Adjective Ordering: a Compositional Distributional Semantics Perspective

Introduction: the problem of adjective ordering

Neutral uses of multiple stacked attributive adjectives in English follow an order dictated by the semantic properties of the various classes of adjectives (1) (Quirk, Greenbaum, Leech, and Svartvik 1972; Dixon 1982). Alternative orders are possible when one of the adjectives is focused, when the two orders yield different meanings (*former famous singer* vs. *famous former singer*) or when the adjectives are conjoined (possibly asyndetically, but with a characteristic intonation pattern). According to Kingsbury and Wellman (1986), the order of the various categories is (2). Scott (2002) proposes an even more detailed classification.

(1) A{small triangular golden / ??triangular small golden / *golden triangular small / ??small golden triangular} jewel

(2) Det > Subjective-Comment > Size > Age > Shape > Color > Nationality/Origin > Material > Compound-Element > Noun

Sproat and Shih (1990) and Scott (2002) argue that the N modifiers that cannot be reduced to relative clauses show the same ordering restrictions in other unrelated languages (e.g. Mandarin, Finnish, Polish, etc.), suggesting a universal status. Given the current attempt to minimize the role of UG (Hauser, Chomsky, and Fitch 2002; Chomsky 2004; Chomsky 2007), it would be highly desirable to reduce this ordering to independently motivated cognitive principles; it has been proposed, for instance, that closeness to the nouns corresponds to more concrete or more perceptually salient properties (Dixon 1982). Yet, this is hard to apply to many of the classes above.

Our approach: testing adjective ordering with distributional semantics

Here we report on preliminary work in which we use compositional *distributional semantics* (DS) (Dumais and Landauer 1997; Lund and Burgess 1996; Erk 2012) to evaluate the semantic difference, if any, between multiple orders of Adj pairs; specifically we want to find out (i) if the dispreferred order can be discriminated from its meaning alone, (ii) if it is 'worse' along pre-established metrics.

DS uses words' cooccurence contexts in a large corpus to build a *vector representation* (VR) which captures the meaning of linguistic expressions as a function of the context in which they are used (Firth 1957). VRs have been found to be able to express word and sentence similarity scores which agree with human on-line and off-line measures (Griffiths, Steyvers, and Tenenbaum 2007; Turney and Pantel 2010). Compositional distributional semantics takes word VRs as primitives and tries to combine them compositionally to obtain VRs of more complex expressions (Baroni, Bernardi, and Zamparelli 2014). Specifically Baroni and Zamparelli (2010) model attributive adjective meanings as linear functions from noun VRs (VR(N)) to VR(Adj+N)) (thus, *red* is a matrix which, applied to e.g. VR(*car*) gives VR(*red car*); this is equivalent to the <<et><et>> analysis in denotational systems). This process can be applied recursively. In our study we replace Baroni&Zamparelli's linear functions with a neural network NN (one for each attributive adjective in the test), trained on <N,Adj+N> pairs. Note that the NNs will not see stacked cases during training. We will then apply two networks in sequence both in the acceptable and non-acceptable order (after checking that the latter is also unattested in a large corpus). So, $\mathrm{NN}_{large}(\mathrm{NN}_{red}(\mathrm{VS}(car)))$ vs $\mathrm{NN}_{red}(\mathrm{NN}_{large}(\mathrm{VS}(car)))$. Since NNs are not linear, a difference outcome can be expected. In cases when this obtains, we will use classifiers to see if the two orders can be distinguished from their output alone. If so, we will compare the outcomes in terms of the measures of semantic deviance uncovered in Vecchi, Marelli, Zamparelli, and Baroni (2017) (who apply the linear-functional method to sensical and nonsensical Adj+N pairs: *yellow water* vs. ??*democratic water*), to see if the unacceptable order in the classes above gives a small but perceivable push toward nonsensicality, in one of its measures (e.g. distance from the original noun vector).

References

Baroni, M., R. Bernardi, and R. Zamparelli (2014). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies 9*(6), 5–110.

Baroni, M. and R. Zamparelli (2010). Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of EMNLP*, Boston, MA, pp. 1183–1193.

Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and beyond*, pp. 104–131. Oxford: Oxford University Press.

Chomsky, N. (2007). Approaching UG from below. In U. Sauerland and H. M. Gärtner (Eds.), *Interfaces + Recursion = Language?*, pp. 1–29. New York: Mouton de Gruyter.

Dixon, R. M. W. (1982). *Where have all the adjectives gone? and other essays in semantics and syntax*. Berlin: Mouton.

Dumais, S. and T. Landauer (1997). A solution to Platos problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological review 104*, 211–240.

Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass 6*(10), 635–653.

Firth, J. R. (1957). *Papers in linguistics, 1934-1951*. Oxford, UK: Oxford University Press.

Griffiths, T., M. Steyvers, and J. Tenenbaum (2007). Topics in semantic representation. *Psychological Review 114*, 211–244.

Hauser, M. D., N. Chomsky, and W. T. Fitch (2002, November). The faculty of language: What is it, who has it, and how did it evolve? *Science 298*, 1569–1579.

Kingsbury, R. and G. Wellman (1986). *Longman Advanced English*. Essex: Longman.

Lund, K. and C. Burgess (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods 28*, 203–208.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1972). *A grammar of contemporary English*. London: Longman.

Scott, G.-J. (2002). Stacked adjectival modification and the structure of nominal phrases. In G. Cinque (Ed.), *Functional Structure in DP and IP. The Carthography of Syntactic Structures*, Volume 1. Oxford University Press.

Sproat, R. and C. Shih (1990). The cross-linguistics distribution of adjective ordering restrictions. In C. Georgopoulos and I. R. (Eds.), *Interdisciplinary approaches to language: essays in honor of Yuki Kuroda*, pp. 565–593. Dordrecht: Kluver.

Turney, P. and P. Pantel (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research 37*, 141–188.

Vecchi, E., M. Marelli, R. Zamparelli, and M. Baroni (2017). Spicy adjectives and nominal donkeys: Capturing semantic deviance using compositionality in distributional spaces. *Cognitive Science 41*(1), 102–136.