete.

For example, one might invoke extralinguistic factors such as contact, prestige, or stigma, to explain why a feasible change took place. But other questions arise: Why was this borrowed but that wasn't? Why didn't bilingualism result?

It's also tempting to invoke economy and say that the change moves downhill so as to minimize some measure of complexity, and that drives the population from the old grammar to the new grammar. However, it isn't clear which mental operations are more or less expensive, and until we know more about the neurological hardware in the brain, principles of economy will remain somewhat speculative.

Finally, there is interest in explaining change as occurring because of shifts in usage, which cause shifts in the primary linguistic data, thereby triggering the new grammar in children (see for example [Lightfoot, 1999]). However, in hindsight, we are left with a chicken-and-egg problem: Did a shift in usage cause this change, or did the change cause the shift in usage?

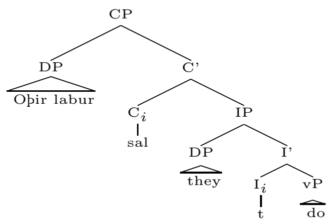
In what follows I will illustrate how mathematical modeling and simulation might help us to confirm proposed answers to the Three Hard Questions and to address the Three Very Hard Questions more completely.

2 Verb-second in Middle English

Take the case of verb-second in Middle English. To begin with, manuscript evidence indicates that there were two (possibly more) regional dialects, each with slightly different word orders.

Northern Middle English was heavily influenced by Scandinavian invaders who spoke Old Norse. The northern word order seems to be borrowed from Old Norse:

Verb-second in northern Middle English



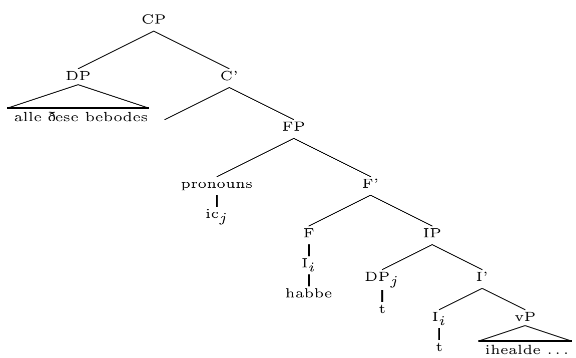
(1) '[Oþir labur] *sal* they do'

(*The Rule of St. Benet*, Fischer et al. [2000, p. 131])

CP-type verb-second, pronouns are full words

Verb-second in southern Middle English is not quite the same as in German (SOV+CPv2) or Icelandic (SVO+IPv2): The finite verb can raise in embedded clauses, although topic fronting only happens in root clauses. Also, there is a space for pronouns between C and I resulting in verb-third sentences as in the example (2). Hence the following tree structure from Fischer et al. [2000, p. 126–8]):

Verb-second in southern Middle English



(2) '[alle ðese bebodes] ic *habbe* ihealde fram childhade'

(*Vices & Virtues,* Fischer et al. [2000, p. 130])

FP-type verb-second, pronouns are special

Three hard questions

- Before: South SVO+FPv2+pronoun slot, North SVO+CPv2
- Northern, Southern, Modern: Subject Vf ...
  Northern, Southern: XP Vf Subject ...
  Southern: XP Pronoun Vf Subject ...
- After: Modern SVO

So why did Middle English switch to modern SVO with no verb-second? The usual explanation is that increased contact between the two regional dialects was responsible for the change [Fischer et al., 2000, Kroch et al., 2000, Pintzuk, 1993, van Kemenade and Vincent, 1997]. Lightfoot [1999] makes a more specific proposal: The speech of southerners would contain fewer plain verb-second sentences than northerners were used to hearing, because southerners would often front the topic plus a pronoun as in (2). Therefore, contact produced an environment in which children learning the northern dialect heard a lower fraction of sentences that served as cues for the northern verb-second word order.

Cue sentences for NME

XP Vf Subject ...

- Assume children determine quickly that NME has underlying SVO
- Cue sentences: Those which can only be parsed with SVO+CPv2
- Learning hypothesis: Northern children acquire a verb-second grammar only if they hear enough sentences of this form. Proposed by Lightfoot [1999].
- Contact with southern ME leads to fewer cue sentences

In the contact environment, children then ought to acquire the modern SVO word order, which further dilutes the number of cue sentences, causing the change to spread north and south.

This hypothesis seems quite reasonable, but is it testable in some quantitative way? Here's a simple mathematical model:

First model: Grammar and learning

- Two grammars, G_1 and G_2, but G_1 is marked (requires substaintial evidence to acquire) and G_2 is the default
- People speak either G_1 or G_2, no diglossia
  (We'll come back to this. . .)
- Learning: Children hear n sentences total; choose G_1 if m or more of them are cues, else choose G_2
- Speakers of G_1 produce cue sentences at rate $p_1 \approx 30\%$
  Speakers of G_2 produce cue sentences at rate $p_2 < 5\%$

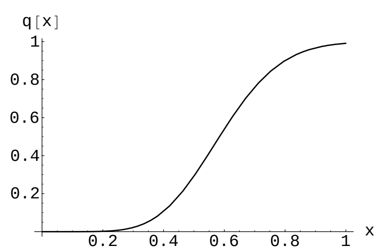
First model: Population structure

- Two regions, North and South; children learn only from neighbors
- x_N = fraction of northerners speaking G_1
- x_S = fraction of southerners speaking G_1
- Mixing parameters α and β: Measure rate at which people move from one region to the other

Mathematical notation

$$q(x) = \mathbb{P}\{\text{child picks } G_1\}$$
$$= \sum_{j=m}^{n} \binom{n}{j} \gamma^j \gamma^{n-j}$$

$\gamma = p_1 x + p_2(1 - x)$
x = fraction speaking G_1
n = # sample sentences = 100
m = min # cue = 20
p_1 = cue freq for G_1 = 0.3
p_2 = cue freq for G_2 = 0.05

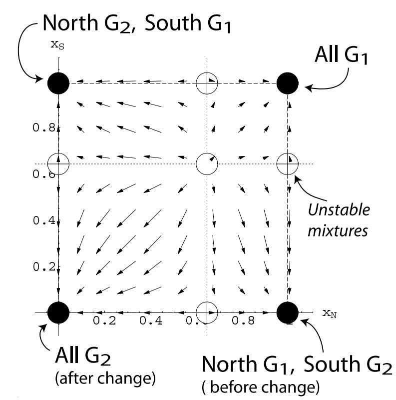


$$\dot{x}_N = q(x_N) - x_N + \alpha(x_S - x_N)$$
$$\dot{x}_S = q(x_S) - x_S + \beta(x_N - x_S)$$

The primary component of the model is a pair of differential equations called a dynamical system [Strogatz, 1994]. The population state is represented by a point (x_N, x_S) in a square called the phase space. The phase space is a square because x_N and x_S both range from 0 (no G_1) to 1 (100% G_1). The dynamical system defines a vector field on the phase space, and populations will change over time by following the vector field. The vector field represents a sort of wind or current that affects populations depending on their location in the phase space.

Below is a phase portrait and vector field for the case $\alpha = 0$ and $\beta = 0$. Since $\alpha = \beta = 0$, the north and south are isolated. The various dots in the phase portrait represent fixed points, which are places where the vector field is zero (as in "no wind"). The black dots are stable fixed points, meaning the wind nearby blows into them. (Think of a black hole.) The white and crossed dots are unstable fixed points, meaning the wind nearby blows away from them. A population near a stable fixed point will move towards it, and a population near an unstable fixed point will eventually move away. The stability of a fixed point is determined by constructing a linear approximation to the vector field nearby using the Jacobian matrix, and computing its eigenvalues. If all eigenvalues have negative real parts, then the fixed point is stable, otherwise it's unstable. The dashed lines are special paths through the vector field that connect the fixed points and are of mathematical interest.

Phase portrait for isolated regions



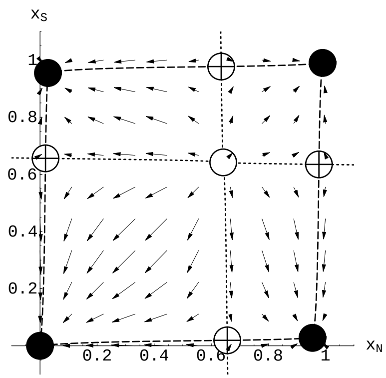
$\alpha = \beta = 0$

This dynamical system allows for stable split populations (upper left and lower right corners) where one region is 100% one grammar and the other region is 100%

the other. The lower right fixed point is roughly analogous to the scenario in Middle English before contact.

If we increase α and β so as to mix the regions, the wind shifts, so to speak, resulting in a slightly different phase portrait. The fixed points shift as well:

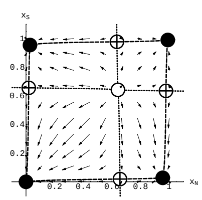
Phase portrait: small mixing



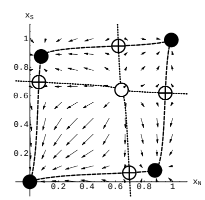
$\alpha = \beta = 0.03$

As we continue to increase α and β, the regions mix more and more, and the stable split fixed points approach nearby unstable fixed points. And at a certain mixing level, they actually collide, as indicated by the squares in the last picture below:

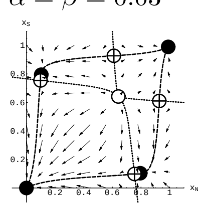
Phase portrait: more and more mixing



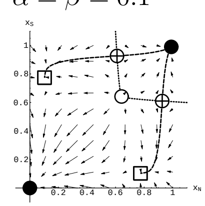
$\alpha = \beta = 0.03$



$\alpha = \beta = 0.1$



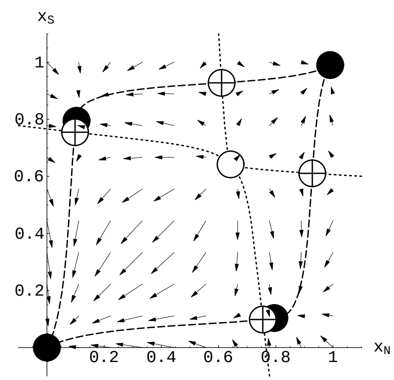
$\alpha = \beta = 0.15$



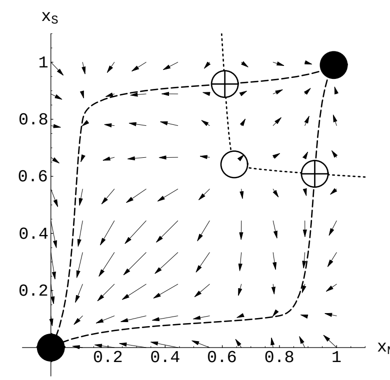
$\alpha = \beta = 0.152985\ldots$

A sudden change in the structure of a phase portrait that happens when parameters in the dynamical system hit a critical value is called a bifurcation. That's what happens here: In each corner, two fixed points collide, and as the mixing parameters continue to increase, they disappear entirely, as shown below. (The term "bifurcation" comes from the fact that most such structural changes involve the creation or destruction of pairs of features.) Bifurcations like the ones shown here are identified by finding parameter values that cause a fixed point to have an eigenvalue with zero real part.

Phase portrait: bifurcation



Before

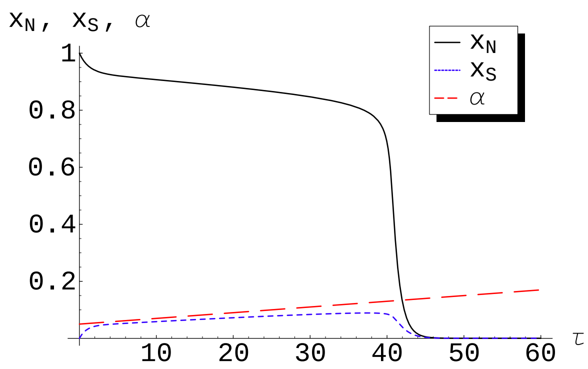


After

Once this bifurcation happens, there is no longer a stable fixed point in the lower right corner, so any population in a split state must flow all the way to the lower left corner, where G_1 is extinct.

Let's put it all together. If the population starts in a split state, and mixing increases, there will be a sharp drop in the fraction of people speaking G_1 when the mixing parameters cause the bifurcation:

Time traces



Horizontal axis: Time (rescaled units, not years)
Vertical axis: Fraction of population speaking G_1
Using $\alpha = \beta$, both increasing linearly in time

3 Improvements to the model

This simple model exhibits a sudden change in the population, and with reasonable values for the various numerical parameters, the model correctly predicts that the marked dialect should go extinct. Thus, the model illustrates that contact between dialects resulting in diluted evidence for the more marked grammar is sufficient to explain the extinction of a dialect as in Middle English.

However, this model is in many ways too simple, and there are many flaws with it that should be addressed:

Needed improvements

- Allow diglossia
- Allow more grammars (SVO+v2, SVO+v2+pronoun slot, vanilla SVO, various kinds of v2, etc.)
- Connect to manuscript data

The possibility of allowing people to speak both grammars in some proportion is necessary to reconcile the model with the observation from manuscripts that individual speakers seem to use a mixture of the old and new grammars, a sort of syntactic doublet. Such dialectic diglossia can be modeled by switching to a function-valued differential equation: Each region is represented by a time-dependent probability density function u, where $u(t, z)$ is the density at time t of people that mix the two grammars such that a fraction z of their utterances are generated by G_1 and the rest of their utterances are generated by G_2.

It turns out that such a model displays essentially the same behavior. (In fact, the average values of u_N and u_S follow essentially the same dynamical system as the first model.) Unfortunately, I was not able to find much in the literature about how children might decide on the proportion to use in the case of a doublet. (If you know of anything, please let me know!) So, to get this model to work, I have to "make up" a reasonable learning approximation. The resulting model is simply too speculative to be useful at this point.

The first model only allows for two grammars, when the actual situation with Middle English seems to require at least three: The northern SVO+v2 grammar, the southern SVO+v2+pronoun slot grammar, and the modern SVO grammar. It may also turn out that variations of these grammars caused by adult/L2 learning and accommodation are also needed to properly simulate the loss of verb-second. Each additional grammar adds two dimensions to the phase space, thereby vastly increasing the difficulty of the mathematics.

The first model is also very abstract. It makes some quantitative predictions (for example, if the learning parameters are set so that G_1 is too hard or too easy to learn, things go wrong) but the main result is purely qualitative: G_1 goes extinct suddenly. Manuscripts suggest that the actual extinction took decades, however there is uncertainty as to how to interpret such data. In particular, it is unclear how quickly changes in the spoken language might appear in the written language.

4 Proposal

In short, a second model is called for.

Second model: a simulation

In collaboration with Anthony Kroch

- Agent based, not a continuous approximation
- Minimalist grammar [Adger, 2003]
- Detailed learning algorithm [Yang, 2002]
- Literacy

This second model will be an agent based simulation, representing individuals, rather than the "mean field" approximation used in the first model. This is known to be critical for correctly modeling certain features of population dynamics [Briscoe, 2000, Durrett and Levin, 1994, Kirby, 2001].

Language will be represented by a sentence generator based on the minimalist program. This will allow for a wide variety of grammars grounded in current syntactic theories.

Agents will acquire language based on a fairly detailed process, most likely something like the proposal in Yang [2002]. The acquisition of phonology, vocabulary and meaning will not be modeled, but syntax will, since the changes of interest are primarily syntactic.

Perhaps most importantly, some of the agents will be literate. The goal is to model literacy such that documents cause literate speakers' language to persist after their death. We would then like to connect the model to the available data from the written language. For example, the constant rate effect Kroch [1989]. This feature allows us to model differences between the spoken and written forms of ancient languages.

The proposed simulation is in the initial stages. With any luck, we will have made some progress within the next year or so.

References

David Adger. *Core Syntax: A minimalist approach*. Oxford University Press, Oxford, 2003.

E. J. Briscoe. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2):245–296, 2000.

Richard Durrett and Simon Levin. The importance of being discrete (and spatial). *Theoretical Population Biology*, 46(3):363–394, December 1994.

Olga Fischer, Ans van Kemenade, Willem Koopman, and Wim van der Wurff. *The Syntax of Early English*. Cambridge University Press, 2000.

S. Kirby. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110, 2001.

Anthony Kroch. Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1:199–244, 1989.

Anthony Kroch, Ann Taylor, and Donald Ringe. The middle english verb-second constraint: A case study in language contact and language change. In Susan Herring, Pieter van Reenen, and Lene Schøsler, editors, *Textual Parameters in Older Languages*, pages 353–391, Philadelphia, 2000. John Benjamins Publishing Company.

D. Lightfoot. *How to Set Parameters: Arguments from Language Change*. MIT Press, Cambridge, MA, 1991.

D. Lightfoot. *The Development of Language: Acquisition, Changes and Evolution*. Blackwell Publishers, Malden, MA, 1999.

Susan Pintzuk. Verb seconding in Old English: verb movement to Infl. *The Linguistic Review*, 10:5–35, 1993.

S. H. Strogatz. *Nonlinear Dynamics and Chaos*. Perseus Books, Reading, MA, 1994.

Ans van Kemenade and Nigel Vincent, editors. *Parameters of Morphosyntactic Change*. Cambridge University Press, 1997.

Charles D. Yang. *Knowledge and Learning in Natural Language*. Oxford University Press, Oxford, 2002.