

Article

Smooth Signals and Syntactic Change

Joel C. Wallenberg ^{1,*} , Rachael Bailes ¹ , Christine Cuskley ¹  and Anton Karl Ingason ^{2,*} 

¹ School of English Literature, Language and Linguistics, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK; rachael.bailes@newcastle.ac.uk (R.B.); christine.cuskley@newcastle.ac.uk (C.C.)

² Hugvísindasvið, Íslensku- og menningardeild, University of Iceland, Sæmundargötu 2, 102 Reykjavík, Iceland

* Correspondence: joel.wallenberg@newcastle.ac.uk (J.C.W.); antoni@hi.is (A.K.I.)

Abstract: A large body of recent work argues that considerations of information density predict various phenomena in linguistic planning and production. However, the usefulness of an information theoretic account for explaining diachronic phenomena has remained under-explored. Here, we test the idea that speakers prefer informationally uniform utterances on diachronic data from historical English and Icelandic. Our results show that: (i) the information density approach allows us to predict that Subject and Object type will affect the frequencies of OV and VO in specific ways, creating a complex Constant Rate Effect, (ii) the bias towards information uniformity explains this CRE and may help to explain others, and (iii) communities of speakers are constant in their average target level of information uniformity over long periods of historical time. This finding is consistent with an understanding of this bias which places it deep in the human language faculty and the human faculty for communication.

Keywords: diachronic syntax; information theory; uniformity; Icelandic (Old and Modern); English (Old, Middle, and Modern)



Citation: Wallenberg, Joel C., Rachael Bailes, Christine Cuskley, and Anton Karl Ingason. 2021. Smooth Signals and Syntactic Change. *Languages* 6: 60. <https://doi.org/10.3390/languages6020060>

Academic Editors: Sina Bosch, Ilaria De Cesare, Ulrike Demske and Claudia Felser

Received: 16 January 2021

Accepted: 10 March 2021

Published: 25 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the early-to-mid 20th century, advances in the theory of computation gave rise to a new mathematical theory of communication (Shannon 1948) now known as Information Theory. This has led to the key insight that language use is shaped by a pressure to distribute information uniformly across utterances (Aylett and Turk 2004; Fenk and Fenk 1980; Levy and Jaeger 2007). In the present study, we show that information uniformity leads to specific hypotheses concerning the effect of syntactic context on the change from head-final *vP*/*VP*s to head-initial *vP*/*VP*s (OV to VO) in English and Icelandic.

We begin in Section 1.1 by explaining the intellectual context for the notion of information uniformity we adopt throughout this paper. We then present the study's main hypothesis (Section 1.3): that the frequencies of OV and VO are modulated throughout the change by an effect of Subject and Object type, where the combination of these variables leads to greater information uniformity in a sentence. We follow our Methods (Section 2) and quantitative Results (Section 3) with a discussion of their main theoretical significance in Section 4: information uniformity has predicted and explained an instance of the "Constant Rate Effect" (Kroch 1989), and one that was not easy to detect in the first place.

1.1. Information Uniformity in Language

In Shannon's (1948) model, a receiver begins with some expected set of possible outcomes about an event. Communication takes place when a signal narrows this set of outcomes, and the extent to which a signal narrows this set is the extent to which that signal is informative. A signal string made up of constituent units will therefore give rise to peaks and troughs of information distributed across the whole signal, as each unit varies in the amount of uncertainty it reduces for the receiver. The probability of a given unit determines

its Shannon information content ($I(u) = -\log_2 p(u)$), such that a less probable unit is high information and, conversely, a more probable unit is low information. This probability is standardly estimated in terms of frequency (see [Pierce 1980](#); [Shannon 1948](#)), e.g., the more frequent word “animal” is low information compared to the less frequent word “porpoise”, which is more informative. The frequency, predictability, and (thus) informativeness of a given word may vary according to contextual parameters of a discourse ([Zarcone et al. 2016](#)) or even characteristics of speakers themselves (e.g., [Degaetano-Ortlieb 2018](#)). Nevertheless, models capture these extralinguistic effects by assuming this central relationship between probability and information value.

[Shannon’s \(1948\)](#) model also outlines the manner in which noise affects the transmission of information: the noisier the channel, the longer a signal needs to be in order to compensate for resulting information loss. Peaks of information in a given string can therefore be made more robust to noise by stretching them out over the transmission time. This observation has given rise to *information density accounts* of language, which suggest the distribution of information in linguistic utterances is optimized for resistance to noise ([Aylett and Turk 2004](#); [Fenk and Fenk 1980](#); [Frank and Jaeger 2008](#); [Genzel and Charniak 2002](#); [Turk 2010](#)).

A growing body of empirical work demonstrates that language is shaped by the pressure to keep information uniformly distributed by ‘smoothing’ peaks of information. This applies to a wide range of linguistic phenomena, such as lengthening and emphasis in prosody ([Aylett and Turk 2004](#); [Bergen and Goodman 2015](#)); phonetic reduction ([Pluymaekers et al. 2005](#)); syntactic reduction ([Frank and Jaeger 2008](#); [Jaeger 2010](#); [Levy and Jaeger 2007](#)); and classifier choice in Mandarin ([Zhan and Levy 2018](#)). For example, one such study of a large corpus of natural speech showed that word frequency alone accounted for 11% of variation in duration; speakers take more time to articulate a word when it appears in a context in which it is unusual (i.e., when it is less frequent and therefore high information) ([Aylett 1999](#)). In the domain of syntactic choice, [Levy and Jaeger \(2007\)](#) and [Jaeger \(2010\)](#) show that complementizer deletion, which concentrates the information of two words into one (e.g., *I know that she is*, vs. *I know she is*), is avoided in cases where the embedded Subject is high information. This suggests that speakers may condition their syntactic choice on the prevention of large peaks or troughs of information in order to maintain a more uniform distribution.

A more recent approach to noise resistance optimization in language considers the role of ordering items in a sentence. Following [Fenk and Fenk \(1980\)](#), [Cuskley et al. \(2020\)](#) suggest that peaks and troughs of information can be avoided through word order (without necessarily stretching the signal in time), and that orderings which yield greater uniformity help to mitigate against catastrophic communication failures due to noise. In other words, if high-information words are clustered in a string, catastrophic information loss is more likely to result from noise disruption. Distributing high information units more uniformly throughout the string, however, improves overall resistance to the same noise ([Cuskley et al. 2020](#)).

This order-based account of information uniformity has wide-ranging applications in the study of language that remain under-explored. Here we present one such application by investigating the effect of syntactic context on the English and Icelandic OV to VO change. With this case, we demonstrate that an information density approach enables the generation of new hypotheses and a mechanistic level of explanation for known phenomena in language change, and for the discovery of new ones.

1.2. Current Study

Our study reports a previously unnoticed (to our knowledge) effect of linguistic context on the Icelandic ([Hróarsdóttir 2000](#); [Ingason et al. 2012](#); [Kiparsky 1996](#); [Rögnvaldsson 1994, 1996](#); [Sigurðsson 1983](#)) and English ([Kiparsky 1996](#); [Kroch and Taylor 2000b](#); [Pintzuk 1991](#); [Pintzuk and Taylor 2006](#)) OV to VO change, and one that we would not have even thought to look for if not for an analysis of information density in language. The OV to

VO change, in both languages, involved the innovation or introduction of a head-initial *vP*/VP structure (“VO”) into an older Germanic system that had originally had head-final *vP*/VPs (“OV”) (Kiparsky 1996). This change in *vP*/VP structure followed a similar change in IP/TP structure, as can be demonstrated for English, at least (Kroch and Taylor 2000b; Pintzuk 1991). Over a period of some hundreds of years, the frequency of the VO *vP*/VPs increased at the expense of OV *vP*/VPs, until the modern Icelandic and English categorical VO systems became the only extant ones in those speech communities. Note that, for English, the change took place from sometime during the Old English period through roughly the late 17th c., and for Icelandic, the period in question lasted from sometime in the early Old Icelandic period through the early 19th c. (see Figure 1).

The examples in (1) demonstrate the OV/VO variation during the period of change in Icelandic. They both contain a finite auxiliary, occupying I/T position in the phrase structure, as well as a main verb and its complement. In the OV order (1a), which is more frequently attested in earlier periods, the (here pronominal) complement precedes the nonfinite main verb, whereas in the VO order (1b), which takes over in the course of history, the Object follows the main verb. The Middle English example in (2) shows the same kind of variation during the period of change in English.

Historical Icelandic:

- (1) a. ...og sannleikurinn mun yður frelsa
 ...and the truth will you free
 “...and the truth will set you free.”
 (Oddur Gottskálksson’s *New Testament*, date: 1540; ID 1540.NTJOHN.REL-BIB, 204.662 from Icelandic Parsed Historical Corpus (IcePaHC))
- b. ...en eg skal sjá yður aftur.
 but I shall see you-PL again
 “...but I shall see you again”
 (Oddur Gottskálksson’s *New Testament*, date: 1540; ID 1540.NTJOHN.REL-BIB, 223.1305 from IcePaHC)

Middle English:

- (2) Mi feader & Mi moder for-þi þt ich nule þe forsaken; habbe forsake me.
 My father and my mother because that I not+would you forsake have forsaken me
 “Because I would not forsake you, my father and mother have forsaken me”
 (St. Juliana, northern Herefordshire/southern Shropshire, date: c1225; ID CMJULIA-M1,106.172 from the Penn Parsed Corpus of Middle English 2 (PPCME2))

The example in (2) not only illustrates both the OV and VO *vP*/VP structures in historical English, but also illustrates the intra-speaker variation characteristic of the period of time when the change was in progress. Like other speakers during the change (perhaps all speakers during the change), the author of this manuscript could readily switch between the OV and VO structures, even within a single sentence, and even when the phrases in question are composed of the same words.

One of the most often and precisely replicated results in language change is the “Constant Rate Effect” (Fruehwald et al. 2013; Kauhanen and Walkden 2018; Kroch 1989; Pintzuk 1991; Santorini 1993, *inter alia*). As a change in a linguistic feature spreads through a population over time, the change may be observed in a number of different surface contexts, and if the rate of change is measured in the various linguistic contexts, it will not significantly differ between them. Even so, the context often still affects the frequency of the old and new variants as a change progresses. For instance, in the English OV to VO change (where the relevant feature is the headedness of *vP*/VP), Pintzuk and Taylor (2006) report that subordinate clauses have a higher frequency of OV, even though the rate of change does not differ between clause types (see also Pintzuk 1991, and our Figure 1).

The current study demonstrates that some such contextual effects derive from the hypothesis that speakers prefer information uniformity. The speaker bias for smoother information distributions results in a preference for OV in contexts where OV results in a smoother information distribution than VO. In turn, this bias also favors VO in contexts where the result is more uniform than OV. In particular, an information density account of linguistic structure turns out to make precise predictions about how the nominal/pronominal status of Subjects interacts with the nominal/pronominal status of Objects to modulate the frequencies of OV and VO in *both* historical English and Icelandic.

The predictions borne out in the historical data would have been difficult to even imagine without information theory, and yet they follow naturally from an account of language in which speakers optimize information density for noise resistance. We calculate precise measurements of information uniformity in all the clauses in the Icelandic data set, using the descriptive statistic developed for this purpose in Cuskley et al. (2020). The present article continues the integration of information theory and linguistics, and leads to deeper explanations for patterns in syntactic optionality and language change.

1.3. Predictions

Our hypotheses follow from the following broad generalizations:

1. Pronouns are closed-class and frequent, and so are low information content, (cf. Shannon 1948) in comparison with nominal DPs (e.g., *you* vs. *the politician*). For example, the average information content value for pronouns in the Penn Parsed Corpus of Modern British English (PPCMBE; Kroch et al. 2016) is 11.7 bits.
2. Nominal DP Objects are high information content in general, because any particular DP will be low probability. There are more common noun lexemes than any other part of speech, and so any given noun is lower probability by virtue of belonging to a very large set. The probability of a given noun *combined with* a determiner and/or modifiers is necessarily lower than the probability of the noun alone. For example, the average information content for nouns in PPCMBE is 13.7 bits, and we can reasonably expect the average information content is much higher for nominal DPs, since they can be of arbitrary complexity.
3. Verbs are mid-level in frequency and so also in information content: lower than nominal DPs, on average, but higher than pronouns. For example, the average information content for lexical verbs in the PPCMBE is 13.5 bits.

Given these generalizations, we developed and tested the following hypotheses.

- When Subject and Object are of the same type (i.e., both pronominal or both nominal), VO clause structure is statistically favored.
- Otherwise, OV clause structure is favored.
- These effects are orthogonal to the OV to VO change, but apply as constants throughout the course of the change.

In cases where both Subject and Object are the same type, and therefore similar in terms of information content, VO will tend to result in more uniform sentences. The verb placed between two arguments will break up and spread a potential peak or trough of information content more evenly across the whole utterance, thus yielding **HIGH-MID-HIGH** in (3) and **low-MID-low** in (5) below. This bias for information uniformity means speakers are biased against OV in contexts with arguments of the same type. In contrast, arguments of different types will bias speakers *towards* OV. When the Subject is pronominal and the Object is nominal, as in (6), speakers should favor S-Aux-O-V because this shape of information distribution is closer to uniform (**low-HIGH-MID**) than a VO order with the same elements; S-Aux-OV yields the maximally asymmetric, informationally back-loaded **low-MID-HIGH**. Similarly, if the Subject is nominal and Object pronominal, as in (4), S-Aux-O-V yields an information distribution of **HIGH-low-MID**. Though not perfectly uniform, this pattern is closer to uniform than the VO version of the same sentence, which would yield the maximally asymmetric, informationally front-loaded **HIGH-MID-low**.

These descriptions of information distributions are intended as a schematic guide to the patterns that different syntactic configurations will *tend* to produce, averaged across many sentences. We will quantify the distributions more precisely in Section 3.2 below, using the summary statistic described in Section 2.

Historical Icelandic:

- (3) grátur mun þryngva ina efstu hluti fagnaðar
Weeping will throng the highest parts joy-GEN
“Weeping will crowd into even the very last bits of joy”
(*Íslensk Hómilíubók*, date: c. 1150; ID 1150.HOMILIUBOK.REL-SER,.1408 from IcePaHC)

HIGH-MID-HIGH is the information distribution shape produced, on average, by the above clause structure.

- (4) ...og sannleikurinn mun yður frelsa
...and the truth will you free
“...and the truth will set you free.”
(Repeated from above, ID 1540.NTJOHN.REL-BIB,204.662 from IcePaHC)

HIGH-low-MID is the information distribution shape produced, on average, by the above clause structure.

Middle English:

- (5) Ne maȝ he nohht riht cnaewenn me
NEG may he not right know me
“He may not rightly know me”
(*Ormulum*, Lincoln, date: 1200; ID CMORM-M1,II,241.2491 from PPCME2)

low-MID-low is the information distribution shape produced, on average, by the above clause structure.

- (6) tu miht ec gastlike laf Onn oþerr wise ȝarrkenn
you might also spiritual loaf in another way prepare
“You might also prepare a spiritual loaf in another way.”
(*Ormulum*, Lincoln, date: 1200; ID CMORM-M1,I,49.493 from PPCME2)

low-HIGH-MID is the information distribution shape produced, on average, by the above clause structure.

In short: if the Object and Subject are of the same type (as in 3 or 5), putting a verb between them creates a more uniform information distribution. If, however, the Object and Subject are of differing types (as in 4 or 6), final placement of the verb will create a more uniform distribution.

We first test the hypotheses above as stated, in terms of the proportions of OV and VO when they co-occur with different Subject and Object types, and over time, on diachronic corpora of English and Icelandic (Section 3.1). We then use a more precise quantification of information uniformity to test the general hypothesis underlying the three above: speakers of all historical times target a basic level of uniformity in their sentences, and use either OV or VO structures in syntactic contexts where the chosen variant leads to a more uniform sentence.

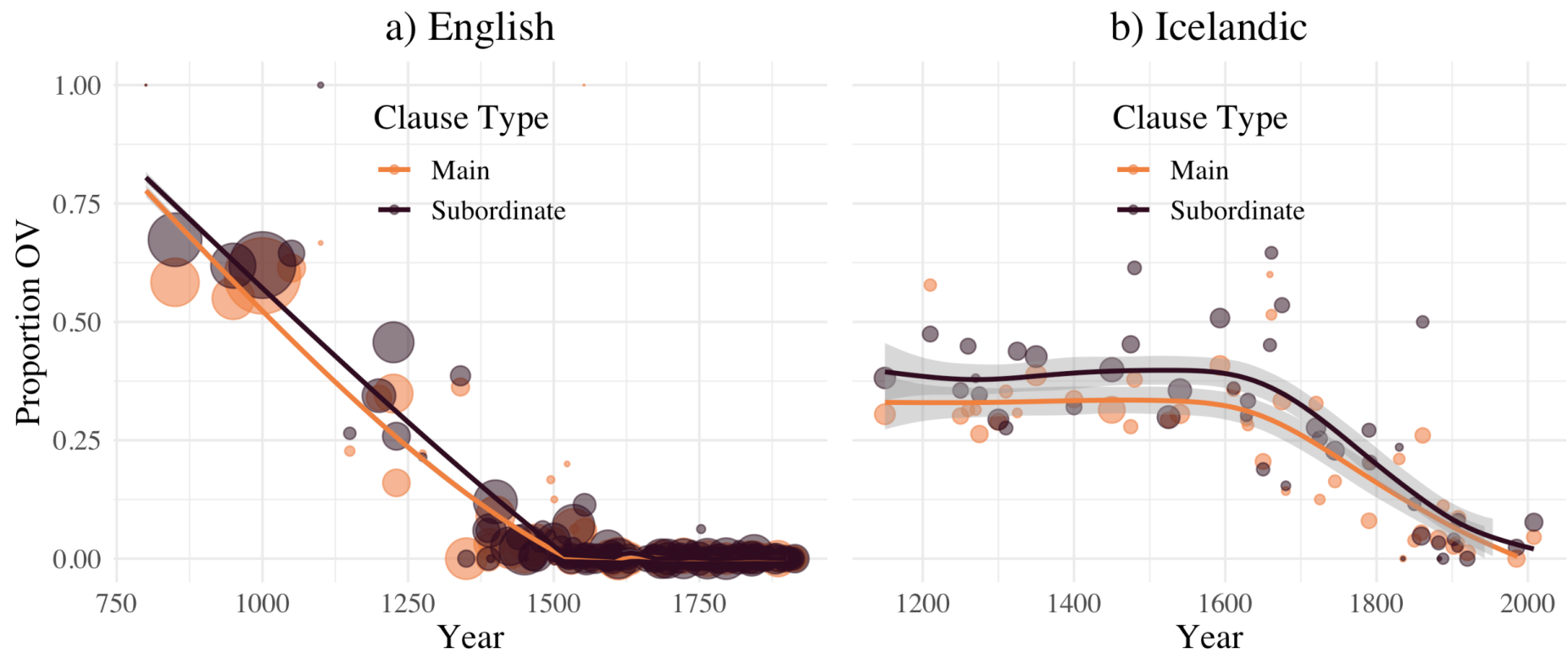


Figure 1. Proportion OV word order by clause type over time, in (a) English and (b) Icelandic. Note the larger data set for English and the difference in timescale. Subordinate clauses favor OV in both English and Icelandic.

2. Materials and Methods

We test these predictions using the Penn Parsed Corpora of Historical English (Kroch et al. 2004, 2016; Kroch and Taylor 2000a) and the Icelandic Parsed Historical Corpus (IcePaHC) (Rögnvaldsson et al. 2012, 2011; Wallenberg et al. 2011). We use CorpusSearch (Randall 2013) to query the English corpora and Treebank Studio (PaCQL query language) (Ingason 2016) to query the IcePaHC. As in previous work, e.g., Pintzuk (1991); Pintzuk and Taylor (2006); Taylor and Pintzuk (2012), we only consider clauses containing a finite auxiliary, so that we can focus on *vP*/VP structure and control for the effects of verb movement to Tense (V-to-T and V-to-T-to-C movement). We therefore take VO word order to be a clause in which a finite auxiliary precedes a nonfinite main verb, which in turn precedes a DP Object. Our definition of OV order is again a clause with a finite auxiliary, but one in which the auxiliary precedes a DP Object, which in turn precedes a nonfinite main verb (we recognize that there are processes which obscure *vP*/VP structure by moving certain Objects across the verb, but none of the effects described in this paper will be affected by these processes; cf. the description of interpreting historical data in this regard in Kroch and Taylor 2000b).

An example of an OV subordinate clause from IcePaHC is shown in (7) and its treebank representation in (8).

- (7) ef þeir vilja trú taka, ...
 if they want faith take ...
 "...if they wish to convert, ..."
 (*Jómsvíkinga saga*, date: 1260; ID 1260.JOMSVIKINGAR.NAR-SAG,.765 from IcePaHC)

- (8) (IP-SUB
 (NP-SBJ (PRO-N þeir))
 (MDPI vilja)
 (NP-OB1
 (N-A trú))
 (VB taka))
 Token ID: 1260.JOMSVIKINGAR.NAR-SAG,.765

A PaCQL query that extracts clauses containing a finite auxiliary, a nonfinite main verb and an Object, coding OV orders as 1 and VO orders as 0, is shown in (9). It defines named regular expressions and matches patterns in the syntactic analysis in terms of immediate dominance (idoms) and sisterwise precedence (sprec).

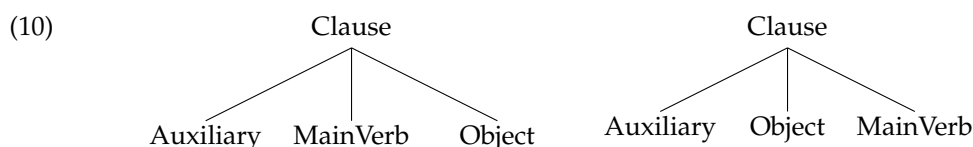
- (9) define:
 clause IP-(MAT|SUB).*
 finaux (MD|HV|BE|DO|RD)[PD][IS]
 mainverb (VB|VAN|VBN)
 object NP-OB[12]

 clause idoms finaux
 clause idoms NP-SBJ

 ov:1
 finaux sprec object
 object sprec mainverb

 ov:0
 finaux sprec mainverb
 mainverb sprec object

The two annotation patterns that are matched by the query are shown schematically in (10). They are coded as 1 (for OV) and 0 (for VO), respectively.



The query furthermore includes a meta coding section, shown in (11), that writes out for each result the genre of the text, the year of manuscript (or an estimate thereof) and the label matched by the clause regular expression. The clause label, IP-MAT or IP-SUB, allows us to include the matrix vs. subordinate clause distinction in our model.

(11) meta:
 text genre
 text year
 node label clause

For inclusion in our models, the query also outputs the type of the Object and Subject of the clause in question (a part of the query omitted here). Objects are coded as pronouns, quantified/negated, or other full nominal DPs, because these categories are known to show different displacement properties across various constructions. In particular, quantified or negated Objects can exceptionally occur to the left of the nonfinite verb position (often analyzed as leftward movement), creating surface Object–Verb orders that are not structurally the same as other Object–Verb orders. This occurs in all historical stages of Icelandic and also in historical English (Pintzuk and Taylor 2004; Rögnvaldsson 1984, 1987; Thráinsson 2007). Subjects are coded as pronouns, full nominal DPs, or “gapped” Subjects (i.e., non-overt because they were extracted, reduced following a conjunction, or rarely, pro-dropped). The output also includes the full text of each sentence that was matched by the query. This text is then lemmatized and each word is mapped to the type frequency of this lemma in the Tagged Icelandic Corpus (Helgadóttir et al. 2012). Note that while we have described a PaCQL query here, the syntax of the CorpusSearch queries used for English is quite similar. The syntax of CorpusSearch is explained in its online documentation.

These frequencies form the empirical basis for computing the Deviation of the Rolling Mean of the sentence (DORM), a descriptive statistic which provides a concise and easily interpretable summary of how uniform or clumpy a particular utterance is. To compute the DORM for a sentence we take the arithmetic mean of every adjacent, overlapping pair of information content values (i.e., a rolling mean), derived from the lemma frequencies in the results. We then compute the sample variance of the resulting list of means (an unbiased estimator, using denominator $n - 1$) and arrive at a single number, the DORM (see Cuskley et al. 2020 for comprehensive description of DORM and its properties).

The DORM values are subsequently calibrated (i.e. adjusted for the particular array of vocabulary items in a sentence so that DORM is comparable between sentences) by applying a Uniform Information Density Optimization (UIDO) algorithm. For any array of information content values, UIDO systematically reorders them, seeking to minimize the DORM for the whole strings. We output a “calibrated” DORM measure for each sentence, i.e., $DORM - DORM(UIDO)$, that allows us to model the relationship between information density (calibrated DORM), the OV/VO variable, and variables that describe the context.

Statistical analyses were conducted in R (R Core Team 2017), and visualizations used Wickham (2009). All statistical analyses and data manipulation can be found here for the results in Section 3.1: <https://raw.githubusercontent.com/joelcw/constantentropy/master/infoTheoryTest.R> (accessed on 11 March 2021), and here for the results in Section 3.2: <https://raw.githubusercontent.com/joelcw/iceBits/main/processIcePaHCinfo.R> (accessed on 11 March 2021).

3. Results

The results from the corpora bear our predictions out precisely, in both English and Icelandic:

- *When Subject and Object are of the same type (i.e., both pronominal or both nominal), VO clause structure is statistically favored.*
- *Conversely, when Subject and Object are of different types, OV clause structure is favored.*
 - These complementary results are summarized in Section 3.1 below, and illustrated by Figure 2. That this pattern of clause structure results in more uniform information distributions is summarized in Section 3.2 and illustrated by Figure 3.
- *These effects are orthogonal to the OV to VO change, but apply as constants throughout the course of the change.*
 - This result is indicated in Section 3.1 below and illustrated by Figure 2. That information density remains stable given the OV to VO change is summarized in Section 3.3, and is illustrated by Figure 4.

3.1. Effect of Syntactic Context on OV/VO Diachrony

Figure 2 shows the decline of OV in the two languages, by Subject type and Object type. As predicted, VO word order is favored when the Subject and Object are the same type, whether nominal or pronominal, and OV is favored when the Subject and Object are of different types. The size of the effect is smaller in the Icelandic data, probably because that data set is smaller (≈ 1 million words across all time periods and 6,190 clauses in our data set, compared with ≈ 10 million words for the combined English corpora and 58,437 clauses in our data set).

To check the effect of Subject and Object type statistically, we fit a mixed effects logistic regression separately to the English and Icelandic data, with fixed predictors for Year of text, Clause type, Subject type, and Object type, and random intercepts for individual texts (since individual texts can be idiosyncratic, and there were multiple observations per text). Unfortunately, no mixed effects models containing interaction terms would converge on this size of data set. However, a mixed effects logistic regression with a simpler structure did converge: $OV \sim (1|ID) + Clause + zYear + SbjType + ObjType$. For English and Icelandic, the inclusion of terms for Subject and Object type did significantly improve the model fit (though estimated slopes are difficult to interpret without any interaction terms); for English: AIC (with Subject and Object type) = 12,529 vs. 13,249 (without), BIC = 12,601 vs. 13,285, $p < 2.20 \times 10^{-16}$ for the model comparison on the Chi-squared distribution; for Icelandic: AIC (with Subject and Object type) = 6455 vs. 6549 (without), BIC = 6515 vs. 6582, $p < 2.20 \times 10^{-16}$. A standard logistic regression containing the Object type and Subject type interaction did show a significant effect of the interaction of Object type and Subject type on the probability of OV, in the expected direction: a pronominal Subject and a pronominal Object decrease the probability of OV (for English: $\beta = -0.661$, $p = 0.0149$; for Icelandic, the effect was borderline but in the predicted direction: $\beta = -0.271$, $p = 0.0848$), and both Subject and Object being nominal decrease the probability of OV with roughly the same effect size (for English: $\beta \approx -0.67$, $p \approx 0.01$, depending on combination of non-pronominal arguments; for Icelandic: same borderline effect as above for non-quantified/non-negated nominal Objects, and a significant effect decreasing OV when the Object is quantified/negated: $\beta = -0.554$, $p = 9.36 \times 10^{-3}$). For both languages, slope estimates for the interaction between argument type and text date were not significantly different from zero, with $0.221 \leq p \leq 0.884$ depending on the argument combinations. Thus, the effect of the arguments together on OV is plausibly constant over time, making the Subject–Object interaction a Constant Rate Effect with respect to the OV to VO change.

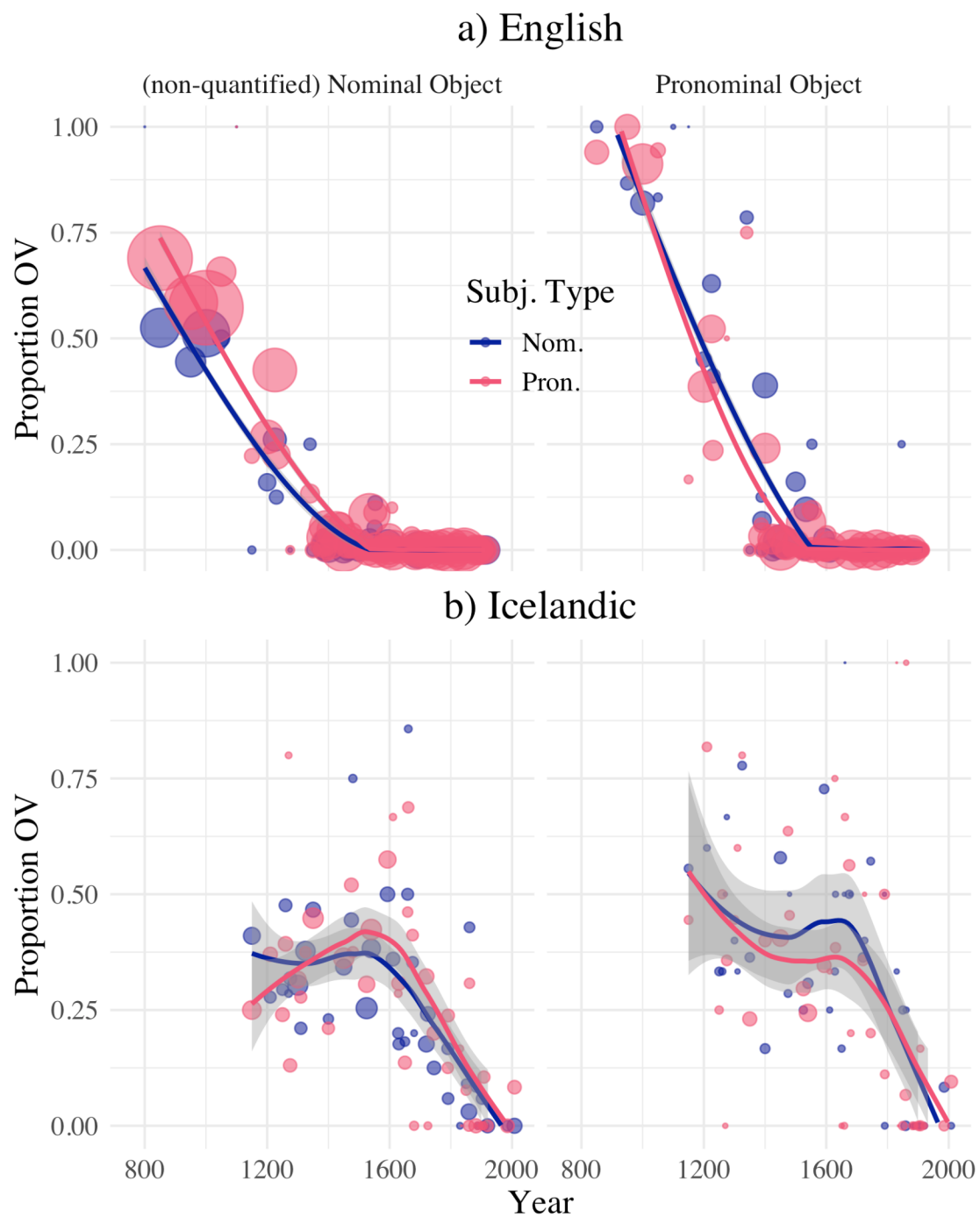


Figure 2. Proportion OV clauses by Subject type over time for all clauses in (a) English and (b) Icelandic. Note the larger data set for English (dot size represents number of data points for a year). In all cases, OV is favored when Object and Subject are of different types, and OV is *disfavored* when Objects and Subject are of the same type. **Upper left:** Given a nominal Object in English, nominal Subjects (blue line) disfavor OV, while pronominal Subjects (pink line) favor OV. **Upper right:** Given a pronominal Object in English, pronominal Subjects (pink line) disfavor OV, while nominal Subjects (blue line) favor OV. **Lower left:** Given a nominal Object in Icelandic, nominal Subjects (blue line) disfavor OV, while pronominal Subjects (pink line) favor OV. **Lower right:** Given a pronominal Object in Icelandic, pronominal Subjects (pink line) disfavor OV, while nominal Subjects (blue line) favor OV.

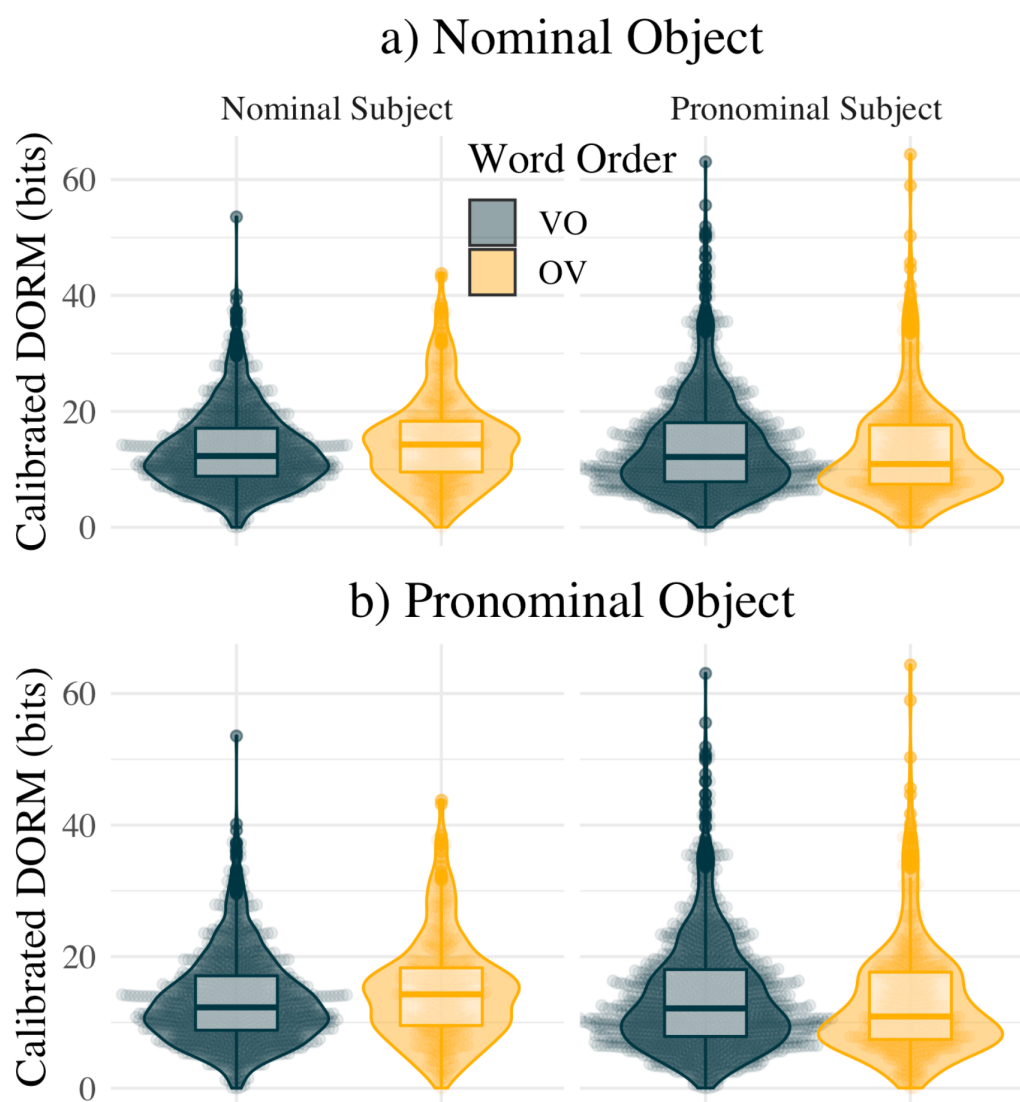


Figure 3. Calibrated DORM values for Icelandic sentences by Subject type, Object type and *v*P/VP word order. Note that where the Subject and Object types are matched in (a), VO results in greater uniformity (lower DORM), and elsewhere OV results in greater uniformity. In (b), the effect is less certain, and may be carried by the tails of the distributions.

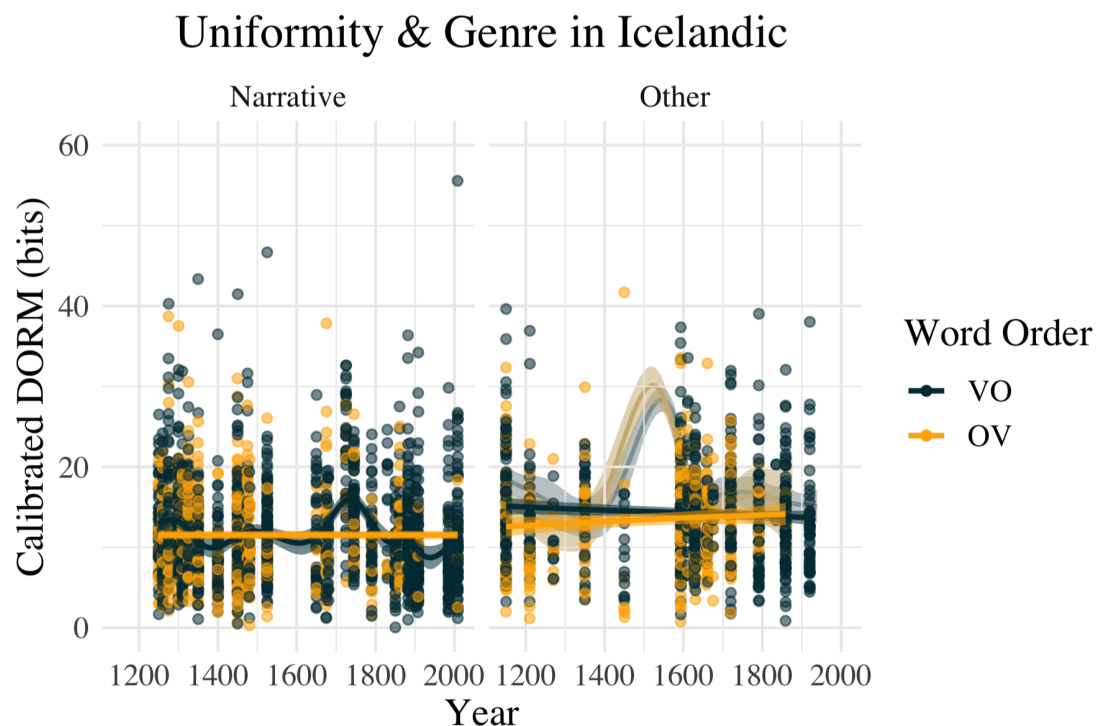


Figure 4. Calibrated DORMs for Icelandic sentences by Year, Word Order, and Genre (narrative vs. other). Note the small but stable effect of narrative texts in favoring more uniform sentences, and the important absence of a Year effect on information uniformity. An outlier, Oddur Gottskálksson’s New Testament translation from 1540, is excluded from the solid lines in “Other”, but included in the shaded lines behind them (see Section 4.3).

3.2. OV/VO and Information Density in Icelandic

Because of the computational infrastructure available for both modern and historical Icelandic, we were able to quantify the information density of strings in IcePaHC based on lemma frequencies, and so test our hypotheses more precisely. Once we had a “calibrated” DORM measure for each sentence based on its string of lemmas (i.e., $DORM - DORM(UIDO)$), we modelled the relationship between calibrated DORM of the sentence, OV/VO, and surrounding syntactic context as a mixed effects linear regression with calibrated DORM as the dependent variable and random intercepts by text. To put the effect sizes below in context, the mean of calibrated whole-sentence DORMs for the sample of 6473 clauses was 12.7 bits, with a range of 1.98×10^{-4} to 64.3 bits, and a standard deviation of 8.18 bits.

The model which most closely matched our hypotheses had the following specification: $SentDormUido \sim (1 | TextId) + Year + OV + Clause + SimpleGenre + ObjType + SbjType + SbjType * ObjType * OV$. The 3-way interaction term directly tested our hypothesis that the effect of OV and VO on sentence uniformity depends on the Subject and Object type (of course, the model also contained the nested 2-way interaction terms implied in the 3-way term). Note that “SimpleGenre” is text genre recoded as a binary variable: narrative texts (the majority of IcePaHC) compared with all non-narrative texts in the corpus. An anonymous reviewer observes that this binary variable does not distinguish between the style of the Old Icelandic Sagas and that of other, more modern narrative texts. This is true, and we acknowledge it as a limitation of the study. However, the binary distinction preserves enough statistical power for the rest of the analysis to be tractable, and it importantly separates narrative texts from the religious and technical “non-narrative” texts, whose styles we know are very different from all of the narrative texts. We also note that later narrative texts in Icelandic are often influenced by the classical Saga style; see reference in Section 4.3.

As we predicted, there was a significant interaction between OV/VO, Subject type, and Object type: OV clauses combined with a pronominal Subject result in more uniform sentences when the Object is not constrained to be pronominal (lower DORMs; $\beta = -1.74$, $p = 4.05 \times 10^{-3}$), but OV combined with a pronominal Subject and a pronominal Object leads to *less* uniform sentences (higher DORMs: $\beta = 2.66$, $p = 0.0136$). The same effect holds if OV is combined with a nominal Subject and a nominal Object ($\beta = 2.66$, $p = 0.0136$). If the Object is nominal (non-quantified/non-negated) and the Subject is pronominal, the estimated effect in combination with OV is the same magnitude but in the opposite direction: nominal Objects and pronominal Subjects predict *greater* uniformity (lower DORM; $\beta = -2.66$, $p = 0.0136$). The model estimates the same effect in the same direction when an OV clause contains a nominal Subject and a pronominal Object. Similarly, OV increased the uniformity of the sentence when the Subject was gapped and the Object was either a non-quantified/non-negated nominal ($\beta = -4.19$, $p = 0.024$) or quantified/negated ($\beta = -6.24$, $p = 0.0124$).

The estimates for the model above also showed significantly non-zero slopes for Clause type ($\beta = 0.823$, $p = 1.65 \times 10^{-5}$). In other words the presence of a subordinate clause predicts slightly less uniform sentence (i.e., higher DORM value). The effect of Genre was also significant: non-narrative texts have less uniformity (i.e., higher DORM values: $\beta = 3.07$, $p = 2.19 \times 10^{-4}$). This model did not show a significant main effect of OV on DORM ($\beta = 0.543$, $p = 0.245$). There were also significant main effects of Object and Subject type being non-pronominal on calibrated DORM values, if the reference level for these variables was set to “pronoun”: nominal Objects (whether quantified/negated or not) and nominal Subjects made sentences less uniform overall, i.e., predicted higher DORM values; quantified/negated Objects: $\beta = 3.22$, $p = 1.73 \times 10^{-9}$; other nominal Objects: $\beta = 4.19$, $p < 2 \times 10^{-16}$; nominal Subjects: $\beta = 2.33$, $p < 3.39 \times 10^{-7}$.

There were also significant interactions between Subject type and Object type: pronominal Objects in combination with either pronominal Subjects ($\beta = -3.09$, $p = 1.77 \times 10^{-8}$) or gapped Subjects ($\beta = -2.67$, $p = 2.77 \times 10^{-3}$) increased the uniformity of a sentence, predicting lower DORM values.

We also considered a more complex model which contained a 4-way interaction between OV, Clause, Subject type, and Object type, and this model estimated a small overall effect of OV decreasing uniformity ($\beta = 2.76$, $p = 1.82 \times 10^{-3}$). However, inconsistent model comparison metrics make it unclear which model to prefer. This model also estimated a significant interaction between Subordinate Clause type and OV in predicting lower (more uniform) DORMs ($\beta = -2.98$, $p = 4.09 \times 10^{-3}$). We report these two effects here for completeness, and because they pertain to our hypotheses, but note that this model is not necessarily more reliable than the 3-way model we discuss above (the other effects we discuss above do not conflict between these models).

3.3. Stability of Information Density over Time

It is important to note that in none of the models we considered was the estimated effect of text date on calibrated DORM significantly different from a slope of 0 ($-9.58 \times 10^{-4} \leq \beta \leq -8.34 \times 10^{-4}$, $0.524 \leq p \leq 0.579$). The year of the Icelandic text was also shown to not be a significant factor in model comparison, and no interactions involving text date made a significant contribution to model fit. Comparing the 3-way interaction model above with a version that omits the Year predictor yields the following comparison metrics: AIC = 44,308 (with Year) vs. 44,307 (without), BIC = 44,505 vs. 44,496, $p = 0.559$.

This result makes it clear that even though the frequencies of OV and VO clauses certainly changed as a function of time, and OV/VO has an effect on calibrated DORM, the date of the text has no effect on calibrated DORM once the OV/VO is statistically controlled for. In other words, date and calibrated DORM are conditionally independent, given OV/VO. Figure 4 shows this graphically: the small Genre effect is apparent, with non-narrative texts containing slightly less informationally uniform sentences, but DORM

over time is essentially flat (the only possible trend is in VO sentences from narrative texts, though this is by no means certain).

4. Discussion

4.1. Effect of Syntactic Context on OV/VO Diachrony

The data from the preliminary study of Subject and Object type on the English and Icelandic OV to VO change supported the hypotheses we constructed based on the uniform information density theory. Pronominal arguments will be lower in information content than any nominal argument, simply because they are closed-class items and any given pronoun will occur with a higher probability than any nominal argument; nominal arguments come in a variety of complexities and with a variety of head nouns, so the probability of encountering any particular one of them will be low (and probably follows a Zipfian distribution). Nonfinite verbs, on the other hand, are always a single word, and there is a limited (though not entirely closed) number of lexical verbs. Thus, their information content will tend to be between that of a pronoun and a nominal DP. Given that, if speakers prefer more uniform information distributions across an utterance (neither front-loading or back-loading information content), VO is a good strategy to break up a sequence of particularly low information arguments (pronouns) or particularly high information ones (nominal DPs). On the other hand, OV is the better strategy when the arguments are of different types, using the final verb to modulate the potential information peak or trough from an adjacent pronoun Object or nominal Object (this is only a potential peak or trough when the Subject is of the opposite type).

This seems to be what occurs in the historical data: neither the Subject type or Object type alone is entirely predictive of whether OV or VO is favored. Rather, there is an interaction between the argument types, leading to the reversals of the colored lines in Figure 2. This interaction is interestingly *not* predicted by a “given”-first “new”-last theory of word order (Gundel 1988), since the lexical verb will often be the newest element in a clause with both a pronominal Subject and Object. There are certainly other documented effects of information structure (e.g., new vs. given, Focus) on the frequencies of OV and VO in both historical English and Icelandic, though these do not drive the OV-to-VO change, and may constitute their own Constant Rate Effect on the change (Taylor and Pintzuk 2012, 2011; see also background references therein). Our results are compatible with these observations in the literature, especially if the analysis in Taylor and Pintzuk (2011) is correct. They argue that information structure is primarily relevant to nominal, non-quantified/non-negated Objects, because these can be extraposed rightward when they are focused. This affects the frequencies of surface OV and VO in Old English and Old Icelandic particularly, by making underlyingly head-final *vP*/VPs appear to be VO. The statistical effects recede over time as the head-initial *vP*/VP phrase structure increases in frequency in English and Icelandic, as this rightward displacement of Objects is often no longer detectable in a VP *vP*/VP. This analysis does not speak to our observations about pronominal Objects one way or the other, as they do not participate in rightward Object extraposition (see, e.g., Kroch and Taylor 2000b; Pintzuk and Taylor 2006 and references therein).

Additionally, the Subject and Object type interaction does not actually affect the course of the change from OV to VO in these languages. It does, however, predict subtle differences in the frequencies of the two variants at every time point along the trajectory of the change. In other words, the lines in Figure 2 largely move together along the change. The Subject and Object is an orthogonal contextual effect, which is present as a constant throughout the OV to VO change. Thus, it is an instance of the “Constant Rate Effect” (Fruehwald et al. 2013; Kroch 1989; Pintzuk 1991; Santorini 1993; Wallenberg 2013 and many others; see the excellent overview of CREs in Kauhanen and Walkden 2018). The CRE describes situations where a syntactic context affects the frequencies of variants involved in a change, but without affecting the direction or rate of the change itself. CREs have now been found in numerous linguistic changes in different grammatical domains,

but particular instances of the CRE are very difficult to explain. In this case, we predicted a CRE from the uniform information density theory, and so information density may actually constitute an explanation for the existence and details of this CRE—we return to this point below.

4.2. OV/VO and Information Density in Icelandic

The additional data from historical Icelandic allows us to directly connect the contextual effects observed above to an actual measure of a sentence's information uniformity: the calibrated DORM, based on the strings of lemmas from a sentence. Interestingly, there was no significant overall effect of OV/VO structure on the information uniformity of the resulting strings, at least according to our preferred model. However, surprisingly, the choice of OV or VO *vP*/VP does have a demonstrable effect on the uniformity of the resulting sentence depending on the choice of Subject and Object, as we predicted. Even in the one model we considered above that does estimate an overall effect of OV/VO on sentence uniformity (the 4-way interaction model), the effect can be ameliorated or reversed by other choices the speaker might make in a given sentence. Under that model, the estimated slopes show that the combination of OV and a pronominal Subject *increases* uniformity at least 1.5 times more ($\beta = -4.17$) than the amount OV in general *decreases* uniformity ($\beta = 2.76$). Of course, the converse is true for VO sentences. It is therefore not at all clear that either OV or VO order is generally more conducive to uniform information density. Rather, speakers can use the different elements in a clause in combination with each other, including OV or VO structures, in order to modulate the information uniformity of the resulting string. These results have three important theoretical consequences: for the role of information uniformity in syntactic planning, for our understanding of the Constant Rate Effect, and for our understanding of OV to VO changes. We outline each of these in turn below.

First, the quantitative results support our hypothesis that there is an interaction between OV/VO, Subject type, and Object Type. This hypothesis was directly suggested by uniform information density theory, and not expected under any existing theory of syntax of which we are aware. The fact this odd prediction was supported constitutes indirect evidence that information density is an important factor in syntactic planning. However, this data also goes further than the initial set of diachronic English and Icelandic data: we can now directly link the interaction between phrase structure and Subject and Object Type to information density quantitatively. Not only did information uniformity predict the interaction, but it correctly predicted the direction of effect on the information density of the resulting sentences. VO structure leads to greater information uniformity when Subject and Object are of the same syntactic type (i.e., both pronominal or both nominal), and therefore more balanced in their information content: in this case, the nonfinite verb can be placed between them and so better balance the overall information distribution of the resulting sentence by mitigating the verb's relative peak in information content. The unbalanced configuration of pronominal Subject and nominal Object, on the other hand, combines better with OV to yield more uniform sentences. When the sentence contains one of the optimal configurations, OV with unbalanced Subject and Object Type or VO with balanced, the DORM of the resulting sentence is lower (more uniform) by about 2.66 bits, a third of the standard deviation of sentence DORMs (see Figure 3). If one of the less optimal configurations occurs, OV with balanced Subject and Object or VO with unbalanced, the resulting sentence is less uniform by about a third of a standard deviation of sentence DORMs. We note also that gapped Subjects patterned with pronominal ones, and it is not surprising that they reduce the information content at the beginning of a clause. In the pronoun case, it is the high frequency of pronouns that means the Subject is low information content; in the gapped case, the Subject is not overt, and so the clause will begin with either a complementizer or a conjunction, both of which are high-frequency closed-class words.

Secondly, as we mentioned above, this interaction is a plausible Constant Rate Effect, and with the DORM data we can link the effect directly to information density. Our results show that a Subject and Object that are matched in type combine with VO to produce more uniform utterances, and unmatched arguments yield greater uniformity with OV. This is exactly the direction of the argument-type effect we observed in Figure 2 and Section 3.1: OV is favored over VO, and vice versa, throughout the time course of the change, only in Subject and Object contexts which result in greater information uniformity (i.e., lower DORMs). This does not mean that speakers get the right combination all of the time; indeed, if they did, we would never be able to use the DORM data to see that the less optimal syntactic configurations increase DORMs. However, an overall speaker bias in favor of greater uniformity will boost the frequencies of OV and VO just slightly in the relevant contexts.

If information uniformity can explain this CRE, perhaps it can explain others. We note that the effect of subordinate vs. matrix clause type on the OV to VO change is a much replicated CRE, and it has been a mystery for some time as to why subordinate clauses favor head-final *vP*/VPs (Hróarsdóttir 2000; Pintzuk and Taylor 2006 for historical Icelandic and English, respectively) more than matrix clauses do (as well as head-final TPs over head-initial, as shown in Pintzuk 1991 for Old English *inter alia*). Indeed, we replicated this CRE as well in our data sets (Figure 1). While we do not currently have a full explanation for the clause type CRE, it is notable that our final model in Section 3.2 does estimate an effect of subordinate clause type interacting with OV to produce informationally more uniform sentences overall. This CRE may also have an explanation in the realm of information uniformity, though we leave a full investigation to future work.

Thirdly, the result should sound a note of caution for theorists hoping to use information density accounts to make broad typological predictions. For instance, Maurits et al. (2010) predict, based on their measure of information uniformity, that “SVO” word orders should be more common than “SOV” orders (“OV”/“VO” here being string descriptions without distinguishing verb types or details of *vP*/VP structure). They then attempt to explain the discrepancy between their prediction and the observed world typology (see also Ferrer-i-Cancho 2017 for a different information theoretic approach to word order typology, and an excellent review of relevant literature). However, any such endeavors should be tempered by this study’s observation that there was no clear overall effect of word order on information density. Even according to the one model that estimated such an effect, it could be reversed and turned into a much larger effect in the opposite direction by manipulating the syntactic characteristics of Subject and Object. Additionally, a snapshot of the current world’s typology may not be the kind of data that can support (or refute) such an explanation. A more appropriate question is: does a given word order variant have an advantage once it happens to be introduced into competition with another variant?

This is not to say that OV does not have an effect on uniformity; it might have a small effect. If it does have a small general effect, this could be seen as a kind of selective pressure in favor of VO *vP*/VPs in the analog of natural selection for language change. However, the fact that the tenuous main effect of OV can be not only cancelled but reversed means that it’s possible for speakers to manipulate various parts of the syntax in concert, and thus maintain a certain level of uniformity in their sentences whether a particular phrase happens to be right-headed or left-headed. The target level of information uniformity is presumably something like the stable average calibrated DORM shown in Figure 4 (≈ 10 bits). Maintaining this level of uniformity is as possible with OV *vP*/VPs as it is with VO ones. If there is a small bias against uniformity in sentences with OV clauses, it could be a selective advantage for VO once it is in competition with OV in a population (and only then). This is worthy of further investigation, particularly if order-based information uniformity is a strategy to prevent catastrophic communication failures in the transmission of a sentence (Cuskley et al. 2020). Once a VO variant is innovated in a population (of speakers and of utterances), the VO and OV variants must be tracked by acquirers in some way, and any systematic effect of noise on the parsing of sentences with an OV *vP*/VP

would lead to a systematic over-representation of the frequency of VO on the part of acquirers. As Heycock and Wallenberg (2013); Yang (2000, 2006) show, a small systematic parsing advantage can lead to the rise of one variant over another, given enough time. However, it could take a very long time indeed, and thus, a very long time to affect the world typology of languages. Furthermore, the selective advantage of VO only would only emerge if that variant is introduced/innovated into an OV population; the probability of its introduction is another matter entirely, and it may never occur in most cases.

4.3. Stability of Information Density over Time

Finally, it is interesting and important to note the lack of any diachronic change in the information density of sentences itself (Figure 4). This is exactly what we would expect if information uniformity is an unconscious strategy on the part of speakers, perhaps to mitigate against potential large noise events that would disrupt communication (Cuskley et al. 2020). If this is a deep property of communication, the bias towards information uniformity in syntactic planning may well be a cognitively deep and evolutionarily old strategy, and therefore constant over historical time. This means that speakers at any time period will use all of the syntactic options available to them, whatever those happen to be at that point in history, in order to maintain a basic level of uniformity in their sentences.

There is an effect of Genre on information density, and we take this to mean that there are additional pressures in favor of dense sentences in genres such as scientific writing (cf. e.g., Biber and Gray 2013; Bizzoni et al. 2020; Degaetano-Ortlieb and Teich 2019), law, and religious texts; these are the non-narrative genres in IcePaHC. This difference in non-narrative genres may be conceived as an effect of ‘distance’ from spoken language, which would be related to (for example) the effect of literary vs. oral genre on information theoretic measures of language (Ortmann and Dipper 2019, 2020). Informationally dense sentences are not necessarily non-uniform, but uniformity will be harder to achieve if there are many fixed, large peaks in lexical information content. However, this simply means that an existing speaker bias towards information uniformity can intersect with a genre bias in certain types of writing. The only possible trend over time in DORM is in VO clauses from narrative texts. As the slope was not significantly non-zero, we should be cautious interpreting it, but it could reflect a very mild stylistic change: the 19th century Icelandic independence movement brought with it an approach to writing that sought to emulate the simple sentence structure and vocabulary choices of the Icelandic Sagas (Ottósson 1990, pp. 54–57). However, this is merely speculation.

Interestingly, there is a spike in calibrated DORM in the non-narrative texts at 1540, which can only be due to the translation of the New Testament by Oddur Gottskálksson. This translation is notable for two reasons: first, it is based primarily on Luther’s German *Septembertestament*. Secondly, Oddur Gottskálksson spent some of their early years in Norway, and may have some semi-non-native features in the morphosyntax of their Icelandic (Einarsson 1988; Helgason 1929; Kvaran et al. 1988; Steinþórsson 2015). We cannot say which of these factors gives rise to the outlying DORM, but both of these factors share a commonality: the influence of a translation and interference from an L1 both have the effect of constraining the syntactic options of the author. According to our interpretation of the data above, speakers unconsciously use a variety of syntactic means in concert to maintain a level of uniformity in their sentences. This requires a great deal of planning and a fluent ability to manipulate a variety of linguistic features simultaneously. It is therefore entirely in line with our analysis that *any* additional constraint placed on a person’s syntactic options, or their ability to rapidly plan a combination of interlocking linguistic features, would lead to an overall decrease in uniformity (higher sentence DORMs).

5. Conclusions

This study used the theory of uniform information density to make some very precise quantitative predictions about effects of syntactic context on the OV to VO change in English and Icelandic. Specifically, we predicted that when Subject and Objects are of

the same type (pronominal or non-pronominal), VO *vP*/VPs would be slightly favored because they yield more informationally uniform sentences in this context. We predicted the converse for OV. The study demonstrated that these hypotheses hold in both English and Icelandic. Furthermore, they constitute an instance of the Constant Rate Effect; the effect of syntactic context does not drive the OV to VO change at all, but simply modulates the frequencies of the variants along the whole time course of the change. We suggested that if uniform information density could explain this CRE, it may be useful to explain others as well.

Finally, we show that the average information uniformity of sentences does not change over the history of Icelandic. Its constancy fits with a theory in which a speaker bias towards informationally uniform utterances is a general property of the human faculty for communication.

Author Contributions: Conceptualization, J.C.W.; methodology, J.C.W. and A.K.I.; software, A.K.I.; statistical analysis, J.C.W.; writing—original draft preparation, J.C.W. and R.B. and A.K.I.; writing—review and editing, J.C.W. and R.B. and C.C. and A.K.I.; data visualization, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Economic and Social Sciences Research Council (ESRC, UK) grant number ES/T005955/1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The secondary data sets generated from the corpora, which we used for statistical analyses, may be found in the following public repositories: <https://github.com/joelcw/constantentropy/tree/master/outputs>; <https://github.com/joelcw/iceBits> (accessed on 23 March 2021).

Acknowledgments: We would like to thank Anthony S. Kroch for inspiring this line of research, and many enlightening discussions on the theoretical background and methodology.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

OV/VO	Object Verb/Verb Object
DORM	Deviation of Rolling Mean
IcePaHC	Icelandic Parsed Historical Corpus
YCOE	York Corpus of Old English Prose
PPCME2	Penn Parsed Corpus of Middle English, Second Edition
PPCEME	Penn Parsed Corpus of Early Modern English
PPCMBE	Penn Parsed Corpus of Modern British English
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion

References

- Aylett, Matthew. 1999. Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and syllabic duration. Paper presented at the ICPHS-99, San Francisco, CA, USA, August 1–7.
- Aylett, Matthew, and Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47: 31–56. [CrossRef] [PubMed]
- Bergen, Leon, and Noah D. Goodman. 2015. The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science* 7: 336–50. [CrossRef] [PubMed]
- Biber, Douglas, and Bethany Gray. 2013. Nominalizing the verb phrase in academic science writing. *The Verb Phrase in English: Investigating Recent Language Change with Corpora* 9: 1–27.

- Bizzoni, Yuri, Stefania Degaetano-Ortlieb, Peter Fankhauser, and Elke Teich. 2020. Linguistic variation and change in 250 years of English scientific writing: A data-driven approach. *Frontiers in Artificial Intelligence, Section Language and Computation*. [CrossRef]
- Cuskley, Christine, Rachael Bailes, and Joel C. Wallenberg. 2020. Noise Resistance in Communication: Quantifying Uniformity and Optimality. Available online: <https://psyarxiv.com/wpvq4> (accessed on 11 March 2021).
- Degaetano-Ortlieb, Stefania. 2018. Stylistic variation over 200 years of court proceedings according to gender and social class. Paper presented at the 2nd Workshop on Stylistic Variation collocated with NAACL HLT 2018, New Orleans, LA, USA, June 1–6; pp. 1–10.
- Degaetano-Ortlieb, Stefania, and Elke Teich. 2019. Toward an optimal code for communication: The case of scientific English. *Corpus Linguistics and Linguistic Theory (Open Access)* 1–33. [CrossRef]
- Einarsson, Sigurbjörn. 1988. Oddur Gottskálksson. In *Nýja testamenti Odds Gottskálkssonar*. Edited by Sigurbjörn Einarsson, Guðrún Kvaran, Gunnlaugur Ingólfsson and Jón Aðalsteinn Jónsson. Reykjavík: Lögberg, pp. VII–XX.
- Fenk, August, and Gertraud Fenk. 1980. Konstanz im kurzzeitgedächtnis-konstanz im sprachlichen informationsfluß. *Zeitschrift für Experimentelle und Angewandte Psychologie* 27: 402.
- Ferrer-i-Cancho, Ramon. 2017. The placement of the head that maximizes predictability. An information theoretic approach. *arXiv* arXiv:1705.09932
- Frank, Austin F., and T. Florian Jaeger. 2008. Speaking Rationally: Uniform Information Density as an Optimal Strategy for Language Production. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Volume 30. Available online: <https://escholarship.org/uc/item/7d08h6j4> (accessed on 11 March 2021).
- Fruehwald, Josef, Jonathan Gress-Wright, and Joel C. Wallenberg. 2013. Phonological Rule Change: The Constant Rate Effect. In *Proceedings of the 40th Annual Meeting of the North East Linguistic Society*. Edited by Seda Kan, Claire Moore-Cantwell and Robert Staubs. Amherst: GLSA (Graduate Linguistic Student Association) Publications.
- Genzel, Dmitriy, and Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, pp. 199–206.
- Gundel, Jeanette K. 1988. Universals of Topic-Comment Structure. pp. 209–39. Available online: <https://benjamins.com/catalog/tsl.17.16gun> (accessed on 11 March 2021).
- Helgadóttir, Sigrún, Ásta Svavarsdóttir, Eiríkur Rögnvaldsson, Kristín Bjarnadóttir, and Hrafn Loftsson. 2012. The tagged Icelandic corpus (MÍM). In *Proceedings of the Workshop on Language Technology for Normalisation of Less-Resourced Languages—SaLTMiL 8*. Istanbul: European Language Resources Association (ELRA), pp. 67–72.
- Helgason, Jón. 1929. *Málið á Nýja Testamenti Odds Gottskálkssonar*. Copenhagen: Hið íslenska fræðafjelag.
- Heycock, Caroline, and Joel C. Wallenberg. 2013. How variational acquisition drives syntactic change: The loss of verb movement in Scandinavisation. *The Journal of Comparative Germanic Linguistics* 16: 127–57. [CrossRef]
- Hróarsdóttir, Thorbjörg. 2000. *Word Order Change in Icelandic. From OV to VO*. Number 35 in Linguistik Aktuell. Amsterdam: John Benjamins.
- Ingason, Anton Karl. 2016. PaCQL: A new type of treebank search for the digital humanities. *Italian Journal of Computational Linguistics* 2: 51–66 [CrossRef]
- Ingason, Anton Karl, Einar Freyr Sigurðsson, and Joel Wallenberg. 2012. A Paper Presented at DiGS. Antisocial Syntax. Disentangling the Icelandic VO/OV Parameter and its Lexical Remains. Available online: <https://www.staff.ncl.ac.uk/joel.wallenberg/papers/antisocial2012.pdf> (accessed on 11 March 2021).
- Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61: 23–62. [CrossRef]
- Kauhanen, Henri, and George Walkden. 2018. Deriving the Constant Rate Effect. *Natural Language & Linguistic Theory* 36: 483–521.
- Kiparsky, Paul. 1996. The Shift to Head-Initial VP in Germanic. In *Studies in Comparative Germanic Syntax II*. Edited by Höskuldur Thráinsson, Samuel David Epstein and Steve Peter. Dordrecht: Kluwer Academic Publishers, pp. 140–179.
- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. 2004. Penn-Helsinki Parsed Corpus of Early Modern English, Release 3. Size 1.8 Million Words. Available online: <https://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-3/index.html> (accessed on 11 March 2021).
- Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199–244. [CrossRef]
- Kroch, Anthony S., Beatrice Santorini, and Ariel Diertani. 2016. Penn Parsed Corpus of Modern British English 2nd Edition, Release 1. Size ~ 2.8 Million Words. Available online: <https://www.ling.upenn.edu/hist-corpora/PPCMBE2-RELEASE-1/index.html> (accessed on 11 March 2021).
- Kroch, Anthony S., and Ann Taylor. 2000a. Penn-Helsinki Parsed Corpus of Middle English. CD-ROM. Second Edition, Release 4. Size: 1.3 Million Words. Available online: <https://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-4/index.html> (accessed on 11 March 2021).
- Kroch, Anthony S., and Ann Taylor. 2000b. Verb-complement order in Middle English. In *Diachronic Syntax: Models and Mechanisms*. Edited by Susan Pintzuk, George Tsoulas and Anthony Warner. Oxford: Oxford University Press, pp. 132–63.
- Kvaran, Guðrún, Gunnlaugur Ingólfsson, and Jón Aðalsteinn Jónsson. 1988. Um þýðingu Odds og útgáfu þessa. In *Nýja testamenti Odds Gottskálkssonar*. Edited by Sigurbjörn Einarsson, Guðrún Kvaran, Gunnlaugur Ingólfsson and Jón Aðalsteinn Jónsson. Reykjavík: Lögberg, pp. XXI–XXXII.

- Levy, Roger P., and Florian T. Jaeger. 2007. Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems*. Boston: MIT Press, pp. 849–56.
- Maurits, Luke, Dan Navarro, and Amy Perfors. 2010. Why are some word orders more common than others? A uniform information density account. In *Advances in Neural Information Processing Systems*. New York: Curran Associates Inc, pp. 1585–93.
- Ortmann, Katrin, and Stefanie Dipper. 2019. Variation between different discourse types: Literate vs. oral. Paper presented at the NAACL-Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial), Minneapolis, MN, USA, June 7; pp. 64–79.
- Ortmann, Katrin, and Stefanie Dipper. 2020. Automatic orality identification in historical texts. In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC)*. Marseille: European Language Resources Association, pp. 1293–302.
- Ottósson, Kjartan. 1990. *Íslensk málhreinsun*. Reykjavík: Íslensk málnefnd.
- Pierce, John R. 1980. *An Introduction to Information Theory: Symbols, Signals and Noise*, 2nd ed. Kent: Dover.
- Pintzuk, Susan. 1991. Phrase Structures in Competition: Variation and Change in Old English Word Order. Ph. D. thesis, University of Pennsylvania, Pennsylvania, PA, USA.
- Pintzuk, Susan, and Ann Taylor. 2004. Objects in Old English: Why and how Early English is not Icelandic. In *York Papers in Linguistics*. Edited by Jonny Butler, Davita Morgan, Leendert Plug and Gareth Walker. Series 2. Heslington York: University of York, Department of Language and Linguistic Science, Issue 1.
- Pintzuk, Susan, and Ann Taylor. 2006. The loss of OV order in the history of English. In *Blackwell Handbook of the History of English*. Edited by Ans van Kemenade and Bettelou Los. Oxford: Blackwell, pp. 247–78.
- Pluymaekers, Mark, Mirjam Ernestus, and Harald Baayen. 2005. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62: 146–159. [CrossRef] [PubMed]
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Randall, Beth. 2013. Corpussearch 2: A Tool for Linguistics Research. Available online: <http://sourceforge.net/projects/corpussearch/> (accessed on 11 March 2021).
- Rögnvaldsson, Eiríkur. 1984. Rightward displacement of NPs in Icelandic—formal and functional characteristics. In *The Nordic Languages and Modern Linguistics*. Edited by Kristian Ringgaard and Viggo Sørensen. Århus: Nordisk Institut, Århus Universitet, Number 5, pp. 361–268.
- Rögnvaldsson, Eiríkur. 1987. OV word order in Icelandic. In *Proceedings of the Seventh Biennial Conference of Teachers of Scandinavian Studies in Great Britain and Northern Ireland*. London: University College, pp. 33–49.
- Rögnvaldsson, Eiríkur. 1994. Breytileg orðaröð í sagnlið. *Íslenskt mál* 16–17: 27–66.
- Rögnvaldsson, Eiríkur. 1996. Word Order Variation in the VP in Old Icelandic. *Working Papers in Scandinavian Syntax* 58: 55–86.
- Rögnvaldsson, Eiríkur, Anton Karl Ingason, Einar Freyr Sigurðsson, and Joel Wallenberg. 2012. The Icelandic Parsed Historical Corpus (IcePaHC). In *Proceedings of LREC*. Istanbul: European Language Resources Association (ELRA), pp. 1977–84.
- Rögnvaldsson, Eiríkur, Anton Karl Ingason, Einar Freyr Sigurðsson, and Joel C. Wallenberg. 2011. Creating a dual-purpose treebank. *Journal for Language Technology and Computational Linguistics* 2: 141–52.
- Santorini, Beatrice. 1993. The rate of phrase structure change in the history of Yiddish. *Language Variation and Change* 5: 257–83. [CrossRef]
- Shannon, Claude Elwood. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27: 379–423. [CrossRef]
- Sigurðsson, Halldór Ármann. 1983. Um frásagnarumröðun og grundvallarorðaröð í forníslensku. Cand.mag. thesis, University of Iceland, Reykjavík, Iceland.
- Steinþórsson, Atli Freyr. 2015. Bækurnar í fjósinu. Um fyrirmyndir Nýja testamentis Odds Gottskálkssonar. BA thesis, University of Iceland, Reykjavík, Iceland.
- Taylor, Ann, and Susan Pintzuk. 2011. The interaction of syntactic change and information status effects in the change from OV to VO in English. *Catalan Journal of Linguistics* 10: 71–94. [CrossRef]
- Taylor, Ann, and Susan Pintzuk. 2012. *Rethinking the OV/VO Alternation in Old English: The Effect of Complexity, Grammatical Weight and Information Status*. Volume The Oxford Handbook on the History of English. Oxford: Oxford University Press, pp. 1199–213.
- Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.
- Turk, Alice. 2010. Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology* 1: 227–62. [CrossRef]
- Wallenberg, Joel C. 2013. Scrambling, LF, and Phrase Structure Change in Yiddish. *Lingua* 133: 289–318. [CrossRef]
- Wallenberg, Joel C., Anton K. Ingason, Einar F. Sigurðsson, and Eiríkur Rögnvaldsson. 2011. Icelandic Parsed Historical Corpus (IcePaHC). Version 0.9. Size: 1 Million Words. Available online: http://www.linguist.is/icelandic_treebank (accessed on 11 March 2021).
- Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12: 231–50. [CrossRef]
- Yang, Charles. 2006. *The Infinite Gift: How Children Learn and Unlearn the Languages of the World*. New York: Scribner.
- Zarcone, Alessandra, Marten van Schijndel, Jorrig Vogels, and Vera Demberg. 2016. Salience and attention in surprisal-based accounts of language processing. *Frontiers in Psychology* 7. [CrossRef]
- Zhan, Meilin, and Roger P. Levy. 2018. *Comparing Theories of Speaker Choice Using a Model of Classifier Production in Mandarin Chinese*. Stroudsburg: Association for Computational Linguistics.