Semantic change

Population-level dynamics and very slow change

Rob Truswell

EGG, Brno, August 2015

Introduction

- Take-home messages from yesterday:
  - Semantic theory has a range of options for describing discrete semantic changes (and, to a lesser extent, gradual changes).
  - Discrete semantic changes need not have catastrophic consequences for comprehension.
  - Cumulative non-catastrophic discrete changes can have very large effects (e.g. Jespersen's cycle).
- Today's problem: grammatical change (including semantic change) doesn't look discrete.
- The solution (here as elsewhere): grammar competition, or competing specifications of form–function correspondences.

S-curves

- S-curves are everywhere. Grammar change very often looks like Fig.1.

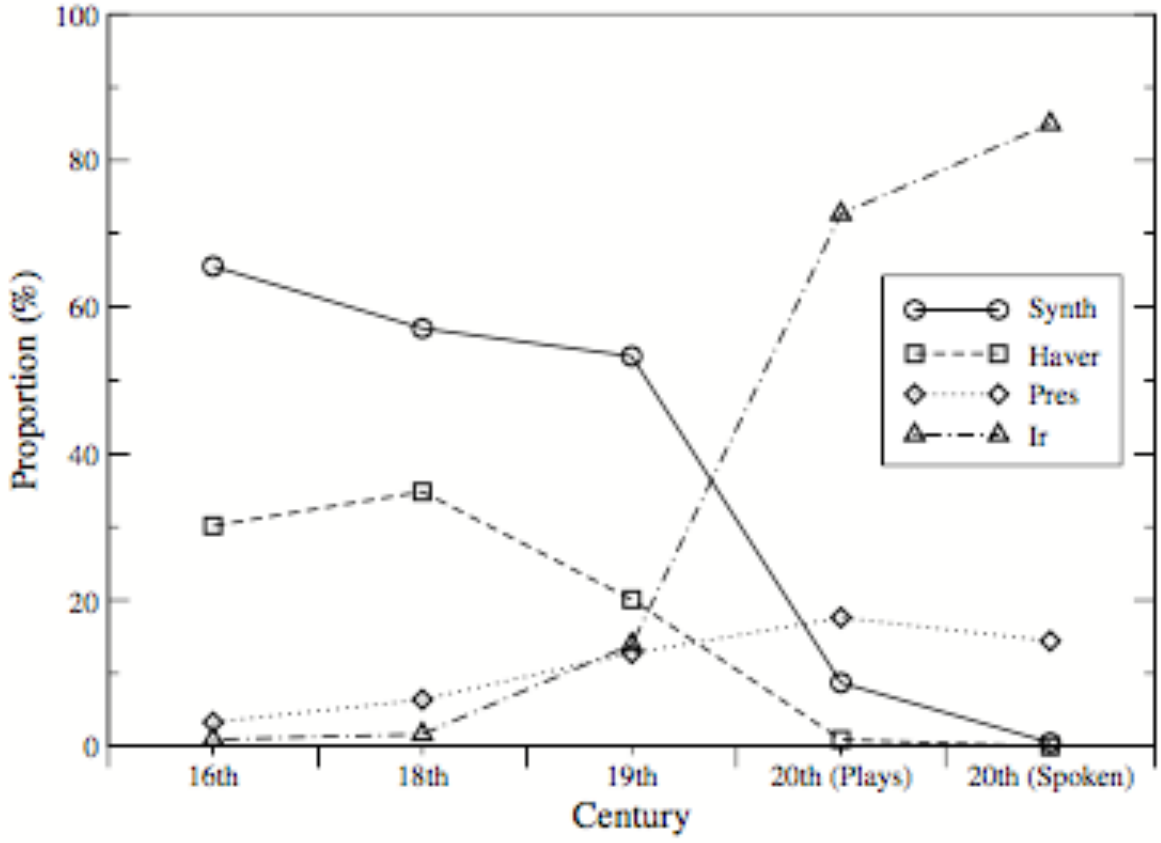


Figure 1 : Future markers in Brazilian Portuguese, from Poplack & Malvar (2007) via Blythe & Croft (2012)

S-curves

- We've known this for a long time.

> *The process of change in the community would most probably be represented by an S-curve. The rate of change would probably be slow at first, appearing in the speech of innovators, or more likely young children; become relatively rapid as these young people become the agents of differential reinforcement; and taper off as fewer and fewer older and more marginal individuals remain to continue the old forms.*
>
> *(Osgood & Sebeok 1954: 155)*

- See also Weinreich et al. (1968), Bailey (1973), Kroch (1989), Yang (2002), Niyogi (2006), Blythe & Croft (2012), ...

Why S?

- The common understanding of the derivation of S-curves is already implicit in Osgood & Sebeok (1954). You need:
  - One (diachronically stable) function, F,
  - Two competing forms, *Old* and *New*.
- As more people use *New* to do F, evidence that you should use *New* to do F increases and evidence that you should use *Old* to do F recedes.
- A simple equation can describe this shape:

$$\frac{p}{1-p} = e^{k+st} \tag{1}$$

(where p is the frequency of one of the the two variants).

- Equivalently:

$$p = \frac{e^{k+st}}{1+e^{k+st}} \tag{2}$$

- Two parameters:
  1. s describes the rate of change (higher = faster);
  2. k describes the intercept.

Varying s and k

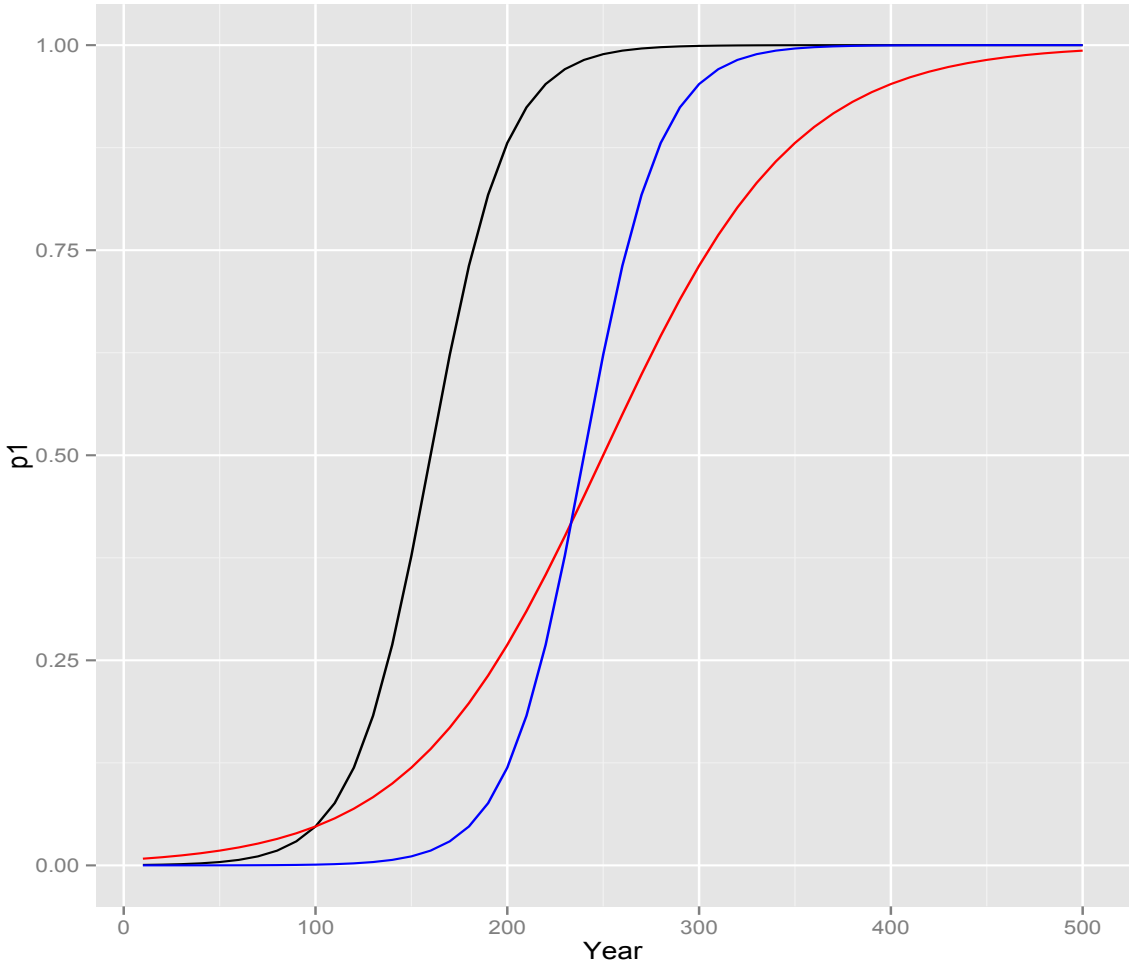


Figure 2 : Logistic functions, $k = -8, s = 0.05$ (black); $k = -12, s = 0.05$ (blue); $k = -5, s = 0.02$ (red)

S-curves and grammar competition

- Kroch (1989): S-curves reflect competition between grammars (an instance of the general claim that S-curves reflect competition between variants in a population).
- Understandable reluctance to countenance competition between full grammars.
  - Imagine I'm somehow involved in n changes right now; does that really mean I have 2^n grammars of English in my head?!.
- This becomes more palatable if we remember that grammars are just bundles of lexical items and some invariant ways of combining them (Borer 1983, Kroch 1994). Two immediate advantages of this view:
  1. It gives a robust theory of grammatical change: CCG, CxG, Minimalism, TAG, etc. are all lexicalist in this sense and so all amenable to conceptualization of grammar competition as lexical competition.
  2. It's more straightforwardly extensible to semantics: grammar competition is essentially competition among bits of specifications of lexical items which will be input to combinatorial systems. Compositional semantics matches that description just as well as syntax.

S-curves and grammar competition

- So S-curves reflect competition between lexical items. E.g. the introduction of *do*-support reflects competition between $T_{[+V]}$ and $T_{\emptyset}$. Or whatever.
- Yang (2002): speakers have multiple such lexical items (because why wouldn't a grammar contain both $T_{[+V]}$ and $T_{\emptyset}$?), some of which are more "central" to a grammar than others. To a first approximation, we can assign a weight p to each lexical item ($0 \leq p \leq 1$), reflecting correspondences between observed linguistic data and the generative capacity of grammars containing that lexical item.
- Yang: s reflects the extent to which evidence favours the incoming grammar.
- k reflects the effect of contextual factors on choice of lexical items (as in classical sociolinguistic variable rule analysis, Kroch 1989). This is the Constant Rate Effect.

Kroch's *do*-support CRE

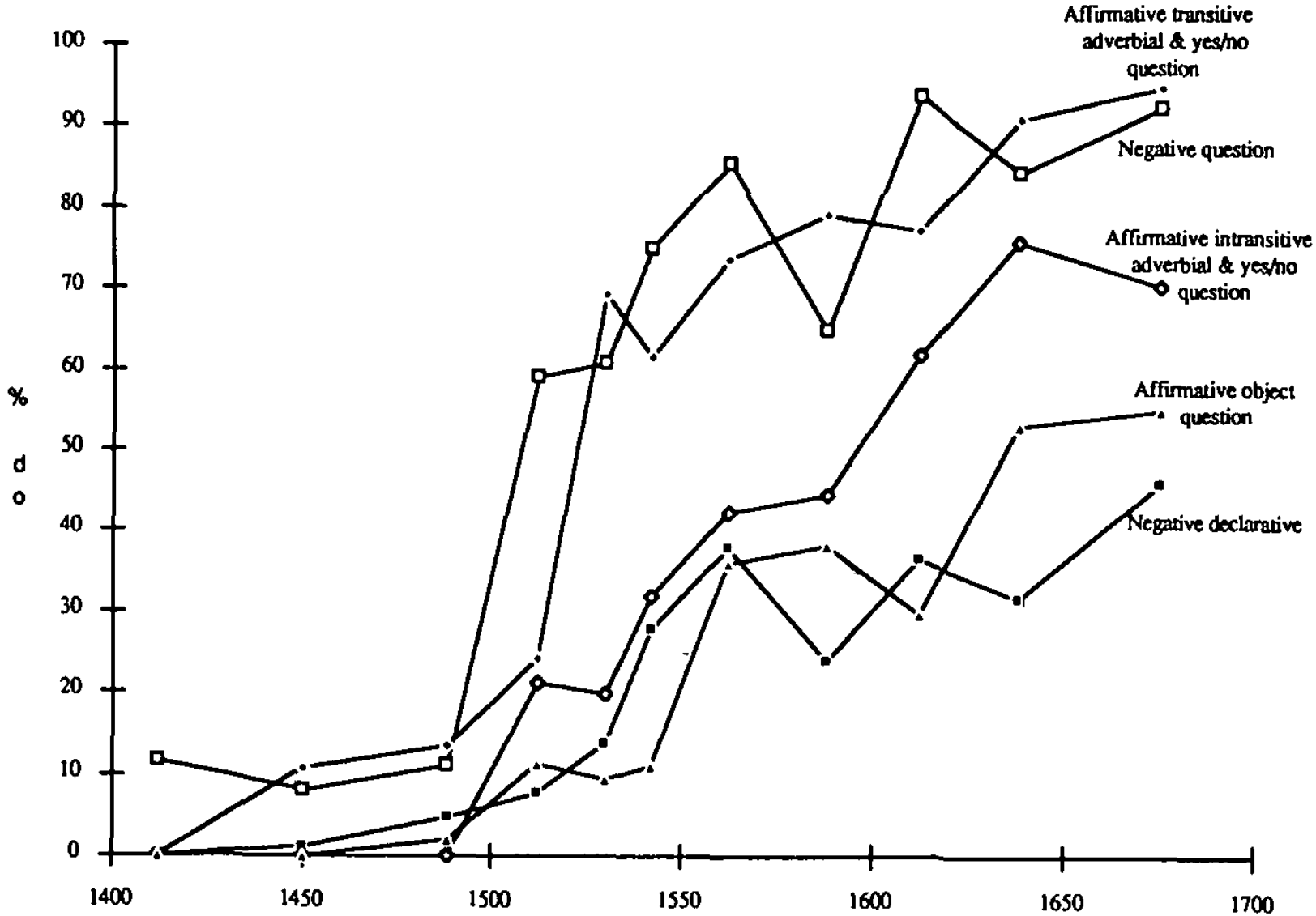


FIGURE 6: The rise of periphrastic *do* (adapted from Ellegård, 1953).

Kroch's *do*-support CRE

	Slope	Intercept
Neg.Decl	3.74	−8.33
Neg.Qn	3.45	−5.57
Aff.Trans.Adv/yn.Qn	3.62	−6.58
Aff.Intr.Adv/yn.Qn	3.77	−8.08
Aff.*wh*-obj.Qn	4.01	−9.26

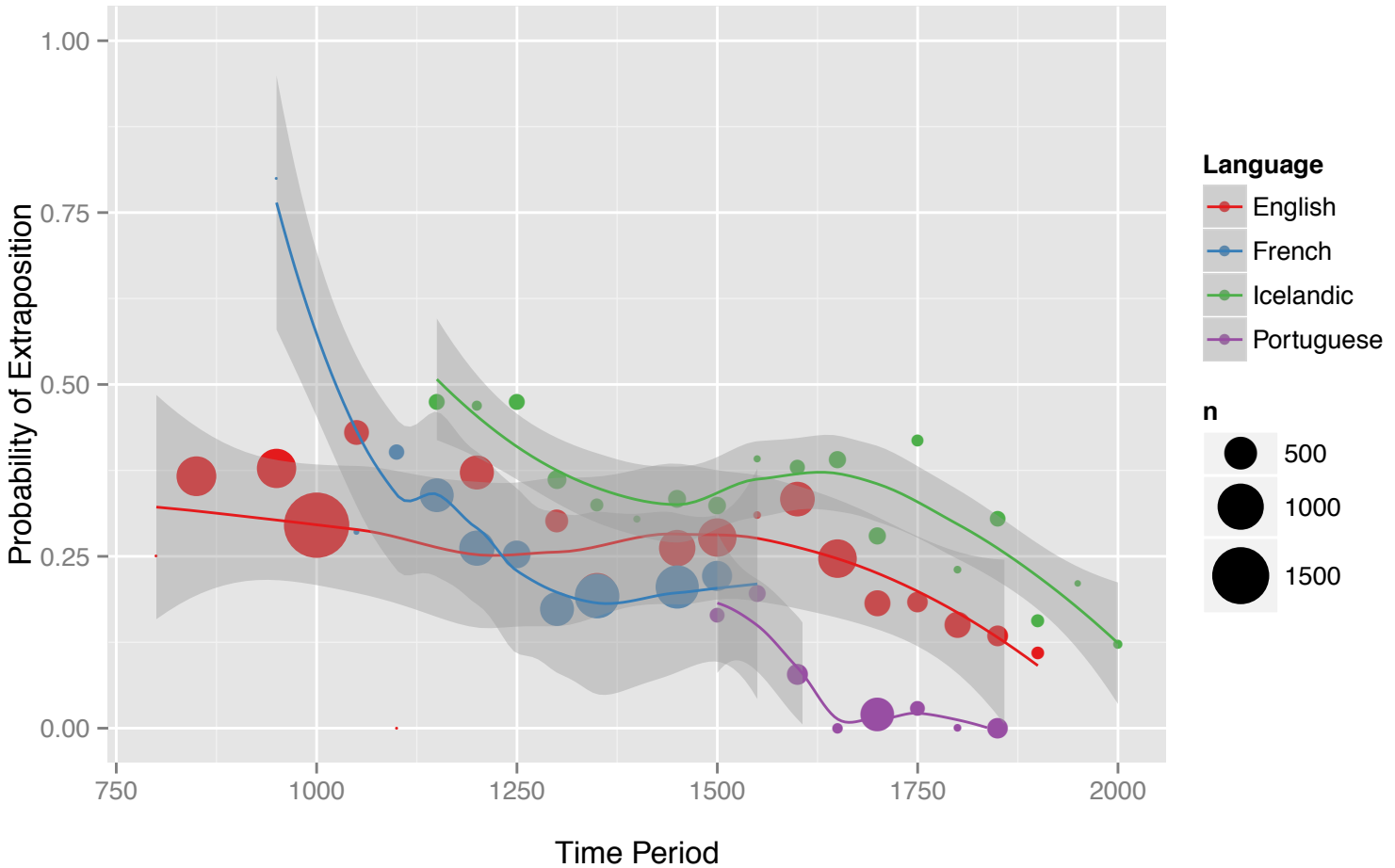
Table 1 : Regression parameters for periphrastic *do* in different contexts over time, based on Kroch 1989, table 4

Discussion

- The Constant Rate Effect is *beautiful*.
- It legitimizes the idea that the introduction of *do*-support is a single change, despite the fact that that change took place at different times in different environments.
- It gives us a way of relating the fact that change is local and near-instantaneous (occurring within an individual, during that individual's lifespan and possibly strictly during acquisition), without contradiction, to the fact that change is a very slow, population-level phenomenon.
- Change driven by grammar competition can be extremely slow. Figure 3 shows a loss of extraposition in four Romance and Germanic languages over 1,250 years, with a Constant Rate Effect between English and Icelandic.

Loss of extraposition

Four Languages (Subj Ex), over time



A CRE without competition

- However, it is not always plausible to reduce Constant Rate Effects to competition between forms, holding a function constant.
- An example from the introduction of headed *wh*-relatives in Middle English: PP-relatives were the first to emerge, followed by argumental relatives. When PP-relatives emerged, there was no competing strategy for relativizing a PP. Argumental *wh*-relatives were competing with *that* or $\emptyset$. Nevertheless, there is a Constant Rate Effect.
- Moreover, this is not too surprising: the introduction of headed *wh*-relatives is and isn't a change in the same senses in which the introduction of *do*-support is and isn't a change.

A CRE without competition

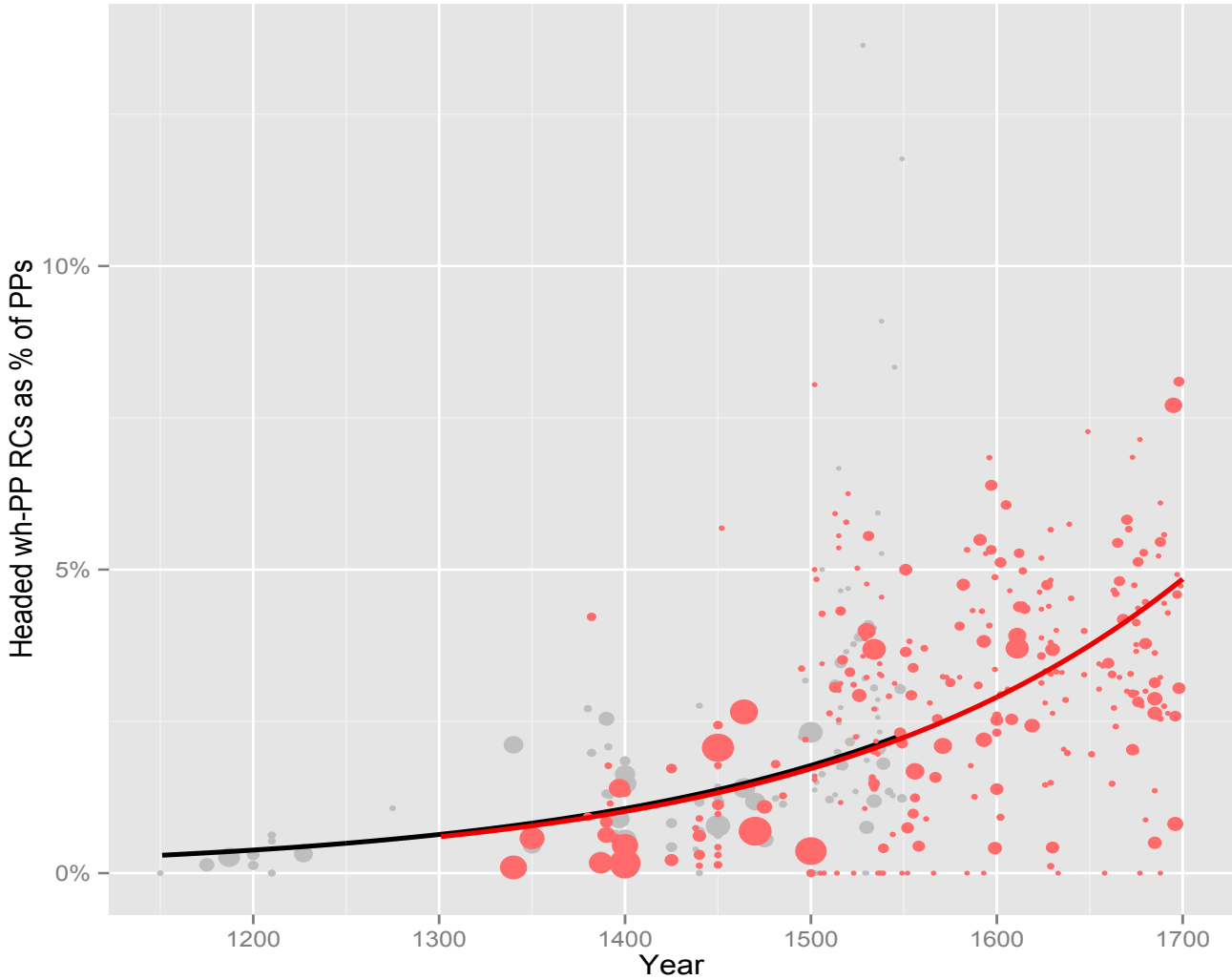


Figure 4 : The rise of headed *wh*-RCs with PP (black) and argumental (red) gaps

Different types of competition

- So we need a way to articulate a logic of S-curves and competition without relying on competition between functionally equivalent forms.
- One possibility: change can also reflect competition between different specifications of the function (i.e. syntactic features and/or denotations) of a given form.
- Consider *do*-support again.

	Decl	Int	"Cause"
∅	X	X	
Do			X
Cause			

→

	Decl	Int	"Cause"
∅	X		
Do		X	
Cause			X

Neither the form *do* nor its function was new. The assignment of forms to functions is new.

- With classic examples (biological evolution, sound change), the "functions" (genes, phonemes) are often taken as given, with a range of variants for each slot.
- Less obvious that that makes sense for syntax and semantics.

Acquisition: "But what does it do?"

- Shipley et al. (1969): children show sensitivity to the distribution of function words before they produce them *or understand them*.
- So recognizing that *the* is a word (with a certain distribution) precedes knowledge of what *the* does.
- Typical models (e.g. Bayesian) would assume a distribution over a range of hypotheses (or candidate denotations) for *the*, with incoming evidence provoking reweighting of those denotations.
- Those denotations are in competition.

Two possible examples

Where

- Earliest headed *wh*-relatives often feature "R-pronoun" forms (e.g. *whereby*).
- Those forms did not exist in OE.
- When they emerge, they are used roughly equally in interrogatives and relatives (but most texts don't use them at all).
- "What does *where* do? Is it a locative pronoun? Is it an R-pronoun?"

Which

- Is *which* inherently definite, or is its definiteness in FRs a result of null δ (Caponigro 2003)?
- Consequences for emergence of headed *wh*-relatives (last lecture).

The fairy tale

1. Start by associating form F_0 with denotation D_0.
2. A learner associates F_0 with D_0 with high probability, and with D_1 with low probability.
3. The learner produces F_0 paired with D_1 occasionally.
4. This increases the evidence for F_0 paired with D_1.
5. Iterating this gives F_0 paired with D_1 (possibly and D_0).

Why wouldn't this always happen?

- That fairy tale seems rigged: mislearning can help introduce new denotations, nothing helps lose old denotations.
- Various factors can help redress the balance:
  1. Mutual exclusivity / principle of contrast (Slobin 1985, Markman & Wachtel 1988, Clark 1993): Learners are biased towards 1–1 form–meaning mappings.
  2. Various learning algorithms (again, e.g. many Bayesian theories) prefer less general grammars.
- So we would expect that mislearning should mainly operate among low-frequency forms.
- That seems accurate (Naro & Lemle 1976 on low-accessibility endogenous change vs. high-accessibility borrowings).

Imperfect cadence

- Turns out that things are not so straightforward.
  Simple-minded simulations of the above ideas give results like this, at best.

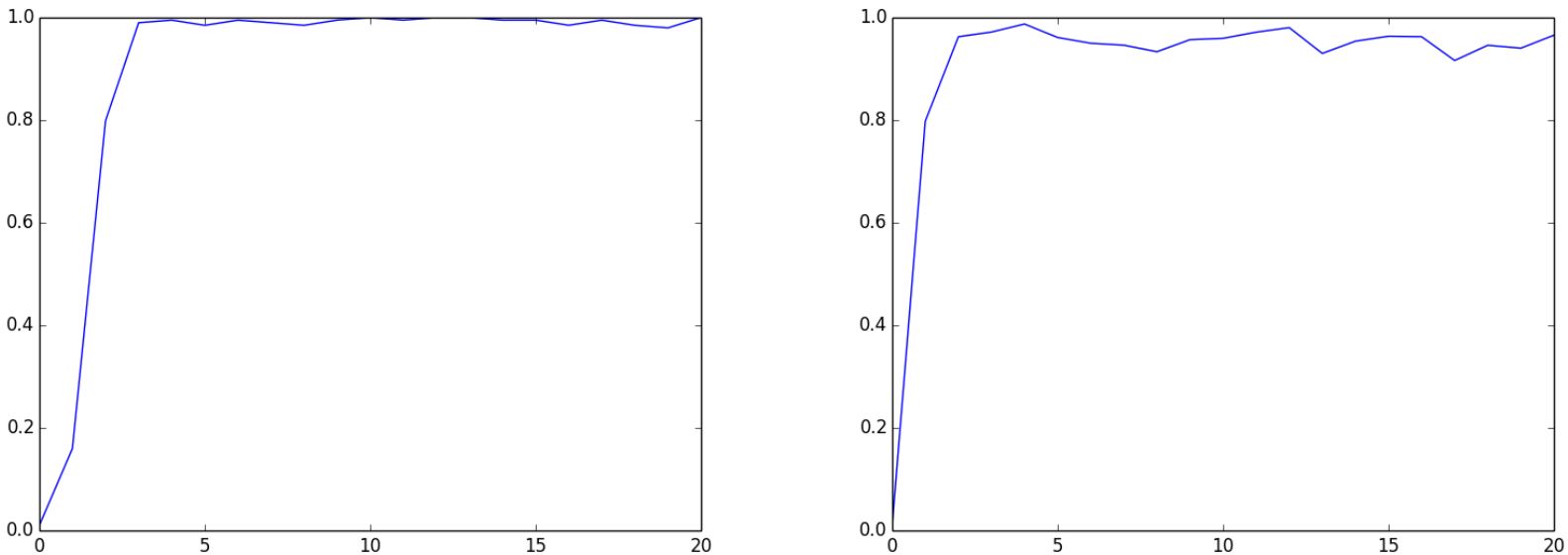


Figure 5 : Spread of a new function for an old form, Bayesian learner (left), and linear reward–penalty learner following Yang (2002) (right)

Imperfect cadence

- Other runs often have the forms disappearing instantly. If a run were to give satisfying S-curve dynamics, it would reflect minuscule tweaking of parameters, not robust results about learning and change.
- This may partly reflect a difference in the dynamics of the two ways of construing competition (thanks to Simon Kirby and Richard Blythe for discussion):
  - The classic construal of competition reduces to "what form do I expect to use in this situation".
  - The alternative suggested here reduces to "When do I expect to use this word".
  - So the "situations" counted in the two cases are not the same.
  - (To a first approximation, one is given by the "comunicative situtations" presented by the world at large; one is given by the structure of the grammar).
- We're still working on better models. Watch this space...

Key ideas

- For all its complexity and gradience, there are clear advantages to maintaining the formal view of semantics as part of grammar, and of grammar as a system which manipulates discrete objects.
- This instantly allows us to borrow large amounts of theoretical machinery from other types of grammar change.
- The major challenge is to account for slow, gradual change.
- The diachrony of *wh*-relatives seems like a particularly recalcitrant example of that, because it resists reduction to classical notions like grammaticalization.
- Synchronic formal semantics has clarified the nature of the change, but the account will be incomplete until a plausible model of the dynamics of the change is found.

Bibliography I

Bailey, C.-J. (1973). *Variation and Linguistic Theory*. Arlington, VA: Center for Applied Linguistics.

Blythe, R. & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, *88*, 269–304.

Borer, H. (1983). *Parametric Syntax: Case Studies in Semitic and Romance Languages*. Dordrecht: Foris.

Caponigro, I. (2003). *Free not to Ask: On the Semantics of Free Relatives and Wh-words Cross-linguistically*. PhD thesis, University of California, Los Angeles, CA.

Clark, E. (1993). Cambridge: Cambridge University Press.

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

Kroch, A. (1994). Morphosyntactic variation. In *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society*, (pp. 180–201).

Markman, E. & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

Naro, A. & Lemle, M. (1976). Syntactic diffusion. In *Proceedings of the Chicago Linguistic Society*, (pp. 221–240).

Niyogi, P. (2006). *The Computational Nature of Language Learning and Evolution*. Cambridge, MA: MIT Press.

Bibliography II

Osgood, C. & Sebeok, T. (1954). *Psycholinguistics: A Survey of Theory and Research Problems*. Baltimore, MD: Waverly Press.

Poplack, S. & Malvar, E. (2007). Modeling linguistic change: The past and the present of the future in Brazilian Portuguese. *Probus*, *19*, 121–169.

Shipley, E., Smith, C., & Gleitman, L. (1969). A study in the acquisition of language: Free responses to commands. *Language*, *45*, 322–342.

Slobin, D. (1985).

Wallenberg, J. (2015). Science of the experimentally possible: Very slow change and language technology. Paper presented at Research Links workshop, Campinas.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for Historical Linguistics* (pp. 95–188). Austin, TX: University of Texas Press.

Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.