Inferring Comparison Classes of Gradable Adjectives

The Role of Informational Goals and Sentence Structure

By

Polina Tsvilodub

Thesis Supervisor:

Prof. Dr. Michael Franke, Institute of Cognitive Science,
Osnabrück University

Thesis Supervisor:

Dr. Michael Henry Tessler, Postdoctoral Associate,
Department of Brain and Cognitive Sciences, MIT

Abstract

Understanding gradable adjectives like "big" requires making reference to a so-called comparison class - a set of objects the referent is implicitly compared to. For example, the utterance '"That Great Dane is big" could mean "That Great Dane is big compared to dogs in general" or "That Great Dane is big compared to other Great Danes"; yet the comparison class is rarely stated explicitly. So how do listeners establish the comparison class, given multiple a priori reasonable options? Research on gradable adjectives has focused on the representation and integration of comparison classes into compositional semantics, but little is known about how human listeners decide upon a comparison class. This work takes a functional perspective on comparison class inference, guided by informational goals that speakers pursue when producing an utterance with a gradable adjective, and how listeners expect these goals to be achieved syntactically. For instance, given simple "Subject Predicate" sentences listeners expect that the subject aids reference (i.e., identifies the target), whereas the predicate accomplishes predication (i.e. asserts a property of the subject). Therefore, the role of the noun for comparison class determination in a simple "Subject Predicate" sentence depends on its syntactic position. A noun appearing in the predicate is more likely to be intended to constrain the comparison class, whereas a noun in the subject can be explained away as intended for reference, leaving comparison class inference to other pragmatic reasoning. Converging evidence from four behavioural experiments supporting this proposal is presented alongside a novel formalisation of the inferential account in a qualitative computational model within the Rational Speech Act framework. This work contributes to the body of research on gradable adjectives, and provides a case study of context-dependent language, emphasizing the complexity of the relation between the form and the meaning of linguistic expressions.

Acknowledgements

I want to thank...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The meaning of natural language expressions heavily depends on the context in which these expressions are used, but speakers rarely explicitly outline which aspects of the context are relevant for their interpretation.

This issue is clearly illustrated by utterances involving gradable adjectives like *big, small, tall, expensive* etc. These adjectives are typically taken to describe a degree to which an object possesses some property, e.g., the degree of bigness (i.e., size), for the adjective 'big', but specific degrees a speaker intends to convey vary a lot depending on the particular referent and context. Intuitively, the utterance "That's big!" denotes quite different size degrees meanings, depending on whether it was uttered in reference to a flower or in reference to a house, while both objects could potentially co-occur in the same context; given this utterance, it is left to the listener to identify the correct referent and size degree. The aspect that goes unsaid and allows for this flexible use of the adjective *big* across referents and contexts is *what the intended referent is big relative to.* Humans easily infer that these two objects might be compared to different things: for instance, it is more likely that the flower is big for this specific kind of flowers or relative to other flowers around it, whereas the house is probably rather being compared to other houses in the neighborhood.

However, speakers rarely explicitly state this comparison class - the set of entities the target is compared against, and it is left to the addressee to establish the relevant comparison set (Solt, 2009). Listeners feature vast general knowledge and experience about the world helping them interpret context-sensitive language (Tessler et al., 2017), but what additional linguistic features do listeners attend to? In particular, how do listeners establish a comparison class in order to interpret a gradable adjective, given infinitely many a priori plausible options for the comparison class?

This work investigates the role of syntactic structure for sentences containing relative gradable adjectives, suggesting that the syntax provides a cue to contex-

tually relevant aspects for adjective interpretation, which is integrated with other cues like perceptual context and world knowledge. [1] In particular, we hypothesize that syntactic structure reflects informational goals interlocutors strive to achieve; they reason about these goals pragmatically when inferring the comparison class of gradable adjectives. Focusing on the informational goals *reference* and *predication*, this work presents a novel **reference-predication trade-off hypothesis** of comparison class inference, contributing to the body of research on gradable adjectives and providing a case study for the relationship between linguistic form and meaning. Evidence from four behavioral experiments is provided in support of this functional hypothesis, as well as a Bayesian model of gradable adjective interpretation, showing that sophisticated pragmatic reasoning about syntactic structure can be captured using the generic probabilistic Rational Speech Act framework (Goodman & Frank, 2016).

---

[1]This thesis summarizes and extends the work by Tessler, Tsvilodub, Snedeker and Levy published in Tessler et al. (2020), that appeared in the *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society.*

Chapter 2

Understanding Gradable Adjectives

Absolute vs. relative gradable adjectives. One particularly interesting case study of context-sensitive language are gradable adjectives. Specifically, it depends on the context what exactly counts as *tall, expensive, small* or *full* - a one-meter tall three-year-old counts as tall, but a one-meter tall redwood tree does not; a three-quarter full cup of coffee counts as full, but a three-quarter full spaceship fuel tank does not. While both examples show context-sensitivity of the adjective's meaning, these two adjectives differ in what exactly about their meaning depends on the context: in case of *relative gradable adjectives* like 'tall' the context determines how much of the feature described by the adjective is required to count as 'tall', whereas in case of *absolute gradable adjectives* like 'full' the context determines how much the degree of the described feature may deviate from total fullness (Aparicio et al., 2016; Hofherr & Matushansky, 2010; Kennedy, 2007).

In particular, the meaning of a relative gradable adjective, for instance 'big', can be described in that 'big' refers to the size of an object, and the size of that object described as 'big' must be at least X, such that it counts as big, relative to some standard of comparison θ. This means, relative gradable adjectives convey a feature, like size, and the degree to which the referent possesses this feature must exceed some threshold θ for the referent to be felicitously described by the respective gradable adjective (e.g., Kennedy, 2007). At the same time this threshold θ can vary across contexts or categories: the minimal size of a flower that counts as big is quite different from the minimal size of a house that counts as big. Moreover, this threshold can vary within categories: the minimal size of a big sunflower is different from the minimal size of a big daisy, although both belong to the category flowers. Hence, this threshold θ is strongly influenced by the set relative to which the object is compared - namely the *comparison class*.

In contrast, the meaning of an absolute gradable adjective, for instance 'full', is debated: some researchers argue that it refers to an endpoint on the feature scale described by the adjective, i.e., 'full' refers to the maximum on the scale of volume for the object under discussion (Aparicio et al., 2016; Kennedy, 2007; Qing & Franke, 2014). Others argue that the meaning of absolute gradable adjectives is also resolved relative to a context-sensitive threshold θ, by mechanisms universal for all gradable adjectives (Lassiter & Goodman, 2017).

Generally, gradable adjectives are *vague* - their meaning is subject to contextual variability, and to other characteristic features of vagueness: there exist so-called borderline cases, and these adjectives give rise to the Sorites paradox (Kennedy, 2007). Specifically, even when a comparison class is set, there are cases where it is unclear whether an object counts as e.g. 'expensive': while a cup of coffee for $1 is clearly cheap, and a cup for $5 is clearly expensive, it might be difficult to say whether a $3.75 coffee is expensive or not - this is a borderline case. Using the same example, the Sorites paradox can be illustrated for gradable adjectives as follows:

P1: A $5 cup of coffee is expensive (for a cup of coffee).

P2: Any cup of coffee that costs 1 cent less than an expensive one is expensive (for a cup of coffee).

C: Therefore, any free cup of coffee is expensive.

It is the vague nature of gradable adjectives that makes it difficult to pinpoint why exactly people accept the premises so easily, and although the argument seems valid, the conclusion is clearly false (see Kennedy, 2007, for more details).

Investigating these important properties in greater detail is outside of the scope of this work: in the remainder, the focus is to investigate the importance of comparison classes, specifically for relative gradable adjectives. Yet characteristics like borderline cases and eliciting the Sorites paradox emphasize that capturing the kind of implicit comparison to a threshold θ which occurs in the positive form of gradable adjectives, while accounting for the existence of these properties is rather difficult. The following sections review state-of-art representations of relative gradable adjective semantics and the role of comparison classes therein. Then, prior related theoretical and experimental work on comprison classes is presented.

2.1 Semantic Representation of Gradable Adjectives

Currently standard theories of gradable adjectives converge in representing gradable adjectives as a function mapping their argument - the referent - to a degree on

an ordered scale representing some feature (e.g., 'big' and 'small' represent size), utilizing degree morphology (Kennedy, 2007). Degree morphology for positive forms of relative adjectives is informed by their comparative form, where the degree of a feature of the referent is explicitly compared to another degree of the same feature, and this comparison is overtly realised by a degree morpheme -*er*. For instance, in the comparative sentence 'Bob is taller than Alice' Bob's height is explicitly compared to Alice's height, expressed by the morpheme -*er* appended to tall. By contrast, unmodified positive forms of relative adjectives which are the focus of this work don't have an overt degree morpheme specifying the comparison to some point of reference; in the currently widely accepted approach (reviewed by Kennedy, 2007) a phonologically silent null degree morpheme *pos* is introduced for this purpose. The morpheme *pos* takes the adjective as an argument and returns a standard of comparison - the context-dependent threshold θ. In Kennedy (2007), the comparison class is assumed to be an argument of the adjective, potentially restricting the domain of entities it applies to - an assumption discussed further in section **??**. Formally, *pos* denotes the following:

$$[\![_{Deg}pos\lambda g\lambda x.g(x)]\!] = \lambda g.\lambda x : g(x) > s(\lambda x : g(x))$$

In other words, the degree to which the referent x possesses the property denoted by the adjective g must exceed some threshold, provided by $s(g)$, where s is "a context-sensitive function that chooses a standard of comparison in such a way as to ensure that the objects that the positive form [of the adjective] is true of 'stand out' in the context of utterance, relative to the kind of measurement that the adjective [i.e., g] encodes" (Kennedy, 2007, p. 17). The contextually relevant aspects providing the threshold can be summarised as the comparison class of the adjective. For example, the expression 'big dog' is true if the size of a target dog exceeds some size-threshold, set by the comparison class. Depending on the context and the comparison class this threshold might vary: the minimal size the dog has to have in order to be described as 'big' is different if the dog is a toy dog, and the comparison class are other toys, than for a dog that is a Great Dane and the comparison class is other Great Danes.

Alternative to the degree-semantics framework, delineation-based formalizations of gradable adjectives treat them as unary predicates, forming partial functions depending on contextually provided comparison classes (Klein, 1980). Such an approach removes degree representations from the semantics, although degrees arguably are an indispensable part of the meaning of gradable adjective (Solt, 2009).

The general issue of outlined semantic representations of gradable adjectives is that they assume the relevant comparison class to be supplied contextually, yet omit-

ting to specify what exactly the comparison class is or how it is determined. While this work assumes a degree-based formalisation, it should be noted that alternative approaches also rely on the notion of contextually appropriate comparison classes, making the question addressed in this work as to how exactly comparison classes are determined a relevant one across different semantic representations.

2.2   Understanding Comparison Classes

Comparison classes can be understood as sets of entities, or reference frames the object described by the adjective is compared against (Bierwisch, 1989; Klein, 1980; Solt, 2009). In the outlined examples comparison classes were assumed to be sets of physical objects like dogs or flowers. But comparison classes need not be comprised of individuals or objects, they can also comprise events or locations: In the utterance "The store is crowded for a Tuesday" the fullness of a particular store is naturally compared to other Tuesdays, rather than to other stores (Solt, 2009). It is crucial that "the comparison class provides statistical information that serves to determine the thresholds [...., and] what is relevant is not only the central value but also some measure of the extend of dispersion of values corresponding to members of the comparison class" (Solt, 2009, p.193). Note that the width of the value distribution is closely related to the specificity of the comparison class: more general categories serving a comparison classes like *basic-level* categories tend to imply a wider distribution than more specific comparison classes, for instance based on *subordinate* categories (Rosch et al., 1976). From a pragmatic perspective, cooperative speakers should tend to use relatively specific comparison classes appropriate in context, since they are more informative with respect to the underspecified threshold θ than more general ones. Listeners assuming cooperative speakers would then tend to infer maximally specific comparison classes, respectively (Tessler et al., 2017).

This naturally leads to the question is how exactly the standard of comparison - the threshold θ - is determined by a given comparison class. For instance, Cresswell (1976) suggested that the threshold θ is the average of the relevant feature over the comparison class, but arguments have been laid against this idea, showing that these thresholds do not seem to comprise a single point on the degree scale, but should rather be represented as comprising a range of values (Kennedy, 2007; Stechow, 1984). One proposal by Solt (2009, p.194) is that this range is computed as an interval around the median $median_{x \in C}$ provided by the comparison class C (which the target referent x is a member of), where the width of this interval is determined by the degree of variability of the feature in the comparison class, as provided by the measure function $MEAS$ and captured by the median absolute deviation (MAD):

$$R_{Std:C} = median_{x \in C}MEAS(x) \pm n \bullet MAD_{x \in C}MEAS(x)$$

However, it is still unclear how the relevant comparison class C is determined. Comparison classes can be expressed overtly using prepositional *for*-phrases, for instance, as in "That Great Dane is *big for a dog*" or in "That shirt is *big for you*". In the first example, additionally to expressing the comparison class, the *for*-phrase acts as a *presupposition trigger*, implying that the Great Dane is also a dog (Bale, 2011; Solt, 2009, cf.). Notably, this is not the case for the second example.

There are several proposals with respect to compositional semantic integration of *for*-phrases. Kennedy (2007) suggested that *for*-phrases introduce a domain restriction on the gradable adjective via direct composition, hence being an argument of the adjective. That is, the comparison class restricts the domain of entities the adjective applies to. But this approach has difficulties accounting for cases when it is not the subject of the sentence that combines with the gradable adjective, or when adjectives appear in what has been labeled by Ebeling and Gelman (1994) as *functional uses*, e.g., "That short is big for you" (Solt, 2009).

An alternative is to interpret *for*-phrases in relation to the *pos*-morpheme, as marking its scope, similar to the relation between *than*-phrases and the comparative morpheme *-er*. In order to account for their presupposition-triggering behavior, the *pos*-morpheme is then assumed to take a comparison class C as an argument, which by presupposition the referent is a member of (Solt, 2009). Formally:

$$[\![POS]\!] = \lambda C_{\langle et \rangle} \lambda P_{\langle d,et \rangle} \lambda x : x \in C. \forall d \in R_{Std:C}[P(x,d)]$$

where $P(x,d)$ denotes the measure function mapping individuals onto respective degrees on the feature scale described by the adjective, and $R_{Std:C}$ is the standard of comparison, e.g., computed as described above. This view follows the proposal by Bartsch and Vennemann (1972), wherein the comparison class is an argument of a function computing the standard of comparison, whatever the nature of this function may be. However, in cases like "John is tall for a gymnast", overt *for*-phrases may provide this argument, but for cases like "Sara reads difficult books for an 8-year-old" Solt (2009) assumed 'books' to be the basis of the comparison class rather incidentally, focusing on the representation of the presupposition triggered by the *for*-phrase.

Finally, another approach to comparison class representation proposes that they "restrict binary relations, and these binary relations form the basis for the construction of [degree] scales [..., which] serve to relativize the calculation of a standard" of comparison (Bale, 2011, p.170). This proposal is based on deriving scales described by gradable adjectives from quasi-orders, i.e., those binary relations, for instance

by creating so-called equivalence classes (sets of objects with equivalent degrees on that scale), which then are ordered based on the original quasi-ordering, and finally by defining a measure function via mapping each element onto its equivalence class in the scale (Bale, 2011). Comparison classes then restricts the quasi-order before formation of the scale, restricting the quasi-orders to "ordered pairs consisting only of members of the comparison class", such that the scale only describes degrees of members of the comparison class (Bale, 2011, p.178). This structure is then passed as an argument to some function returning the standard of comparison, analogous to approaches described above. One feature of this approach is the possibility to introduce a scale for gradable adjectives which are not inherently connected to some metric scale, e.g., for adjectives like 'beautiful' or 'interesting' (Bale, 2011).

In cases where no overt *for*-phrase is used, it is assumed that the argument of *pos* is a contextually appropriate implicit comparison class, e.g., supplied syntactically by the nominal modified by the adjective. Many assume the modified noun to supply the comparison class universally (cf. e.g. Cresswell, 1976; Heim, 2000; Kamp, 1975), while Solt (2009) restricts this mechanism in terms of the *pos*-morpheme scope, proposing that comparison class saturation is local given a modified nominal, but involves raising in case of *for*-phrases. This leaves open the origin of comparison class arguments in sentences where the adjective appears predicatively without a *for*-phrase - a question focused on in sections 2.3, 2.4.

This work focuses on the determination of relevant comparison classes even before they are integrated compositionally, so no commitment to a specific compositional approach shall be made here.

Gradable adjectives and comparison classes, respectively, have also been addressed from a developmental and psychological perspective, in particular as a case study of children's developing understanding of context. Barner and Snedeker (2008) have shown that by the age of 4 years, children are able to track statistical regularities of a property described by an adjective (e.g., height described by 'tall') in a novel population of toys ('pimwits') and flexibly adjust their use of the adjective according to changes of the property distribution.

Ebeling and Gelman (1994) distinguish three prominent uses of gradable adjectives children are exposed to, which can be loosely related to distinct linguistic constructions they tend to occur in and how the comparison class may be supplied, namely occurrences of adjectives where the comparison class is supplied *normatively, perceptually* or *functionally*. Normative comparison classes are based on a mental representation of the referent, for example it can comprise general world knowledge about the kind of things the referent belongs to. One could hypothesize that here the relevant knowledge might remain implicit and require interlocutors to infer relevant cues from context. Perceptual comparison classes are based on other objects

of the same type as the referent physically co-present at the moment of utterance (Ebeling & Gelman, 1994). The notion of perceptual comparison classes could naturally be extended to incorporate perceptually co-present objects of other kinds, in general. These comparison class uses might require less implicit general knowledge, but might still require figuring out which aspects of context are relevant. Finally, functional comparison class uses reference the intended use of the object, as in the aforementioned example "This shirt is *big for you*" (Ebeling & Gelman, 1994; Sera & Smith, 1987). While 'functional' comparison classes may be an exception in that they are very often stated overtly via the prepositional *for*-phrase, both normative and perceptual comparison classes often remain implicit, left to the listener to infer from their world knowledge or relevant contextual aspects. A preliminary study shows that adults might use syntactic structure of the utterance containing the adjective to signal the intended comparison class in such underspecified cases, consistent with the reference-predication trade-off hypothesis proposed in this work (Sinelnikova, 2020, discussed in greater detail in Chapter 6).

2.3 Semantic and Syntactic Aspects of Gradable Adjective Interpretation

former 'compositional accounts' section; maybe do a prior-work chapter?

While the notion of relative gradable adjectives as interpreted in reference to a comparison class has a long tradition (e.g., Bartsch & Vennemann, 1972; Bierwisch, 1989), there is little agreement on how exactly relevant comparison classes are identified when not supplied overtly. Prior work reviewed in this section has mainly focused on how syntactic and semantic properties of adjectives determine them.

One line of work on how comparison classes might be determined approaches this question from a purely compositional perspective. In particular, the noun the adjective combines with is said to be at least a very salient contextual cue towards the comparison class (Kamp, 1975). Simple compositional accounts propose that the nominal syntactically modified by the adjective necessarily stipulates the comparison class, such that 'small watch' resolves to 'the watch is small for a watch' (Cresswell, 1976; Kamp, 1975). More sophisticated ideas involve syntactic aspects of interpreting the *pos*-morpheme (see section 2.2). Yet, a lot of examples have been laid against such a simple mapping of the modified noun to the comparison class: intuitively, 'John is a rich Fortune 500 CEO' doesn't mean that he is *rich for a Fortune 500 CEO*; 'Kyle's car is an expensive BMW' doesn't mean that his car is *expensive relative to other BMWs* (Kennedy, 2007).

However, such syntactic theories focus on gradable adjectives occuring attribu-

tively, not accounting for their flexibility in occuring both attributively and predicatively (for example, attributive: 'That's a big dog'; or predicative: 'That dog is big'; cf., Hofherr and Matushansky (2010), McNally and Kennedy (2008)). Furthermore, attributive adjectives can occur prenominally (e.g., 'visible stars') and postnominally (e.g., 'stars visible [tonight]') (Hofherr & Matushansky, 2010). In English, the common basic position of attributive adjectives is prenominal, but post-nominal in e.g. Italian (Cinque, 2010); for this work focusing on English, post-nominal cases will be disregarded.

The exact syntactic relation between attributive and predicative occurences of adjectives is widely discussed; prior work attempted to derive one kind of syntactic construction from the other(e.g., Cresswell, 1976). For instance, predicative adjectives might be seen as elliptical uses derived from underlying attributive adjectives (e.g., 'The dog is big' derived from 'The dog is a big dog', cf., Kamp (1975)). This implies the simplest generalisation of these compositional syntactic accounts to predicative adjectives: one would posit that the noun of the sentence generally sets the comparison class, such that the utterance "That Great Dane is big" would be taken to mean "That Great Dane is big for a Great Dane" (Tessler et al., 2020). Yet, similar intuitive counter-examples might be put forward here. Therefore, although the noun the adjective combines with is arguably a salient cue to the comparison class, the degree to which it restricts the comparison class might vary across different utterances and contexts.

modification accounts , cf google doc table

Another distinction along the lines of adjective position potentially relevant for comparison class determination is the difference between intersective and non-intersective (or subsective) adjective readings (Hofherr & Matushansky, 2010; Kennedy, 2012). Intersective adjective interpretations emerge when the target is interpreted as a member of the intersection of two sets: the one denoted by the noun and the one denoted by the adjective (Kennedy, 2012). For example, the sentence 'Bobby is an expensive dog' implies that Bobby is both expensive (in general) and that he is a dog. In contrast, subsective interpretations emerge when the referent is interpreted as a member of a subset of the set denoted by the noun, returned by the adjective combining with the noun: For example, the sentence "John is a skillful surgeon" implies that he is a surgeon, but not necessarily that he is generally skillful - it only implies that he is skillful as a surgeon (Kennedy, 2012). Nonsubsective adjectives like 'former' can be disregarded for our purposes.

Although it is argued in the literature that specifically prenominal attributive adjectives give rise to ambiguity between the two readings (cf. 'Olga is a beautiful dancer'; Hofherr and Matushansky, 2010), it seems plausible to a priori treat gradable adjectives occuring in either position (attributively or predicatively) as eliciting

intersective interpretations (i.e. both amount to the reading 'the referent is both a dog and is generally big'), therefore leaving the comparison class underspecified. Positing a subsective reading for either sentence frame (i.e. 'the referent is big for a dog, but not generally big') would amount to the simple syntactic hypothesis wherein the noun sets the compison class, which intuitively can not hold in general (especially given examples like "John is a rich Fortune 500 CEO": positing a subsective reading would translate to the sentence "John is rich for a Fortune 500 CEO, but not rich in general", which intuitively isn't correct). The ambiguity mentioned above would then rather be ascribed to the comparison class ambiguity, than to the ambiguity of the denotation set.

Syntactic accounts outlined above stipulate that the meaning of an utterance involving gradable adjectives is fully specified by its words. Yet it was shown that several other pragmatic components like context of the utterance and listeners' world knowledge seem to come together when gradable adjectives are interpreted. Psycholinguistic studies investigated the role of these pragmatic factors for gradable adjectives and comparison class determination.

2.4 Pragmatic Aspects of Gradable Adjective Interpretation

Several visual world studies addressed the role of context for relative adjective interpretation, specifically focusing on manipulating the distractor set and the feature degrees of the distractors in the context in which a target described by an adjective was presented (e.g., Aparicio et al., 2016; Sedivy et al., 1999). They considered prenominally modified nouns in predicative position for the goal of reference resolution, showing that targets were identified faster when the context supported restrictive interpretation of the predicate (Sedivy et al., 1999; Aparicio, Xiang, Kennedy, 2015). That is, then context included another object of the same kind as the target, but with a different degree of the feature described by the adjective, such that the adjective was interpreted as helping 'restrict' the pair of objects to the target . By contrast, this effect was not shown in contexts including distractors of other categories than the target (Aparicio et al., 2016). Notably, both studies assume that comparison classes of gradable adjectives were supplied contextually, and more so when the context was more homogenous (i.e. included several objects of the same type) (cf. Aparicio et al., 2016), showing the context-dependent nature of comparison class determination.

Another recent study addressed the role of world knowledge for comparison class inference (Tessler et al., 2017). The authors showed that listeners flexibly adjust

comparison classes of gradable adjectives based on their world knowledge, when encountering simple utterances like "He's tall" said of targets about which listeners typically have strong expectations regarding the degree of the feature described by the adjective. They showed that listeners are more likely to infer that an utterance like "He's tall" said of a basketball player means "He's tall for a person" (i.e. relative to a general, basic-level category), whereas the utterance "He's short" rather means "He's short for a basketball player" (i.e. relative to the target's subordinate category). This pattern was clearly shown for targets of those categories which exhibit a rather high or low degree of the feature (e.g. basketball players, whose height is generally quite large; and jockeys, whose height is generally rather low) (Tessler et al., 2017). The present studies build on this experimental paradigm, making use of listeners' expectation about such categories highly salient with respect to some feature. When adjectives describing this certain feature scale are attributed of those categories, the basic-level comparison class is a priori more likely to provide a felicitous expression than the subordinate comparison class. However, the study by Tessler et al., 2017 only considered simple utterances, appearing without much context or a noun.

Chapter 3

A Fucntional Perspective on Comprison Class Inference

This section aims to combine the two lines of research and address both the role of the noun in the sentence as well as the role of pragmatic cues like perceptual context and world knowledge for comparison class inference. Specifically, the role of the noun is considered from a pragmatic perspective, since purely compositional approaches don't account for interpretative differences arising in different situations.

In this work, the issue of comparison class determination is approached from a functional perspective, based on the question what informational goals speakers might pursue when producing an utterance, and how these goals might influence listeners' comparison class inferences (Tessler et al., to appear). The proposed approach is an inferential account of comparison class determination, informed by the idea of recursive social reasoning mechanisms, applied to language in Gricean tradition (Goodman and Frank, 2016): Speakers have basic informational goals which guide how they craft their utterance in order to facilitate the message interpretation for a listener. Listeners, in turn, infer the most likely comparison class in light of those speaker goals (Tessler et al, ta). In particular, in order to communicate a property about a referent, speakers must achieve two goals: reference - identifying the right target - and predication - attributing a property to the target, specifically, communicating the degree of the feature denoted by the gradable adjective (Reboul, 2001; Tessler et al, ta). For the two stated informational goals, it is reasonable to posit that listeners generally expect the subject to be sufficient in order to establish reference - independent of the predicate asserted to hold of the subject - and that speakers aim to satisfy this general expectation (Reboul, 2001; Tessler et al, ta). This expectation might be based on general information structural reasons: In order to predicate a property of a target, i.e. potentially provide a comment in the sense of adding some information about the target, this target must be clear (Searle, 1969;

Krifka, 2008). There are exceptions to this tendency: e.g. the pronominal he can be resolved not only after applying the predicate, but also only taking into account the context of the sentence "The boss fired the worker because he was a convinced communist": He can either refer to the boss or to the worker (Reboul, 2001). But these syntactic expectations are a general enough heuristic holding in many contexts. These expectations have implications for comparison classes of gradable adjectives insofar as speakers have the liberty to choose from a priori similar sentence options to communicate the same message. For example, in order to tell some 4-year-old kids on a playground in winter that they built a big snowman, a speaker has the liberty to say "That's big!" pointing at the snowman, "That snowman you built is big!", "You built a big snowman!" and many other options. Consequently, the choice of a particular sentence over other equivalent options might respond to particular informational - communicative needs.

From this perspective, the influence of the noun on the comparison class in a simple "Subject Predicate" sentence depends on its position in the sentence (Tessler et al, ta). If the noun appears in the predicate of the sentence (e.g. in "That's a big Great Dane"), it can naturally be explained as produced by a speaker intending to constrain the comparison class. By contrast, if the noun appears in the subject of the sentence (e.g. in "That Great Dane is big"), it can potentially be explained away as produced by a speaker intended to support reference (especially via combining it with the deictic 'that'), and hence it is a weaker cue towards the comparison class, leaving it to other pragmatic cues like world knowledge (Tessler et al., 2017) or perceptual context to guide the comparison class inference. Hence, the utility of the noun as constraining the comparison class is the result of a trade-off between reference and predication, such that the noun is integrated with syntactic and contextual cues in order to infer the comparison class (Tessler et al, ta).

3.1   Understanding Reference and Predication

This reference-prediction trade-off hypothesis takes a rather linguistic point of view on phenomena treated in the literature from various perspectives. Searle (1969) conceptualizes both reference and predication as particular kinds of propositional acts, defining conditions to be fulfilled in order to accomplish them. Of particular importance for accomplishing reference is that the expression intended for reference isolates the target referent for the listener (Searle, 1969). Studies have shown that speakers are sensitive to contextual variability and adjust the informativity of their referential expression correspondingly, such that this requirement is satisfied (Graf et al., 2016). One of the requirements for accomplishing predication is that the

same sentence contains a reference to the intended target of predication (Searle, 1969; Reboul, 2001); together, these two conditions support the presented hypothesis. Whereas it is debated which linguistic expressions actually establish reference (Reboul, 2001; Michaelson and Reimer, 2019), it should be noted that the trade-off hypothesis specifically takes advantage of the flexibility of nouns with respect to both informational goals: combining with the deictic 'that' which acceptedly is taken to be referential (Reboul, 2001; Braun, 2017), the noun can accomplish reference; but the bare deictic 'that' can also be sufficient for reference if supported by the context, such that the noun of the sentence may then be 'left over' for accomplishing predication (Reboul, 2001).

3.2   Experimental Operationalization

In present studies, this property of nouns leads to the operationalization of the reference-predication trade-off hypothesis via a syntactic manipulation, wherein the noun (N) appears either in the subject or in the predicate of a sentence involving a gradable adjective ADJ (i.e. "That N is ADJ" (subject N), "That's a ADJ N" (predicate N); Experiment 1 - Experiment 3) (Tessler et al, ta). The critical question addressed by this distinction is how speakers and listeners treat these syntactic frames, asserting the ADJ of referents for whom they are felicitous given one comparison class, but not another (e.g. a normal-sized Great Dane can felicitously be described as 'big' given the comparison class 'dogs', but not 'Great Danes') (Tessler et al, ta). The reference-predication trade-off hypothesis predicts that nouns that are more likely to establish reference are less likely to constrain the comparison class. Conversely, when the noun is taken to contribute to predication, the noun is rather expected to be consistent with the comparison class felicitous in order to describe the target: i.e. the basic-level category label is more appropriate for setting the comparison class when describing a normal-sized target from a large-subordinate category as 'big', than the subordinate category label (cf. Tessler et al., 2017).

Note that although the differences in comparison class inference is approached through the lens of this syntactic manipulation, the underlying communicative goals are the primary driving force in comparison class inference, to which the syntax is just a cue, as supported by exploratory analyses of Experiment 3 (see Section 4; Tessler et al, ta) and Experiment 4. There might well be other syntactic realisations of these informational goals (Reboul, 2001): The sentence "What is big is that Great Dane" seems appropriate in a context where generally big things are discussed; in this utterance reference is accomplished from the predicate, and because of this referential pressure, under the trade-off hypothesis the noun would not be expected

to constrain the comparison class, although it appears in the predicate, supporting the view that the syntactic position of the noun is dissociable from the intended communicative goals.

To show that informational goals are primary for comparison class inference as opposed to specific syntactic properties of the adjectival phrase, another syntactic manipulation is employed in the fourth experiment wherein the noun appearing in the subject or predicate of the sentence is always syntactically modified by the adjective. This manipulation allows to disentangle the effect of the noun position from the effect of syntactic modification of the noun. For example, the critical sentences are "That big Great Dane is a prize-winner" (subject-N) or "That prize-winner is a big Great Dane" (predicate-N). The trade-off hypothesis predicts that even directly modified nouns in the subject position contribute to reference, and thus should be less likely to constrain the comparison class.

This work presents four behavioural experiments exploring the reference-predication trade-off hypothesis, specifically investigating the use of the size adjectives 'big' and 'small'. These two adjectives are chosen for practical reasons: size is a visually accessible feature, allowing for easy presentation and manipulation of the context in web-based experiments. Furthermore, humans usually have strong expectations about typical size distributions of different natural categories, from which the target referents were sampled for the experiments. Three distinct dependent measures were used to assess the influence of various cues on comparison class inference. Finally, the hypothesis is formalized computationally within the Rational Speech Act framework (Frank and Goodman, 2012). Extending a state-of-art model accounting for flexible use of comparison classes (Tessler et al., 2017), the model incorporates reasoning about the most likely comparison class via reasoning about the referential utility of possible utterances in context, aiming to qualitatively fit experimental data.

Chapter 4

Experiments

The reference-predication trade-off hypothesis was tested in four behavioural web-based experiments employing different dependent measures. The crucial manipulation in all experiments was the varying position of the critical noun - it appeared either in the subject (e.g., "That N is ADJ" or "That ADJ N is N2") or in the predicate ("That's a ADJ N" or "That N2 is a ADJ N") of the sentences presented in the experiments. These sentences described an object which appeared in visual context. These objects were sampled from five different basic-level categories (Rosch et al., 1976): dogs, birds, flowers, trees and fish. Within each basic-level category, at least two subordinate categories were chosen which exhibit a rather high or rather low amount of the feature described by the gradable adjectives under investigation - that is, those subordinate categories were chosen which people expect to be rather large or rather small representatives of their basic-level categories. For example, for the dog-category the large-subordinate category Great Danes and the small-subordinate category pugs were chosen. As shown by Tessler et al. (2017), when encountering representatives of such categories described by the adjective consistent with participants' prior expectations about the degree of feature-under-discussion, people are a priori more likely to infer the basic-level comparison class than the subordinate comparison class. For example, when encountering the sentence "It's big" said of a Great Dane (a large-subordinate category for the basic-level category dogs), humans are more likely to infer that the Great Dane is big relative to other dogs in general, than big relative to other Great Danes. Following the design of Tessler et al. (2017) allows to test the effect of syntactic position of the noun on how strong the noun is taken to constrain the comparison class: The reference-predication trade-off hypothesis predicts that nouns in the predicate position constrain the comparison class more strongly than in the subject position, such that a priori using the basic-level noun in predicate position is more felicitous in order to describe a normal-sized large-subordinate object (e.g. a Great Dane) than using a subordinate-label of the object

in predicate position. Both nouns would be felicitous in the subject position. In all experiments, the referents were described by the adjective matching prior feature-degree expectations, i.e. Great Danes and sunflowers were always described as big, and pugs or daisies as small. Table 1 shows all the stimuli used in all experiments.

The structure of all experiments was similar. First, participants completed a bot-check trial: Participants read a sentence where a named speaker asks a named listener: "It's a beautiful day, isn't it?". The speaker and listener names are sampled from lists of ten most popular male and female English names. For example, the sentences say: "John says to Mary: "It's a beautiful day, isn't?"; Who is John talking to?". Participants are asked to fill-in in lowercase who the listener is talking to. Participants are provided feedback and have maximally 3 attempts to fill-in the correct name. They are only allowed to proceed, if they successfully complete the bot check. Then, participants read instructions and complete practice trials, before completing main trials. After the main trials, they complete a socio-demographic post-test questionnaire, where they are asked to indicate their native language and optionally further information. For all experiments, participants were recruited via the crowd-sourcing platform Amazon's Mechanical Turk; only participants with IP addresses in the United States and work approval rating of at least 95

The first experiment (E1, Sentence Rating Experiment) was a sentence rating experiment, wherein participants had to rate two sentences which differed in the position of the noun and the specificity of the noun, as describing an object in context. In the second experiment (E2, Noun Production Experiment), participants had to fill-in the missing noun of a sentence describing the size of a referent in context. The position of the missing noun was varied. In the third experiment (E3, Comparison Class Inference Experiment), participants provided the inferred comparison classes via a free-production paraphrase, given sentences which varied by the noun and its position, as describing a referent in different contexts (Tessler et al., 2020). Finally, the fourth experiment (E4, Direct Modification Experiment) investigated a potential confound - the syntactic modification of the critical noun - and gathered inferred comparison classes in a paradigm akin to experiment three, but from sentences wherein the critical noun appearing in subject or in predicate was always syntactically modified by the adjective. All experimental materials can be found under https://github.com/polina-tsvilodub/refpred, the preregistrations can be found under . . . . . All experiments were realized using the _magpie - framework The data from these experiments provides a suite of evidence for the role of world knowledge, context and syntactic cues on the interpretation of the gradable adjectives "big" and "small".

4.1 Experiment 1: Sentence Rating Experiment

Results: main, by-participant / by-item variation, different exclusion criteria

The aim of the sentence rating experiment was to investigate whether participants prefer one syntactic frame over the other, given two truth-conditionally equivalent sentences, involving different nouns. The type of the noun and its syntactic position differed within-subjects. First, participants completed two warm-up trials to familiarize themselves with the slider rating procedure. On one trial, participants read: "Imagine you see this basketball" above a picture of an orange basketball, and read below the question: "How well does each of the sentences describe it? (Please click on the slider to provide a rating)". Two sentences appeared below: "The basketball is orange" and "The basketball is green", to be rated on sliders ranging from "very bad" to "very well". In the background, the ratings are mapped onto a scale ranging from 0 to 100. The slider is light gray, with a round handle appearing upon clicking on the slider track. The same sliders were used in the main trials. On the other trial, participants read: "Imagine you see this chair" above a picture of a purple chair. The sentences to be rated appearing below were: "The chair is yellow", and "The chair is blue". The order of the warm-up trials was randomized. Then, participants completed six main trials. Participants read "You and your friend see the following:" above a basic-level context picture (e.g. a group of dogs). The picture consists of six other members of the same basic-level category as the referent of the trial, including two other members of the same subordinate category as the referent. The six members consist of two members of a large-subordinate, a medium-sized subordinate, and a small-subordinate within the basic-level category each (e.g. the dog-context consists of two Great Danes, two poodles and two pugs). The context is used to set the overall reference comparison class for the targets. It also sets the visual reference frame. Below, they read the sentence "You also see this SUB_N", where SUB_N is the subordinate label of the target referent, which appears depicted below. The pictures depict referents a little smaller than members of the same subordinate category in the context, such that the felicitous comparison class is pushed towards the basic level category of the target. Below, the question about the critical sentences appears: "How well does each of the sentences describe it? (Click on the slider to provide a rating)". Then, the two critical sentences appear left of the sliders one below the other. The sliders range from "very bad" to "very well". On every trial, in one of the sentences the noun appears in the subject (e.g. "That N is big, small"), in the other in the predicate position ("That's a big, small N"). The order in which these syntactic conditions appear is randomized between-subjects. On half of the trials the noun is the basic-level target label (e.g. dog), on the other half it is the subordinate

target label (e.g. Great Danes), balanced within-subjects. Participants see each of the six possible context once, and for each context, one of the two possible targets (large-subordinate vs. small-subordinate category representatives) is sampled, balanced within-participants. Example trial view. The reference-predication trade-off hypothesis predicts that sentences with a basic-level noun in the predicate position should receive a higher rating than sentences with a subordinate noun in predicate position, but there should be no difference in the ratings of the sentences with a basic-level and subordinate noun in the subject position.

4.1.1 Participants:

113 participants were recruited and 33 were excluded for indicating a native language other than English, failing the practice trials or providing the same responses on every trial (see Appendix X). The experiment took about 5 minutes and participants were compensated $0.80. If partial data was missing from a participant, available data was used for analyses.

4.1.2 Results:

A Bayesian linear mixed-effects regression model was fit, predicting the sentence rating from the syntactic condition of the sentence (subject vs predicate noun), the noun type (basic-level vs. subordinate target label), their interaction and random by-participant and by-target random intercepts and random effects of syntax, noun type and their interaction. An exploratory model including a main effect of syntactic condition order was also fit, revealing no effect of syntactic condition order [...], so the data was collapsed across the two conditions. Both predictors are deviation coded, coding both the subject-noun and the basic-level noun as 1 and the other levels as -1, respectively. Consistent with the predictions, participants substantially dispreferred sentences with a subordinate noun in the predicate compared to the subordinate position, but no effect of syntax was found for the basic-level nouns, as indicated by the syntax X noun-type interaction [...]. Additionally, an overall preference for basic-level nouns [...] and the subject-noun syntactic structure [...]. In an exploratory analysis including a predictor of target size (small-subordinate vs. large-subordinate category) revealed that .... . Furthermore, a relatively high by-subjects and by-target variance revealed that .... .

4.2 Experiment 2: Noun Production Experiment

classification of responses results: main, by-participant / by-item, by-size

The goal of the noun production experiment is to investigate whether in a free-production setting participants produce nouns of different categories, given different syntactic frames. The noun of the critical sentences in the main trials appeared either in the subject position (e.g. in "That N is big,small") or in the predicate position (e.g. in "That's a big,small N"), manipulated between-subjects.

In this experiment, participants completed two blocks, each consisting of three warm-up trials and three main trials. In the warm-up trials participants familiarize themselves with the subordinate categories used in the main trials. They see pictures of a member from a large-subordinate and a small-subordinate category within one of the basic-level categories used in the main trials (e.g. a Great Dane and a pug). Participants are prompted to provide labels for these pictures. Below they are prompted to provide a common label for both pictures (i.e. dogs). They are provided feedback for the labels and can proceed upon adjusting their labels to correct responses. The number of attempts participants needed until they filled-in the correct labels is recorded. In this experiment, four additional subordinate categories were used, which can be found in Table 1 marked with *. For each participant, six out of ten possible contexts are sampled. Three of these contexts and their corresponding targets appear in the first experimental block, and the other three in the second. The trial order within the warm-up block and the main block is randomized. On the main trials, participants read: "You see the following:" above a basic-level context picture, akin to the contexts used in experiment 1. Below, they read "You also see this one:" and see a picture of the target referent. Then they read: "You say to your friend:", prompting them to fill-in the missing noun in the sentence: for the subject-noun condition, the template is "That＿＿ is big, small", for the predicate-noun condition, the template to be completed is "That's a big, small ＿＿ ". The size of the target referent was balanced within-participants: on three trials, participants see referents from a small-subordinate category, and on three, they see referents from a large-subordinate category. For each context, participants see only one of the possible targets (e.g. the large or the small subordinate target). Example trial view / example warm-up views The reference-predication hypothesis predicts that participants should be less likely to produce subordinate target labels in the predicate compared to the subject position.

4.2.1 Participants:

242 participants were recruited, and 52 were recruited for indicating a native language other than English or for failing the warm-up trials. The exclusion criterion was taking more than four attempts on any warm-up trial to provide the correct label upon correction. The experiment took about 7 minutes and participants were

compensated $1.00.

4.2.2 Results:

The responses provided by participants were categorized manually into basic-level or subordinate-level labels of the targets. X % were superordinate referent labels (i.e. more general labels like animals) and were collapsed with basic-level labels. 16 (1.4%) uncategorizable responses were excluded from analysis. A logistic generalized mixed-effects regression model was fit, regressing the response category (basic-level. vs subordinate target label) against the syntax of the sentence, random by-participant and by-referent intercepts and random by-referent slope effects of syntax. Consistent with predictions, a strong effect of syntactic position of the noun was found, indicating that participants were appreciably more likely to use basic-level labels in the predicative position (2.25 [0.74, 4.01]). Furthermore, different participants showed different sensitivity to the syntax, as indicated by... Additionally, by-referent variation was found, which could be attributed to differing namability of the targets. .... An exploratory model including a main effect of referent size (large-subordinate vs. small-subordinate category) revealed that...

4.3 Experiment 3: Comparison Class Inference Experiment

The two previous experiments support the reference-predication trade-off view, by showing that participants disprefer sentences like "That's a big Great Dane" in order to describe a normal-sized Great Dane. The goal of this comparison class inference experiment was to measure comparison class inferences more directly, as influenced by the position of the critical noun in the sentence, the type of noun and the visual context of the sentence. All three factors are manipulated within-subjects. When participants don't have access to visually assessing the size of a referent and have to infer the comparison class from the sentence, they should be sensitive to linguistic cues like the sentence structure. According to the outlined hypothesis, they should be more likely to take the noun as a cue to the comparison class when the noun appeared in the predicate of that sentence, than when it appeared in the subject. When the noun appeared in the subject, comparison class inference can be driven by other pragmatic inference, e.g. from world knowledge and visual context. First, participants completed a comparison class paraphrase practice trial, akin to the paradigm employed in the main trials. Participants were told that on the main trials they will see a sentence containing a word that is relative, and their task will be

to figure out what this word is relative to. They read an example task: "Speaker A: 'The Empire State building is tall.' What do you think speaker A meant?". Below they saw a paraphrase template where they provided the inferred comparison class of the adjective tall: "The Empire State building is tall relative to other_" (blank to be completed with the inferred comparison class). Participants were provided feedback on their response and had to correct it to one of the possible options among buildings, skyscrapers, houses, constructions. Then, participants completed two blocks consisting of labeling warm-up trials and main paraphrase trials. Three of the six basic-level categories used in this experiment (table 1) are sampled for the first block, with the respective subordinate category members appearing in the warm-up trials, the other three categories appear in the second block. These labeling warm-up trials are of the same kind as in Experiment 2. In this experiment, for the main trials there were basic-level and subordinate-level contexts for each possible referent. Basic-level contexts were identical to the contexts of respective categories in Experiment 1 and Experiment 2; the subordinate contexts consisted of six other representatives of the same subordinate category as the target referent. For example, the subordinate context for a Great Dane consisted of a picture of a group of six other Great Danes. Within each main trial block, there were six trials, wherein for each of the three sampled categories, one possible referent appeared in the corresponding basic-level context (e.g. for the category flowers, the sunflower appeared in basic-level flower context), and the other possible referent appeared in the corresponding subordinate context (i.e., then the daisy appeared in subordinate daisies-context). The referent was described by a critical sentence in which the noun could appear in the subject or in the predicate of the sentence. The noun could be either the basic-level (e.g. dog) or the subordinate label of the referent (e.g. Great Dane). Furthermore, a baseline condition with an anaphoric 'one' in the noun position was included, in order to measure the baseline influence of the visual context on comparison class inference (Goldberg and Michaelis, 2017). Crossing the visual context (basic vs. subordinate), the syntax (subject-N vs. predicate-N) and the possible nouns (basic vs. subordinate vs. 'one') results in a 2x2x3 design, yielding 12 unique conditions. Each participant saw each condition once during the total of 12 main trials. On main trials, participants read "You and your friend see the following:" above a context picture. Below, they read: "Your friend runs far ahead of you, and you see him in the distance:". The illusion of distance was created contextually in order to disguise the perceptual size of the target referent and push participants towards inferring the size of the referent from the sentence, rather than perceptually. This illusion was supported by the picture appearing below, wherein the small target referent was depicted next to a small person (as compared to the context, i.e. appearing in distance). Below, participants read: "Your friend says:",

followed by the critical sentence. Participants were asked "What do you think your friend meant?", followed by the paraphrase template "It is big, small relative to other __", blank to be completed with the inferred comparison class.

4.3.1   Participants:

245 participants were recruited and 45 were excluded for indicating a native language other than English, or failing either the comparison class inference practice trial or the labeling warm-up trials more than four times upon correction. The experiment took about 9 minutes and participants were compensated $1.20.

4.3.2   Results:

by-target / by-participant variation by-size? Participants' responses were manually classified into basic-level and subordinate comparison classes. X superordinate comparison classes were collapsed with the basic-level responses. 39 (1.6%) uncategorizable responses were excluded from the analysis. A Bayesian generalized logistic mixed-effects regression model was used, regressing the response category against the syntactic condition (subject-N vs. predicate-N), the noun category (basic vs. subordinate vs. 'one'), the context (basic vs. subordinate), their two-way and three-way interactions and maximal random effect structure appropriate for this experimental design (footnote formula). The results indicate that participants flexibly adjust the inferred comparison class according to many factors. First and foremost, a large effect of visual context going above and beyond other factors was found, providing evidence against a simple purely syntactically-oriented theory of gradable adjective interpretation; as indicated by the inferences drawn from the baseline condition anaphoric 'one'. Furthermore, an effect of noun regardless of its position in the sentence was found: participants were more likely to infer basic-level comparison classes from basic-level nouns than from subordinate nouns. Notably, the subordinate comparison class was the minority response given a subordinate noun in the basic-level context, speaking against a modificational view of adjective comparison classes. Crucially, a credible syntax X noun interaction was found, supporting the reference-predication trade-off hypothesis: more subordinate comparison classes were inferred from subordinate nouns appearing in predicate position than in the subject position, compared to basic-level nouns. Exploratory model with the main effect of context; comparison to one – subordinate noun driving the interaction

30

4.4 Experiment 4: Direct Modification Experiment

In order to keep a simple and interpretable operationalization of the reference-predication distinction, a potential confound was introduced in Experiments 1-3. The position of the noun was perfectly confounded with whether the noun was syntactically modified by the adjective (predicate-N condition) or not (subject-N condition). However, the reference-predication trade-off view predicts that referential pressure takes off some weight from the noun used for reference and decreases its strength in constraining the comparison class independent of the syntactic modification. This prediction was investigated in this Direct Modification experiment. In this experiment, the position of the critical noun in the sentence was varied, and the noun was always directly modified by the adjective big or small. The critical nouns were always subordinate referent labels. To create respective sentences, a second noun was used which described a visually salient feature of the referent. For example, the referents for one of the dog contexts were prize-winners, as indicated by prize-bows depicted on the referents. So the critical sentence was either "That prize-winner is a big Great Dane" (predicate-N) or "That big Great Dane is a prize-winner" (subject-N). The referents appeared in a basic-level context, which included two other members of the same subordinate category as the referent, and two other individuals with the feature described by the second noun of the sentence, e.g. in the dog-context there were two other prize-winners. Because the reference-predication trade-off is based on explaining away a noun via its potential referential use, through this context manipulation the referential utilities of the two nouns of the sentence is equated, such that only the noun's syntactic position and combination with the deictic 'that' could provide a cue towards referential use. Therefore the critical subordinate noun is expected to constrain the inferred comparison class more strongly when it appears in the predicate of the sentence than in the subject. The experimental set-up was similar to the set-up of experiment 3. Five different contexts were used in this experiment: there were two dog contexts, a flower, a bird and a tree context (Table 2). Four out of five contexts were sampled for each participant. Participants completed two experimental blocks, each consisting of warm-up and main trials using two of the sampled categories. In the first block, participants first completed three rounds of labeling warm-up trials. A round consisted of a demonstration trial where participants saw two subordinate members of a basic-level category used in this block and read their labels. For example, they saw pictures of a sunflower and a daisy next to each other and read "This is a sunflower" and "This is a daisy", respectively. They could proceed after 3.5 seconds to the

next trial where they had to label other instances of the same categories themselves. They also had to provide a common label for the pictures (i.e. flowers). The order of the pictures was randomized between-participants. They were provided feedback on their labels and could proceed only after correcting their labels. After two labeling warm-up rounds, participants completed two demonstration trials of at least 3.5 seconds each learning about the additional features of the referents described by the second noun of the critical sentences in main trials. For example, participants saw a picture depicting the sunflower and the daisy in pots with bows, and read: "These flowers are gifts. Notice the bow on the pots.". Finally, participants completed a comparison class paraphrase practice trial, identical to the one used in experiment 3. The warm-up trials in the second experimental block were identical, but there was no paraphrase practice trial. Then, participants completed four trials - two main and two filler trials, in randomized order, where a filler trials was always the first trial of the block. In the main trials, a subordinate referent with an additional feature (e.g. a prize-winner bow) appeared in the corresponding context, as described above. Participants read different context stories for each context (table 2). For example, for a flower context, they read "You and your friend are at their garden and you see the following:" above the context picture. Below, they read "Your friend runs far ahead of you. You see your friend in the distance:", followed by a depiction of the referent with the additional feature next to a person; to induce the illusion of distance, both were small relative to the context picture. Then they read "Your friend says:", followed by the critical sentence. Finally, they were asked: "What do you think your friend is saying it is big, small relative to?", introducing the paraphrase template, like in experiment 3. For a given category, one of the possible targets appeared in this critical trial (e.g. the sunflower). The other possible target (i.e. the dandelion) then appeared in a filler trial in the same block. Filler trials were identical to main trials with basic-level contexts from experiment 3. The size of referent (i.e. large-subordinate vs. small-subordinate) was counterbalanced across syntactic conditions and trial types within-participant, resulting in 8 unique conditions. Each participant saw each condition once, resulting in eight main trials.

4.4.1 Participants:

4.4.2 Results:

Chapter 5

A Bayesian Reference-Predication Model

Chapter 6

Discussion

Declaration

I declare that..

Appendix A

Appendix

Bibliography

Aparicio, H., Xiang, M., & Kennedy, C. (2016). Processing gradable adjectives in context: A visual world study, In *Semantics and linguistic theory*.

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

Barner, D., & Snedeker, J. (2008). Compositionality and statistics in adjective acquisition: 4-year-olds interpret tall and short based on the size distributions of novel noun referents. *Child development*, *79*(3), 594–608.

Bartsch, R., & Vennemann, T. (1972). Semantic structures: A study in the relation between semantics and syntax. *Athenäum-Skripten Linguistik Bd*, *9*.

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

Cinque, G. (2010). *The syntax of adjectives: A comparative study* (Vol. 57). MIT press.

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

Ebeling, K. S., & Gelman, S. A. (1994). Children's use of context in interpreting "big" and "little". *Child Development*, *65*(4), 1178–1192.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818–829.

Heim, I. (2000). Degree operators and scope, In *Semantics and linguistic theory*.

Hofherr, P., & Matushansky, O. (2010). *Adjectives: Formal analyses in syntax and semantics*. John Benjamins Publishing Company.

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

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

Kennedy, C. (2012). Adjectives. In D. Fara & G. Russell (Eds.), *The routledge companion to philosophy of language*. Routledge. https://books.google.de/books?id=cV9LvekiKKAC

Klein, E. (1980). A semantics for positive and comparative delition. *Linguistics and Philosophy*, *4*(1), 1–46.

Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, *194*(10), 3801–3836.

McNally, L., & Kennedy, C. (2008). *Adjectives and adverbs: Syntax, semantics, and discourse*. Oxford University Press.

Qing, C., & Franke, M. (2014). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model, In *Semantics and linguistic theory*.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382–439.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/s0010-0277(99)00025-6

Sera, M., & Smith, L. B. (1987). Big and little:"nominal" and relative uses. *Cognitive Development*, *2*(2), 89–111.

Sinelnikova, A. (2020). *Cues to comparison classes in child-directed language* (M. Eng. Thesis). Massachusetts Institute of Technology.

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

Stechow, A. v. (1984). Comparing semantic theories of comparison. *Journal of Semantics*, *3*(1-2), https://academic.oup.com/jos/article-pdf/3/1-2/1/9835673/1.pdf, 1–77. https://doi.org/10.1093/jos/3.1-2.1

Tessler, M. H., Lopez-Brau, M., & Goodman, N. D. (2017). Warm (for winter): Comparison class understanding in vague language, In *39th annual meeting of the cognitive science society*.

Tessler, M. H., Tsvilodub, P., Snedeker, J., & Levy, R. P. (2020). Informational goals, sentence structure, and comparison class inference, In *42th annual meeting of the cognitive science society*.