

Subjectivity predicts adjective ordering preferences

Gregory Scontras, Judith Degen, Noah D. Goodman

February 5, 2016

Regularities in the behavior of speakers and speech communities provide a window onto the psychology of language. Here we take up one such regularity: adjective ordering. Speakers and listeners exhibit strong ordering preferences when two or more adjectives are used to modify a noun, as in “the big blue box” or “the good smooth purple plastic chair.” Deviate from the preferred order, and the construction becomes odd. Something feels particularly unwieldy about “the blue big box,” even more so with “the plastic good purple smooth chair.” Why do most strings of adjectives have tightly-constrained order? We investigate the role of adjective meaning, specifically the subjectivity of the properties that the adjectives name, in predicting ordering preferences.

Adjective ordering preferences stand as a particularly striking case of regularity in language. More remarkable than their robustness in English is their cross-linguistic systematicity: we continually find the *same* preferences across the world’s languages. Hungarian (Uralic), Telugu (Dravidian), Mandarin Chinese, and Dutch are just a handful of languages with pre-nominal adjectives (i.e., languages where adjectives precede nouns) reported to have the same ordering preferences as English (Martin, 1969b; Hetzron, 1978; Dixon, 1982; Sproat and Shih, 1991; LaPolla and Huang, 2004). In languages like Selepet (Papuan) and Mokilese (Micronesian) with post-nominal adjectives (i.e., where adjectives follow nouns), these preferences are preserved in the reverse (Hetzron, 1978; Dixon, 1982; Sproat and Shih, 1991)—stable preferences determine the linear distance of an adjective from the noun it modifies.

There have been two general approaches to the investigation of adjective ordering preferences. As part of a larger project mapping the syntax and semantics of adjectives (a “cartographic” approach, one could say), the linguistics literature advances a universal hierarchy of semantic classes of adjectives. Leading the charge, Dixon (1982) set out to uncover language-internal structure by which to organize ordering preferences. The preferences were assumed to be hard-coded in the grammar; the researcher’s job was simply to uncover them. Building on the ordering of semantic classes proposed by Dixon, Cinque (1994) advanced a fully syntactic account of ordering preferences under which different classes of adjectives populate dedicated syntactic categories which inhabit specialized projections in the syntactic tree. For example, color adjectives project a Color Phrase, shape adjectives project a Shape Phrase. The Shape Phrase syntactically dominates the Color Phrase; with left-branching structure, hierarchical dominance results in linear precedence. The ultimate source of this rigid structure was immaterial; at issue was a comprehensive and deterministic account of the facts (see Scott, 2002 for a similar proposal).

Before the grammatical approaches, which map, as it were, the terrain of adjective *structure*, psychological approaches advanced the idea that aspects of adjectives’ *meaning* determine their relative order. The trouble lies in deciding precisely which aspects of meaning are relevant. Kicking off the enterprise in 1898, Sweet proposed that adjectives which are more closely connected with the noun in meaning occur closer to the noun, and that adjectives with a more specialized meaning occur closer to the noun. Similarly, Whorf (1945) proposed that adjectives describing more “inherent” properties occur closer to the noun. Ziff (1960) proposed that adjectives with less context-dependent meaning occur closer to the noun, and that adjectives that felicitously describe a narrower set of nouns occur closer to the noun. Recent compositional approaches have argued that the fundamental factor in predicting adjective ordering is whether or not an adjective forms a new concept with the noun it modifies (McNally and Boleda, 2004; Svenonius, 2008): first you form the concepts (e.g., “wild rice” or “bad apple”), then you modify them (e.g., “Minnesotan wild rice”). These proposals and others like them circle around similar aspects of adjective meaning in their account of ordering preferences; unfortunately, operationalizing metrics like meaning distance, specificity, inherence, and context-dependence is not a trivial task (but see the attempt in Martin, 1969a). Lacking clear empirical measures of the relevant aspects of adjective meaning, these psychological accounts gave way to the more descriptive, grammatical ones that settle for innate syntax as the ultimate arbiter.

We revisit the idea that ordering preferences emerge from aspects of adjective meaning, attempting to provide more thorough empirical grounding to these notions; from the grammatical approach we adopt the strategy of using semantic classes of adjectives to structure our investigation and smooth our data. Distilling the psychological proposals that precede us into a single feature, we advance the hypothesis that it is the *subjectivity* of the property named that determines ordering preferences (Hetzron, 1978; Quirk et al., 1985; Hill, 2012). Less subjective adjectives are reliably more useful at communicating the speaker’s intended message; the chance of miscommunication decreases with decreasing subjectivity. Perhaps for this reason, less subjective adjectives occur closer to the substantive head of the nominal projection, that is, to the modified noun? In “the big blue box,” judgments about bigness are likely less consistent than judgments about blueness; “blue” is less subjective than “big,” and so, according to this theory, it occurs closer to the noun “box.”

To test the hypothesis that adjective subjectivity predicts ordering preferences, we created and validated empirical measures of the ordering preferences themselves and of an adjective’s subjectivity. With reliable estimates of both, we then evaluated the predictive power of subjectivity in adjective ordering preferences. Next, we investigated the role of specific noun information in the determination of ordering preferences. Finally, we extended our empirical coverage to a larger set of adjectives and nouns.

1 Experiment 1: Establishing the measures

1.1 Ordering preferences

We began by measuring preferences in adjective ordering. We selected a sample of 26 relatively frequent, imageable adjectives from seven different semantic classes discussed in the

Table 1: Adjectives and their semantic classes tested in all experiments, and nouns and their classes tested in the behavioral experiments.

Adjective	Class	Adjective	Class	Noun	Class
old	age	good	value	apple	food
new	age	bad	value	banana	food
rotten	age	round	shape	carrot	food
fresh	age	square	shape	cheese	food
red	color	big	dimension	tomato	food
yellow	color	small	dimension	chair	furniture
green	color	huge	dimension	couch	furniture
blue	color	tiny	dimension	fan	furniture
purple	color	short	dimension	TV	furniture
brown	color	long	dimension	desk	furniture
wooden	material	smooth	physical		
plastic	material	hard	physical		
metal	material	soft	physical		

grammatical literature (dimension, value, age, physical, shape, color, material). We then elicited naturalness judgments on adjective-adjective-noun object descriptions, permuting the relative order of the adjectives.

Participants. We recruited 50 participants through Amazon.com’s Mechanical Turk crowd-sourcing service. Participants were paid \$0.30 for their participation.

Design and methods. Participants were asked to indicate which of two descriptions of an object sounded more natural. Each description featured a noun modified by two adjectives, for example “the red small chair” or “the small red chair”. Descriptions were random combinations of two adjectives and a noun from the list in Table 1, with the constraint that no description contained adjectives from the same semantic class. Description pairs contained the same words, with relative adjective order reversed. On each trial, participants indicated their choice by adjusting a slider with endpoints labeled with the competing descriptions; an example trial appears in Fig. 1. Participants completed 26 trials. On each trial, we measured the distance of the slider from each endpoint; values ranged between 0 and 1. Only native speakers of English with IP addresses located within the United States were included in the analyses; we analyzed data from 45 participants.

Results. For each adjective, we computed its mean naturalness score by averaging ratings of configurations in which it appeared in first position, farthest from the noun. Fig. 2 (*naturalness*) plots these mean naturalness scores by adjective class; greater values signal that a class’s adjectives are preferred in first position, while lower values indicate that a class’s adjectives are preferred in second position, closer to the noun. This preferred distance

Progress:

Which description of the chair sounds more natural?

"the small red chair" "the red small chair"

Adjust the slider to indicate your preference.

Continue

Figure 1: Example trial from *Expt. 1.1 Ordering preferences*; participants indicated the more natural of two adjective-adjective-noun descriptions on a sliding scale.

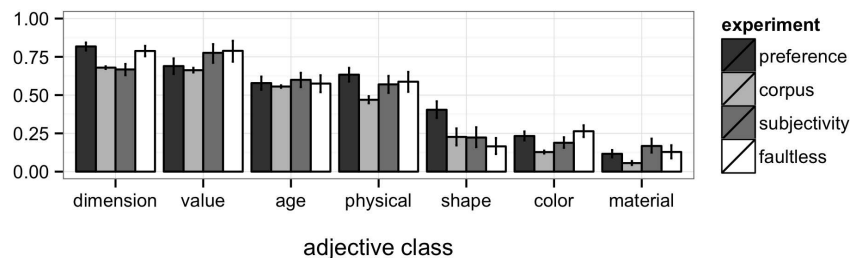


Figure 2: Mean distance from noun inferred from naturalness ratings (*preference*), mean distance from noun calculated from corpus counts (*corpus*), mean subjectivity ratings (*subjectivity*), and mean faultless disagreement ratings (*faultless*) for adjectives grouped by their semantic class. Error bars represent bootstrapped 95% confidence intervals drawn from 10,000 samples of the data (DiCiccio and Efron, 1996).

measure closely tracks class-level ordering hierarchies reported in the literature (Dixon, 1982; Sproat and Shih, 1991).

Corpus validation. To validate our behavioral measure of ordering preferences, we conducted a corpus study on the same 26 adjectives and measured their mean distance from the noun in phrases with two adjectives. We used TGrep2 (Rohde, 2005) and the TGrep2 Database Tools (Degen and Jaeger, 2011) to extract all “A A N” NPs that contained one of the 26 adjectives in Table 1 from the Penn Treebank subset of the Switchboard corpus of telephone dialogues (Godfrey et al., 1992), as well as from the spoken and the written portions of the British National Corpus (BNC, see <http://www.natcorp.ox.ac.uk/>). For these cases, we computed the distance of each occurrence of our 26 target adjectives from the modified noun, yielding results for a total of 38,418 adjective tokens. For each adjective, mean distance from the noun was computed (where the position directly preceding the noun was coded as 0, and the position preceding that was coded as 1). The procedure described above was used to compute mean distance by adjective class.

Mean distance from the noun for each adjective class is shown in Fig. 2 (*corpus*); higher values indicate that adjectives from the specified class tend to appear farther from the mod-

ified noun. The corpus measure closely tracks the qualitative pattern we measured in our naturalness experiment; quantitatively, the two measures are highly correlated ($r^2 = 0.83$, 95% CI [0.63, 0.90]), in spite of the fact that the corpus measure includes cases from a superset of the nouns tested in our naturalness experiment. Our naturalness ratings thus operationalize both immediate ordering preferences and speakers’ preferences in natural usage.

Looking for noun effects. We next asked to what extent ordering preferences depend on the modified noun. Analyses of adjective semantics hold that adjectives that form idiomatic concepts or natural kinds (e.g., “bad apple” or “blue cheese”) compose first with nouns, before run-of-the-mill intersective adjectives, thus enforcing a strict ordering (McNally and Boleda, 2004; Svenonius, 2008); we might therefore expect to find that adjectives that compose with a noun to yield an idiomatic concept are preferred linearly closer to the modified noun. To evaluate the role of specific noun information in determining ordering preferences, we performed a nested linear model comparison. The models predicted naturalness ratings either by ADJECTIVE (i.e., the adjective farthest from the noun) only, or by ADJECTIVE together with its interaction with NOUN (i.e., the modified noun). The model comparison revealed that noun-specific ratings did not explain any additional variance in ordering preference above and beyond adjective-level ratings ($F(1, 234) = 1.10, p < 0.15$). Thus, we fail to find evidence of noun-specific effects on ordering preferences in our materials. However, there remains the possibility that our materials (i.e., nouns and adjectives) did not yield sufficiently idiomatic object descriptions that would deliver these noun-specific effects. We follow up on this possibility in Experiment 2.

1.2 Subjectivity

With clear estimates of ordering preferences, we then measured the subjectivity of the adjectives that were tested in the ordering preferences experiment. We started with a direct measure: participants were asked explicitly to rate the “subjectivity” of our adjectives.

Participants. We recruited 30 participants through Amazon.com’s Mechanical Turk crowdsourcing service. Participants were paid \$0.30 for their participation.

Design and methods. Participants were shown a series of adjectives and asked to indicate how “subjective” each one was on a sliding scale with endpoints labeled as “completely objective” (coded as 0) and “completely subjective” (coded as 1; Fig. 3). Participants completed a total of 26 trials, one for each adjective in Table 1. The order was randomized for each participant. Only native English speakers with IP addresses located within the United States were included in the analyses; we analyzed data from 28 participants.

Results. We averaged the subjectivity scores for each adjective; greater values indicate greater subjectivity. These averages were used in the analyses reported below. Fig. 2 (*subjectivity*) show these scores by adjective class.

Progress:

Consider the following description:

brown cheese

How subjective is the description "brown cheese"?

completely objective completely subjective

Figure 3: Example trial from *Expt. 1.2 Subjectivity*; participants rated the subjectivity of adjectives.

Faultless disagreement validation. Because subjectivity may be an ambiguous, or even subjective, property, we explored a second measure that may have greater ecological validity. We operationalized subjectivity as the potential for faultless disagreement between two speakers, which captures potential uncertainty about assessment criteria and assessment outcomes (Kölbel, 2004; Kennedy, 2013; Barker, 2013).¹ We had participants ($n=40$) evaluate whether two speakers could both be right while the speakers produced conflicting object descriptions. For example, an experimental trial would have Mary assert, “That apple is old,” then have Bob counter with “That apple is not old;” participants rated whether both Mary and Bob could be right, or whether one of them must be wrong. This measure, the faultless disagreement potential for the adjective at issue, serves as an empirical estimate of adjective subjectivity. Fig. 2 (*faultless*) plots these scores by adjective class, where a value of 1 signals that a class’s adjectives are always amenable to faultless disagreement (i.e., maximally subjective), and a value of 0 indicates that a class’s adjectives are never amenable to faultless disagreement (i.e., minimally subjective). The results of this method were highly correlated with our direct “subjectivity” scores ($r^2 = 0.91$, 95% CI [0.86, 0.94]), suggesting that they measure a common underlying value—adjective subjectivity.

1.3 Predicting adjective order

To evaluate the power of subjectivity in predicting adjective ordering preferences, Fig. 4 plots mean naturalness ratings (Expt. 1.1) against mean adjective subjectivity scores (Expt. 1.2). Adjective subjectivity scores account for 85% of the variance in the naturalness ratings ($r^2 = 0.85$, 95% CI [0.75, 0.90]). The faultless disagreement scores also perform well, accounting for 88% of the variance ($r^2 = 0.88$, 95% CI [0.77, 0.95]). Using either measure, more subjective adjectives are preferred farther from the noun; subjectivity indeed predicts adjective ordering preferences.

One might worry that conducting our analysis at the level of individual adjectives obscures information about the specific adjective-adjective configurations that participants rated in our naturalness experiment. We therefore computed a subjectivity difference score

¹See MacFarlane (2014) for more discussion of the many factors, both “semantic” and “pragmatic,” that contribute to faultless disagreement effects. For a different approach, see Hill (2012), who builds on previous corpus work (Wulff, 2003) to infer adjective subjectivity from surface features of strings.

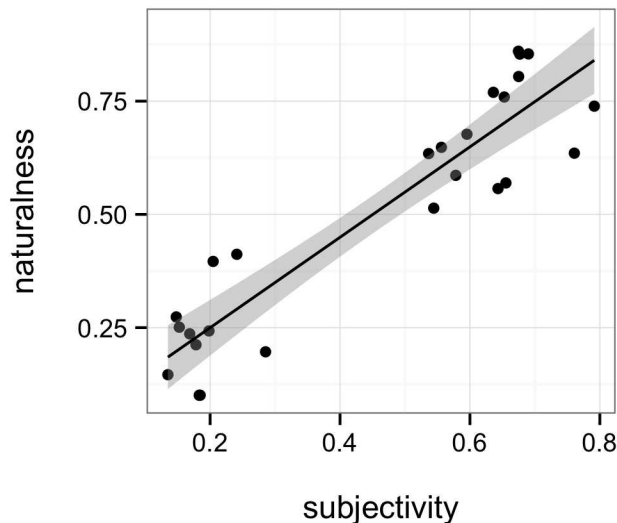


Figure 4: Mean naturalness ratings plotted against mean subjectivity scores for each of the 26 adjectives tested in Expt. 1.

for each adjective class configuration (i.e., an ordered pairing of two adjective classes, CLASS1-CLASS2) by subtracting the mean subjectivity score for CLASS2 from the mean subjectivity score for CLASS1. Higher difference scores indicate that the adjective class closer to the noun is less subjective than the class farther away. Fig. 5 plots mean naturalness ratings for adjective class configurations against these subjectivity difference scores; the two measures are highly correlated ($r^2 = 0.80$, 95% CI [0.68, 0.88]).²

1.4 Discussion

We found that adjective subjectivity scores account for a great deal of the variance in naturalness ratings, for several different analyses, strongly supporting our hypothesis that less subjective adjectives occur closer to the noun. We generalize this finding to a broader set of adjectives in Expt. 3, but first we explore in more detail whether the modified noun affects ordering preferences of the adjectives that modify it.

2 Experiment 2: In search of noun effects

Compositional accounts of ordering preferences hold that the fundamental factor in predicting adjective ordering is whether or not an adjective is used to form a complex concept/subkind description: first you form the concept, then you modify it with additional adjectives (McNally and Boleda, 2004; Svenonius, 2008).³ This would imply that an interac-

²If we analyze faultless disagreement difference scores instead, we find that they also account for significant variance in the configuration naturalness ratings ($r^2 = 0.82$, 95% CI [0.71, 0.88]).

³Bouchard (2005) makes a similar claim, namely that the formation of complex concepts can override adjective ordering preferences.

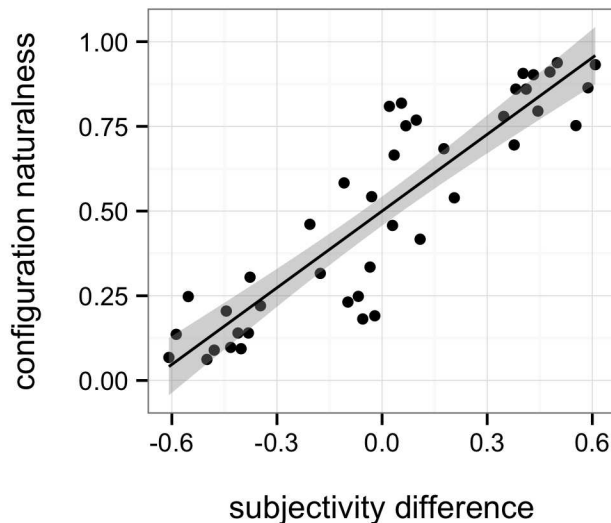


Figure 5: Mean configuration naturalness ratings plotted against subjectivity difference scores for each pair of adjective classes tested in Expt. 1.

tion between the noun and a modifying adjective—whether they combine to form a complex concept—should have a large influence on adjective ordering. Indeed, a more general hypothesis is that *some* interaction between a noun and adjective will influence how closely the adjective is placed to that noun. This interaction could be caused by concept-formation, differential subjectivity, or other factors. We tested for such an interaction in Expt. 1 and found that noun-specific naturalness did not explain any variance in ordering preference above and beyond adjective-level naturalness. Looking in more detail, there were two adjective-noun pairs in our data with trends in the predicted direction: the naturalness ratings for *hard* and *soft* suggested a preference to occur closer to the noun *cheese*. (Plausibly because hard and soft cheeses are natural kinds.) While these adjective-noun interactions do not survive correction for multiple comparisons in our statistical analysis, they do indicate that a different set of materials might reveal by-noun effects on ordering preference. Here we follow up on this result with a new set of materials that were chosen to maximize the probability of noun effects.

2.1 Ordering preferences

This experiment was a direct replication of *Expt. 1.1 Ordering preferences*, using a different set of nouns. We aimed to choose nouns that formed idiomatic, complex concepts with our previous set of adjectives. Complex concepts tend to be described using the two-word name, yielding more occurrences of this bigram than would be expected from the unigram frequencies of the noun and adjective. This provided a way to extract candidate complex concepts from corpora.

Participants. We recruited 50 participants through Amazon.com’s Mechanical Turk crowd-sourcing service. Participants were compensated for their participation.

Design and methods. The design was identical to our original naturalness ratings experiment: participants were asked to indicate which of two object descriptions sounded more natural, using a sliding scale. Each description featured a noun modified by two adjectives; description pairs contained the same words with the relative adjective order reversed (e.g., “the big blue thing” vs. “the blue big thing”). Adjectives were chosen at random from the original set of 26. The nouns were a smaller set of five (compared to the original ten). Nouns were chosen to maximize the probability of detecting noun-specific effects on adjective ordering preferences. In particular, we expected that nouns that are likely to form complex concepts should be highly collocational with that adjective. We thus searched for nouns that occur in particular adjective-noun phrases more frequently than predicted by the individual noun and adjective probabilities; in other words, nouns whose adjective-noun combinations were under-predicted by their individual word probabilities.

To find these nouns, we estimated the probability $p(A)$ of each adjective from our set of 26 by computing its relative frequency in an adjective-noun sequence in the BNC. We then computed the relative frequency of each noun $p(N)$ occurring in an adjective-noun sequence. Finally, we estimated the predicted joint probability of each adjective-noun combination by taking the product of each individual probability estimate: $\hat{p}(A, N) = p(A) \cdot p(N)$. Comparing $\hat{p}(A, N)$ to the empirically estimated $p(A, N)$ establishes which adjective-noun combinations are under-predicted—more collocational—and thus likely to name complex concepts. We then restricted nouns to those 50 that maximize the observed range of under-predictedness while simultaneously requiring that each noun be attested to occur with at least 11 of the 26 adjectives; from these 50 nouns, we selected the following four: *apple*, *cheese*, *eyes*, *hair*.⁴ (Recall that *cheese* occurred in our original materials, where it suggested possible by-noun effects with the adjectives *hard* and *soft*.) To these four nouns we added a fifth: *thing*. While *thing* did not occur in the top 50, it did occur naturalistically with the most adjectives (23) out of the set of 26, thus allowing it to serve as a filler for the various object descriptions. The selected nouns, together with the number of adjectives they occur with, their range of ratios of empirical to predicted joint probabilities, and their minimum / maximum ratios, are shown in Table 2.

Results. To evaluate the role of specific noun information in determining ordering preferences, we performed the same nested linear model comparison from our original naturalness ratings experiment. The models we compared predicted naturalness ratings either by ADJECTIVE (i.e., the adjective farthest from the noun) only, or by ADJECTIVE together with its interaction with NOUN (i.e., the modified noun). The model comparison revealed that noun-specific ratings did not explain any additional variance in ordering preference beyond adjective-level ratings ($F(1, 225) = 0.93, p < 0.75$). Thus, we again fail to find evidence of noun-specific effects on ordering preferences in our new materials.

⁴We restricted to a small subset of the 50 target nouns in order to maximize statistical power to identify noun effects on ordering.

Noun	# of adjectives	range of ratios	minimum ratio	maximum ratio
thing	23	10.4	0.1	10.5
eyes	18	120.6	0.12	120.7
hair	15	82.9	0.03	83.0
cheese	13	114.0	0.4	114.4
apple	11	674.0	1.1	675.1

Table 2: For each chosen noun, the number of adjectives (out of 26) that it occurs with; and for each adjective A that the noun occurs with, the range of ratios $p(A, N)/\hat{p}(A, N)$ (empirical to predicted probability of occurrence); the minimum ratio; and the maximum ratio.

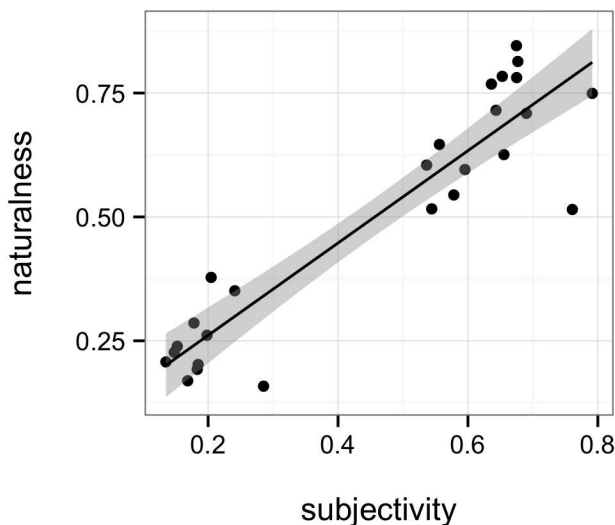


Figure 6: Mean naturalness ratings plotted against mean subjectivity scores for each of the 26 adjectives tested in Expt. 2.

Given the lack of noun effects on ordering preferences, we should continue to find that adjective subjectivity predicts ordering preferences. Indeed, it does: adjective subjectivity scores (obtained in *Expt 1.2 Subjectivity*) account for 85% of the variance in the new naturalness ratings ($r^2=0.85$, 95% CI [0.64, 0.93]; Fig. 6). As with our original materials, more subjective adjectives are preferred farther from the noun.

2.2 Discussion

We failed to find evidence in support of compositional accounts of ordering preferences, which hold that the most important factor in determining order is whether or not an adjective forms a complex concept with the noun it modifies. Using nouns chosen to maximize the probability of complex concepts formed with our 26 adjectives, we failed to find evidence that adjective ordering preferences depend on the modified noun. We do continue to find that

subjectivity predicts ordering preferences. It is quite possible that noun-effects (and hence effects of concept composition) would be found with a different set of adjectives. However, this already suggests that the effect of modified nouns explains only a small part of the overall story of adjective order; subjectivity seems to do better.

3 Experiment 3: Generalizing our findings

One might worry that the observed success of subjectivity in predicting ordering preferences is an artifact of the set of 26 adjectives we have so-far tested, and might not generalize to a broader set of adjectives. Therefore, we next consider a much larger set of adjectives. We aimed to construct a set of adjectives that are attested in multi-adjective constructions and that span both grammatical classes, and a broad spectrum of frequencies and lengths. The set of 78 adjectives we ultimately used includes many adjectives that are traditionally overlooked in investigations of ordering preferences.

3.1 Ordering preferences

Participants. We recruited 495 participants through Amazon.com’s Mechanical Turk. Participants were compensated for their participation.

Materials To arrive at a new set of adjectives we began by extracting every unique adjective appearing in an “A A N” NP from the Penn Treebank subset of the Switchboard corpus; there were 350 unique adjectives. Then, two of the authors (GS and JD) independently coded these 350 adjectives according to semantic classes from the literature (e.g., Dixon, 1982; Sproat and Shih, 1991), indicated in Table 3. We added a miscellaneous category “X” containing non-intersective adjectives (cf. Cinque, 2014). After ignoring adjectives for which there was no obvious semantic class, as well as adjectives on which our classifications did not agree, we were left with 196 unique adjectives from 13 different classes.

To further refine this list of 196, we calculated adjective frequency, and on the basis of frequency divided the adjectives within each class into four quartiles. We also calculated adjective length on the basis of number of characters, and classified adjectives as either short or long. Then, for each adjective class with more than eight members, for each frequency quartile, we sampled a random adjective from each length (e.g., one short extremely frequent dimension adjective and one long extremely frequent dimension adjective). There were three cases without any observations from which to sample: “nationality” (Q3 short and Q4 short) and “physical” (Q4 short). We filled in these cases by randomly sampling from the leftovers in those two classes. Finally, we added an additional shape adjective, “square”, because previously there was only “circular”. This process yielded the 78 unique adjectives in Table 3.

Nouns were chosen in a similar fashion. From the set of “A A N” NPs in Switchboard, we extracted the 295 unique nouns, then restricted this set by 1) removing all plural or clearly collective nouns, and 2) removing nonsense words or words that might be confused with participles. This process yielded 166 unique nouns.

Table 3: Adjectives used in Expt. 3.

Adjective	Class	Adjective	Class	Adjective	Class
junior	age	professional	human	sweet	physical
new	age	sad	human	circular	shape
old	age	selfish	human	square	shape
old-time	age	strict	human	fast	speed
senior	age	closest	location	slow	speed
young	age	internal	location	speedy	speed
black	color	overhead	location	current	temporal
blonde	color	corduroy	material	daily	temporal
blue	color	crocheted	material	everyday	temporal
green	color	gold	material	historical	temporal
purple	color	wooden	material	best	value
red	color	brazilian	nationality	exciting	value
white	color	english	nationality	favorite	value
yellow	color	european	nationality	lavish	value
biggest	dimension	hispanic	nationality	plain	value
large	dimension	international	nationality	pleasant	value
long	dimension	japanese	nationality	prestigious	value
mini	dimension	national	nationality	strange	value
narrow	dimension	vietnamese	nationality	designated	X
open	dimension	creamy	physical	different	X
thick	dimension	curly	physical	individual	X
thin	dimension	frozen	physical	last	X
civilized	human	lacy	physical	mixed	X
creative	human	smooth	physical	potential	X
entrepreneurial	human	solid	physical	token	X
playful	human	spicy	physical	unique	X

Design and methods. The design was identical to our previous naturalness rating experiments (Expts. 1.1 and 2.1): participants indicated which of two object descriptions sounded more natural, choosing between adjective-adjective-noun permutations that varied the relative order of the adjectives. Adjectives were chosen at random from the set in Table 3, with the constraint that adjectives from the same class were not paired together. Participants completed 30 trials. On each trial, participants indicated their choice by adjusting the slider between endpoints labeled with the competing descriptions. Additionally, participants were able to indicate if a particular description did not make sense by checking a box labeled “Neither option makes sense.” Only native speakers of English with IP addresses located within the United States were included in the analyses; we analyzed data from 473 participants.

Results. Participants demonstrated little preference for adjective order when the descriptions were nonsense. For this reason, we excluded responses to nonsensical descriptions from the analyses of subjectivity below; this exclusion process removed 2295 observations (16% of the total 14190). Fig. 7 (*preference*) plots mean naturalness scores by adjective class for the descriptions that made sense to participants. Greater values signal that a class’s adjectives are preferred in first position, farther from the modified noun; values around 0.5 indicate that neither order is preferred.

As with the previous ordering preference experiments, we performed a nested model comparison to investigate the role of specific noun information in determining ordering preferences. The models predicted naturalness ratings either by ADJECTIVE only, or by ADJECTIVE together with its interaction with NOUN. Once again, the model comparison revealed that noun-specific ratings did not explain any additional variance in ordering preference above and beyond adjective-level ratings ($F(1, 10100) = 0.98, p < 0.87$).

3.2 Subjectivity

Next, we evaluated the subjectivity of our new set of adjectives using the direct “subjectivity” task from Expt. 1.2.

Participants. We recruited 198 participants from Amazon.com’s Mechanical Turk. Participants were compensated for their participation.

Design and methods. The design was identical to our previous direct “subjectivity” experiment: participants rated the subjectivity of an adjective on a sliding scale with endpoints labeled “completely objective” (coded as 0) and “completely subjective” (coded as 1). Participants completed a total of 30 trials. On each trial an adjective was chosen at random from the set of 78 in Table 3. Only native speakers of English with IP addresses located within the United States were included in the analyses; we analyzed data from 189 participants.

Results. Fig. 7 (*subjectivity*) plots mean subjectivity scores by adjective class; greater values indicate that a class’s adjectives are judged to be more subjective. To evaluate the power of subjectivity in predicting adjective ordering preferences, we compared subjectivity scores

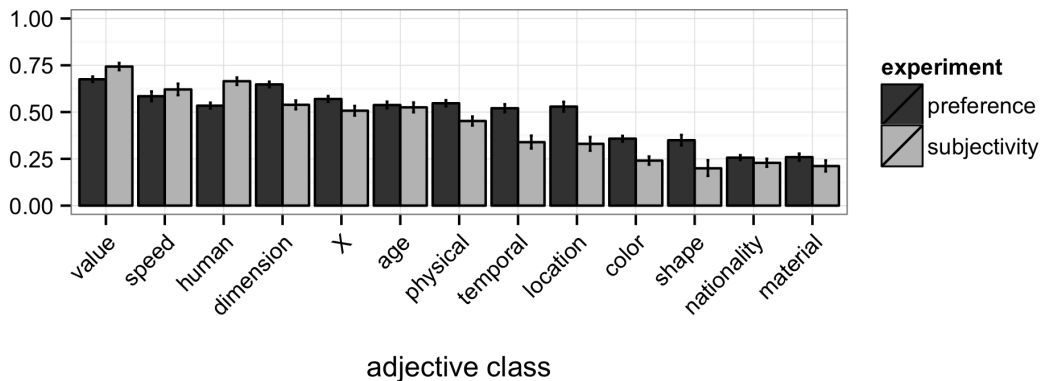


Figure 7: Mean subjectivity ratings grouped by adjective class.

with the naturalness ratings (Fig. 8). Adjective subjectivity scores account for 51% of the variance in the naturalness ratings ($r^2 = 0.51$, 95% CI [0.32, 0.66]). Four observations clearly stood out in Fig. 8, corresponding to the superlatives *best*, *biggest*, *closest*, and *last*. Indeed, superlatives have been observed to eschew adjective ordering preferences, occurring farthest from the modified noun regardless of class or subjectivity (Dixon, 1982); our naturalness ratings reflect this fact. Removing superlatives, we find that subjectivity scores perform markedly better, accounting for 61% of the variance ($r^2 = 0.61$, 95% CI [0.46, 0.71]). At the level of adjective class configurations, subjectivity difference scores account for 74% of the variance in the configuration ratings ($r^2 = 0.74$, 95% CI [0.66, 0.79]; Fig. 9).⁵

A post-hoc look at our data revealed a small number of outlier adjectives in addition to the four superlatives. To systematically detect these outlier adjectives, we fit a linear regression predicting naturalness ratings by subjectivity scores, then calculated the absolute difference between the actual naturalness ratings and the model’s predicted values. Setting the cutoff for this difference score at $3 \times$ standard deviation, four adjectives stood apart as outliers: *entrepreneurial*, *solid*, *current*, and *daily* (labelled in blue in Fig. 8). Without the four outlier adjectives (and the four superlatives), adjective subjectivity scores account for 70% of the variance in the naturalness ratings ($r^2 = 0.70$, 95% CI [0.58, 0.78]).

We also looked at the contribution of frequency and length in predicting ordering preferences. Treating subjectivity, frequency, and length as predictors in a linear regression predicting naturalness ratings (excluding superlatives), the model accounts for 70% of the variance ($r^2 = 0.70$). If we remove outlier adjectives that fall more than three standard deviations away from the predicted value of the model (there were six: *mini*, *frozen*, *solid*, *current*, *daily*, *designated*), the model performs better, accounting for 76% of the variance ($r^2 = 0.76$).

⁵This analysis and the plot in Fig. 9 exclude superlatives. If we include superlatives in the class configuration analysis, subjectivity difference scores account for 69% of the variance in the naturalness ratings ($r^2 = 0.69$, 95% CI [0.60, 0.76]).

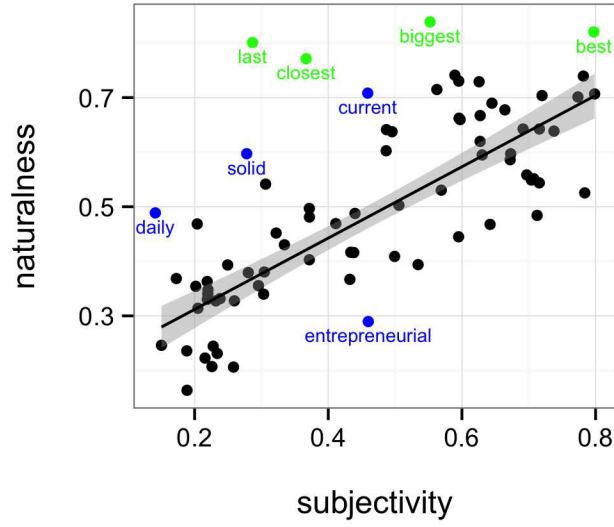


Figure 8: Mean naturalness ratings plotted against mean subjectivity scores for each of the 78 adjectives tested in Expt. 3. Superlatives are labeled in green; outlier adjectives are labeled in blue.

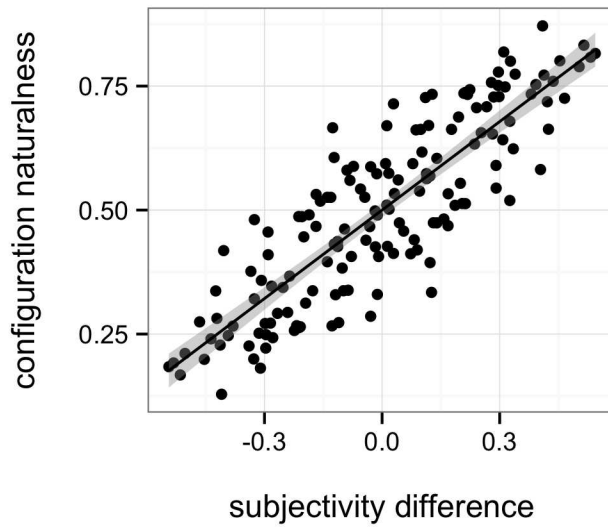


Figure 9: Mean configuration naturalness ratings plotted against subjectivity difference scores for each pair of adjective classes tested in Expt. 3.

3.3 Discussion

The results of the current experiment demonstrate that subjectivity predicts ordering preferences not just for our original set of 26 adjectives, but in a much larger set of materials drawn from naturally-occurring examples. At worst, subjectivity accounts for more than half of the variance in the naturalness ratings for our set of 78 adjectives. Once we exclude superlatives, whose semantics likely dictates their position in strings of nominal modifiers, as well as four outlier adjectives, subjectivity accounts for 70% of the variance in this set of 70 adjectives. While adjective frequency and length contribute to the observed preferences, we saw that subjectivity alone accounts for the vast majority of the variance in our data.

There remains the question of precisely *why* the four outlier adjectives—*entrepreneurial*, *solid*, *current*, and *daily*—performed so poorly with respect to the predictions of subjectivity. Perhaps the most notable feature of this set of adjectives is its heterogeneity: we fail to find clear groupings by semantic class, relative frequency, or length. However, length likely does factor into the observed behavior of *entrepreneurial*, the longest adjective tested, which was the only outlier *under*-predicted by its subjectivity: participants preferred *entrepreneurial* closer to the noun than its subjectivity alone would predict. Indeed, relative length has long been known to affect the order of constituents, even in the domain of adjective ordering (Wulff, 2003): longer constituents appear later. Once we factor length into the equation predicting ordering preferences, *entrepreneurial* no longer stands out.

4 General discussion

Adjective ordering preferences have received considerable attention throughout the history of generative grammar and cognitive psychology, owing to their remarkable stability within and across languages. Something so robust, the reasoning goes, must evidence a deep principle of the cognitive architecture that shapes language. Yet while descriptions of the phenomenon abound, an explanation has proven elusive. Grammatical theories that posit a rigid syntax of adjective classes offer little more than a codification of the facts, and psychological approaches stumble when it comes to operationalizing the specific aspects of adjective meaning at play.

In our investigation, we established two empirical constructs: the preferences themselves, which we measured using naturalness ratings and validated with corpus statistics; and adjective subjectivity, which we measured directly and corroborated with potential for faultless disagreement. An adjective’s semantics predicts its distance from the modified noun, such that less subjective adjectives occur linearly closer to nouns they modify. This preference is not deterministic; non-preferred orderings of adjectives can serve a communicative purpose, for example to establish contrastiveness in discourse (Martin, 1969a, 1970; Hill, 1958; Vendler, 1963). This contrastiveness follows straightforwardly from a manner implicature (Levinson, 2000): marked forms (i.e., non-preferred orderings of adjectives) yield marked interpretations (i.e., atypical modification constituency). The work lies in determining the preferred orderings from which contrastive uses depart. Indeed, many other situational factors are likely to influence ordering (e.g., phonological shape, noun semantics, word and bi-gram frequencies; cf. Wulff, 2003); it is the more general tendencies we are concerned with here.

Adjectives are just one of many elements that may occur in complex nominal constructions. Other classes of elements include demonstratives (e.g., *this* and *that*) and numerals. In his Universal 20, Greenberg observes that the relative order of these higher-order classes is also stable cross-linguistically (Greenberg, 1963; Culbertson and Adger, 2014), suggesting that subjectivity interacts with additional constraints from semantic composition in the determination of word order. Indeed, we saw hints of such interactions in Expt. 3, where superlatives stood apart from run-of-the-mill adjectives. Beyond nominals, adverbs (e.g., *honestly*, *probably*, *carefully*) are reported to exhibit regular orderings cross-linguistically (Cinque, 1999; Ernst, 2002). Understanding these orderings would likely benefit from a systematic empirical treatment similar to the one we have advanced here.

Our results suggest that ordering preferences likely emerge, at least partially, from a desire to place less subjective content closer to the substantive head of a nominal construction (i.e., closer to the modified noun). For now we can only speculate about the ultimate source of this desire: Subjective content allows for miscommunication to arise if speakers and listeners arrive at different judgments about a property description. Hence, less subjective content is more useful at communicating about the world. An explanation along these lines, based on pressures to facilitate successful reference resolution, would have to depend on the hierarchical, not linear, ordering of adjectives (cf. the mirroring of preferences in pre- vs. post-nominal languages). Whatever its source, the success of subjectivity in predicting adjective ordering preferences provides a compelling case where linguistic universals, the regularities we observe in adjective ordering, emerge from cognitive universals, the subjectivity of the properties that the adjectives name.

References

- Barker, C. (2013). Negotiating Taste. *Inquiry* 56(2-3), 240–257.
- Bouchard, D. (2005). Sériation des adjectifs dans le SN et formation de concepts. *Recherches linguistiques de Vincennes* 34, 125–142.
- Cinque, G. (1994). On the Evidence for Partial N-Movement in the Romance DP. In R. S. Kayne, G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, and R. Zanuttini (Eds.), *Paths Towards Universal Grammar. Studies in Honor of Richard S. Kayne*, pp. 85–110. Washington DC: Georgetown University Press.
- Cinque, G. (1999). *Adverbs and Functional Heads*. Oxford: Oxford University Press.
- Cinque, G. (2014). The semantic classification of adjectives: A view from syntax. *Studies in Chinese Linguistics* 35(1), 1–30.
- Culbertson, J. and D. Adger (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences of the United States of America* 111(16), 5842–5847.
- Degen, J. and T. F. Jaeger (2011). The TGrep2 Database Tools.

- DiCiccio, T. J. and B. Efron (1996). Bootstrap confidence intervals. *Statistical Science* 11, 189–228.
- Dixon, R. (1982). *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton.
- Ernst, T. (2002). *The Syntax of Adjuncts*. Cambridge: Cambridge University Press.
- Godfrey, J. J., E. C. Holliman, and J. McDaniel (1992). SWITCHBOARD: A Telephone Speech Corpus for Research and Development. In *Proceedings of ICASSP-92*, pp. 517–520.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Human Language*, pp. 73–113. Cambridge, MA: MIT Press.
- Hetzron, R. (1978). On the relative order of adjectives. In H. Seller (Ed.), *Language Universals*, pp. 165–184. Tübingen: Narr.
- Hill, A. A. (1958). *Introduction to linguistic structures*. New York: Harcourt, Brace.
- Hill, F. (2012). Beauty Before Age? Applying Subjectivity to Automatic English Adjective Ordering. In *Proceedings of the NAACL HLT 2012 Student Research Workshop*, pp. 11–16.
- Kennedy, C. (2013). Two Sources of Subjectivity: Qualitative Assessment and Dimensional Uncertainty. *Inquiry* 56(2-3), 258–277.
- Kölbel, M. (2004). Faultless Disagreement. *Proceedings of the Aristotelian Society* 104, 53–73.
- LaPolla, R. J. and C. Huang (2004). Adjectives in Qiang. In R. M. W. Dixon and A. Y. Aikenvald (Eds.), *Adjective Classes: A Cross-Linguistic Typology*, pp. 306–322. Oxford: Oxford University Press.
- Levinson, S. C. (2000). *Presumptive Meanings*. Cambridge, MA: MIT Press.
- MacFarlane, J. (2014). *Assessment Sensitivity*. Oxford: Clarendon Press.
- Martin, J. (1969a). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior* 8, 697–704.
- Martin, J. E. (1969b). Some Competence-Process Relationships in Noun Phrases with Prenominal and Postnominal Adjectives. *Journal of Verbal Learning and Verbal Behavior* 8, 471–480.
- Martin, J. E. (1970). Adjective Order and Juncture. *Journal of Verbal Learning and Verbal Behavior* 9, 379–383.
- McNally, L. and G. Boleda (2004). Relational adjectives as properties of kinds. *Empirical Issues in Formal Syntax and Semantics* 5, 179–196.

- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985). *A Comprehensive Grammar of the English Language*.
- Rohde, D. L. T. (2005). *TGrep2 User Manual*, Version 1.15.
- Scott, G.-J. (2002). Stacked adjectival modification and the structure of nominal phrases. In G. Cinque (Ed.), *The cartography of syntactic structures, Volume 1: Functional structure in the DP and IP*, pp. 91–120. Oxford: Oxford University Press.
- Sproat, R. and C. Shih (1991). The cross-linguistic distribution of adjective ordering restrictions. In C. Georgopoulos and R. Ishihara (Eds.), *Interdisciplinary approaches to language: Essays in honor of S.-Y. Kuroda*, pp. 565–593. Dordrecht: Kluwer Academic Publishers.
- Svenonius, P. (2008). The position of adjectives and other phrasal modifiers in the decomposition of DP. In L. McNally and C. Kennedy (Eds.), *Adjectives and Adverbs: Syntax, Semantics, and Discourse*, pp. 16–42. Oxford: Oxford University Press.
- Sweet, H. (1898). *A New English Grammar, Logical and Historical*. London: Clarendon Press.
- Vendler, Z. (1963). The transformational grammar of English adjectives.
- Whorf, B. L. (1945). Grammatical Categories. *Language* 21(1), 1–11.
- Wulff, S. (2003). A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8(2), 245–282.
- Ziff, P. (1960). *Semantic Analysis*. Ithaca, NY: Cornell University Press.