Signal or noise? Biases, interactions and the Constant Rate Effect

Any linguist trying to access I-language through the lens of production data – including all historical linguists for whom I-language is the desired object of study – requires at least an implicit theory of how reflexes of internalized grammatical knowledge manifest themselves in corpus data. Simple inferences can be drawn without dwelling on the details: for instance, if a sentence type is attested multiple times in reliable sources, we can infer that it is grammatical. However, given the paucity of our data, it is desirable to go as far beyond these basic inferences as we can – and to do so we need a linking theory that must ultimately include how grammatical knowledge interacts with non-I-language factors of production that are uncontroversially involved in skewing the production data that we see.

A well-known attempt to make sophisticated inferences about I-language on the basis of usage data is Kroch’s (1989) Constant Rate Hypothesis. This hypothesis is based on the intuition that a structure that differs in frequency across two or more contexts may nevertheless reflect the operation of a single grammatical rule or the result of a single parameter setting. In situations of change, when two grammatical options compete and one replaces the other over time, the Constant Rate Hypothesis predicts that rate of change will be the same in all contexts. To date, this hypothesis has been studied in a number of languages and data sets and has accumulated enough support to be referred to as the Constant Rate Effect or CRE (e.g. Pintzuk 2003). Despite this wealth of empirical studies, however, the Constant Rate Hypothesis has never been explicated formally in a detailed mathematical model of change that takes both grammatical competition and contextual effects into account.

The standard way of detecting CREs in corpus data is to fit a number of independent logistic curves, one per each context of interest: if the slope parameters of these curves turn out to be more or less similar, a CRE is diagnosed (see, e.g., Kroch 1989; Santorini 1993; Taylor 1994; Pintzuk 1995; Postma 2010). Rather than stemming from any first principles of the dynamics of language change, this procedure is data-driven and suffers from three problems:

(P1) Fitting a number of independent logistic curves to a set of contexts leaves variation in the time dimension entirely unexplained: as long as two logistics agree in their slope parameter, the difference in the intercept parameters may be arbitrarily large, making it possible to establish a ‘CRE’ across two contexts where the change goes to completion in one before it even takes off in another.

(P2) The slope parameters of two logistics may be similar for reasons other than underlying grammatical unity: ‘CREs’ can be demonstrated across languages (Wallenberg 2015), geographical areas (Willis 2015) and gender (Corley 2014).

(P3) There exists no non-problematic statistical procedure for evaluating whether a given difference in the slopes of two (or more) logistic curves is significantly different, and thus more generally, there exists no non-problematic litmus test for the CRE (Paolillo 2011).

Together, these problems imply that the standard operationalization of CREs is not sufficient for assuming that a single underlying change has occurred, undermining Kroch’s (1989) original motivation for using CREs as a link between corpus data and abstract, underlying grammatical changes.

In this talk, we aim to overcome the problems and to shed light on the nature of causation in language change by introducing a mathematical model of the CRE that is more tightly constrained, and therefore makes stronger (more restricted) empirical predictions, than the standard formulation. Starting with Yang’s (2000) mathematical model of grammar competition, we augment the model with production biases across an arbitrary number of linguistic contexts, and show that this extension of Yang’s framework naturally gives rise to the CRE both analytically and in computer simulations. We additionally probe the long-term dynamics of this system by conducting a full bifurcation analysis of the two-grammar case, working out the exact quantitative relation between grammatical advantage (in the sense of Yang 2000) and contextual biases, and point out how the model can be generalized beyond Yang’s framework as long as underlying logisticity can be assumed.

The model we propose solves problem (P1) above: it is a theorem of our model that the time separation between the propagation curves of an innovatory grammatical parameter in two different contexts always necessarily has a finite upper bound which is inversely proportional to the rate of the underlying change (the *Time Separation Theorem*). With this result, we are able to tell actual CREs apart from
pseudo-CREs – developments that on the surface proceed at similar rates but that are not unified by an underlying change (Fig. 1). We thus show that a more constrained, theoretically motivated model of the CRE can both fit historical data no worse than a less constrained one, and diagnose certain pseudo-CREs which for the standard operationalization remain false positives.

Our answer to problems (P2) and (P3) is essentially negative. We argue that two reflexes of one and the same underlying change may actually propagate at slightly different rates, provided that the production biases for the two contexts in question have suitable magnitudes; in fact, identical rates of change are precisely not to be expected when two contexts are unified by an underlying change, in most cases. This follows from the fact that usage frequencies must tail off towards 0 and 1 at the start and end of the change; the ensuing nonlinearity in the propagation curves, combined with the effect of the contextual production biases, then gives rise to slightly different ‘slopes’, a result which in our model can again be derived as a theorem (Disunity of Slopes). Somewhat paradoxically, then, and pace Kroch (1989), a model that derives the CRE from first principles leads us to conclude that in a CRE, slopes need not always be identical across contexts (Fig. 2).

Disunity of Slopes reveals (P2) to be a pseudo-problem: in tracking CREs, a historical linguist should not be looking for identical rates of change across contexts, but for contextual reflexes of a (hypothetical) underlying change which are related to the underlying change (and, by proxy, to each other) in a subtle way which can be quantified exactly but whose detection in noisy corpus data presents a serious and nontrivial challenge. In particular, we show that when artificially generated data from our model is fed into a logistic regression, significant interactions between contextual predictors and time will emerge (Fig. 3). This finding is to be expected in light of the Disunity of Slopes, but it is surprising in view of the standard assumption that any interaction between time and context is evidence for a lack of underlying grammatical unity (see e.g. Fruehwald, Gress-Wright & Wallenberg 2009; Zellou & Tamminga 2014), and strongly suggests that the standard assumption should be rejected.

Problem (P3) thus remains a serious challenge, and we conclude that the only non-problematic way of detecting CREs in corpus data is to fit models which are theoretically motivated and strong in the sense that they generate restrictive empirical predictions, and to examine the residual errors of fits so obtained.

**Fig. 1.** Residual error of our model (grey) and the standard formulation (black) for three different CREs – rise of do-support in English (Kroch 1989), Jespersen’s Cycle in English (Welage 2013), and loss of final fortition in Early New High German (Fruehwald, Gress-Wright & Wallenberg 2009) – and for a pseudo-CRE obtained by combining the latter two changes. Our more constrained model gives a fit which is essentially no worse than the standard, but correctly diagnoses the pseudo-CRE.

**Fig. 2.** CRE in a computer simulation. Note that slopes of the green and purple contexts are not exactly identical, even though derived from one and the same underlying change.

**Fig. 3.** Likelihood (p-value in two-tailed Z-test) of time:context interaction in a logistic regression model when fed data generated by our model, assuming two contexts with bias strengths $b_1$ and $b_2$: maxima over 50 experiments. The change was tracked for 50 generations, with each generation outputting $T$ tokens. The change was tracked for 50 generations, with each generation outputting $T$ tokens. The change was tracked for 50 generations, with each generation outputting $T$ tokens.