Testing Processing Explanations of Word Order Universals

Michael Hahn (Stanford), Richard Futrell (UCI)
mhahn2@stanford.edu

A prominent class of language universals are the *harmonic word order correlations* (Greenberg, 1963): E.g., whether a language puts verbs before or after objects strongly determines whether it places auxiliaries before or after verbs, whether it has prepositions or postpositions, etc. A prominent explanation is in terms of processing efficiency (Hawkins, 1990): Grammars of languages are viewed as solutions to the problem of optimizing for processing efficiency measures that are valid across languages. We computationally evaluate such explanations.

Approach We specify a simple probabilistic model of grammar as it relates to word order, where grammars have parameters that can be optimized for metrics of processing efficiency. Applying these grammars to tree structures found in dependency corpora of 51 languages (Figure 1), we optimize the grammar parameters for three different metrics of processing efficiency. We test whether the resulting grammars exhibit the harmonic order correlations.

Word Order Models We specify simple parametric word order grammars that transduce unordered dependency trees into strings of words (Figure 1, cf. Gildea & Jaeger, 2015). In our model, the position of a word in relation to its head and to its siblings is determined by the the label of the syntactic relation, taken from the largely language-independent inventory of Universal Dependencies (UD, Nivre et al., 2017). The grammar is parameterized by distance and direction parameters for each dependency. We say a word order grammar exhibits a word order correlation between two dependencies if the direction parameters for the dependencies have the same sign. **Processing Metrics** We consider three processing metrics, each of which have been argued to be optimized by word orders found in natural language: (1) *Minimizing the length of syntactic dependencies* (Rijkhoff, 1986; Hawkins, 1990; Futrell et al., 2015); (2) *minimizing the surprisal* of each word in context (Hale, 2001; Gildea & Jaeger 2015); and (3) *maximizing the accuracy of syntactic parsing* (Hawkins, 1990). We additionally consider the sum of (2) and (3). We implement (2) and (3) using neural-network-based methods that underlie state-of-the-art NLP systems (Jozefowicz et al., 2016; Dozat et al., 2017). Word order grammars are optimized for each of the three metrics by taking tree topologies from a dependency corpus and applying stochastic gradient descent to optimize the average value of the metric over linearized trees (Figure 1). **Setup** We use all UD 2.1 languages for which at least one training set was available, a total of 51 languages. For each UD language, we created optimized grammars for each of the three criteria, repeating 8 times to control for variance in optimization.

Results Which of the Greenbergian correlations are reproduced by the optimized grammars? We base our evaluation on the list of correlations in Dryer (1992). Discarding correlations that could not be formalized in UD and merging some that could only be formalized together, we obtained 10 formalized correlations in UD (Table 1). Eight of these were supported by the orders found in the majority of UD corpora. We evaluated which of the correlations were instantiated in significantly more than 50% of the optimized languages for each processing metric, with random effects for languages and language families. Eight correlations were predicted by Dependency Length Minimization. Surprisal predicted four, parsability six of the correlations. Metrics (1) and (3) largely made overlapping predictions, and (2) made complementary predictions. All but two correlations were predicted by at least one of the metrics. **Conclusion** Our results (Table 1) show that a large subset of the Greenbergian word order correlations, to the extent that they can be formalized in UD, are predicted by minimizing dependency length, minimizing surprisal, and/or maximizing parsability. This supports processing-based explanations of word order universals.

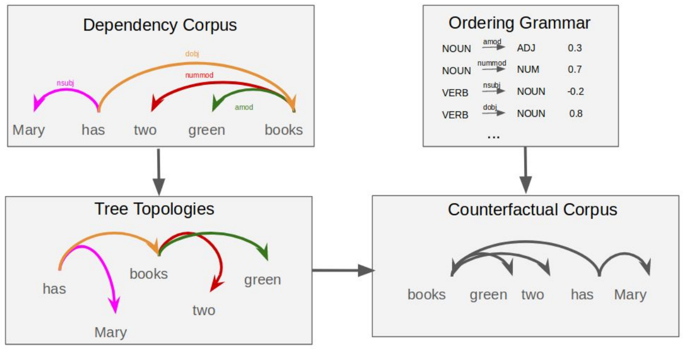

Figure 1: Applying ordering grammars to tree topologies from dependency corpora, we obtain counterfactual corpora. We optimize the weights of ordering grammars to optimize for processing metrics as computed on the obtained counterfactual corpora.

Correlates with... verb	object	Operationalization	Real	Dep.L.	Surp.	Pars.	Surp.+Pars.
adposition	NP	case	86	**81*****	**40****	**77*****	**67*****
copula	NP	cop	94	**80*****	59	**72*****	**63****
auxiliary	VP	aux	88	**74*****	**81*****	51	**65****
noun	genitive	nmod	80	**82*****	49	**71*****	**68*****
noun	relative clause	acl	80	**85*****	43	**82*****	**71*****
complementizer	S	mark	76	**85*****	**57***	**76*****	**72*****
verb	PP	obl	88	**78*****	**71*****	46	**68*****
want	VP	xcomp	88	**90*****	**76*****	**89*****	**85*****
verb	subject	nsubj	33	**29*****	53	**7*****	**23*****
verb	manner adverb	advmod	35	51	**18*****	47	**39****

Significance levels: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

Table 1. Greenbergian Correlations: In Dryer (1992), all correlations are relative to the order of verbs and objects. Thus, each correlation is stated in terms of a pair of a *'verb patterner'* and an *'object patterner'* (the *'Correlates with...'* column), whose relative order correlates with that of verbs and objects. We organize the correlations by the category of the 'verb patterner'. For each correlation, we give our *operationalization* in terms of UD. We then report how many (in %) of the UD languages satisfied it (*'Real'*). We then report, for each correlation and each processing metric, how many (in %) of the optimized grammars satisfy the correlation, with the two-sided significance level in a logistic mixed-effects analysis across languages and language families.

References

Dozat et al. 2017. *Stanford's Graph-based Neural Dependency Parser at the CoNLL 2017 Shared Task*. Proceedings of the CoNLL 2017 Shared Task. Dryer. 1992. *The Greenbergian Word Order Correlations*. Language. Futrell et al. 2015. *Large-scale evidence of dependency length minimization in 37 languages*. PNAS. Gildea and Jaeger. 2015. *Human languages order information efficiently*. arXiv:1510.02823 Greenberg. 1963. *Some universals of grammar with particular reference to the order of meaningful elements.* Universals of language. Hale. 2001. *A probabilistic Earley parser as a psycholinguistic model.* NAACL. Hawkins. 1990. *A Parsing Theory of Word Order Universals.* Linguistic Inquiry. Jozefowicz et al. 2016. *Exploring the Limits of Language Modeling*. arXiv:1602.02410 Nivre et al. 2017 *Universal Dependencies 2.1*.