Information Theoretic Constant Rate Effects

This paper presents new empirical and theoretical results on one of the most robust findings of quantitative historical linguistics, the Constant Rate Effect [CRE; Kroch, 1989, Pintzuk, 1991, Santorini, 1993, Fruehwald et al., 2013, Zimmermann, 2017, *inter alia*]. The CRE is an often replicated observation that, during a change in which one linguistic rule or parameter setting replaces another, the replacement progresses at the same rate in all of the various linguistic contexts or surface reflexes of the variable in question. Each context, however, often favors one variant over the other during the change in progress (i.e., the y-intercept, k, of each context may differ), even though the change proceeds at the same rate in all contexts. There is usually no clear explanation for these contextual effects in a given change.

Earlier work has shown that contextual effects found in examples of the CRE can be explained by an information theoretic notion of information spread across sentences Wallenberg et al. [2021]. Here, we add further empirical work on the CRE in changes to Icelandic word order, showing that the verb-second phenomenon (V2) interacts with the right- and left-headed vP variants such that combinations are favored which result in more dispersed spreads of information. Additionally, we suggest that an information theoretic understanding of CREs supports and helps explain the "time separation theorem" argued for in Kauhanen and Walkden [2018].

Background: The information values of linguistic units can be quantified based on their probabilities [Shannon, 1948]. The distribution of information across an utterance thus exhibits peaks and troughs that correspond to a sentence's low and high probability units. In order for communication to be successful in the presence of noise, information ought to be spread across the sentence as uniformly as possible (Fenk & Fenk, 1980, see also; Fenk-Oczlon, 2001) . Recent work has shown that the ordering of words optimises for information uniformity by preventing peaks and troughs of information from clustering together Cuskley et al. [2021].

Based on this account of information uniformity, Wallenberg et al. [2021] predicted a previously undetected CRE in the English and Icelandic OV-to-VO changes. When Subjects and Objects are both pronominal DPs (low information) or nominal DPs (high information), VO prevents a clustered peak or trough of information. However, when Subjects and Objects are of sufficiently differing information values, OV derives the most uniform distribution. This pattern was demonstrated throughout the change from OV-to-VO in both languages, and was also shown to result in measurably more uniform distributions of information in Icelandic. In short, depending on linguistic context, speakers favor whichever variant yields the most uniform information distribution.

Current Study: While Wallenberg et al. showed that a pressure for information uniformity creates contextual effects in the OV-to-VO change, they did not account for how Subject-Aux inversion under V2 interacts informationally with OV/VO. Here, we amend this limitation of that study, show a new contextual effect of V2 within a CRE, and replicate the previous findings for Icelandic.

This paper presents preliminary results that V2 orders such as those in (3)-(4) interact with the changing headedness of vP in the history of Icelandic to produce peaks of information in the OV case, and more uniform distributions of information in the VO case. This creates the contextual effects of V2 vs non-V2 on the CRE in the Icelandic OV-to-VO change, shown in Fig.1. In the full talk, we will also quantify the dispersal of information in all the resulting clauses, as in Wallenberg et al. [2021].

VO – no fronting

(1) Jón hefur keypt bók í dag.
Jón has bought a.book in today
'Jón has bought a book today.'

OV – no fronting

(2) Þessi sami riddari vildi eigi gaum gefa
This same knight wanted not attention give
'This same knight didn't want to pay attention [to...].'
(1475.AEVINTYRI.NAR-REL,.933)

VO – adjunct fronting

(3) Í dag hefur Jón keypt bók.
in today has Jón bought a.book
'Today, Jón has bought a book.'

OV – adjunct fronting

(4) Aldrei hafði kóngsson slíkan grip séð.
never has a prince such a.thing seen
'Never has a prince seen such a thing'
(1450.VILHJALMUR.NAR-SAG,21.327)

The left column (without adjunct fronting) contrasts with the right column (with adjunct fronting): in the latter cases, V2 triggers Subject-Auxiliary inversion. The V2 constraint thus places the subject adjacent to the object if the vP is right-headed, as in (4), leading to two high information (i.e. low probability) phrases next to each other: a clustered peak of information. It follows, then, that speakers should prefer VO over OV in adjunct-fronted V2 contexts. In contrast, OV yields a more uniform information distribution across the clause in Subject-initial contexts.

This study also goes beyond Wallenberg et al. [2021] in advancing our theoretical understanding of the CRE, in line with proposals by Kauhanen and Walkden [2018]. K & W derive the CRE from a model of language acquisition, and their CRE model places a bound on the magnitude of contextual effects during a change in progress (in other words, an upper bound of the distance between k intercepts for each linguistic context). They further show that this limit, the "time separation theorem", is consistent with contextual effects in empirically observed CREs. Our work suggests a natural mechanism for the limit on contextual effects: when one variant replaces another in several contexts, each variant-context pairing changes the distribution of information across the resulting sentence. The total possible change in information spread is constrained by the possible permutations of context + variant (and the sentence's vocabulary items Cuskley et al. 2021). If CRE contextual effects result from speakers trying to maintain a certain uniformity of information spread, then those effects are necessarily limited in magnitude.
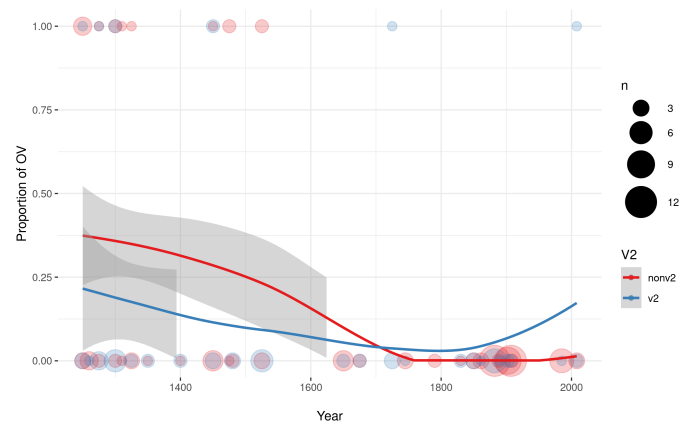

Figure 1: Proportion of OV in V2 vs non-V2 contexts, narrative texts from IcePaHC

Abbreviated references: • Cuskley, Bailes & Wallenberg, 2021. Noise resistance in communication: Quantifying uniformity and optimality. • Fenk & Fenk, 1980. Konstanz im kurzzeitgedächtnis-konstanz im sprachlichen informationsfluß. • Fruehwald, Gress-Wright, & Wallenberg, 2013. Phonological Rule Change: The Constant Rate Effect. • Kauhanen & Walkden, 2018. Deriving the Constant Rate Effect. • Kroch, 1989. Reflexes of grammar in patterns of language change. • Pintzuk, 1991. Phrase structures in competition: Variation and change in Old English word order. • Santorini, 1993. The rate of phrase structure change in the history of Yiddish. • Shannon, 1948. A mathematical theory of communication. • Wallenberg, Bailes, & Cuskley, 2021. Smooth signals and syntactic change. • Zimmermann, 2017. Formal and quantitative approaches to the study of syntactic change: Three case studies from the history of English.