Once again on norms and comparison classes

ELENA TRIBUSHININA

Abstract

A central assumption about relative adjectives (e.g., big, old) is that their positive form is interpreted vis-à-vis a class-specific reference point located in the midzone of a series (norm). More recently, functional-cognitive studies argued that other reference points (e.g., argumentative zero, endpoints) are more relevant in actual language use than a norm. This paper argues that the two positions are not necessarily irreconcilable and experimentally tests a hypothesis that a norm is a default reference point used for the interpretation of relative adjectives in zero contexts. Experiment 1 addressed the location of a norm in the midzone of a series and its category dependence. As predicted by the traditional semantic studies, the cutoff point between "big" and "small" is located around the midpoint of a scale. Furthermore, its location is category-dependent and sensitive to prototypicality effects. The results further indicate that adults easily compute a contextually relevant norm by integrating their world knowledge with the visually provided information. Experiment 2 investigated the relevance of a norm in the online processing of relative adjectives. The results suggest that language users indeed exploit norms for the interpretation of relative adjectives in real time.

1. Introduction

Relative adjectives such as *tall* have varying denotations. By way of illustration, consider the following examples from the British National Corpus (BNC):

(1) *All at once I could see Galway Bay.* ***Tall****, dim* ***mountains*** *on a curve of horizon, a winking lighthouse and the open sea.*

(2) *Their houses, the* ***tall****, crumbling* ***tenements*** *with their cracked roof tiles and their creaking balconies huddle together round the church . . .*

Linguistics 49–3 (2011), 525–553
DOI 10.1515/LING.2011.016

0024–3949/11/0049–525

(3) *Even today Scots tend to be wary of clever* ***women****; but in those days, to be young, female,* ***tall****, beautiful, witty, talented and intelligent — and a Queen — was like writing one's own death sentence.*
(4) *He wanted to lie down, in the sunshine, lost in this* ***tall****, scented* ***grass*** *and go to sleep.*

A tall mountain in (1) is probably thousands of meters tall. Tall tenements in (2) might be between 5 and 10 meters tall. A woman in (3) is no more than two meters tall. And the tallness of grass in (4) is likely to be measured in centimeters. Nonetheless, language users seem to have no trouble with assigning these rather different values to the same adjective or recalibrating adjective meaning through the local contexts created by the head nouns (Kamp and Partee 1995; Partee 1995, 2007). This holds not only for *tall*, but also for other relative adjectives, such as *big, old* and *cheap* (Kamp 1975). This intriguing property of relative adjectives explains why they have been a focus of semantic studies for several decades. Semanticists have argued on numerous occasions that the interpretation of relative adjectives in the positive (i.e., noncomparative) form includes locating a property on a gradual scale vis-à-vis a class-specific reference point. This reference point has traditionally been called a *norm* (Apresjan 1974; Arutjunova 1999; Bierwisch 1967, 1989; Lang 1989; Lehrer and Lehrer 1982; Leisi 1975; Lyons 1977; Sapir 1944; Taylor 1992). In other, especially more recent theories the terms *standard value* (H. Clark 1973; Pander Maat 2006; Rotstein and Winter 2004; Vendler 1968), *pivotal region* (Cruse 1986; Paradis 1997), *relative standard of comparison* (Kennedy 2007; Kennedy and McNally 2005; Syrett 2007) and *cognitive zero* (Šabes 1989; Tribushinina 2008a) are used. For perspicuity, I will confine myself to the most wide-spread term — *norm*.

A norm is usually defined as an average value of the property denoted by an adjective and is claimed to be located in the midzone of a scale between the ranges of the antonymous terms (Bierwisch 1967, 1989; Katz 1972; Lang 1989; Lyons 1977; Sapir 1944; Siegel 1980; Vendler 1968, *inter alia*). Another crucial property of a norm is its category dependence. Norms are defined in relation to a specific *comparison class*, i.e., "a subset of the universe of discourse which is picked out relative to a context of use" (Klein 1980: 13). Comparison classes are intrinsically subjective. For instance, a person from Western Europe will probably have a different expectation about average winter temperatures than a resident of Siberia and will, therefore, assign a different value to the adjective *warm* in a sentence like *This winter is surprisingly warm*. Thus, norms largely rely on what we experience as an average value of a property, rather than on some objective parameters in the world.

A comparison class may be provided explicitly, as in (1)–(4), but it can also be supplied from a broader linguistic and/or nonlinguistic context (world

knowledge). On different occasions, the same sentence (e.g., *They have a big house*) can be interpreted with reference to different comparison classes (e.g., all houses a speaker has ever seen or lived in, all houses in a particular town, neighborhood, street). Thus, comparison classes and their average values are dynamically construed in specific communicative situations on the basis of various linguistic and nonlinguistic cues (cf. Partee 2007). Barner and Snedeker (2008) found that even four-year-old children are able to define and redefine comparison classes and to compute average values associated with them. When presented with nine novel objects (called *pimwits*) of different size, the subjects labeled about one third of them *tall* and about one third *short*; the remaining objects constituted the norm — objects of average size that were neither *tall* nor *short*. In this case, the norm was located exactly in the middle of the scale because the whole comparison class was visually given and because the values were evenly distributed. Most interestingly, when the comparison class was changed (by adding taller or shorter objects to the series), the four-year-olds' scalar judgments changed accordingly.

It is important to remark that only the positive form of relative adjectives is claimed to be norm dependent. Comparative and superlative adjectives, as well as adjectives used in *how*-questions (e.g., *How tall is John?*), constructions with measure phrases (e.g., *two meters tall*) and consequential grading constructions (e.g., *too tall, tall enough*) are not contingent upon norms and evoke other points of reference such as a zero plane, incidental landmarks, maximum and minimum values (Tribushinina 2008a).

Recently, however, the well-established idea that the positive form of a relative adjective without a complement is always associated with a norm (cf. Bierwisch 1989) has been increasingly called into question by linguists working in the usage-based (functional-cognitive) paradigm (e.g., Pander Maat 2003, 2006; Taylor 1992; Tribushinina 2008a). Using naturally produced language data (corpora), these studies have shown that norms cannot account for all uses of relative adjectives in the positive form and that other reference points are more important to the interpretation of relative adjectives in real-life communicative situations than a norm.

Inspired by this recent debate, the purpose of this paper is to re-examine the long-standing hypothesis that relative adjectives in the positive form are interpreted vis-à-vis a class-specific norm using new methodology (offline and online comprehension experiments).

The structure of the paper is as follows. Section 2 summarizes the claims regarding the applicability of a norm made by the two camps (traditional semantics *vs.* functional-cognitive linguistics) and introduces a new unified account which brings the two approaches together by postulating a default status of a norm. Section 3 further elaborates on the objectives of the present study. Section 4 reports the results of two comprehension experiments investigating

the relevance of a norm in the interpretation of relative adjectives. General conclusions are presented in Section 5.

2. Theoretical background

In this section, I will review the claims made by the two camps regarding the relevance of a norm. Section 2.1 presents arguments for maintaining the applicability of norms in the semantic analysis of relative adjectives. Section 2.2 summarizes relevant findings from corpus studies indicating that other reference points may be more important to the interpretation of relative adjectives in the positive than a norm. In Section 2.3, I introduce an alternative view suggesting that a norm is a *default* reference point for relative adjectives.

2.1. *Traditional view: explanatory power of a norm*

It is not without a reason that analyses of relative adjectives in terms of norms and comparison classes have been most influential for, at least, the last sixty years. In this section, I will summarize the advantages of using a norm in the semantic analysis of relative adjectives. Two aspects will be given primary attention: context dependence and antonymy.

2.1.1. *Context dependence.* Most authors maintaining the division of gradable adjectives into relative and absolute consider context dependence of a reference point as a basic criterion for distinguishing between the two groups of adjectival words (Kamp and Partee 1995; Katz 1972; Kennedy 2007; Kennedy and McNally 2005; Lehrer and Lehrer 1982; Nelson and Benedict 1974; Partee 1995, 2007; Rips and Turnbull 1980; Rotstein and Winter 2004; Sapir 1944; Syrett 2007; Syrett et al. 2010; Vendler 1968). Absolute gradable adjectives, it is argued, are interpreted vis-à-vis a context-independent standard of comparison. For example, a container can be dubbed *empty* if there is nothing left in it. What is empty for a glass is also empty for a pan. By contrast, a standard of comparison for relative gradable adjectives (norm) is determined by context. What is *big* for a glass is not necessarily big for a pan.

Rips and Turnbull (1980) investigated processing consequences of this distinction. They asked subjects to verify sentences containing either relative or absolute adjectives used predicatively and attributively. Reaction times and error rates for relative adjectives decreased when the comparison class was made explicit by combining an adjective with a noun denoting the superordinate category of a subject (i.e., in attributive uses). This shows that computation of an implicit comparison class incurs a processing cost in the case of

relative adjectives, but not in the case of absolute adjectives. Rips and Turnbull conclude that reference points for relative adjectives depend on a contextually relevant comparison class, whereas reference points for absolute adjectives are independent of such a standard.

Class specificity of a norm accounts for the fact that the same adjective may have varying denotations, as shown in (1)–(4). Category-dependence also explains why sentences like (5) are not interpreted indexically (Kennedy 1999; Klein 1980; Ludlow 1989).

(5) *That elephant is large and that flea too.*

In (5) we do not compare the flea directly to the elephant. Rather, the adjective *large* is assigned two different values with respect to two different norms — one for elephants and one for fleas. Thus, the elephant is large for elephants and the flea is large with respect to fleas.

2.1.2. *Antonymy.* A norm is also of paramount importance to the account of antonymy. Relative adjectives are opposites *par excellence* that fully satisfy Lyons' (1977) definition of antonyms. True antonyms establish converse relations in comparative constructions allowing for inferences of the following kind: *Boris is taller than Mary* $\Rightarrow$ *Mary is shorter than Boris*. The assertion of one term implies the negation of its antonym: *It is tall* $\Rightarrow$ *It is not short*. The negation of one term, however, does not entail the assertion of the other: *It is not tall* $\not\Rightarrow$ *It is short*. This type of semantic relations is called *contrariety*. Norms play an important role in the establishment of contrary semantic relations between polar antonyms. The negation of one term does not lead to the assertion of the other, since there is also a midzone, where neither of the antonymous adjectives applies. This midzone, also known as *extension gap* (Klein 1980), *zone of indifference* (Lyons 1977) and *range of indeterminacy* (Partee 1995), is where a norm is located.

There are, however, more aspects of antonymy that can be elucidated in terms of norms. As has often been pointed out in the literature, perfect antonyms must denote values equidistant from a norm (Croft and Cruse 2004: 166; Cruse 1976: 282; De Schutter 1976: 24; Lehrer and Lehrer 1982: 487; Murphy 2003: 186; Murphy and Andrew 1993: 302; Sapir 1944: 133). The adjectives, *big* and *small*, for example, are better opposites than *big* and *tiny*. The reason is that the ranges denoted by *big* and *tiny* are not symmetrically disposed about a norm. *Big* profiles the "big" zone, which is disposed symmetrically with the "small" zone, while *tiny* profiles the "very small" zone, which is disposed symmetrically with the "very big zone" on a gradual scale.

Yet another argument in favor of a norm hinges on the observation that a positive-pole adjective can be noncanonically contrasted with a negative-pole

adjective denoting a different dimensional axis, because (+Pol) and (−Pol) terms are located on the opposite sides of the scale vis-à-vis a norm (Murphy 2004). For instance, the antonym of *tall* is *short*. Indeed, the two adjectives are often used as opposites of each other in the corpus; see, for instance, (6) and (7). This, however, does not preclude the use of the (−Pol) terms *low* or *small* as the opposites of *tall*, as in (8) and (9):

(6) *Seems so funny having a **tall** dog after having **short** dogs*. (BNC)
(7) *Alright he's probably blindfolded her after that, but at least she should have some kind of er Yeah whether he were dark hair or light hair or **tall** or **short**.* (BNC)
(8) *Each line had its speciality; it might comprise **exceptionally small** girls, or **very tall** ones, or talented acrobats, but whatever the speciality, all could sing and dance.* (BNC)
(9) ***Lower** stances, for instance, are more stable than **taller** ones, and, being solid and powerful, are naturally resistant to sudden attacks*. (BNC)

The basis of antonymy in (8) and (9) is the fact that the terms *small* and *low* are different from *tall* in one crucial respect, namely the profiling of the negative region on a scale and placement of the reference value in the area below a norm. It is important to observe that co-occurrence of canonical and noncanonical antonyms is a lexical matter rather than a semantic property. However, lexical decisions are motivated by conceptual facts. It is not just a matter of co-association and co-occurrence that *tall* and *short* (and sometimes *small* and *low*) belong together. These relations have a solid semantic basis, norm being an important part of it.

In a similar vein, Bartlett (1976) found that polarity of dimensional adjectives is more prominent than their dimensional features early in development. Therefore, so the argument goes, children first learn the dimensional hyperonyms *big* : *small*. Other dimensional adjectives are acquired as more elaborate (in the spatial sense) hyponyms of *big* and *small*. In other words, *short* is acquired as a hyponym of *small*, rather than antonym of *tall* or *long* (cf. E. Clark 1973; Eilers et al. 1974). These relations between (+Pol) and (−Pol) terms seem to persist in the language of adults as well (Paradis et al. 2007, 2009). The fact that the position vis-à-vis a norm dominates over objective spatial properties in the construals of dimensional relations reinforces the importance of this reference point in the treatment of relative dimensional adjectives.

To summarize, in this section I have reviewed some evidence suggesting that a norm plays an important role in the semantics of relative adjectives. First, the application of norms accounts for the fact that the interpretation of relative adjectives is context dependent. Second, the presence of a norm in the midzone of a scale explains why the negation of one term does not imply the assertion of its antonym. Third, the position of the antonyms' semantic ranges

vis-à-vis a norm provides a good explanation for an intuition that *huge* is a better antonym of *tiny* than, for instance, *big*. And, finally, a norm divides the scale into a (−Pol) and a (+Pol) region, which sanctions non-canonical contrast relations between different-polarity adjectives.

Before closing this section, it is important to notice that not all semanticists would take an extreme stance that relative adjectives in the positive form are *always* associated with a norm (cf. Apresjan 1974; Bierwisch 1967, 1989; Katz 1972; Vendler 1968). However, what seems to be the case for most semantic studies is that they consider a norm if not the only then, at least, the most important reference point in the semantic make-up of relative adjectives (cf. Arutjunova 1999; Leisi 1975; Lyons 1977). In this respect, "traditional" semantic studies are different from the functional-cognitive studies postulating a primary status of reference points other than a norm. I turn to the latter line of research in the following section.

2.2. *Functional-cognitive approaches to relative adjectives*

Cognitive psychologists and linguists taking a usage-based approach to language have recently presented a number of counterarguments to the overall applicability of a norm and suggested that other reference points are more relevant to the interpretation of relative adjectives in natural-language contexts.

2.2.1. *Problem of identification.* The first problem of the norm-based approach that is commonly mentioned in usage-based studies is that it is often quite difficult, if not impossible, to identify the comparison class involved, let alone the average value associated with it (Hutchinson 1993; Pander Maat 2003, 2006; Taylor 1992; Tribushinina 2008a). For instance, it is not at all clear which comparison class should be used for interpreting a sentence like *The universe is large* (Pander Maat 2006). And even if the comparison class is identified, it is even more difficult to define its average value. For one thing, in the case of relative adjectives denoting non-measurable properties (e.g., *good*) it is almost impossible to find an average.

2.2.2. *Norm-free readings.* Another major argument against the norm-based approach is that the positive form of relative adjectives is often interpreted vis-à-vis reference points other than a norm. For instance, Ebeling and Gelman (1994) suggest that a sentence like *The hat is big* may have, at least, three different interpretations: a *normative* one (bigger than an average hat), a *perceptual* one (e.g., bigger than the hat next to it) and a *functional* one (e.g., too big for a tiny doll). Notice that only one of these three readings — the normative one — has to do with a norm. The other two are not contingent on an

average value. Yet neither the perceptual, nor the functional readings are odd. On the contrary, they are quite normal and frequent (Tribushinina 2008a). Moreover, Ebeling and Gelman (1994) have shown that even 2-year-old children are sensitive to the difference between the three readings. They, however, switch more easily from a normative context to a perceptual/functional one than the other way around, probably due to abstractness of the former and visual salience of the latter. This suggests that perceptual and functional readings are not only usual, they even tend to be more cognitively salient than norm-related readings, which is consistent with the accounts positing a nonprimary status of a norm (e.g., H. Clark 1973; Pander Maat 2006).

2.2.3. *Alternatives to a norm.* If a norm is not primary to the semantics of relative adjectives, which reference point *is* primary then? H. Clark (1973) argues that a zero plane at the ground level is a primary reference point for all uses of spatial adjectives. The following quotation deserves citing in full:

> Consider the adjectives *high* and *low*. To say *The balloon is high* (or *low*) is really to say *The balloon is high* (or *low*) *off the ground*. Implicit in such simple statements is a zero point, an origin, the point of reference from which all measurement is taken. *High* and *low* happen to have a particular reference plane — ground level — unless some other reference plane is mentioned explicitly. This origin or zero point could be called the primary point of reference. Adjectives also have a secondary point of reference. *High* and *low*, to continue the example, both refer to height off the ground, but *high* indicates a distance that surpasses some implied standard, and *low* indicates a distance that fails to meet that standard. This standard depends very strongly on what exactly is being measured, as many linguists have pointed out, for one would describe a balloon as high in a room when it was perhaps 6 feet high, but in a large auditorium perhaps only if it was 10 to 20 feet high. The main point here is that *high* has two implicit reference points: ground level (the primary one) and some standard height (the secondary one). (H. Clark 1973: 36)

Thus, H. Clark (1973) suggests that both reference points — a zero value and a norm — are relevant to the interpretation of spatial adjectives. However, a zero plane is argued to be primary and a norm is attributed a secondary status. Tribushinina (2009b) argues that a zero point can be relevant not only for adjectives of vertical extent/position (such as *high* and *low*), but also for non-spatial relative adjectives, since the domain of height serves as a basic cognitive schema structuring our experience with and understanding of, other, more abstract scales (cf. Johnson 1987; Lakoff 1987).

Pander Maat (2006) posits the primary status of what he terms an *argumentative zero*. This argumentative reference point is a degree of the property which is sufficient to support a particular conclusion. By this analysis, (10) is interpreted as ‘too high to climb’ rather than ‘higher than other mountains’ (cf. Ebeling and Gelman’s [1994] functional reading). Pander Maat maintains that

all uses of relative adjectives are anchored by an argumentative reference point rather than by a norm (see also Verhagen 2005: 13).

(10) *Wow, this mountain is high!*

Using English and Russian corpora, Tribushinina (2008a) shows that conceptual specifications of relative adjectives are frequently anchored by *incidental landmarks*, i.e., a degree of the property in another entity, as in comparatives and superlatives (cf. Ebeling and Gelman's [1994] perceptual reading). See, for instance example (11):

(11) *For example, a person witnesses the following events in a swimming pool: A* ***tall adolescent boy*** *walks purposefully up behind a* ***small coloured child*** *and pushes him strongly into the pool.* (BNC)

A norm does not seem to be relevant to the interpretation of the positive forms *tall* and *small* in (11). The speaker does not claim that the adolescent is tall for his age and the child is small vis-à-vis the average size of children of his age. What probably matters here is that the adolescent is tall with respect to the child, who, in his turn, is small as compared to the adolescent. Another possibility is that the adjectives in (11) are used categorically, in the sense that little boys are always small and adolescent boys are always tall (Kristen Syrett p.c.).

In summary, several psychological and cognitive linguistic studies have argued that a norm may not always be the primary reference point for relative adjectives, since it is difficult to identify and often irrelevant to the interpretation of relative adjectives in the positive form. Other reference points, such as an absolute zero and an argumentative reference point were argued to be primary. In what follows, I will focus on the primary status of a norm contested by the recent functional-cognitive studies. The norm-free readings and alternative reference points briefly discussed in this section will not be further studied in this paper. For research along these lines the reader is referred to Paradis (2005), Tribushinina (2008a) and Holleman and Pander Maat (2009).

2.3. *A unified approach: construals with multiple reference points*

Although the functional-cognitive studies reviewed above explicitly contrast their view with the traditional semantic research when arguing against the overall applicability of norms, I would like to suggest that the two views are not necessarily incompatible (cf. Partee 2007). Pursuing the approach introduced in Tribushinina (2008a), I argue that each adjective can, in principle, be interpreted vis-à-vis several different reference points such as a norm, incidental landmarks, prototypes (Tribushinina 2009a), endpoints of a scale (Tribushinina 2009b) and the self (Tribushinina 2008c). Not all reference points are

equally salient at the same time — one of them usually gains a primary status. Local choices of reference points and their relative salience with respect to each other is a matter of dynamic construal, i.e., alternate conceptualisations of the same objective scene (cf. Paradis 2005). A reference point may be primary *by default* or by virtue of *contextual salience.* Default reference points anchor context-free uses of adjectives and motivate their basic entailment patterns (cf. Section 2.1). However, a default can be easily overridden in actual communication where non-default reference points often gain primary salience under contextual or constructional constraints.

For instance, the default reference point for absolute gradable adjectives such as *full* and *empty* is an endpoint of a scale — maximum fullness for *full* and maximum emptiness for *empty* (Kennedy and McNally 2005). Frazier et al. (2008) have shown that this default reference point is involved in the online processing of sentences containing absolute adjectives. However, in some contexts this default reference point may be overridden by another reference point, as in (12) and (13):

(12) *The restaurant is very full tonight*.
(13) *The restaurant is very empty tonight*.

Kennedy and McNally (2005: 371), among others, suggest that, when modified by *very*, absolute adjectives have to be interpreted vis-à-vis a relative standard of comparison (norm) rather than the endpoints of the scale. So, (12) does not depict a situation where all the tables are occupied; in this case the unmodified *full* or the adjectival phrase *absolutely full* would be used. Nor does (13) construe a situation with no visitors at all. Rather, what is meant in (12) is that the usual number of guests is considerably surpassed. Likewise, a plausible interpretation of (13) is that there were fewer visitors tonight than usual. Syrett et al. (2006, 2010) have shown that the use of a non-default reference point, as in (12) and (13), incurs an additional processing effort. Therefore, it takes subjects longer to compute a relative standard of comparison for, by default, absolute adjectives as compared to intrinsically relative terms such as *big* and *long*.

In the same vein, I claim that a norm is a default reference point for relative adjectives (cf. Rips and Turnbull 1980: 171). As shown in Section 2.1, orientation to a norm determines a number of crucial semantic properties of relative adjectives such as category dependence, contrariety and patterns of negation. This default can, however, be overridden in the actual communication process. Firstly, context can render another reference point more relevant than a norm, as shown in (10) and (11). Secondly, a different reference point can become primary through constructional constraints. For instance, a zero point is primary in constructions with measure phrases (e.g., *six feet tall*) and questions with *how* (e.g., *How tall is he?*), since in such cases the starting point for measurement is much more relevant than a norm.

This view accounts for the contradictory claims made in the traditional semantic literature *vs.* functional-cognitive studies. The traditional approach investigates properties of objects and often focuses on somewhat artificial uses of adjectives in zero contexts. No wonder that these studies found evidence of the default reference point. The functional-cognitive studies, on the other hand, deal with real uses of words in natural communication and focus on adjective meanings *in context*. In view of the fact that a default can be easily overridden by context, these studies have discovered a number of other reference points that are more salient than a norm in natural (and thus contextualized) language use.

3. The objectives of this study

Taking as a starting point the assumption that a norm is a default reference point for relative adjectives, I hypothesize that we should be able to find evidence of a norm in zero contexts and in constructions that are not associated with a different reference point. This hypothesis will be tested by means of two comprehension experiments. Experiment 1 targets two fundamental properties of a norm that are usually mentioned in the semantic literature — location around the midpoint of a series and category dependence. Experiment 2 aims to determine whether a norm is involved in the processing of adjectives in real time.

The research reported in this paper builds on earlier studies of adjective comprehension and contributes to the existing literature in several important ways. First, comprehension of relative adjectives was primarily investigated by psychologists and (less frequently) linguists studying language acquisition (e.g., Barner and Snedeker 2008; Bartlett 1976; Ebeling and Gelman 1988, 1994; Eilers et al. 1974; Nelson and Benedict 1974; Sera and Smith 1987; Smith et al. 1986, 1988; Syrett 2007; Syrett et al. 2006, 2010). Most of these studies used adult subjects as a control group. The study reported in this article will add to this line of research by focusing specifically on adjective interpretation by adult language users in order to test the long-standing semantic claims about the applicability of a norm in adjective comprehension.

Second, there is crosslinguistic evidence that dimensional adjectives reveal prototypicality effects, in the sense that language users associate adjectives with culturally determined best exemplars of a property (Dirven and Taylor 1988; Tribushinina 2008a, 2009a; Vogel 2004; Weydt and Schlieben-Lange 1998). For instance, towers are known to be prototypically tall, elephants stand out as prototypes of *big* and a mouse is known to be a characteristically small entity. Best exemplars can be traced by means of elicitation tests (e.g., *as small as* . . .) and corpus analyses. Furthermore, Tribushinina (2008b, 2011) found

that the acquisition of relative adjectives by children is strongly linked to the knowledge of best exemplars in the sense that by far the most frequent referents of relative adjectives in early child language and in the parental input are objects that are judged as prototypes of the property by adult language users. Given this recent evidence, it is important to establish whether prototypicality effects influence the assignment of norms in the process of adjective interpretation. Smith et al. (1986) embarked upon this path by comparing adults' and children's judgments of what is considered to be *high* or *low* for a bird (typically located high up in the sky) and a bunny (typically located low on the ground). Their results suggest that adults and older children (5-year-olds) shift reference points depending on a comparison class involved. More specifically, the reference point was located higher for a prototypically high object (bird) and lower for a characteristically low object (bunny). It is important to develop this line of research and study the effects of a greater variety of reference classes, including reference classes that are not associated with best exemplars (i.e., prototype-neutral entities). This is the main goal of Experiment 1.

Third, the experiments reported below will investigate the ability of adults to dynamically integrate more than one reference point in the process of adjective interpretation. Rips and Turnbull (1980) demonstrated that the interpretation of relative adjectives by adults involves combining two kinds of reference points — an average value for the superordinate category and an average value for everyday objects. Notice that both of these reference points are part of our world knowledge. I would like to go a step further and establish whether language users are able to compute the contextually relevant norm by integrating their world knowledge with other sources of information, such as a visually presented comparison class. This question will be pursued in Experiment 1.

Fourth, although several recent studies have used online experimental techniques for investigating the interpretation of relative adjectives in real time (Huang and Snedeker 2008; Rips and Turnbull 1980; Sedivy et al. 1999; Syrett et al. 2010), it is still not clear whether a norm is actually exploited in the online process of adjective interpretation. This question will be taken up in Experiment 2.

And last but not least, relevant studies in the past have largely focused on adjective comprehension in English. However, recent semantic research shows that relative adjectives display intriguing crosslinguistic variation (Koptjevskaja-Tamm and Rakhilina 2006; Tribushinina 2008a, 2008c, 2009b), which makes extension of this type of research to languages other than English particularly worthwhile. The experiments reported below target comprehension of two Dutch adjectives — *groot* 'big' and *klein* 'small'.

Before proceeding to the experiments, it should be stressed again that, for reasons of feasibility, this study only targets one part of the reference-point

model (Tribushinina 2008a) — the part positing the default status of a norm. The idea that a default reference point can be overridden by other reference points in actual contexts of use is beyond the scope of this investigation. It will be a matter for future research to subject the second part of the reference-point model to experimental scrutiny.

4. Experiments

4.1. *Experiment 1*

4.1.1. *Aim and hypotheses.* Recall that semanticists traditionally mention two properties of a norm — location around the midpoint of a series and category dependence. Experiment 1 tests psychological reality of these theoretical constructs on the basis of adult judgments about the application of the Dutch antonyms *groot* 'big' and *klein* 'small' to a series of same-kind objects incrementally increasing/decreasing in size. If a norm is located in the midzone of a scale as predicted by the semantic studies, the acceptance of an adjective should significantly decrease around the midpoint of an ordered set (cf. Barner and Snedeker 2008; Syrett et al. 2006). And, further, if a norm is category dependent, the position of the cutoff point between the ranges of the antonymous adjectives should vary depending on referent categories.

As explained above, it is expected that adult language users dynamically construe norms on the basis of their world knowledge and other sources of information, such as a visually given context. In order to trace this interaction, I will manipulate reference classes, but keep the sizes of the visually presented ranges constant. The pictures of objects from various reference classes (e.g., elephants, mice, balloons) will all be of exactly the same size range (from 1 to 7 cm). In addition, all objects in the test pictures will be smaller than in reality. Therefore, if subjects only make use of their world knowledge, but not of the visually provided context, they will judge all stimuli small and none of them big. Further, if subjects only use the contextually given comparison classes (visual range), but not their background knowledge of average object sizes in reality, there will be no significant differences between the ranges dubbed 'big' and 'small' across various referent categories. And, finally, if subjects are able to integrate two kinds of reference points (one from their background knowledge and one from the visually provided context), they will dub more objects 'small' than 'big'. Furthermore, the ranges of 'big' and 'small' should vary by object category. The "small" zone should be larger for the objects that are very big in reality (like elephants and hippos) than for objects that are relatively small (like mice and gnomes). And, conversely, the range dubbed 'big' ought to be larger for very small objects and smaller for very big objects.

Recall also that relative adjectives in general and size adjectives in particular reveal prototypicality effects, in the sense that some entities are seen as best exemplars of a property within a particular culture (Dirven and Taylor 1988; Tribushinina 2008a, 2009a; Vogel 2004; Weydt and Schlieben-Lange 1998). For instance, towers and giraffes are known to be prototypically tall, elephants are considered to be characteristically big, mice and bugs are small, etc. In this experiment, I will pursue the question whether knowledge of prototypes (best exemplars) influences adjective assignment to elements of a series (offline measure) and the ease with which scalar judgments are made (online measure). Offline, I will compare the number of objects dubbed 'big' and 'small' across three experimental categories — prototypically big objects, prototypically small objects and prototype-neutral objects.

In order to trace the ease with which scalar judgments are made across different adjectives and referent categories, I will analyze subjects' reaction times (RTs) during the scalar judgment process. Based on the findings from previous research, two predictions can be made. First, cognitive effort needed to process relative adjectives may vary by object category. Several psychological studies have shown that RTs increase when people have to make incongruent judgments, such as which of the two big objects (e.g., an elephant or a hippo) is smaller or which of the two small objects (e.g., a mouse or a moose) is larger. This is an instantiation of the so-called *semantic congruity effect* first described by Shipley et al. (1945) and later studied, among others, by Audley and Wallis (1964), Banks and Root (1979), Jamieson and Petrusic (1975), Holyoak (1978), and Ryalls and Smith (2000). With reference to these studies, it can be hypothesized that incongruent scalar judgments will be more effort demanding (and will therefore take longer to process) than congruent judgments.

Second, it is possible that the negative-pole adjective *klein* 'small' will take longer to process than its positive counterpart *groot* 'big'. Evidence of greater cognitive complexity of negative-pole terms was repeatedly found in adults (Clark 1971, 1972), children (e.g., Barner and Snedeker 2008; Sera and Smith 1987; Smith et al. 1988) and aphasics (Drummond et al. 1981). Therefore, it is plausible that the adult subjects in this experiment will need more time to make a decision when the target adjective is *klein* than when the target adjective is *groot*.

4.1.2. *Participants.* The subjects were twenty undergraduate students attending Utrecht University (12 female, 8 male; age range: 19−27). They completed the experiment for a course credit. Each subject signed an informed consent form.

4.1.3. *Materials, design and procedure.* Experiment 1 involved a Scalar Judgment Task (Smith et al. 1986, 1988; Syrett et al. 2006; Syrett 2007). The

Figure 1. *Example of a visual stimulus used in Experiment 1 (ascending trial)*

subjects saw pictures of seven same-kind objects on a computer screen and heard a question of the type *Welke X vind je groot/klein*? 'Which X do you find big/small', where X was the name of an object category in plural. On the descending trials, the test pictures incrementally decreased in vertical size from 7–1 cm at one centimeter intervals. On the ascending trials, the pictures increased in vertical size from 1–7 cm at one centimeter intervals (see Figure 1).

Three experimental categories were included in this study: prototypically big entities (elephants, hippos, houses, planes), prototypically small entities (babies, chickens, gnomes, mice) and prototype-neutral entities that are not particularly associated with either *groot* 'big' or *klein* 'small' (balloons, cakes, monkeys, umbrellas). The selection of test objects was made on the basis of the previous studies that established which objects are seen as prototypically big and small in the Dutch culture (Tribushinina 2008b, 2011). Prototype-neutral objects were selected on the basis of two criteria — 1. they are not unequivocally associated with either *groot* 'big' or *klein* 'small'; and 2. they are equally often dubbed *groot* and *klein* by adults in the Dutch corpora (Corpus of Spoken Dutch, Groningen Corpus, Van Kampen Corpus). It is important to notice that prototypicality is a matter of culturally determined construals rather than objective properties of objects. For instance, entities that are known to be best exemplars of smallness are not necessarily smaller than prototype-neutral entities. What matters is that certain objects are assigned the status of best exemplars within a particular language/culture. The effects of such best exemplars in language use and language development proved robust (Tribushinina 2008b, 2009a, 2011). This experiment aims to determine whether prototypicality effects qua best exemplars also affect adults' scalar judgments.

Each of the 12 object categories was presented in four types of trials: *groot*-descending, *groot*-ascending, *klein*-descending, *klein*-ascending (cf. Smith et al. 1986). This produced the total of 48 experimental trials.

The subjects were tested individually in a quiet room. The experiment started with two pre-test items. The subjects first saw a picture of eleven balloons of different colors and answered the question *Welke ballonnen vind je*

mooi? 'Which balloons do you find pretty?' After that, they saw a picture of six cars and answered the question *Welke auto's vind je lelijk?* 'Which cars do you find ugly?' During the pre-test phase, the participants were instructed to point to the objects on the screen if they thought an object could be assigned the corresponding property. They were also informed that there was no upper or lower bound: they were allowed to point to *all* the objects or to *none* of them, if they found appropriate. After the completion of the pre-test phase, the subjects were presented with the experimental trials, which were pseudorandomized with respect to two factors: the side of the relevant pole (left or right) and the adjective (*groot* or *klein*). Prerecorded audio stimuli (e.g., *Welke olifanten vind je klein?* 'Which elephants do you find small?') were automatically presented immediately after the corresponding visual stimulus appeared on the screen. To keep the scalar judgment process as natural as possible, the subjects were *not* instructed to make speeded judgments.

The experimental sessions were videotaped using a JVC Everio Camcorder and later analyzed using ELAN 3.8.1 software. Two analyses of the data were conducted. The first is an analysis of the ranges dubbed *groot* 'big' and *klein* 'small' across the three conditions — prototypically big entities, prototypically small entities and prototype-neutral entities (offline measure). The second is an analysis of RTs during the scalar judgment process (online measure).

4.1.4. *Results.* As far as the offline measures are concerned, there were no effects due to scale direction (*groot*: $p = .29$; *klein*: $p = .75$). Therefore, the data were collapsed across these two conditions. The mean ranges dubbed *groot* 'big' and *klein* 'small' are presented in Table 1.

The figures in the table show that, as predicted, the acceptance of the antonymous adjectives decreased around the midpoint of the scale. The subjects dubbed about a third of the range *groot* and about a third *klein*, which is consistent with the findings from previous research on English (Barner and Snedeker 2008; Smith et al. 1986; Syrett et al. 2006). In line with the claims made in the semantic literature, this result shows that a norm dividing a scale into the realms of (−Pol) and (+Pol) adjectives is indeed located in the midzone of a contextually relevant series.

Further, the ranges dubbed *groot* 'big' and *klein* 'small' were not equal. The zone labeled *klein* was significantly larger than the *groot* zone: $Z = 3.1$, $p < .005$ (Wilcoxon Signed-ranks test). This finding corroborates the hypothesis that

Table 1. *Mean number of objects dubbed* groot *and* klein

Adjective	Mean range	Std. deviation
groot 'big'	1.65	0.9
klein 'small'	2.6	1.3

subjects compute norms using, at least, two kinds of information — a contextually given comparison class (i.e., visually presented series) and their world knowledge of average object sizes in reality. All objects in the pictures were a lot smaller than in real life; therefore, the subjects more readily called them 'small' than 'big'. Notice that the observed difference cannot be merely a result of manipulating visual perceptions, because all sets of stimuli in the test pictures had exactly the same size range. If the subjects had only relied on the visually given sets, they would not have assigned different values to *groot* 'big' and *klein* 'small' the way they did. On the other hand, if the subjects had only relied on their world knowledge and had not used the contextually relevant information (i.e., visual range), they would not have labeled any of the stimuli *groot*, because all sets of stimuli were smaller than in reality. This leads us to conclude that the participants used two kinds of cues in their scalar judgments — visual context and background knowledge of what they experience as a typical entity size in reality.

In order to determine whether the subjects' scalar judgments were influenced by prototypicality qua best exemplars, we need to compare the ranges dubbed *groot* 'big' and *klein* 'small' across the three conditions — prototypically big entities, prototypically small entities and prototype-neutral entities. The mean numbers of objects dubbed *groot* and *klein* per category are presented in Figure 2.

As predicted, the values judged *groot* 'big' and *klein* 'small' were clearly category dependent. The *groot* zone was the biggest for objects from the *prototypically small* category and the smallest for objects from the *prototypically big* category: χ^2 (2) = 9.5, $p < .05$ (Friedman test). A Wilcoxon Signed-ranks test indicated that there were significant pair-wise differences between all three

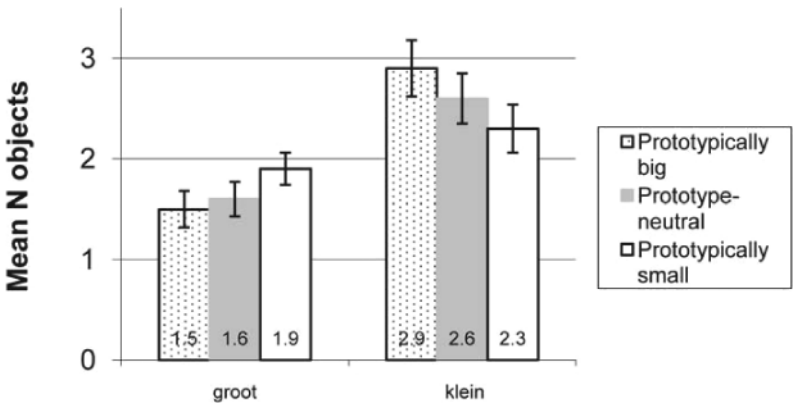
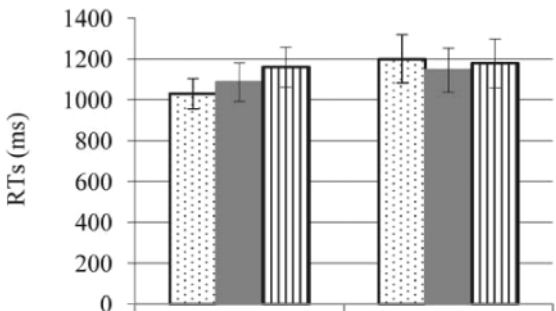


Figure 2. *Mean number of objects labeled* groot *and* klein *across the three conditions*

conditions: *prototypically big : prototype neutral* ($Z = 2.03$, $p < .05$), *prototypically small : prototype neutral* ($Z = 2.03$, $p < .05$), *prototypically big : prototypically small* ($Z = 2.65$, $p < .05$). And, conversely, the broadest *klein* ranges were found in the *prototypically big* category, whereas the least number of *klein* objects was assigned to the stimuli from the *prototypically small* category: χ^2 (2) = 14.4, $p < .005$ (Friedman test). And again, a Wilcoxon Signed-ranks test revealed that all pair-wise differences were significant: *prototypically big : prototype neutral* ($Z = 2.9$, $p < .005$), *prototypically small : prototype neutral* ($Z = 2.08$, $p < .05$), *prototypically big : prototypically small* ($Z = 3.08$, $p < .005$). As is evident from Figure 2, the ranges of *groot* and *klein* are mirror reflections of each other, which illustrates the inverse relations between antonymous adjectives. This result is consistent with the claim often made in the semantic literature that a reference point dividing the scale into the realms of (−Pol) and (+Pol) terms is determined by our knowledge of standard dimensions of a specific reference class. Furthermore, the subjects were sensitive to prototypicality effects qua best exemplars.

The next question is whether there were differences in processing effort across the three conditions. As explained in Section 4.1.1, it is plausible that incongruent situations (e.g., Which elephants do you find small?) take longer to process. It was also predicted that the negative-pole term *klein* 'small' might be more cognitively demanding than its antonym *groot* 'big'. These two questions will be considered in order.

Figure 3 shows mean RTs for the two adjectives. The measure targeted for analysis are RTs from adjective onset to the first pointing move.



	groot	klein
Prototypically big	1030	1200
Prototype-neutral	1086	1145
Prototypically small	1160	1179

Figure 3. *Mean reaction times (adjective onset to the first pointing move)*

Although reactions to prototype-incompatible questions (e.g., Which elephants do you find small? / Which mice do you find big?) took slightly longer than reactions to prototype-compatible questions (e.g., Which elephants do you find big? / Which mice do you find small?), the difference was only significant for *groot*: $Z = 2.7$, $p < .05$, but not for *klein*: $p = .12$ (Wicoxon Singed-ranks test). The observed difference between *groot* and *klein* is probably related to the fact that all stimuli were, in fact, smaller than in reality, which led to the clash of two reference points in the case of *groot*: the stimuli were big within a visually given range, but small for a real-life object of that kind. Overall, the results suggest that adult language users easily compute a contextually relevant norm for relative adjectives (cf. Syrett et al. 2010).

The second question addressed by measuring subjects' RTs was whether the negative-pole term *klein* 'small' was more difficult to process than the positive-pole adjective *groot* 'big'. The prediction that the processing of *klein* will cost more cognitive energy is not borne out by the data. Although RTs on the *klein* trials were somewhat longer than on the *groot* trials (1175 ms *vs.* 1092 ms), the difference was not significant: $p = .13$. We may, therefore, conclude that the processing of the negative-pole adjective *klein* was no more cognitively demanding than the processing of its positive counterpart *groot*. Holleman (1999) and Chessa and Holleman (2007) report similar results for the contrast pair *forbid* : *allow*.

The observation that might offer support for a greater conceptual complexity of negative-pole terms is that only in the case of *klein* there was a significant effect of scale direction. A pair-wise application of the Wilcoxon test shows that RTs on descending trials with the adjective *klein* are reliably longer than RTs on ascending trials with the same adjective ($Z = 2.1$, $p < .05$). It is possible that series of ascending format (from small to big) are easier to process because they are deeply entrenched in our cognition (e.g., counting, growing). Since no effect of scale direction was observed on the *groot* trials ($p = .29$), the difference in RTs on the *klein* trials may have been caused by a combined effect of a less natural scale direction and more complex semantics of *klein* (cf. H. Clark 1973: 38; Tribushinina 2008a: 291–293).

To summarize: as predicted by the prior semantic research, the cutoff point between the realms of the antonymous adjectives is indeed located around the midpoint of a scale. Furthermore, values assigned to the relative adjectives in this experiment were clearly category dependent. The subjects easily computed a relevant norm by combining two sources of information — an average size of a visually given series and background knowledge of normal object sizes in reality. Prototypicality qua best exemplars influenced both reference assignment and the ease with which scalar judgments were made. This experiment provided only indirect evidence of a greater cognitive complexity of *klein* 'small' over *groot* 'big'.

4.2. *Experiment 2*

4.2.1. *Aim and hypothesis.* The previous experiment demonstrated that the class-dependent norm in the midzone of a series is reflected in the scalar judgments of language users. Notice, however, that the procedure used in Experiment 1 does not inform us about the involvement of a norm in the actual processing of relative adjectives. It could be the case that a norm determines internal scale structure without being directly involved in the online interpretation of relative adjectives. In order to establish whether a norm plays a role in the processing of relative adjectives we need an online experimental technique that can keep track of the interpretation processes in real time. Therefore, in the present experiment I will record subjects' eye movements during the scale judgment process.

A central assumption in eye-tracking research is that eye movements immediately following a linguistic stimulus reflect the underlying interpretation processes (Tannenhaus et al. 1995). If a norm is indeed as important in the interpretation of relative adjectives as has been often suggested in the semantic literature, it should be actively involved in the process of assigning adjectival labels to the elements of a scale. It is plausible to assume that the subjects will repeatedly switch their gaze between the object under judgment and the midzone of a series in order to estimate the distance from the target item to the midzone. This hypothesis is based on the recent finding that relative adjectives can only be used if the degree of the property in an object *significantly* diverges from the contextually relevant reference point (Graff 2000; Kennedy 2007).

We can also hypothesize that the number of looks to the midzone will increase in the course of a trial, because objects located farther from the endpoints of the scale will raise more doubts as to the acceptance of the adjectives *groot* 'big' and *klein* 'small'. The subjects will presumably have little trouble with naming the biggest object *groot* and the smallest object *klein*. However, they are likely to have some doubts as to whether the second and the third objects can still be dubbed *groot* or whether they belong to the gray area of entities that are neither big nor small. This increasing hesitation is likely to cause an increase in the number of looks to the midzone.

To summarize, I make two predictions about the applicability of a norm:

(i) The subjects will be switching their gaze from the object under judgment to the midzone.
(ii) The looks to the midzone will become more frequent in the course of the trial.

4.2.2. *Participants.* Twelve subjects with normal or corrected to normal vision took part in this experiment (5 male, 7 female; age range: 23−26). All of them were MA students attending Utrecht University and participated in the

study for course credits. They were all monolingual speakers of Dutch. None had participated in the previous experiment.

4.2.3. *Materials and procedure.* The subjects first saw a blank screen and heard an instruction to point to big or small objects. As shown by a pilot study, this order of stimuli is crucial, because we would like to know how the subjects evaluate the scale from the moment they hear the target word and it is, therefore, important not to give them an opportunity to get familiarized with the presented series before they hear a relevant instruction.

Notice that only predicative adjectives had been used in the previous three experiments. One might argue that a norm proved relevant in these cases, because predicative uses are referent modifying, i.e., they ascribe new properties to already given referents. Attributive uses, in contrast, are reference modifying, i.e., an adjective takes part in referent identification from the set of entities denoted by the noun. In the latter case, there is more interaction between adjectival and nominal components (e.g., Bhat 1994: 49−54; Bolinger 1967; Ferris 1993: 39; Lewis 1976: 10−11; Taylor 1992: 7). In order to find out whether a norm is also relevant to attributive adjectives, two thirds of the trials in this experiment included adjectives used attributively. The target adjective was used in one of the following sentence frames: an imperative (e.g., *Wijs de kleine muisjes aan.* 'Point to the small mice-DIM.'), a *which*-question (e.g., *Welke olifanten zijn groot?* 'Which elephants are big?') or a *where*-question (e.g., *Waar zijn de grote huizen?* 'Where are the big houses?').

The visual stimuli were six series of objects incrementally increasing or decreasing in size. Three trials included the adjective *groot* 'big'; the test objects on these trials were houses, elephants and cakes. The other three trials included the adjective *klein* 'small' used with reference to dice, rabbits and mice. In order to reduce a strategy bias, scale direction was randomly manipulated: three trials contained objects incrementally decreasing in size; the other half of the trials involved ranges increasing in size. The visual stimuli were on the screen for 9 seconds.

Each participant was tested separately in an experimental cabin (160 cm wide, 250 cm deep, 210 cm high). The binocular Tobii 1750 eye tracker with remote cameras was used to record the eye movements. The subjects were seated on a medical chair in front of the 17″ TFT monitor (resolution 1280 × 1024 pixels) on which the test items were presented. The experiment started after calibration of the eye-tracker system and general instructions given by the investigator.

4.2.4. *Results.* For the purposes of the present study, the test ranges were divided into three areas of interest: the target zone (two biggest objects for *groot*-trials and two smallest objects for *klein*-trials), the competitor zone (two

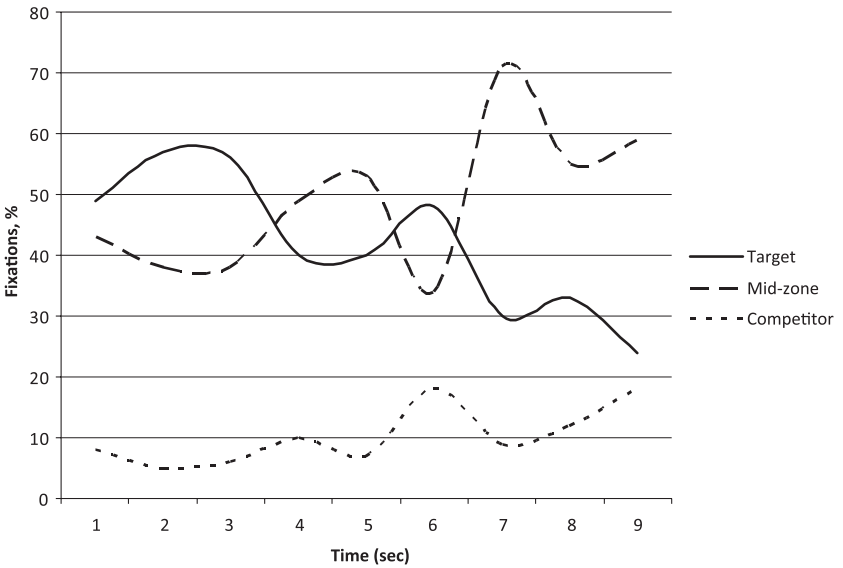


Figure 4. *Proportion of fixations in the areas of interest*

biggest objects for *klein*-trials and two smallest objects for *groot*-trials) and the midzone (three objects between the target and the competitor zones). This division was determined by the offline scalar judgments of the subjects in this experiment — the mean cutoff point for both *groot* and *klein* (averaged across all trials and subjects) was the third object (counted from the relevant endpoint).

Eye-gaze data were coded in terms of the number of fixations in each of the three areas of interest per second. The assumption is that the number of fixations in each area reflects the relative importance of the area in the interpretation process. The proportion of looks to the target, the competitor and the midzone averaged across all trials and subjects is shown in Figure 4.

As is evident from Figure 4 the subjects barely look at the competitor zone (small items for *groot* and big items for *klein*). The proportion of eye fixations in the competitor area does not exceed 10% in the first five seconds of the trials and is below 20% in the rest of the trial time. The mean number of fixations in the competitor region is significantly different from both the target and the midzone: $t1\ (11) = 7.45$, $p < .001$; $t2\ (11) = 5.32$, $p < .001$. This result is perfectly reasonable: the subjects barely look to small objects when asked to point to big ones and *vice versa*.

Perhaps the most remarkable finding is that the subjects looked to the midzone as often as they looked to the target items. The mean number of eye gazes in the target area is not significantly higher than the number of fixations in the

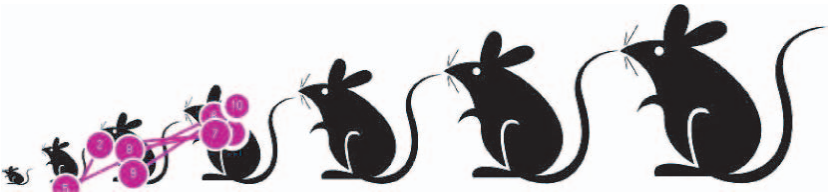


Figure 5. *Gaze plot of one subject (target:* klein *'small'*)

midzone: t (11) = .02, p = .49. Furthermore, the gaze patterns for the target and the midzone are in a complementary distribution, which means that the subjects switched their gaze from the target object to the midzone repeatedly throughout a trial (see Figure 4). This switching behavior may be taken as evidence of the reference-point status of a norm: the subjects presumably look at the midzone in order to estimate the distance from the reference point to the item under judgment, since the distance from the reference point is crucial to assigning adjectival labels to referents (Graff 2000; Kennedy 2007).

Figure 5 captures this switching behavior in one subject during one trial (the numbers indicate the order of fixations). The target adjective was *klein* 'small'. Therefore, the subject does not look at the big objects (competitor zone) at all. She immediately labels the smallest object *klein*. From the second object on, she starts switching her gaze from the object under judgment to the midpoint (the fourth mouse) and back. This switching behavior was attested in all subjects.

Furthermore, as is evidenced by Figure 4, the subjects increasingly looked to the midzone from the seventh second on. This result is consistent with the hypothesis that the participants would more often switch their gaze between the target zone and the midzone when they doubt whether the object can still be dubbed *groot* or *klein*. In order to establish whether there is a relation between time and proportion of fixations in the midzone, I defined three time intervals of 3 seconds each (Interval 1: 1−3 sec, Interval 2: 4−6 sec, Interval 3: 7−9 sec) and calculated the number of fixations in the target region and in the midzone per interval. A repeated measures ANOVA was performed with zone and interval as within-subject factors. There was a significant zone by interval interaction: F (1, 11) = 76.27, p < .05. Partial eta squared = .402. This result shows that the subjects indeed looked more to the midzone as they moved further up or down the scale.

The observed pattern might have been caused by two factors. One possible factor is the direction of scale evaluation. The subjects started evaluating the scale from the relevant endpoint (the biggest object for *groot* 'big' and the smallest object for *klein* 'small') and gradually moved to the opposite end of the range. As a result, the number of fixations in the midzone increased as the

subjects moved further down the scale. One might even argue that the subjects were performing pair-wise comparisons and increasingly looked to the midzone by sheer virtue of the fact that they needed to consider objects that lie ahead (I thank Kristen Syrett p.c., for this observation). However, this cannot be the only factor, since we do not observe a similar increase in the proportion of looks to the competitor area. Recall also that the midzone proved relevant from the very beginning of the scale judgment process, and not only towards the end of a trial. Therefore, the increasing attention to the midzone in the final stage of a trial is probably also related to the reference-point function of the norm during the judgment of less clear-cut cases located around the cutoff points. In order to disentangle the effects of these two factors, future research will have to use sets of more than seven objects.

To conclude, I hypothesized that the subjects would switch their gaze from the target to the midzone repeatedly throughout a trial, because a norm would anchor their scalar judgments. I also expected that the number of fixations in the midzone would grow in the course of a trial, as the subjects increasingly get confronted with less clear-cut cases. Both these predictions are borne out by the eye-tracking data in this experiment. This result may be taken to support the view that a norm in the midzone of a series is an important reference point that is involved in the online processing of relative adjectives. It is also important to observe that a norm proved relevant not only on predicative, but also on attributive trials. Therefore, the interpretation of attributive adjectives is also, by default, contingent upon a relative standard of comparison defined for a specific reference class.

5. Conclusion

This study has subjected the well-known assumption that relative adjectives are interpreted vis-à-vis a norm to experimental scrutiny. Two comprehension experiments targeted the relevance of a norm in the interpretation of the Dutch antonyms *groot* ‘big’ and *klein* ‘small’. Offline, I tested two major properties of a norm usually posited in the semantic literature — location in the midzone of a series and category dependence. Experiment 1 provided evidence that the cutoff point between positive- and negative-pole adjectives is indeed located in the midzone of a series and that its location is category dependent. The results indicate that the subjects took two references points into account — an average value of the visually given range and their background knowledge of normal object sizes in reality. Since all test items were smaller than in reality, the subjects more readily labeled them *klein* ‘small’ rather than *groot* ‘big’.

Furthermore, the subjects applied *groot* ‘big’ more often to objects known as best exemplars of smallness than to objects from a prototypically big category.

By contrast, *klein* 'small' was more frequently applied to prototypically big objects than to best exemplars of smallness. The objects from prototype-neutral categories took a position between the two extremes. Prototypicality effects also influenced the ease with which scalar judgments were made. It should be stressed that prototypes are rooted in cultural construals rather in objective parameters of the real world, since objects traditionally associated with, say, smallness, are not necessarily smaller than prototype-neutral entities. This study has shown that prototypicality qua best exemplars is important not only in language acquisition (cf. Tribushinina 2008b, 2011), but also in adjective comprehension by adults.

Experiment 2 focused on the applicability of a norm in the online interpretation process. The results suggest that language users actively exploit the reference point in the midzone of a series for assigning adjectival labels to objects. The subjects frequently switched their gaze from the target objects to the midzone when deciding whether an object can be dubbed *groot* 'big' or *klein* 'small'. This switching pattern became more frequent as the subjects turned to less clear-cut cases closer to the midpoint of the scale. A norm is, therefore, relevant not only to offline scalar judgments on scale structure, but also to online interpretation of relative adjectives.

Taken as a whole, these results support the experimental hypothesis that a norm is a default reference point used in the interpretation of relative adjectives when no further contextual or constructional constraints are involved.

Received 17 August 2009
Revised version received 5 July 2010

University of Antwerp
Utrecht University

Note

* The research reported in this paper was supported by the Netherlands Organisation for Scientific Research (NWO), grant 446-08-018. I would like to thank Kristen Syrett, two anonymous reviewers, and the responsible editor for valuable comments on an earlier version of this paper. Pim Mak, Linda de Leeuw and Nina Versteeg kindly helped with the implementation of the experiments. Correspondence address: Utrecht Institute of Linguistics (OTS), Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands. E-mail: e.tribushinina@uu.nl.

References

Apresjan, Jurij D. 1974. *Leksičeskaja semantika: sinonimičeskie sredstva jazyka* [Lexical semantics: Synonymic means in language]. Moskva: Nauka.

Arutjunova, Nina D. 1999. *Jazyk i mir čeloveka* [Language and human world]. Moskva: Yazyki russkoi kul'tury.

Audley, Robert J. & C. P. Wallis. 1964. Response instructions and the speed of relative judgments. I. Some experiments on brightness and discrimination. *British Journal of Psychology* 55. 59−73.
Banks, William P. & Maria Root. 1979. Semantic congruity effects in judgments of loudness. *Perception and Psychophysics* 26(2). 133−142.
Barner, David & Jesse Snedeker. 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.
Bartlett, Elsa J. 1976. Sizing things up: The acquisition of the meaning of dimensional adjectives. *Journal of Child Language* 3. 205−219.
Bierwisch, Manfred. 1967. Some semantic universals of German adjectivals. *Foundations of Language* 3. 1−36.
Bierwisch, Manfred. 1989. The semantics of gradation. In Manfred Bierwisch & Ewald Lang (eds.), *Dimensional adjectives. Grammatical structure and conceptual interpretation*, 71−261. Berlin, etc: Springer.
Bhat, Darbhe Narayana Shankara. 1994. *The adjectival category: Criteria for differentiation and identification*. Amsterdam & Philadelphia: John Benjamins.
Bolinger, Dwight. 1967. Adjectives in English: Attribution and predication. *Lingua* 18. 1−34.
Chessa, Antonio, C. & Bregje C. Holleman. 2007. Answering attitudinal questions: Modelling the response process underlying contrastive questions. *Applied Cognitive Psychology* 21. 203−225.
Clark, Eve V. 1973. What's in a word? On the child's acquisition of semantics in his first language. In Timothy E. Moore (ed.), *Cognitive development and the acquisition of language*, 65−110. New York & London: Academic Press.
Clark, Herbert H. 1971. More about "adjectives, comparatives, and syllogisms": A reply to Huttenlocher and Higgins. *Psychological Review* 78. 505−514.
Clark, Herbert H. 1972. Difficulties people have in answering the question "Where is it?". *Journal of Verbal Learning and Verbal Behavior* 11. 265−277.
Clark, Herbert H. 1973. Space, time, semantics, and the child. In Timothy E. Moore (ed.), *Cognitive development and the acquisition of language*, 27−63. New York & London: Academic Press.
Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
Cruse, D. Alan. 1976. Three classes of antonym in English. *Lingua* 38. 281−292.
Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
De Schutter, Georges. 1976. *Een semantisch-syntactische beschrijving van adjektieven in het Nederlands* (Antwerp Papers in Linguistics 7). Antwerp: University of Antwerp.
Dirven, René & John R. Taylor. 1988. The conceptualisation of vertical space in English: The case of *tall*. In Brygida Rudzka-Ostyn (ed.), *Topics in cognitive linguistics*, 379−402. Amsterdam & Philadelphia: John Benjamins.
Drummond, Sakina S., Tanya M. Gallagher & Russel H. Mills. 1981. Word-retrieval in aphasia: An investigation of semantic complexity. *Cortex* 17. 63−82.
Ebeling, Karen S. & Susan A. Gelman. 1988. Coordination of size standards by young children. *Child Development* 59. 888−896.
Ebeling, Karen S. & Susan A. Gelman. 1994. Children's use of context in interpreting "big" and "little". *Child Development* 65. 1178−1192.
Eilers, Rebecca E., D. Kimbrough Oller & Judy Ellington. 1974. The acquisition of word-meaning for dimensional adjectives: The long and short of it. *Journal of Child Language* 1. 195−204.
Ferris, D. Connor. 1993. *The meaning of syntax: A Study in the adjectives of English*. London & New York: Longman Publishing Group.
Frazier, Lyn, Charles Clifton, Jr. & Britta Stolterfoht. 2008. Scale structure: Processing minimum standard and maximum standard scalar adjectives. *Cognition* 106(1). 299−324.

Graff, Delia. 2000. Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics* 28(1). 45−81.
Holleman, Bregje C. 1999. The nature of the forbid/allow asymmetry: Two correlational studies. *Sociological Methods and Research* 28(2). 209−244.
Holleman, Bregje C. & Henk Pander Maat. 2009. The pragmatics of profiling: Framing effects in text interpretation and text production. *Journal of Pragmatics* 41(11). 2204−2221.
Holyoak, Keith J. 1978. Comparative judgments with numerical reference points. *Cognitive Psychology* 10. 203−243.
Huang, Yi Ting & Jesse Snedeker. 2008. Use of referential context in children's language processing. In Brad C. Love, Ken McRae & Vladimir M. Sloutsky (eds.), *Proceedings of the 30th annual meeting of the Cognitive Science Society*, 1381−1385. Austin, TX: Cognitive Science Society.
Hutchinson, Larry G. 1993. The logic of relative adjectives. In Mushira Eid & Gregory Iverson (eds.), *Principles and predication: The analysis of natural language*, 105−118. Amsterdam & Philadelphia: John Benjamins.
Jamieson, Donald G. & William Petrusic. 1975. Relational judgments with remembered stimuli. *Perception and Psychophysics* 18(6). 373–378.
Johnson, Mark. 1987. *The body in the mind. The bodily basis of meaning, imagination and reason.* Chicago & London: The University of Chicago Press.
Kamp, Hans. 1975. Two theories about adjectives. In Edward L. Keenan (ed.), *Formal semantics for natural languages*, 123–155. Cambridge: Cambridge University Press.
Kamp, Hans & Barbara Partee. 1995. Prototype theory and compositionality. *Cognition* 57. 129–191.
Katz, Jerrold J. 1972. *Semantic theory*. New York: Harper & Row Publishers.
Kennedy, Christopher. 1999. Gradable adjectives denote measure functions, not partial functions. *Studies in the Linguistic Sciences* 29(1). 65–80.
Kennedy, Christopher. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30. 1−45.
Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81. 345−381.
Klein, Ewan A. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4. 1−45.
Koptjevskaja-Tamm, Maria & Ekaterina Rakhilina. 2006. Some like it hot: On semantics of temperature adjectives in Russian and Swedish. *Sprachtypologie und Universalienforschung* 59(2). 253−269.
Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.
Lang, Ewald. 1989. The semantics of dimensional designation of spatial objects. In Manfred Bierwisch & Ewald Lang (eds.), *Dimensional adjectives. Grammatical structure and conceptual interpretation*, 263−417. Berlin, etc.: Springer.
Lehrer, Adrienne & Keith Lehrer. 1982. Antonymy. *Linguistics and Philosophy* 5. 483−501.
Leisi, Ernst. 1975. *Der Wortinhalt: Seine Struktur im Deutschen und Englischen.* Heidelberg: Quelle & Meyer.
Lewis, David. 1976. General semantics. In Barbara Partee (ed.), *Montague Grammar*, 1−50. New York: Academic Press.
Ludlow, Peter. 1989. Implicit comparison classes. *Linguistics and Philosophy* 12. 519−533.
Lyons, John. 1977. *Semantics.* Cambridge: Cambridge University Press.
Murphy, Gregory L. & Jane M. Andrew. 1993. The conceptual basis of antonymy and synonymy in adjectives. *Journal of Memory and Language* 32. 301−319.

Murphy, M. Lynne. 2003. *Semantic relations and the lexicon: Antonymy, synonymy, and other paradigms*. Cambridge: Cambridge University Press.
Murphy, M. Lynne. 2004. The development of size adjective meaning: What antonym use reveals. Paper presented at the Seminar on Child Language, Bristol, 12−14 July.
Nelson, Katherine & Helen Benedict. 1974. The comprehension of relative, absolute, and contrastive adjectives by young children. *Journal of Psycholinguistic Research* 3(4). 333−342.
Pander Maat, Henk. 2003. Reference point concepts and document quality. *Document Design* 4. 1−17.
Pander Maat, Henk. 2006. Subjectification in gradable adjectives. In Angeliki Athanasiadou, Costas Canakis & Bert Cornillie (eds.), *Subjectification: Various paths to subjectivity*, 279−322. Berlin & New York: Mouton de Gruyter.
Paradis, Carita. 1997. *Degree modifiers of adjectives in spoken British English* (Lund Studies in English 92). Lund: Lund University Press.
Paradis, Carita. 2005. Towards a theory of lexical meaning as ontologies and construals. *Axiomathes* 15. 541−573.
Paradis, Carita, Caroline Willners & M. Lynne Murphy. 2007. Dimensions of meaning and antonym canonicity. Paper presented at the First Conference of the Swedish Association for Language and Cognition, Lund, 29 November–1 December.
Paradis, Carita, Caroline Willners & Steven Jones. 2009. Good and bad opposites: Using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon* 4(3). 380−429.
Partee, Barbara H. 1995. Lexical semantics and compositionality. In Daniel Osherson (series ed.), Lila Gleitman & Mark Liberman (volume eds.), *Invitation to cognitive science. Part I: Language*, 311−360. Cambridge: MIT Press.
Partee, Barbara H. 2007. Compositionality and coercion in semantics: The dynamics of adjective meaning. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 145−161. Amsterdam: Royal Netherlands Academy of Arts and Sciences.
Rips, Lance J. & William Turnbull. 1980. How big is big? Relative and absolute properties in memory. *Cognition* 8. 145−174.
Rotstein, Carmen & Yoad Winter. 2004. Total adjectives *vs.* partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* 12. 259−288.
Ryalls, Brigette O. & Linda B. Smith. 2000. Adults' acquisition of novel dimension words: Creating a semantic congruity effect. *The Journal of General Psychology* 127(3). 279−326.
Šabes, Vladimir J. 1989. *Sobytie i tekst* [Event and text]. Moskva: Vyssšaja škola.
Sapir, Edward. 1944. Grading: A study in semantics. *Philosophy of Science* 11. 93−116.
Sera, Maria & Linda B. Smith. 1987. *Big* and *little*: "Nominal" and relative uses. *Cognitive Development* 2. 89−111.
Sedivy, Julie C., Michael K. Tannenhaus, Craig G. Chambers & Gregory N. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71. 109−147.
Shipley, Walter C., Judith I. Coffin & Kathryn Hadsell. 1945. Affective distance and other factors determining reaction time in judgments of color preference. *Journal of Experimental Psychology* 35. 206−215.
Siegel, Muffy E. A. 1980. *Capturing the adjective*. New York & London: Garland Publishing, Inc.
Smith, Linda B., Nancy J. Cooney & Carol McCord. 1986. What is "high"? The development of reference points for "high" and "low". *Child Development* 57. 583−602.
Smith, Linda B., Mary Jo Rattermann & Maria Sera. 1988. "Higher" and "lower": Comparative and categorical interpretations by children. *Cognitive Development* 3. 341−357.
Syrett, Kristen. 2007. *Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives*. Evanston, IL: Northwestern University dissertation.

Syrett, Kristen, Evan Bradley, Christopher Kennedy & Jeffrey Lidz. 2006. Shifting standards: Children's understanding of gradable adjectives. In Kamil Ud Deen, Jun Nomura, Barbara Schulz & Bonnie D. Schwartz (eds.), *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition — North America, Honolulu, HI,* 353−364. Cambridge, Mass: University of Connecticut Occasional Papers in Linguistics 4.

Syrett, Kristen, Christopher Kennedy & Jeffrey Lidz. 2010. Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics* 27(1). 1−35.

Tannenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 1632−1634.

Taylor, John R. 1992. Old problems: Adjectives in Cognitive Grammar. *Cognitive Linguistics* 3(1). 1−35.

Tribushinina, Elena. 2008a. *Cognitive reference points: Semantics beyond the prototypes in adjectives of space and colour* (LOT Dissertation Series 192). Utrecht: LOT.

Tribushinina E. 2008b. Do small elephants exist?: Prototypicality effects in the acquisition of dimensional adjectives. *Toegepaste Taalwetenschap in Artikelen* 79(1). 31−41.

Tribushinina, Elena. 2008c. EGO as a reference point: The case of *nevysokij* and *nizkij*. *Russian Linguistics* 32(3). 159−183.

Tribushinina, Elena. 2009a. On prototypicality of dimensional adjectives. In Jordan Zlatev, Mats Andrén, Marlene Johansson Falck & Carita Lundmark (eds.), *Studies in language and cognition*, 111−128. Newcastle upon Tyne: Cambridge Scholars Publishing.

Tribushinina, Elena. 2009b. The linguistics of zero: A cognitive reference point or a phantom? *Folia Linguistica* 43(2). 417−461.

Tribushinina, Elena. 2011. Spatial adjectives in Dutch child language: Towards a usage-based model of adjective acquisition. To appear in Carita Paradis, Jean Hudson & Ulf Magnusson (Eds.), *Conceptual spaces and the construal of spatial meaning: Empirical evidence from human communication*. Oxford: Oxford University Press.

Vendler, Zeno. 1968. *Adjectives and nominalizations*. The Hague & Paris: Mouton.

Verhagen, A. 2005. *Constructions of intersubjectivity: Discourse, syntax, and cognition*. Oxford: Oxford University Press.

Vogel, Anna. 2004. *Swedish dimensional adjectives*. Stockholm: Almqvist & Wiksell International.

Weydt, Harald and Brigitte Schlieben-Lange. 1998. The meaning of dimensional adjectives. Discovering the semantic process. *Lexicology* 4(2). 199−236.