This is the pre-published version of an article accepted for publication in *Semantics of Natural Languages.*

Sentiment-linked quantification among adjectives[1]

Lelia Glass
School of Modern Languages, Georgia Institute of Technology (USA)
lelia.glass@modlangs.gatech.edu

May/June 2026

Abstract

Yoon (1996) observed that *The cups are clean* is taken by default to describe all the cups, while *The cups are dirty* might describe only some of them. This paper grounds Yoon's *clean/dirty* contrast in the sentiment (valence) of these adjectives: *clean* is desirable, *dirty* undesirable. Sentiment is linked to quantification through the longstanding idea, known as the Anna Karenina Principle, that a good situation must be *all* good, while *any* bad element makes it bad. Assuming that "Is the current situation okay?" is a common Question Under Discussion across contexts, positive-sentiment adjectives are easily taken to answer this question affirmatively, conveying that the current situation is *all* okay; while negative-sentiment adjectives are taken to answer it negatively, conveying that there is at least *something* wrong. Using quantitative data, this analysis is argued to explain not just Yoon's *clean/dirty* contrast, but also further instances of variable quantificational force among adjectives in the domains of degrees and dimensions. This paper breaks new ground in deriving (*a priori* emotionless) logical inferences from sentiment, a relatively less-explored topic in formal semantics.

Keywords: sentiment, adjectives, homogeneity, quantification, Anna Karenina Principle

1 Introduction

This paper revisits an observation from Yoon (1996) that (1b) – to a greater extent than (1a) – allows an interpretation whereby the predicate of the sentence applies to some-but-not-all of the cups.

(1) a. The cups are **clean**. ⇝ **All** of the cups are clean. (adapted from Yoon 1996, p. 217)
b. The cups are **dirty**. ⇝ **Some** of the cups are dirty.

[1]This paper was reviewed and accepted at *Natural Language Semantics*, but now appears in the inaugural issue of its open-access successor, *Semantics of Natural Languages*. I am very grateful to three reviewers at *Natural Language Semantics* (and two before that at the *Journal of Semantics*) as well as the Editor, Florian Schwarz, for very constructive feedback. This paper has benefited from conversations with Larry Horn, Joshua Knobe, Poppy Mankowitz, Louise McNally, and audiences at Yale and the Valence Asymmetries Conference organized in Barcelona (March 2026) by Isidora Stojanovic. My thinking on this topic is also shaped by my former advisors Beth Levin, Christopher Potts, and Christopher Kennedy.

In other words, sentences with definite plural subjects can in some cases receive an existential interpretation[2] rather than the usual universal interpretation. The availability of such interpretations seems to be connected to the lexical meaning of the predicate.

Krifka (1996) points out that the existential versus universal interpretation of definite plurals is not purely a lexical property of words such as *clean* versus *dirty*, but depends on how the sentence answers the Question Under Discussion (Roberts 2012) in context. If the QUD asks whether the speaker was able to reach the safe, the same predicate *open* can be interpreted as universal (2a) when *all* doors must be open to reach it, but as existential when it is reachable through *any* open door (2b).

(2) I was able to reach the safe because the doors were **open**. (adapted from Krifka 1996, ex. 12)

a. Context: The safe is behind a series of three doors; you must pass through all of them to reach it.
   ⇝ **All** of the doors were open.

b. Context: The safe is behind a wall with three adjacent doors; you can pass through any of them to reach it.
   ⇝ **Some** of the doors were open.

The same point is echoed by Lasersohn (1999), who notes that (3a) may be accepted as a rough scene-setting description if only 90% of the townspeople are asleep, while in the context of a scientific sleep study, (3b) may require 100% of the research subjects to be asleep. Again, a sentence with a definite plural subject varies in its quantificational force, and/or its tolerance to exceptions, depending on the Question Under Discussion.

(3) a. The townspeople are asleep. (adapted from Lasersohn 1999, p. 522)
   ⇝ **Most** of the townspeople are asleep.

b. The research subjects are asleep.
   ⇝ **Every single one** of the research subjects is asleep.

As elaborated below (Section 2), various theories have been proposed to explain the truth conditions of sentences with definite plural subjects and how these truth conditions interact with the QUD. But such theories have largely set aside Yoon's original observation about the role of particular predicates such as *clean* versus *dirty*. The data from Krifka and Lasersohn show that the existential/universal contrast should not be hardwired into the meaning of such lexical items, but

[2]The existential interpretation (1b) is also described as partial (Yoon 1996), weak (Yoon 1996), non-maximal (Dowty 1987; Brisson 2003), or imprecise (Lasersohn 1999), in contrast to the total, strong, maximal, precise interpretation of (1a).

Yoon's data can be taken to suggest that different lexical items raise different expectations about how the utterance addresses the QUD, in ways that remain unexplored.

This paper revisits the contrast between *clean* and *dirty*, asking which adjectival predicates tend to be interpreted as universal versus existential and why, and thus how different predicates address the QUD in different ways. While logical quantification may seem dispassionate, I pursue the hypothesis (laid out in Section 3) that the *clean/dirty* contrast is rooted in the sentiment (emotional valence) of these words: *clean* is desirable, *dirty* is undesirable.

The universal versus existential interpretations of (1) in the domain of individuals constitute just one among several examples of variable quantificational force among adjectives. As reviewed in Section 2, variable quantificational force is also found among degrees (*clean* means zero dirt, while *dirty* denotes at least some dirt; Cruse 1980; Rotstein & Winter 2004; Kennedy 2007) and across dimensions of multi-dimensional adjectives (*healthy* is conjunctive in that it quantifies across all organ systems and biometrics, while *sick* is disjunctive in that it describes a deficiency in at least one; Yoon 1996; Sassoon 2013). I argue (Section 3) that these phenomena are also rooted in sentiment.

Synthesizing work in social psychology (e.g., Kanouse 1984; Baumeister et al. 2001; Rozin & Rozyman 2001), I link sentiment to quantification through the longstanding and widely motivated Anna Karenina Principle – that is, the idea that a good thing must be *all* good, while *any* bad element makes it bad. This principle explains why the questions (4) are vastly more common than their opposites (5) in any corpus, and predicts that they should also be more likely as unstated QUDs.

(4) a. Is everything okay?
    b. Is something wrong?

(5) a. Is something okay?
    b. Is everything wrong?

The quantificational asymmetry of (4) thus shapes our expectations about how this QUD is answered by sentences predicating positive-sentiment versus negative-sentiment adjectives of definite plurals. Positive-sentiment adjectives favor a universal interpretation because they are easily taken to answer the QUD with "Everything is okay," whereas negative-sentiment adjectives are more likely to allow an existential interpretation because they are easily taken to answer it with, "Something is wrong."

Turning to degrees, the Anna Karenina Principle is argued to explain an old observation from Cruse (1980) that negative-sentiment adjectives such as *dirty* often describe a minimal (existential) degree of dirt, while their positive-sentiment antonyms such as *clean* may describe a maximal

(universal) degree of cleanliness. Assuming that the lexicon evolves to encode the distinctions that speakers find useful, the Anna Karenina Principle predicts that speakers find it most useful to distinguish things that are *all* clean (everything is okay) from those with *some* amount of dirt (something is wrong), rather than those with *some* cleanliness (something is okay).

Turning to dimensions, the same logic explains why positive-sentiment multidimensional adjectives such as *healthy* tend to quantify conjunctively (universally) across all dimensions of health, while their negative-sentiment antonyms such as *sick* quantify disjunctively (existentially) across some dimension of sickness. Here too, the Anna Karenina Principle predicts that speakers find it most useful to distinguish states of full health across all dimensions (everything is okay) from sickness along some dimension (something is wrong) than the reverse, and thus why these distinctions are reflected in the structure of the lexicon.

This analysis yields predictions tested in quantitative data at scale (Section 4). In the domain of individuals, a judgment study shows, as predicted, that existential interpretations are more likely for negative-sentiment adjectives such as *dirty*. In the domain of degrees, a corpus study shows that negative-sentiment adjectives such as *dirty* are more often modified by minimum-standard modifiers (*slightly*). Across dimensions, it is shown that the adjectives characterized by Sassoon (2013) as disjunctive (*sick*) have lower sentiment than those characterized as conjunctive (*healthy*). Overall, this paper offers to explain multiple distinct instances of sentiment-linked variable quantificational force among adjectives, involving both token-level pragmatic inferences and type-level meaning.

The Anna Karenina Principle has been discussed for centuries under various names, dating back to Aristotle, and Cruse (1980) noted decades ago that *clean* describes maximal (desirable) cleanliness while *dirty* indicates a minimal degree of undesirable dirt. But this paper is the first to observe the effect of adjectival sentiment on quantification across individuals and dimensions; to offer a unified explanation for variable quantification across individuals, degrees, and dimensions; and to test these claims in quantitative data. Behind Horn (2022, p. 827) and Glass (2023), it is also one of the only papers to invoke the Anna Karenina Principle to explain sentiment-linked patterns within semantics and pragmatics.

Stepping back (Section 5), this paper connects lexical semantics to discourse pragmatics by exploring how the type-level meaning of an adjective drives inferences about the discourse context of its tokens, and how the structure of the lexicon reflects the distinctions that speakers find useful. Moreover, sentiment has been studied widely in sibling (sub)fields such as psychology (Osgood et al. 1957), natural language processing (Pang & Lee 2008), corpus linguistics (Sinclair 1991), and cognitive linguistics (Paradis et al. 2012), and in semantic research on clearly emotive phenomena such as expressives (e.g., Potts 2007), but has only recently attracted attention in the formal literature on adjectives (e.g., Nouwen 2024; Gotzner & Mazzarella 2024; Beltrama 2025).

This paper thus advances a broader endeavor to connect the logical and emotional dimensions of meaning.

2 Background

This section situates Yoon's *clean/dirty* contrast within the literature about variable quantificational force, adjectival antonym pairs, and sentiment.

2.1 Variable quantificational force

Yoon's *clean/dirty* contrast is connected to homogeneity (Fodor 1970; Löbner 1987, 2000; Križ 2016, 2019; Champollion et al. 2019; Bar-Lev 2021): the idea that sentences with definite plural subjects are typically taken to describe a set of individuals who all match (i.e., are homogeneous) with respect to the relevant predicate. In (6), by default, either all the kids laugh or none of them do. The universal interpretation (6a) is a strong default for positive (non-negated) sentences with definite plural subjects.[3]

(6) **Homogeneity** (adapted from Bar-Lev 2021, p. 1046)

a. The kids laughed. $\rightsquigarrow$ **All** of the kids laughed.

b. The kids did **not** laugh. $\rightsquigarrow$ **None** of the kids laughed.

Homogeneity means that the truth conditions of the positive and negated forms of (6) do not exhaust the logical space – that is, they are contraries rather than contradictories in the sense of Aristotle (Horn 2025). We are left with a gap of murky judgments ("varied" across speakers and contexts, and "guarded" with hesitation, according to Rooth 1987, p. 254) in non-homogeneous "mixed" scenarios where some kids laugh and others don't. These murky judgments, in turn, create the potential for existential/non-maximal interpretations like those observed above, whereby sentences like (7) may be judged true, depending on the context, when only some of the kids laughed.

(7) **Non-maximality** (adapted from Bar-Lev 2021, p. 1047)

a. Q: Was the clown funny?

b. A: The kids laughed. $\rightsquigarrow$ **Some/most** of the kids laughed.

Homogeneity (6) and non-maximality (7) can thus be seen as "two sides of the same coin"

[3]This observation was originally named the "all or none" presupposition by Fodor (1970, p. 162), who noted similar issues among sentences with bare plural subjects. I return to bare plurals at the end of Section 4.

(Križ & Spector 2021, p. 1132) because both properties reflect delicate judgments, sensitive to the Question Under Discussion, about sentences with plural subjects in mixed scenarios. To capture such judgments, the literature offers three families of analyses, characterized by Bar-Lev (2021, p. 1047) as 1–3.

1. The Ambiguity Approach (Krifka 1996; Lasersohn 1999; Malamud 2012; Križ & Spector 2021): *The kids laughed* has family of potential meanings, filtered down in context (inspired by the Strongest Meaning Hypothesis of Dalrymple et al. 1994, 1998) to those that are logically strongest and/or most pragmatically informative in light of the QUD.

2. The Trivalence Approach (Križ 2016; Champollion et al. 2019): *The kids laughed* is undefined when some kids laughed and others didn't, but can be used anyway when it resolves the QUD in the correct way.

3. The Implicature Approach (Bar-Lev 2021; Chierchia 2022): *The kids laughed* has essentially existential truth conditions, but can be pragmatically strengthened to a universal interpretation when that interpretation would resolve the QUD.

Rather than choosing among these approaches, this paper explores how the potential for non-maximality is linked to the lexical meaning of the predicate, as shown in Yoon's *clean/dirty* contrast.

As Chierchia (2022) notes, definite plurals constitute just one example among many where we observe variable quantificational force in the absence of any overt quantifier; other examples include bare plurals (Carlson 1977; Leslie 2008), locative prepositions (Mador-Haim & Winter 2015), markers of free choice and negative polarity (Giannakidou 2001), and more. Representing a further instance of variable quantificational force in the absence of a quantifier, plural predication also raises questions about distributive versus collective interpretations of the predicate (Link 1983; Roberts 1987; Champollion 2010). On a distributive interpretation (8a), a predicate is inferred to apply individually (to 'distribute' down) to each member of a plural subject. On a collective interpretation (8b), the predicate applies to the plural subject as a whole, but not to each member individually.

(8) The boxes are **heavy**. (adapted from Scontras & Goodman 2017, p. 294)

   a. **Distributive:** $\rightsquigarrow$ Each box is individually heavy.
   b. **Collective:** $\rightsquigarrow$ The boxes are jointly heavy but not individually so.

Most adjectival predicates (including *clean, dirty, healthy, sick*, and so on) require a distributive interpretation (i.e., are "stubbornly distributive" in the sense of Schwarzschild 2011), presumably because they describe properties that inherently characterize individuals (Roberts 1987;

Glass 2018). Collective interpretations of adjectival predicates (8b) are rarer, restricted to certain predicates that can characterize groups above and beyond their individual members, and require more contextual support (Scontras & Goodman 2017; Glass 2018). But as Brisson (2003, p. 131) observes, both distributive and collective interpretations of a predicate can involve universal or existential ('non-maximal') interpretations of the definite plural subject.

2.2 Adjectives

Variable quantificational force is also found in other facets of adjective meaning. Gradable adjectives are commonly taken to map individuals to their degree on a scale (Bartsch & Venneman 1972; Cresswell 1977; Kennedy 1997; Kennedy & McNally 2005) – an ordered dimension, such as height for the adjective *tall*. As observed by Unger (1975); Cruse (1980); Rotstein & Winter (2004); Kennedy & McNally (2005) and Kennedy (2007), some scales are open-ended: height extends infinitely, yielding vague relative gradable adjectives such as *tall*, which can be modified by relative adverbs such as *very* and *fairly*, but resist those describing maxima (*perfectly*) or minima (*slightly*). *Tall*'s antonym *short* also arguably extends indefinitely (Kennedy 2007, p. 35), even though an entity could in principle have zero height, as evidenced by the fact that *short* also cannot easily be modified by *perfectly* or *slightly*.

Other scales have a clear maximum standard: a cup can be so *clean* that it could not be cleaner. Maximum-standard adjectives such as *clean* can therefore be modified by adverbs describing maximal degrees (*perfectly, totally*). On the other hand, *dirty* is a minimum-standard adjective: a cup counts as *dirty* once it has any noticeable dirt on it. So *dirty* can be modified by adverbs describing minimal degrees (*slightly, partially*). While pairs such as *clean/dirty* have one maximum standard and one minimum standard, there are other pairs such as *full/empty* where both ends show a maximum standard. Table 1 presents a typology of scale structures for gradable adjectives adapted from Kennedy (2007), as diagnosed by their compatibility with maximum- versus minimum-standard modifiers – although, as observed by Kamoen et al. (2011), such judgments may not be as clear-cut as the literature suggests.

We saw above (1) that *clean* leads to an inference of universal quantification across individuals when applied to a plural subject (all the cups are clean). When applied to a singular subject, it similarly evokes a maximal (universal) degree of cleanliness (9a). On the other hand, *dirty* leads to an inference of existential quantification across individuals when applied to a plural subject (some of the cups are dirty), as well as an existential degree of dirt when applied to a singular (9b).

(9) **Maximum-standard versus minimum-standard adjectives** (Cruse 1980; Rotstein & Winter 2004; Kennedy 2007)

a. The cup is clean. $\rightsquigarrow$ The cup is **totally/perfectly** clean.

Scale structure	**Left-side antonym**	**Right-side antonym**
Open at both ends:	*(✗slightly, ✗perfectly) short, light, cheap, sad* (relative)	*(✗slightly, ✗perfectly) tall, heavy, expensive, happy* (relative)
Closed at one end:	*(✗slightly, ✓perfectly) clean, healthy, straight, safe, dry* (maximum-standard)	*(✓slightly, ✗perfectly) dirty, sick, bent, dangerous, wet* (minimum-standard)
Closed at both ends:	*(✗slightly, ✓perfectly) empty, transparent* (maximum-standard)	*(✗slightly, ✓perfectly) full, opaque* (maximum-standard)

Table 1: Typology of scale structures for gradable adjectives, adapted from Kennedy (2007, p. 33, p. 36).

b. The cup is dirty. $\rightsquigarrow$ The cup is **somewhat/partially** dirty.

As for their semantic representation, Kennedy (2007) proposes that all gradable adjectives (relative adjectives such as *tall*, absolute maximum-standard adjectives such as *clean*, and absolute minimum-standard adjectives such as *dirty*) denote measure functions, mapping an individual to its degree along the adjective's associated scale. In the positive (i.e., non-comparative, non-superlative) form, gradable adjectives are argued to combine with a silent morpheme POS (Bartsch & Venneman 1972), which requires that this degree exceeds a contextual standard **s** for what "counts as" tall, clean, or dirty in the context.

(10) **Semantics for gradable adjectives in the positive form** (adapted from Kennedy 2007, p. 20)

a. $[\![tall_{\text{POS}}]\!] = \lambda x[\textbf{tall}(x)] \succeq \textbf{s}(\textbf{tall})$

b. $[\![clean_{\text{POS}}]\!] = \lambda x[\textbf{clean}(x)] \succeq \textbf{s}(\textbf{clean})$

c. $[\![dirty_{\text{POS}}]\!] = \lambda x[\textbf{dirty}(x)] \succeq \textbf{s}(\textbf{dirty})$

All three types of adjectives (relative, maximum-standard, and minimum-standard) have the same semantics, even though they differ in their interpretation: *tall* is vague, while by default, *clean* describes a maximal degree of cleanliness and *dirty* describes a minimal degree of dirt. To explain these differences, Kennedy (2007) points to their scale structure (Table 1). He argues that vague relative adjectives such as *tall* are anchored only to the discourse context, while absolute maximum-standard adjectives such as *clean* are anchored to the maximal point of the scale, and

absolute minimum-standard adjectives such as *dirty* are anchored to the minimum. The unified semantics in (10) interacts with the typology of scale structures (Table 1) to explain the varying quantificational force of these adjectives in the realm of degrees.

I adopt this analysis as far as it goes. But it remains an open question, or a matter of lexical idiosyncrasy, which adjectives are associated with what sort of scale and why.

Another quantificational dimension of adjective meaning is found among adjectives describing properties that comprise multiple dimensions[4], such as *healthy*, which involves the state of all one's organs and biometrics – in contrast to unidimensional adjectives such as *clean/dirty*, which describe the single dimension of cleanliness/dirt. By default, someone who is *healthy* is taken to satisfy all these dimensions (11a), while someone who is *sick* may be deficient in only one (11b).

(11) **Multidimensional adjectives** (adapted from Sassoon 2013)

a. My grandma is healthy. $\rightsquigarrow$ My grandma is healthy in **every** respect.

b. My grandma is sick. $\rightsquigarrow$ My grandma is sick in **some** respect.

To represent these adjectives semantically, Sassoon (2013, ex. 9) suggests that *healthy* by default quantifies universally over dimensions of health, while its antonym *sick* asserts the existence of at least one dimension lacking health. Sassoon's quantificational analysis can be synthesized with Kennedy's semantics for gradable adjectives in the positive (i.e., non-comparative) form to arrive at denotations such as (12).

(12) a. $[\![healthy_{\text{POS}}]\!] = \lambda x[\forall Q \in DIM(\textbf{healthy})[\textbf{health-wrt}(Q,x) \succeq \textbf{s}]]$
"For all dimensions of health Q, x's health with respect to Q exceeds a contextual standard **s**."

b. $[\![sick_{\text{POS}}]\!] = \lambda x[\exists Q \in DIM(\textbf{healthy})[\neg\ \textbf{health-wrt}(Q,x) \succeq \textbf{s}]]$
"For some dimension of health Q, x's health with respect to Q does not exceed (i.e., falls short of) a contextual standard **s**."

As observed by Sassoon (2013), the facets of quantification noted in (9) and (11) are connected: someone who is *healthy* shows a maximal degree of health across all dimensions thereof. Moreover, Yoon (1996) classifies *healthy* as a 'total' predicate like *clean*, meaning that it is taken to quantify universally across members of a definite plural subject (13a). Thus *healthy* quantifies universally across degrees, dimensions, and individuals. On the other hand, someone who is *sick* exceeds a minimum standard of sickness on at least one dimension, and Yoon (1996, p. 227) argues that (13b) can be used when only some of the kids are sick. Thus *sick* quantifies existentially

[4]Such adjectives were originally described as nonlinear by Kamp (1975) and Klein (1980), and renamed multidimensional by Sassoon (2013) drawing on Kennedy (1997, §1.1.2).

across degrees, dimensions, and individuals.

(13) a. The kids are **healthy**. ⇝ **All** of the kids are healthy.
b. The kids are **sick**. ⇝ **Some** of the kids are sick.

Across the domains of individuals, degrees, and dimensions, it remains an open question why the quantificational force of *healthy* is universal while that of *sick* is existential. Yoon (1996) frames it as a lexical idiosyncrasy, which is not explanatory; Krifka (1996) shows that it is not purely lexical in the first place. We are left to wonder: Why then do *healthy* and *sick* behave as they do? Which further predicates will pattern in which ways?

My answer will leverage the structure of pairs such as *clean/dirty* and *healthy/sick*. Such pairs illustrate that the lexicon of adjectives is organized not just around scales/degrees, but also around antonyms (Sapir 1944; Boucher & Osgood 1969; Cruse 1980; Lehrer & Lehrer 1982; Lehrer 1985) that frame the underlying scale from opposite perspectives. One half of the pair is typically considered 'unmarked' (a term critiqued by Haspelmath 2006 for being confusingly polysemous) – in the sense of describing a scale along its increasing dimension (*tall* describes increasing height, while *short* describes decreasing height); being morphologically basic rather than complex (*even* versus *uneven*); learned earlier by children (Clark 1993, p. 32); listed earlier in binomial constructions (*happy or unhappy* is more common than *unhappy or happy*; Cooper & Ross 1975, p. 65); more flexible in terms of the contexts that it can occur in (*how tall are you?* can be used in a range of contexts, while *how short are you?* is restricted to contexts where shortness is presupposed); more frequent overall; and/or – I argue, crucially – more positive in terms of sentiment. This paper posits that sentiment helps to explain all the adjective-related quantificational inferences described here.

2.3 Sentiment

Researchers have long observed that words carry emotional as well as denotational meaning (Osgood et al. 1957). In psychology, the opposition between good and bad (linked by Boucher & Osgood 1969 to the yin/yang of Eastern philosophy) is thought to organize the human experience at a fundamental level (Taylor 1991; Rozin & Rozyman 2001; Baumeister et al. 2001; Barrett 2006), including the meanings of words. Unlike denotational meaning, sentiment is narrow enough to plot on a single axis, but also connects to broader topics in emotion and cognition.

In natural language processing, a long tradition of work aims to automatically identify the sentiment of text (Pang & Lee 2008). Adjectives in particular are noted as a locus of textual sentiment (Wiebe 2000; Taboada 2016), and the lexicon of adjectives is organized around antonym pairs with opposing sentiment: *clean* and *healthy* are desirable, *dirty* and *sick* are undesirable.

Sentiment is long recognized in linguistics too. In corpus linguistics, scholars study SEMANTIC

PROSODY: patterns of collocational sentiment spanning multiple words, as when *set in* tends to describe unpleasant states of affairs (Sinclair 1991, p. 73–74).

In the study of adjectival antonyms in particular, Horn (1989, p. 274) credits Zimmer (1964) with the observation that negative affixes such as *un-* tend to attach to positive-sentiment adjectives to form negative-sentiment antonyms (*happy/unhappy*), connecting morphological markedness to negative sentiment. Lehrer (1985, p. 399) similarly counts a positive connotation as one of several attributes that tend to characterize the unmarked half of an antonym pair.

Sentiment is important in cognitive linguistics, where researchers aim to connect language to psychology. Paradis et al. (2012) use the Implicit Association Test (Greenwald et al. 1998) to probe the emotional valence of eight antonym pairs of relative gradable adjectives describing physical properties that are not explicitly sentiment-coded, such as *large/small, heavy/light*, and *fast/slow*. They find that each antonym pair is strongly divided by sentiment, one half being emotionally positive and the other negative, consistent with the idea from Boucher & Osgood (1969) that antonyms are organized around sentiment opposition. They also find that the positive-sentiment half of the pair is usually the one describing more of the relevant property (more size, more speed), further connecting positive sentiment to the unmarked half of an antonym pair. But they also find exceptions: *heavy* and *thick* describe greater weight/size but are found to be more emotionally negative, at least out of context, than their antonyms *light* and *thin*.

To complicate matters further, a given adjective type can describe a property that is desirable for some objects but undesirable for others: *cold beer* is good while *cold pizza* is bad (Fahrni & Klenner 2008); a *thick blanket* is cozy while *thick fog* is dangerous. Using classifications of adjectives from Bierwisch (1989, p. 88) and Dixon (2004, p. 3-4), it is especially common among dimensional adjectives (i.e., those describing size) for a given property to vary in desirability depending on what it is applied to. In contrast, among adjectives describing value judgments (*beautiful/ugly*), non-dimensional physical properties (*clean/dirty*), and human propensities (*clever/stupid*), desirability is more consistent across contexts.

Cruse (1980) uses sentiment to explain the scale structure of maximum/minimum-standard pairs such as *clean/dirty* (Table 1). "The desirable state is zero value of the property" (i.e., zero dirt), he writes (p. 23), "and this is signaled by one of the terms of the opposition" (i.e., *clean*). Moreover, "any positive value" – of dirt – "represents an unsatisfactory state, and thus is signaled by the other term of the opposition" (*dirty*). Cruse (1980, p. 21) describes *clean* as E-POSITIVE (e- for evaluative), *dirty* as E-NEGATIVE (using terms from Boucher & Osgood 1969, p. 1). Cruse (1980, p. 15) uses the term Q-POSITIVE (q- for quantity) for adjectives describing a greater quantity of the relevant property. Thus *clean* is q-negative (describing less dirt) but e-positive (desirable); *dirty* is q-positive (describing a greater quantity of dirt) but e-negative (undesirable).

As Cruse explains, maximum/minimum-standard pairs such as *clean/dirty* show different gram-

matical and inferential patterns from relative gradable antonyms such as *tall/short*. Relative gradable antonyms such as *tall/short* allow an intermediate "zone of indifference" of objects that are neither *tall* nor *short* (Sapir 1944, p. 98), while pairs such as *clean/dirty* arguably exhaust the logical space by describing zero versus any amount of dirt.

As for scales such as *full/empty* that are closed at both ends (Table 1), Cruse (1980, p. 15) says these are "relatively infrequent" and would "require quite a separate treatment." Such pairs are often divided by sentiment (*full* is more positive than *empty*), but sentiment does not drive any difference in quantification over degrees, because both antonyms equally describe maximum standards.

Within formal semantics, Cruse (1980) is credited by Rotstein & Winter (2004) for many of the observations defining the class of absolute gradable adjectives like *clean/dirty*. Rotstein & Winter (2004) inspire the influential analysis by Kennedy (2007) of the lexical semantics of relative versus absolute gradable adjectives, and Kennedy (2007) lays the foundation for the work of Sassoon (2013) on multidimensional absolute adjectives such as *healthy/sick*. But none of these papers, nor Yoon (1996), mention Cruse's observation that antonym pairs of absolute gradable adjectives such as *clean/dirty* and *healthy/sick* are organized around sentiment. In other words, while *clean/dirty* and *healthy/sick* clearly contrast in sentiment, sentiment has not featured prominently in the semantics literature about such adjectives. The effect of sentiment on variable quantification among degrees is discussed by Cruse, but hardly pursued in later work; and its effect across individuals and dimensions has not been noted at all.

Sassoon (2010, p. 144) does mention sentiment, but only to dismiss it as irrelevant to her point about which half of an antonym pair can appear with measure phrases or ratio comparisons (*five feet* {*tall/*short*}*, twice as* {*tall/*short*}). Similar to Paradis et al. (2012), she observes exceptional cases such as *old/young* (or *thick/thin*) where the adjective describing more of the relevant property is arguably less desirable in some contexts (i.e., q-positive but e-negative in Cruse's terms), and says that she aims to explain linguistic facts about measure phrases and ratio comparisons that pattern with the adjective describing more of the relevant property, rather than cultural facts about which property is more desirable.

Recently, though, some work in the formal semantics of adjectives has started to confront the role of sentiment. Glauch (2024) studies adjectives carrying both denotational and emotional meaning ('mixed expressives'; Gutzmann 2011) such as *shitty*, arguing that negative appraisal is key to explaining their felicity conditions. Nouwen (2024) strives to explain why the sentiment of an adjective predicts the meanings of adverbs built from it, as when *pleasantly warm* denotes moderate warmth while *unpleasantly warm* denotes an extreme. Gotzner & Mazzarella (2024) show experimentally, as predicted by Ducrot (1973, p. 125), that negated positive-sentiment adjectives are more likely to be pragmatically strengthened to entail their antonym: *not very clean* is

taken to convey *dirty*, whereas *not very dirty* does not necessarily mean *clean*. As discussed below (Section 3.3.3), Beltrama (2025) explores adjectives describing mildly positive evaluations such as *adequate*, arguing that they describe a minimally sufficient degree of the relevant property given some goal or norm.

Adjectives with strong sentiment are also often subjective in the sense of Lasersohn (2005), allowing faultless disagreement (Kölbel 2004): two people can disagree about whether roller coasters are *fun* or *scary* without either of them being wrong (see, e.g., Stojanovic & Kaiser 2022 for discussion of how subjectivity and sentiment overlap). Subjectivity is found to shape syntax (Scontras et al. 2017, 2019), in that subjective adjectives occur further away from the head noun in multi-adjective attributive strings (*fun blue toy* is preferred over *?blue fun toy*), and more likely to appear in predicative position (*The red dress is cute* is a more likely utterance than *The cute dress is red*; Glass 2025). Such work has begun to uncover both grammatical and inferential consequences of the emotional side of the lexicon.

While it has only blossomed recently, this connection has deep historical roots. In work published after his death in 1939, discussing comparatives such as *more than* versus *less than*, Sapir (1944, p. 111-112) observes:

> "Here, as in every other phase of linguistic inquiry, we find that the more closely we study actual linguistic forms, the more we are driven to realize that they never express merely static, affectively neutral, concepts and judgments, but classes of concepts and judgments in which nuclear notions, capable of logical definition, are colored by unavowed dynamic and affective determinants."

Over eighty years later, this paper takes inspiration from Sapir's insight that the logical and emotional dimensions of language are linked.

3 From sentiment to quantification

The study of quantification may seem to involve dry, emotionless logic.[5] But this paper hypothesizes that the varied quantificational force observed among adjectives is shaped by sentiment.

3.1 The Anna Karenina Principle

In particular, I argue that such inferences can be rooted in the insight – discussed under many names in diverse fields, sometimes called the Anna Karenina Principle (14) – that for a composite

[5]This framing takes inspiration from Potts (2011), who notes that logical negation (*not*) might seem dispassionate, but tends to appear in unfavorable film reviews.

entity or situation to be good, *all* of its components must be good; equivalently, the entity/situation goes wrong if *any* component is wrong. This principle links quantification to sentiment.

(14) **The Anna Karenina Principle**
A good situation must be *all* good; *any* bad element makes it bad.

Ranging from the renowned first line of Tolstoy's novel *Anna Karenina*, to insights dating back to Aristotle (15), various thinkers have noted that happiness and success require a conjunction of multiple favorable elements, such that the (disjunctive) absence of any of them suffices for unhappiness/failure. The human geographer Jared Diamond (1997) named the Anna Karenina Principle[6] and used it to explain why many animals are unsuitable for domestication, and thus why some societies developed faster than others.

(15) (adapted from Glass 2023, p. 24)

a. "All happy families are alike; each unhappy family is unhappy in its own way." – Tostoy (2004 [originally published 1874])

b. "It is possible to fail in many ways [. . .], while to succeed is possible only in one way (for which reason also one is easy and the other difficult – to miss the mark easy, to hit it difficult)." – Aristotle (1925 [originally written 350 B.C.E.])

While the literary allusion may seem esoteric, I use the name Anna Karenina Principle for consistency with prior work across many fields.

In social psychology, the Anna Karenina Principle is known as negativity dominance (Peeters & Czapinski 1990; Taylor 1991; Rozin & Rozyman 2001; Baumeister et al. 2001): the finding that the overall assessment of a composite object is not just a sum or average of its traits, but is as bad as its worst element. Negativity dominance is reflected in folk sayings such as "a chain is as strong as its weakest link." It is viscerally illustrated in the domain of food: "one rancid ingredient spoils the finest soup" (Kanouse 1984, p. 705); an elaborate meal is ruined by a single cockroach, while there is no parallel anti-cockroach whose presence would redeem it (Rozin & Rozyman 2001, p. 296). Along the same lines, a person who murders once is a murderer, while one must worry often to be considered a worrier (Leslie 2008, p. 15), because murder is appalling while worrying is benign. A faithful spouse is faithful always, while a single affair qualifies them as unfaithful.[7] Echoing the literary roots of the Anna Karenina Principle, Rozin & Rozyman (2001) link negativity dominance to the "tragic flaw" from Ancient Greek tragedies: *any* single defect overcomes *all* of a hero's advantages.

[6] Please see https://en.wikipedia.org/wiki/Anna_Karenina_principle.

[7] Thanks to Larry Horn for this example.

Negative events trigger cognitive processing as well as gut emotion, perhaps because people are motivated to figure out how the event could be avoided next time (Taylor 1991, p. 73). Weiner (1985) finds that bad outcomes prompt greater causal reasoning than good ones, and Knobe (2003) shows that people attribute greater intention to agents who cause bad outcomes rather than good ones.

Negativity dominance can also be seen as the flip side of the POLLYANNA HYPOTHESIS of Boucher & Osgood (1969), named after a famously optimistic fictional heroine. Boucher & Osgood (1969) observe that positive-sentiment word tokens are more common in text (a widely replicated finding; Dodds et al. 2015), which they take to show that people tend to "look on the bright side of life" (p. 1). While this finding may seem to conflict with negativity dominance, Taylor (1991, p. 77) weaves both observations together by suggesting that the normal state of affairs is mildly positive, which is why positive-sentiment word tokens are more frequent, but also why any single bad element dominates our attention by disrupting this homeostasis.

Negativity dominance is connected to negative differentiation: the observation that negative experiences are more diverse and distinctive than positive ones (Rozin & Rozyman 2001; Baumeister et al. 2001), as presaged by Tolstoy and Aristotle. In astrobiology, it is noted that the emergence of life requires various parameters – heat, atmospheric chemicals, and so on – to all fall within a habitable "Goldilocks zone" (Gonzalez et al. 2001): not too much, not too little, representing the "doctrine of the [golden] mean" from Aristotle (Kraut 2022) and Confucius (Legge 1861). A pleasant meal must likewise be neither be too spicy nor too bland (Unkelbach et al. 2020). Even on a single dimension such as temperature or spice, there are two ways to fail (too much, too little), but only one way to succeed (just right). As previewed above, Nouwen (2024) uses the "Goldilocks zone" to explain why adverbs built from positive-sentiment adjectives denote intermediate degrees (*pleasantly warm*) while those built from negative-sentiment adjectives denote extremes (*unpleasantly warm*).

Moving to multiple dimensions, if *any* bad element suffices for a composite object to be deemed bad, then bad things are also more distinctive because each bad thing may have its own unique reason for badness.

In pragmatics, Horn (2022, p. 827) invokes the Anna Karenina Principle to explain why, when two clauses are conjoined, it is often the negatively-valenced one that occupies the position of information-structural prominence at the end of the sentence. He spotlights the order of conjuncts in Tolstoy's (15a) as well as another famous literary opening:

(16) "It was the best of times, it was the worst of times." – Dickens (1859)

Horn suggests that negative information receives structural prominence because it is more distinctive and interesting. Social psychologists might add that negative information dominates our

assessment of a composite.

In semantics, Glass (2023) uses the Anna Karenina Principle to explain why the seemingly neutral verb *cause* tends to occur with negative-sentiment complements in corpora (*cause problems, cause cancer*). Glass (2023) draws on findings from social psychology (Kun & Weiner 1973) that successes are taken to require multiple necessary-but-individually-insufficient factors (hard work *and* luck), such that the absence of any of them suffices for failure, and casts the Anna Karenina Principle in terms of causal models inspired by Pearl (2009). But the key insight extends beyond events and causal relations: more generally, good *things* – composite objects as well as events – require a conjunction of good components, such that the (disjunctive) absence of any of them suffices for badness.

3.2 The Anna Karenina Principle across individuals

This paper aims to use the Anna Karenina Principle to explain Yoon's *clean/dirty* contrast in the realm of individuals, via pragmatic reasoning about how the QUD is addressed by a sentence attributing a predicate to a definite plural subject.

Beginning with the semantic representation of such sentences, we saw above (Section 2.1) that the literature offers various possible analyses for predicates applied to definite plurals. But all approaches agree that such sentences allow both universal and existential interpretations as determined by the QUD. For concreteness, I illustrate with the analysis of Champollion et al. (2019), which builds on Križ (2016). These authors argue that a sentence such as (17) is true in a world $w_{\text{all open}}$ in which all the doors are open, false in a world $w_{\text{none open}}$ in which none of them are open, and undefined in a "mixed" scenario $w_{\text{some open}}$ where some are open and others closed. Here, of course, *some* must read as "some but *not all*."

Although its truth value is undefined in a mixed scenario, Champollion et al. (2019) argue, (17) can be used felicitously when it's "true enough" (echoing Lasersohn 1999, p. 523) – when the worlds are indistinguishable with respect to the QUD ($w_{\text{some open}} \approx w_{\text{all open}}$), so that (17) answers the QUD in the same way whether it's actually true, or undefined but "true enough."

(17) The doors are open.

a. $w_{\text{all open}}$: True

b. $w_{\text{some open}}$: Undefined

c. $w_{\text{none open}}$: False

Returning to the example from Krifka (1996), let us imagine that the QUD asks whether the speaker is able to reach the safe. If the safe is behind three side-by-side doors, such that the speaker can pass through any of them to reach it, then $w_{\text{some open}} \approx w_{\text{all open}}$ because the QUD about reach-

ing the safe is resolved affirmatively in both scenarios. On the other hand, if the safe is behind three sequential doors, then $w_{\text{some open}} \not\approx w_{\text{all open}}$ because these situations would resolve the QUD in different ways: the speaker can't reach the safe in $w_{\text{some open}}$, but can reach it in $w_{\text{all open}}$. This analysis explains why (17) can be interpreted as universal or existential depending on the arrangement of the doors. It also explains why, as noted above, the homogeneous w_{all} interpretation is strongly preferred, because the sentence is actually true in that scenario. In contrast, the heterogeneous/mixed interpretation w_{some} requires more context and may be met with confusion because speakers have to recover the QUD to decide whether the sentence counts as true enough in context even though it is technically undefined.

The same intuition could be re-worked in other frameworks. If (17) is compatible with a variety of meanings as determined by the QUD, then such reasoning could explain which interpretations are available. And if (17) has essentially existential truth conditions that can be strengthened to universal ones in light of the QUD, then this reasoning could explain when such strengthening is needed. Most important is the observation that the some-but-not-all interpretation may be available or unavailable, depending on whether w_{some} and w_{all} would give equivalent or distinct answers to the QUD. Secondarily, one would have to explain why the homogeneous w_{all} interpretation is favored above the mixed w_{some} interpretation, which on the theory of Križ (2016) and Champollion et al. (2019) arises because the sentence is true in the w_{all} scenario but undefined in the w_{some} one. If the sentence is instead analyzed to be true in the w_{some} scenario, then these facts would need a different explanation.

In any case, such an analysis captures how universal versus existential interpretations of definite plurals depend on the QUD. But here the QUD is specific to the imagined context about the safe, and the variability hinges on the stipulated position of the doors. Next, I propose to generalize this analysis to contexts such as Yoon's *clean/dirty* example, where the QUD is not stated explicitly and the variability hinges on the sentiment of the adjective.

For Roberts (2012), conversations circle around a Big Question, "What is the way things are?," raising a series of nested sub-questions that interlocutors pursue as guided by their domain goals.Roberts thus elaborates what Grice (1989, p. 45) means when he refers to the "current purposes of the exchange." The sub-question, "Is the current situation okay?," fits naturally under the overarching Big Question and arguably serves various domain goals. Assuming via the Anna Karenina Principle that a good thing must be all good while any bad element makes it bad, asking whether the current situation is okay amounts to asking whether *everything* is okay, or whether *something* is wrong.

In contrast, the opposite set of questions – whether *something* is okay; whether *everything* is wrong – are less useful. It is under-informative to know whether *something* is okay, because this still does not tell us whether the whole situation is okay. It is over-informative to know whether

everything is wrong, because a situation can go wrong even without every part of it being so.

As previewed above, the questions (18) are thus predicted to be vastly more common than their opposites (19) in any corpus, and presumably also far more likely as implicit Questions Under Discussion across a variety of contexts.

(18) a. Is everything okay?
b. Is something wrong?

(19) a. Is something okay?
b. Is everything wrong?

As a result, when readers scour their background knowledge to interpret a decontextualized example such as (20), they may reason that "Is the current situation okay?" (18) is a likely QUD, and that this question is answered affirmatively by a sentence with a positive-sentiment predicate. The sentiment of the adjective offers a clue as to how it may address this common QUD.

(20) a. The cups are **clean**. $\rightsquigarrow$ **All** of the cups are clean. (= (1))
b. The cups are **dirty**. $\rightsquigarrow$ **Some** of the cups are dirty.

The Anna Karenina Principle tells us that a situation is only okay if it is *all* okay. A scenario with all clean cups is therefore not equivalent to one with some clean cups ($w_{\text{all}} \not\approx w_{\text{some}}$) because these would provide different resolutions to the QUD of whether the current situation is okay (Table 2). Thus (20a) is only true in the $w_{\text{all clean}}$ scenario, not "true enough" in the mixed $w_{\text{some clean}}$ scenario. So (20a) is interpreted universally, which is also the default interpretation for sentences with definite plural subjects overall.

On the other hand, readers may assume that this QUD ("Is the current situation okay?") is answered negatively by a sentence with a negative-sentiment predicate. According to the Anna Karenina Principle, the situation is not okay if any part of it is not okay. Thus, a situation with some dirty cups is equivalent to one with all dirty cups ($w_{\text{some dirty}} \approx w_{\text{all dirty}}$) because the situation is not okay in either case (Table 2). (20b) is thus "true enough" in the mixed $w_{\text{some dirty}}$ scenario, so (20b) can receive an existential interpretation. As noted above, speakers are more hesitant to give judgments in such scenarios when the cups are "mixed" with respect to cleanliness, so the existential interpretation is more fragile ("varied and guarded"; Rooth 1987) than the universal one. But this analysis at least explains why this existential interpretation is more likely for negative-sentiment adjectives.

This proposed explanation begins by positing "Is the current situation okay?" as the QUD. One might wonder, what if the QUD instead asked about the mathematical facts of the situation (21a)? In fact, (21b) is a very awkward response to that (explicit) QUD, perhaps because the answer to the

The cups are clean (situation is good).		**The cups are dirty** (situation is bad).	
$w_{\text{all clean}}$: True		$w_{\text{none dirty}}$: False	
$w_{\text{some clean}}$: False-ish $w_{\text{none clean}}$: False	$w_{\text{some}} \approx w_{\text{none}}$	$w_{\text{some dirty}}$: True-ish $w_{\text{all dirty}}$: True	$w_{\text{some}} \approx w_{\text{all}}$

Table 2: Visualization of the proposed analysis in the domain of individuals denoted by definite plurals.

QUD typically carries prosodic focus (Roberts 2012), whereas the quantificational effect of (21b) cannot bear focus because it arises implicitly.

(21) a. How many (what percent) of the cups are clean?
b. #The cups are clean.

This explanation is intended to capture the behavior of *clean/dirty* in a typical context in which *clean* is desirable, *dirty* undesirable, consistent with the sentiment-linked dimension of their meaning. For example, while setting up for a party, we want clean cups ready for guests. Such typical contexts are most likely to be imagined by hearers/readers out of context, explaining the inference patterns observed by Yoon (1996). But in a highly atypical context such as (22) in which dirtiness is for some reason desired (inspired by Lasersohn 1999, who also discusses the interpretation of definite plurals in medical testing contexts), then we predict the inferences to flip: *dirty* should be interpreted as universal, while *clean* should be more likely to receive an existential interpretation.

(22) Context: We are running a medical test in which patients are asked to drink juice, followed by measurements of their blood sugar. We hope that all the cups are dirty and juice-stained, because it's important that all the patients have drunk the juice. If any cups are clean, it means the nurse has made a mistake.

a. There's been a mistake; the cups are **clean**! $\rightsquigarrow$ **Some/most** of the cups are clean.
b. Everything is fine; the cups are **dirty**. $\rightsquigarrow$ **All** of the cups are dirty.

Such contextual variability is predicted on any theory that derives the interpretation of definite plurals from pragmatic reasoning about the QUD. Originally, Krifka's point about the importance of the QUD was cast in tension with Yoon's observation about the role of the lexical predicate. But on my proposal, both elements work together: in typical contexts, the type-level sentiment of the lexical predicate usually determines the desirability of the property described by a given token of that adjective, which in turn interacts with the Anna Karenina Principle to shape the quantificational force of the answer to the QUD about whether the current situation is okay. This analysis captures the *clean/dirty* contrast observed by Yoon (1996) in typical contexts as well as its exceptions (22) in highly atypical ones.

Beyond maximum- versus minimum-standard pairs such as *clean/dirty*, the same explanation is predicted to extend to relative-standard adjectives as in (23), for example in a context where a CEO is asking how ten employees feel about a new policy. Here too, the positive-sentiment adjective answers "Yes" to the question of whether the current situation is okay, which – according to the Anna Karenina Principle – means it is *all* okay, favoring a universal interpretation. The negative-sentiment version answers "No" to that question, which – according to the Anna Karenina Principle – means it is at least *somewhat* bad, yielding a greater potential for an existential interpretation.

(23) a. The employees are **happy**. $\rightsquigarrow$ **All** of the employees are happy.
b. The employees are **upset**. $\rightsquigarrow$ **Some** of the employees are upset.

Beyond adjectival antonym pairs, the same logic extends to predicates of any lexical category that can be divided into desirable/undesirable pairs, either based on the sentiment of their type-level meaning (*succeed/fail, win/lose, love/hate*, *spare/damage* – the latter of which is mentioned by Yoon 1996, p. 217) or on the particular discourse context of a given token (*walk/run* in a situation where one or the other is desired). While the overall prediction is broader, this paper focuses on adjectives to explain the original data from Yoon (1996), and to use antonym-swapped minimal pairs to explore at scale how various inferences are grounded in the desirability of the properties that they describe.

But when two antonyms do not clearly differ in the desirability of the property that they describe, the Anna Karenina Principle cannot explain their interpretation. Both *awake* and *asleep* have middling ratings in the Hedonometer sentiment lexicon of Dodds et al. (2011, 2015) (5.74 and 6.14, respectively, on a 1-to-9 scale). Here, it may be less clear out of context which antonym describes a more desirable property, and thus less clear how each half of the antonym pair would address the QUD "Is the current situation okay?", yielding no clear prediction about how people will interpret such sentences. In other words, my proposed analysis of Yoon's *clean/dirty* contrast makes predictions about the hundreds of predicate pairs that clearly diverge in sentiment, but not about pairs describing properties whose desirability is indistinguishable. In such cases, one would need to know which one is more desirable in context.

In sum, we predict:

(24) **Hypothesis: Sentiment-driven quantification across individuals**
Sentences of the form *The Xs are P* should tend to:
a. Favor a universal interpretation (*All of the Xs are P*) when *P* is positive in sentiment.
b. Be more likely to allow an existential interpretation (*Some-but-not-all of the Xs are P*) when *P* is negative in sentiment.

The statistical effect may be subtle, because the universal interpretation is a strong default; because the existential interpretation describes a "mixed" scenario in which speakers hesitate to provide clear judgments; because human judgment studies are almost never categorical in the first place; because some antonym pairs (*awake/asleep*) do not clearly differ in sentiment; and because speakers may be somewhat uncertain about the implicit QUD. But the direction of the effect is predicted clearly by the Anna Karenina Principle in combination with pragmatic reasoning about how the sentence answers the QUD of whether the current situation is okay or not.

3.3 The Anna Karenina Principle across degrees

Beyond individuals, the Anna Karenina Principle also yields predictions about the inferential behavior of adjectives in the domains of degrees and dimensions, this time through the mechanism of lexicalization – the mapping between concepts and word types (Talmy 1985; Brinton & Traugott 2005; Levin & Rappaport Hovav 2019) – rather than pragmatic reasoning about how the utterance of a particular token answers the QUD.

3.3.1 Proposed analysis of degree semantics of *clean/dirty*

Regarding degrees, Cruse (1980) observes that there are adjectives describing "all okay" scenarios (*clean*) and "somewhat bad" scenarios (*dirty*) with respect to degrees of cleanliness, but no words describing "somewhat okay" scenarios (possessing at least some cleanliness). To the extent that *filthy* can be modified by maximum-standard adverbs such as *perfectly* or *utterly* (the latter of which is preferred for adjectives describing undesirable properties; Louw 1993, p. 160), perhaps this word describes maximal filth, but a word meaning 'minimally clean' seems to be truly absent. This observation aligns with Paradis (1997, p. 91), Bylinina & Zadorozhny (2012, p. 36), and Bogal-Allbritten (2012, p. 80), who find that minimum-standard modifiers often combine with adjectives describing undesirable properties: *a bit awkward; slightly tall* in a context where height is unwanted.[8]. Minimum-standard modifiers are associated with undesirable properties ("somewhat bad").

As noted above (Table 1), there are also maximum/maximum-standard pairs such as *full/empty*, for which both poles of the scale are closed; and relative/relative pairs such as *tall/short* and *happy/sad*, for which both poles extend indefinitely. These pairs also diverge in sentiment, but the two poles match in their quantificational force across degrees, so there is no difference here to be explained. As for the pair *famous/unknown*, the negative-sentiment antonym (*unknown*) can be modified by maximum-standard adverbs (*totally unknown*) while the positive-sentiment pair-

[8] Bylinina & Zadorozhny (2012) suggest that "negative evaluative" adjectives such as *lazy* combine with minimizers for politeness reasons, to downplay the social face threat associated with unflattering judgments.

mate *famous* extends indefinitely. While *clean/dirty* represents a maximum/minimum-standard pair, *famous/unknown* seems to be a relative/maximum-standard pair, to which the Anna Karenina Principle also does not clearly apply.[9]

The precise semantic analysis (discussed in Section 2.2) does not matter as much as the observation that different adjectives are associated with different scale structures, which drives differences in their interpretation and the adverbial modifiers available to them. Whatever semantics one chooses, the Anna Karenina Principle offers a partial answer to a question that otherwise remains idiosyncratic: which adjectives are associated with what sorts of scales and why.

As above, I argue that *The cup is clean* is easily taken to provide an affirmative answer to the common QUD of whether the current situation is okay. Following the Anna Karenina Principle, the current situation is only okay if it is all (maximally) okay, meaning that the cup is perfectly clean. With respect to this QUD, a scenario in which the cup is only somewhat clean is not equivalent to one in which it is perfectly clean ($w_{\text{somewhat clean}} \not\approx w_{\text{maximally clean}}$; Table 3). Therefore, *The cup is clean* is interpreted to describe a maximal degree of cleanliness.

On the other hand, *The cup is dirty* would provide a negative answer to the common QUD of whether the current situation is okay. Following the Anna Karenina Principle, the current situation is not okay if there's anything wrong, i.e., any dirt on the cup. With respect to this QUD, a scenario in which the cup is somewhat dirty is equivalent to one in which it is maximally dirty ($w_{\text{somewhat dirty}} \approx w_{\text{maximally dirty}}$; Table 3). So *The cup is dirty* can be used when the cup has any dirt on it.

The cup is clean (situation is good).		**The cup is dirty** (situation is bad).	
$w_{\text{maximally clean}}$: True		$w_{\text{zero dirt}}$: False	
$w_{\text{some clean}}$: False-ish $w_{\text{zero clean}}$: False	$w_{\text{some}} \approx w_{\text{zero}}$	$w_{\text{some dirt}}$: True-ish $w_{\text{maximally dirty}}$: True	$w_{\text{some}} \approx w_{\text{max}}$

Table 3: Visualization of the proposed analysis in the domain of degrees.

Above, this line of reasoning was used to explain the interpretation of adjective tokens predicated of definite plurals, which are known to be sensitive to the QUD. Here, I use the same reasoning to explain the scale structure of adjective types, abstracted across their specific token contexts. This explanation assumes that the lexicon evolves to encode the distinctions that speakers regularly find useful. The word *electrification* emerged as speakers found it useful to talk about proliferating

[9] As noted above (Section 2.3), *un-* tends to combine with positive-sentiment adjectives to form negative-sentiment antonyms (*unhappy* but not **unsad*; Zimmer 1964; Horn 1989). But *un-* is more productive with past participles (and verbs suffixed with *-able*) regardless of their sentiment/desirability (*unknown, unreadable*; Horn 1989, p. 275). In general, *unhappy*-type *un-* adjectives show relative scale structure (*very/?perfectly unhappy*), whereas those formed from participles and verbs suffixed with *-able* tend to describe maximum standards (*perfectly/totally unknown, perfectly/totally unreadable*). In other words, *unknown* shares the *un-* prefix with negative-sentiment adjectives such as *unhappy*, and shares a maximum-standard scale structure with positive-sentiment adjectives such as *clean*.

electrical machines (Quirk et al. 1985, p. 1526). The word *headache* exists but **thighache* does not (Brinton & Traugott 2005, p. 47) because speakers find it more useful to discuss common complaints than uncommon ones. The word *planet* originally described any non-stagnant celestial object, but evolved to denote only the large bodies orbiting the sun (Ludlow 2014, p. 41) as people found it useful to draw distinctions from advancing science. Using this logic, I suggest that the antonyms *clean/dirty* have evolved to describe the distinctions that people find useful according to the Anna Karenina Principle combined with the common QUD of whether everything is okay: that is, to distinguish full cleanliness (all okay) from any dirt (something wrong). In contrast, a pair of antonyms such as *filthy/clean-ish*, distinguishing maximal filth (all bad) from any cleanliness (some good), is conceptually possible but less useful.

3.3.2 An alternative framing in terms of vagueness

An anonymous reviewer suggests an alternative analysis of these facts. Imagine that an object might possess any quantity of dirt ranging from zero to infinity, similar to height, so that *clean/dirty* are vague just like *tall/short*. The question would be how we choose thresholds for *clean* versus *dirty*, and the Anna Karenina Principle would ground a theory of vagueness (alongside those proposed by Kennedy 2007, Lassiter & Goodman 2013, Burnett 2014, and so on to explain why we often set the threshold for *clean* at zero dirt, and the one for *dirty* at some dirt. Namely, the Anna Karenina Principle says these thresholds are useful for separating the desirable state of affairs (zero dirt) from undesirable ones (any amount of dirt). The Anna Karenina Principle would thus aim to explain vagueness (i.e., setting thresholds) rather than scale structure (shaping the concept of cleanliness/dirt). By framing *clean* as vague, perhaps this analysis would more easily capture the cases (described as imprecise by Kennedy 2007, p. 23) where *clean* describes a non-maximal degree, for example when a parent accepts their teenager's hastily-tidied room as *clean* (enough) even though it could be cleaner.[10]

On this analysis, perhaps the thresholds for *short/tall* aren't set at zero versus any amount of height because possessing zero height, for a physical entity, is usually neither plausible nor desirable.

But this analysis is largely synonymous to the one I propose: both leverage the Anna Karenina Principle to ground a utility-based explanation of why we need concepts/scales, words, and/or consistent thresholds to distinguish all-good versus any-bad situations, that is, all-clean from some-dirt rather than the reverse (all-dirt versus some-clean). That is my main claim.

[10] As a less promising alternative analysis, one might imagine that the concept of cleanliness/dirt is already anchored to zero at one end, distinguishing zero from some dirt. Given that dirt is undesirable, this analysis assumes, but does not explain, that the zero point is the desirable one.

3.3.3 From *slightly dirty* to *slightly sad*

Beyond *clean/dirty*, I suggest that this analysis also illuminates the less-discussed scale structure of subjective adjectives describing value and human propensity in the terms of Dixon (2004, p. 3-4), or what Bierwisch (1989, p. 88) calls "evaluative" adjectives: *happy/sad; beautiful/ugly*.

Bylinina & Zadorozhny (2012, p. 36) and Bogal-Allbritten (2012, p. 80) note that "negative evaluative adjectives" easily combine with minimum-standard modifiers (*slightly sad/ugly*), suggesting a lower-bounded scale structure like the one associated with *dirty*. But while *dirty*'s antonym *clean* represents a prototypical maximum-standard adjective, the scale structure of positive-sentiment evaluative adjectives such as *happy* and *beautiful* is far less clear. Is there an upper bound to these properties, or do they extend indefinitely? Empirically, such adjectives are at least sometimes modified by maximum-standard modifiers (*perfectly happy*), which might frame them as maximum-standard adjectives like *clean*. On the other hand, *clean* and *dirty* arguably exhaust the logical space (i.e., with no in-between), whereas *happy/sad* pattern like prototypical relative adjectives such as *tall/short* in allowing a "zone of indifference" (Sapir 1944, p. 98) of individuals who are neither happy nor sad. So we find conflicting evidence about the scale structure of pairs such as *happy/sad*.

To make sense of these facts, I build on the analysis by Beltrama (2025) of mildly positive evaluations, such as *adequate* and *decent*. Beltrama (2025, p. 191) says these adjectives convey that an entity exceeds a minimal "functional" standard for what counts as good enough for current purposes (drawing on Kagan & Alexeyenko 2011 and Nadathur 2023). Although Beltrama does not emphasize this point, this "minimal" functional standard, ironically, shows *maximum*-standard modification behavior: {*perfectly/??slightly*} *adequate*.

Broadening Beltrama's proposal, I suggest that *all* positive-sentiment adjectives can (but need not always) be interpreted to evoke a functional standard for what counts as good enough in context. This interpretation is compatible with maximum-standard modifiers, because a good-enough situation is "all good" according to the Anna Karenina Principle and thus maximal in some sense. Conversely, negative-sentiment adjectives can be taken to convey that an entity falls short of this functional standard, and can thus be modified by minimum-standard modifiers because, per Anna Karenina, even a minimal deviation from "good-enough" counts as bad.

Assume in Figure 4 that I have to be pretty happy in order to be happy enough for current purposes. An intermediate situation (neither happy nor sad)[11] is not equivalent to one in which I'm happy ($w_{\text{intermediate}} \not\approx w_{\text{pretty happy}}$) because the QUD, "Is the current situation okay?", is answered differently in each case. In contrast, this intermediate situation *is* equivalent to a situation in which

[11] In Table 4, *I am happy* and *I am sad* are taken to be indeterminate in the situation where the speaker is neither happy nor sad, reflecting the longstanding idea – e.g., Kennedy (2007, p. 2) – that vague adjectives yield "borderline" cases.

I'm pretty sad ($w_{\text{intermediate}} \approx w_{\text{pretty sad}}$) because in both cases, the QUD is answered negatively: the situation is not okay. Assuming that negative-sentiment predicates are taken to convey that the situation is not okay, *slightly sad* characterizes this unsatisfactory intermediate zone in a way that is "true-enough" with respect to this QUD, while *slightly happy* does not.

I am happy enough (situation is good).		**I am sad** (situation is bad).	
$w_{\text{extremely happy}}$: True $w_{\text{pretty happy}}$: True		$w_{\text{extremely happy}}$: False $w_{\text{pretty happy}}$: False	
$w_{\text{intermediate}}$: False-ish $w_{\text{pretty sad}}$: False $w_{\text{extremely sad}}$: False	$w_{\text{intermediate}} \approx w_{\text{sad}}$	$w_{\text{intermediate}}$: True-ish $w_{\text{pretty sad}}$: True $w_{\text{extremely sad}}$: True	$w_{\text{intermediate}} \approx w_{\text{sad}}$

Table 4: Visualization of the proposed analysis applied to *happy/sad*.

In sum, my proposed analysis of sentiment-linked quantification among degrees extends beyond prototypical maximum/minimum-standard pairs such as *clean/dirty*, to further encompass adjectives with more debatable scale structure. Here too, I suggest that negative-sentiment adjectives combine more easily with minimum-standard modifiers (*slightly sad*), because it is useful to distinguish fully satisfactory levels of desirable properties (all good) from even slightly unsatisfactory ones (any bad).

3.3.4 Predicted typology of scale structure by sentiment

This analysis captures Cruse's observation that if a pair is divided into maximum/minimum standard pair-mates, then it is the minimum-standard antonym that describes the undesirable property (*slightly dirty*), rather than the reverse. It also captures the observation from Bogal-Allbritten (2012) and Bylinina & Zadorozhny (2012) that minimum-standard modifiers often appear with negative-sentiment adjectives more generally (*slightly sad*). Empirically, we thus predict:

(25) **Hypothesis: Sentiment-driven quantification across degrees**
Minimum-standard adjectives – that is, those modified by minimum-standard modifiers (*slightly, partially*) – should tend to be negative in sentiment.

As illustrated in Table 5, the prediction only runs one way. We do not predict the reverse, that negative-sentiment adjectives all describe minimum standards; in fact, some negative-sentiment adjectives describe relative standards (*slow*) or maxima (*empty*). As for positive-sentiment adjectives, these may describe maximum standards (*clean, full*) or relative standards (*fast*), just not minima:

But among minimum-standard adjectives (bottom row of Table 5), the prediction is clear: those describing *any* bad are more useful, and thus more likely to be lexicalized, than hypothetical alter-

	Positive-sentiment	**Negative-sentiment**
Relative	*(✗slightly, ✗perfectly) fast*	*(✗slightly, ✗utterly) slow*
Max-standard	*(✗slightly, ✓perfectly) clean, full, healthy, happy*	*(✗slightly, ✓utterly) filthy, empty, starving, moribund, devastated, distraught*
Min-standard	[I predict to be absent]	*(✓slightly, ✗utterly) dirty, hungry, sick, sad*

Table 5: Typology of scale structures for positive- versus negative-sentiment adjectives, as identified by their potential to combine with different modifiers. I use *perfectly* as a maximum-standard modifier for positive-sentiment adjectives, and *utterly* for negative-sentiment ones.

natives describing *any* good.

3.4 The Anna Karenina Principle across dimensions

Regarding dimensions, the same lexicalization pressure would predict that speakers find it useful to discuss maximal health (across all dimensions) versus its absence (on any dimension), but might find it less useful to discuss health along at least one dimension. Here too, assuming that the lexicon is shaped by communicative utility, it is argued to encode the distinctions that speakers find most useful following the Anna Karenina Principle combined with the common QUD of whether things are okay.

Sassoon (2013) suggests that her "conjunctive" adjectives such as *healthy* are also maximum-standard adjectives, while their "disjunctive" antonyms such as *sick* are minimum-standard adjectives. She offers (in her Section 1.3.2) various reasons that adjectives show matching quantificational force across both degrees and dimensions (universal/maximal for *healthy*, minimal/existential for *sick*), but she does not invoke the desirability of the property denoted by the adjective, nor does she explain which adjectives would show which type of quantificational force and why.

Sassoon (2013) also notes that some relative-standard adjectives can vary contextually as to whether they quantify universally/conjunctively or existentially/disjunctively across dimensions: a person can be *intelligent/stupid* in every way, or *intelligent/stupid* in a single area such as mathematics. Just as these adjectives pick out a vague, relative standard for what counts as intelligent/stupid in context, they also vary contextually in their quantificational force across dimensions. But here, because both antonyms equally can be conjunctive or disjunctive, there is no asymmetry to be explained.

Regarding quantification across dimensions, the prediction stemming from the Anna Karenina Principle thus applies only to the small handful of adjectives such as *healthy/sick* for which

one antonym describes a maximal degree of a multidimensional property such as health – not a unidimensional one such as cleanliness or height, nor one that can be multidimensional or unidimensional depending on the context, such as intelligence/stupidity. The semantic representation (sketched in Section 2.2) does not matter as much as the phenomenon itself: that some adjectives quantify universally/conjunctively across dimensions, while their antonyms quantify existentially/disjunctively.

Representing this limited set of adjectives, *My grandma is healthy* is easily taken to provide an affirmative answer to the common QUD of whether the current situation is okay. Following the Anna Karenina Principle, the current situation is only okay if it is *all* okay; that is, Grandma is healthy along all dimensions. With respect to this QUD, a scenario in which Grandma is healthy across all dimensions is not equivalent to one in which she is healthy on some dimensions but sick on another ($w_{\text{all healthy}} \not\approx w_{\text{some health}}$) because the QUD is resolved differently in these scenarios. Therefore, *My grandma is healthy* conveys that she is healthy along all dimensions.

In contrast, *My grandma is sick* is taken to provide a negative answer to the common QUD of whether the current situation is okay. Following the Anna Karenina Principle, the current situation is not okay if anything is wrong. Here, the QUD is resolved in the same way in a scenario in which Grandma is sick along some dimension as when she is sick along all dimensions ($w_{\text{some sickness}} \approx w_{\text{all sick}}$; Table 6). So *My grandma is sick* conveys that she is sick along at least some dimension.

Grandma is healthy (situation is good).		**Grandma is sick** (situation is bad).	
$w_{\text{all healthy}}$: True		$w_{\text{no sickness}}$: False	
$w_{\text{some health}}$: False-ish $w_{\text{no health}}$: False	$w_{\text{some}} \approx w_{\text{none}}$	$w_{\text{some sickness}}$: True-ish $w_{\text{all sick}}$: True	$w_{\text{some}} \approx w_{\text{all}}$

Table 6: Visualization of the proposed analysis in the domain of dimensions.

On this analysis, conjunctive positive-sentiment adjectives and disjunctive negative-sentiment ones are therefore more useful than their alternatives (hypothetical conjunctive negative-sentiment adjectives or disjunctive positive-sentiment ones), thus more worthy of being lexicalized. We therefore predict:

(26) **Hypothesis: Sentiment-driven quantification across dimensions**

a. Multidimensional conjunctive adjectives should tend to be positive in sentiment.

b. Their disjunctive antonyms should tend to be negative in sentiment.

In sum, when combined with pragmatic reasoning about how sentences answer the common QUD of whether the current situation is okay, the Anna Karenina Principle grounds a series of hypotheses that aim to explain why sentiment shapes quantification among adjectives across individuals, degrees, and dimensions. The next section tests these predictions empirically. To foreshadow,

these hypotheses are all indeed consistent with the data reported below.

4 Testing the hypotheses

Across individuals, degrees, and dimensions, this section tests the predictions set out above. The key independent variable of lexical sentiment is measured through the Hedonometer lexicon of human annotations gathered by Dodds et al. (2011, 2015), ranging from 1 (most negative) to 9 (most positive). *Clean* has a rating of 6.9, versus 3.2 for *dirty*. All data and code are shared through the Open Science Framework at `https://osf.io/9x4ks/`. The Appendix to this paper also provides full lists of the adjectives used in these studies.

4.1 Quantification over individuals

In the domain of individuals, I elicited binary-choice inference judgments of the form in (27) – where each item was randomly assigned to one half or the other of the antonym pair (*clean* or *dirty*):

(27) Your friend Ian is working with ten different cups.
You ask Ian: "Are the cups {clean/dirty}?"
Ian says, "Yes, the cups are {clean/dirty}."
What do you infer?

a. Some of them are {clean/dirty}.
b. All of them are {clean/dirty}.

These items were constructed from 55 adjectival antonym pairs (110 adjectives) drawn from the literature (Lehrer 1985; Yoon 1996; Rotstein & Winter 2004; Kennedy 2007; Kamoen et al. 2011), given in full in the Appendix. These adjectives comprise not just maximum/minimum-standard pairs such as *clean/dirty*, but also relative/relative (*happy/sad*) and maximum/maximum pairs (*full/empty*). Any antonym pair mentioned in the semantics literature was included, as long as both halves of the pair have sentiment ratings in the Hedonometer lexicon. I excluded pairs including quite rare words, such as *ferocious/meek* (mentioned by Lehrer 1985) and *pure/impure* (Rotstein & Winter 2004; Kennedy 2007), because at least one member was absent from the Hedonometer.

For each antonym pair, I chose a plural noun that makes sense with both halves of the antonym pair as something that an individual might be responsible for keeping in good condition – typically artifacts (*clean/dirty* predicated of *cups*) and social roles (*happy/sad* predicated of *customers*). As in Yoon's original example, (27) thus evokes the QUD of whether the cups are in good condition

or not – that is, whether the current situation is okay. This QUD is not stated explicitly, because my hypothesis is that it is commonly inferred out of context, as in Yoon's original example.

These items were intermixed with fourteen fillers using sentiment-neutral non-gradable adjectives. Seven used colors such as *orange*, seven used materials such as *wooden*. Combining the 55 target items with these fillers, there were 69 questions overall. To prevent fatigue, each experimental participant judged 40 randomized items, of which approximately 8 (20%) were fillers.

With approval from the Institutional Review Board of the author's institution, these items were presented in a random order (with target items randomly assigned to one half or the other of the antonym pair) to 40 participants (self-identified native English speakers located in the United States) recruited for pay on the Prolific web platform.

To test the hypothesis (28), I ran a mixed-effects logistic regression using the lme4 package (Bates et al. 2015) in R (R Core Team 2012) predicting binary response (*all* or *some*) as a function of sentiment, using random intercepts for each participant and each antonym pair.

(28) **Hypothesis: Sentiment-driven quantification across individuals** (= (24))
Sentences of the form *The Xs are P* should tend to:

a. Favor a universal interpretation (*All of the Xs are P*) when *P* is positive in sentiment.
b. Be more likely to allow an existential interpretation (*Some of the Xs are P*) when *P* is negative in sentiment.

As predicted (Figure 1), adjectives with higher sentiment (*clean*) are more likely to receive the universal *all* interpretation ($\beta = 0.14, z = 2.60, p < 0.01$), meaning that every one-point increase in sentiment corresponds to a 15% increase (i.e., exp(0.14)=1.15) in the odds of a universal interpretation. The universal interpretation is a strong default overall, chosen in 80% of all target items, reflecting the fact (Section 2.1) that readers/hearers prefer to imagine all members of the definite plural to be homogeneous with respect to the property described by the adjective, and are reluctant to assign "mixed" interpretations. The effect size of the predicate's sentiment is relatively small (Figure 1), especially in comparison to the huge variability found across participants (random effect variance = 7.94 on a log-odds scale), again perhaps because people resist interpretations describing "mixed" scenarios and because they may be somewhat uncertain about the QUD or how it is answered by adjectives of middling sentiment. But the default universal interpretation is less likely for negative-sentiment adjectives, as predicted by (28).

Overall, the data are consistent with this paper's proposed explanation for Yoon's *clean/dirty* contrast: that the sentiment of an adjective combines with the Anna Karenina Principle to shape our expectations about how an assertion answers the QUD of whether the current situation is okay.

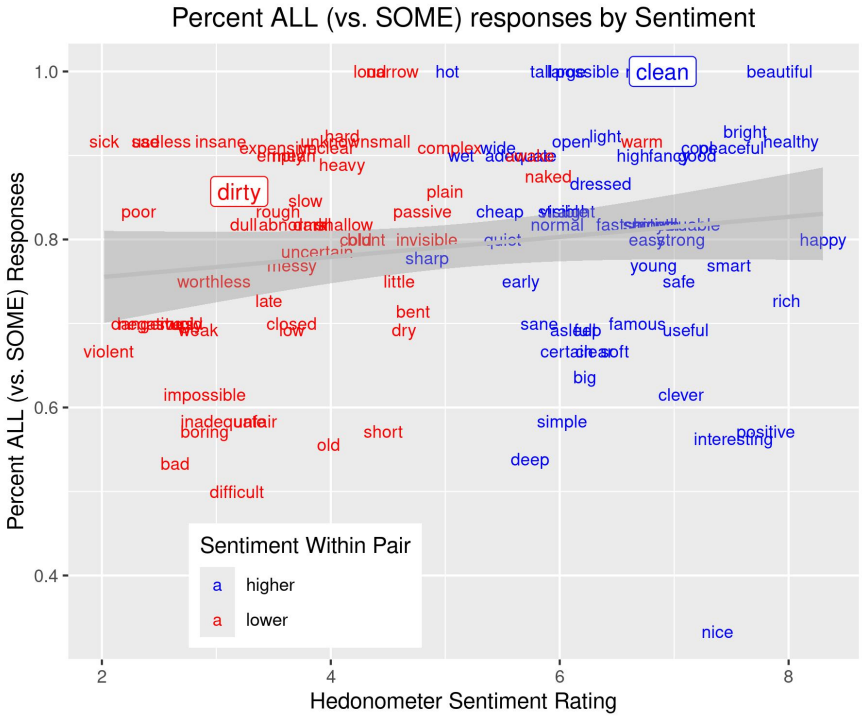

Figure 1: Percentage of *all* (versus *some*) responses for each predicate as a function of sentiment.

4.2 Quantification over degrees

In the domain of degrees, I explored modifiers of the same 55 antonym pairs (110 adjective types) in corpus data. Beyond a few clear headline examples such as *clean/dirty*, it is not easy to label these adjective types for whether their associated scales are minimum-standard (*dirty*), maximum-standard (*clean*), or open (*tall*). Studying the acceptability of adjectives with modifiers said to indicate maximum-standard scale structure (e.g., *almost/completely easy*), Kamoen et al. (2011) find that acceptability varies widely depending on the surrounding textual context of the token, which makes it difficult to label decontextualized adjective types. If scale structure were labeled by the author, the data might be further confounded by confirmation bias. So instead of labeling scale structure by hand, I took inspiration from Sassoon (2013) in approximating it from the frequency of modifiers in corpus data.

Using one hundred million words of English text from the AskReddit web forum from January 2018 (Baumgartner et al. 2020)[12], I extracted all tokens (tagged as adjectives by the part-of-speech tagger built into the NLTK Natural Language Toolkit of Loper & Bird 2002) of these 110 adjective types. For each adjective token, I checked its syntactic dependents according to a dependency parser (Honnibal & Johnson 2015) and recorded its modifiers: dependents part-of-speech-tagged

[12]It is more difficult to access Reddit data now that Reddit wants to be paid for data used to train large language models. I used data from January 2018 because I had downloaded it while it was freely available for researchers.

as adverbs; those linked to the adjective by the *npadvmod* dependency to capture *(a) bit*, which is tagged as a noun phrase; and the adjective-modifying adjective *little*. Of all modifiers, I counted those that the literature classifies as indicating a maximum, minimum, or relative standard (Rotstein & Winter 2004; Kennedy & McNally 2005; Kennedy 2007; Kamoen et al. 2011; Sassoon 2013):

(29) a. **Maximum-standard modifiers:** *absolutely, almost, completely, entirely, fully, half, nearly, perfectly, totally, utterly*
b. **Minimum-standard modifiers:** *(a) bit, (a) little, partially, partly, slightly*
c. **Relative-standard modifiers:** *extremely, fairly, incredibly, pretty, quite, really, relatively, so, very*

Adapting an idea from Sassoon (2013), I computed the number of minimum-standard modifiers as a percentage of all (maximum-, minimum-, and relative-standard) modifiers. There are 24 tokens of *dirty* with minimum-standard modifiers, out of 239 tokens of *dirty* with any of the modifiers listed in (29), so *dirty* appears with minimum-standard modifiers 24/239=10% of the time (versus 5/440=1% for *clean*). For comparison, I also computed the number of maximum-standard modifiers as a percentage of all modifiers (82/440=19% for *clean*, 2/239$\leq$1% for *dirty*).

To test the hypothesis (30), I conducted a linear regression predicting the percentage of minimum-standard modifiers as a function of sentiment.

(30) **Hypothesis: Sentiment-driven quantification across degrees** (=(25))
Minimum-standard adjectives – that is, those modified by minimum-standard modifiers (*slightly, partially*) – should tend to be negative in sentiment.

As predicted, positive-sentiment adjectives are less often modified by minimum-standard modifiers (Figure 2; $\beta = -1.6\%, t(108) = -3.53, p < 0.001$, adjusted $R^2 = 0.095$). In other words, every one-point increase in a predicate's sentiment means that its percentage of minimum-standard modifiers is predicted to decrease by 1.6 on a 100-point scale (the meaning of β), and sentiment explains almost ten percent of the variability in the percentange of minimum-standard modifiers that we find for each predicate (the meaning of adjusted R^2).

The hypothesis (30) makes no predictions about the behavior of *maximum*-standard modifiers. And indeed (Figure 2), in another linear regression predicting the percentage of such maximum-standard modifiers as a function of sentiment, sentiment has no significant effect; some adjectives favoring maximum-standard modifiers are positive in valence (*clean*) while others are negative (*worthless, impossible, empty*).

Overall, the data are consistent with the hypothesis (30) that the lexicon is shaped by communicative utility, as shaped in turn by the Anna Karenina Principle in combination with the common

QUD of whether the current situation is okay.

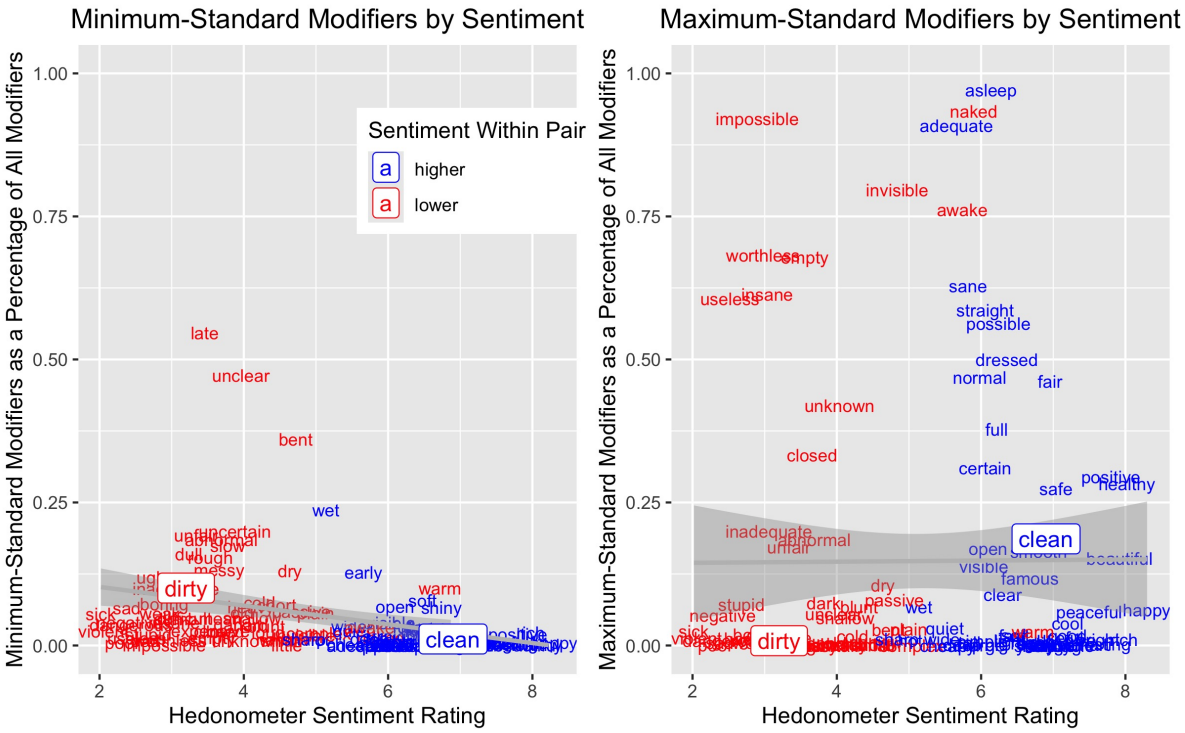

Figure 2: Frequency of minimum- and maximum-standard modifiers for each predicate as a function of sentiment.

4.3 Quantification over dimensions

In the domain of dimensions, I gathered six adjectives (*good, healthy, identical, normal, typical, similar*) that Sassoon (2013) deems to be conjunctive multidimensional adjectives, and four more (*bad, sick, abnormal, different*) that she classifies as disjunctive[13]. *Healthy* is conjunctive, quantifying universally over all dimensions of health (all organs, all biometrics), while *sick* is disjunctive, quantifying existentially over any dimension thereof.

To test (31), I compared the sentiment ratings of Sassoon's conjunctive adjectives versus disjunctive adjectives.

(31) **Hypothesis: Sentiment-driven quantification across dimensions** (=(31))

[13]Some of Sassoon's adjectives – *atypical, dissimilar* – had to be excluded because they do not appear in the Hedonometer sentiment ratings. Moreover, Sassoon defines conjunctivity and disjunctivity from corpus data, so that a conjunctive adjective is one that frequently occurs in the construction *ADJ except DIMENSION* (e.g., *healthy except for blood pressure*), and a disjunctive adjective is one that favors the construction *not ADJ except DIMENSON* (*not sick except a cold*). She finds that *similar* and *good* usually occur in constructions associated with the conjunctive interpretation, but also occur to some extent in those associated with a disjunctive one.

a. Multidimensional conjunctive adjectives should tend to be positive in sentiment.
b. Their disjunctive antonyms should tend to be negative in sentiment.

As predicted (Figure 3), the conjunctive adjectives such as *healthy* are far higher in sentiment (mean = 6.35 out of 9) than the disjunctive ones such as *sick* (mean = 3.43). These results are significant in Welch's (Welch 1947) unpaired[14], two-sided t test ($t = 3.43$ with 5 degrees of freedom; $p = 0.019$).

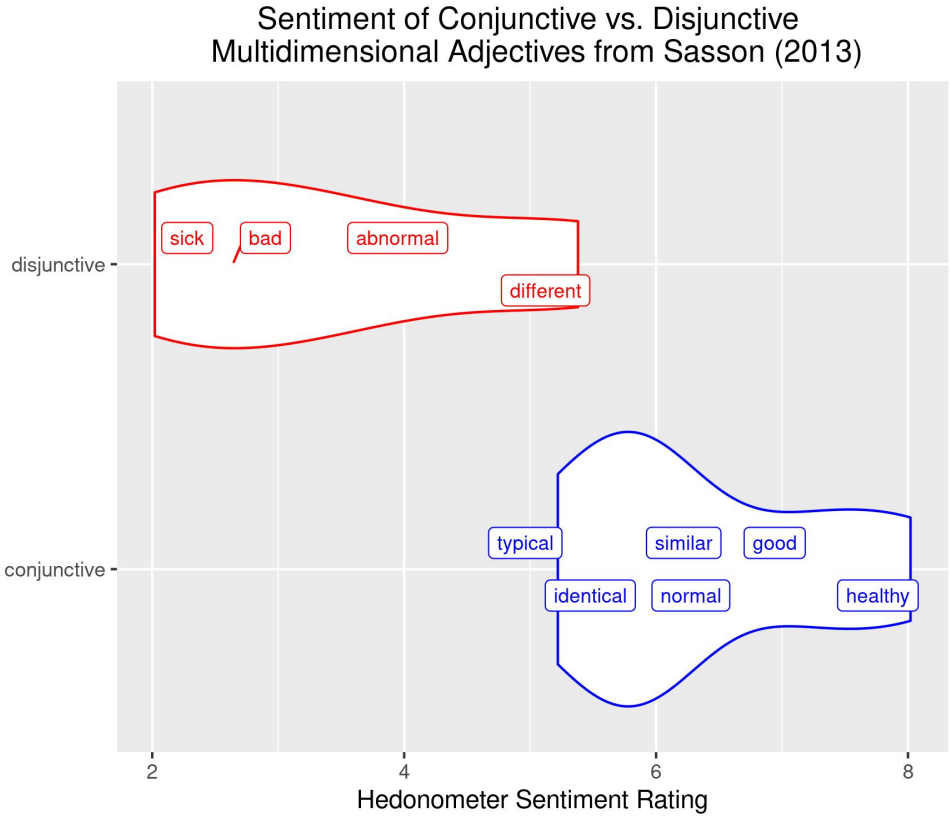

Figure 3: Sentiment of multidimensional adjectives classified by Sassoon (2013) as conjunctive versus disjunctive.

These findings are consistent with the hypothesis (31) that among adjectives describing multidimensional properties, the lexicon encodes the distinctions that are most useful according to the Anna Karenina Principle alongside the common QUD, "Is the current situation okay?"

[14]I ran an unpaired t test rather than a paired one in order to keep data from adjectives such as *similar* and *typical* whose antonyms (*dissimilar, atypical*) do not appear in the Hedonometer ratings.

4.4 Discussion

Aiming to explain Yoon's *clean/dirty* contrast in the realm of individuals described by definite plurals, this paper leverages the Anna Karenina Principle, alongside the QUD of whether the current situation is okay, to predict that positive-sentiment adjectives should be interpreted as universal while negative-sentiment adjectives should be more likely to allow an existential interpretation. Consistent with this hypothesis, a judgment study of 55 antonym pairs (110 adjective types) finds that existential interpretations of definite plurals are more likely for negative-sentiment adjectives: that is, *The cups are dirty* is more likely to be taken to describe some-but-not-all cups.

The proposed explanation for Yoon's *clean/dirty* contrast makes predictions that extend beyond definite plurals. Among degrees, I find (among the same 110 adjective types) that negative-sentiment adjectives are more often modified by minimum-standard modifiers in corpus data (e.g., *slightly dirty*). These results are consistent with the suggestion from Cruse (1980) that the lexicon encodes the distinctions that are most useful according to the Anna Karenina Principle with respect to the common QUD of whether the current situation is okay. Among dimensions, I find (in a study of ten adjective types) that the adjectives that are interpreted as conjunctive/universal (*healthy*) are positive in sentiment, while those that are interpreted as disjunctive/existential are more negative (*sick*). These results are consistent with the same hypothesized lexicalization pressure.

In other words, the Anna Karenina Principle is argued to shape both token-level pragmatic inferences (the interpretation of definite plurals), *and* type-level semantic features (the distinctions between maximum- versus minimum-standard adjectives, and between conjunctive versus disjunctive multidimensional ones). It is a point in favor of the proposed analysis that it explains such diverse phenomena.

Beyond adjectives, the same reasoning might explain observations from Leslie (2008) about generic sentences describing dangerous properties, as when (32) is judged true even though only a tiny percentage of mosquitoes are infected with West Nile.

(32) Mosquitoes carry West Nile virus. (Leslie 2008, p. 15)

Leslie (2008, p. 15) suggests that (32) is judged true because it describes a "harmful, dangerous, or appalling" property that one would aim to avoid. To generalize this insight, I would suggest that (32) answers "No" to the implicit question, "Are mosquitoes safe/okay?" even though it describes a tiny minority of mosquitoes; following the Anna Karenina Principle, a safe situation is perfectly safe, while any amount of danger makes it unsafe.

More broadly, I suggest, *clean* and *dirty* are just the tip of the iceberg. Future work may find that further inferences studied in semantics are also connected to the emotional valence of the properties described in language.

5 Conclusion

In the literature on the formal semantics of adjectives, various papers observe grammatical and inferential differences between adjectives such as *clean* versus *dirty* and *healthy* versus *sick* – without confronting the striking difference in sentiment across these pairs. This paper revisits such data from the lens of sentiment, using the widely-recognized Anna Karenina Principle alongside QUD-based pragmatic reasoning to explain why – across individuals, degrees, and dimensions – positive-sentiment adjectives tend to be interpreted as universal, while negative-sentiment adjectives are more likely to be interpreted as existential. The Anna Karenina Principle is not new (although it is relatively new to the semantics literature); also not new is Cruse's observation that maximum/minimum pairs such as *clean/dirty* align with desirability in the domain of degrees. But this paper breaks new ground by observing sentiment-driven quantificational asymmetries among individuals and dimensions as well as degrees, and by proposing a unified explanation for all these phenomena.

Empirically, this explanation is consistent with a study of human judgments as well as corpus data handled with tools from natural language processing. Theoretically, this paper brings together formal semantics (the interpretation of predicates applied to definite plurals) with lexical semantics (the meanings of specific adjectival predicates), discourse pragmatics (reasoning about the QUD), and usage-based ideas about how the lexicon is shaped by communicative function, in turn rooted in theories from social psychology about emotional valence. The overall research program is grounded in the conviction that the study of linguistic meaning should mutually illuminate the human experience described therein.

From that perspective, quantification may seem dispassionate in decontextualized examples, but in real conversation, quantities and degrees are often discussed as instantiating more or less than would be desired. So it is no surprise that our inferences may be shaped by the desirability of the property that they describe.

Appendix: Adjectives used in empirical studies

Here are the 110 adjectives (55 antonym pairs) used in the experiment (Section 4.1) and the corpus study (Section 4.2) of minimum- versus maximum-standard modifiers. Adjectives are listed with the more positive-sentiment antonym first, following the annotations in the Hedonometer sentiment lexicon (Dodds et al. 2011, 2015) – hence why I list *light/heavy* and *cool/warm* rather than the reverse ordering. In the experiment (Section 4.1), each adjective is paired with a noun, given here in parentheses.

1. active/passive (employees)
2. adequate/inadequate (hotel rooms)
3. awake/asleep (toddlers)
4. beautiful/ugly (centerpieces)
5. big/little (cookies)
6. bright/dark (streets)
7. certain/uncertain (itineraries)
8. clean/dirty (cups)
9. clear/unclear (tutorials)
10. clever/stupid (advertisements)
11. closed/open (doors)
12. cool/warm (beers)
13. deep/shallow (arguments)
14. dressed/naked (babies)
15. early/late (buses)
16. easy/difficult (tests)
17. expensive/cheap (bikes)
18. fair/unfair (exams)
19. famous/unknown (actors)
20. fancy/plain (meals)
21. fast/slow (internet routers)
22. full/empty (vases)
23. good/bad (podcast episodes)
24. happy/sad (customers)
25. healthy/sick (parrots)
26. high/low (quality metrics)
27. hot/cold (lunches)
28. interesting/boring (videos)
29. large/small (pizzas)
30. light/heavy (desserts)
31. neat/messy (bookshelves)
32. nice/mean (teachers)
33. normal/abnormal (chemical levels)
34. peaceful/violent (protests)
35. positive/negative (meetings)
36. possible/impossible (goals)
37. quiet/loud (kids)
38. rich/poor (schools)
39. safe/dangerous (parks)
40. sane/insane (clients)
41. sharp/blunt (knives)
42. shiny/dull (silver platters)
43. simple/complex (policies)
44. smart/dumb (interview questions)
45. smooth/rough (field trips)
46. soft/hard (chairs)
47. straight/bent (rods)
48. strong/weak (athletes)
49. tall/short (basketball players)

50. useful/useless (trainings)
51. valuable/worthless (gemstones)
52. visible/invisible (campaign messages)
53. wet/dry (towels)
54. wide/narrow (seats)
55. young/old (dogs)

As for the ten adjectives used in the study of conjunctive versus disjunctive interpretations (Section 4.3), these are:

(33) a. **Conjunctive adjectives:** good, healthy, identical, normal, typical, similar
b. **Disjunctive adjectives:** bad, sick, abnormal, different

Statement about the use of Artificial Intelligence

I personally wrote all the words in this paper with my human brain. I used code snippets adapted from Gemini (appearing in AI-assisted search results) in statistical analyses.

References

Aristotle. 1925 [originally written 350 B.C.E.]. *The Nichomachean ethics*. Oxford: Oxford University Press. Translated by Sir William David Ross.

Bar-Lev, Moshe E. 2021. An implicature account of homogeneity and non-maximality. *Linguistics and Philosophy* 44. 1045–1097. doi:10.1007/s10988-020-09308-5.

Barrett, Lisa Feldman. 2006. Valence is a basic building block of emotional life. *Journal of Research in Personality* 40(1). 35–55. doi:10.1016/j.jrp.2005.08.006.

Bartsch, Renate & Theo Venneman. 1972. The grammar of relative adjectives and comparison. *Linguistische Berichte* 20. 19–32. doi:10.1007/3-540-07016-8_11.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). doi:10.18637/jss.v067.i01.

Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer & Kathleen D. Vohs. 2001. Bad is stronger than good. *Review of General Psychology* 5(4). 323–370. doi:10.1037//1089-2680.5.4.323.

Baumgartner, Jason, Savvas Zannettou, Brian Keegan, Megan Squire & Jeremy Blackburn. 2020. The PushShift Reddit dataset. In *International AAAI (Association for the Advancement of Artificial Intelligence) conference on web and social media*, vol. 14, 830–839. doi:10.1609/icwsm.v14i1.7347.

Beltrama, Andrea. 2025. Evaluation, thresholds, and practical commitments: The grammar of adjectival mildness. *Natural Language Semantics* 33(2). 169–205. doi:10.1007/s11050-025-09230-1.

Bierwisch, Manfred. 1989. The semantics of gradation. In Manfred Bierwiech & Ewald Lang (eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation*, 71–261. Berlin and Heidelberg: Springer-Verlag. doi:10.1007/978-3-642-74351-1_3.

Bogal-Allbritten, Elizabeth. 2012. Slightly coerced: Processing evidence for adjectival coercion by minimizers. In Andrea Beltrama, Tasos Chatzikonstantinou, Jackson L. Lee, Mike Pham & Diane Rak (eds.), *Proceedings of the 48th meeting of the Chicago Linguistic Society*, 77–92. Chicago: Chicago Linguistic Society.

Boucher, Jerry & Charles E. Osgood. 1969. The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior* 8(1). 1–8. doi:10.1016/s0022-5371(69)80002-2.

Brinton, Laurel J. & Elizabeth Closs Traugott. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press.

Brisson, Christine. 2003. Plurals, 'all', and the nonuniformity of collective predication. *Linguistics and Philosophy* 26(2). 129–184.

Burnett, Heather. 2014. A delineation solution to the puzzles of absolute adjectives. *Linguistics and Philosophy* 37(1). 1–39. doi:10.1007/s10988-014-9145-9.

Bylinina, Lisa & Stas Zadorozhny. 2012. Evaluative adjectives, scale structure, and ways of being polite. In Maria Aloni, Vadim Kimmelman, Floris Roelofsen, Galit Weidman Sasson, Katrin Schulz & Matthijs Westera (eds.), *Logic, language and meaning: Proceedings of the 18th Amsterdam Colloquium*, 133–142. Berlin: Springer. doi:10.1007/978-3-642-31482-7_14.

Carlson, Gregory N. 1977. *Reference to kinds in English*. Amherst, MA: University of Massachusetts dissertation.

Champollion, Lucas. 2010. *Parts of a whole: Distributivity as a bridge between aspect and measurement*. Philadelphia, PA: University of Pennsylvania dissertation.

Champollion, Lucas, Dylan Bumford & Robert Henderson. 2019. Donkeys under discussion. *Semantics and Pragmatics* 12. 1–1. doi:10.3765/sp.12.1.

Chierchia, Gennaro. 2022. 'People are fed up; don't mess with them:' non-quantificational arguments and polarity reversals. *Journal of Semantics* 39(3). 475–521. doi:10.1093/jos/ffac006.

Clark, Eve V. 1993. *The lexicon in acquisition*. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511554377.

Cooper, William E. & John Robert Ross. 1975. Word order. In Robin E. Grossman, L. James San & Timothy J. Vance (eds.), *Proceedings of the 11th annual meeting of the Chicago Linguistic Society (CLS 11): Parassession on Functionalism*, 63–111. Chicago, IL: Chicago Linguistic Society.

Cresswell, Maxwell John. 1977. The semantics of degree. In Barbara Partee (ed.), *Montague grammar*, 261–292. New York: Academic Press. doi:10.1016/b978-0-12-545850-4.50015-7.

Cruse, D. Alan. 1980. Antonyms and gradable complementaries. In Dieter Kastovsky (ed.), *Perspektiven der lexikalischen Semantik: Beiträge zum Wuppertaler Semantikkolloquium vom 2-3 Dezember 1977 [Perspectives on lexical semantics: Contributions to the Wuppertal semantics colloquium of December 2-3, 1977]*, 14–25. Bonn: Bouvier Verlag Herbert Grundmann.

Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo & Stanley Peters. 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy* 21. 159–210. doi:10.1023/a:1005330227480.

Dalrymple, Mary, Makoto Kanazawa, Sam Mchombo & Stanley Peters. 1994. What do reciprocals mean? In Mandy Harvey & Lynn Santelmann (eds.), *Semantics and Linguistic Theory (SALT)*, vol. 4, 61–78. doi:10.3765/salt.v0i0.2466.

Diamond, Jared M. 1997. *Guns, germs, and steel: A short history of everybody for the last 13,000 years*. New York and London: W. W. Norton and Company.

Dickens, Charles. 1859. *A tale of two cities*. London: Chapman & Hall.

Dixon, Robert M. W. 2004. Adjective classes in typological perspective. In Robert M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Adjective classes: A cross-linguistic typology*, 1–49. Oxford: Oxford University Press. doi:10.1093/oso/9780199270934.003.0001.

Dodds, Peter Sheridan, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdoomian, Matthew McMa-

hon, Brian Tivnan & Christopher Danforth. 2015. Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences* 112(8). 2389–2394. doi:10.1073/pnas.1411678112.

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss & Christopher M. Danforth. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS one* 6(12). e26752. doi:10.1371/journal.pone.0026752.

Dowty, David R. 1987. Collective predicates, distributive predicates, and 'all'. In Fred Marshall, Ann Miller & Zheng sheng Zhang (eds.), *Proceedings of the third Eastern States Conference on Linguistics (ESCOL)*, The Ohio State University.

Ducrot, Oswald. 1973. *La preuve et le dire [Proof and statement]*. Paris: Mame.

Fahrni, Angela & Manfred Klenner. 2008. Old wine or warm beer: Target-specific sentiment analysis of adjectives. In Chris Mellish, Fiorella de Rosis, Isabella Poggi & Ielka van der Sluis (eds.), *Proceedings of the artificial intelligence and simulation of behaviour (AISB) symposium on affective language in human and machine*, Aberdeen: The Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Fodor, Janet Dean. 1970. *The linguistic description of opaque contexts*: Massachusetts Institute of Technology dissertation. doi:10.4324/9781315880303.

Giannakidou, Anastasia. 2001. The meaning of free choice. *Linguistics and Philosophy* 24(6). 659–735. doi:10.1023/a:1012758115458.

Glass, Lelia. 2018. Deriving the distributivity potential of adjectives via measurement theory. *Proceedings of the Linguistic Society of America* 3. 49: 1–14. doi:10.3765/plsa.v3i1.4343.

Glass, Lelia. 2023. Using the Anna Karenina Principle to explain why 'cause' favors negative-sentiment complements. *Semantics and Pragmatics* 16. 6–1. doi:10.3765/sp.16.6.

Glass, Lelia. 2025. *The red dress is cute*: Why subjective adjectives are more often predicative. *Corpus Linguistics and Linguistic Theory* 21(2). 401–432. doi:10.1515/cllt-2024-0044.

Glauch, Kalle. 2024. *Expressive meaning and emotional valence*. Bochum, Germany: Ruhr University of Bochum dissertation.

Gonzalez, Guillermo, Donald Brownlee & Peter Ward. 2001. The galactic habitable zone: Galactic chemical evolution. *Icarus* 152(1). 185–200. doi:10.1006/icar.2001.6617.

Gotzner, Nicole & Diana Mazzarella. 2024. Negative strengthening: The interplay of evaluative polarity and scale structure. *Journal of Semantics* ffae004. doi:10.1093/jos/ffae004.

Greenwald, Anthony G., Debbie E. McGhee & Jordan L. K. Schwartz. 1998. Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology* 74(6). 1464. doi:10.1037/0022-3514.74.6.1464.

Grice, H. P. 1989. Logic and conversation. In *Studies in the way of words*, Cambridge: Harvard University Press. doi:10.1163/9789004368811_003. Originally published as Grice (1975), in *Syntax and semantics* (edited by Peter Cole and Jerry L. Morgan), volume 3: Speech acts, 41–58. New York: Academic Press.

Gutzmann, Daniel. 2011. Expressive modifiers and mixed expressives. In Olivier Bonami & Patricia Cabredo Hofherr (eds.), *Empirical issues in syntax and semantics*, vol. 8, 123–141. Colloque de Syntaxe et Sematique de Paris.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1). 25–70. doi:10.1017/s0022226705003683.

Honnibal, Matthew & Mark Johnson. 2015. An improved non-monotonic transition system for dependency parsing. In Lluís Màrquez, Chris Callison-Burch & Jian Su (eds.), *Empirical Methods in Natural Language Processing (EMNLP)*, 1373–1378. Lisbon, Portugal: Association for Computational Linguistics (ACL) Anthology. doi:10.18653/v1/d15-1162.

Horn, Laurence R. 1989. *A natural history of negation*. Chicago: University of Chicago Press.

Horn, Laurence R. 2022. Contrast and clausal order: On beyond Behaghel. *Language* 98(4). 812–843. doi:10.1353/lan.2022.0022.

Horn, Laurence R. 2025. Contradiction. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University.

Kagan, Olga & Sascha Alexeyenko. 2011. Degree modification in Russian morphology: The case of the suffix'-ovat'. In Ingo Reich, Eva Horch & Dennis Pauly (eds.), *Proceedings of sinn und bedeutung*, vol. 15, 321–336. Saarbrücken: Saarland University Press.

Kamoen, Naomi, Bregje Holleman, Rick Nouwen, Ted Sanders & Huub van den Bergh. 2011. Absolutely relative or relatively absolute? The linguistic behavior of gradable adjectives and degree modifiers. *Journal of Pragmatics* 43(13). 3139–3151. doi:10.1016/j.pragma.2011.05.017.

Kamp, Hans. 1975. Two theories about adjectives. In Edward Keenan (ed.), *Formal semantics of natural language*, 123–155. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511897696.011.

Kanouse, David E. 1984. Explaining negativity biases in evaluation and choice behavior: Theory and research. In Thomas C. Kinnear (ed.), *North American advances in consumer research*, vol. 11, 703–708. Provo, UT: Association for Consumer Research.

Kennedy, Christopher. 1997. *Projecting the adjective: The syntax and semantics of gradability and comparison*: University of California, Santa Cruz dissertation. doi:10.4324/9780203055458.

Kennedy, Christopher. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30. 1–45. doi:10.1007/s10988-006-9008-0.

Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 345–381. doi:10.1353/lan.2005.0071.

Klein, Ewan. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4. 1–45. doi:10.1007/bf00351812.

Knobe, Joshua. 2003. Intentional action and side effects in ordinary language. *Analysis* 63(3). 190–194. doi:10.1111/1467-8284.00419.

Kölbel, Max. 2004. Faultless disagreement. *Proceedings of the Aristotelian Society* 104(1). 53–73. doi:10.1111/j.0309-7013.2004.00081.x.

Kraut, Richard. 2022. Aristotle's Ethics. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University Fall 2022 edn.

Krifka, Manfred. 1996. Pragmatic strengthening in plural predications and donkey sentences. In Teresa Galloway & Justin Spence (eds.), *Semantics and Linguistic Theory*, vol. 6, 136–153. doi:10.3765/salt.v6i0.2769.

Križ, Manuel. 2016. Homogeneity, non-maximality, and 'all'. *Journal of Semantics* 33(3). 493–539. doi:10.1093/jos/ffv006.

Križ, Manuel. 2019. Homogeneity effects in natural language semantics. *Language and Linguistics Compass* 13(11). e12350. doi:10.1111/lnc3.12350.

Križ, Manuel & Benjamin Spector. 2021. Interpreting plural predication: Homogeneity and non-maximality. *Linguistics and Philosophy* 44. 1131–1178. doi:10.1007/s10988-020-09311-w.

Kun, Anna & Bernard Weiner. 1973. Necessary versus sufficient causal schemata for success and failure. *Journal of Research in Personality* 7(3). 197–207. doi:10.1016/0092-6566(73)90036-6.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 522–551. doi:10.2307/417059.

Lasersohn, Peter. 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy* 28(6). 643–686. doi:10.1007/s10988-005-0596-x.

Lassiter, Daniel & Noah D. Goodman. 2013. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In Todd Snider (ed.), *Semantics and linguistic theory (SALT) 23*, 587–610. doi:10.3765/salt.v0i0.2658.

Legge, James (ed.). 1861. *The Chinese classics*, vol. 1. London: Trübner.

Lehrer, Adrienne. 1985. Markedness and antonymy. *Journal of Linguistics* 21(2). 397–429. doi:10.1017/s002222670001032x.

Lehrer, Adrienne & Keith Lehrer. 1982. Antonymy. *Linguistics and Philosophy* 483–501. doi:10.1007/bf00355584.

Leslie, Sarah-Jane. 2008. Generics: Cognition and acquisition. *Philosophical Review* 117(1). 1–47. doi:10.1215/00318108-2007-023.

Levin, Beth & Malka Rappaport Hovav. 2019. Lexicalization patterns. In Robert Truswell (ed.), *Oxford handbook of event structure*, 395–425. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199685318.013.18.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use, and interpretation of language*, Berlin: De Gruyter. doi:10.1515/9783110852820.302.

Löbner, Sebastian. 1987. The conceptual nature of natural language quantification. In Imre Ruzsa & Anna Szabolcsi (eds.), *Proceedings of the Debrecen symposium on logic and language*, 81–94. Budapest: Akadèmiai Kiadò.

Löbner, Sebastian. 2000. Polarity in natural language: Predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy* 23. 213–308. doi:10.1023/a:1005571202592.

Loper, Edward & Steven Bird. 2002. NLTK: The natural language toolkit. In Chris Brew, Michael Rosner & Dragomir Radev (eds.), *Proceedings of the Association for Computational Linguistics (ACL) Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, 63–70. Philadelphia, PA. doi:10.3115/1118108.1118117.

Louw, Bill. 1993. Irony in the text or insincerity in the writer?—the diagnostic potential of semantic prosodies. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, 157–176. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Ludlow, Peter. 2014. *Living words: Meaning underdetermination and the dynamic lexicon*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198712053.001.0001.

Mador-Haim, Sela & Yoad Winter. 2015. Far from obvious: the semantics of locative indefinites. *Linguistics and Philosophy* 38. 437–476. doi:10.1007/s10988-015-9175-y.

Malamud, Sophia A. 2012. The meaning of plural definites: A decision-theoretic approach. *Semantics and Pragmatics* 5. 3–1. doi:10.3765/sp.5.3.

Nadathur, Prerna. 2023. *Actuality inferences: Causality, aspect, and modality*, vol. 15. Oxford: Oxford University Press. doi:10.1093/oso/9780192849885.001.0001.

Nouwen, Rick. 2024. The semantics and probabilistic pragmatics of deadjectival intensifiers. *Semantics and Pragmatics* 17. doi:10.3765/sp.17.2.

Osgood, Charles Egerton, George J. Suci & Percy H. Tannenbaum. 1957. *The measurement of meaning*. Urbana-Champaign, IL: University of Illinois Press.

Pang, Bo & Lillian Lee. 2008. Opinion mining and sentiment analysis. In *Foundations and trends in information retrieval*, vol. 2, 1–135. Hanover, MA: NOW Publishers. doi:10.1561/1500000011.

Paradis, Carita. 1997. *Degree modifiers of adjectives in spoken British English*, vol. 92 Lund Studies in English. Lund University Press.

Paradis, Carita, Joost Van de Weijer, Caroline Willners & Magnus Lindgren. 2012. Evaluative polarity of antonyms. *Lingue e linguaggio* 11(2). 199–214.

Pearl, Judea. 2009. *Causality*. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511803161.

Peeters, Guido & Janusz Czapinski. 1990. Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology* 1(1). 33–60. doi:10.1080/14792779108401856.

Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(2). 165–198. doi:10.1515/tl.2007.011.

Potts, Christopher. 2011. On the negativity of negation. In Nan Li & David Lutz (eds.), *Semantics and Linguistic Theory (SALT)*, vol. 20, 636–659. Cornell Linguistics Circle Publications. doi:10.3765/salt.v0i20.2565.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik & David Crystal. 1985. *A grammar of contemporary English*. London: Longman.

R Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. `http://www.R-project.org/`.

Roberts, Craige. 1987. *Modal subordination, anaphora, and distributivity*. Amherst, MA: University of Massachusetts dissertation.

Roberts, Craige. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. doi:10.3765/sp.5.6.

Rooth, Mats. 1987. Noun phrase interpretation in montague grammar, file change semantics, and situation semantics. In Peter Gärdendors (ed.), *Generalized quantifiers: Linguistic and logical approaches*, vol. 31 Studies in Linguistics and Philosophy (SLAP), 237–268. Dordrecht: Reidel. doi:10.1007/978-94-009-3381-1_9.

Rotstein, Carmen & Yoad Winter. 2004. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* 12. 259–288. doi:10.1023/b:nals.0000034517.56898.9a.

Rozin, Paul & Edward B. Rozyman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* 5(4). 296–320. doi:10.1207/s15327957pspr0504_2.

Sapir, Edward. 1944. Grading: A study in semantics. *Philosophy of Science* 11(2). 93–116. doi:10.1086/286828.

Sassoon, Galit Weidman. 2010. The degree functions of negative adjectives. *Natural Language Semantics* 18. 141–181. doi:10.1007/s11050-009-9052-8.

Sassoon, Galit Weidman. 2013. A typology of multidimensional adjectives. *Journal of Semantics* 30(3). 335–380. doi:10.1093/jos/ffs012.

Schwarzschild, Roger. 2011. Stubborn distributivity, multiparticipant nouns and the count/mass distinction. In Suzi Lima, Kevin Mullin & Brian Smith (eds.), *Proceedings of the North East Linguistic Society (NELS)*, vol. 39 2, 661–678. Amherst, MA: Graduate Linguistics Students Association, University of Massachusetts.

Scontras, Gregory, Judith Degen & Noah D. Goodman. 2017. Subjectivity predicts adjective ordering preferences. *Open Mind* 1(1). 53–66. doi:10.1162/opmi_a_00005.

Scontras, Gregory, Judith Degen & Noah D. Goodman. 2019. On the grammatical source of adjective ordering preferences. *Semantics and Pragmatics* 12. 7–1. doi:10.3765/sp.12.7.

Scontras, Gregory & Noah D Goodman. 2017. Resolving uncertainty in plural predication. *Cognition* 168. 294–311. doi:10.1016/j.cognition.2017.07.002.

Sinclair, John McHardy. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stojanovic, Isidora & Elsi Kaiser. 2022. Exploring valence in judgments of taste. In Jeremy Wyatt, Julia Zakkou & Dan Zeman (eds.), *Perspectives on taste*, 231–259. New York: Routledge. doi:10.4324/9781003184225-15.

Taboada, Maite. 2016. Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics* 2(1). 325–347. doi:10.1146/annurev-linguistics-011415-040518.

Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 3, 36–149. Cambridge: Cambridge University Press.

Taylor, Shelley E. 1991. Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin* 110(1). 67–85. doi:10.1037//0033-2909.110.1.67.

Tostoy, Leo. 2004 [originally published 1874]. *Anna Karenina*. New York: Penguin Classics. Translated by Richard Pevear and Larissa Volokhonsky.

Unger, Peter. 1975. *Ignorance: A case for scepticism*. Oxford: Clarendon Press. doi:10.1093/0198244177.001.0001.

Unkelbach, Christian, Hans Alves & Alex Koch. 2020. Negativity bias, positivity bias, and valence asymmetries: Explaining the differential processing of positive and negative information. In Bertram Gawronski (ed.), *Advances in experimental social psychology*, vol. 62, 115–187. Amsterdam, The Netherlands: Elsevier. doi:10.1016/bs.aesp.2020.04.005.

Weiner, Bernard. 1985. 'Spontaneous' causal thinking. *Psychological Bulletin* 97(1). 74–84. doi:10.1037/0033-2909.97.1.74.

Welch, Bernard L. 1947. The generalization of Student's problem when several different population varlances are involved. *Biometrika* 34(1-2). 28–35. doi:10.1093/biomet/34.1-2.28.

Wiebe, Janyce. 2000. Learning subjective adjectives from corpora. In *Seventeenth national conference on artificial intelligence*, vol. 20 0, 0. Austin, TX: Association for the Advancement of Artificial Intelligence (AAAI).

Yoon, Youngeun. 1996. Total and partial predicates and the weak and strong interpretations. *Natural Language Semantics* 4(3). 217–236. doi:10.1007/bf00372820.

Zimmer, Karl E. 1964. Affixal negation in English and other languages: An investigation of restricted productivity. *Word* 20. Monograph No. 5.