Adjective Semantics, World Knowledge and Visual Context: Comprehension of Size Terms by 2- to 7-Year-Old Dutch-Speaking Children

Elena Tribushinina

Abstract   The interpretation of size terms involves constructing contextually-relevant reference points by combining visual cues with knowledge of typical object sizes. This study aims to establish at what age children learn to integrate these two sources of information in the interpretation process and tests comprehension of the Dutch adjectives *groot* 'big' and *klein* 'small' by 2- to 7-year-old children. The results demonstrate that there is a gradual increase in the ability to inhibit visual cues and to use world knowledge for interpreting size terms. 2- and 3-year-old children only used the extremes of the perceptual range as reference points. From age four onwards children, like adults, used a cut-off point in the mid-zone of a series. From age five on, children were able to integrate world knowledge and perceptual context. Although 7-year-olds could make subtle distinctions between sizes of various object classes, their performance on incongruent items was not yet adult-like.

Keywords   Size terms · Language acquisition · Reference points · Adjectives

Introduction

Processing and Representation of Spatial Adjectives: System of Reference Points

This paper studies children's developing ability to combine different sources of information for interpreting spatial adjectives such as *big* and *small*. Adult language users commonly employ a variety of cues for processing linguistic information. Semantic information from words and sentences is easily integrated with world knowledge and perceptual cues (Hagoort et al. 2004). Spatial expressions were shown to be a prime example of linguistic items whose interpretation hinges on a complex system of reference points dynamically construed by integrating lexical semantics with information supplied by our world knowledge and perceptually

E. Tribushinina (✉)
Department of Dutch Language and Culture, Utrecht Institute of Linguistics, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands
e-mail: e.tribushinina@uu.nl

given context (Bryant 1997; Clark 1973; Coventry et al. 1994, 2001; Coventry and Garrod 2004; Garrod and Sandford 1989; Holyoak 1978; Šetić and Dominjan 2007). For example, research on processing and representation of spatial prepositions demonstrates that adults rely not only on geometric cues (i.e. perceptual information), but also, and sometimes primarily, on extra-geometric knowledge of, for instance, canonical object functions and interaction schemes (Coventry et al. 1994, 2001).

In a similar vein, semantic studies of spatial adjectives (e.g. *big, small, high, low*), and relative adjectives in general, typically claim that their interpretation is contingent on a variety of context-specific reference points (Clark 1973; Holyoak 1978; Kennedy 2007; Paivio 1975; Rips and Turnbull 1980; Tribushinina 2008a). Most researchers of relative adjectives suggest that the primary reference point used for interpreting these words is a so-called norm, i.e. knowledge of average dimensions of a specific comparison class (Bierwisch 1967, 1989; Kamp 1975; Lang 1989; Lehrer and Lehrer 1982; Leisi 1975; Lyons 1977; Sapir 1944). A big elephant is big with respect to the class of elephants, and a big mosquito is big with respect to the class of mosquitoes. Consequently, an adjective such as *big* may denote quite different values depending on the comparison class being described.

However, a norm is not the only reference point that may be involved in the interpretation of spatial adjectives. Ebeling and Gelman (1994) show that a sentence such as *The hat is big* may be interpreted with respect to one of the three different reference points—a norm (bigger than an average hat), a perceptual standard (bigger than another hat in the visual range) or a functional standard of comparison (too big for a tiny doll). The experiments reported in Ebeling and Gelman (1988, 1994) and Gelman and Ebeling (1989) demonstrate that adults can easily choose one of these reference points and effortlessly switch between them.

However, it is not the case that people always have to choose one reference point for adjective interpretation. Very often, their spatial judgments would rely on a combination of reference points (for a detailed review, see Tribushinina 2008a). For instance, Rips and Turnbull (1980) demonstrate that adult language users easily combine two conceptual reference points—knowledge of average sizes of a specific compassion class (e.g. flowers) and knowledge of average sizes of various objects people daily deal with. Therefore, incongruent sentences such as *This poinsettia is tall* (where the referent is tall with respect to indoor plants, **but** short with respect to most objects in the human environment) take longer to process than congruent items such as *This spruce is tall* (where the referent is tall within the class of trees **and** tall with respect to everyday objects). Notice that both reference points involved in this case are part of our world knowledge.

It is, however, also common to use two different sources of information—conceptual knowledge and perceptual context—for assigning meanings to spatial adjectives. For instance, Clark (1973: 36) claims that the adjective *high* in the sentence *The balloon is high* is interpreted with respect to two different reference points at the same time—the visually perceived ground level (or another reference plane) and conceptual knowledge of how high a balloon can typically be located under particular circumstances, "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 10–20 feet high" (Clark 1973: 36). Therefore, the application of *high* in the above sentence is assumed to be contingent on a context-specific reference point that is provided by an interaction of the perceptual and conceptual information.

Clark's (1973) claim that the interpretation of spatial adjectives often involves interaction of perceptual and conceptual reference points has been recently supported by comprehension experiments reported in Tribushinina (2010). In these experiments, Dutch-speaking adults were shown series of pictures incrementally decreasing or increasing in size. The subjects were asked to indicate either big or small entities within each series. Object categories were

manipulated across trials (e.g. elephants vs. mosquitoes), whereas the perceptual sizes of the series remained constant across trials. A major result of interest was that the judgments were clearly category-dependent. The subjects called more elephants 'small' than 'big', whereas for mosquitoes, the pattern was reversed. This difference is clearly attributable to the fact that all elephants in the test pictures were smaller than in real life, whereas all mosquitoes were bigger than in reality. The observed pattern suggests that the adult subjects in these experiments interpreted the adjectives *groot* 'big' and *klein* 'small' by integrating two different reference points—an average size of the visually presented comparison class (the visual range) and average sizes of the conceptually represented comparison class (world knowledge of elephants/mosquitoes). If they had only used the visual context, they would have called an equal number of pictures 'big' and 'small', because all test pictures were of exactly the same size range. As against this, if they had only used their conceptual knowledge, they would not have judged any of the elephants 'big' and any of the mosquitoes 'small', because all of the test elephants were smaller and all of the mosquitoes bigger than in real life. The fact that the subjects chose several elephants and several mosquitoes as being 'big' and 'small', but the ranges labelled 'big' and 'small' were asymmetric, strongly suggests that their judgments were based on a context-specific reference point constructed by integrating the reference class in the visual context with their knowledge of object sizes in reality.

Tribushinina (2011) extended this work by including a greater variety of typically big and typically small categories, as well as middle-sized categories. In that study, all test pictures were smaller than in reality. As in the previous experiments, the adults applied the adjective *groot* 'big' more often to typically small entities (e.g. mice, gnomes), whereas the adjective *klein* 'small' was more often used for typically big entities (e.g. elephants, hippos). As expected, the middle-sized entities (e.g. umbrellas, cakes) occupied an intermediate position between typically small and typically big categories. It is also noteworthy that incongruent questions such as *Which gnomes do you find big?* incurred an additional processing effort resulting in longer reaction times.

To summarize, both traditional semantic studies and recent psycholinguistic experiments suggest that adults interpret spatial adjectives by integrating the available visual cues with their world knowledge of typical object sizes. Adults can easily integrate perceptual and conceptual information for assigning meanings to spatial adjectives, often without being aware of this mental operation. However, situations where these two sources conflict rather than converge (e.g. small for a conceptual reference point, **but** big within the visually given range) are more complex and, therefore, require more processing time (Paivio 1975; Šetić and Dominjan 2007). This paper goes a step further and investigates children's developing ability to integrate different sources of information for interpreting spatial adjectives using the scalar judgment procedure developed in Tribushinina (2011).

Before proceeding to the description of the experiment, let us briefly review some relevant studies investigating children's ability to use different reference points for interpreting spatial adjectives.

The Acquisition of Reference Points for Size Terms

Ebeling and Gelman (1988) showed that 2-year-old children are able to use two different standards of comparison for interpreting size terms. In their study, the children were first asked whether objects showed one at a time were *big* or *little*. Toddlers as young as age 2;6 performed above chance and their judgments were clearly category-dependent. For instance, they would dub a 10-cm big egg *big*, but a 10-cm big box of cereal *little*. In another session, the same object was presented next to a smaller or a bigger object of the same kind.

The subjects were able to shift their judgments when a perceptual reference point was introduced. For example, a 10-cm egg presented next to an even bigger egg was re-labelled *little* (cf. Syrett et al. 2010). Thus, by age three children are able to use either a conceptual or a perceptual reference point for adjective interpretation. It is, however, not clear whether they are also able to integrate the two sources of information into a more complex context-sensitive reference point the way adults do (Tribushinina 2010, 2011).

To the best of my knowledge, the only study that has addressed this issue so far is Smith et al. (1986). In a series of experiments, these authors investigated the ability of 3- to 5-year-old children to construe contextually relevant reference points for *high* and *low* on the basis of their knowledge of object classes and the extent of the visually given range. They found that children initially attach size terms to the extremes of a series and later extend categories to cover a broader range of values (cf. Berndt and Caramazza 1978; Clark 1970; Ehri 1976; Smith et al. 1988). More precisely, most 3-year-olds accepted only the highest object as *high* and only the lowest object as *low* in more than 50 % of cases. In contrast, 4- and 5-year-olds defined broader categories and tended to divide the whole range into the regions of *high* or *low*.

Smith and collaborators also report a growing ability to use contextual information in the interpretation of size adjectives. 3-year-olds only took one perceptual factor into account—extremes of the visually presented range; their judgments were not contingent on either conceptual knowledge of object categories or a perceptually given range of variation. 4-year-olds were capable of shifting the cut-off points for *high* and *low* depending on an object category, but in a non-target-like way. Birds, i.e. animals prototypically located high in the sky, were judged *high* more often than bunnies typically located low on the ground. Furthermore, 4-year-olds were not able to take the range of variation into account: their judgments about objects moving along a 4-feet backdrop did not significantly differ from the judgments made about objects on a 6-feet backdrop (but see Barner and Snedeker 2008). In contrast, 5-year-old children and adults took both the range of perceptual variation and the object category into account. They dubbed more birds *low* and more bunnies *high*, because what is high for a bunny is not necessarily high for a bird. Furthermore, more objects were labelled *high* or *low* when the backdrop was 6-feet high than when it was 4-feet high.

In summary, it appears that children start to integrate the visually present comparison-class information with their knowledge of reality from age 5 onwards. However, there are still a number of important open questions to be addressed.

Aims of this Study

The experiment reported below builds on the work of Smith et al. (1986, 1988) and extends it in several important ways. Firstly, in order to study the emerging ability to integrate perceptual and conceptual reference points, we need a sharper contrast between standards of comparison provided by the visual range and those supplied from world knowledge. Smith and colleagues worked with objects that moved along a 4- to 6-feet high backdrop. In that sense, the location of the target objects (birds and bunnies) was very similar, and sometimes identical, to their location in reality. This is especially evident for birds that are only sometimes located high in the sky: very often they would merely sit on the ground or in the bushes. And, similarly, some positions of the bunnies in the experiment were identical to their normal positions in reality. In this case, the information provided by the perceptual range significantly coincides with the information provided by world knowledge. Thus, it is not completely clear whether the children simply relied on one source of information

(visual range or world knowledge) or used a combination of the two. The answer to this question can only be obtained if we tease visual context and world knowledge apart. Therefore, in the experiment reported below, all objects in the test pictures will be smaller than in reality so that the subjects will be forced either to choose between the two reference points or to integrate them into a more complex reference point.

Secondly, in order to determine whether sensitivity of older children to prototypes (birds and bunnies) reported by Smith and colleagues can be generalized to other prototypes, we need to employ a greater variety of prototypical, as well as prototype-neutral stimuli. Previous semantic research shows that, across languages, dimensional adjectives reveal prototypicality effects in the sense that they are intrinsically associated with best exemplars, such as an elephant for *big*, a mouse for *small* and a tower for *tall* (Dirven and Taylor 1988; Tribushinina 2008a; Vogel 2004; Weydt and Schlieben-Lange 1998). Such best exemplars were shown to play an important role in the acquisition of size terms. Analyses of longitudinal transcripts of spontaneous speech of English- and Dutch-speaking children revealed that early in development dimensional adjectives are overwhelmingly used with reference to best exemplars by both children and their caregivers (Tribushinina in press). Later, the proportion of such prototypical referents decreases (Tribushinina 2008b). And in spoken adult corpora the proportion of adjective uses with reference to best exemplars is very low, due to redundancy of such modifications (Dirven and Taylor 1988; Tribushinina 2008a: Ch. 9). Given the important role of best exemplars in the acquisition of size terms, it is crucial to continue the line of research started by Smith et al. (1986) by testing the generalizability of the results to a larger set of prototypes and by comparing children's performance on prototypes and prototype-neutral categories.

The experiment reported in this paper will do just this by investigating the comprehension of the Dutch adjectives *groot* 'big' and *klein* 'small' by 2- to 7-year-old children. The reason to confine this research to only two size terms is that these are the least complex dimensional adjectives and the first ones acquired by Dutch children (Tribushinina in press). *Groot* 'big' and *klein* 'small' denote overall size and emerge in child speech around children's second birthdays, whereas other size terms denoting more specific dimensions (e.g. *hoog* 'high/tall', *diep* 'deep', *breed* 'wide') are more complex and acquired later (Bartlett 1975; Brewer and Stone 1975; Eilers et al. 1974; Tribushinina in press). The inclusion of more complex size terms would make the task too difficult for the youngest subjects in this study.

Hypotheses

Based on previous research reviewed in the introduction, the following predictions can be made.

Hypothesis 1: From Distinct Endpoints to a Common Midpoint

As in prior studies on English (Berndt and Caramazza 1978; Clark 1970; Ehri 1976; Smith et al. 1986, 1988), Dutch-speaking children are expected to start by attaching adjectives only to the extreme values of a scale and later discover a reference point around the mid-zone of a series. From age four onwards, children are likely to use a cut-off point between 'big' and 'small' located around the midpoint of a series, because at this age they typically learn to order series (Ehri 1976) and realize that antonymous adjectives are converse relations of one another sharing a common reference point in the middle of the scale (Smith et al. 1988).

Hypothesis 2: Integration of Perceptual and Conceptual Reference Points

If language learners are able to integrate perceptual information with their world knowledge of object classes from age five onwards (cf., Smith et al. 1986), this will be manifest in the number of objects labelled 'big' and 'small'. More precisely, children younger than age five should select an equal number of objects for 'small' and 'big'. From age five onwards, children, like adults in the previous experiments (Tribushinina 2010, 2011), will call more objects 'small' than 'big', because all test stimuli are smaller than in reality.

Hypothesis 3: Prototypicality of Object Categories

If children are able to use prototypicality information in an adult-like way from age five onwards (cf. Smith et al. 1986), they will call more objects from the prototypically big categories (e.g. elephants) *klein* 'small' and more objects from the typically small categories (e.g. mice) *groot* 'big'. The number of objects dubbed 'big' and 'small' in the prototype-neutral categories (e.g. umbrellas) should be between the two prototypical categories. At younger ages, language learners are expected to make category-independent judgments, operationalized as an equal number of objects labelled 'big' and 'small' across the experimental categories.

Method

Participants

Participants were 150 children (76 male, 74 female). All participants were native speakers of Netherlandic Dutch. The children were divided into six age groups (25 subjects per group)—2-year-olds ($M = 2;7$, range $= 1;10–2;11$), 3-year-olds ($M = 3;6$, range $= 3;1–3;11$), 4-year-olds ($M = 4;6$, range $= 4;0–4;11$), 5-year-olds ($M = 5;4$, range $= 5;0–5;11$), 6-year-olds ($M = 6;6$, range $= 6;0–6;11$) and 7-year-olds ($M = 7;5$, range $= 7;0–8;0$). The children came primarily from middle-class families and attended either a primary school or a day nursery in the Amsterdam area. The control group were 25 adults (10 male, 15 female) recruited in the Amsterdam and Utrecht areas. Informed consent was obtained from all adult participants and parents/guardians of the children.

Materials

The procedure developed by Tribushinina (2011) for research with adults was adopted in this study and adjusted for research with children. The test materials were 48 coloured computer-generated images of seven same-kind objects incrementally increasing or decreasing in size (see Fig. 1). The pictures were presented on a 15-inch computer screen.

Each picture was accompanied by a pre-recorded question uttered by a child-friendly female voice. The target question was always: *Welke X vind je groot/klein?* 'Which X do you find big/small?', where X was the name of the object category in plural. This formulation was chosen for two reasons. First, the construction with the verb *vinden* 'find' is the most common way to express subjective judgements in both adult and child Dutch. Second, pilot studies with adults showed that a more objective formulation of the question, such as *Welke X zijn groot/klein* 'Which X are big/small?', made some of the participants think that they were supposed to give consistent answers across all trials. Crucially, however, in our daily
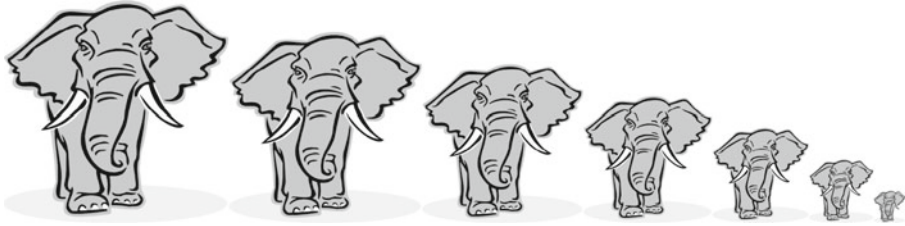
Fig. 1 Example of a test stimulus (typically big, descending)

practices we produce and interpret dimensional terms in a very subjective, situation-specific way. In order to underline that the subjects were expected to give subjective judgments of every situation, the more subjective sentence frame with the verb *vinden* 'find' was opted for.

The best exemplars of 'bigness' and 'smallness' used in this study were selected from the Dutch picture-books meant to teach the meaning of dimensional adjectives to toddlers. Two further selection criteria were applied. First, the nouns denoting the target categories had to be known and used by 2-year-old children. This was established on the basis of longitudinal transcripts of seven Dutch-speaking children from the Groningen Corpus (Bol 1995) available in the CHILDES archive (MacWhinney 2000). The familiarity of the nouns to the subjects was also tested in a pre-experimental trial (see below).

The second criterion was strength of association with bigness and smallness in child speech and child-directed speech in the Dutch CHILDES corpora (Tribushinina in press). For the prototypically big categories, more than 70 % of the size descriptions in child and caregivers' speech were to involve the adjective *groot* 'big'. For the typically small entities, more than 70 % of the size descriptions in the corpus had to contain the adjective *klein* 'small'. Prototype-neutral entities were selected from a range of categories described by *groot* and *klein* about equal number of times (50 %) in the longitudinal transcripts.

The categories obtained from these procedures were judged by 15 adult speakers of Dutch in a scaling task. The subjects, none of whom participated in the main experiment, were asked to indicate how big the entities were on a 10-point scale. Only the categories with a mean score above 7 were selected as typically big entities. Similarly, only the categories with a mean score below 3.5 were selected as best exemplars of smallness. The prototype-neutral categories had a mean score between 3.5 and 7. This procedure provided twelve test categories—four best exemplars of bigness (elephants, hippos, houses, planes), four best exemplars of smallness (mice, chickens, gnomes, babies) and four prototype-neutral categories that are not intrinsically associated with either of the properties (balloons, cakes, monkeys, umbrellas).

Test items from each category were presented in two different orders—ascending and descending. On the ascending trials, the stimuli increased in vertical size from 1 to 7 cm, at one centimetre intervals. On the descending trials, the vertical size of the entities decreased from 7 to 1 cm, at 1 cm intervals. Each picture was shown two times, once with *groot* 'big' and once with *klein* 'small' as a target adjective. This produced the total of 48 experimental trials.

Procedure

The children were tested individually in a quiet room at their nursery or school. The adults were tested either in a quiet room at the university or at home. The subjects were sitting in front of the computer and the experimenter was sitting next to the subject. Before the experiment started, each child received a pre-test and two training trials.

The pre-test was used to make sure that the children knew the object categories used in the main experiment. The subjects were shown pictures of several creatures including the ones used in the study, and asked to show the investigator each of the creatures by touching them on the screen (e.g. *Where is the elephant? Can you touch the elephant?*). In order to activate children's knowledge of size differences between the categories, the size of the pictures proportionally reflected the actual size differences between the real-life categories. For example, a mouse and a gnome in the pre-test pictures were much smaller than an elephant and a hippo. And the size of the prototype-neutral stimuli was between the two extremes. This activation was necessary because the test pictures of all categories in the main experiment were of the same size range.

For some of the categories (e.g. monkeys), more than one exemplar picture was included in a series. This was done in order to demonstrate that there always can be more than one target object in the pictures. For the same reason, all nouns denoting referent categories were always presented in the plural form. This was important because younger children are known to attach size terms only to the extreme points of a range (Smith et al. 1986, 1988).

The pre-test was followed by two training trials. The first training trial involved pictures of eleven balloons—five pink ones, four blue ones and two yellow ones. The subjects were asked to point to the balloons that they found pretty (*Welke ballonnen vind je mooi?* 'Which balloons do you find pretty?'). And in the second training trial the participants saw pictures of six different cars and were asked to show the experimenter the cars that they found ugly (*Welke auto's vind je lelijk?* 'Which cars do you find ugly?'). In this way, the participants received some practice in making subjective judgments about a range of objects.

The training trials were followed by 48 experimental trials pseudo-randomized with respect to two factors: the side of the target response (left or right) and the target adjective (*groot* or *klein*). Younger children—2-, 3- and 4-year-olds—took two short breaks during which they could choose a small present. Older children—5-, 6- and 7-year-olds—had one break halfway through the experiment. The adult subjects fulfilled the whole experiment without breaks. The sessions were videotaped using a JVC Everio Camcorder.

Two 2-year-old children failed the pre-test and were excluded from the study. One 2-year-old child and one 3-year-old child passed the pre-test, but failed to understand the main test. They always pointed at all seven objects on both 'big' and 'small' trials. The data from these children were not included in the analysis.

For each trial, the objects dubbed *groot* or *klein* were registered. For example, if the child selected the 1-, 2- and 3-cm tall objects as *klein* 'small', objects 1 through 3 were registered, as well as the total number of selected items (three, in this case). By age group, the proportion of subjects selecting each of the seven objects as being either big or small was obtained. The mean number of objects dubbed *groot* and *klein* was calculated for each of the experimental categories (typically big, typically small and prototype-neutral entities).

Results

In this section, I will first consider which objects children chose for *groot* 'big' and *klein* 'small' to test whether there is indeed a transition from using two distinct endpoints to a common reference point in the middle of the scale (Hypothesis 1). Then I will compare the mean number of objects dubbed *groot* 'big' and *klein* 'small' across all categories to see from what age children start taking two sources of information into account (Hypothesis 2). Finally, I will compare the mean number of objects labelled *groot* 'big' and *klein* 'small' in the three experimental conditions to establish whether children are sensitive to prototypicality

of objects in terms of size (Hypothesis 3). Paired-samples T tests showed that across all age groups there were no significant differences between ascending and descending trials (all p values above .05). Therefore, the data were collapsed across these two conditions.

From Distinct Endpoints to a Common Midpoint

The first hypothesis to be explored is whether Dutch-speaking children, like the English-speaking subjects in the previous experiments, start by attaching the size terms only to the endpoint of a scale and later come to use a cut-off point located around the midpoint of a series. In order to test this, the percentages of trials on which each of the seven objects was selected as either *groot* 'big' or *klein* 'small' to the total number of trials were calculated. Figure 2 presents the results for adults.

As in previous experimental studies (Smith et al. 1986; Syrett et al. 2010; Tribushinina 2011), adults placed a cut-off point between 'big' and 'small' around the midpoint of a series. Further, as predicted, the cut-off point was skewed towards the bigger pole, because all test objects were smaller than in reality. Now compare the adult data in Fig. 2 with the judgments obtained from children (Fig. 3).

For the adjective *groot* 'big', a 7×7 Mixed ANOVA with age (2-, 3-, 4-, 5-, 6-, 7-year-olds, adults) as a between-subjects factor and object size (1, 2, 3, 4, 5, 6, 7 cm) as a within-subjects factor conducted on the proportion of applications of the adjective *groot* 'big' indicated a significant effect of age ($F(1, 164) = 7.1, MSE = 1,306.9, p < .001, \eta_p^2 = .2$) and of object size ($F(6, 984) = 638.9, MSE = 350.9, p < .001, \eta_p^2 = .8$), as well as a significant object by age interaction ($F(36, 984) = 10.4, p < .001, \eta_p^2 = .28$). Likewise, for *klein* 'small', a 7×7 Mixed ANOVA with age (2-, 3-, 4-, 5-, 6-, 7-year-olds, adults) as a between-subjects factor and object size (1, 2, 3, 4, 5, 6, 7 cm) as a within-subjects factor revealed a significant effect of age ($F(1, 164) = 10.4, MSE = 1,532.1, p < .001, \eta_p^2 = .28$) and of object size ($F(6, 984) = 574.9, MSE = 377.7, p < .001, \eta_p^2 = .78$), as well as a significant object by age interaction ($F(36, 984) = 11.7, p < .001, \eta_p^2 = .3$).

The pattern of results in Fig. 3 demonstrates that 2- and 3-year-old children attach the size terms to the extremes of a range. As in previous research (Berndt and Caramazza 1978; Clark 1970; Ehri 1976; Smith et al. 1986, 1988), children younger than age four agreed to dub only the biggest object *groot* and only the smallest one *klein* in more than 50 % of the cases. The only significant difference between 2- and 3-year-olds was that the younger children were about 80 % correct on the extreme objects, whereas the older subjects were 94 % correct on the *groot* trials and 95 % correct on the *klein* trials. Two-year-olds, like adults and 7-year-old children, but unlike children aged 3–6, sometimes did not point to the biggest
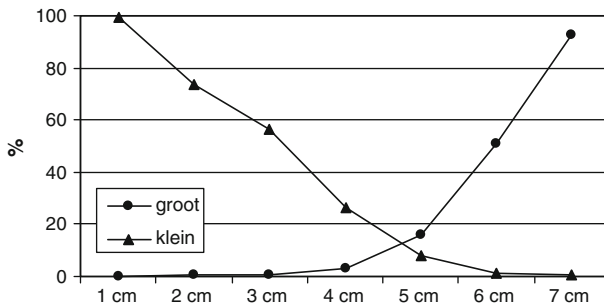

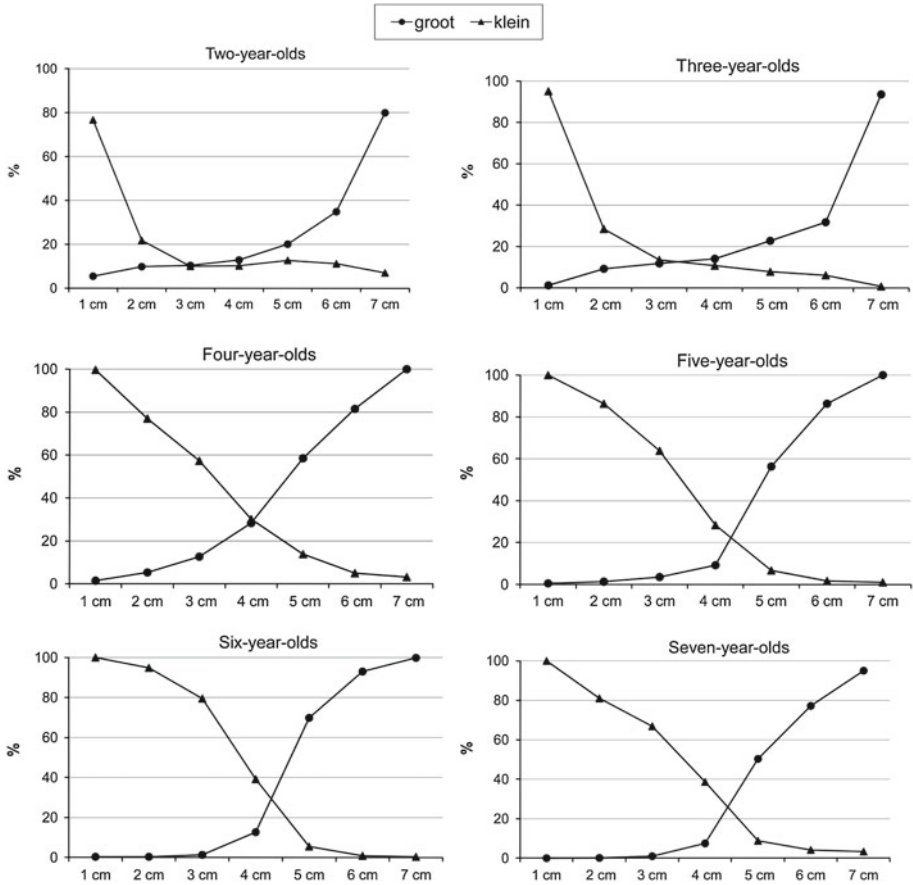**Fig. 2** Percentage of test pictures dubbed *groot* and *klein* (adults)

Fig. 3 Percentage of test pictures dubbed *groot* and *klein* (children)

object when the stimulus was *groot* 'big'. Even though the performance of the youngest group seems similar to that of the oldest groups, the motivation for this performance was not the same. The adult subjects and the oldest children were reluctant to call the test objects 'big' and sometimes refused to apply *groot* to any stimuli because they were all smaller than in real life. As against this, the youngest children sometimes failed to choose the biggest object as 'big' and the smallest as 'small' and pointed to a different object instead, either to a non-extreme object or to the opposite pole (the biggest object for 'small' and the smallest object for 'big'). This non-target-like performance of the youngest group is most likely due to immature knowledge of adjective semantics and to a relatively limited processing capacity.

Four-year-olds were 100 % correct with the extreme objects. Furthermore, they clearly used a cut-off point located precisely in the middle of the scale to divide the range into the realms of *groot* and *klein*. The fact that the ranges of *groot* and *klein* assigned by 4-year-old subjects were perfectly symmetric suggests that they only used a reference point constructed on the basis of the perceptual range and did not bring in their knowledge of typical sizes in reality. From age five onwards, we observe an increasingly asymmetric distribution of *groot* and *klein*, which might be evidence of integrating world knowledge and perceptual cues. This observation will be tested in the following subsection.

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

| | *groot* 'big' | | *klein* 'small' | |
	M	SD	M	SD
2-year-olds	1.64	1.5	1.5	1.35
3-year-olds	1.86	1.68	1.63	1.36
4-year-olds	2.87	1.4	2.91	1.51
5-year-olds	2.57	0.98	2.88	1.18
6-year-olds	2.77	0.82	3.2	0.97
7-year-olds	2.37	1.04	3.13	1.38
Adults	1.64	0.94	2.65	1.32

Integration of Perceptual and Conceptual Reference Points

As explained above, the ability to integrate perceptual and conceptual comparison classes was operationalized as a difference between the mean number of objects labelled *groot* 'big' and *klein* 'small'. More precisely, if children are able to integrate the two sources of information for interpreting spatial adjectives, they will call more objects 'small' than 'big', because all test objects were smaller than in real life. Table 1 presents the mean number of objects dubbed *groot* and *klein* by age group.

A 7 × 2 Mixed ANOVA with age group (2-; 3-; 4-; 5-; 6-; 7-year-olds; adults) as a between-subjects factor and target adjective (groot; klein) as a within-subjects factor revealed significant main effects of age ($F(6, 164) = 27.4$, $MSE = 21, 417.6$, $p < .001$, $\eta_p^2 = .5$) and adjective ($F(1, 164) = 32.02$, $MSE = 152.8$, $p < .001$, $\eta_p^2 = .163$), as well as a significant age by adjective interaction ($F(6, 164) = 8.2$, $p < .001$, $\eta_p^2 = .23$). Posthoc T tests with Bonferroni adjustment showed that the difference between the mean numbers of objects labelled *groot* and *klein* were not significantly different in the group of 2- ($p = .2$), 3- ($p = .6$) and 4-year-olds ($p = .8$). Significantly more objects were dubbed *klein* than *groot* by 5- ($t(24) = 2.7$, $p = .01$), 6- ($t(24) = 2.6$, $p = .02$) and 7-year-olds ($t(24) = 3.4$, $p = .002$), as well as by adults ($t(24) = 5.2$, $p < .001$).

A quadratic fit of the data from 2 to 7 years (see Fig. 4) shows that the mean number of objects selected as *klein* 'small' grows until 6 years with a plateau after that ($R^2 = 0.89$). By contrast, the mean number of objects dubbed *groot* 'big' grows until age 5 and decreases afterwards ($R^2 = 0.81$). This pattern is consonant with the hypothesis that children will be able to construct reference points by combining different sources of information from around age five. It is plausible that adults and older children constructed a novel reference point by combining the mid-point of the visual range with their knowledge of object sizes in reality. This integration motivated a shift from the exact midpoint (as attested in 4-year-olds) towards the bigger end of the scale. As a result, from age 5 onwards children are more reluctant to apply *groot* 'big' to objects that are smaller than in real life.

Prototypicality of Object Categories

Recall that if children are able to use object-class information in a target-like way, then the degree of reference-point shift towards the bigger pole in their judgments should be contingent upon the typical size of the specific categories. Put another way, the mean number of objects dubbed 'big' and 'small' should be category-dependent. The mean number of
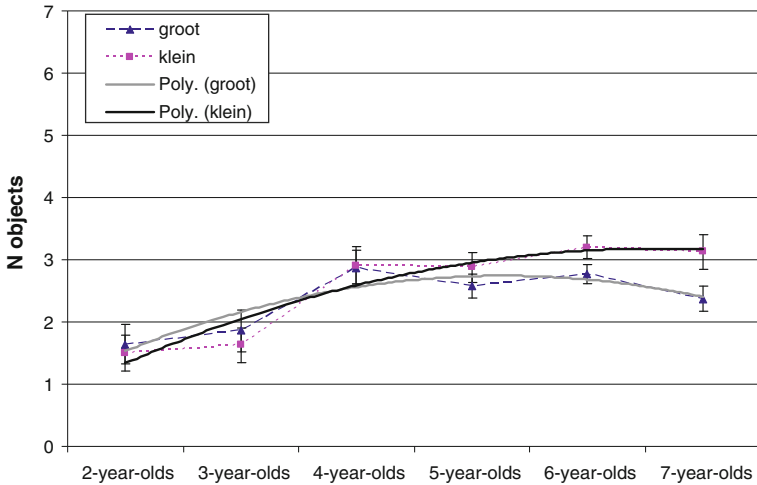
Fig. 4 Quadratic fit of the mean number of objects selected as *groot* 'big' and *klein* 'small'
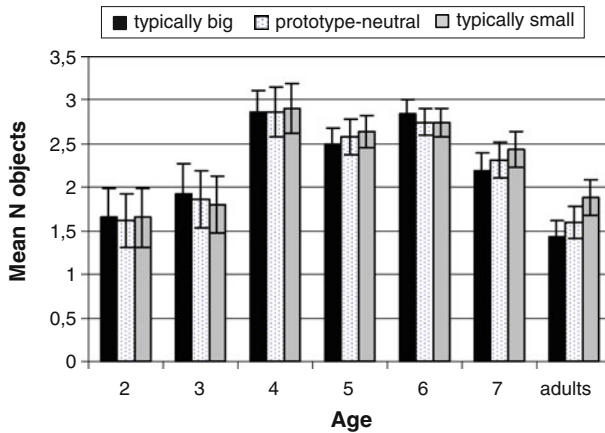

Fig. 5 Mean number of objects dubbed *groot* 'big' by category and age group

entities from the TYPICALLY BIG, TYPICALLY SMALL and PROTOTYPE-NEUTRAL categories labelled *groot* 'big' is presented in Fig. 5.

The application of *groot* by the adult subjects was clearly category-dependent, which replicates the results reported in Tribushinina (2011). The *groot* zone was the biggest for objects from the typically small category and the smallest for objects from the typically big category. The prototype-neutral entities occupied a position between the two extremes. A Repeated Measures ANOVA with object category (typically big; typically small; prototype-neutral) as a within-subjects factor shows that there are significant differences between the three experimental conditions: $F(2, 48) = 9.7$, $MSE = 8.02$, $p < .001$, $\eta_p^2 = .287$. Post-hoc Bonferroni pair-wise comparisons show significant differences between typically large and typically small entities ($p = .005$) and between prototype-neutral and large entities ($p = .02$). The difference between prototype-neutral and typically small entities was not significant ($p = .06$).
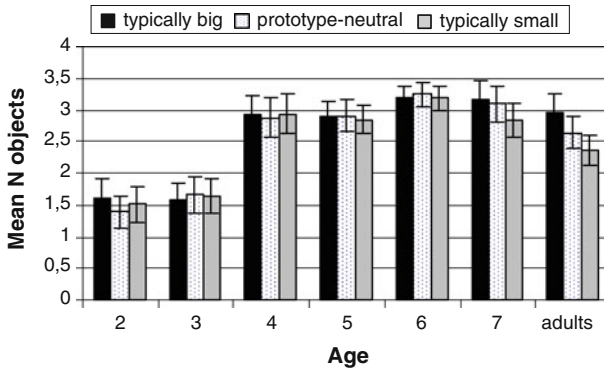
Fig. 6 Mean number of objects dubbed *klein* 'small' by category and age group

Unlike in adults, the judgments of 2- to 6-year-old children were not category-dependent (all *p* values above .05). Only at age seven, we observe an adult-like pattern. Seven-year-olds, like adults, were reluctant to call typically big objects *groot*. And, conversely, they tended to dub more entities from the typically small category *groot*. However, the differences between the three conditions were not significant: $F(2, 48) = 2$, $MSE = 7.1$, $p = .1$.

Figure 6 shows the mean number of objects labelled *klein* 'small' by condition and age group.

Yet again, the adult subjects applied the adjective *klein* in a category-dependent way. The broadest *klein* ranges were found in the typically big category, whereas the least number of *klein* objects was attested in the typically small category: $F(2, 48) = 9.7$, $p < .001$, $MSE = 14.9$, $\eta_p^2 = .287$. Posthoc Bonferroni pair-wise comparisons revealed significant differences between typically large and typically small entities ($p = .008$), as well as between prototype-neutral and typically large categories ($p = .008$). The difference between typically small and prototype-neutral entities was not significant ($p = .07$).

As in the case of *groot*, only the group of 7-year-old children showed an adult-like pattern with *klein*. The difference between the three conditions proved significant: $F(1, 48) = 4.1$, $MSE = 10.4$, $p = .02$, $\eta_p^2 = .161$. Posthoc Bonferroni comparisons revealed a significant difference between typically large and typically small entities ($p = .02$). The differences between prototype-neutral entities and two prototypical groups were not significant ($p = 1$ and $p = .2$).

To summarize, only the oldest child group in this study (7-year-olds) started making subtle distinctions between various categories when applying their world knowledge to the interpretation of the size terms *groot* 'big' and *klein* 'small'. The difference between the three conditions (typically big, typically small and prototype-neutral) was more pronounced for the adjective *klein* 'small'.

Conclusion and Discussion

Developing Ability to Construct Reference Points

This study targeted the developing ability to interpret size terms by integrating world knowledge with the visually provided cues. Based on previous research, three predictions were made. First, it was hypothesized that children would start with two distinct reference points

Table 2 Age differences in size term interpretation

	Primary reference point	Available sources of reference points	Integration
2-year-olds	Endpoints (best exemplars)	Perceptual range; world knowledge	No
3-year-olds	Endpoints (best exemplars)	Perceptual range; world knowledge	No
4-year-olds	Midpoint (average)	Perceptual range; world knowledge	No
5-year-olds	Midpoint (average)	Perceptual range; world knowledge	Yes
6-year-olds	Midpoint (average)	Perceptual range; world knowledge	Yes
7-year-olds	Midpoint (average)	Perceptual range; world knowledge	Yes
Adults	Midpoint (average)	Perceptual range; world knowledge	Yes

for 'big' and 'small' (endpoints of the visual range) and later discover a common reference point in the mid-zone of a scale. Second, children were expected to construct reference points by combining information provided by a visually present comparison class and a conceptually represented comparison class from age five onwards. Third, 5-year-old children were also expected to use object-class information in a target-like way by making different size judgments about typically big/small entities and for object classes not particularly associated with either of the size terms. An overview of the results, combined with insights from prior research (Ebeling and Gelman 1988, 1994; Gelman and Ebeling 1989; Smith et al. 1986, 1988) is presented in Table 2. The results for each of the hypotheses will be discussed in turn.

Hypothesis 1: From Distinct Endpoints to a Common Midpoint

As expected, 2- and 3-year-old children applied the adjectives *groot* 'big' and *klein* 'small' only to the endpoints of the scale in more than 50 % of cases. This result replicates earlier findings on children acquiring English (Berndt and Caramazza 1978; Clark 1970; Ehri 1976; Smith et al. 1986, 1988). Only at age four, children seem to discover the common reference point for 'big' and 'small' located around the midpoint of a series. Ehri (1976) claims that this transition is related to the developing ability to order objects by size. Children younger than age four are able to sort objects categorically by putting small and big things in two different groups, but they are not yet capable of ordering them. The ability to order objects develops between ages four and five and was shown to affect children's comprehension of size terms (Ehri 1976).

Before language learners have discovered the common reference point for 'big' and 'small', they do not understand the inverse relations between the antonymous terms and use comparatives immaturely by applying 'bigger' only to big objects and 'smaller' only to small objects, and not along the whole scale, as in adult language (Smith et al. 1988). Thus, the transition from two distinct reference points at the extreme poles of the scale to a common reference point in the middle of a series heralds an important transition to more mature adjective semantics.

Hypothesis 2: Integration of Perceptual and Conceptual Reference Points

Taking as a starting point the findings reported in Smith et al. (1986), it was hypothesized that the capacity to construct contextually-relevant reference points by combining different sources of information emerges around age five. In line with this prediction, children

in this study took two sources of information—world knowledge and the visual context—into account from age five onwards. Before that time, children labelled the same number of objects *groot* 'big' and *klein* 'small', even though all test items were smaller than in real life. From age five on, we observe asymmetric distributions of the antonymous terms. More precisely, the cut-off point is shifted from the midpoint of the scale towards the bigger pole. These results are most likely related to the fact that adults and older children were reluctant to call a lot of test pictures—that were all smaller than in real life—*groot* 'big'. This pattern is consistent with the claim that from age five onwards children are able to use multiple reference points for interpreting relative terms.

An alternative explanation of this pattern might be that children come to label more objects 'small' because they become bigger as they grow and the things in the world become smaller and smaller. However, there are two counterarguments to this explanation. First, one of the experiments reported in Smith et al. (1986) showed that the 3- to 5-year-old children did not define the denotations of *high* and *low* depending upon their own vertical position (either sitting on the floor or standing on a ladder). Based upon these results, the authors suggest that children have a primary external rather than egocentric definition of spatial adjectives.

Secondly, it is important that the response patterns of children aged five to seven are similar to those of adults. Recall also that the adult subjects in a prior experiment where test pictures were bigger than objects in real life (Tribushinina 2010) revealed an opposite pattern by calling more objects 'big' than 'small'. This finding confirms the idea that adults construe contextually relevant reference points from two different comparison classes—the visual range and knowledge of object classes in real life. It is very likely that 5-, 6- and 7-year-olds in the present experiment did the same and, therefore, called significantly more objects *klein* 'small' than *groot* 'big'. Notice, however, that the experiment reported in this paper only included test items that were smaller than in real life. In order to rule out the possibility that 5-year-olds come to call more objects 'small' because they grow themselves and their perspective on the world changes, future research may conduct a similar experiment with pictures of entities that are all bigger than in real life (e.g. mosquitoes, pins, lady-birds).

Hypothesis 3: Prototypicality of Object Categories

The idea that older children in the present experiment labelled more objects 'small' than 'big' because they started integrating reference classes provided by the visual knowledge with their background knowledge of object classes in real life is also supported by the finding that the oldest children in the experiment (7-year-olds) were able to make subtle distinctions between various object categories, which is a plausible further developmental step. The hypothesis that children would be able to distinguish between prototypically big, prototypically small and prototype-neutral entities from age five onwards was not supported by the experimental data. It is only from age seven that the children's scalar judgments were contingent on an object category. The greatest number of 'small' referents was assigned in the TYPICALLY BIG category and the smallest number in the TYPICALLY SMALL category. In the case of 'big', the distribution was reversed.

However, even 7-year-olds were not completely adult-like—they had more trouble with the incongruent (big) trials. As explained in the Introduction, incongruent trials incur an additional processing effort and result in longer reaction times even in adults (Tribushinina 2011). On these trials, the subjects were asked to indicate objects that were small for a conceptually available comparison class, **but** big within the visually given comparison class. As against this, on the small-trials both sources of information converged (small for a real-life class **and** small for a visual range). Consequently, it does not come as a surprise that children

had more trouble integrating the two sources of information when they conflicted (i.e. on the big-trials) than when they converged (i.e. on the small-trials). Therefore, the difference between the three experimental categories was more pronounced on the 'small' trials than on the 'big' trials.

Why is Reference-Point Integration Difficult?

The fact that younger children in this study were not able to integrate visual and conceptual reference points into a complex context-specific standard of comparison does not mean they cannot use perceptual and conceptual reference points. Prior research has repeatedly shown that toddlers as young as age 2;6 are capable of using both normative (i.e. conceptual) and perceptual standards of comparison, one at a time (Ebeling and Gelman 1988, 1994; Gelman and Ebeling 1989). However, it is not until age five that they learn to construe complex reference points by combining two sources of information—world knowledge and perceptual context (cf. Smith et al. 1986). The question is then why younger children fail to integrate reference points while they can use them one at a time.

It is plausible to assume that the ability to integrate different sources of information for adjective interpretation is facilitated by the development of executive functions, including working memory, inhibition and cognitive control. Executive functions are strongly associated with the prefrontal cortex, which is the last cortical region to reach full development (Kanemura et al. 2003). The prefrontal cortex is crucial for integrating multiple relations (Walz et al. 1999), for integrating word meaning and world knowledge (Hagoort et al. 2004) and for controlling interference from perceptual and semantic distracters during relational processing (Krawczyk et al. 2008). These skills play a key role in the process of interpreting spatial adjectives by construing contextually relevant reference points.

The core components of executive function show a developmental spurt between 3 and 6 years of age (Garon et al. 2008). In this period, children learn to integrate strategies (so-called 'rules') for solving problems. For example, when 3-year-olds are asked to sort things by, say, colour (If red, put in box A. If green, put in box B.), they tend to perform this task successfully irrespective of the dimension being used. However, when they are asked to play a new game with the same objects using a new pair of rules (If triangle, put in box A. If circle, put in box B.), most 3-year-olds fail to shift to the new rule (e.g. shape) and keep using the pre-switch rule (colour). Furthermore, Zelazo et al. (1996) found that 3-year-olds continue to use pre-switch rules even though they are perfectly aware of the new rule and can express knowledge of that rule both manually and verbally. By contrast, 4- and 5-year-olds can sort by one and by one of the two varying dimensions (rule switch condition). And from age five on children are able to sort by two dimensions simultaneously (Fischer and Roberts 1989; Frye et al. 1995). This means that from age five onwards children are able to integrate multiple rules for solving problems. This is exactly what is needed for interpreting relative adjectives vis-à-vis several reference points at the same time. In line with this developmental trend, the 5-year-old subjects in this experiment were able to take both conceptual and perceptual comparison classes into account. Children under age five are not yet able to integrate strategies (Frye et al. 1995; Zelazo et al. 1996; Zelazo and Reznik 1991). Rather they use one particular strategy at a time. In the present experiment, the subjects younger than age five made scalar judgments on the basis of the perceptual range alone.

According to the Cognitive Complexity and Control theory (Zelazo and Frye 1998), younger children cannot reflect on rules they know to integrate conflicting pieces of knowledge into a more complex, hierarchical rule system (Zelazo 2004; Zelazo et al. 1996: 57). And this is exactly what is needed to be able to construct complex reference points for relative

adjectives. Children know which elephants are big or small within the visually provided contexts. They also know that elephants are prototypically big entities. And they know that a picture elephant is smaller than a real elephant. However, children younger than age five are not able to integrate these pieces of knowledge into a more complex reference-point system.

Another important question is why children who are not yet able to integrate reference points choose for a perceptual comparison class rather than for a conceptual one, if both are available to them? This pattern is most likely related to the fact that visual context is perceptually salient and, therefore, more accessible and more difficult to inhibit than conceptual knowledge. The inhibition capacity shows a developmental spurt between 3 and 6 years of age (Bialystok and Senman 2004; Davidson et al. 2006). Therefore, on a scalar judgement task like the one used in this study, children younger than age five make scalar judgments on the basis of the perceptual range alone, rather than on the basis of world knowledge alone. In the same vein, Ebeling and Gelman (1994) report that both children and adults switch more easily from a normative to a perceptual interpretation of size terms than the other way around.

Although the results of the present study are consistent with the idea that the ability to integrate reference points is closely related to the development of executive functions (see also Sassoon 2011), this idea has not been subjected to experimental scrutiny yet. To explore this idea, future studies will need to correlate a child's cognitive control and inhibition capacities with her ability to interpret adjectives with respect to several reference classes at the same time. It is also possible that bilingual children who are used to inhibiting one of their languages would have an advantage over monolingual children or bilinguals speaking two languages in two different environments (e.g. one at school and the other at home) (cf. Costa et al. 2009). By contrast, elderly people with a decreased cognitive control and flexibility might not be able to integrate visual knowledge and world knowledge the way adult subjects do (cf. Zelazo et al. 2004). We are currently exploring these possibilities in our lab.

Generalizability of the Results

This study focused on comprehension of only two size terms—*groot* 'big' and *klein* 'small', which allowed to administer the same task with children of all age groups. As explained earlier in this paper, using more specific dimensional adjectives, such as 'long', 'high', or 'wide' would have made the task too difficult for the youngest age groups (Bartlett 1975; Brewer and Stone 1975; Eilers et al. 1974; Tribushinina in press). An important question to ask then is to what extent the results can be generalized to other dimensional adjectives and to relative adjectives in general.

Another important question concerns ecological validity of the experiment. Recall that the subjects had to take two kinds of comparison classes into account—the conceptual comparison class (knowledge of how big, say, hippos are) and the perceptual comparison class (the visually given range). Although people often use these two types of comparison classes for making dimensional judgments, the operationalization of the perceptual comparison class under laboratory conditions is different from what it would usually be in natural settings. Imagine that you see seven elephants in the zoo and that all of them are relatively small compared to a mental representation of an average-sized elephant (i.e. relative to the conceptual reference point). In this case, you can legitimately call the elephants 'small' by just relating their size to a stored image of a normal elephant. But if the size of the elephants in that visually given comparison class varies and you would like to make more specific judgments about the size of a particular elephant, which happens to be bigger than an average elephant in that zoo, you may need to fine-tune your spatial judgments by coordinating two reference points—an average height of the elephants at that particular place (perceptual reference point) and an

average size of a typical elephant stored in your long-term memory (conceptual reference point). It might be the case that a natural situation like this would provide more valuable cues and enable a child to coordinate the two types of reference points at a somewhat younger age. However, it is difficult, if not impossible, to conduct comprehension experiments of this kind in natural settings and to keep all variables constant across subjects.

Nevertheless, there are reasons to assume that younger children' inability to over-ride the perceptual cues and to choose an interpretation driven by more abstract, conceptual factors is an across-the-board phenomenon. Firstly, recall that the experiments reported in Smith et al. (1986, 1988) using objects (rather than pictures) moving in real space provided very similar results. It is only at age five that children in these experiments were able to take both the perceptual range and their conceptual knowledge of object classes into account for assigning meanings to the adjectives *high* and *low*. The fact that testing a different pair of adjectives in a different language using a different procedure provided largely comparable results suggests that the findings from the present study may be generalized to other dimensional adjectives. I leave this issue for future experimental scrutiny.

Secondly, the present results are also in line with the findings on the extension of nouns. Several experiments by Landau, Smith and Jones (Landau et al. 1996, 1998; Smith et al. 1996) showed that children younger than age five have trouble integrating conceptual knowledge and perceptual cues in object-naming tasks. Even if the relevant conceptual information (e.g. object function) is accessible, younger children still generalize on the basis of the visual cues (e.g. shape) without bringing in their conceptual knowledge of object functions (see also Gentner 1878; Tomikawa and Dodd 1980). By contrast, adults and older children generalize strongly on the basis of function. Thus, before age five children tend to rely primarily on the perceptually salient visual cues even if they have relevant conceptual knowledge. Age five appears to be an important milestone in the development of a child, because at this age children learn to take multiple perspectives on the same object and to combine strategies for solving problems (De Mulder 2011; Frye et al. 1995; Zelazo et al. 1996; Zelazo and Reznik 1991). One manifestation of this emerging capacity is the ability of 5-year-olds to integrate perceptual information with conceptual knowledge for making context-specific interpretations of spatial adjectives. The inhibition capacity also shows significant development between ages three and six (Davidson et al. 2006), allowing children to over-ride the perceptually salient cues and to involve more abstract, and therefore less accessible, conceptual knowledge.

Taken as a whole, the results show that semantic development of adjectives may have a protracted time course. Although size terms such as *big* and *small* are among the first adjectives acquired by children (Blackwell 2005; Nelson 1976; Ravid et al. 2010; Saylor 2000), their interpretation is not adult-like until, at least, age seven. In situations where dimensional adjectives are interpreted with respect to only one standard of comparison, 4-year-old children may make target-like interpretations. However, an adult-like interpretation of relative terms crucially hinges on the ability to dynamically construe context-specific reference points by combining their world knowledge with the information provided by the perceptual context. The development of this ability extends far beyond age three.

References

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*, 594–608.

Bartlett, E. J. (1975). Sizing things up: The acquisition of the meaning of dimensional adjectives. *Journal of Child Language, 3*, 205–219.

Berndt, R. S., & Caramazza, A. (1978). The development of vague modifiers in the language of pre-school children. *Journal of Child Language, 5*, 279–294.

Bialystok, E., & Senman, L. (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development, 75*, 562–579.

Bierwisch, M. (1967). Some semantic universals of German adjectivals. *Foundations of Language, 3*, 1–36.

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives. Grammatical structure and conceptual interpretation* (pp. 71–261). Berlin: Springer.

Blackwell, A. A. (2005). Acquiring the English adjective lexicon: Relationships with input properties and adjectival semantic typology. *Journal of Child Language, 32*, 535–562.

Bol, G. W. (1995). Implicational scaling in child language acquisition: The order of production of Dutch verb constructions. In M. Verrips & F. Wijnen (Eds.), *Papers from the Dutch–German colloquium on language acquisition* (pp. 1–13). Amsterdam: Institute for General Linguistics.

Brewer, W., & Stone, J. B. (1975). Acquisition of spatial antonym pairs. *Journal of Experimental Child Psychology, 19*, 299–307.

Bryant, D. J. (1997). Representing space in language and perception. *Mind and Language, 12*, 239–264.

Clark, H. H. (1970). The primitive nature of children's relational concepts. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 269–277). New York: Wiley.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York and London: Academic Press.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*, 135–149.

Coventry, K. R., Carmichael, R., & Garrod, S. C. (1994). Spatial prepositions, object-specific function, and task requirements. *Journal of Semantics, 11*, 289–309.

Coventry, K. R., & Garrod, S. C. (2004). *Saying, seeing and acting: The psychological semantics of spatial prepositions*. Hove: Psychology Press.

Coventry, K. R., Prat-Sala, M., & Richards, L. (2001). The interplay between geometry and function in the comprehension of over, under, above, and below. *Journal of Memory and Language, 44*, 376–398.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory. Inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.

De Mulder, H. L. N. M. (2011). *Putting the pieces together: The development of theory of mind and (mental) language*. Utrecht: LOT.

Dirven, R., & Taylor, J. R. (1988). The conceptualisation of vertical space in English: The case of *tall*. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 379–402). Amsterdam and Philadelphia: John Benjamins.

Ebeling, K. S., & Gelman, S. A. (1988). Coordination of size standards by young children. *Child Development, 59*, 888–896.

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

Ehri, L. C. (1976). Comprehension and production of adjectives and seriation. *Journal of Child Language, 3*, 369–384.

Eilers, R. E., Oller, D. K., & Ellington, J. (1974). The acquisition of word-meaning for dimensional adjectives: The long and short of it. *Journal of Child Language, 1*, 195–204.

Fischer, K., & Roberts, R. (1989). *The development of classification skills in the preschool years: Developmental level and errors*. Harvard: Harvard University.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60.

Garrod, S. C., & Sandford, A. J. (1989). Discourse models as interfaces between language and the spatial world. *Journal of Semantics, 6*, 147–160.

Gelman, S. A., & Ebeling, K. S. (1989). Children's use of nonegocentric standards in judgments of functional size. *Child Development, 60*, 920–932.

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development, 49*, 988–998.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438–441.

Holyoak, K. J. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology, 10*, 203–243.

Kamp, H. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal semantics for natural languages* (pp. 123–155). Cambridge: Cambridge University Press.

Kanemura, H., Aihara, M., Araki, T., & Nakazawa, S. (2003). Development of the prefrontal lobe in infants and children: A three-dimensional magnetic resonance volumetric study. *Brain and Development, 25*, 195–199.

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

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., Miller, B. L., & Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia, 46*, 2020–2032.

Landau, B., Smith, L. B., & Jones, S. (1996). Naming in young children: A dumb attentional mechanism. *Cognition, 60*, 143–171.

Landau, B., Smith, L. B., & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language, 38*, 1–27.

Lang, E. (1989). The semantics of dimensional designation of spatial objects. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives. Grammatical structure and conceptual interpretation* (pp. 263–417). Berlin: Springer.

Lehrer, A., & Lehrer, K. (1982). Antonymy. *Linguistics and Philosophy, 5*, 483–501.

Leisi, E. (1975). *Der Wortinhalt: Seine Struktur im Deutschen und Englischen*. Heidelberg: Quelle and Meyer.

Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum.

Nelson, K. (1976). Some attributes of adjectives used by young children. *Cognition, 4*, 1–31.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory and Cognition, 3*(6), 635–647.

Ravid, D., Tribushinina, E., Korecky-Kröll, K., Xanthos, A., Kilani-Schoch, M., Laaha, S., et al. (2010). *The first year of adjectives: A cross-linguistic study of the emergence of a category*. Poster presented at the Child Language Seminar, London, June 24–25, 2010.

Rips, L. J., & Turnbull, W. (1980). How big is big? Relative and absolute properties in memory. *Cognition, 8*, 145–174.

Sapir, E. (1944). Grading: A study in semantics. *Philosophy of Science, 11*, 93–116.

Sassoon, G. W. (2011). Adjectival versus nominal categorization processes: The rule versus similarity hypothesis. In G. Bochner, P. De Brabanter, M. Kissine, & D. Rossi (Eds.), *Cognitive and empirical pragmatics: Issues and perspectives* (pp. 104–147). Amsterdam: John Benjamins.

Saylor, M. M. (2000). Time-stability and adjective use by child and adult English speakers. *First Language, 20*, 91–120.

Smith, L. B., Cooney, N. J., & McCord, C. (1986). What is "high"? The development of reference points for "high" and "low". *Child Development, 57*, 583–602.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60*, 143–171.

Smith, L. B., Rattermann, M. J., & Sera, M. (1988). "Higher" and "lower": comparative and categorical interpretations by children. *Cognitive Development, 3*, 341–357.

Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics, 27*, 1–35.

Šetić, M., & Dominjan, D. (2007). The influence of spatial orientation on property verification. *Language and Cognitive Processes, 22*, 297–312.

Tomikawa, S. A., & Dodd, D. H. (1980). Early word meanings: Perceptually or functionally based? *Child Development, 51*, 1103–1109.

Tribushinina, E. (2008a). *Cognitive reference points: Semantics beyond the prototypes in adjectives of space and colour*. 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, E. (2010). The cognitive status of the norm. In G. I. Lushnikova, L. P. Prokhorova, S. V. Omelichkina, E. N. Ermolaeva, N. S. Sokolova, & E. J. Mokrushina (Eds.), *Concept and culture* (pp. 494–500). Kemerovo: Kemerovo University Press.

Tribushinina, E. (2011). Once again on norms and comparison classes. *Linguistics, 49*(3), 525–553.

Tribushinina, E. (in press). Spatial adjectives in Dutch child language: Towards a usage-based model of adjective acquisition. In C. Paradis, J. Hudson, & U. Magnusson (Eds.), *Conceptual spaces and the construal of spatial meaning: Empirical evidence from human communication*. Oxford: Oxford University Press (to appear).

Vogel, A. (2004). *Swedish dimensional adjectives*. Stockholm: Almqvist and Wiksell International.

Walz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes, Santos M., Thomas, C. R., & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119–125.

Weydt, H., & Schlieben-Lange, B. (1998). The meaning of dimensional adjectives. Discovering the semantic process. *Lexicology, 4*(2), 199–236.

Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences, 8*, 12–17.

Zelazo, P. D., Craik, F. I. M., & Booth, L. (2004). Executive function across the life span. *Acta Psychologica, 115*, 167–183.

Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: The development of exectuive function. *Current Directions in Psychological Science, 7*, 121–126.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.

Zelazo, P. D., & Reznik, J. S. (1991). Age-related asynchrony of knowledge and action. *Child Development, 62*, 719–735.