

Word order in a grammarless language: A ‘small-data’ information-theoretic approach

Nicholas A. Lester (nlester@umail.ucsb.edu)

Department of Linguistics, South Hall 3432

Santa Barbara, CA 93106 USA

Fermín Moscoso del Prado Martín (fmoscoso@umail.ucsb.edu)

Department of Linguistics, South Hall 3432

Santa Barbara, CA 93106 USA

Abstract

David Gil has argued that Riau Indonesian (Sumatra, Indonesia) has no syntax, or at least not much. This controversial analysis undermines all current models of grammar, especially those describing acquisition and on-line processing. To test the strength of this analysis, we computed the *information gain* holding between unigram and bigram models of regular and randomized samples of English and Riau Indonesian. English samples were included as a relatively syntax-heavy baseline. We then correlated information gain values with language (*English vs. Riau Indonesian*), text type (*original vs. randomized*), and their interaction within a linear mixed-effects regression. The results suggest (a) that English and Riau Indonesian have the same amount of bigram informativity and (b) that randomization eliminates this effect in both languages. These findings do not support Gil’s syntax-free analysis; rather, they point to some kind of productive constraints on Riau Indonesian word order.

Keywords: Indonesian; word classes; n-gram models; information gain; entropy.

Introduction

Most theories of grammar posit a layer of categorization that discriminates words into different functional types, or *word classes*.¹ These word classes define the combinatorial possibilities of words, and so serve a central function within the syntax of any given language. Examples of common word classes include nouns, verbs, adjectives, adverbs, and adpositions. The widespread recurrence of these classes cross-linguistically has led some researchers to argue for their either being cognitively basic or universally available as part of the genetically endowed human linguistic apparatus. Over the past few decades, however, a growing body of evidence from languages across the globe has come to challenge the basic or universal status of even the most common word classes, including nouns and verbs (cf. Rijkhoff & van Lier, 2013). The most radical argument to emerge from this line of research comes from David Gil and his analysis of Riau Indonesian (RI), a colloquial variety of Bahasa Indonesian (BI) spoken on the central-eastern coast of Sumatra.

According to Gil (1994; 2013), RI essentially lacks any grammatically relevant word classes. In itself, this claim is

not so shocking, as similar arguments have been made for other languages. For example, Hengeveld (2013) points out the difficulty of assigning Samoan words to default word classes. In Samoan, most words can occur in any of the morphosyntactic frames defining *predication*, *reference*, and *modification*. This phenomenon has been widely discussed as a form of *lexical flexibility* (van Lier & Rijkhoff, 2013). But Gil goes further, arguing that RI has largely unconstrained word order and no inflectional morphology; that is, he argues that RI has no syntax. This last point sharply differentiates RI from languages like Samoan, in that the latter indeed provides systematic, *grammatically encoded* cues for reconstructing meaning: interpretation of a word’s functional-semantic contribution to the sentence is guided by its morphosyntactic positioning. By contrast, according to Gil, RI has no set of syntactic constructions to which regular interpretations could be linked.

The existence of a language like RI contradicts the most basic assumptions underlying most (if not all) established theories of grammar, especially in the domains of processing and acquisition. For example, linguists generally assume that all humans acquire their respective native language(s) by means of the same perceptual-cognitive system. In one formulation, this apparatus constitutes a genetically pre-specified Language Acquisition Device (LAD) which mediates the interaction between linguistic experience and the inborn set of grammatical categories and rules known as Universal Grammar (UG). Other formulations emphasize the role of domain-general (i.e., not purely linguistic) statistical learning mechanisms during language acquisition. According to these theories, all humans bring the same simple learning algorithms to bear on the problem of parsing input sequences (e.g., the ‘chunking’ of sequences with high internal transitional association strengths). Under either theory, knowledge of word sequences, either in the form of innate constraints within UG or statistical generalizations over the likelihood of candidate word combinations, stands at the heart of linguistic competence. But how could a language like RI, which lacks reliable constraints on word order, arise from – or be acquired by – a cognitive system that depends so heavily on sequence-driven (i.e., syntactic) generalizations?.

In the next section, we describe the grammatical analysis of RI as developed by David Gil over the past two decades. We then introduce a corpus methodology capable of measuring differences in the regularity of local syntactic

¹ I use the term ‘functional’ in its broadest sense to refer to both the syntactic and semantic properties of words.

structure between two languages on the basis of extremely small samples ($n < 3000$ words). Next, we apply this comparative method to naturalistic data from RI and English and present the results. Finally, we discuss these results with reference to usage-based theories of language structure.

Riau Indonesian: A language without (much) grammar

Like other Austronesian languages of western Indonesia, RI is close to the ideal isolating language: its words exhibit few, if any, inflectional variants. RI does make limited use of a sparse and functionally heterogeneous set of affixes (e.g., the common Indonesian voice-altering prefixes *N-* and *di-*) and some other morphological processes (e.g., reduplication, compounding; Gil, 2008). However, these processes are never inflectional in RI, nor are they ever obligatory. Rather, they represent ‘optional semantic embellishments’ (Gil, 2008: 127).

Unlike other closely related Austronesian languages, RI also closely approximates a hypothetical *monocategorial* language (Gil, 2008). Monocategoriality is related to the notion of lexical flexibility introduced earlier, though it differs in one crucial respect. In highly flexible languages, ‘precategoryal’ words are combined with morphosyntactic templates to derive specifically referential, predicative, or modificational interpretations for each item (Hengeveld & Rijkhoff, 2005). Monocategorial languages, on the other hand, offer no such links between semantic interpretation and morphosyntactic form. Words in these languages are therefore not ‘flexible’ in the sense that they can appear in a number of contrastive syntactic functions; they are ‘monocategorial’ in the sense that they are *at all times* equivalent in ontological and syntactic status. They are in principle free to occur in any syntagmatic combination with any interpretation (provided a few principled exceptions).

To be clear, Gil does not argue that RI lacks word order *altogether*. In fact, he proposes one syntactic and several competing extra-syntactic forces to account for what he sees as apparent syntactically-driven constraints on order (Gil, 2005). Syntactically, he argues that RI has a vanishingly small set of functionally heterogeneous particles. These particles must precede the constituent with which they combine. However, these forms are few in number and may precede (as well as follow) any other word form (that is, they are collocationally unrestricted). Apart from syntax, Gil proposes several soft constraints related to functional interpretations and discourse-pragmatics. Thus, conceptual heads tend to precede their modifiers, topical and discourse-old information tends to surface earlier within ‘clauses’ (though he does not recognize any qualitative difference between individual words and clauses), and so on. However, these principles are only optionally applied. Furthermore, as with the particles, they do not dictate the choice of lexical item: any word may serve any syntactic and thematic function in any position relative to any other word. Finally, these tendencies have been implicated within more rigid syntactic systems, as well, only there they are treated as

parameters governing choices of constructional alternatives (e.g., the dative alternation; Arnold, Wasow, Lonsongco, & Ginstrom, 2000). Therefore, their presence in RI does not rule out the possibility that they are linked directly to syntactic representations.

To the extent that RI word order is unconstrained (with respect to surface tokens), local sequences of words should approach near-random distribution. For any given pairing of words (w_1, w_2), we should find an equal probability of observing the order w_1+w_2 as the alternative w_2+w_1 . In other words, knowledge of the preceding word should not help much in guessing the word to follow. We can refer to this situation as one of *minimal bigram informativity*. This would be the case for purely randomly distributed strings of words, and so can be considered a theoretical lower-bound for bigram structure. Given Gil’s account of the presence of at least some ordering effects, we expect RI to exceed this lower-bound. However, based on Gil’s claims of lexical flexibility at both the syntactic and thematic levels of structure, we can predict that RI should come closer to the minimal threshold than a language like English, which is well known to exhibit robust syntactic constraints on word order. In the next section, we describe an information-theoretic measure known as *information gain*, which can be used to estimate the degree of bigram informativity of a sample text, relative to the baseline probabilities of words. Using this measure, we can compare RI with English to determine just how flexible RI word order is.

Information gain

Information gain is calculated in three stages. First, a unigram model and a bigram model are estimated on the basis of a training text (e.g., some connected subset of a corpus of writing or transcribed speech). Next, an information-theoretic measure known as *cross entropy* is computed for each model to assess the fit of the trained model for an unseen test sample. The cross entropy represents the average number of bits needed to code an event from one distribution as if it belonged to a different distribution. Applying this to the problem of model fit, we can measure the number of bits needed to code a target event according to the distribution predicted by a model trained on part of a text (the training sample) when that event actually belongs to the distribution instantiated by the remainder of that text (the test sample). The resulting value measures the incompatibility of the expectations of the model with the observed properties of the test sample.

Cross entropy is expressed formally as a variation of Shannon’s entropy (Shannon, 1948). Shannon’s entropy represents the average uncertainty of observing an event which belongs to a given probability distribution. It is expressed as the statistical expectation of the minus-log probability of events in some distribution. Cross entropy, on the other hand, is expressed as the negative sum of the probabilities of events within some distribution P times the log probabilities of those events within a comparison distribution Q . The uncertainty associated with any event e

in Q is weighted by its probability in P with the effect that cross entropy $H(P, Q)$ increases as P and Q diverge. The equation is presented as Eq. 1 below.

$$H(P, Q) = - \sum_e P(x) \log_2 Q(x) \quad (1)$$

In the present context, Q represents the probabilities as estimated on the basis of the trained model, P represents the probability distribution observed within the test sample, and x represents a target event (i.e., the occurrence of a particular word type). $P(x)$ and $Q(x)$ represent the probabilities of event x as defined within the respective distributions. As mentioned earlier, $H(P, Q)$ increases as the shape of probability distribution Q diverges from that of P . High cross-entropy values thus reflect a poor fit of Q as a model of P . Conversely, $H(P, Q)$ approaches $H(P)$ as the probabilities of events within Q approximate the associated probabilities within P .

To calculate IG, we need only subtract the bigram cross entropy from the unigram cross entropy (the models must be trained and tested on the same texts). The resulting value reflects the compression rate in bits attributable to knowledge of the immediately preceding word. IG is positive by definition (unigram models necessarily have higher entropies than the corresponding bigram models).

IG can be used to compare the contribution of bigrams to models of different languages, a task which is not entirely straightforward. For instance, languages might differ in their baseline unigram entropies (and bigram entropies for that matter) for any number of reasons (e.g., as a function of morphological complexity of words). IG counters this difficulty by relativizing the performance of the bigram model to that of the unigram baseline.

Another useful feature of IG is that it can be applied to relatively small text samples, where all comparison models are trained on samples of the same size. This property is critical considering that the RI corpus published by Gil and colleagues contains only 2,700 words, well below what is typically needed to draw reliable statistical inferences.

One might argue that bigram models trained on such small samples will always perform remarkably poorly. However, the goal of this analysis is not to create an optimal model, but to test to what extent bigram knowledge improves classification in RI and English. And yet the magnitude of this increase in informativity naturally depends on model performance. Therefore, comparing IG estimates for different samples requires that we take careful steps to ensure the comparability of our samples and interpretability of our results. First, we must match all samples for size (larger samples will *ceteris paribus* yield more accurate models). Second, we must establish a common baseline against which the IG increase can be assessed. One way to obtain a common baseline is to derive the corresponding minimal informativity distribution for

each text (i.e., by randomizing the texts). IG can then be computed for the randomized and non-randomized versions for each sample, and scaling up, for each language. Comparing the IG of a sample against its corresponding minimal (randomized) IG tells us how much information is contained in the bigram word order of our samples above and beyond what would be contributed by chance for that same sample. In this way, IG values observed for different samples can be anchored to a common lower bound, and so are rendered comparable in magnitude.

Following Gil's analysis, we should expect that the IG value for an RI text should approximate that of the corresponding randomized text. Furthermore, the difference between the IG values for randomized and non-randomized samples should be smaller for RI than for English, if only because English is assumed to have a much less fluid lexicon and much more rigid constraints on word ordering. We dub this set of predictions the *Variable Order Hypothesis* to reflect the fact that word order in RI – irrespective of functional interpretation – allows for more variability than English. If no such difference were to be found between RI and English, then we can conclude that the apparent syntactic structure of RI introduced but rejected by Gil (1994; 2005) may in fact play a much more serious role in structuring RI speech. We dub this alternative the *Regular Order Hypothesis* to indicate that under this scenario, words in RI tend to be placed into regular patterns of syntagmatic association.

Experiment: IG in RI and English

We test the above hypotheses with respect to surface (unstemmed) n-grams in RI and English. We apply the logic of IG as formulated above to size-matched samples of RI and English transcribed speech. To further compensate for the size of the corpora under investigation, we introduce a bootstrapping procedure designed to maximize the reliability of small-sample IG estimates. We then model the resulting IG estimates as a function of language (RI vs. English) and text type (original vs. randomized) using linear mixed-effects regression.

Data

The data for RI were taken from Gil's corpus as published in four text files on the Max Plank Institute for Evolutionary Anthropology Jakarta Field Station website². Each file contains a stretch of transcribed naturalistic face-to-face interactions, including the telling of stories and extended riddles. The transcriptions have been broken into lines, and each line is annotated for the following set of representational dimensions: orthography, phonetic form, morphological parse, and morpheme-by-morpheme English gloss. Based on the orthographic tier, the RI corpus comprises 2,727 words (unintelligible speech, marked with sequences of 'x', was removed prior to the analysis).

² http://lingweb.eva.mpg.de/jakarta/data_PKN.php: data accessed 7/11/2014.

All four recordings were taken from the same native speaker, making the corpus less than ideal for purposes of generalization. However, given the strength of Gil's claims, it is not clear how the idiosyncratic behavior of a single individual could result in the appearance of bigram structure when the underlying grammar remains as free and underspecified as the system described above.

For the purposes of this analysis, the four individual files were combined into a single text. This means that the corpus is multiply discontinuous as there are three unauthentic junctures. This heterogeneity of content could be problematic for the generation of n -gram models. For instance, it might be that the training sample (randomly selected) comprises text from three of the four text files, while the test sample comes from the fourth. Differences in topic, register, and genre, among others, might lead such a model to underperform in the classification of the unseen text given the over- or under-representation of certain distributionally biased unigrams or bigrams. As it turns out, however, these issues are not all that serious for this particular corpus considering the goals of this study. First, all four of the texts are more or less closely related in terms of register (informal speech) and genre (story telling vs. riddles). Therefore, we should not expect the files to differ substantially in their basic structural properties because of these discourse-level features. Second, as mentioned above, the four files were all collected from the same speaker. While undoubtedly harmful for generalization to the whole population, this fact actually works to solve the problem of discontinuity. Because we are not dealing with different speakers, the concatenation of these four files can be considered a proxy for the evolution of topics across extended conversation. Finally, even if the discontinuities present in the collapsed corpus did lead to artificially high cross entropies for the individual models, this would only bias the results in favor of the monocategorical analysis, and so serve to increase the conservativeness of the estimates.

In addition to the RI data, we extracted 19 English comparison samples. All English data were sampled from the Santa Barbara Corpus of Spoken American English (Du Bois, Chafe, Meyer, Thompson, Englebretson, & Martey, 2000-2005), which contains approximately 249,000 words of transcribed spoken English from 60 different naturalistic face-to-face interactions. Each sample was matched to length of the RI corpus. Full files with word counts between 2500 and 3000 words were taken whole. There were twelve such files. The remaining seven samples consist of the first 2750 words of seven files (respectively), which were selected randomly from among those having length greater than 3000 words.

The reasoning behind selecting exactly 19 comparison samples is that a difference in behavior of the one RI text out of twenty total texts would mirror the 5% α -threshold commonly adopted for significance testing. In this way, we guard against the possibility that non-representative idiosyncrasies of individual English files could lead to spurious results. This is especially important, given that the

English samples represent diverse genres, ranging from intimate conversations to evangelical sermons. They therefore vary along a number of dimensions, including formality, mono-/dialogicality, and the total number of speakers and/or other justified conversational participants.

Procedure

We generated IG estimates for randomized and non-randomized versions of each of the 20 samples. This required several steps. First, we created a randomized version of each sample by scrambling the words contained therein at the level of the complete text. In so doing, we eliminated (a) any non-random, semantically motivated proximity effects and (b) any principled ordering of words/constituents. This process left us with a total of 20 randomized texts (19 English, 1 RI) and 20 non-randomized texts (19 English, 1 RI).

Next, we estimated the IG values for all 40 samples (random and non-random). In order to maximize the representativeness of the IG value, we calculated ten cross entropies for each sample by using ten-fold leave-one-out cross-validation. All models were trained using the n -gram model estimator from the Natural Language Toolkit (NLTK) module for Python (cf. Bird, Klein, & Loper, 2009) with Lidstone probability smoothing ($\gamma = 0.2$) to account for the existence of unigram/bigram types not observed in the sample. This process yielded 400 unigram cross entropy estimates (10 estimates per 40 samples; 200 random and 200 original). We then repeated these steps, substituting bigram models (with unigram backoffs), to generate a total of 800 n -gram cross entropy estimates. Following the procedure outlined above, we subtracted each bigram cross entropy from its associated unigram cross entropy, creating a total of 400 IG estimates (380 English, 20 RI). These estimates serve as the dependent variable in the regression analysis, discussed below.

Finally we computed the type-token ratios for each sample. We calculated these values on the basis of surface forms (i.e., we did not apply any stemming in my estimation of the number of types) by dividing the size of the unique set of lexical forms by the overall size of the sample.

Results

To evaluate the hypotheses laid out above, we computed a linear mixed effects model with IG estimate as dependent variable, language (RI vs. English), text type (original or randomized), and type-token ratio as fixed effects, and filename as a random effect. In order to test the Variable Order Hypothesis, language and text type were allowed to interact as an additional fixed effect. All possible effects were included in the initial model. Non-significant contributors to the model were removed through a hierarchical backward elimination of factors: the predictor with the largest non-significant p -value ($\alpha = 0.05$) was eliminated so long as it did not participate in any higher order interaction. This process continued until only

significant predictors or predictors participating in significant higher-order effects remained.

Only type-token ratio ($F(1, 18) = 20.81; p < .0001$) and text type (original vs. random; $F(1, 378.68) = 2739.73; p < .0001$) emerged as significant predictors. Importantly, no significant interaction was observed between language and text type. Further, no significant main effect was observed for language after removing the interaction of language and text type.

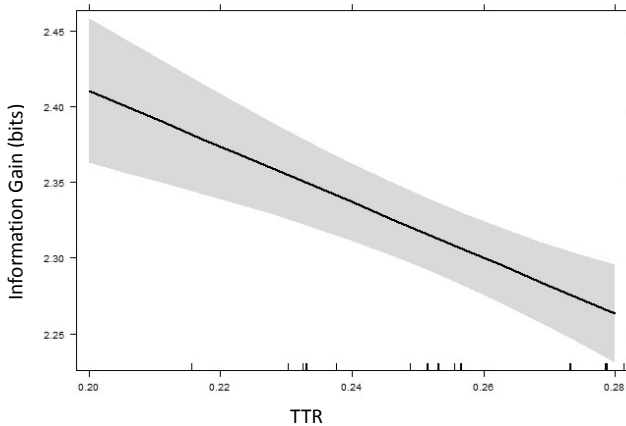


Figure 1: Main effect of type-token ratio (TTR) on information gain (measured in bits)

As shown in Figure 1, TTR correlated negatively with IG ($\beta = -1.84, s.e. = .4$). Therefore, as expected, models based on more lexically diverse samples (i.e., samples with higher TTR values) tended to benefit less from knowledge of the previous word. This term was only included in the model as a control, and so will not be considered further.

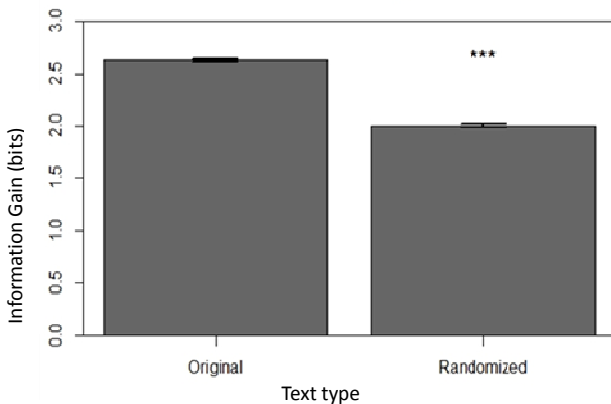


Figure 2: Main effect of text type (was the text randomized or not) on information gain values

Above and beyond the effect of type-token ratio, the regression returned a main effect of text randomization ($\beta = -.62, s.e. = .01$). Figure 2 shows that, overall, IG scores from properly ordered (*original*) texts are approximately 0.7 bits

higher than those from randomized texts. Crucially, this relationship holds for both English and RI.

Discussion

The analysis did not reveal any interaction between language and text type. Therefore, the Variable Order Hypothesis was not supported: RI and English exhibit equivalent preferences for ordering of word pairs. Notice that, given our design, this interaction between language and text randomization should have been the only way for language to impact IG scores in this model. The reasoning is this: randomization should have the same effect on all samples (so long as sample size and TTR have been controlled for). A main effect of language, however, would mean that both non-randomized and randomized IG values were consistently higher for one of the two language groups. Such a finding would suggest either that our measure is sensitive to properties of the samples beyond those included in our design or that our randomization strategy left unequal residues of non-random sequencing within the samples.

We observed a significant effect of text type, as predicted by the Regular Order Hypothesis. The advantage in IG boosting effects found for bigram models of non-randomized texts in English can be attributed to generalized syntactic constraints. Should we impute the same source to the effect observed for RI? While the answer to this question must for several reasons remain tentative, the fact that bigram sequencing was identically informative requires explanation. Gil's monocategorial analysis could either attribute this effect to the small class of particles described above or else rely exclusively on semantic and pragmatic regularity. The former case would only offer a compelling counter if the internal speciation of the particle class was on par with that of English word classes generally (i.e., to admit the existence of highly differentiated word classes in RI). The latter case would suggest that preferences on the combination of certain semantic types within certain pragmatic contexts translate into particular sequencing of the words instantiating those types. This contingency between meaning and form would be indistinguishable from that posited of the fundamental syntactic constituent of Construction Grammar (and other usage-based theories), namely the *construction* (see Goldberg, 1995). In either case, explaining these results would push the monocategorial analysis ever closer to the more traditional syntactic accounts of word order.

Conclusions

Previous studies have argued that RI lacks generalized word classes or syntactic constraints on word order. This analysis constitutes an important departure from our generally accepted understanding of the human language faculty. The present study showed that if this analysis is correct, then it is of little practical relevance to RI speech. An examination of surface-derived bigram models of RI and English speech revealed that these languages show an *indistinguishable* benefit (IG increase) over unigram models trained on and

applied to identical samples. English word order is largely syntactically determined, raising the question of whether this similarity in surface structure can be attributed to the same (syntactic) source for RI. On the one hand, we see that RI behaves *as if* it were syntactically constrained (in comparison with English), and we know that closely related varieties of Indonesian and Malay exhibit something like ‘traditional’ syntax (Sneddon, 1996). On the other hand, we know that word order is co-determined by several other factors, including prosodic, segmental, semantic, and pragmatic variables. Further evidence is needed to tease this network of effects apart. However, these factors have themselves been implicated in the grammaticalization process whereby syntactic forms emerge out of local contingency of form and meaning. It is therefore unclear whether these additional factors can ever be fully disentangled from syntax in the diachrony of a language, much less in the synchrony. Indeed, many frameworks argue for these parameters being directly encoded in syntactic representations (e.g., Goldberg, 1995).

Whatever the explanation, the degree of structure observed here suggests the presence of pervasive functional grouping in RI. Depending on the quantity and distinctiveness of these groups, and of course their diachronic depth, RI might present a case of intermediate grammaticalization of syntactic structures, whereby probabilistic biases have begun to crystallize into constituent order constraints. On the other hand, this apparent structure might simply be the product of simplification via widespread second language acquisition (McWhorter, 2006). In this scenario, lower level structural features like the co-occurrence probabilities examined here might have persisted while overt morphosyntax wore away. In either case, Gil’s monocategorial analysis does not appear capable of eluding the specter of grammar.

On another level, this finding provides a more reasonable solution to the underspecification problem faced by speakers of a language like RI. Mired in ambiguity, they appear to have adopted methods to structure their utterances in consistent, conventionalized ways. As more richly annotated corpora become available, we can begin to examine whether these conventionalized orderings correlate with conventionalized interpretations. Based on Gil (2005), we should expect that, indeed, particular orderings impose at least the tendency for particular interpretations.

Methodologically, this study has demonstrated that even small corpora of understudied languages can provide for relatively rich analyses of probabilistic structure. Particularly, the relativization vis-à-vis (1) the English to Riau sample ratio and (2) the unigram-bigram differential ensured that comparisons could be established on the basis of a normalized baseline. Without this additional layer, it would be difficult to determine what magnitude of IG score should be treated as indicating a meaningful difference in amount of structure.

References

- Arnold, J. E., Wasow, T., Losongco, A., and Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76, 28-55.
- Baayen, R. H., Milin, P., Filipović-Đurđević, D., Hendrix, P. & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118, 438-482.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. Sebastopol: O’Reilly.
- Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., and Martey, N. (2000-2005). *Santa Barbara corpus of spoken American English, Parts 1-4*. Philadelphia: Linguistic Data Consortium.
- Gil, D. (1994). The structure of Riau Indonesian. *Nordic Journal of Linguistics*, 17, 179-200.
- Gil, D. (2002). The prefixes *di-* and *N-* in Malay/Indonesian dialects. In F. Wouk & M. Ross (Eds.), *The history and typology of Western Austronesian voice systems* (241-283). Canberra: Pacific Linguistics.
- Gil, D. (2005). Word order without syntactic categories: How Riau Indonesian does it. In A. Carnie, H. Harley, and S. A. Dooley (Eds.), *Verb first: On the syntax of verb-initial languages* (243-263). Amsterdam: John Benjamins.
- Gil, D. (2008). How Much Grammar Does It Take to Sail a Boat. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (123-130). Oxford University Press, Oxford.
- Gil, D. (2013). Riau Indonesian: A language without nouns and verbs. In J. Rijkhoff & E. van Lier (Eds.), *Flexible word classes: Typological studies of underspecified parts of speech* (89-130). Oxford: Oxford University Press.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Hengeveld, K. (2013). Parts-of-speech system as a basic typological determinant. In J. Rijkhoff & E. van Lier (Eds.), *Flexible word classes: Typological studies of underspecified parts of speech* (31-52). Oxford: Oxford University Press.
- McWhorter, J. (2006). *Language interrupted: Signs of non-native acquisition in standard language grammars*. New York: Oxford University Press.
- Rijkhoff, J. & van Lier, E. (Eds.). (2013). *Flexible word classes: A typological study of underspecified parts-of-speech*. Oxford: Oxford University Press.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 28(3), 379-423.
- Sneddon, J. N. (1996). *Indonesian: A comprehensive grammar*. New York: Routledge.