Subjectivity-based adjective ordering maximizes communicative success

Michael Franke, Gregory Scontras, Michael Simonič

adjective ordering preferences

pre-nominal	post-nominal
big brown **bag**	**bag** brown big
English	Arabic
Hungarian	Indonesian
Turkish	Basque
Hindi	Selepet
Telugu	Mokilese
Tagalog	Farsi
⋮	⋮

robustly attested in a host of unrelated languages, both with pre- and post-nominal adjectives

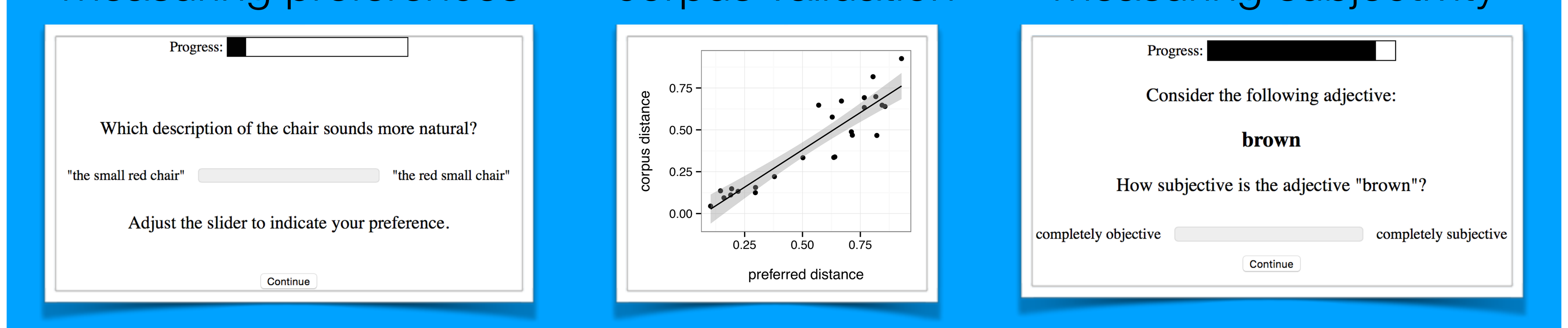
but why?

subjectivity predicts adjective ordering preferences

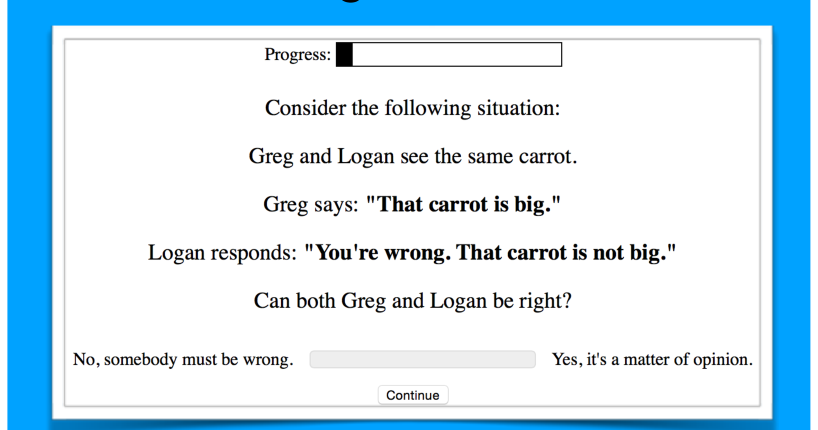
measuring preferences

corpus validation

measuring subjectivity



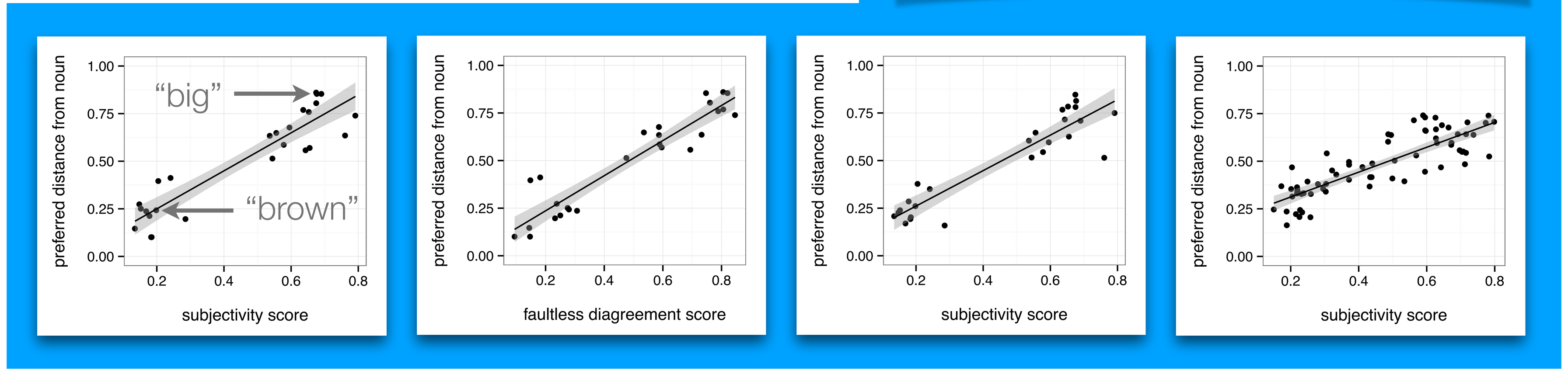
faultless disagreement validation



Scontras et al. (2017) showed adjective **subjectivity** to be a **robust predictor** of ordering preferences in English

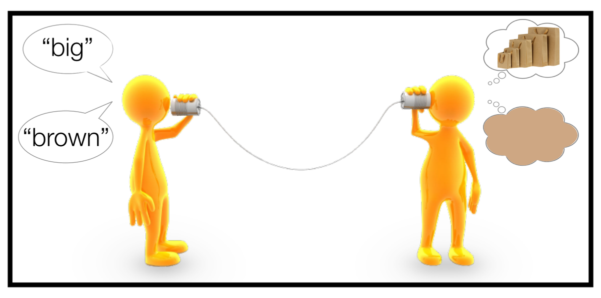
big brown bag

subjectivity decreases

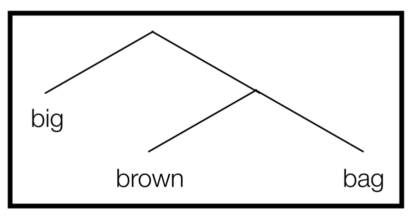


why subjectivity?

Less subjective adjectives are **more useful** for establishing reference; speakers prefer to **consolidate the more useful material around the the modified noun**



Composition tracks the hierarchical, not linear, **structure** of multi-adjective strings; **adjectives closer to the noun compose earlier** semantically



Simonič 2018 offers a Rational Speech Act model of reference resolution with multi-adjective strings

$$P(L^{L,C}_{adj} \mid L^{S,C}_{adj}) \propto \begin{cases} 1 & if\, L^{S,C}_{adj} = L^{L,C}_{adj} \\ \varepsilon_{adj} & \text{otherwise} \end{cases}$$

The speaker assume that the listener might have a different lexical meaning for each adjective. If $L^{S,C}$ is the speaker's lexical entry in context C, the speaker believes that the listener has lexical entry $L^{L,C}$ as in the box above

Scontras et al. to appear treat adjective subjectivity as potential noise in the semantics of an adjective, similar to Simonič

$$[\![\text{ADJ}]\!]^C = \lambda x \in C.\ \text{if } \texttt{ADJ}(x) \text{ then } \texttt{flip}(1-\varepsilon_{adj}), \text{ else } \texttt{flip}(\varepsilon_{adj})$$

Based on a ground-truth of objective adjective meaning, each agent (speaker or listener) will incorrectly classify each potential referent in the current context C with an error rate ε_{adj}, which is assumed to increase with the size of C

semantic assumptions

Schmidt et al. 2009 model adjective semantics using **relative height by range**: any object that falls within the top k% of the range of heights in C counts as tall in C

$$[\![\text{tall}]\!]^C = \{o \in C \mid \texttt{tall}(o) \geq \texttt{max} - \theta \cdot (\texttt{max} - \texttt{min})\}, \text{ where } \theta = k/100.$$

$$[\![\text{adj}_\text{i}\ [\text{adj}_\text{j}\ N]]\!]^C = [\![\text{adj}_i]\!]^{[\![\text{adj}_j]\!]^{C \cap [\![N]\!]}}$$

Following Simonič and Scontras et al., we assume that iterated adjectival modification triggers **sequentially intersective context updates**

motivating example

Subjective properties are more likely to lead to **deviation between the ground truth** (i.e., the true context) **and an agent's representation**

By deviating from the ground truth, more subjective properties are also more likely to lead to **deviations between two agent representations**

These **deviations and our awareness of their potential** are what **lead to perceived subjectivity** as measured by a faultless disagreement task

If the speaker wants to describe a bag that is big and brown **according to her subjective representation** of the context, **would it be better to describe it as "big brown bag" or "brown big bag"** if the listener **interprets either phrase from their own subjective perspective**?

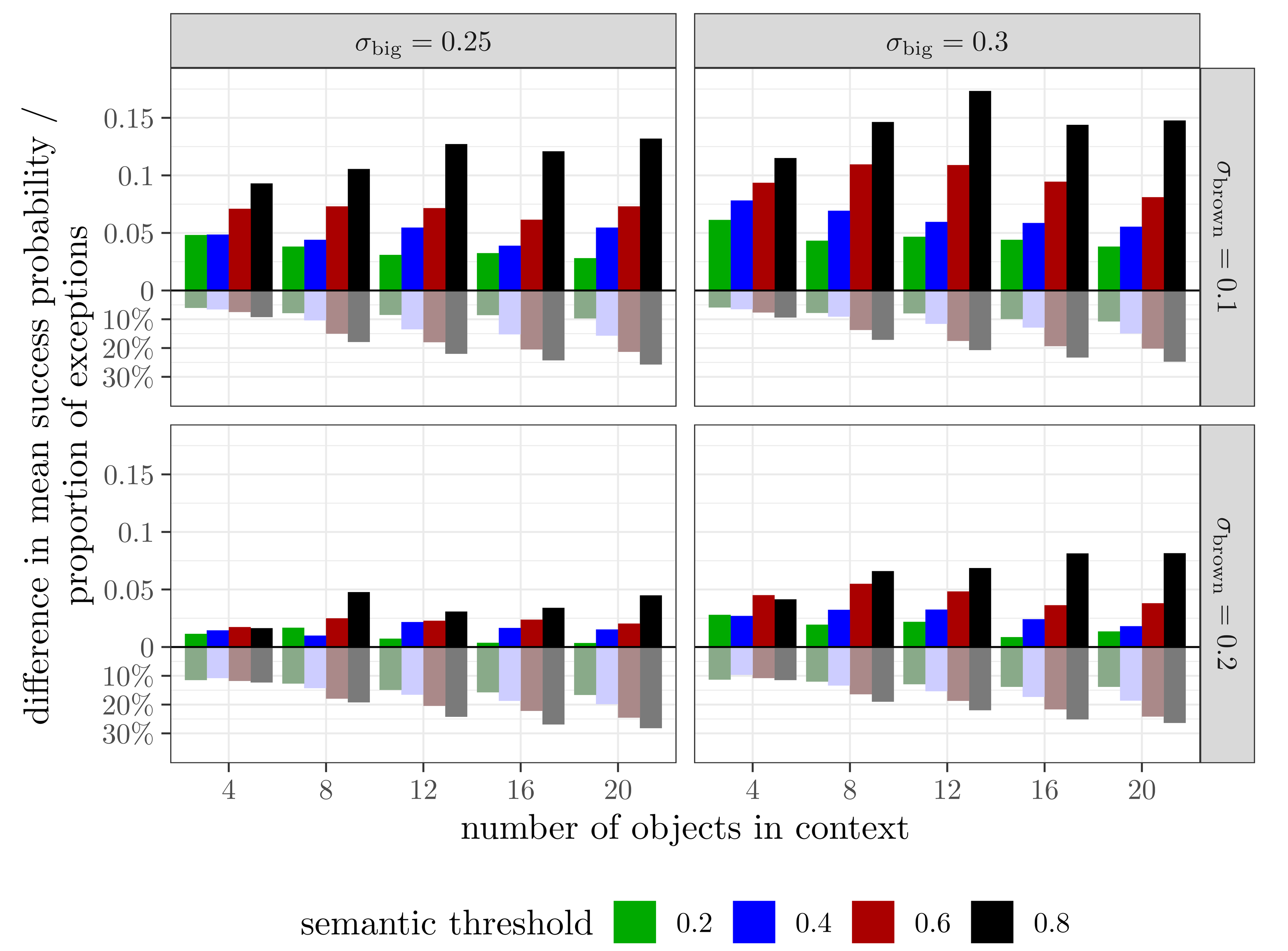
computing average communicative success

We use a **Monte Carlo simulation** to estimate the **difference in expected referential success** between phrases "big brown bag" and "brown big bag"; we calculate this value by **averaging over many different contexts** with different numbers of objects and varying degrees of subjectivity for the properties involved

A single run of the Monte Carlo simulation proceeds as follows:

1. We first sample a number n of bags in the current context uniformly at random from 4 to 20.
2. We then sample the degree to which each object is brown and the degree to which it is big. Samples are independent draws from a standard normal distribution. This yields a representation of the *actual context* C as an $n \times 2$ matrix of feature values for the n objects. The probability of sampling context C for fixed n is
$$P(C \mid n) = \prod_{i=1}^{n} \prod_{j=1}^{2} \mathcal{N}(C_{ij} \mid \mu = 0, \sigma = 1).$$
3. Agent X's (speaker's or listener's) subjective representation C^X of C is derived from C by assuming normally distributed noise around the property degrees in C, with a fixed standard deviation for each adjective. The probability of obtaining a subjective representation C^X from true C is
$$P(C^X \mid C) = \prod_{i=1}^{n} \prod_{j=1}^{2} \mathcal{N}(C^X_{ij} \mid \mu = C_{ij}, \sigma = \sigma_j).$$
The standard deviations $\sigma_{1,2}$ are obtained by sampling two numbers uniformly from the interval $[0;0.5]$ and assigning the higher number to the more subjective ("tall") and the lower to the less subjective adjective ("brown").
4. A *semantic threshold* θ is sampled uniformly at random from the unit interval. We apply the context-dependent threshold semantics in (3) from Schmidt et al. (2009) with the incrementally intersective context update in (4), using each agent's context representation, to yield each agent's subjective interpretation of each referential phrase.
5. We then sample the *speaker-intended referent object* i^* randomly from the set $[\![\text{adj}_1]\!]^{C^S} \cap [\![\text{adj}_2]\!]^{C^S}$ (i.e., an object that is both brown and big from the point of view of the speaker). If there is no such object, the run is discarded.
6. If the listener's interpretation of the phrase "[adj$_\text{i}$ [adj$_\text{j}$]]" from his subjective point of view is $I = [\![[\text{adj}_\text{i}\ [\text{adj}_\text{j}]]]\!]^{C^L}$, the probability of recovering the intended referent is $|I|^{-1}$ if $i^* \in I$ and 0 otherwise. We record the probability of recovery for both adjective orders and evaluate their distribution over all samples obtained in this way.

results



References: