

Uniform processing difficulty is a poor predictor of cross-linguistic word order frequency

Brennan Gonerig

Linguistics Department
University of California, Davis
btgonering@ucdavis.edu

Emily Morgan

Linguistics Department
University of California, Davis
eimorgan@ucdavis.edu

Abstract

Some have argued that word orders which are more difficult to process should be rarer cross-linguistically. Our current study fails to replicate the results of Maurits, Navarro, and Perfors (2010), who used an entropy-based Uniform Information Density (UID) measure to moderately predict the Greenbergian typology of transitive word orders. We additionally report an inability of three measures of processing difficulty - entropy-based UID, surprisal-based UID, and pointwise mutual information - to correctly predict the correct typological distribution, using transitive constructions from 20 languages in the Universal Dependencies project (version 2.5). However, our conclusions are limited by data sparsity.

1 Introduction

Cross-linguistic typologies of word order preferences have shown a robust pattern for orderings of subjects, objects, and verbs (see the World Atlas of Language Structures (WALS) data in Table 3 and Figure 4) (Greenberg, 1963; Hawkins, 2014). Several hypotheses have been proposed to explain this pattern, including the Animate First Principle, Verb-Object Bonding, and the Theme First Principle (Tomlin, 1986). However, these are somewhat circular in their reasoning: they are motivated based on the prevalence of subject-first languages and languages where the verb and object are adjacent. These principles are then used to explain the very same word order patterns that motivated them in the first place. With the goal of achieving a more robust explanation, some have argued in more recent work that processing difficulty (broadly construed) is inversely proportional to typological frequency (Hawkins, 2004, 2014; Fedzechkina et al., 2018). Languages are claimed to evolve from word orders that are more difficult to process to orders that are easier to process. Therefore, a word order that is more difficult to process should be rarer

cross-linguistically. Within this context, Maurits et al. (2010) attempted to use the Uniform Information Density (UID) hypothesis to explain the distributional patterns of subject-verb-object constructions. Maurits et al. (2010) (henceforth MNP) constructed simplified sets of events, ranking each possible ordering of subject, verb, and object by deviation from UID ideal.

MNP's results showed a consistent high ranking for VSO and VOS, and a consistent low ranking for SOV, contrary to the observed cross-linguistic frequency. While MNP used an entropy-based UID measure, there are other measures of processing difficulty that could be applied to the underlying question of whether languages evolve from difficult-to-process word orders to easy-to-process word orders. This paper therefore seeks to test the claim that languages evolve in this way, by comparing the predicted processing difficulty of the six possible subject-verb-object word orders. To do this, we use the following three metrics of processing difficulty, defined in the following sections: deviation from a UID ideal based on changes in entropy, deviation from a UID ideal based on surprisal, and total pointwise mutual information.

To that end, this paper is divided into two major components. In the first, we attempted to replicate the entropy-based UID results MNP obtained on their toyworld and on child-directed corpus data. In the second, we evaluate the ranking of word orders produced by each of the three processing difficulty measures using event datasets derived from corpora in 20 different languages.¹

2 Processing difficulty measures

2.1 Entropy

The UID hypothesis models human language as a noisy communication channel. A speaker/signer

¹All code and data for this project are available at https://github.com/ucdavis/gonerig_morgan-con112020

should attempt to convey as much information as possible for a comprehender to reasonably reconstruct their message, given the possibility for misunderstanding, interference, cognitive misrepresentation, or literal noise in the surrounding environment (Levy and Jaeger, 2007; Frank and Jaeger, 2008). Under the UID hypothesis, the optimal solution is for a speaker/signer to spread the information as evenly as possible across the component parts of a message. Theoretically, this increases the likelihood of successful reconstruction by the comprehender (Levy and Jaeger, 2007; Frank and Jaeger, 2008). Additionally, because language is modelled as a noisy channel, there exists some maximum speed at which a language user can communicate successfully – termed the channel capacity. The UID hypothesis therefore further claims that as a language user approaches this capacity, they will make linguistic choices that distribute the information density of their utterance more uniformly. This should enable a language user to communicate both quickly and without errors (Levy and Jaeger, 2007).

“Information” can be conceptualized as the amount of uncertainty that is reduced by a message, calculated using entropy. If we model words in an incoming language stream as random variables which identify a random event, we can calculate how much each word reduces the uncertainty (entropy) about what event is being described. We can thereby quantify how much information each word contains. For the purposes of investigating transitive word order preferences, we adopt here the same simplified model as MNP. Each event is composed of an agent², a , drawn randomly from a set of possible agents, A ; a verb, v , drawn randomly from a set of possible verbs, V ; and a patient, p , drawn randomly from a set of possible patients, P . We then define an event, (a, v, p) , as a variable drawn randomly from a set of possible events, (A, V, P) . We can then assume a particular event is discussed with probability $P(a, v, p)$. Because we represent an event using just these three components, the words of an utterance based on such an event can be artificially permuted to derive all six possible word orders.

In this model, we assume that the comprehender always begins with the maximum uncertainty about

²In part because semantic and grammatical roles are conflated in our event sets, as well as to avoid some issues with notation, we refer to subjects in our model as agents and objects as patients.

what will be discussed:

$$H_0 = - \sum_{a,v,p \in A,V,P} P(a, v, p) \log_2(P(a, v, p))$$

After all words are communicated (and sometimes before), the entropy will always be zero. This signifies that the comprehender has uniquely identified the event discussed out of all the other potential events. In between, each word provides an amount of information to decrease the comprehender’s uncertainty about the event being discussed. The information provided by a word, then, is the change in uncertainty, calculated as the difference in entropy before and after the word was communicated: for word i , $I_i = H_{i-1} - H_i$.

Given the simplified three word utterance modelled here, an utterance that achieves the UID ideal would have exactly $\frac{1}{3}$ of the total entropy, H_0 , provided by each word. How far a particular utterance deviates from this ideal is therefore dependent on how much information each word provides above or below this $\frac{1}{3}$ mark. Figure 1 illustrates the deviation from this entropy-based UID ideal for two events taken from MNP’s toyworld, (Eve,ate,rice) and (Eve,ate,bread) (see Table 1), in SOV and VOS order. By summing the individual deviations of each word, we can arrive at an overall deviation score for a particular utterance. Multiplying the current deviation sum by $\frac{3}{4}$ rescales the maximum deviation value to 1, deriving a final equation:

$$D(I) = \frac{3}{4} \sum_{i=1}^3 \left| \frac{I_i}{H_0} - \frac{1}{3} \right|$$

A weighted average of the deviation scores for each utterance in a word order, using the probability of the event as the weight, creates an overall score for that word order. This metric allows us to systematically compare how close each of the six possible orderings of subjects, verbs, and objects approach the ideal of uniform information density. This is the deviation metric that MNP used, and which we adopt here as well.

2.2 Surprisal

Surprisal is defined as the negative log of a probability (i.e. $S(x) = -\log(P(x))$) and has been posited as a good model of incremental language processing, particularly in capturing difficulty of processing a sentence (Hale, 2001; Levy, 2011). Surprisal is closely related to entropy and is compatible with the UID hypothesis as well. In this case the UID hypothesis would state that the total surprisal of an event should be distributed equally

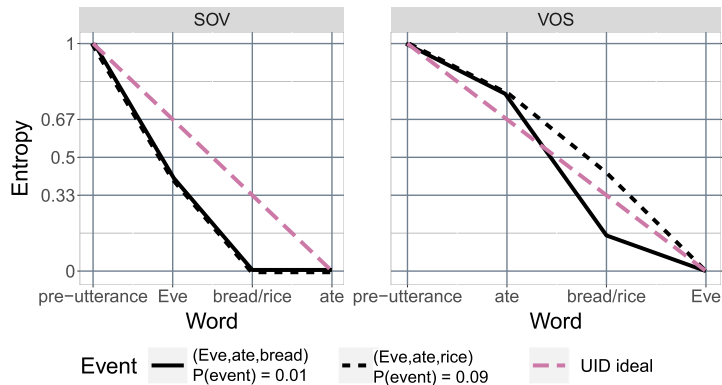


Figure 1. The deviation of the higher-probability event (Eve,ate,rice) (dotted black lines) and the lower-probability event (Eve,ate,bread) (solid black lines) from the UID ideal (dashed purple lines), using entropy to quantify the amount of information per word. Example utterances and probabilities were taken from the MNP toyworld.

across the words of an utterance. As such, surprisal provides a measure that is sensitive to the probabilities of each word of an utterance, and the probability of the utterance as a whole, whereas the MNP entropy-based measure was not. More specifically, the surprisal of an event is given by $S(event) = -\log_2 P(event)$, and the surprisal of the individual words that make up an utterance is the negative log probability of that word, given the history of words previously encountered (i.e. $S(word) = -\log_2 P(word|history)$).

We define here a surprisal-based deviation from UID ideal, analogous to the entropy-based deviation proposed by MNP:

$$D(S_{utt}) = \frac{3}{4} \sum_{i=1}^3 \left| \frac{S_{word_i}}{S_{event}} - \frac{1}{3} \right|$$

Again, a weighted average of the deviation scores for each utterance in a word order, using the probability of the event as the weight, creates an overall score for that word order. Figure 2 illustrates the deviation from this surprisal-based UID ideal for two events (see Table 1) in SOV and VOS order. Fig. 2 also underscores that despite passing similarities in the underlying equations, these two processing difficulty measures may indeed lead to ultimately different results.

2.3 The information locality hypothesis

So far we have discussed the information structure of utterances within the framework of a single theory, UID. The UID hypothesis is not without critique though. Specifically, UID ignores findings from traditional linguistics about how the information structure of utterances is specifically nonuniform, e.g. new information and contrastive focus (Rizzi, 1997; Choudhury, 2015). Information local-

ity on the other hand, offered as an explanation for a range of syntactic phenomena, claims that words with high mutual information are placed as close as possible to each other in order to ease processing for a comprehender (Futrell, 2019; Futrell et al., 2019; Gibson et al., 2019).

In the context of the current paper, information locality would predict that the components of a transitive utterance should be placed such that the pointwise mutual information of adjacent words is maximized:

$$\begin{aligned} pmi_{utt} &= pmi(word_1, word_2) + pmi(word_2, word_3) \\ &= \log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)} + \log_2 \frac{P(word_2, word_3)}{P(word_2)P(word_3)} \end{aligned}$$

Note that because of this formulation, this measure does not distinguish between mirror word orders (i.e. SOV has the same overall PMI value as VOS). One method, which we adopt here, of converting these PMI values into a metric that can be compared across event sets is to z-score the pointwise mutual information values. We can use this standardized PMI score to rank word orders from greatest total PMI score to lowest. Figure 3 illustrates the total pointwise mutual information for two events (see Table 1) in SOV/VOS, SVO/OVS, and VSO/OSV orders.

3 Attempted replication of MNP

3.1 Background

MNP used three different methods to construct sets of events and event probabilities: they constructed an artificial set of events as a “toyworld,” they extracted transitive utterances from corpora, and finally they asked participants to rank the likelihood of certain events occurring. We focus here on the first two methods they used. Their corpora-derived

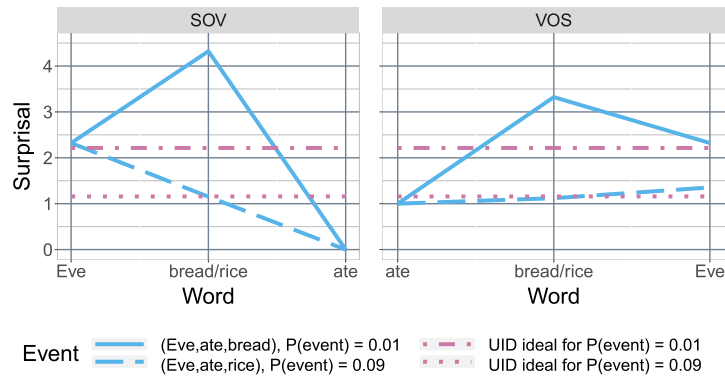


Figure 2. The deviation of the higher-probability event (Eve,ate,rice) (dashed blue lines) and the lower-probability event (Eve,ate,bread) (solid blue lines) from the UID ideal (dotted and dot-dashed purple lines), using surprisal to quantify the amount of information per word.

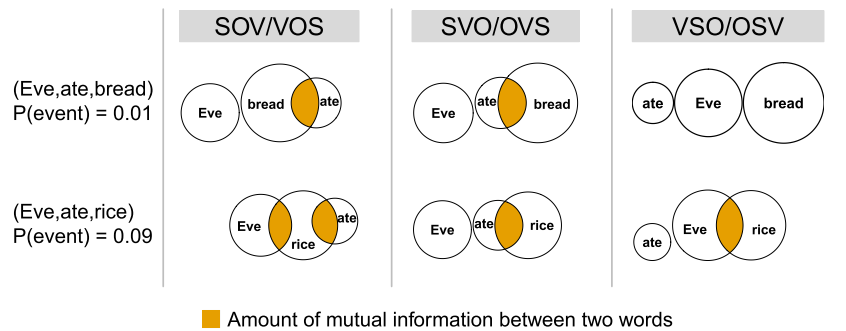


Figure 3. The pointwise mutual information of the events (Eve,ate,bread) and (Eve,ate,rice). Each circle represents the information content of the labeled word, measured as $-\log_2(P(\text{word}))$, and the orange-shaded intersections represent the pointwise mutual information between two adjacent labeled words.

event sets were constructed by extracting transitive utterances from the Brown corpus of English child-directed speech from the CHILDES database. From the Brown corpus, MNP extracted transitive events from adult utterances in the transcripts related to just one child (of the three in the Brown corpus), “Adam.” They discarded events involving pronouns that could refer to a broad range of agents or patients (e.g. “it” and “this”), resulting in a total of 544 transitive utterances. Each unique utterance was defined as an event, and these events were assigned a probability based on the frequency that an utterance involving the event was observed in the corpus. MNP’s results showed a consistent low ranking for SOV and consistent high rankings for both VSO and VOS, in contrast to their actual distributions (see Table 2).

3.2 Methods

We tested the entropy-based measure on the original toyworld events from MNP and a dataset of event probabilities generated by extracting transitive utterances from the Brown corpus of child-directed speech (Brown, 1973; MacWhinney, 2000). We created two separate, exact implementations of the entropy-based UID algorithm described

by MNP, one in the Java programming language and the other in R. In recreating the MNP toyworld, we used the exact probability distributions specified in their supplementary materials, except for two events with clear typographical errors. We then input these event probabilities into both the Java and R algorithm implementations, to derive weighted mean entropy-based UID deviation scores for each of the six possible word orders. The word orders were finally ranked from least deviation to greatest.

In attempting to replicate MNP’s results on child-directed speech data, we decided to use all utterances from all transcripts in the Brown corpus (Brown, 1973; MacWhinney, 2000). This was motivated by a desire to lessen the potential impact of data sparsity. We also did not exclude utterances with broad-referent pronouns (such as “it” or “what”), as MNP did, so as not to exclude highly probable, naturalistic speech. In total, 30,166 transitive utterances were automatically extracted using an algorithm we created for that purpose, with 14,381 unique utterance types. In comparison, MNP used just 544 utterance tokens. We then assigned each transitive utterance a probability based

	Eat				Drink			
	Apple	Bread	Cake	Rice	Coffee	Cola	Juice	Water
Alice	0.05	0.00	0.03	0.02	0.07	0.03	0.00	0.00
Bob	0.02	0.00	0.04	0.04	0.02	0.04	0.02	0.02
Eve	0.00	0.01	0.00	0.09	0.03	0.01	0.00	0.06
Mallory	0.04	0.04	0.01	0.01	0.00	0.01	0.09	0.00
Trent	0.02	0.00	0.01	0.07	0.02	0.03	0.03	0.02

Table 1. MNP’s toyworld. Each cell represents the joint probability $P(a,v,p)$ of an event.

on the proportion of times it occurred in the corpus, similarly to MNP. Then, we input these event probabilities into the Java algorithm implementation to derive weighted mean entropy-based UID deviation scores for each of the six of the possible word orders. The word orders were finally ranked from least deviation to greatest.

3.3 Results

Both the Java and R implementations of the MNP entropy-based UID model failed to replicate their toyworld results. However, both of these implementations did produce the exact same ranking and the exact same numerical results (Table 2). Having used their exact event probabilities and an exact recreation of their algorithm (as best we understand it), we expected to get the exact same rankings as they did. We note though that they did not release their numerical results for us to compare to. We further failed to replicate MNP’s ranking from the “Adam” transcripts-derived data using our full Brown-derived data (Table 2). The difference between our Brown corpus based rankings and MNP’s “Adam” transcripts based rankings seems larger than the differences between the two toyworld rankings. Although the Brown-derived dataset we used was larger than MNP’s, we still expected substantial overlap in the word order rankings.

3.4 Discussion

Despite failing to replicate MNP’s results exactly, as expected, we are inclined to trust our results. Both of our separate implementations of their model – in two different programming languages – produced exactly identical numerical and rank order results. This is also backed by our further failure to replicate MNP’s results using child-derived speech. Although we did not expect perfect alignment of results, our ranking appears, qualitatively, to be quite different from that reported by MNP (Table 2). Taken together, these results call into question all of the rankings reported by MNP. Ad-

ditionally, our results cast doubt on MNP’s main conclusion that processing difficulty, as conceptualized by the Uniform Information Density hypothesis, is likely to have played a role in the evolution of subject-verb-object word order preferences across the world’s languages.

4 Universal Dependencies languages

4.1 Methods

Having evaluated just one processing difficulty measure on data from a single language, we next sought to test all three measures on event probabilities derived from corpus data from a wide variety of languages. For this task, the Universal Dependencies (UD) project (version 2.5) was chosen (Zeman et al., 2019). The claim that languages evolve from orders that are difficult to process to orders that are easier to process assumes that certain word orders are inherently easier or more difficult to process. The dominant word order of a language from which we derive event data should therefore not matter. Nevertheless, to try to reduce this potential confound, we chose 20 languages from the UD project, from as many different dominant word orders as possible, based on two primary criteria: (1) genetic diversity, and (2) a large dataset in the UD treebanks. The languages we chose consisted of six SVO-dominant languages (English, Mandarin, Mbya Guarani, Indonesian, Vietnamese, and Wolof), five SOV-dominant languages (Basque, Hindi, Japanese, Korean, and Turkish), five VSO-dominant languages (Breton, Irish, Scottish Gaelic, Tagalog, and Welsh), and four languages considered to have no single dominant subject-verb-object word order (Dutch, German, Hungarian, and Warlpiri) according to WALS (Dryer, Matthew S., 2013). Transitive constructions were thus extracted from all treebanks that included the underlying text and the CoNLL-U parse annotation. Lemma-level information from the CoNLL-U annotation, instead of word-level information, was extracted whenever possible. In

MNP’s toyworld rankings	Our toyworld replication rankings	MNP’s rankings from the “Adam” transcripts	Our Brown corpus rankings
VSO	VOS, 0.206	VSO, 0.38	SVO, 0.427
VOS	VSO, 0.288	SVO, 0.41	SOV, 0.439
SVO	SVO, 0.291	VOS, 0.48	VSO, 0.585
OVS	SOV, 0.500	SOV, 0.64	VOS, 0.588
SOV	OVS = OSV, 0.530	OSV, 0.78	OVS = OSV, 0.684
OSV		OVS, 0.79	

Table 2. Side-by-side comparison of MNP’s results and our attempted replication.

addition to the words of each utterance, we also automatically determined the word order of each construction extracted.

Each extracted event was assigned a probability based on the frequency it was observed in the UD treebanks for a given language. Despite the large size of many of the underlying treebanks, this method proved to create a very sparse model. This led to underestimating the probabilities of events that occurred rarely in the corpora while overestimating the probabilities of events that occurred more frequently. This also influences the amount of information each word provides: if an event occurs rarely, the words that make up an utterance corresponding to it will better disambiguate the event than those words otherwise would. Across our subset of languages, more than 46% of utterances tested in each language were deterministically predicted by the first two words alone. Due to this, we eliminated events where any one of the words only occurred once in a language’s corpus. This improved the problem of sparsity somewhat, dropping the proportion of deterministically predictable utterances from a minimum of 46% to 8%. However, some languages still had 100% of utterances being deterministically predicted by the first two words, with a mean across the subset of 72%. The consequences of this will be discussed in more detail in section 4.4.

The weighted mean entropy-based and surprisal-based UID scores for each word order in each language of our subset were aggregated, then averaged over languages to derive an overall deviation score for each word order. In doing so, we made the assumption that a subject is as predictive of a verb in one language as in every other; that a verb is as predictive of an object in one language as in every other; etc. In order to compare pointwise mutual information scores across languages, raw PMI scores for every utterance were z-scored within each language. This is necessary because different sized event sets in each language would lead to different

magnitudes of raw PMI scores. A weighted mean for each word order within each language was then taken. The means for each word order were finally averaged across languages.

4.2 Results

As shown in Table 3, the entropy-based and surprisal-based UID measures both generated the same ordinal rankings for word orders, though their deviation scores differed. These results differ partially from the observed distribution of transitive word orders. To test whether means for each word order were significantly different from each other or not, a one-way ANOVA with unequal variance was performed for each processing difficulty measure. Word orders did not show significant differences in means for either the entropy-based or the surprisal-based UID measures, ($F(5,53) = 1.40, p = 0.239$ and $F(5,53) = 1.91, p = 0.108$, respectively).

Our mutual information measure likewise failed to produce a ranking of word orders that resembles the actual cross-linguistic distribution (Table 3). A one-way ANOVA with unequal variance indicated a significant effect of word order for this measure ($F(5,53) = 3.65, p = 0.00649$). Post hoc two-tailed Welch’s t-tests indicated that SOV/VOS orders had a significantly greater mean PMI value than either VSO/OSV ($t(34) = 2.54, p = 0.0157$) or SVO/OVS ($t(37) = 2.98, p = 0.00508$) at the 95% confidence level. However, there was no significant difference in mean PMI values between VSO/OSV and SVO/OVS ($t(37) = 0.700, p = 0.488$).

Under the hypothesis we test, our three measures of processing difficulty should produce a distribution of highest-ranked word orders that, when aggregated, roughly matches the observed typological distribution. Figure 4 shows the proportion of word orders that were ranked highest by each processing difficulty measure across our subset of 20 languages from the UD project.

WALS	Entropy-based UID	Surprisal-based UID	Pointwise mutual information
SOV (47.6%)	VSO, 0.583 (0.163)	VSO, 0.604 (0.141)	SOV = VOS, -0.00677 (0.229)
SVO (41.1%)	VOS, 0.590 (0.172)	VOS, 0.612 (0.146)	VSO = OSV, -0.166 (0.160)
VSO (8.0%)	SVO, 0.628 (0.183)	SVO, 0.652 (0.153)	SVO = OVS, -0.204 (0.187)
VOS (2.1%)	SOV, 0.648 (0.179)	SOV, 0.674 (0.147)	
OVS (0.9%)	OVS, 0.694 (0.205)	OVS, 0.716 (0.175)	
OSV (0.3%)	OSV, 0.699 (0.197)	OSV, 0.720 (0.169)	

Table 3. Rankings of word orders generated by our three processing difficulty measures, compared to the observed distribution of word orders as catalogued by WALS. Mean values are indicated after the word order with standard deviations in parentheses.

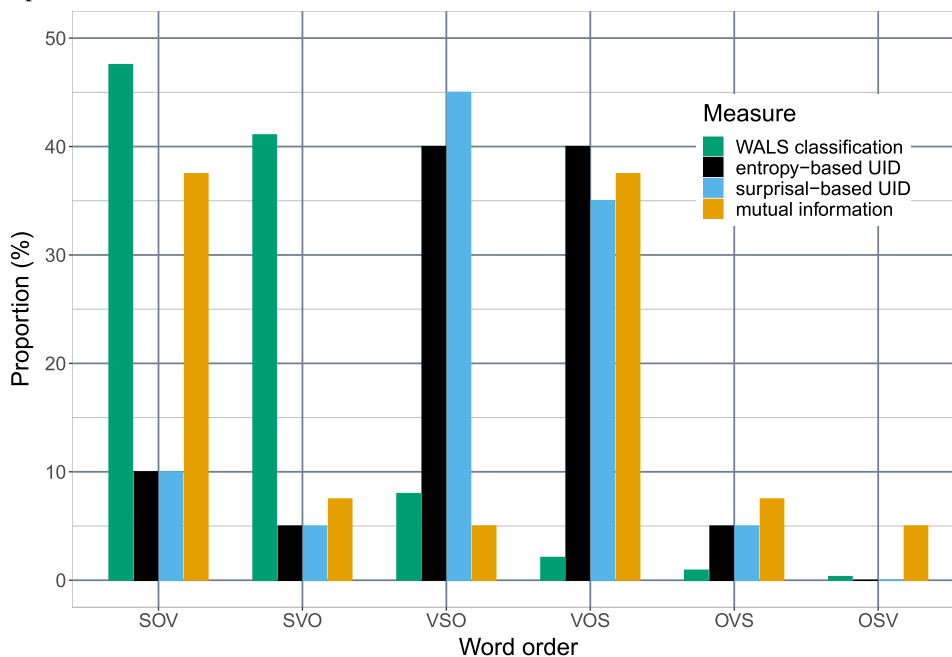


Figure 4. The proportion of word orders predicted across the 20 languages of our UD subset, by the entropy-based UID measure (black), the surprisal-based UID measure (blue), and the pointwise mutual information measure (orange), compared to the observed proportion of word orders catalogued by WALS (cyan).

4.3 Discussion

The rankings from our UID-based measures, and their overprediction of VSO and VOS orders, would suggest that verb-initial languages should be most frequent, contrary to the actual typological distribution. The rankings of our mutual information measure, and its overprediction of VOS word order, would suggest an abundance of languages with dominant SOV and VOS word orders and a dearth of languages with any other order. Although the prediction about SOV does accord with its typological frequency, these results do not accord with the observed frequencies of the other word orders. This is most notable in regards to SVO and VOS orders, which are greatly underpredicted and overpredicted, respectively. A limitation of the mutual information measure we used though, is that mirror word orders are indistinguishable. There may be other ways to define such a measure that

would distinguish between mirror orders in future work. That said, the underprediction of SVO orders presents a serious challenge to the hypotheses we tested.

Overall, our raw data would suggest tendency for languages to evolve to be verb-initial. Evidence in favor of this is rather weak or nonexistent though (Givn, 1979; Li, 2014). Only one measure ranked SOV highly, which is both the most typologically frequent order and has been argued to be a target order for language evolution. Strong evidence that languages evolve towards SVO order contradict the results from all of our three measures (Givn, 1979; Li, 2014; Maurits and Griffiths, 2014). However, our results do agree with Maurits and Griffiths (2014)'s suggestion that processing difficulty is not the primary motivator of language change.

Given that ANOVA tests indicated no significant difference between the means of the six word or-

ders in the entropy-based and surprisal-based UID measures, our data further suggest that UID either plays no role in language evolution of transitive constructions, or that it exerts equal pressure toward every possible ordering. This conclusion is weakly supported by the ANOVA and t-tests conducted on the PMI results. These PMI results would suggest a strong pressure for languages to evolve towards SOV or VOS order. While we can't rule out the possibility that other, stronger, forces act to push language evolution in specific directions, there does not appear to be evidence for languages evolving in such ways.

4.4 Limitations

As mentioned previously, these event data were very sparse, even after removing utterances where any word occurred just once in the corpus. Because of the semantic of some verbs, we would expect a non-zero percentage of utterances where the first two words deterministically predict the third. Nevertheless, the levels obtained in some languages though were unacceptably high. This disproportionately affected SOV and OSV utterances more (80.5% of utterances) compared to VOS and OVS (70.1% of utterances), and compared to VSO and SVO (67.4% of utterances). Both the entropy-based and surprisal-based UID deviation scores are likely to be higher for these utterances, potentially depressing the rankings of SOV, OSV, VOS, and OVS compared to VSO and SVO. Conversely, the PMI values for these utterances may be inflated. This is a considerable limitation, and as such, our conclusions should be considered tentative. This problem even affected the largest corpora we used, but could be addressed in future work using a smoothing algorithm.

Another criticism of this work could be that the use of events and event probabilities is an inadequate model of human language. Although we concede that this model fails to capture the infinitely productive nature of language, it is still a suitable, if imperfect approximate model of speech. The benefit of using a corpus to generate model events is the potential to capture naturalistic speech. There is also evidence that people draw on their knowledge of events when comprehending sentences, suggesting that information about events is an important aspect of language use (Nieuwland, 2015; Tieu et al., 2019).

5 General discussion

With regard to the original MNP paper, we identify three main claims: (1) that object-initial languages have the least uniform information structure; (2) that SVO and VSO have the most uniform information structures; and (3) that UID influenced the distribution of word orders beyond what would be expected by chance. The raw orders from our MNP replication attempts and the entropy-based and surprisal-based UID rankings from our subset of UD languages supports the first point, in that OVS and OSV were consistently placed as the last orders in all rankings. However, the statistical tests we conducted on our UD-derived entropy-based and surprisal-based UID scores suggest there was no difference between the deviation scores of all six word orders. Our pointwise mutual information results are similarly murky on this question, with both our rankings and statistical tests indicating object-initial word orders performing worse than object-medial orders, but with no distinction possible between SVO and OVS or between VSO and OSV.

On the second claim of MNP, our raw UD derived entropy-based and surprisal-based UID orderings support the claim that VSO order results in the most uniform distribution of information, while again our statistical tests fail to support this. SVO, on the other hand, consistently performed worse in our raw rankings than in MNP's, suggesting that SVO does not have one of the most uniform information structures – and again, our statistical tests suggesting that SVO is no more or less uniform in its information structure than any other order. Meanwhile, our PMI results contradict MNP's claim with regard to both VSO and SVO. For both of these first claims, the discrepancies between what the raw rankings suggest and what the statistical tests indicate could be due to not testing enough languages, or due to the other limitations outlined in section 4.4.

On claim three of MNP, the data presented here suggest that uniform processing difficulty may not have played a large role in the evolution of transitive constructions. This follows other lines of evidence that suggest word order evolution is lineage specific (Dunn et al., 2011; Maurits and Griffiths, 2014). Evidence also points to language evolution being subject to a number of complex, interacting factors, including the number of speakers (Lupyan and Dale, 2010; Raviv et al., 2019) and geography

(Norris, 1998; Lupyan and Dale, 2010). Noisy-channel surprisal models further strongly predict SVO word order, as it reliably encodes which noun is the subject and which is the object (Gibson et al., 2019). This suggests that the pressure for a language to transition to SVO order is based in faithful transmission of an utterance rather than in easing processing costs. This pressure for faithful communication may itself be influenced by the factors mentioned earlier.

The contribution of language contact cannot be ignored either, and researchers have long been aware of areal distributions in typological patterns. A growing idea is that bilinguals are the cause of these areal effects, by inducing language change in one or both of their language communities (see e.g. Fernández et al., 2017; Kootstra and ahin, 2018). Examining the language processing and mental representations of bilinguals will be a fruitful area of research to shed more light on this issue. Indeed, some studies are already showing that an L2 can exert influence on a bilingual's L1 (Carando, 2015; Higby, 2016; de Oliveira et al., 2017).

Finally, it is pertinent we point out that many languages with typologically infrequent word orders are spoken by minoritized peoples, oftentimes ethnically or racially. Arguments that these languages are difficult to process can reinforce assimilationist narratives that such languages are inferior and their speakers should speak supposedly superior dominant languages instead. Given that many settler-colonial powers have instituted programs to this effect, and continue to, with justifications based on the supposed deficiencies or inferiorities of minoritized languages, this warning is not hyperbole (DeGraff, 2005; Baloy, 2011; Zentella, 2014; Campbell, 2016; Rosa and Flores, 2017; Flores, 2020). Additionally, such explanations can white-wash the potential role of imperialism and colonialism in causing a dearth of non-SVO/SOV languages. There is value in examining the communicative and cognitive pressures that shape human language, but we must be diligent in how we frame our research questions and findings so as not to uphold oppressive language ideologies.

References

- Natalie J. K. Baloy. 2011. *We Can't Feel Our Language: Making Places in the City for Aboriginal Language Revitalization*. *American Indian Quarterly*, 35(4):515.
- Roger Brown. 1973. *A first language: The early stages*. Harvard U. Press, Cambridge, MA.
- Rebecca Ann Campbell. 2016. *Reification, Resistance, and Transformation? The Impact of Migration and Demographics on Linguistic, Racial, and Ethnic Identity and Equity in Educational Systems: An Applied Approach*. Ph.D. thesis, University of South Florida.
- Agustina Carando. 2015. *The emergence of L1 innovations in Spanish-English bilinguals: Evidence from cross-linguistic structural priming*. *CUNY Academic Works*.
- Arunima Choudhury. 2015. *Interaction between prosody and information structure: experimental evidence from Hindi and Bangla*. Ph.D. thesis, University of Southern California.
- Michel DeGraff. 2005. *Linguists' most dangerous myth: The fallacy of Creole Exceptionalism*. *Language in Society*, 34(04).
- Dryer, Matthew S. 2013. *Order of subject, object and verb*. In Matthew S. Dryer and Martin Haspelmath, editors, *The world atlas of language structures online*. Max-Planck-Institut für evolutionäre Anthropologie., Leipzig.
- Michael Dunn, Simon J. Greenhill, Stephen C. Levinson, and Russell D. Gray. 2011. *Evolved structure of language shows lineage-specific trends in word-order universals*. *Nature*, 473(7345):79–82.
- Maryia Fedzechkina, Becky Chu, and T. Florian Jaeger. 2018. *Human Information Processing Shapes Language Change*. *Psychological Science*, 29(1):72–82.
- Eva M. Fernández, Ricardo A. De Souza, and Agustina Carando. 2017. *Bilingual innovations: Experimental evidence offers clues regarding the psycholinguistics of language change*. *Bilingualism: Language and Cognition*, 20(2):251–268. Publisher: Cambridge University Press.
- Nelson Flores. 2020. *From academic language to language architecture: Challenging raciolinguistic ideologies in research and practice*. *Theory Into Practice*, 59(1):22–31.
- Austin F Frank 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, page 7.
- Richard Futrell. 2019. *Information-theoretic locality properties of natural language*. In *Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 2019)*, pages 2–15, Paris, France. Association for Computational Linguistics.
- Richard Futrell, Peng Qian, Edward Gibson, Evelina Fedorenko, and Idan Blank. 2019. *Syntactic dependencies correspond to word pairs with high*

- mutual information. In *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*, pages 3–13, Paris, France. Association for Computational Linguistics.
- Edward Gibson, Richard Futrell, Steven P. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. [How Efficiency Shapes Human Language](#). *Trends in Cognitive Sciences*, 23(5):389–407.
- Talmy Givn. 1979. *On understanding grammar*. Academic Press, London.
- Joseph H Greenberg. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In *Universals of Language*, pages 73–113. MIT Press.
- John Hale. 2001. [A probabilistic Earley parser as a psycholinguistic model](#). In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–8. Association for Computational Linguistics.
- John A. Hawkins. 2004. *Efficiency and complexity in grammars*. Oxford University Press on Demand.
- John A. Hawkins. 2014. *Cross-linguistic variation and efficiency*. OUP Oxford.
- Eve Higby. 2016. [Native Language Adaptation to Novel Verb Argument Structures by Spanish-English Bilinguals: An Electrophysiological Investigation](#). *CUNY Academic Works*.
- Gerrit Jan Kootstra and Hlya ahin. 2018. [Crosslinguistic structural priming as a mechanism of contact-induced language change: Evidence from Papiament-Dutch bilinguals in Aruba and the Netherlands](#). *Language*, 94(4):902–930.
- Roger Levy. 2011. [Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results](#). In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 1055–1065. Association for Computational Linguistics.
- Roger Levy and T Florian Jaeger. 2007. [Speakers optimize information density through syntactic reduction](#). In *Advances in Neural Information Processing Systems*.
- Charles N Li. 2014. *Mechanisms of syntactic change*, 3rd edition. University of Texas Press.
- Gary Lupyan and Rick Dale. 2010. [Language Structure Is Partly Determined by Social Structure](#). *PLoS ONE*, 5(1):e8559.
- Brian MacWhinney. 2000. [The CHILDES Project: Tools for Analyzing Talk \(third edition\): Volume I: Transcription format and programs, Volume II: The database](#). *Computational Linguistics*, 26(4):657–657.
- Luke Maurits and Thomas L. Griffiths. 2014. [Tracing the roots of syntax with Bayesian phylogenetics](#). *Proceedings of the National Academy of Sciences*, 111(37):13576–13581.
- Luke Maurits, Amy Perfors, and Danielle Navarro. 2010. [Why are some word orders more common than others? A uniform information density account](#). In *Advances in Neural Information Processing Systems*, pages 1585–1593.
- Mante S. Nieuwland. 2015. [The Truth Before and After: Brain Potentials Reveal Automatic Activation of Event Knowledge during Sentence Comprehension](#). *Journal of Cognitive Neuroscience*, 27(11):2215–2228.
- Mary Jane Norris. 1998. [Canada’s aboriginal languages](#). *Canadian social trends*, 51(36):19–27.
- Cndido Samuel Fonseca de Oliveira, Ricardo Augusto de Souza, and Fernando Luiz Pereira de Oliveira. 2017. [Bilingualism effects on L1 representation and processing of argument structure](#). *Journal of the European Second Language Association*, 1(1):23–37.
- Limor Raviv, Antje Meyer, and Shiri Lev-Ari. 2019. [Larger communities create more systematic languages](#). *Proceedings of the Royal Society B: Biological Sciences*, 286(1907):20191262.
- Luigi Rizzi. 1997. [The fine structure of the left periphery](#). In *Elements of Grammar*, pages 281–337. Springer.
- Jonathan Rosa and Nelson Flores. 2017. [Unsettling race and language: Toward a raciolinguistic perspective](#). *Language in Society*, 46(5):621–647.
- Lyn Tieu, Philippe Schlenker, and Emmanuel Chemla. 2019. [Linguistic inferences without words](#). *Proceedings of the National Academy of Sciences*, 116(20):9796–9801.
- Russell S. Tomlin. 1986. [Basic word order – functional principles](#). London: Croom Helm.
- Daniel Zeman, Joakim Nivre, Mitchell Abrams, Noëmi Aepli, Željko Agić, Lars Ahrenberg, Gabrielè Aleksandravičiūtė, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, Colin Batchelor, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A.

Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaz Erjavec, Aline Etienne, Wograiné Evelyn, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Johannes Heinecke, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Takumi Ikeda, Radu Ion, Elena Irimia, Olájdé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Markus Juutinen, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Arne Köhn, Kamil Kopacewicz, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phng Lê H`ông, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Maria Livovina, Yuan Li, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Mackentanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Tomohiko Morioka, Shinsuke Mori, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horňiáček, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lng Nguy`ên Thị, Huy`ên Nguy`ên Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayo Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvreliid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng,

Cenel-Augusto Perez, Guy Perrier, Daria Petrova, Slav Petrov, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Ivan Riabov, Michael Riebler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djámé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh Shohibussirri, Dmitry Sichinava, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Takaaki Tanaka, Isabelle Tellier, Guillaume Thomas, Lisi Torga, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uriá, Hans Uszkoreit, Andrius Utká, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilan Wendt, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Manying Zhang, and Hanzhi Zhu. 2019. *Universal dependencies 2.5*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Ana Celia Zentella. 2014. *TWB (Talking while Bilingual): Linguistic profiling of Latina/os, and other linguistic torquemadas*. *Latino Studies*, 12(4):620–635.