Is Structure Dependence Shaped for Efficient Communication?: A Case Study on Coordination

Kohei Kajikawa[1,2], Yusuke Kubota[2], Yohei Oseki[1]
[1]The University of Tokyo, [2]NINJAL
{kohei-kajikawa, oseki}@g.ecc.u-tokyo.ac.jp
kubota@ninjal.ac.jp

Abstract

Natural language exhibits various universal properties. But why do these universals exist? One explanation is that they arise from functional pressures to achieve *efficient communication*, a view which attributes cross-linguistic properties to domain-general cognitive abilities. This hypothesis has successfully addressed some syntactic universal properties such as compositionality and Greenbergian word order universals. However, more abstract syntactic universals have not been explored from the perspective of efficient communication. Among such universals, the most notable one is *structure dependence*, that is, grammar-internal operations crucially depend on hierarchical representations. This property has traditionally been taken to be central to natural language and to involve domain-specific knowledge irreducible to communicative efficiency.

In this paper, we challenge the conventional view by investigating whether structure dependence realizes efficient communication, focusing on coordinate structures. We design three types of artificial languages: (i) one with a structure-dependent reduction operation, which is similar to natural language, (ii) one without any reduction operations, and (iii) one with a linear (rather than structure-dependent) reduction operation. We quantify the communicative efficiency of these languages. The results demonstrate that the language with the structure-dependent reduction operation is significantly more communicatively efficient than the counterfactual languages. This suggests that the existence of structure-dependent properties can be explained from the perspective of efficient communication.

1 Introduction

To understand the universals of natural language, it is crucial to address *why* such universals exist, as well as *how* such universals can be theoretically described. This raises the question: what kinds of pressures shape these universals?

One explanation is that the universals of natural language are shaped as a result of functional pressures to achieve *efficient communication* (Zipf, 1949; Jaeger and Tily, 2011; Christiansen and Chater, 2016; Kemp et al., 2018; Gibson et al., 2019; Futrell and Hahn, 2022; Fedorenko et al., 2024). Efficient communication refers to a situation where the amount of information conveyed is maximized while the effort required for production and comprehension is minimized under human cognitive constraints. If some structural property of languages is shaped to achieve efficient communication, it can be optimized under two competing functional pressures: the need to be as *simple* as possible and the need to be as *informative* as possible.

The hypothesis that two competing pressures for utilities shape the form of human language has long been assumed in linguistics (Hawkins, 1994, 2004; Haspelmath, 2008). In recent years, methodologies have been established to examine this hypothesis by quantifying the *simplicity* and *informativeness* of languages by using information-theoretic criteria (Kemp et al., 2018; Gibson et al., 2019; Futrell and Hahn, 2022). To date, this hypothesis has been successfully examined at the lexical level (Ferrer i Cancho and Solé, 2003; Kemp and Regier, 2012; Piantadosi et al., 2011, 2012; Regier et al., 2015; Zaslavsky et al., 2018; Mollica et al., 2021; Steinert-Threlkeld, 2021; Denić et al., 2022; Trott and Bergen, 2022; Uegaki, 2022; Chen et al., 2023; Pimentel et al., 2023; van de Pol et al., 2023; Denić and Szymanik, 2024, *inter alia*). At the syntactic level, there are pieces of empirical evidence that grammar itself is shaped to achieve efficient communication (Gildea and Jaeger, 2015; Futrell et al., 2020a,b; Hahn et al., 2020, 2021; Clark et al., 2023), and it has been shown that the existence of syntactic universals such as compositionality and

Greenbergian word order universals (Greenberg, 1963) can be explained by this hypothesis (Kirby et al., 2015; Hahn et al., 2020).

But we do not yet know whether this type of competition-based account can be extended to more abstract types of linguistic knowledge that go beyond mere sensitivity to structure such as compositionally and word order. A representative type of such abstract knowledge comes from cases of what one might call *structure dependence*, by which we broadly refer to operations that directly manipulate structural representations at some level of linguistic representation. Structure dependence has traditionally been taken to be a characteristic and central property of human language (Chomsky, 1957, 1965; Everaert et al., 2015); in fact, a key underlying theme throughout the whole history of mainstream generative grammar is extreme skepticism of the idea that such properties can be reduced to communicative principles. The dogma in this line of thought has it that abstract syntactic properties of language are thought to be governed by domain-specific efficient computation necessary for deriving the structure of language (Hauser et al., 2002; Chomsky, 2005; Berwick and Chomsky, 2016), while communication is viewed as *not* essential to the core linguistic competence (Chomsky, 2002; Hauser et al., 2002). It is thus crucial to investigate whether even such syntactic properties can be accounted for from the perspective of domain-general efficient communication.

In this paper, we directly address this issue by examining structure dependence. Specifically, we investigate whether structure dependence realizes efficient communication by focusing on coordinate structures. We design three types of languages: (i) one with a structure-dependent reduction operation, which has coordinate structures similar to those in natural language, (ii) one without any reduction operations, and (iii) one with a linear (rather than structure-dependent) reduction operation. The latter two are conceptually possible but counterfactual languages. We adopted White and Cotterell's (2021) artificial probabilistic context-free grammars (PCFGs) to create the three languages. Then we quantify the *simplicity* and *informativeness* of these languages and compare their communicative efficiency. The results demonstrate that the languages with a structure-dependent reduction operation are significantly more communicatively efficient than their counterfactual counterparts. This suggests that the structure-dependent properties in

human language can be explained in terms of efficient communication.

2 Background

2.1 Efficient communication hypothesis

In recent years, many researchers in cognitive science and computational psycholinguistics have increasingly focused on attributing cross-linguistic properties to domain-general cognitive functions. The central thesis of this strand of research is that natural language is shaped to achieve efficient communication (Zipf, 1949; Jaeger and Tily, 2011; Christiansen and Chater, 2016; Kemp et al., 2018; Gibson et al., 2019; Futrell and Hahn, 2022; Fedorenko et al., 2024). Communicatively efficient structures are more likely to be learned because they may be used more frequently and are easier to process during learning. This can drive changes in the language that further enhance communicative efficiency (Jaeger and Tily, 2011; Fedzechkina et al., 2012). Alternatively, the intergenerational transmission bottleneck in cultural evolution might lead to the selection of communicatively efficient languages (Christiansen and Kirby, 2003; Kirby et al., 2015). In either case, if functional pressures for efficient communication are at work, languages are expected to be optimized for efficient communication.

To test this hypothesis, one approach is to quantify and compare the communicative efficiency of real languages with logically possible but unattested counterfactual languages. For example, Hahn et al. (2020) showed that real languages reach an optimal word order under the trade-off between simplicity and informativeness. Simplicity refers to how simple the sentences of a language are as strings, while informativeness indicates how accurately the meaning can be reconstructed from the sentences of that language. The communicative efficiency of a language is then defined as the weighted sum of simplicity and informativeness. They created counterfactual languages for each of the 51 natural languages by changing word order patterns while maintaining the projectivity of dependency structures and calculated the communicative efficiency of all the languages. They found that almost all of the real languages had significantly higher communicative efficiency than their counterparts. This indicates that natural language is shaped by the pressure to enhance communicative efficiency.

2.2 Structure dependence

It has long been argued that natural language syntax exhibits structure dependence, the sensitivity to a hierarchical syntactic structure rather than a linear sequence of words (Chomsky, 1957, 1965; Everaert et al., 2015). Grammatical operations are thus applied based on syntactic structures, not on linear strings. For instance, yes-no questions in English are a well-known syntactic phenomenon that requires a structure-dependent grammatical rule. In English yes-no questions, the auxiliary of the main clause moves to the front of the sentence. If we were to formulate the rule for forming yes-no questions as *move the leftmost auxiliary verb to the front*, which is not structure-dependent, we would incorrectly transform a sentence *The man who is running is happy* into *Is the man who running is happy?* The correct transformation should move the second *is* from the main clause, resulting in *Is the man who is running happy?*

In the same way, it has traditionally been assumed that coordination, which is the focus of this study, also requires a structure-dependent grammatical operation (Chomsky, 1957, 1975 (=1955); Ross, 1967). For example, the coordinated sentence *John ran and swam* is derived from *John ran and John swam*, and *Mary called and praised John* is derived from *Mary called John and Mary praised John*, through a structure-dependent grammatical operation known as *Conjunction Reduction* (Figure 1). Conjunction Reduction is formulated as follows:

$$(Y + X_1 + Z) + \text{CC} + (Y + X_2 + Z)$$
$$\rightarrow Y + (X_1 + \text{CC} + X_2) + Z,$$

(Chomsky, 1957, p.113, with slight modifications)

where X represents any syntactic category, Y and Z represent any syntactic category or string, CC represents any conjunction, and $+$ denotes concatenation. Conjunction Reduction captures the fact that coordination is possible between identical syntactic categories for any syntactic category.[1]

---

[1]In the subsequent linguistic literature, Conjunction Reduction has been replaced by a more sophisticated approach known as *Generalized Conjunction* (Gazdar, 1980; Partee and Rooth, 1983), which overcomes some important limitations of Conjunction Reduction (e.g., Partee, 1970). However, the fundamental insight of structure dependence in the classical formulation of Conjunction Reduction is fully retained in Generalized Conjunction, since the latter can essentially be viewed as a reformulation of the former at the level of semantic rep-
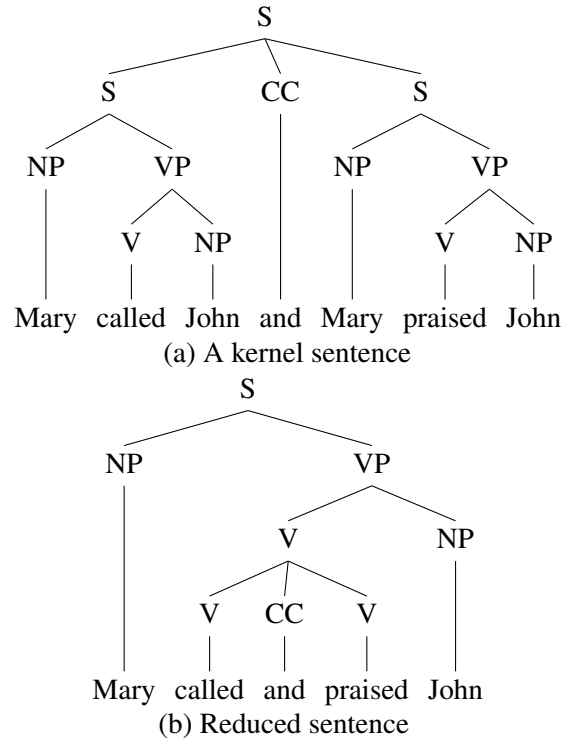

(a) A kernel sentence

(b) Reduced sentence

Figure 1: A coordinate structure (b) is derived by applying Conjunction Reduction, a structure-dependent reduction operation to a sentence-level coordinated kernel sentence (a).

We aim to investigate whether this structure-dependent reduction operation contributes to communicative efficiency in natural language.

3 Experiment

3.1 Data

Design of languages We designed the following three types of languages to investigate the impact on communicative efficiency when a language does not have a structure-dependent reduction operation at all:

1. `No-reduction` language: A language with no reduction. Only sentence-level coordination is possible.

2. `Structure-reduction` language: A language with structure-dependent reduction. Coordination is possible between identical syntactic categories.

---

resentation using lambda calculus and higher-order functions. This study focuses on the operations applied to structures at any level. Therefore, the difference between Conjunction Reduction and Generalized Conjunction is orthogonal to the following discussion.

CFG rules
S $\rightarrow$ NP$_{\text{Subj}}$ VP
VP $\rightarrow$ IVerb | TVerb NP$_{\text{Obj}}$ | Verb$_{\text{Comp}}$ S$_{\text{Comp}}$
S$_{\text{Comp}}$ $\rightarrow$ Comp S
NP $\rightarrow$ Adj NP | NP PP | NP Rel VP
NP$_{\text{Subj}}$ $\rightarrow$ Noun Case$_{\text{Subj}}$ | Pronoun$_{\text{Subj}}$
NP$_{\text{Obj}}$ $\rightarrow$ Noun Case$_{\text{Obj}}$ | Pronoun$_{\text{Obj}}$
PP $\rightarrow$ Prep NP
X $\rightarrow$ X CC X, where X ={NP, Adj, IVerb, TVerb}

Table 1: Overview of the grammatical rules equipped in White and Cotterell's (2021) PCFG. For simplicity, features such as tense and number are omitted.

3. Linear-reduction language: A language with linear (rather than structure-dependent) reduction where repeated expressions in the same sentence are deleted in a coordinate structure.

Data generation For each language, the sentences to be evaluated were created using a set of PCFGs defined by White and Cotterell (2021). The PCFGs are equipped with six switches to reverse the linear order of specific heads and dependents, which results in a total of $2^6 = 64$ word order patterns in the artificial languages.

The PCFG includes the following basic syntactic categories: verb, noun, pronoun, adjective, conjunction, preposition, particle, sentential complementizer, and relativizer. Some features such as tense (present and past), number (singular and plural), and grammatical relation (subject and object) are assigned to categories. The grammatical rules are defined by the categories, as shown in Table 1. The combination of categories and features results in a total of 44 syntactic categories and a lexicon consists of 1,254 words. Although this grammar is much simpler than that of real natural languages, it is sufficiently sophisticated for our purpose of comparing structurally different languages. Moreover, it allows us to simultaneously take into consideration typologically diverse word order patterns.

We used this PCFG to create corpora of the artificial languages with 64 different word orders. We then constructed the no-reduction, structure-reduction, and linear-reduction languages defined above for each of these word orders. The structure-reduction language is the direct output of the PCFG. We then ex-

(a) structure-reduction language

(b) no-reduction language
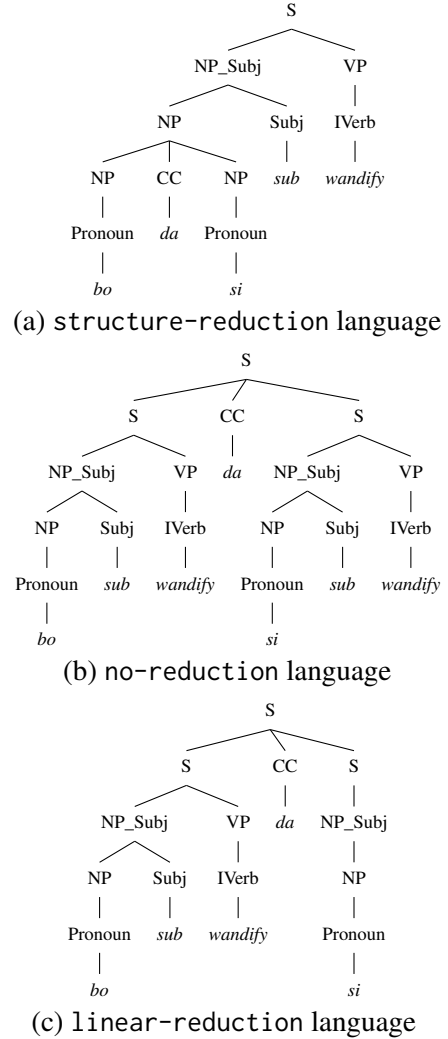

(c) linear-reduction language

Figure 2: Examples of the three languages expressing the same meaning. The word order is set with all six switches being strictly head-final as in Japanese. For simplicity, information on number and tense has been omitted from the syntactic categories in these figures.

panded all of the coordinate structures in the structure-reduction language to a sentence level to create the no-reduction language. Furthermore, we applied a linear reduction to the no-reduction language by deleting all repeated words in the same coordinate structure to create the linear-reduction language. The examples of the tree structures of the three types of artificial languages are shown in Figure 2.

We measure the communicative efficiencies for these 64 word orders $\times$ 3 types = 192 kinds of languages.

3.2 Estimating communicative efficiency

Definition of communicative efficiency For the simplicity/informativeness trade-off, follow-

ing Hahn et al. (2020), we evaluate simplicity as a property of the linear sequence, in terms of how easily the next word in an utterance can be predicted, i.e., *predictability*, and informativeness in terms of how well the syntactic structure behind an utterance can be reconstructed, i.e., *parsability*.

Predictability is specifically defined as the negative entropy, $-H(\mathcal{U})$, of all utterances u in a language:

$$-H(\mathcal{U}) = \sum_{u \in \mathcal{U}} p(u) \log p(u). \tag{1}$$

Here, we define the probability of utterance u as the product of the probabilities of the words that constitute the utterance. When the sample size is sufficiently large, entropy can be estimated as the mean word-by-word surprisal. Surprisal is a metric that empirically predicts human behavioral (e.g., Demberg and Keller, 2008; Smith and Levy, 2013; Shain et al., 2024) and neural (e.g., Frank et al., 2015; Lopopolo et al., 2017; Brennan and Hale, 2019; Shain et al., 2020) data. The mean negative word-by-word surprisal represents the ease of incremental sentence processing on average under surprisal theory (Hale, 2001; Levy, 2008).

Parsability is defined as the negative conditinal entropy, $-H(\mathcal{T}|\mathcal{U})$, of the underlying syntactic structure t given an utterance u:[2]

$$-H(\mathcal{T}|\mathcal{U}) = \sum_{t \in \mathcal{T}, u \in \mathcal{U}} p(t, u) \log p(t|u). \tag{2}$$

Since semantic calculation in compositional semantics crucially depends on the building of syntactic structures (Montague, 1970; Heim and Kratzer, 1998), we employ a metric of informativeness that captures how unambiguously the underlying syntactic structure can be reconstructed—both temporally and globally—as an indicator of how accurately the intended meanings of utterances can be recovered.

Then, following Ferrer i Cancho and Solé (2003) and Hahn et al. (2020), we defined a communicative efficiency function as the weighted sum of predictability and parsability:

$$\Omega(\lambda) := \lambda \text{predictability} + (1 - \lambda)\text{parsability} \tag{3}$$

$$= -\lambda H(\mathcal{U}) - (1 - \lambda)H(\mathcal{T}|\mathcal{U}), \tag{4}$$

where λ is a trade-off parameter ranging from 0 to 1, which represents the contribution of each term. The objective function that captures the trade-off between the cost of linguistic expressions and the likelihood of meaning given the expression has been used in previous studies (e.g., Ferrer i Cancho and Solé, 2003; Frank and Goodman, 2012; Kemp and Regier, 2012; Regier et al., 2015; Hahn et al., 2020).

Recurrent Neural Network Grammars To obtain the values of predictability and parsability, we adopted Recurrent Neural Network Grammars (RN-NGs; Dyer et al., 2016). RNNGs are a generative model of sentences that explicitly models hierarchical structures by processing the action sequence of shift-reduce parsing. RNNGs can be used for both language modeling and parsing with the same model parameters, which is suitable for our purpose here.

In this study, we used the left-corner stack-only RNNGs (Kuncoro et al., 2018) implemented with PyTorch[3] by Noji and Oseki (2021). We used a two-layer LSTM, where both the hidden layer and the input layer have 256 dimensions.[4]

Left-corner parsing is considered reasonable for human incremental sentence processing from the perspective of memory capacity (Abney and Johnson, 1991; Resnik, 1992) and is often assumed as a model of human sentence processing (e.g., Lewis and Vasishth, 2005; van Schijndel et al., 2013). Additionally, it has already been pointed out that a simple bottom-up strategy without a predictive process cannot explain the human incremental processing of coordinate structures in English at least when the parser conducts a serial parsing (Sturt and Lombardo, 2005; Stanojević et al., 2023). This motivates our choice of a left-corner as a parsing strategy.

We performed a beam search with a beam size of 100 for inference. We also used word-synchronous beam search (Stern et al., 2017) with a size of 10.

---

[2]Hahn et al. (2020) cenceptually defined parsability as mutual information $I(\mathcal{U}; \mathcal{T}) = H(\mathcal{T}) - H(\mathcal{T}|\mathcal{U})$ between an utterance and its syntactic structure. However, they actually estimated the value of parsability as $-H(\mathcal{T}|\mathcal{U})$ on the assumption that $H(\mathcal{T})$ is constant.

[3]https://github.com/pytorch/pytorch/releases/tag/v1.12.1

[4]Other hyperparameters are as follows: random seeds are {3435, 3436, 3437}, optimizer is Adam (Kingma and Ba, 2015), learning rate is 0.001, dropout is 0.3, and batch size is 128. The code of RNNGs we employed (Noji and Oseki, 2021) is available at https://github.com/aistairc/rnng-pytorch.

Estimation of communicative efficiency Predictability for the languages can be obtained by

$$-H(\mathcal{U}) = \sum_{u \in \mathcal{U}} p(u) \log p_\phi(u), \qquad (5)$$

and following Hahn et al. (2020), we define the log-likelihood of each utterance u as

$$\log p_\phi(u) := \sum_{i=1}^{N} \log p_\phi(w_i|w_{<i}), \qquad (6)$$

where w_i represents the i-th word composing the utterance and ϕ represents the parameters of the RNNGs. We can approximate the negative entropy of u with its Monte Carlo estimate on test data:

$$-H(\mathcal{U}) \approx \frac{1}{|\text{Test Data}|} \sum_{u \in \text{Test Data}} \log p_\phi(u). \quad (7)$$

We calculated the values of predictability according to the formula above.

Parsability can be calculated by

$$-H(\mathcal{T}|\mathcal{U}) = \sum_{t \in \mathcal{T}, u \in \mathcal{U}} p(t, u) \log p_\phi(t|u), \quad (8)$$

and in the same way, the conditional entropy can be approximated with its Monte Carlo estimate on test data:

$$-H(\mathcal{T}|\mathcal{U}) \approx \frac{1}{|\text{Test Data}|} \sum_{t, u \in \text{Test Data}} \log p_\phi(t|u). \qquad (9)$$

Here, we define the log-likelihood of the conditional probability of the tree structure t given each utterance u as

$$\log p_\phi(t|u) := \sum_{i=1}^{N} \log p_\phi(t_{\text{best}}|w_{\leq i}). \qquad (10)$$

Again, w_i and ϕ represent the i-th word of the utterance and the parameters of RNNGs, respectively. t_{best} refers to the most likely constituency parse in the word-synchronous beam at each word.

For 192 types of artificial languages, we generated 20,000 sentences for each and divided them into an 8-1-1 train-dev-test split for training and evaluation. For all languages, we trained RNNGs on word-by-word using Adam (Kingma and Ba, 2015) for 10 epochs each with multiple random seeds. Then, we calculated the values of predictability and parsability, normalized by the number of words, to ensure valid comparisons across languages with inherently different sentence lengths.

4 Results

A distribution of the values of communicative efficiency for the three types of languages, as calculated by RNNGs, is shown in Figure 3. For an interpretation of the trade-off parameter λ, the predictability and parsability values of all languages are z-transformed (i.e., centered and divided by the standard deviation) before being substituted into Eq 4. The lines in the figure show the transitions of the value of communicative efficiency for λ with a 95% confidence interval (CI). By finding the coordinates where the lines intersect, we can observe the behavior of communicative efficiency for each language depending on the value of λ. The lower bound of the 95% CI for `structure-reduction` intersects with the upper bound of the 95% CI for `linear-reduction` at $\lambda = 0.18$. In the same way, the upper bound of the 95% CI for `structure-reduction` intersects with the lower bound of the 95% CI for `linear-reduction` at $\lambda = 0.93$. This indicates that `structure-reduction` languages are the most communicatively efficient, at least within the range of λ values between 0.18 and 0.93.

Additionally, when we examine only predictability or parsability, their distributions are shown in Figure 4 and Figure 5, respectively. The mean values for predictability were ordered as `no-reduction` > `structure-reduction` > `linear-reduction`, while for parsability, the order was `linear-reduction` > `structure-reduction` > `no-reduction`, in which all of the pairs have a statistically significant difference ($p < 0.05/3$ by paired t-test with Bonferroni correction). This indicates that when only predictability or parsability is individually considered, the `structure-reduction` language may not always be the best option. However, to satisfy both criteria simultaneously, i.e., to consider the weighted sum of predictability and parsability under the parameter $\lambda \in [0.18, 0.93]$, the `structure-reduction` language achieves the highest score of communicative efficiency as shown in Figure 3.

To further interpret the results of predictability and parsability, we plotted least squares regression lines between word position in the sentence and each word-by-word value (Figure 6). For predictability, only the `linear-reduction` languages significantly decline towards the latter part of the sentence. As for parsability, although the
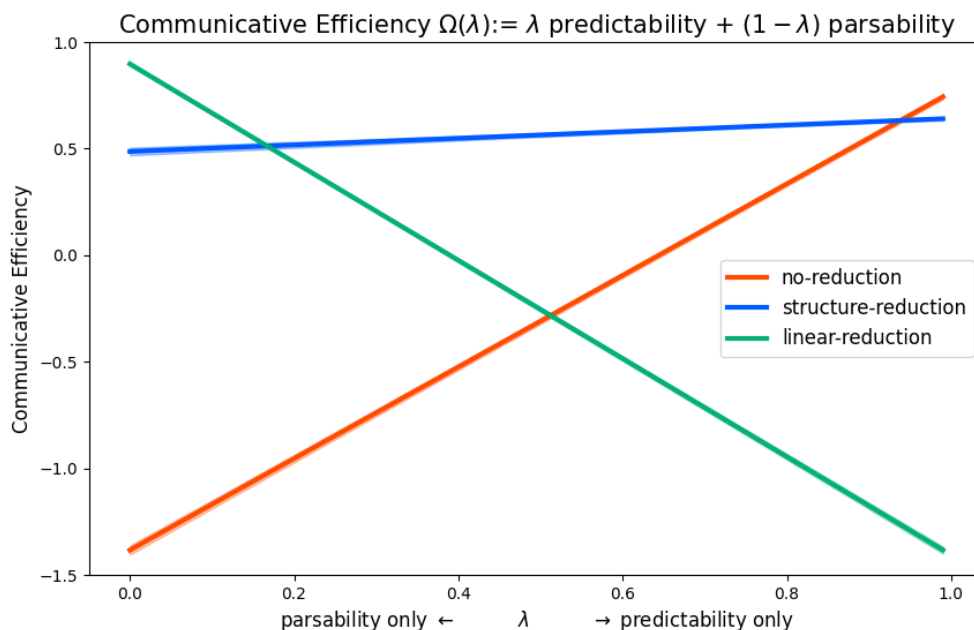
Figure 3: Distribution of **communicative efficiency** for the three types of languages with 95% CI. The *x*-axis and *y*-axis represent the trade-off parameter λ and communicative efficiency, respectively. Both predictability and parsability are *z*-transformed for an interpretation of λ. The `structure-reduction` languages are the most communicatively efficient under the parameter $\lambda \in [0.18, 0.93]$ for 95% CI.
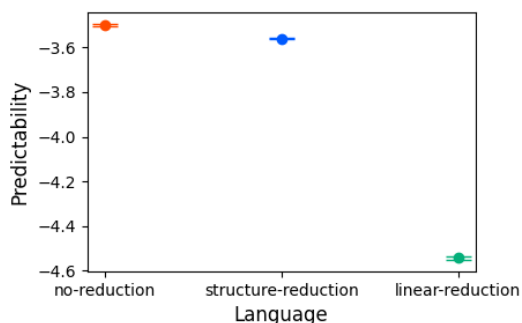

Figure 4: Distribution of **predictability** for the three types of languages. Error bars indicate 95% CI.
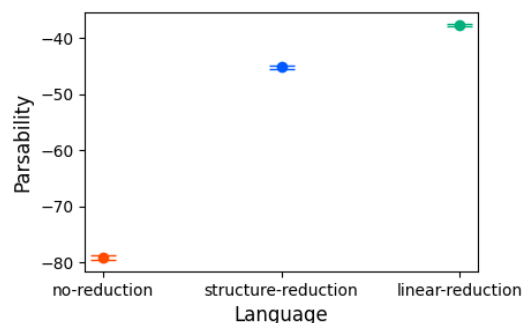

Figure 5: Distribution of **parsability** for the three types of languages. Error bars indicate 95% CI.

`linear-reduction` languages experience a faster decrease, the other two languages, which have longer expression lengths, achieve a lower overall score.

5   Discussion

We demonstrated that the `structure-reduction` languages, which have the same structure-dependent reduction operation as natural language, had significantly higher communicative efficiency compared to the conceptually possible but counterfactual `no-reduction` and `linear-reduction` languages when we calculated the scores by RN-

NGs with a trade-off parameter $\lambda \in [0.18, 0.93]$. This suggests that the structure-dependent reduction operation prevalent in the natural language syntax may exist due to functional pressures to support efficient communication along lines discussed in Section 2.1. It should be noted that, as Figure 3 shows, when λ is extremely small or large, the `no-reduction` or `linear-reduction` languages achieve the highest efficiency scores. However, λ represents the relative contribution of the two terms to the overall efficiency score. It is difficult to assume a reasonable scenario where only one term is emphasized. While we do not aim to estimate a spe-
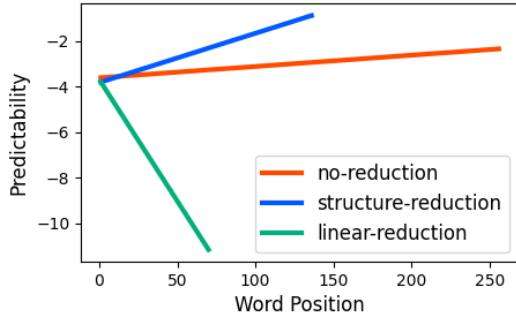
(a) Relationship between **predictability** and word position.
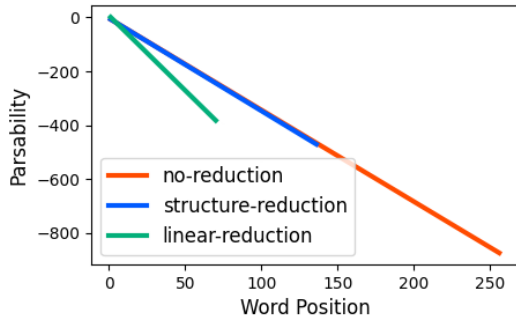
(b) Relationship between **parsability** and word position.

Figure 6: Relationship between predictability/parsability and word position for the three types of languages. Predictability and parsability here refer to the negative surprisal and the negative log-likelihood of the best parse for each word, respectively. The lines represent the fit of a least squares regression model for these data.

cific value of λ, empirically, Ferrer i Cancho and Solé (2003) found in their simulation experiment that Zipf's law emerged when $\lambda \approx 0.41$. In addition, Hahn et al. (2020) demonstrated that grammars optimized at $\lambda \approx 0.47$ captured all 8 of the Greenberg correlations they investigated, whereas optimizing solely for predictability or parsability did not account for all of them.[5]

When we consider only one of the terms constituting communicative efficiency, that is, predictability (simplicity) or parsability (informativeness), one of the two counterfactual languages achieves the highest score. The `no-reduction` language is the easiest in terms of prediction, i.e., the estimation of the next word, among the three types. This is because the language has no reduction, and all sen-

tences are fully represented, so the number of local patterns for the next word is limited, which makes prediction easier. For example, in a `no-reduction` language where the part of speech at the beginning of a sentence is always X in its word order pattern, the part of speech following a conjunction must be X, while `structure-reduction` and `linear-reduction` languages have more variations, which lead to higher entropy of strings. However, the `no-reduction` language is not well-suited for parsing. As sentence length increases, the number of potential parses grows exponentially (Church and Patil, 1982). Since this language lacks reduction and results in longer overall expressions, the possible parses at each word position rapidly increase, as shown in Figure 6(b). Consequently, it is not an optimal design from the perspective of estimating the underlying structure of an utterance. In contrast, the `linear-reduction` language has shorter overall expressions, resulting in fewer possible parses at each word position.[6] As a result, it is superior in terms of estimating tree structures. However, the reduction is too radical, making it challenging to maintain predictability. As shown in Figure 6(a), this issue is evident in the latter parts of sentences in this language, where predictions become increasingly difficult due to the need to consider the possibility that previously mentioned words might have been reduced. In short, a reduction operation is necessary to enhance parsability, but it should be applied *restrictively* so as not to sacrifice next-word predictability. Balancing the trade-off between the two, a structure-dependent reduction is the most preferred design for maximizing communicative efficiency.

It should be noted that even in natural language, there are instances of linear reduction operations such as *stripping* (Hankamer and Sag, 1976), for instance shown below, or sentences that retain sentence-level coordination without reduction for pragmatic purposes such as emphasis.

(11)   Mary took a walk in the park, and Bill too.

However, these phenomena occur as alternative choices in a language that has a structure-dependent

---

[5]Hahn et al. defined a communicative efficiency function as $\Omega(\lambda) := \lambda \text{predictability} + \text{parsability}$ and set $\lambda = 0.9$. In other words, while we assigned weights of λ and $1 - \lambda$ to predictability and parsability, respectively, they assigned weights of 0.9 and 1. Solving the equation $\lambda/(1-\lambda) = 0.9/1$ gives $\lambda \approx 0.47$.

[6]The parsability metric used here may overestimate the informativeness of `linear-reduction` languages. This metric does not account for whether the estimated tree structure is correct, nor does it fully capture the inherent ambiguity in the language. A more accurate approach could involve quantifying informativeness using mutual information (e.g., Ferrer i Cancho, 2005; Futrell, 2017; Zaslavsky et al., 2018; Hahn et al., 2020), though this would not change our conclusion.

reduction operation. Our claim is that a language lacking a structure-dependent reduction operation entirely is not preferable from the perspective of efficient communication.

The results of this study have interesting implications for theoretical linguistics research. Structure dependence, a syntactic universal property we addressed here, has traditionally been argued to be one of the characteristic features of human language (Chomsky, 1957, 1965; Everaert et al., 2015). A prominent view in the mainstream generative grammar argues that natural language involves domain-specific predispositions and that syntactic properties of language—including structure dependence—are best explained from the perspective of 'efficient *computation*' reflecting such predispositions genetically hard-wired in the human brain (Hauser et al., 2002; Chomsky, 2005; Everaert et al., 2015; Berwick and Chomsky, 2016). Under this view, communication is taken to be a kind of epiphenomenon, not essential to the core linguistic competence (Chomsky, 2002; Hauser et al., 2002). However, our results suggest that at least some structure-dependent properties present in natural language (such as coordination) can be explained from the perspective of efficient *communication*. This does not immediately refute the dominant research program attempting to explain linguistic properties from a 'computational' perspective, but it does indicate that abstract properties in syntax may not necessarily need to be explained solely from that perspective. This aligns with the existing body of research that attempts to explain various aspects of natural language from the perspective of efficient communication (Gibson et al., 2019; Fedorenko et al., 2024).

6   Conclusion

In this paper, we investigated whether structure dependence, one of the syntactic universals, reflects the optimization for efficient communication. To address this issue, we focused on coordinate structures and designed three types of artificial languages: (i) one with a structure-dependent reduction operation, (ii) one without any reduction operations, and (iii) one with a linear (rather than structure-dependent) reduction operation. We quantified the communicative efficiency of these languages and compared them. The results demonstrated that the languages with a structure-dependent reduction operation were significantly more communicatively efficient than their counterfactual counterparts. This suggests that the structure-dependent properties of natural language can be explained from the functional perspective of efficient communication.[7]

Limitations

There is room for improvement in the objective function for communicative efficiency. Although we used the mean word-by-word surprisal, conditioned on all preceding words, as a measure of predictability, human language processing is subject to short-term memory constraints (Gibson, 1998; Lewis and Vasishth, 2005; Isono, 2024). Thus, it is preferable to model predictability in a way that incorporates *lossy memory representation* (Futrell et al., 2020a; Hahn et al., 2021, 2022). Moreover, the psychological plausibility of the parsability metric should be critically evaluated, both conceptually and empirically. Since parsability relies on an intermediate representation—syntactic structures—it does not fully capture the direct relationship between linguistic expressions and their meanings, suggesting that there is room for further conceptual refinement.

To the best of our knowledge, this study is the first to investigate communicative efficiency with respect to structure dependence. We focused on coordinate structures with the artificial language paradigm as a starting point. Of course, a deeper understanding of structure dependence in language will require using natural language data and extending the analysis to phenomena such as agreement and movement that are argued to be relevant to structure dependence. In future work, we plan to test the relationship between structure dependence and communicative efficiency by applying the methodology proposed here to a broader range of syntactic constructions, using treebanks like Universal Dependencies (Nivre et al., 2020).

Ethical considerations

We used all tools and datasets following their respective terms and licenses. We employed ChatGPT and Grammarly for writing assistance and utilized ChatGPT for writing experimental code. We used these tools in compliance with the ACL 2023 Policy on AI Writing Assistance.

---

[7]Code for reproducing our experiments is available at https://github.com/kohei-kaji/coordination.

Acknowledgments

References

Steven P. Abney and Mark Johnson. 1991. Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, 20(3):233–250.

Robert C. Berwick and Noam Chomsky. 2016. *Why Only Us: Language and Evolution*. The MIT Press, Cambridge, MA.

Jonathan R. Brennan and John T. Hale. 2019. Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLoS ONE*, 14(1):e0207741.

Sihan Chen, Richard Futrell, and Kyle Mahowald. 2023. An information-theoretic approach to the typology of spatial demonstratives. *Cognition*, 240:105505.

Noam Chomsky. 1957. *Syntactic Structures*. Mouton.

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. MIT Press.

Noam Chomsky. 1975 (=1955). *The Logical Structure of Linguistic Theory*. Springer New York, NY.

Noam Chomsky. 2002. An interview on minimalism. In Adriana Belletti and Luigi Rizzi, editors, *On Nature and Language*, pages 92–161. Cambridge University Press.

Noam Chomsky. 2005. Three Factors in Language Design. *Linguistic Inquiry*, 36(1):1–22.

Morten H. Christiansen and Nick Chater. 2016. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39:e62.

Morten H. Christiansen and Simon Kirby. 2003. Language evolution: consensus and controversies. *Trends in Cognitive Sciences*, 7(7):300–307.

Kenneth Church and Ramesh Patil. 1982. Coping with syntactic ambiguity or how to put the block in the box on the table. *American Journal of Computational Linguistics*, 8(3-4):139–149.

Thomas Hikaru Clark, Clara Meister, Tiago Pimentel, Michael Hahn, Ryan Cotterell, Richard Futrell, and Roger Levy. 2023. A cross-linguistic pressure for Uniform Information Density in word order. *Transactions of the Association for Computational Linguistics*, 11:1048–1065.

Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.

Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. 2022. Indefinite pronouns optimize the simplicity/informativeness trade-off. *Cognitive Science*, 46(5):e13142.

Milica Denić and Jakub Szymanik. 2024. Recursive numeral systems optimize the trade-off between lexicon size and average morphosyntactic complexity. *Cognitive Science*, 48(3):e13424.

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. 2016. Recurrent neural network grammars. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 199–209, San Diego, California. Association for Computational Linguistics.

Martin B.H. Everaert, Marinus A.C. Huybregts, Noam Chomsky, Robert C. Berwick, and Johan J. Bolhuis. 2015. Structures, not strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Sciences*, 19:729–743.

Evelina Fedorenko, Steven T. Piantadosi, and Edward A. F. Gibson. 2024. Language is primarily a tool for communication rather than thought. *Nature*, 630:575–586.

Maryia Fedzechkina, T. Florian Jaeger, and Elissa L. Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44):17897–17902.

Ramon Ferrer i Cancho. 2005. Zipf's law from a communicative phase transition. *The European Physical Journal B - Condensed Matter and Complex Systems*, 47:449–457.

Ramon Ferrer i Cancho and Ricard V. Solé. 2003. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3):788–791.

Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Stefan L. Frank, Leun J. Otten, Giulia Galli, and Gabriella Vigliocco. 2015. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140:1–11.

Richard Futrell. 2017. *Memory and Locality in Natural Language*. Ph.D. thesis, MIT.

Richard Futrell, Edward Gibson, and Roger P Levy. 2020a. Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, 44(3):e12814.

Richard Futrell and Michael Hahn. 2022. Information theory as a bridge between language function and language form. *Frontiers in Communication*, 7.

Richard Futrell, Roger P. Levy, and Edward Gibson. 2020b. Dependency locality as an explanatory principle for word order. *Language*, 96(2):371–412.

Gerald Gazdar. 1980. A cross-categorial semantics for coordination. *Linguistics and Philosophy*, 3(3):407–409.

Edward Gibson. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Edward Gibson, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5):389–407.

Daniel Gildea and T. Florian Jaeger. 2015. Human languages order information efficiently. *Preprint*, arXiv:1510.02823.

Joseph H. Greenberg. 1963. *Universals of language*. MIT press.

Michael Hahn, Judith Degen, and Richard Futrell. 2021. Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. *Psychological Review*, 128:726–756.

Michael Hahn, Richard Futrell, Roger Levy, and Edward Gibson. 2022. A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, 119(43):e2122602119.

Michael Hahn, Dan Jurafsky, and Richard Futrell. 2020. Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, 117(5):2347–2353.

John Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Jorge Hankamer and Ivan A. Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry*, 7:391–428.

Martin Haspelmath. 2008. Parametric versus functional explanations of syntactic universals. *The Limits of Syntactic Variation*, 132:75–107.

Marc D. Hauser, Noam Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579.

John A. Hawkins. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press.

John A. Hawkins. 2004. *Efficiency and Complexity in Grammars*. Oxford University Press.

Irene Heim and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Wiley-Blackwell.

Shinnosuke Isono. 2024. Category locality theory: A unified account of locality effects in sentence comprehension. *Cognition*, 247:105766.

T. Florian Jaeger and Harry Tily. 2011. On language 'utility': processing complexity and communicative efficiency. *WIREs Cognitive Science*, 2(3):323–335.

Charles Kemp and Terry Regier. 2012. Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054.

Charles Kemp, Yang Xu, and Terry Regier. 2018. Semantic typology and efficient communication. *Annual Review of Linguistics*, 4(1):109–128.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the International Conference Learning Representations*, San Diego, CA, USA. Conference Track Proceedings.

Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.

Adhiguna Kuncoro, Chris Dyer, John Hale, Dani Yogatama, Stephen Clark, and Phil Blunsom. 2018. LSTMs can learn syntax-sensitive dependencies well, but modeling structure makes them better. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1436, Melbourne, Australia. Association for Computational Linguistics.

Roger Levy. 2008. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Richard L. Lewis and Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3):375–419.

Alessandro Lopopolo, Stefan L. Frank, Antal van den Bosch, and Roel M. Willems. 2017. Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain. *PLoS ONE*, 12(5):e0177794.

Francis Mollica, Geoff Bacon, Noga Zaslavsky, Yang Xu, Terry Regier, and Charles Kemp. 2021. The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, 118(49):e2025993118.

Richard Montague. 1970. Universal grammar. *Theoria*, 36(3):373–398.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.

Hiroshi Noji and Yohei Oseki. 2021. Effective batching for recurrent neural network grammars. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4340–4352, Online. Association for Computational Linguistics.

Barbara Partee and Mats Rooth. 1983. Generalized conjunction and type ambiguity. *Formal semantics: The essential readings*, pages 334–356.

Barbara Hall Partee. 1970. Negation, conjunction, and quantifiers: Syntax vs. semantics. *Foundations of Language*, 6:153–165.

Steven T. Piantadosi, Harry Tily, and Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529.

Steven T. Piantadosi, Harry Tily, and Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition*, 122(3):280–291.

Tiago Pimentel, Clara Meister, Ethan Wilcox, Kyle Mahowald, and Ryan Cotterell. 2023. Revisiting the optimality of word lengths. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2240–2255, Singapore. Association for Computational Linguistics.

Terry Regier, Charles Kemp, and Paul Kay. 2015. Word meanings across languages support efficient communication. In *The Handbook of Language Emergence*, chapter 11, pages 237–263. John Wiley & Sons, Ltd.

Philip Resnik. 1992. Left-corner parsing and psychological plausibility. In *COLING 1992 Volume 1: The 14th International Conference on Computational Linguistics*.

John R. Ross. 1967. *Constraints on variables in syntax*. Ph.D. thesis, MIT.

Cory Shain, Idan Asher Blank, Marten van Schijndel, William Schuler, and Evelina Fedorenko. 2020. fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, 138:107307.

Cory Shain, Clara Meister, Tiago Pimentel, Ryan Cotterell, and Roger Levy. 2024. Large-scale evidence for logarithmic effects of word predictability on reading time. *Proceedings of the National Academy of Sciences*, 121(10):e2307876121.

Nathaniel J. Smith and Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3):302–319.

Miloš Stanojević, Jonathan R. Brennan, Donald Dunagan, Mark Steedman, and John T. Hale. 2023. Modeling structure-building in the brain with CCG parsing and Large Language Models. *Cognitive Science*, 47(7):e13312.

Shane Steinert-Threlkeld. 2021. Quantifiers in natural language: Efficient communication and degrees of semantic universals. *Entropy*, 23(10).

Mitchell Stern, Daniel Fried, and Dan Klein. 2017. Effective inference for generative neural parsing. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1695–1700, Copenhagen, Denmark. Association for Computational Linguistics.

Patrick Sturt and Vincenzo Lombardo. 2005. Processing coordinated structures: Incrementality and connectedness. *Cognitive Science*, 29(2):291–305.

Sean Trott and Benjamin Bergen. 2022. Languages are efficient, but for whom? *Cognition*, 225:105094.

Wataru Uegaki. 2022. The Informativeness/Complexity Trade-Off in the Domain of Boolean Connectives. *Linguistic Inquiry*, pages 1–23.

Iris van de Pol, Paul Lodder, Leendert van Maanen, Shane Steinert-Threlkeld, and Jakub Szymanik. 2023. Quantifiers satisfying semantic universals have shorter minimal description length. *Cognition*, 232:105150.

Marten van Schijndel, Andy Exley, and William Schuler. 2013. A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science*, 5(3):522–540.

Jennifer C. White and Ryan Cotterell. 2021. Examining the Inductive Bias of Neural Language Models with Artificial Languages. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 454–463, Online. Association for Computational Linguistics.

Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. 2018. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942.

George K. Zipf. 1949. *Human behavior and the principle of least effort*. Addison-Wesley.