The cross-linguistic order of adjectives and nouns may be the result of iterated pragmatic pressures on referential communication

Dhara Yu, Brandon Waldon, Judith Degen
Stanford University
{dharakyu, bwaldon, jdegen}@stanford.edu

Abstract

The world's languages differ in how they order adjectives and nouns relative to each other. We ask whether cross-linguistic variation and systematicity in adjective-noun order can be explained by the iterated pressure for pragmatic referential communication. To this end, we apply the Rational Speech Act framework with an an iterated learning mechanism to study how cooperative pressures may shape typological regularities in referential communication. First, we show that the less informative adjectives are relative to nouns, the more likely they are to occur post-nominally. This is the case when informativeness is manipulated via the composition of the lexical space (i.e., changing the relative number of adjectives vs. nouns that are available for reference), and via the inherent referential utility of adjectives vs. nouns. Secondly, we show that under the assumption that nouns are on average more informative than adjectives, the model predicts a cross-linguistic distribution of ordering preferences that qualitatively resembles the empirical one, with these biases becoming further entrenched with iterated language use. Taken together, these results suggest a possible pathway for syntactic preferences to be calcified over time as the result of pragmatic communicative pressures on language.

Keywords: probabilistic pragmatics; Rational Speech Act; word order; iterated learning; referential communication

Introduction

Languages vary in their grammatical properties, but also exhibit a remarkable degree of systematicity (Greenberg, 1963). Some have proposed that observed typological regularity arises from innate cognitive biases (Goldin-Meadow et al., 2008; Culbertson et al., 2012; Futrell et al., 2015). Other work has sought to explain it as the result of functional pressures of efficient communication under cognitive resource constraints (Maurits et al., 2010; Hahn & Xu, 2022).

This paper is concerned with one typological feature, the order of adjectives and nouns relative to one another. By one count, approximately 27% of documented languages have prenominal adjectives (e.g., English, "red pin"), 64% have postnominal adjectives (e.g., Spanish, "tachuela roja"), and 8% exhibit no clear preference (Dryer, 2013).

Building on a line of work suggesting that languages realize an optimal tradeoff between ambiguity and complexity in the domain of grammar (Hahn et al., 2020), we ask whether documented preferences for adjective-noun order can emerge from incremental *pragmatic reasoning*. Pragmatic reasoning, the process by which what is said literally becomes contextually enriched, makes use of expectations that speakers be informative, while avoiding expressions that are too costly (Grice, 1975). We test whether pressures from pragmatic reasoning may explain cross-linguistic variation and systematicity in syntactic features, and whether language repeatedly produced under such pressures gives way to ossified syntactic conventions.

Previous large-scale corpus studies investigating efficiency-based pressures on the emergence of typological syntactic patterns have relied on proxy terms for capturing the notion of "informativeness," which inherently requires grounding in a meaning space. For example, Hahn et al. (2020) used parseability as a proxy. To capture this grounding component more directly, we use reference games, a foundational and well-studied context for pragmatic language use (Clark & Wilkes-Gibbs, 1986; Frank & Goodman, 2012; Brown-Schmidt & Tanenhaus, 2008; Peloquin et al., 2020), as a testbed for the hypothesis that iterated incremental pragmatic reasoning may drive the cross-linguistic syntactic preference for postnominal adjectives.

To test this hypothesized relationship between pragmatics and syntax, we use the Rational Speech Act (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016), a Bayesian framework for modeling pragmatic production and comprehension, to formalize the relevant notion of incremental referential pragmatic reasoning. We extend the model with a learning mechanism to capture the formation of pragmatic conventions in an *iterated learning* paradigm (Kirby, 2001; Kirby et al., 2014). There are two key features of the proposed RSA model that allow us to provide a proof of concept for exploring how typological conventions can emerge from pragmatics: first, it affords a well-defined notion of *informativeness* (which can vary as a function of the referential context, keeping the lexicon and individual word informativeness constant; or as a function of variable inherential utility, keeping the context constant) and *cost*. Second, it computes utterance utility incrementally, word by word, enabling the breaking of the symmetry in production probabilities between the NOUN-ADJECTIVE and ADJECTIVE-NOUN word orders.

The key results are that i) lacking a syntactic convention, the model is more likely to place more informative words earlier in an utterance in a single-shot utterance production setting; and ii) this bias is calcified in iterated production settings. Thus, as long as nouns are on average more informative than adjectives, the model predicts a cross-linguistic typological distribution that resembles the empirical one.

Word order preferences emerge from varying informativeness

The basic model

In this section, we formalize the factors that lead to the emergence of a word order preference. To do so, we leverage the RSA framework, which models pragmatic communication between speakers and listeners as recursive reasoning about each others' mental states. RSA can be understood as a formalization of Gricean maxims (Grice, 1975) regarding assumptions about speaker behavior. In this framework, the utility of an utterance produced by the speaker is defined as a tradeoff between the utterance's contextual informativeness and its cost. We define a *literal listener* $L_0(o|u)$ and a *pragmatic speaker* $S_1(u|o)$:

$$L_0(o|u) \propto [[u]](o) \cdot P(s) \tag{1}$$

$$S_1(u|o) \propto \exp[\alpha(\log L_0(o|u) - C(u))] \tag{2}$$

where $[[\cdot]]$ is a semantic denotation function for utterance alternatives $u \in U$ applied to objects $o \in O$, and $C(u)$ defines the cost of u.[1] The pragmatic speaker chooses utterances soft-optimally based on their utility, which is realized as a tradeoff between informativeness and cost. An utterance's informativeness is defined as the log probability of the literal listener correctly inferring the intended referent. Thus, the more likely the semantics of an utterance is to lead the literal listener to correctly infer the intended referent in context, the more likely the speaker is to produce that utterance; the more costly the utterance is, the less likely it is to be produced. The parameter α is the softmax temperature, which captures the optimality of the speaker.

Two important limitations of the standard RSA model are that i) all words are assumed to be equally useful for disambiguating, and ii) utilities are calculated globally (for entire utterances), rather than word by word (e.g., Frank & Goodman, 2012). This means that the standard model cannot produce an asymmetry in the predicted preference for the order of adjectives and nouns: it assigns equal utility, and hence equal production probability, to NOUN-ADJECTIVE and ADJECTIVE-NOUN.

Our model makes use of two extensions to break symmetry. First, we use a *continuous* rather than discrete semantic function (Degen et al., 2020), which captures the intuition that utterances may be true (or that objects may exhibit properties denoted by nouns and adjectives) to a gradient extent. This function $[[\cdot]]^C$ is defined as follows:

$$[[u]]^C(o) = \prod_{i=1}^{|u|} \mathcal{L}^C(o, u_i) \tag{3}$$

$$\mathcal{L}^C(o, u_i) = \begin{cases} v & \text{if } u_i \text{ is true of } o \\ 1-v & \text{otherwise} \end{cases} \tag{4}$$

where $\mathcal{L}^C$ is a function mapping object-utterance pairs to a semantic value $v \in [0,1]$, and u_i indicates the ith lexical item of the utterance u.

The second extension is *incremental* planning. Motivated by both the computational inefficiency of basic RSA and the psycholinguistic fact that language is processed incrementally (Sedivy et al., 1999), incremental versions of RSA have recently been proposed in which pragmatic reasoning is modeled at the level of subunits, rather than over a whole utterance (Cohn-Gordon et al., 2019). An incremental model that performs inference at the level of individual words results in qualitatively different predictions compared to a "global" model that considers full utterances.

To account for cross-linguistic variation in referring expressions, specifically differences in the rate of redundant modifiers (Rubio-Fernández, 2016) and in visual search strategies (Rubio-Fernandez et al., 2020), Waldon and Degen (2021) propose *combining* continuous semantics with incremental processing over word units, constituting the *continuous-incremental* RSA (CI-RSA) model (for an alternative model, see Jara-Ettinger & Rubio-Fernández, 2021).

Likewise in our setting, breaking the symmetry between word orders requires *both* continuous and incremental semantics, motivating the use of the CI-RSA model. This requires specifying an incremental literal listener ${L_0}^{INC}(o|c,u_i)$ that takes into account the partially produced utterance $c = [u_0,..,u_{i-1}]$ and the next lexical item u_i:

$${L_0}^{INC}(o|c,u_i) \propto \chi^C(o,c,u_i) \cdot P(o) \tag{5}$$

$$\chi^C(o,c,u_i) = \frac{\sum [[u']]^C(o) : u' \text{ is a cont. of } c+u_i}{u' : u' \text{ is a cont. of } c+u_i} \tag{6}$$

Here, $\chi^C(o,c,u_i)$ is a string interpretation function that returns a (continuous) semantic value for a partial utterance c concatenated with the next lexical item u_i. This function captures the intuition that when there are more possible continuations of c, any particular u_i is more informative to a listener than when there are fewer.

The resulting pragmatic speaker $S_1(u|o)$ takes the product of the incremental utterance utilities defined by $S_1^{INC}(u_j|c,o)$:

$$S_1(u|o) \propto \prod_{j=1}^{|u|} S_1^{INC}(u_j|c,o) \tag{7}$$

$$S_1^{INC}(u_j|c,o) = \exp[\alpha(\log {L_0}^{INC}(o|c,u_i) - C(u_i))] \tag{8}$$

where $C(u_i)$ denotes the cost associated with producing utterance u_i. Intuitively, CI-RSA can produce asymmetric probabilities for different word orders because it is formulated as a *greedy* model: it assigns higher probability to utterances for which lexical items with higher utility appear earlier in the string.[2]

[1] Utterance cost in RSA captures speaker-side production costs, which may be influenced by retrieval difficulty, complexity (phonetic, phonological, morphological, or syntactic), and/or other meaning-independent factors that make the utterance more costly to produce (Goodman & Frank, 2016; Degen, 2023).

[2] An implementation of the CI-RSA model as well as code to run all simulations in the paper is available at `https://github.com/dharakyu/crosslinguistic-order`.

Two notions of informativeness

At its core, the RSA framework captures the communicative tradeoff between *informativeness* (how useful an utterance is for conveying the intended meaning) and *cost* (the effort expended on producing the utterance). There are at least two different ways of modulating utterance informativeness in RSA, resulting in two different notions of informativeness that all have the same causal effect in the model.

We call the first notion of informativeness **inherent informativeness** or inherent referential utility. This corresponds to the semantic value parameter in continuous semantics RSA. The higher this value, the more informative the utterance.[3]

We call the second notion of informativeness **contextual informativeness**. This captures the utility of an utterance for disambiguating, given the space of possible utterances and the composition of the referring context.

To make this idea concrete, consider a speaker with lexicon {blue, red, pin, dress, truck} in a referring context depicted in Figure 1A. Producing the lexical item *red* is not useful for identifying the referent since all of the objects are red. The item *truck*, on the other hand, is useful because it precisely identifies one object. This is an instance of color adjectives having low (or no) contextual informativeness and nouns having high contextual informativeness.

Now consider the same speaker in a minimally different referring context, as in Figure 1B. Producing a color modifier in this context is more useful than in the previous one because identifying an object's color restricts the referential domain to half the size (assuming v^{adj} is close to 1). Here, color adjectives have relatively higher contextual informativeness compared to the previous context. However, color adjectives are still less contextually informative than nouns, since identifying the type of object reduces the space of possible targets to one-third of the original size. This notion of utterance informativeness has long featured in computational models of referring expression production (Dale, 1989; Dale & Reiter, 1995; Van Gompel et al., 2019).

Manipulating inherent informativeness through semantic values

To develop an intuition for how an ordering preference can emerge via manipulation of inherent informativeness let us consider the simple example in Fig. 2. Critically, this example assumes higher inherent informativeness (higher semantic values) for object category-denoting nouns than for adjectival modifiers. That is, we assume that adjectives are "noisier" than such nouns. This assumption is consistent with Gentner's (1982) *Natural Partitions* hypothesis, which states that highly concrete, "cognitively preindividuated" concepts

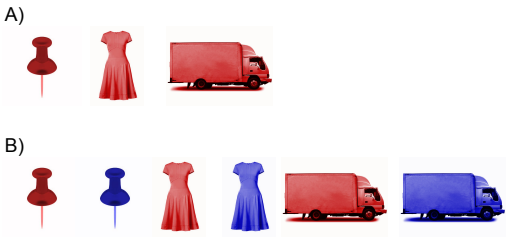

Figure 1: Referring contexts in which the adjective "red" is minimally contextually informative (A) vs. somewhat informative (B), but less informative than a noun.

tend to be lexicalized as nouns rather than in other, predicative syntactic domains (Gentner & Boroditsky, 2001, p. 212). Among the motivations for this hypothesis is the observation that between languages, form-to-meaning mappings are most consistent in the nominal domain (Gentner, 1981): that is, there is considerable cross-linguistic stability when it comes to how object-denoting nouns partition conceptual space. For example, the concept denoted by the English noun *bottle* is equivalently lexicalized in Spanish, but the verb *float* is translated as a periphrastic construction that indicates manner of motion – see also Talmy (1975, 1978). Our analysis assumes that, in referring contexts, the same ontological properties of nominal concepts that explain these regularities across speech communities are also exploited pragmatically by speakers within a particular speech community in referring contexts: that is, because such concepts are often derived from "highly cohesive collections of percepts" (Gentner, 1982, p. 324) we assume that speakers are relatively certain that such concepts can be invoked to establish reference and are less certain when it comes to those concepts lexicalized by adjectives. [4]

A model that assumes a higher semantic value for nouns relative to adjectives predicts an overall preference for the NOUN-ADJ over the ADJ-NOUN order. For the specific semantic values of $v^{\text{noun}} = 0.99, v^{\text{adj}} = 0.95$, the probability of producing a NOUN-ADJ utterance to refer to an item in the context, i.e. $S_1(\text{"dress red"}|r_{\text{red-dress}})$, is 0.512, demonstrating a slight preference for the postnominal adjective ordering.

This simple example illustrates that inducing a difference in the informativeness of adjectives and nouns by varying the semantic values leads to the breaking of the symmetry be-

[3] Degen et al. (2020) offer a more in-depth discussion of what semantic values may represent. Some possibilities they put forth include the difficulty of determining whether the property denoted by the utterance is true of the object; the strength of prior beliefs about correlations between objects and properties; and the past probability of communicative success using the given utterance.

[4] Further motivation for the assumption of greater adjective noise compared to nouns comes from the child language acquisition literature: for example, young children learn more nouns than adjectives in their early vocabulary (Nelson, 1973). Moreover, 2-year-olds cannot yet identify novel words as adjectives, whereas 4-year-olds can (Hall et al., 1993). Young children's difficulty with adjectives may be the result of greater ambiguity in adjectives relative to concrete nouns (Gentner & Boroditsky, 2001). Furthermore, young learners have difficulty integrating adjectives in adjective-noun referring expressions to disambiguate between items (Fernald et al., 2010). This pattern holds both for children learning English, a prenominal adjective language, and for children learning Hebrew, a postnominal adjective language (Ninio, 2004), suggesting that the challenge of integrating adjectives is invariant to the word order convention of a young learner's native language.

tween the two possible word orders even in the absence of cost differences.

Possible referents: $r_{\text{red-pin}}, r_{\text{blue-pin}}, r_{\text{red-dress}}, r_{\text{blue-dress}}$
Lexical items: *red, blue, pin, dress*
Possible utterances: "pin", "dress", "red", "blue", "pin red", "red pin", "pin blue", "blue pin", "dress red", "red dress", "dress blue", "blue dress"
Semantic value of noun concepts (v^{noun})**:** 0.99
Semantic value of adjective concepts (v^{adj})**:** 0.95
Optimality parameter α**:** 5
Cost: $C(\text{noun}) = C(\text{adj}) = 0$
Speaker probabilities:
$S_1(\text{"pin red"}|r_{\text{red-pin}}) = 0.512$
$S_1(\text{"red pin"}|r_{\text{red-pin}}) = 0.426$

Figure 2: Parameter values for a simple example context, and resulting model predictions. Note that these probabilities do not add up to 1 because there is a nonzero probability of producing the utterances *red* and *pin*.

Manipulating contextual informativeness through varying scene composition

Building on this simple example, we ran a series of simulations to characterize how varying **contextual** informativeness of utterances affects the model's preferred word order, and how the two notions of informativeness interact with each other. Recall that contextual informativeness refers to the disambiguation potential of an utterance in a given referring context. The most intuitive way to manipulate contextual informativeness is to change the composition of the referring context, such that certain lexical items pare down the space of possible referents to a varying extent. Accordingly, in our simulations we do exactly this.

Given a lexicon containing n adjectives and k nouns in the lexicon, we assume that the referring context contains $n \cdot k$ objects. The space of possible utterances consists of each of the n unique adjectives, k unique nouns, and $2nk$ combinations of adjective-noun pairs, which can appear in the NOUN-ADJ or ADJ-NOUN order. Thus, contextual informativeness can be operationalized as the *ratio* between the disambiguation potential of adjectives vs. that of nouns in a given context.

Figure 3 shows that the less informative an adjective is relative to the noun, the more likely it is to occur postnominally. This is true when the informativeness is varied via semantic values, or via the composition of the referring context. When the inherent informativeness of nouns is higher than that of adjectives (i.e., in the upper left quartile of each facet), the model predicts a preference for the NOUN-ADJ order. As the contextual informativeness of adjectives increases, going from the leftmost facet to the rightmost, the preference for ADJ-NOUN increases. Furthermore, these two different notions of informativeness may counteract each other: for example, in the 5 adjective, 3 noun referring context, there are parameter values for which ADJ-NOUN is predicted even when the inherent informativeness of nouns is higher than that of adjectives, due to countervailing pressures from the high contextual informativeness of adjectives.

Other ways to manipulate contextual informativeness

In the previous section, we demonstrated *one* way to vary contextual informativeness, namely by changing the composition of the referring context. There are several other ways to induce this variation in contextual informativeness. For instance, another way of doing so would be to change the lexicon and semantics while holding the referring context constant. Concretely, consider a language with separate lexical items for the colors red and orange, vs. a language without such a distinction, given a fixed referring context containing both red and orange items: according to our model, speakers of the first language would be more inclined to produce the adjective first because adjectives are more useful for disambiguation.

Additionally, contextual informativeness can also be varied by adding multiple items of the same type and property to the referring context, which in turn differentially affects the contextual informativeness of specific lexical items. Introducing duplicates of $r_{\text{red-pin}}$ to a referring context has the effect of *reducing* the contextual informativeness of the specific noun *pin* because it becomes less discriminating, while *increasing* the contextual informativeness of other nouns, because indicating that an object is something other than a pin is useful in a referring context dominated by pins.

We emphasize that the core pattern of the noun-first order being preferred over the adjective-order will generalize across conditions in our model, as long as noun utility is greater than that of adjective utility.

Pragmatic biases are perpetuated by iterated learning

Iterated learning model

Language evolution has been widely construed as *iterated learning*, in which learning biases and behaviors emerge from observing past generations of learners (Kirby et al., 2014). Cast in formal terms, iterated learning can be modeled as a process of rational Bayesian agents sampling from a posterior distribution of hypotheses over languages given observed data, where this posterior distribution in generation i is used as the prior for generation $i+1$ (Griffiths & Kalish, 2007). Inspired by this approach, we define an iterated learning model for word order preference where the posterior distribution over possible word orders is approximated by a likelihood

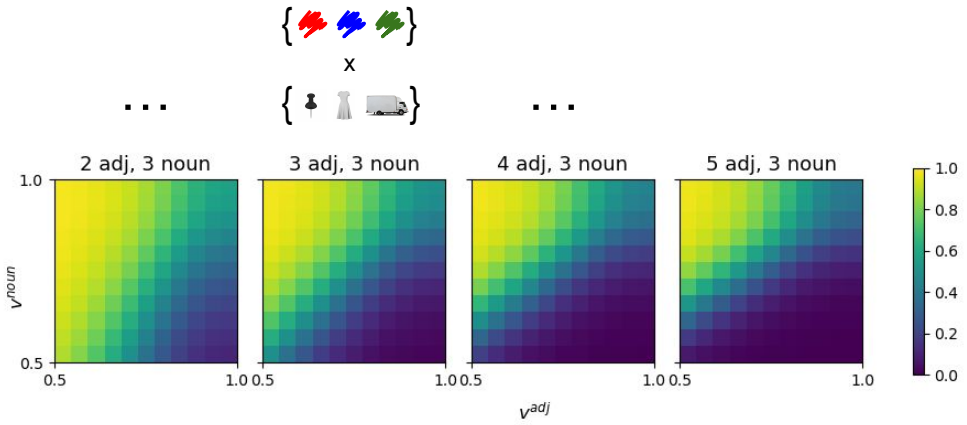

Figure 3: S_1(NOUN-ADJ), i.e. probability of producing a prenominal referring expression, varying $v^{\text{noun}}, v^{\text{adj}}$ and composition of the referring context. Adjectives are more informative as one moves rightward through the facets. Probabilities of producing NOUN-ADJ and ADJ-NOUN are normalized to sum to 1 (i.e., ignoring NOUN or ADJ productions). An example reference context, consisting of 9 objects instantiating the Cartesian product of 3 adjectives and 3 nouns, is shown above the second facet.

term and an utterance prior over word orders. Formally, we can define the posterior distribution $S_1(u|o)$, as follows:

$$S_1(u|o) \propto \underbrace{\left(\prod_{j=1}^{|u|} S_1^{INC}(u_j|c,o)\right)}_{\text{likelihood from RSA}} \cdot \underbrace{W(u|o)}_{\text{utt. prior}} \tag{9}$$

Note that this is the same as equation (7) defining CI-RSA, with the addition of the utterance prior term $W(u|o)$, which represents the prior probability of producing utterance u given the intended referent is object o. It follows that $\sum_{u \in U} W(u|o) = 1$, where U is the set of all utterances such that the utterance is true of object o.

To perform the posterior update step, we sample over the categorical distribution $S_1(u|o)$ for all $u \in U$, which yields observed utterance u'. We then take the utterance type of u' (i.e. NOUN-ADJ or ADJ-NOUN), and increment the count of observed utterances of that type by 1 (representing a single observation of that utterance type). The normalized count across utterance types then becomes the prior distribution over utterance types for the next step.

Simulating language evolution

We simulate the process of language evolution and convention formation in the iterated learning model. Motivated by the single-shot informativeness results, we assume that nouns are *on average* less noisy than adjectives. Specifically, we sample the semantic values $v^{\text{adj}} \sim \text{Beta}(45,5), v^{\text{adj}} \sim \text{Beta}(47.5,2.5)$ with expectations of 0.9 and 0.95, respectively. We assume the same simple referring context as in Figure 2, which consists of $r_{\text{red-pin}}, r_{\text{blue-pin}}, r_{\text{red-dress}}, r_{\text{blue-dress}}$.

Figure 4A shows that the initial average bias for the NOUN-ADJ order magnifies over the course of learning. The iterated learning model predicts an overall convergence to the NOUN-ADJ order, but there are instances in which the production of ADJ-NOUN is ultimately favored. Figure 4B shows the probability of NOUN-ADJ after 300 generations of learning for each of the 1000 runs. Most of the probability mass is to the right of 0.5, indicating a general preference for NOUN-ADJ, yet there are still runs in which the final preferred order is ADJ-NOUN.

These iterated learning simulations yield a distribution over word orders that is consistent with the qualitative pattern observed in the real world: an overall, but not universal, preference for NOUN-ADJ. Furthermore, the model captures the documented pattern of certain languages that cannot be obviously classified as pre- or postnominal languages, as indicated by the probability mass around 0.5 in Fig. 4B. The fact that the model predicts the existence of "suboptimal" languages (like English) that do not realize the NOUN-ADJ order, as well as languages without a clear preference, suggests that iterated learning is a suitable framework for modeling the evolution of syntactic convention. Overall, our results suggest that *iterated* pragmatic pressures may shape the syntactic structure of noun-adjective referring expressions.

Investigating the role of cost in iterated learning

In previous simulations, we assumed that all words in the lexicon, whether adjectives or nouns, have a uniform cost of 0. However, work in psycholinguistics suggests that not all words are equally easy to produce, and suboptimal yet more "accessible" lexical items may be produced earlier in the linear order of an utterance, as a consequence of bounded cognitive resources (V. S. Ferreira & Dell, 2000; F. Ferreira & Patson, 2007; Goldberg & Ferreira, 2022),.

This points to cost, as operationalized in the RSA framework, as an additional factor that can affect word order preferences. Cost is a countervailing pressure against informativeness, meaning that a sufficiently high cost associated with adjectives can induce a preference for the NOUN-ADJ order,

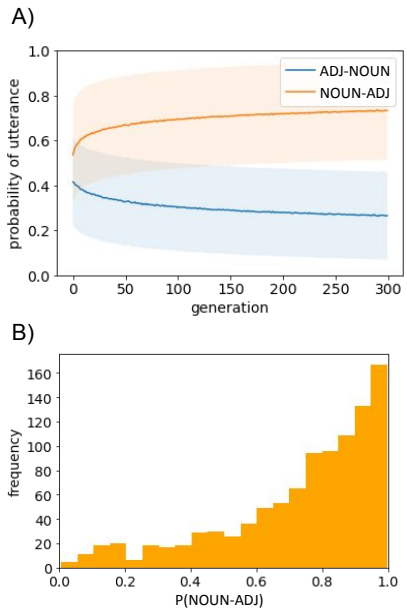

Figure 4: A) Production probabilities for NOUN-ADJ and ADJ-NOUN utterances in the 2 noun, 2 adjective referring context over 300 generations. Bold line represents the mean of 1000 runs, with the lightly shaded area representing the standard deviation. Probabilities of the two possible word orders are normalized to sum to 1. B) Histogram of $P(\text{NOUN-ADJ})$ after the 300th generation for each of the 1000 transmission chains.

even if the adjectives are more informative, broadly speaking, than nouns.

In principle, our model could capture the same qualitative pattern of an overall preference for the postnominal order if we were to operationalize the asymmetry between adjectives and nouns with cost, rather than with informativeness. A cost-based explanation would suggest that ordering conventions may arise from a general pressure to minimize production cost, rather than the more specific pragmatic pressure to be informative. We believe that the cost-based explanation is less likely, given, e.g., the importance of informativeness in explaining related reference phenomena (e.g., Degen et al., 2020). But distinguishing between these explanations is an important avenue of future research.

General Discussion

In this work, we provided a proof of concept that in-the-moment incremental pragmatic reasoning may shape cross-linguistic syntactic conventions, focusing on the case study of the order of adjectives and nouns in referential contexts. By formalizing the iterated pragmatic pressures on referential communication within the Rational Speech Act framework, we showed that a preference for the empirically documented NOUN-ADJ order is predicted as long as nouns are on average more informative than adjectives. Our model also captured the existence of languages without a strong preference for either order, consistent with empirical observations. The emergence of syntactic conventions in a language is a complex process, presumably the result of many different forces. Our results suggest that pragmatic reasoning may be one *possible* causal mechanism for syntactic ossification.

There are multiple limitations to this work. The most important one is that our results are entirely based on computational modeling, so they should be bolstered by empirical validation. One possible approach to doing so could be to quantify the contextual informativeness of different word types across languages through corpus analysis, and then to correlate this data with the observed word order conventions. Another avenue for empirical study could be through artificial language learning experiments, in which participants learn a new language from scratch as informativeness levels of different lexical items are systematically varied.

Another limitation is that we assumed a fixed lexicon, i.e. that there is an established mapping between lexical items and objects in the world. This is a deviation from naturalistic conditions, where semantics and syntax co-evolve with one another (for example, in the village sign language Central Taurus Sign Language, Ergin et al., 2021). Future work could explore the multi-directional interaction between semantics, pragmatics and syntax, as well as the role of functional pressures for communicative success (following an approach similar to Brochhagen et al., 2018).

Furthermore, our focus on one particular typological feature raises the question: to what extent do these findings generalize to other ordering phenomena, such as the order of multi-adjective referring expressions, or basic word order? Indeed, our result that cross-linguistic adjective-noun ordering preferences can be explained by more informative words appearing earlier is in tension with the observed trend of more subjective adjectives occurring earlier in multiple-adjective referring expressions in prenominal adjective languages (Scontras et al., 2017, but note that there is no tension in postnominal languages). Thus, it may be the case that the tendency for more subjective adjectives to occur further from the head noun may be a separate functional pressure that interacts with the pressure to place words with higher disambiguation potential at the beginning of utterances. Additionally, investigating basic word order will require moving beyond the reference game setting; accordingly, moving forward we hope to generalize our methodological approach to investigate ordering preferences for a broader set of typological features.

In conclusion, the work reported here provides a proof of concept that pragmatic pressures on referential communication may have been involved in shaping syntactic patterns of adjective-noun ordering cross-linguistically. The generalizability of this formal investigation to other phenomena and an empirical test of the assumptions and results are exciting avenues for future research.

Acknowledgments

We thank Jiayi Lu and other members of the Stanford ALPS Lab, the audience members of Culture Conference 2022, and the anonymous reviewers for useful feedback.

References

Brochhagen, T., Franke, M., & van Rooij, R. (2018). Coevolution of lexical meaning and pragmatic use. *Cognitive Science*, *42*(8), 2757-2789.

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive science*, *32*(4), 643–684.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Cohn-Gordon, R., Goodman, N., & Potts, C. (2019). An incremental iterated response model of pragmatics. In *Proceedings of the society for computation in linguistics (SCiL) 2019* (pp. 81–90). doi: 10.7275/cprc-8x17

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*, 306-329.

Dale, R. (1989). Cooking up referring expressions. In *27th annual meeting of the association for computational linguistics* (pp. 68–75).

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive science*, *19*(2), 233–263.

Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics*, *9*(1), null. doi: 10.1146/annurev-linguistics-031220-010811

Degen, J., Hawkins, R. X. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to 'overinformative' referring expressions. *Psychological Review*, *127*, 591–621. doi: 10.1037/rev0000186

Dryer, M. S. (2013). Order of adjective and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Ergin, R., Kürşat, L., Hartzell, E., & Jackendoff, R. (2021, Jan). *Central taurus sign language: On the edge of conventionalization.* PsyArXiv. doi: 10.31234/osf.io/x9emd

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive Psychology*, *60*(3), 190-217. doi: https://doi.org/10.1016/j.cogpsych.2009.12.002

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71-83.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296-340.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998 - 998.

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, *136*, 215-221. doi: https://doi.org/10.1016/j.cognition.2014.11.022

Gentner, D. (1981). Some interesting differences between nouns and verbs. *Cognition and Brain Theory*, *4*, 161-178.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Language*, *2*, 301-334.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (p. 215–256). Cambridge University Press. doi: 10.1017/CBO9780511620669.010

Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. *Trends in Cognitive Sciences*, *26*(4), 300-311.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 9163-8. doi: 10.1073/pnas.0710060105

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818-829.

Greenberg, J. H. (1963). Universals of language.

Grice, H. P. (1975). Logic and conversation. In M. Ezcurdia & R. J. Stainton (Eds.), *The semantics-pragmatics boundary in philosophy.* Broadview Press.

Griffiths, T. L., & Kalish, M. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, *31*, 441-480.

Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, *117*(5), 2347-2353. doi: 10.1073/pnas.1910923117

Hahn, M., & Xu, Y. (2022). Crosslinguistic word order variation reflects evolutionary pressures of dependency and information locality. *Proceedings of the National Academy of Sciences*, *119*(24), e2122604119. doi: 10.1073/pnas.2122604119

Hall, D. G., Waxman, S. R., & Hurwitz, W. M. (1993). How two- and four-year-old children interpret adjectives and count nouns. *Child Development*, *64*(6), 1651–1664.

Jara-Ettinger, J., & Rubio-Fernández, P. (2021). The social basis of referential communication: Speakers construct physical reference based on listeners' expected visual search. *Psychological review*.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emer-

gence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102-110. doi: 10.1109/4235.918430

Kirby, S., Griffiths, T., & Smith, K. (2014, 10). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. doi: 10.1016/j.conb.2014.07.014

Maurits, L., Navarro, D., & Perfors, A. (2010). Why are some word orders more common than others? a uniform information density account. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems* (Vol. 23). Curran Associates, Inc.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*(1/2), 1–135.

Ninio, A. (2004). Young children's difficulty with adjectives modifying nouns. *Journal of Child Language*, *31*(2), 255–285. doi: 10.1017/S0305000904006191

Peloquin, B. N., Goodman, N. D., & Frank, M. C. (2020). The interactions of rational, pragmatic agents lead to efficient language structure and use. *Topics in Cognitive Science*, *12*(1), 433-445.

Rubio-Fernandez, P., Mollica, F., & Jara-Ettinger, J. (2020, 06). Speakers and listeners exploit word order for communicative efficiency: A cross- linguistic investigation. *Journal of Experimental Psychology General*. doi: 10.1037/xge0000963

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? an efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*.

Scontras, G., Degen, J., & Goodman, N. D. (2017). Subjectivity predicts adjective ordering preferences. *Open Mind*, *1*(1), 53–66.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109-147. doi: https://doi.org/10.1016/S0010-0277(99)00025-6

Talmy, L. (1975). Figure and ground in complex sentences. In *Proceedings of the first annual meeting of the berkeley linguistics society.*

Talmy, L. (1978). The relation of grammar to cognition–a synopsis. In *Theoretical issues in natural language processing-2*. Retrieved from https://aclanthology.org/T78-1003

Van Gompel, R. P., Van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. J. (2019). Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological review*, *126*(3), 345.

Waldon, B., & Degen, J. (2021, February). Modeling cross-linguistic production of referring expressions. In *Proceedings of the society for computation in linguistics 2021* (pp. 206–215). Online: Association for Computational Linguistics.