A Population Dynamic Mechanism for the Constant Rate Effect (and Beyond)

Henri Kauhanen & George Walkden

The University of Manchester

22nd ICHL, Naples
31 July 2015

Introduction

Aims of this talk:

- Outline some problems with the Constant Rate Effect (CRE) as standardly described
- Provide a mechanism for the CRE using a modified version of the learner in Yang (2000)
- Show that this mechanism has a neat consequence regarding the time separation of curves

Part I

The Constant Rate Effect

S-curves and the logistic equation

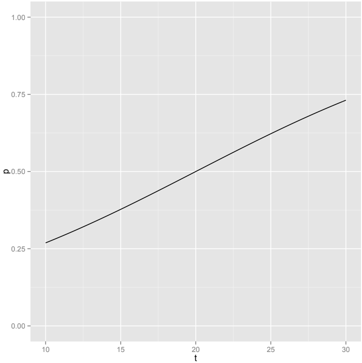
- Altmann et al. (1983), Kroch (1989): change is governed by the logistic equation.

$$p(t) = \frac{1}{1 + e^{-s(t-k)}} \tag{1}$$

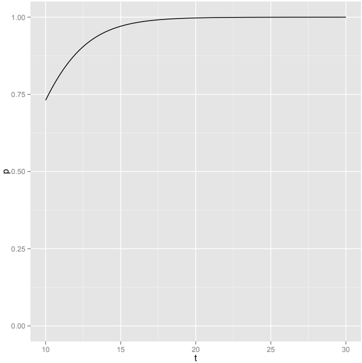
- The s parameter (**slope**) gives the overall steepness of the curve.
- The k parameter (**intercept**) tells us where, in time, the curve is situated.

Varying the parameters

Varying the slope s



Varying the intercept k

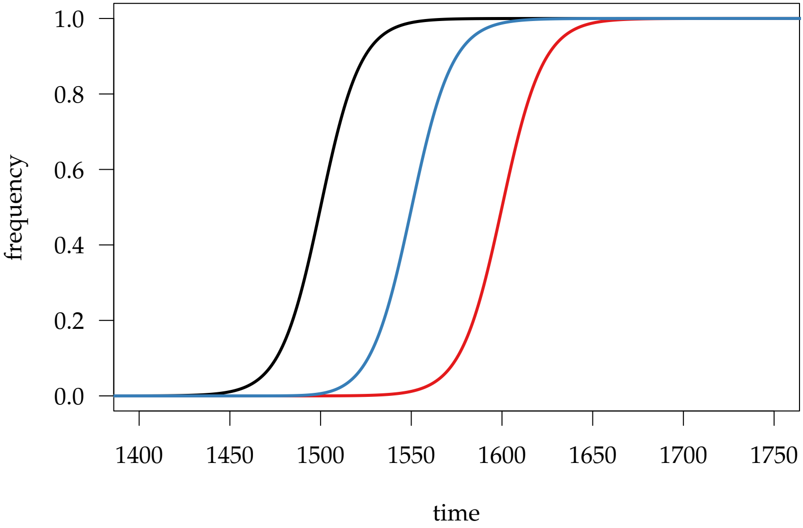


Constant Rate Effect (CRE)

"[W]hen one grammatical option replaces another with which it is in competition across a set of linguistic contexts, the rate of replacement, properly measured, is the same in all of them." (Kroch, 1989, 200)

- In terms of the logistic equation: the *s* values (slopes) are the same, but the *k* values (intercepts) may differ.
- Since Kroch (1989), many empirical studies have found this sort of situation.
- Key intuition (Kroch 1989): a single syntactic parameter governs all contexts. Some contexts may favour the new form, others disfavour it, but those (dis)favouring effects are **constant**.
- Relies on individual speakers having access to multiple, probabilistically-weighted *competing grammars*.

Different linguistic contexts

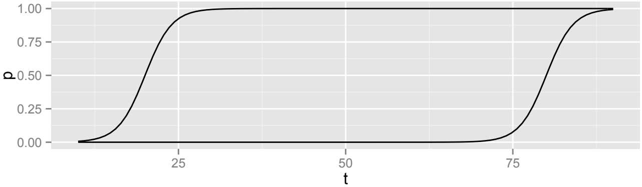


Problems, 1: no mechanism

- Kroch (1989, 204) adopts the logistic equation because of its use in biology, but admits that it does not follow from anything.
- Since then, Yang (2000) has derived the logistic equation from his model of syntactic acquisition and change (on which more later).
- But no population-dynamic mechanism has so far been presented **to account for the CRE**: how exactly are the contexts tied to an underlying change?

Problems, 2: time separation

- For a CRE, the *s* (slope) values must not differ significantly. But what about large differences in the intercept (*k*)?



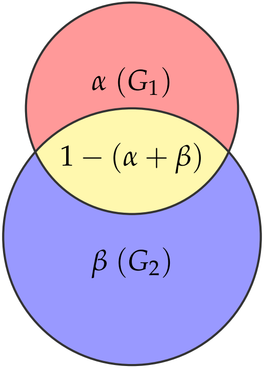
- If the change only gets off the ground in one context 50 years after it has run to completion in another context (or 500, or 5,000,000), how plausible is it that we are dealing with reflexes of a single parametric change? (Think of the acquirer!)
- But this would still qualify as a case of a CRE.

Part II

The model

Yang (2000)

Assume grammar G_1 with probability p, and G_2 with $q = 1 - p$.
Parameter setting with linear reward–penalty learning.



PLD distribution:

	A	B	C
G_1	αp	0	$(1-\alpha)p$
G_2	0	βq	$(1-\beta)q$

Evolution:

$$q' = \frac{\beta q}{\alpha p + \beta q} = \frac{1}{1 + \rho \frac{1-q}{q}} \qquad (2)$$

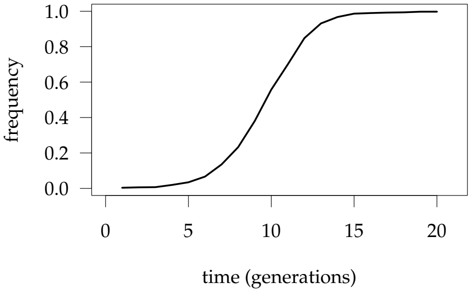
with $\rho = \alpha/\beta$. Holds in the limit of infinite maturation time ($N \to \infty$).

Yang (2000)

Solution (in the limit $N \to \infty$):

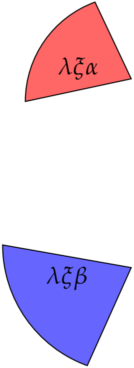
$$q(t) = \frac{1}{1 + e^{\log(\rho)\left(t + \frac{1}{\log(\rho)} \log\left(\frac{1}{q(0)} - 1\right)\right)}} \tag{3}$$

i.e. logistic with $s = -\log(\rho)$ and $k = -\frac{1}{\log(\rho)} \log\left(\frac{1}{q(0)} - 1\right)$.



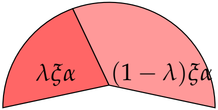
Extension

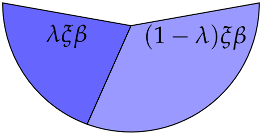
Now introduce two contexts, 1 and 2, for both G_1 and G_2.



Extension

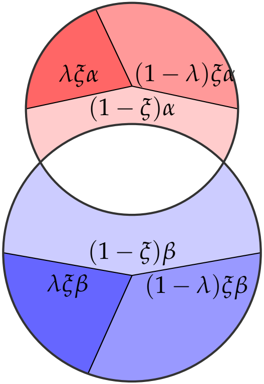
Now introduce two contexts, 1 and 2, for both G_1 and G_2.





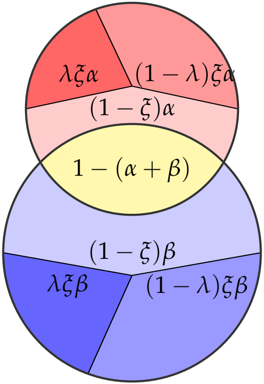
Extension

Now introduce two contexts, 1 and 2, for both G_1 and G_2.

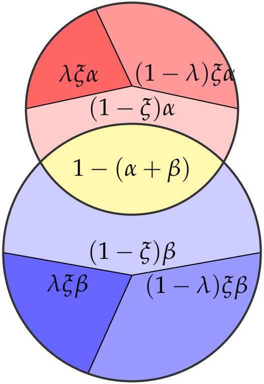


Extension

Now introduce two contexts, 1 and 2, for both G_1 and G_2.



Extension



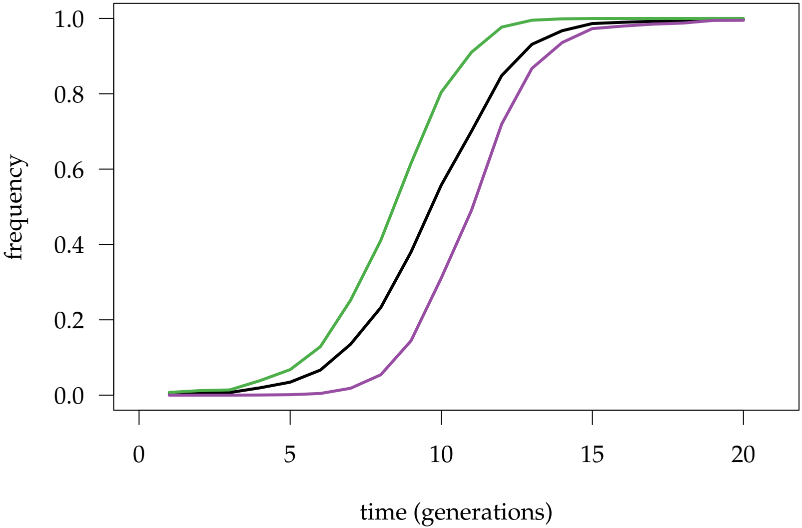
Suppose that learner

1. first sets parameters in the Yangian fashion, arriving at values for p and q (*acquisition phase*)
2. but then applies different (but **constant**) production biases to G_1 and G_2 in different contexts (*production phase*)

Let w_j = bias for G_2 (over G_1) in context j. Then

1. probability to use G_2 in context j is $q + w_j qp$
2. probability to use G_1 in context j is $p - w_j pq$

Constant Rate Effect



Approximation

It can be shown that we have the evolution equation

$$q' = \left(1 + \rho \frac{(1-q)(1-\xi q(\lambda w_1 + (1-\lambda) w_2))}{q(\xi(1-q)(\lambda w_1 + (1-\lambda) w_2) + 1)}\right)^{-1} \qquad (4)$$

in the limit $N \to \infty$.

General solution not known, but reduces to the Yangian (2)

1. if $\xi = 0$ (no feedback from biases to next generation)
2. if $\lambda w_1 = -(1-\lambda) w_2$ (context biases cancel each other)

We can then approximate $q(t)$ with a logistic and derive the context curves from it in closed form in order to test the model against data.

Time separations

The logistic approximation also gives us a nice handle to the following result:

Theorem (The time separation theorem)

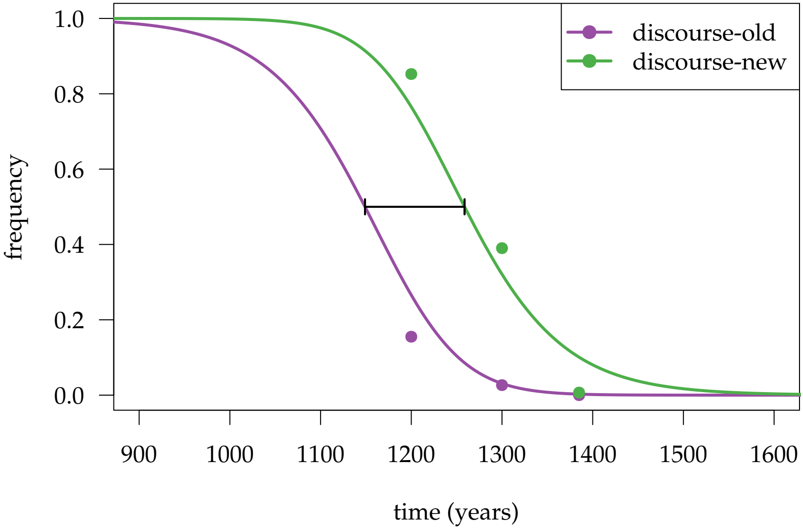
For any two contexts of one underlying change proceeding at rate s, the maximal time separation at tipping points is

$$-\frac{2}{s}\log(\sqrt{2}-1)$$

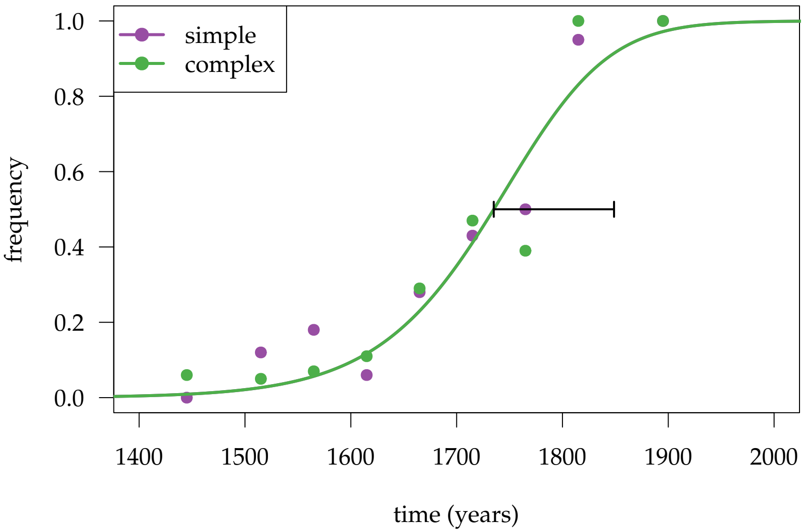
or equivalently

$$\frac{2}{\log(\rho)}\log(\sqrt{2}-1).$$

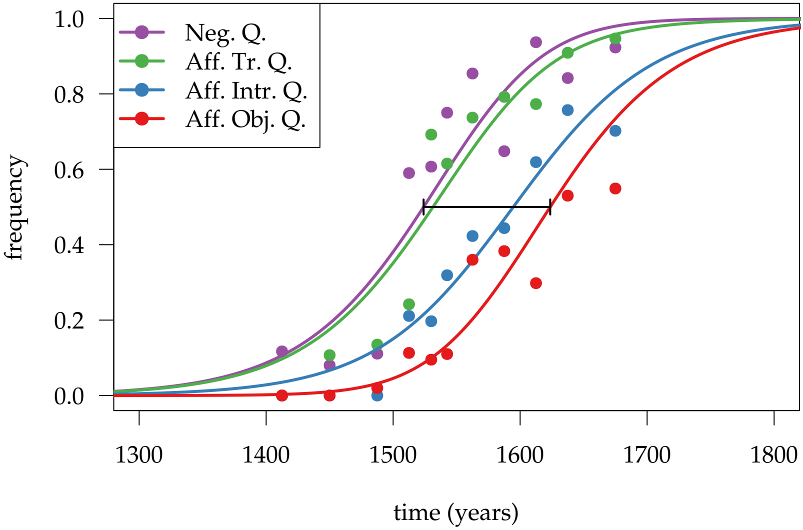
Evaluation: English Jespersen cycle (Wallage, 2013)



Evaluation: Position of T in Yiddish (Santorini, 1993)

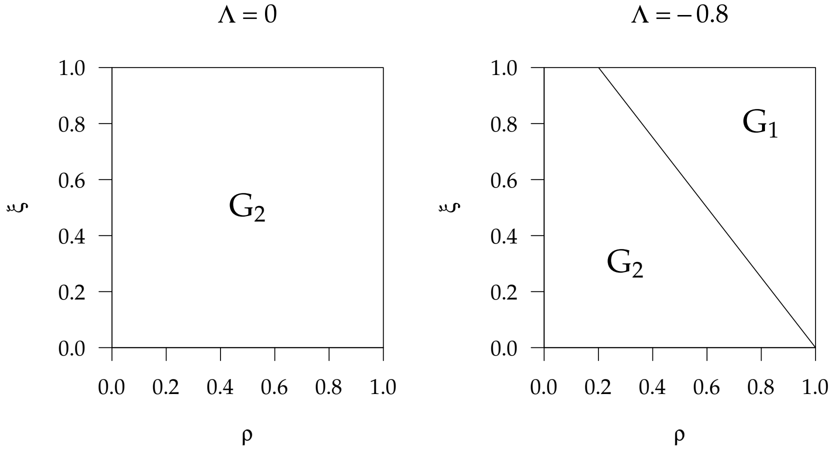


Evaluation: *do*-support in English (Kroch, 1989)



Advantage vs. bias

Let $\Lambda = \lambda w_1 + (1 - \lambda) w_2$. A bifurcation:



Part III

Conclusions

Conclusions

- The CRE is a robust empirical finding in historical syntax, and the intuition is an important one, but the mathematics used to investigate it (simple logistic which may differ in k but not s) raises as many questions as it answers.
- We have presented a **mechanism** for the CRE, derived from simple ingredients:
  - The variational learner of Yang (2000)
  - Iterated learning (in an infinite, well-mixing population...)
  - Production biases with fixed weights
- The model gives a good fit for a number of case studies.
- A corollary of this is that an upper limit can be placed on the time separation of context curves (traditionally k).
- This provides a new, restrictive, and motivated way of continuing the search for parametric clustering!

Acknowledgements

Ricardo Bermúdez-Otero
(for astute comments)

Deepthi Gopal
(for algebra)

References

Altmann, G., von Buttlar, H., Rott, W. & Strauss, U. (1983). A law of change in language. In B. Brainerd (ed.), *Historical linguistics*, 104–115. Bochum: Brockmeyer.

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

Santorini, B. (1993). The rate of phrase structure change in the history of Yiddish. *Language Variation and Change*, 5: 257–283.

Wallage, P. (2013). Functional differentiation and grammatical competition in the English Jespersen Cycle. *Journal of Historical Syntax*, 2: 1–25.

Yang, C. (2000). Internal and external forces in language change. *Language Variation and Change*, 12: 231–250.