Model-based simulation

Christoph Wolk

April 5, 2016

Introduction

Assumption (1): Grammatical variation is sensitive not (only) to categorical constraints, but to multiple and typically conflicting probabilistic constraints, be they formal, semantic, or phonological in nature. […]

Assumption (2): Linguistic knowledge includes knowledge of probabilities, and language users have powerful predictive capacities.

Assumption (3): Corpus-based regression models match speakers' predictive abilities.

(Szmrecsanyi 2013; references omitted)

Variation

- These probabilistic constraints are not universal and constant, but (at least to a degree) malleable across speakers, time and space. Some of this variation appears to be robust.

- Therefore they have to be – at least partially – learned from other speakers.

- Where does this variation come from?
  - is there 'probabilistic drift'?
  - are there environmental causes?
    - what would they look like? how can we know?

Idea

I have doubts that we can completely solve these issues - at least, I don't know how.

What I do think might be worth a try is formalizing the issues in such a way that we can evaluate how they play out.

Idea

To recap:

- speakers have probabilistic knowledge and powerful predictive capacities
- speakers can use them to generate new material
- speakers learn the parameters by observing speakers

- regression models have probabilistic knowledge and powerful predictive capacities
- regression models can use them to generate new material
- regression models learn the parameters by from observing speakers
- … or other regression models?

Artificial artificial language learning

- **regression models as minimal probabilistic learners?**
- we can train models, then use their output to train a new generation of models.
- goal: simulate changes and developments over time, such as responses to changes in the input distributions of predictors

A simple community of models model

1. start with a model of a linguistic phenomenon; this model is our "patient zero"
2. create a generation of clones, with some random variation added
3. get some data about the models (coefficients …)
4. use models to create new datasets, where
   - response is (randomized) model prediction for the data set
5. use datasets to train a new community of models
6. loop to 3.

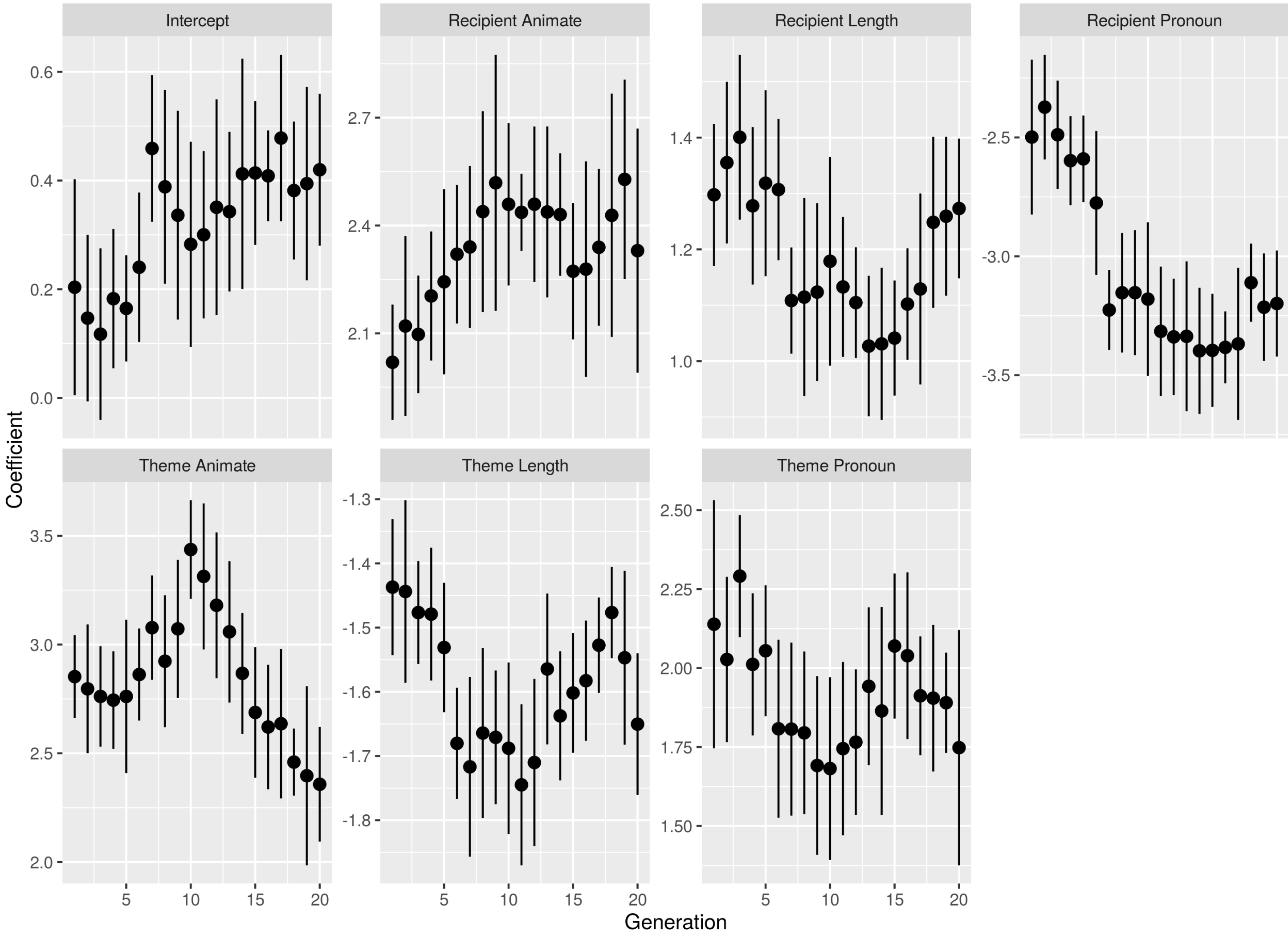
The data

- Wolk et al. (2013) datives, with some modifications to make the model simpler: definiteness reduced to pronoun vs. non-pronoun
- model simplified: animacy, length and pronominality, each for recipient and theme
- full set of random effects for the initial generation, which are not used for prediction. Later generations only have a random slope for speaker (model)

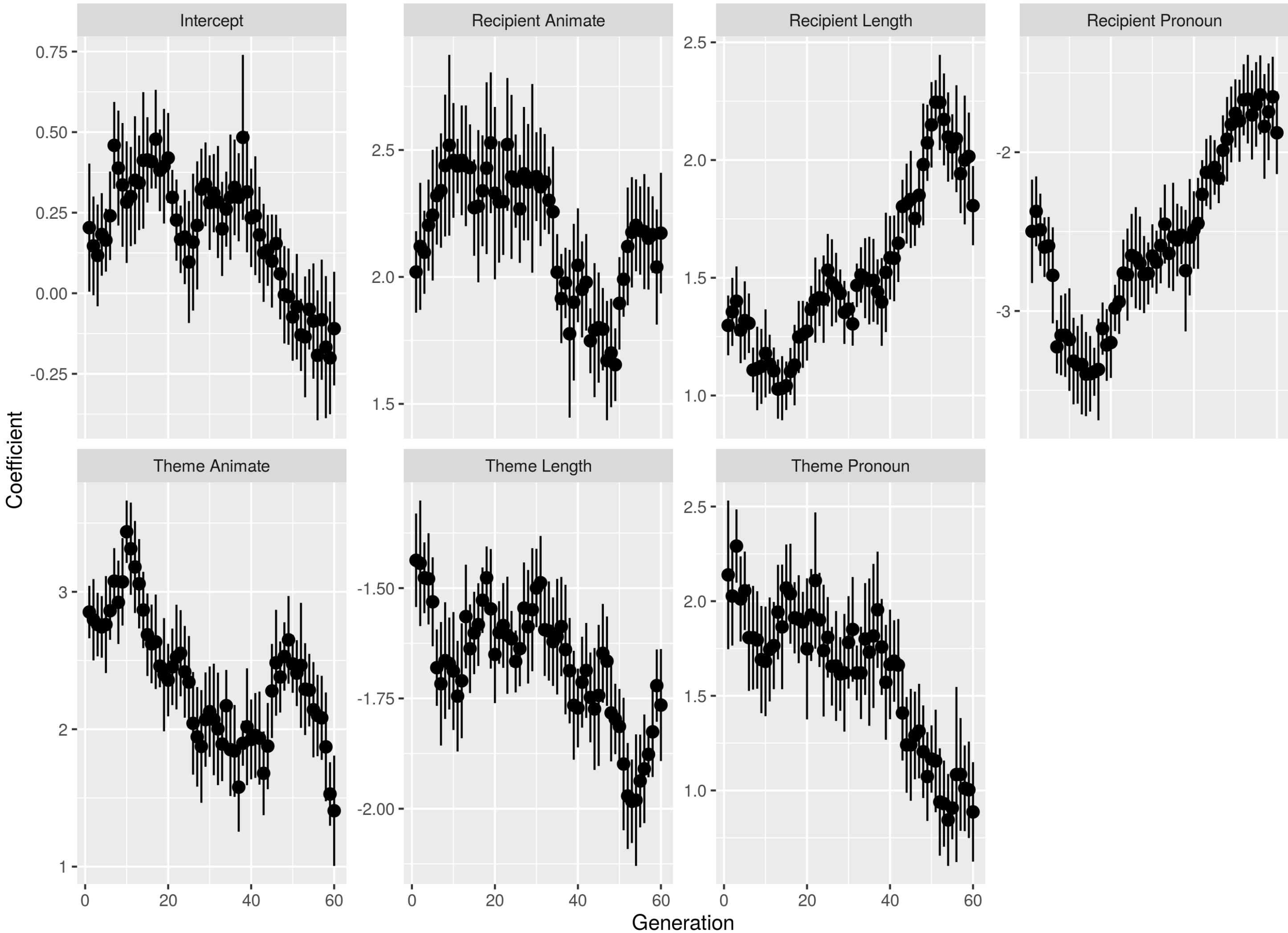
The process

- 25 models per generation
- egalitarian distribution: each model has the same probability of passing on individual realization to each member of the next generation
- each new realization is chosen at random according to the predicted probability
- generations after the first are exposed to the data set 7 times; input models may vary

10 models, 3 datasets, 20 generations



10 models, 3 datasets, 60 generations



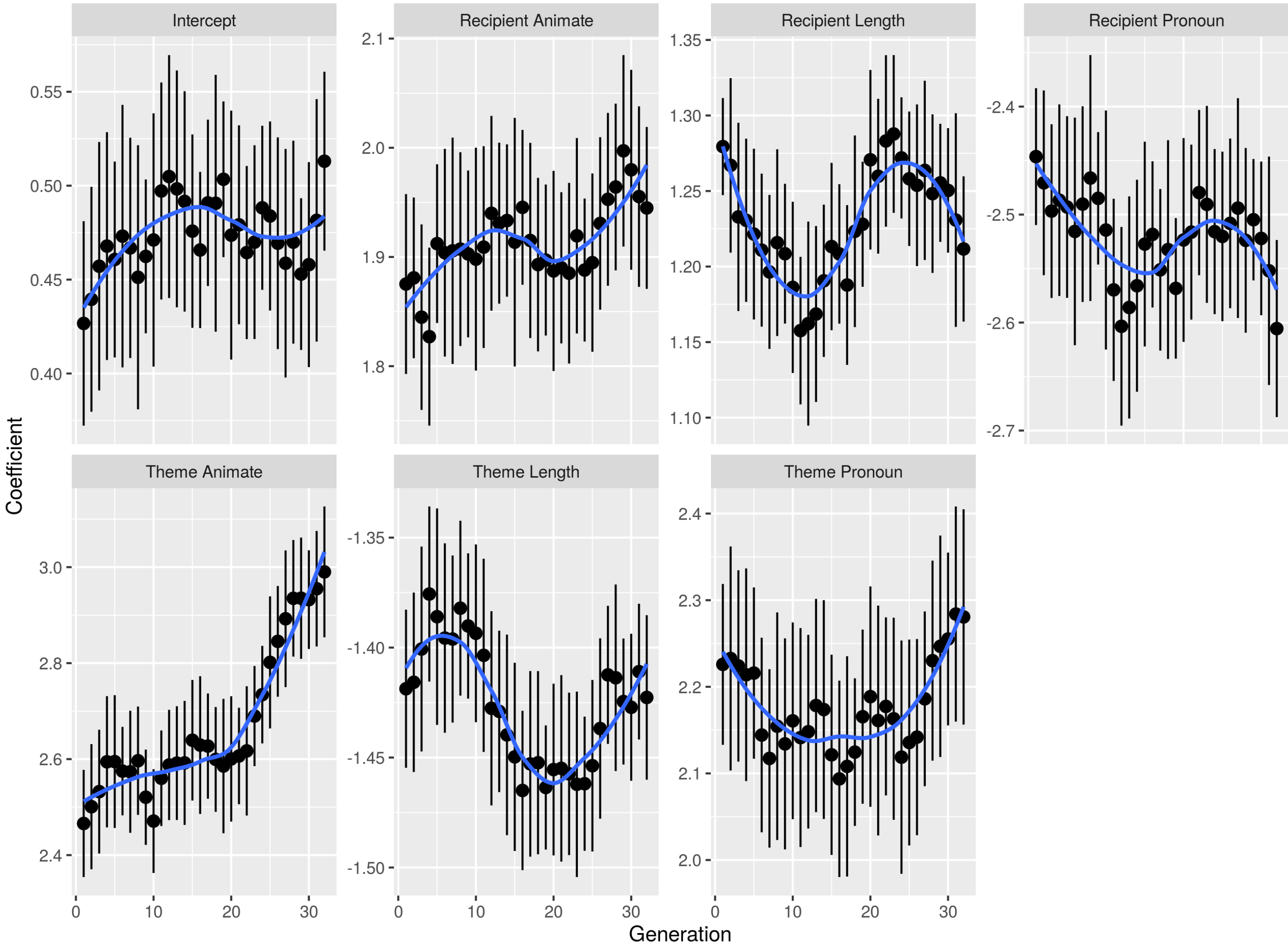
Discussion

The values seem implausibly variable, at least to my eye. In actual data, we did find diachronic developments, but not so many that were so large.

Interestingly, the changes do not appear to be a pure random walk, but keep direction for a while. Closer investigation shows that some features change in lockstep - recipient pronominality weakens as recipient length increases, for example.

- -> Increase amount of data presented to models, increase number of speakers

Regular parameters



Discussion

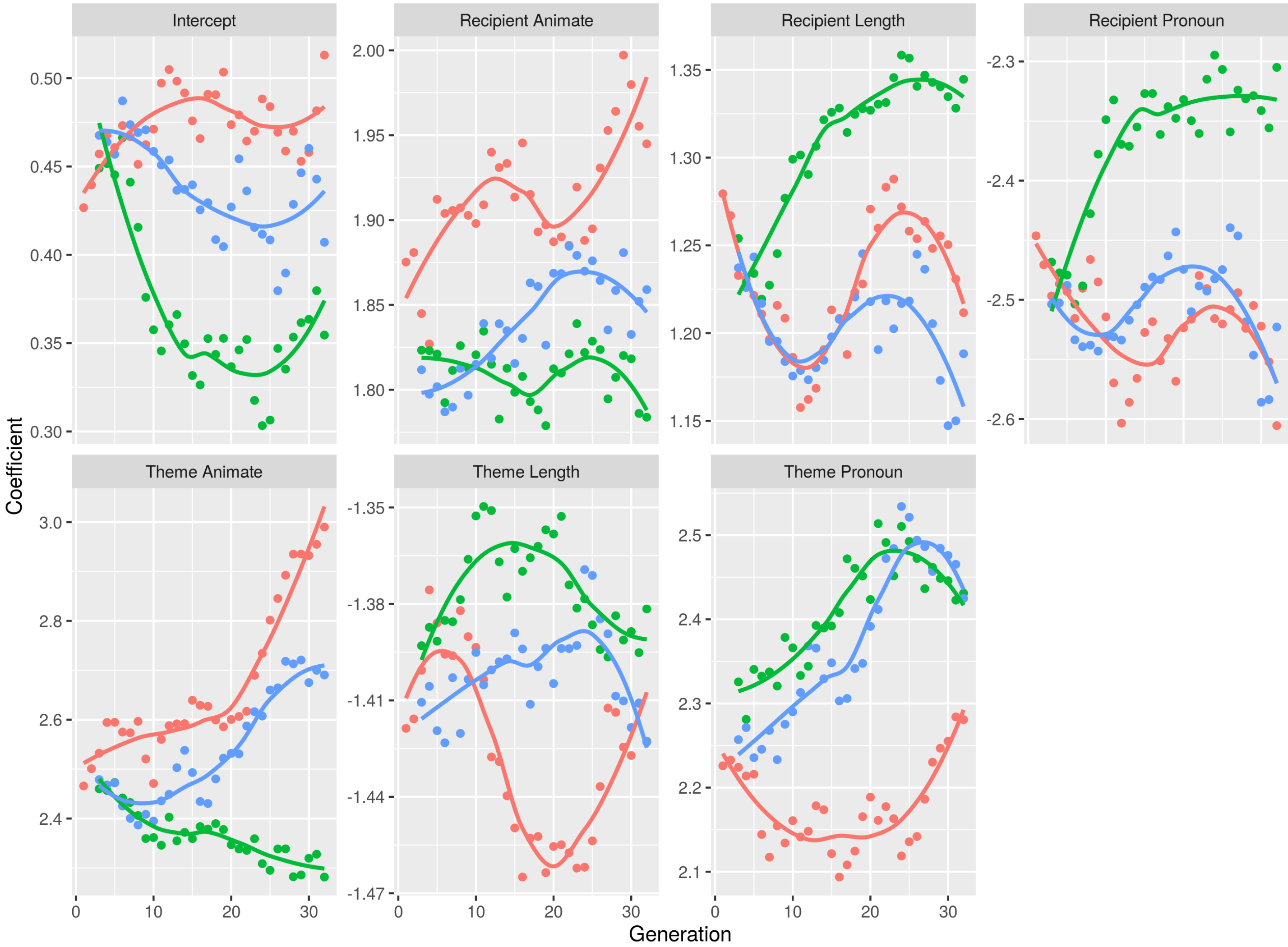
- Here, the changes over time seem more plausible. Most of them are rather small, although individual features can move quite a bit.
- The linear drift still seems to happen, albeit less clearly.
- *How do the models know that a change is underway?*

Across Dialects and Varieties

Let's create varieties!

We can at any point in the simulation train a second (etc.) strain from one generation of models.

'World-wide'



Intercept
Recipient Animate
Recipient Length
Recipient Pronoun
Theme Animate
Theme Length
Theme Pronoun
Coefficient
Generation
0.50
0.45
0.40
0.35
0.30
2.00
1.95
1.90
1.85
1.80
1.35
1.30
1.25
1.20
1.15
-2.3
-2.4
-2.5
-2.6
3.0
2.8
2.6
2.4
-1.35
-1.38
-1.41
-1.44
-1.47
2.5
2.4
2.3
2.2
2.1
0
10
20
30

Discussion

- The varieties develop independently. Overall, the differences remain small, but for individual predictors we may find quite a bit of differences between them.
- But would a simulated linguist find them?
- Idea: sample predictions predictions from each variety for the original dataset, then build a model with by-variety interactions.
- However, in the majority of runs/comparisons, there are no significant interactions. If we pick generations toward the end of the simulation, there is a significant effect in roughly every other randomized sample

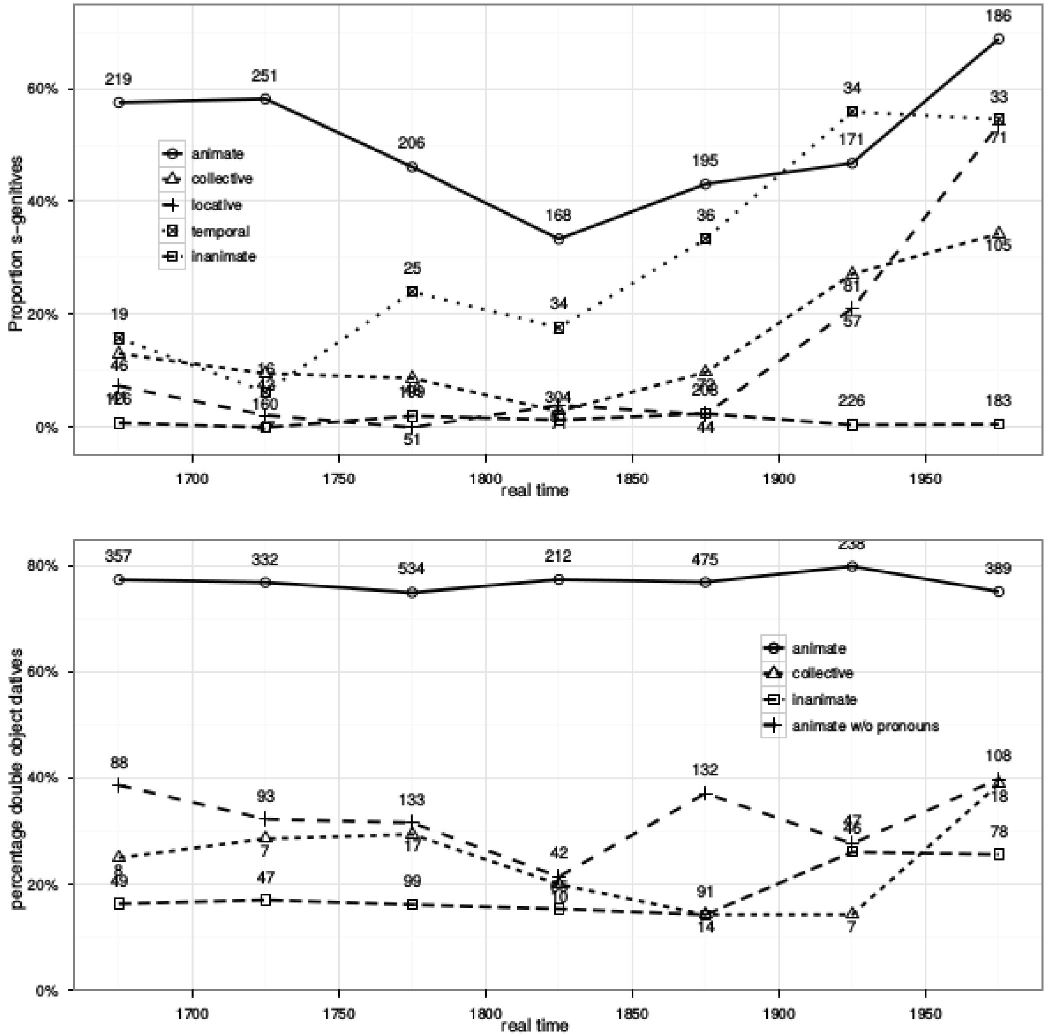
Interim summary

- Is there probabilistic drift?
- The probabilistic learner does exhibit such drift
- However, it seems that such drift requires implausible finetuning, or is too active, or is hard to detect with the tools we use

Hypothetical environmental causes

Collective and/or inanimate possessors/recipients over time begin to appear more often first in genitive/dative constructions (Wolk et al. 2013)

Visualized


Proportion s-genitives
real time
animate
collective
locative
temporal
inanimate
percentage double object datives
real time
animate
collective
inanimate
animate w/o pronouns

Accounting for the changes

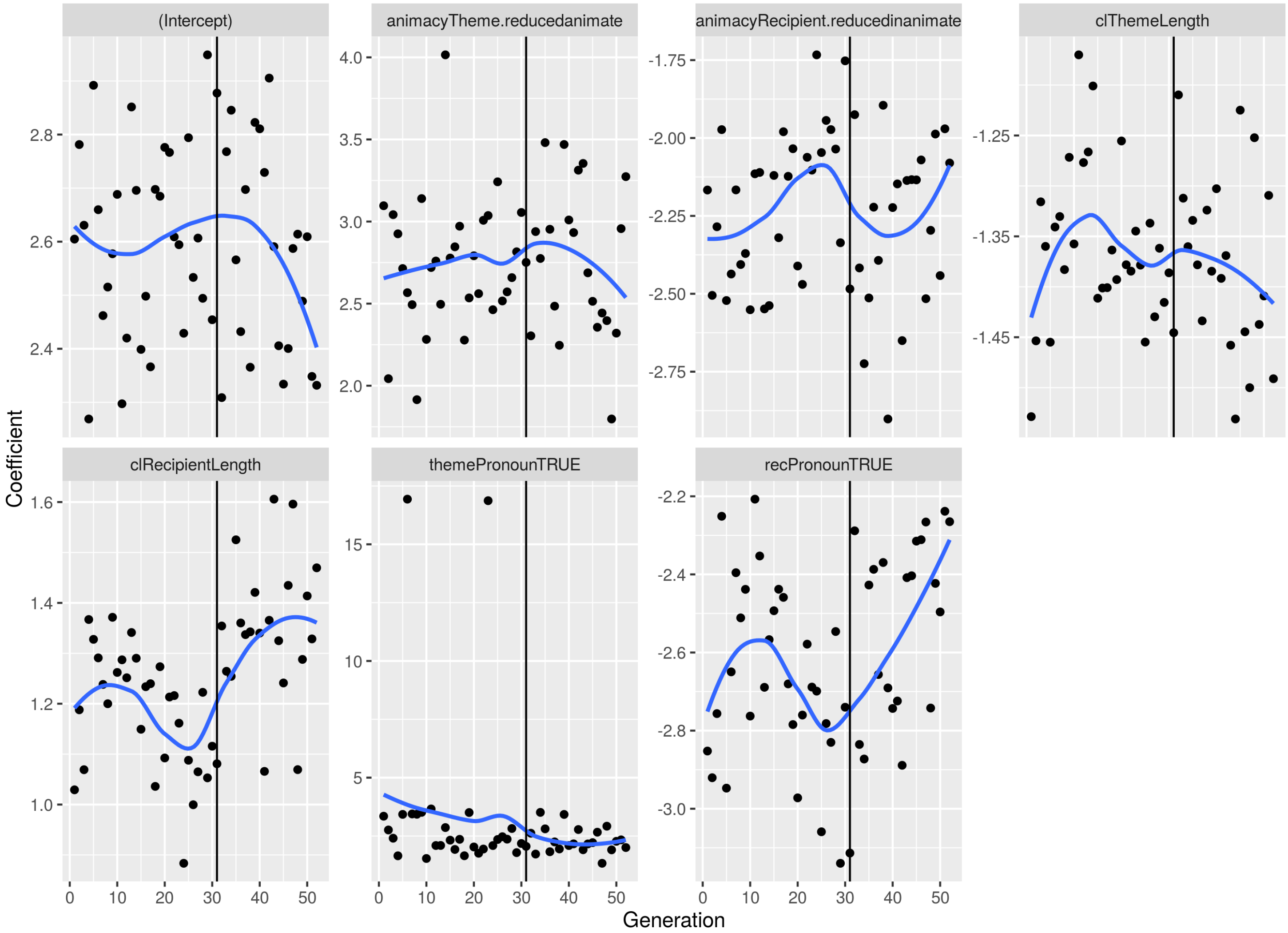
- probabilistic grammatical change: constraint simply changes.
  - drift or semantic bleaching or …
- constraint only appears to change, category changes
- constraint only appears to change, distribution changes
- could interact (distribution change facilitates category change …)
  - highly complex structures of cause and effect
- Let's attempt to simulate

Animacy changes

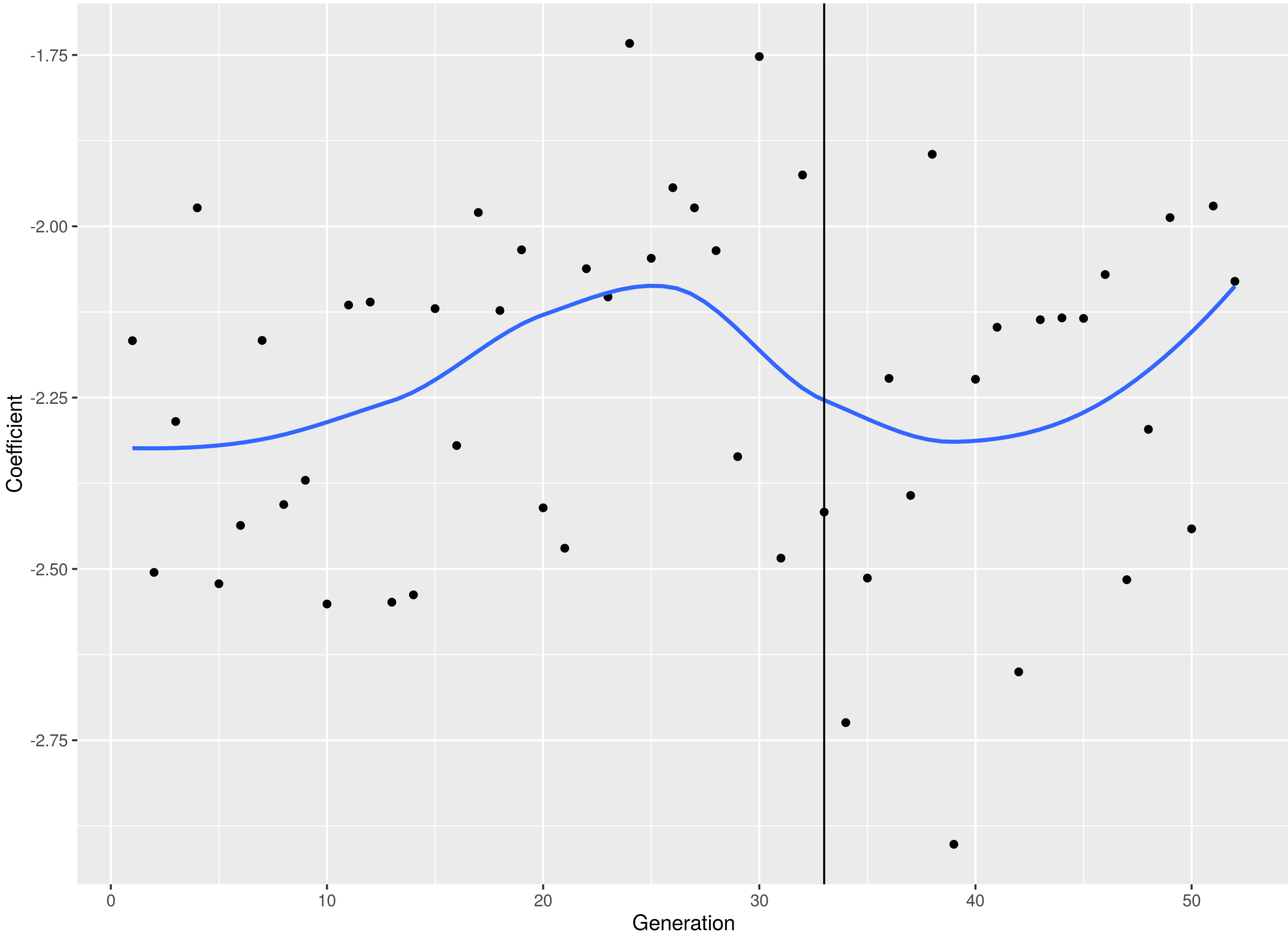
My current formalization is as follows:

- animacy is really a gradient phenomenon, with continuous values ranging from 0 (prototypical inanimate) to 1 (prototypical animate).
- in our five-way classification
    - animate is uniformly distributed in 0.9-1.0; inanimates 0.0-0.1
    - the other three were roughly chosen based on the size of the coefficients in a simple regression model: collective 0.2-0.3, locative 0.3-0.5, temporal 0.0-0.1
- at some point, the context changes, and collectives are now selected from the range 0.3-0.4
- how does this change the simulation
- crucially, how does this affect modeling using the original classification?

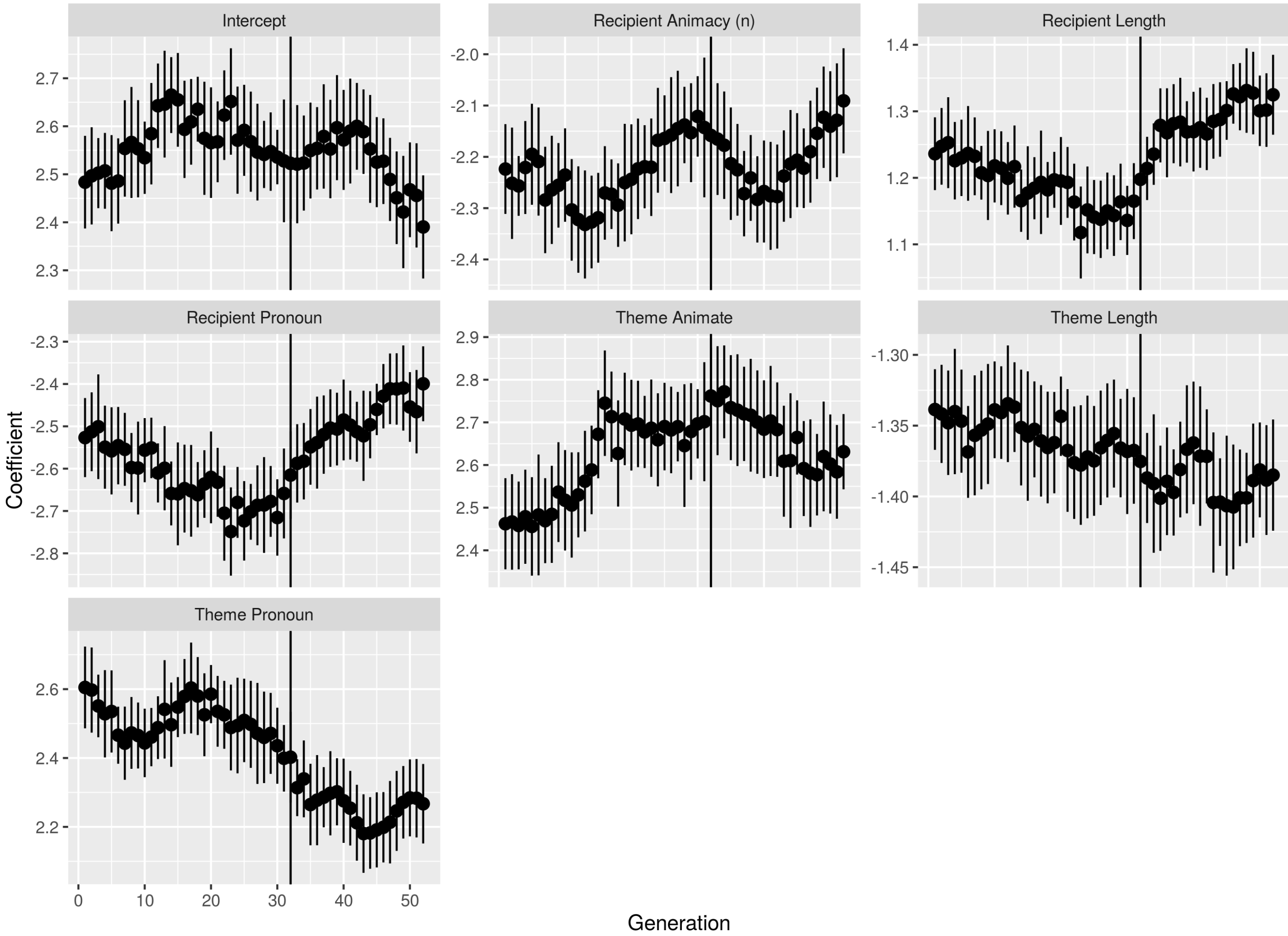
Visualization (categorical)



Zoom in



The simulated models



Discussion

The change did have the predicted effect: changing the numeric predictor did affect the categorical predictor.

- Would we find this change?
- As before: Not reliably.
- This likely depends on the size of the change in the category

Ideas

- The social structure (completely random interactions) is completely implausible. What would happen if we imlement more reasonable social networks
- The simulated speakers lead rather boring lives. Maybe the children should be allowed to play together?
- It would be great to have actual new data for the models, instead of constantly reusing the same data.
  - is there a good way to create (distributionally) realistic data sets? do they even have to be?
- Does the type of model (glmer, glm, Bayesian glm, random forest?)

Summary

- Probabilistic drift does happen in a vacuum
- It's unlikely to be the only story
- Models of communities through communities of models?

Thank you!