Comparing efficiency-based theories of adjective order in dependency treebanks

Richard Futrell, William Dyer, and Gregory Scontras

We take up the scientific question of what determines the preferred order of English adjectives in phrases such as *big blue box* where multiple adjectives modify a following noun. We implement and test four quantitative theories, all of which are theoretically motivated in terms of efficiency in human language production and comprehension. We evaluate theories based on their ability to predict orders of unseen adjectives in hand-parsed and automatically-parsed dependency treebanks of naturalistic speech. We find that all theories provide strong predictors, with suggestive evidence for a two-factor account relying on both semantic information and information about common collocations.

Theories of adjective order. We test four theories: (1) **Subjectivity** (Hetzron, 1978; Scontras et al., 2017) states that more subjective adjectives appear closer to a noun. (2) **Information locality** (Futrell & Levy, 2017; Kirby et al., 2018) claims that words with high *pointwise mutual information* should appear closer to a noun. (3) **Integration cost** (Dyer, 2017) holds that adjectives that greatly lower the entropy over nouns should appear close to the noun. (4) **Information gain** formalizes Danks and Glucksberg's (1971) intuition that adjectives with high "discriminating potential" should appear close to their nouns; we formulate this notion mathematically in terms of the reduction in entropy over nouns caused by partitioning nouns into those which can and cannot be modified by a certain adjective.

Data. Adjective subjectivity is estimated using an extension of Scontras et al.'s (2017) experimental methodology, implemented as a questionnaire on Amazon.com's Mechanical Turk (N=264). We use dependency treebanks to estimate information-theoretic quantities (pointwise mutual information, entropy, and information gain), and to test how well each of our theories can explain adjective order. We use two corpora: (1) the hand-parsed English Web Treebank, part of the Universal Dependencies project; and (2) an automatically-parsed very large subset of the English Common Crawl corpus (Buck et al., 2014). The Common Crawl corpus gives us access to an unprecedented quantity of naturally-occurring multi-adjective strings (N=842,714).

Method. We fit logistic regressions to predict adjective order given our predictors. From the dependency corpora, we collect *AN pairs* (pairs of adjectives and nouns) to estimate our information-theoretic predictors; we then collect *AAN triples* (triples of two adjectives modifying a following noun) to test how well we can predict the order of the two adjectives given our predictors. We estimate our information-theoretic predictors based on maximum likelihood estimates from counts of wordforms.

Results. We find that subjectivity and information locality are the best predictors of word order in our large automatically-parsed corpus; we are unable to distinguish among the predictors in the smaller hand-parsed data. Accuracies for individual predictors are shown in Figure 1 below. Furthermore, we find tentative evidence (see Table 1) for a two-factor account of adjective order, with subjectivity reflecting a semantic factor, and pointwise mutual information reflecting a factor based on a preference to put words in common collocations close to each other. It appears that any theory of adjective ordering will need to

take into account both types of factors. Having demonstrated the viability of this analysis pipeline, the task turns next to cross-linguistic extensions.

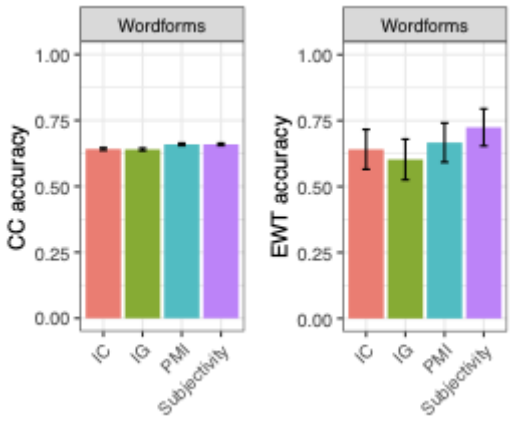

Figure 1. Accuracies of predictors of adjective order in AAN triples drawn from two dependency treebanks. We test predictors on a different set of AAN triples from those used to fit the logistic regressions (Common Crawl (CC) test set size *N*=41,822; English Web Treebank (EWT) *N*=155).

A_1	A_2	N
major	bad	behaviors
large	outstanding	debts
classical	logical	fallacy
dark	ulterior	motives
minor	fine	tuning

(a) Ordered correctly by wordform PMI, but not by wordform subjectivity.

Table 1. AAN triples ordered correctly by some predictors, but not others. Those ordered correctly by PMI (information locality) are often collocations, such as "outstanding debts" and "ulterior motives".

References

C. Buck, K. Heafield, and B. Van Ooyen. 2014. N-gram counts and language models from the common crawl. In *LREC*, volume 2, page 4.

J. H. Danks and S. Glucksberg. 1971. Psychological scaling of adjective orders. *Journal of Verbal Learning and Verbal Behavior*, 10(1): 63–67.

W. E. Dyer. 2017. *Minimizing integration cost: A general theory of constituent order*. Ph.D. thesis, University of California, Davis, Davis, CA.

R. Futrell and R. Levy. 2017. Noisy-context surprisal as a human sentence processing cost model. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 688–698, Valencia, Spain.

R. Hetzron. 1978. On the relative order of adjectives. In H. Seller, editor, *Language Universals*. Narr, Tübingen, Germany.

S. Kirby, J. Culbertson, and M. Schouwstra. 2018. The origins of word order universals: Evidence from corpus statistics and silent gesture. In *The Evolution of Language: Proceedings of the 12th International Conference (Evolangxii)*. NCU Press.

G. Scontras, J. Degen, and N. D. Goodman. 2017. Subjectivity predicts adjective ordering preferences. *Open Mind: Discoveries in Cognitive Science*, 1(1): 53–65.