Informational goals constrain how listeners infer comparison classes

Anonymous CogSci submission

Abstract

Most Great Danes are big dogs, but some are also big Great Danes. Understanding a gradable adjective (e.g., "big") requires making reference to a *comparison class*, a set of objects or entities against which the referent is implicitly compared (e.g., big for a Great Dane), but how do listeners decide upon a comparison class? Standard theories of semantic composition stipulate that the adjective combines with a noun, which necessarily becomes the comparison class (e.g., "That Great Dane is big" means big for a Great Dane). We investigate an alternative hypothesis: the noun in a sentence is a cue to the comparison class, which must be integrated with other cues, like syntax, for a listener to infer the intended comparison class. We theorize that the utility of a noun in an adjectival utterance can be either for reference (getting the listener to attend to the right object) or predication (describing a property of the referent; in the case of adjective understanding, this amounts to setting the comparison class). Therefore, we hypothesize that when nouns are in the subject position ("That NP is big"), they are more likely to be used for reference (e.g., via composition with the deictic 'That') and hence, less likely to set the comparison class; in contrast, nouns in the predicate position ("That's a big NP") are less likely to be used for reference and more likely to set the comparison class. Across three pre-registered experiments, we find evidence that listeners integrate the noun of a sentence with syntactic information and world knowledge to infer comparison classes, consistent with a trade-off between reference and predication.

Keywords: comparison class; adjectives; information structure; reference; predication

Introduction

The meanings of linguistic expressions can change dramatically depending on the context. But determining what are the relevant aspects of context that a listener should use to understand a message is far from trivial. Pointing and saying "That Great Dane is big" informs the listener than the referent (a Great Dane) has a relatively large size, but what the speaker is using as a basis of comparison (*larger relative to what?*) goes unsaid: The Great Dane could be *big for a Great Dane*, *big for a dog*, *big for a four-legged creature*, *big for a furry animal*, as well as an infinity of other possibilities. How do human listeners determine the intended comparison class when faced with many *a priori* reasonable options?

Standard theories of semantic composition posit that adjectives combine with nouns to produce an interpretable adjectival phrase (e.g., BIG(CAR), SMALL(WATCH); ?, ?, ?). Under such an account, the noun in the sentence necessarily stipulates the comparison class (e.g., *big for a car*, *small for a watch*). Many arguments have been laid against such a simple mapping between the noun in the sentence and the comparison class (e.g., ?, ?, ?): A "big snowman" said of a snowman that a 4-year-old built probably means something like *big relative to snowmen that 4-year-olds can build*; a "rich Fortune-500 CEO" might not be *rich relative to other Fortune-500 CEOs*. Theoretical work on comparison classes has focused on how comparison classes are integrated into a compositional semantics and what representations might be preferred (?, ?, ?, ?). Yet, little is known about how human listeners decide upon a comparison class.

We examine the problem from a functional perspective—what goals are speakers trying to achieve when crafting their utterance?—and derive novel predictions about comparison class inferences via the interaction of syntactic cues, world knowledge, and the perceptual context. In particular, we take advantage of the fact that noun phrases (NPs) can be used both for reference (i.e., getting the listener to attend the object that the speaker intends) and predication (i.e., describing a property of that referent) to predict interpretative differences between *attributive* uses of adjectives (e.g., "That's a big Great Dane") and *predicative* uses (e.g., "That Great Dane is big"). For attributive uses, the NP is likely to be used for predication, i.e. communicating the speaker's perspective on the referent (e.g., *I see this thing as a Great Dane*) and thus should serve as a strong cue to the comparison class in case of gradable adjectives (i.e., the Great Dane is *big for a Great Dane*). In predicative uses, on the other hand, the NP is likely to be used for reference (e.g., an NP combined with a deictic such as "That Great Dane is big"), and thus, should less strongly constrain the comparison class. When the NP provides a weak cue to the comparison class, world knowledge can more strongly influence the comparison class (e.g., the Great Dane is big for a dog ?, ?).

We tested this reference-predication trade-off hypothesis in three, pre-registered behavioral experiments with different dependent measures: sentence rating, NP production, and comparison class inference. In Experiment 1, participants provide acceptability ratings of syntactically-distinct sentences in which the NP appears in either the subject or the predicate of the sentence (e.g., "That's a big Great Dane" vs. "That Great Dane is big"). In Experiment 2, participants produce noun phrases in different syntactic frames (e.g., "That's a big __" vs. "That __ is big"). Finally, in Experiment 3, participants are asked to paraphrase a speaker's ad-

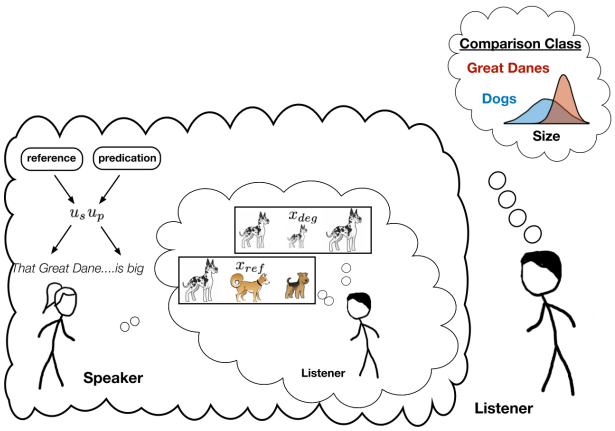

Figure 1: Pragmatic reasoning: the listener infers the comparison class by reasoning about a speaker trading off the utility of the noun phrase for reference (us) and predication (up). The speaker constructs the utterance in order for the listener to pick out the correct referent and assign the correct size to it.

jectival sentence with an explicit comparison class (e.g., "It's big relative to other __"), building on the empirical paradigm of ? (?). Across these diverse measures, we find consistent evidence for a reference-predication trade-off guiding comparison class inferences when interpreting adjectival sentences.

Behavioral Experiments

Our guiding hypothesis is that the greater the extent to which a noun phrase's usage can be explained away by its utility in reference, the less likely it will be that the noun phrase sets the comparison class. We operationalize utility in reference via the syntactic frame in which the noun phrase appears, which we examine using the subject vs. predicate contrast in where the NP appears. If the noun appears in the subject of the sentence (That NP is ADJ), it is likely to be used for reference and less likely to set the comparison class. If the noun appears in the predicate of the sentence (That's an ADJ NP), it is unlikely to be used for reference and more likely to set the comparison class. Experiment 1 tests if participants prefer one noun position in the sentence over the other to describe an object in context depending on the felicity of the noun as the comparison class. Experiment 2 tests if speakers produce different nouns in different syntactic frames to describe the same object. Finally, Experiment 3 investigates listeners' comparison class inferences as driven by the syntax, the noun phrase and the immediate perceptual context in which the sentence is uttered.

Much of the materials and the design of the three experiment are shared. The participants with US IP addresses were recruited on Amazon's Mechanical Turk and restricted to those with at least 95% work approval ratings. The preregistrations and the experimental materials can be viewed at tinyurl.com/rcsyz9f.

Materials Our experiments used the positive- and negative-form gradable adjectives describing size: *big* and *small*. In all experiments, the critical adjectival sentence(s) with the gradable adjective describe a target object presented visually alongside other objects (the *visual context*; Figure **??**).

We chose stimuli from five basic-level categories in the animal and plant domains: dogs, birds, fish, flowers, trees. Within each basic-level category, we chose target objects from subordinate level categories about which people have strong expectations concerning the size of members of those subordinate level categories (Table **??**). For example, Great Danes are generally big relative to other dogs; goldfish are generally small relative to other fish. In all experiments, targets are described using the size adjectives consistent with these general expectations (e.g,. *Great Dane-big*, *goldfish-small*). Thus, given world knowledge, both the subordinate-level or basic-level comparison class could be felicitous, though the basic-level is more likely *a priori* (i.e., the Great Dane is likely to be *big for a dog*, but could also be *big for a Great Dane*).

Experiment 1: Syntax Rating

In this experiment participants rate how well each of two sentences differing in the position of the noun phrase (NP) describe the target in context. The NP is either the basic-level or the subordinate target label (within-subjects).

Participants We recruited 113 participants and excluded 3 due to native languages other than English, 3 to failing a comprehension check and 27 to providing the same responses on every trial. The experiment took about 5 minutes.

Procedure Participants completed two comprehension check trials and six main trials. In the comprehension check trials, participants see a picture (1: a purple chair; 2: an orange basketball), read pairs of sentences describing it (1: "The chair is blue" and "The chair is yellow"; 2: "The basketball is orange" and "The basketball is green"), and are asked to rate how well each of the sentences describes the referent on a scale ranging from "very bad" to "very well", in a randomized order.

In the main trials, participants read: "You and your friend see the following:" above a basic-level context picture (e.g., a group of dogs; Figure **??**A). We created six different basic-level contexts from the basic-level categories shown in Table **??**; they depict groups of several members of the same basic-level category (i.e. dogs) belonging to different subordinate categories (i.e. dogs of several breeds, representatives of the target and filler subordinate categories, e.g. great danes, pugs and poodles). Below the context, they read "You also see this *subordinate label*" (e.g., "great dane") and see the referent pictured below. The visual size of the picture of the referent is manipulated such that it appears incongruent with the general expectations of that subordinate category

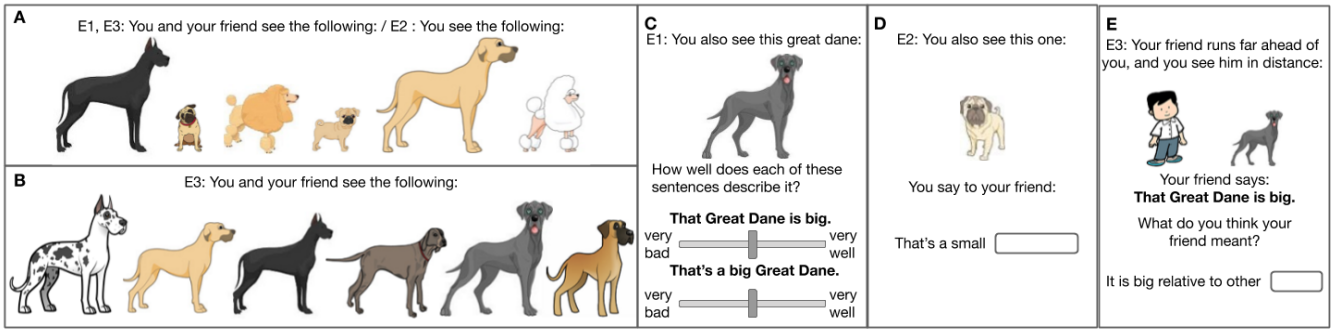

Figure 2: Overview of Experiments 1-3. A - B: Example context stimuli. A: Basic-level contexts used in Expts. 1-3. B: Subordinate context from Expt. 3. C - E: Example test questions with referents. C: Syntax Rating trial (Expt. 1) with a referent from a large-subordinate category referred to with a subordinate NP. D: NP Production trial (Expt. 2) with a referent from a small-subordinate category described with a predicate-NP syntactic frame. E: Comparison Class Inference trial (Expt. 3) with a referent from a large-subordinate category described with a subject-NP syntactic frame using a subordinate-NP label.

Table 1: Experimental items: each basic-level context has two potential targets from an either saliently small or saliently big subordinate category within the basic-level class.

Experiments	Basic-level category	Small referent	Big referent
Experiments 1, 2, 3	Dogs	Pug	Great dane
	Dogs	Chihuahua	Doberman
	Birds	Hummingbird	Eagle
	Fish	Goldfish	Swordfish
	Flowers	Dandelion	Sunflower
	Trees	Bonsai	Redwood
Experiment 2	Birds	Sparrow	Goose
	Birds	Canary	Swan
	Fish	Clownfish	Tuna
	Flowers	Daisy	Peony

(i.e., a Great Dane, which we would expect to be large, is actually small relative to other Great Danes in the context; Fig **??**C compared to context in Fig. **??**A). This subtle visual disparity was imposed in order to enhance the difference in felicity between the adjective used with a basic-level vs. subordinate level comparison class (e.g., the Great Dane is *big for a dog* but probably not *big for a Great Dane*).

Below the visual context, participants rate two sentences differing in whether the NP appears in the subject or predicate of the sentence (e.g., Predicate NP: "That's a big Great Dane"; Subject NP"That Great Dane is big"). The order in which the syntactic frames appeared on the page was randomized between-subjects. In total, participants completed six trials, which differed in whether the referent label was the subordinate category label (e.g., *Great Dane*) or the basic-level category label (e.g., *Dog*); the order of these six trials was randomized. Each participant saw only one of the two possible targets for each context (e.g. either the Great Dane or the pug for the dog basic-level context). [mht: i feel like the information about the small referent and big referent should appear here?] Participants were asked to rate how well each of the two critical sentences described the target on a scale ranging from *very bad* to *very well*.

Results We found no effect of the syntactic conditions presentation order, so the data was collapsed across the two conditions for all analyses. Figure **??** shows the mean ratings provided for subordinate and basic-level NPs in the subject and predicate-NP syntactic frames. Consistent with our prediction, a significant proportion of participants dispreferred sentences with the subordinate NP in predicate position compared to the subject position, confirmed by a Bayesian generalized linear mixed-effects model with main effects of syntax (contrast coded, subject vs. predicate NP), the noun phrase (contrast coded, basic-level vs. subordinate), and, of critical interest, their interaction. By-participant and by-item (referent) random intercepts, effects of syntax, NP and their interaction effects were included. We find an interaction between the syntax and the NP ($\beta = -4.0[-5.8, -2.3]$), as well as an overall preference for the basic-level NPs ($\beta = -5.41[-8.06, -2.82]$) and predicate syntax ($\beta = -2.66, [-4.78, -0.59]$). [mht: if you look at

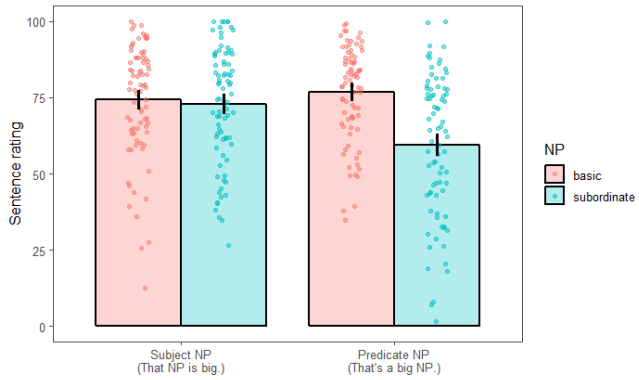

Figure 3: Experiment 1: Means and 95% bootstrapped confidence intervals of ratings of how well the sentences describe the referent when different nouns (color) appear in different syntactic frames (x-axis). Points represent participant means within condition. [mht: if participants complete both trials in red and blue bars, we can connect the dots by participant with a line]

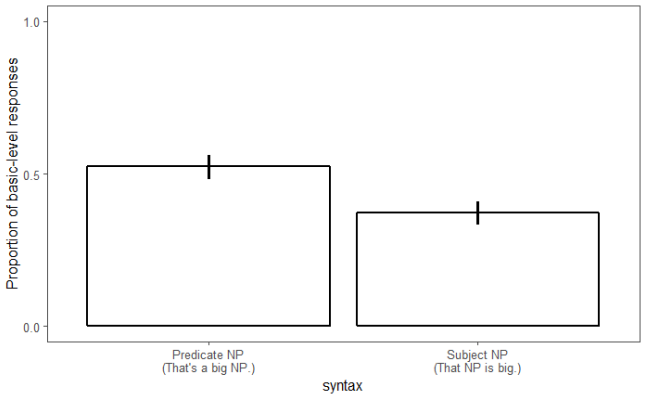

Figure 4: Experiment 2: Means and 95% bootstrapped confidence intervals of produced basic-level labels (e.g. 'dog' when the referent is a great dane) in different syntactic frames (x-axis).

the pairwise contrast, subsetting the data to only basic-NP, of predicate vs. subject, do you find a difference? I.e., are the two red bars different from each other?]

Participants discriminate the felicity of sentences involving the same NPs and same adjectives, depending on the syntactic position of the NP. This result is consistent with the hypothesis that the syntactic position of the NP modulates the strength of the cue that the NP provides towards the comparison class. 2

Experiment 2: Noun free-production

If the syntactic position of the NP modulates the NP-cue strength towards the comparison class, we would expect participants to produce different NPs depending on the syntactic position of the NP, which we test here.

Participants We recruited 242 participants, 1 was excluded for glitches, 15 for native languages other than English and 36 for failing warm-up trials more than 4 times after correction. The experiment took about 7 minutes.

Procedure

The procedure was largely similar to that of Expt. 1. One difference was that the main trials were divided into two blocks, and participants completed warm-up trials before each block to motivate them to produce noun phrases at different levels of abstraction. In the warm-up trials, participants labeled pictures of members of the two target subordinate categories within one basic-level category (e.g., a Great Dane and a pug) by filling-in three sentence frames aimed to elicit these categories. Two sentences appeared below each target reading "This is a __", and one sentence appeared at the bottom of the page reading "These are both __." If participants supplied an incorrect label, they were provided the correct label and were required to correct their response before proceeding. These referents were used as target referents in the main trials. The trial order within each warm-up and main block was randomized. We used the same contexts as in Experiment 1 and created four additional basic-level contexts (Table **??**). Six of the ten contexts were randomly sampled for each participant.

On the main trials, subjects read "You see the following:" and saw the context picture of basic-level instances (as in Expt. 1). Below, they read "You also see this one:" and saw the picture of the target referent (e.g. a great dane or a pug). Below the target picture participants were told "You say to your friend:", followed by either a subject-NP or predicate-NP sentence frame, where the NP was omitted (e.g., "That __ is big" vs. "That's a big __ "). Each participant saw only either the big or the small target for each basic-level context. We categorized the free-production responses by hand into subordinate or basic-level labels of the referent. 16 invalid responses (1.4%) were excluded from the analysis.

Results Consistent with our prediction, participants produced basic-level nouns at a higher rate in the predicate than in the subject position (Figure **??**). The prediction is confirmed by a logistic Bayesian mixed-effects regression model, predicting the response category (basic-level vs. subordinate) by an intercept, the main effect of syntax (contrast coded, subject vs. predicate NP) and by-participant and by-referent random intercepts and a by-referent random slope effect of syntax. Participants are significantly more likely to use basic-level labels in the predicate position ($\beta = 2.25[0.74, 4.01]$). Participants are more likely to produce an NP corresponding to the felicitous - basic-level - comparison class in the predicate syntactic frame.

[mht: same exploratory analyses: (1) separate out big vs. small; (2) by item effects (perhaps collapsed across big vs. small to increase power)]

Experiment 3: Comparison Class Inference

In this experiment we examined the comparison classes inferred by listeners given different contexts, NPs and syntactic frames of the critical utterance (within-subjects).

Participants We recruited 243 participants, 10 were excluded for being non-native speakers of English, 2 for failing a task comprehension check and 31 for failing warm-up trials more than 4 times after feedback. The experiment took about 9 minutes.

Procedure The procedure largely mirrored that of Expt. 2, with the main trials appearing in two blocks with warm-up trials before each. Before these blocks, participants read instructions about the task and were required to complete a paraphrase of the kind used in the main trials, wherein they rephrase what a speaker said by using an explicit comparison colass. Participant read that a speaker said "The Empire State Building is tall", were asked what they thought the speaker meant, and were given a sentence frame to complete: "The Empire State Building is tall relative to other __". As in the warm-up trials, participants had to correct their answer if they provided an infelicious comparison class (viable responses included: buildings, skypscrapers, constructions, houses). Following this comprehension test, participants completed the two blocks of warm-up and main trials, akin to the structure of Expt. 2.

In a main trial, participants read "You and your friend see the following:" above a context picture, which was either a subordinate-level or basic-level context (Fig. **??**A,B). Below the context picture, they read "Your friend runs far ahead of you, and you see him in the distance" and saw a cartoon of a person standing next to the referent (e.g., a Great Dane) in the distance; in order to create the illusion of distance, the person and the target were small compared to the items in the context pictures, such that the referent size could not be inferred visually (Fig. **??**E). Participants read "Your friend says: *critical sentence*", which could vary by both syntactic position of the NP (subject- vs. predicate-NP) as well as the NP label. The NP label could be the subordinate target label (e.g., Great Dane), basic level label (dog), or the underspecified noun "one" (e.g., "That one is big"). We used "one" in order to measure the baseline effect of visual context on comparison class inferences. After reading the sentence, participants were asked "What do you think your friend meant?", to which they responded by the sentence frame: "It is {big, small} relative to other __" with their inferred comparison class (Fig. **??**E).

Participants completed 12 trials in total, seeing one trial in each condition exactly once (syntactic frame [subject vs. predicate], visual context [subordinate vs. basic], NP label [subordinate vs. basic vs. "one"]), the order of which was randomized with one constraint. For a given basic-level set (e.g., dogs), either the big or small target appeared with the basic-level context while the other appeared with the subordinate level context (e.g., Great Dane in the context of other dogs; pug in the context of other pugs); in order to take advantage of the labeling warm-up trials, both of these trials appeared in the same block.

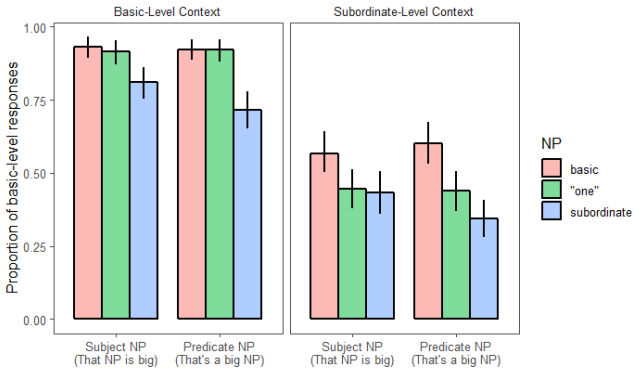

Figure 5: Experiment 3: Means and 95% bootstrapped confidence intervals of inferred basic-level comparison class proportions (e.g. "...big relative to other dogs") when the referent (e.g., a Great dane) appeared in basic-level visual context (e.g. other breeds of dogs, left panel) or in subordinate context (other great danes, right panel) from a sentence where the NP is "one", the subordinate ('Great Dane') or the basic-level ('dog') label of the referent (colors), appearing in different syntactic positions (x-axis).

Results The responses were categorized as basic-level and subordinate target labels. 39 invalid responses were excluded from the analysis. [mht: again, what percentage of total responses is this? Were there any superordinate comparison classes?]

We observed substantial flexibility in what participants decide is the most appropriate comparison class, varying by context, NP label, and syntactic frame (Figure **??**). As a manipulation check, we find the inferences drawn from the NP "one" were affected by the visual context: in the basic-level context, only the basic-level comparison class was inferred. In contrast, although the subordinate context pointed to the subordinate comparison class, both subordinate-level and basic-level comparison classes were inferred, consistent with a bias towards basic-level comparison classes.

In the basic-level visual context, the subordinate NP provided a slightly stronger cue to the comparison class (i.e., more subordinate comparison classes) when the NP appeared in the predicate than when it appeared in the subject of the sentence; a parallel asymmetry could not be observed for the basic-level label, because basic-level comparison class inferences were already at ceiling as measured by the "one"-NP condition. [mht: what stats to show?] We see a similar pattern of results in the subordinate-level visual context, wherein the subordinate NP provided a stronger cue to the comparison class when it appeared in predicate than in the subject position. This interaction is confirmed by a regression analysis that assumes no context interactions: Interestingly, we do not

observe the parallel effect with the basic-level NP (i.e., more basic-level comparison classes) in the subordinate-level context, even though the probability of a basic-level comparison class is not at ceiling (again, as indexed by the "one"-NP condition).

The basic-level NP provides a more general comparison class than the subordinate context, such that a higher proportion of inferred comparison classes was basic-level given the basic-level NP in predicate position than given the NP 'one'. Consistent with the assumption that agents reason about a trade-off between referential and predicative utility of the noun phrase, more basic-level comparison classes were inferred from the basic-level NP in subject position compared to 'one' since the NP cannot be explained be reference - the referential utility of the basic-level NP in subordinate context is low, hence the NP served comparison class communication.

Some of our predictions are confirmed using a Bayesian logistic mixed-effects regression model: the response (categorized into basic-level vs. subordinate labels of the target) was regressed against an intercept, main effects of syntax (contrast coded, subject vs. predicate NP), context (treatment coded, subordinate vs. basic-level), the NP (treatment coded, basic-level vs. 'one' and subordinate vs. 'one') and, critically, their interactions. By-participant and by-referent random intercepts and random slope effect of syntax, NP, context and their interactions were included, the correlation of random effects was set to 0 [1]. Confirming our prediction, a significantly larger proportion of basic-level comparison classes was inferred from the basic-level context ($\beta = -3.56[-4.77, -2.42]$); basic-level comparison classes were more likely to be inferred from the basic-level NPs than from 'one' ($\beta = 1.74[0.22, 3.62]$) and subordinate comparison classes from subordinate NPs compared to 'one' ($\beta = -1.05[-1.80, -0.27]$).

Discussion

References

Bale, A. C. (2011). Scales and comparison classes. *Natural Language Semantics*, *19*, 169–190.

Bierwisch, M. (1989). The semantics of gradation. *Dimensional adjectives*, *71*(261), 35.

Chalnick, A., Billman, D. (1988). Unsupervised learning of correlational structure. In *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510–516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cresswell, M. J. (1976). The semantics of degree. In *Montague grammar* (pp. 261–292). Elsevier.

Kamp, J. A. W. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal semantics of natural language.* Cambridge University Press, Cambridge, England.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy*, *30*(1), 1–45.

Newell, A., Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Solt, S. (2009). Notes on the Comparison Class. In *International workshop on vagueness in communication.*

[1] This is the maximal random effect structure that converges