Alice Barlassina & Emanuela Sanfelici

2022

The Semantics of Gradable Adjectives in Italian-speaking Children and Adults

in

Robin Hörnig, Sophie von Wietersheim, Andreas Konietzko & Sam Featherston (eds.),

Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data

pp. 363–379

Tübingen: University of Tübingen

https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

In R. Hörnig, S. von Wietersheim, A. Konietzko & S. Featherston (eds.) (2022), *Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data*. Tübingen: University of Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

The Semantics of Gradable Adjectives in Italian-speaking Children and Adults

Alice Barlassina & Emanuela Sanfelici
University of Padua
alice.barlassina@gmail.com; emanuela.sanfelici@unipd.it

1 Introduction

This paper investigates the interpretation of subsective relative gradable adjectives, such as 'big', 'small', 'long', and 'short', by Italian children and adults. Whereas much acquisition research has demonstrated that from very early on children use syntactic cues to distinguish adjectives and nouns and make mappings between syntax and semantics, very little is known on how children combine the meaning of words to derive the meaning of phrases. Indeed, only few studies have explored the emergence of compositional semantics, without however reaching a consensus as to whether children use the same principles as adults to guide the interpretation of new word combinations (see e.g., Schmidt et al., 2009; Panzeri et al., 2013; Weicker & Schulz, 2019). For instance, whereas Hamburger & Crain (1984) have argued that children make mappings between syntax and semantics in an adult-like fashion, other works have concluded that children have a simpler non-compositional interpretation. Testing ordinal and color adjectives, Matthei (1982) found that children do not use compositional rules to interpret adjective-noun combinations, rather they interpret them incrementally, restricting the denotation of the expression as each word is encountered. Other studies have proposed that children might learn the meaning of a complex phrase directly via ostension and store it as a whole (Lieven et al., 2003). Since children's interpretation of *long table* is based on an association between that particular NP, i.e. table, and tables of a certain size, such children might easily point out *long tables* when asked, but they might be incapable of interpreting *long* with other nouns, if not previously paired with the adjective *long*. Accordingly, it remains open whether children use compositional rules and the same principles as adults to guide the interpretation of a complex phrase.

This paper investigates how children interpret adjective-noun combinations by restricting our domain to subsective relative gradable adjectives. We believe that these adjectives provide the ideal testing ground for our research question, since they are always interpreted relative to the noun which they modify (Partee, 1995). Subsective relative gradable adjectives relate an individual to a degree on a scale whose nature depends on the property expressed by the adjective (e.g., height, weight, etc.). An adjective like 'long' maps objects on a scale that represents degrees of length. Such a mapping results in an ordered array of objects, based on which we can establish which object is longer than the other (McNally & Weideman-Sassoon, 2018, a.o.). To decide whether something is 'long', the ability to build object orderings must be paired with the ability to perform statistical operations over a set of objects: we establish a standard of comparison by comparing the sizes of the object in the set. The standard is thus relative to the kind of set members under consideration, and it also changes depending on the size of the objects we are evaluating. At the same time, subsective relative gradable adjectives are vague, i.e. the evaluation of a relative gradable adjective admits a grey area (Kennedy, 1999, 2012; van Rooij, 2011, a.o.).

Other studies have tested whether Italian children can identify big or long items in a single novel array (e.g., Panzeri et al., 2013), but this study is the first one manipulating the distribution of object sizes (or length) across conditions. Such a manipulation allowed us to explore whether children take properties of whole sets into consideration when applying gradable adjectives. In this respect, our study is similar to experiment 1 in Barner & Snedeker (2008) on the acquisition of two English subsective relative gradable adjectives, 'tall' and 'short' and to Tribushinina (2013) on the acquisition of 'big' and 'small' by Dutch children. As for the vagueness property of subsective relative gradable adjectives, no studies we are aware of have investigated it in Italian children.

We therefore designed a card selection task, inspired by Barner & Snedeker's (2008) experiment 1. Sixteen children and 20 adults were tested on the comprehension of Italian novel nouns modified by four subsective relative gradable adjectives *lungo* 'long', *corto* 'short', *grande* 'big', *piccolo* 'small'. By testing children's understanding of complex adjective-noun combinations in which the noun was a novel expression, we assessed whether they compute standards for novel sets of objects, while ensuring that previous ostensive learning could not account for their behavior. Our findings reveal that children set a standard of comparison based on the perceptual class of depicted objects and are able to compute standards of comparison to evaluate whether a novel object can be considered 'long/big' or 'short/small'. In addition, we demonstrate that children apply the adjective to the depicted objects, which can be taken as evidence that they interpret the adjective-noun combinations compositionally. Finally, our results show that the vagueness property of gradable adjectives is instantiated in children's and adults' responses, which identify similar object(s) as borderline cases.

The paper is organized as follows. In Section 2 we revise the theoretical semantic properties of subsective gradable adjectives. Section 3 summarizes the main findings in the acquisition literature. Our experimental design is described in Section 4 and our results are outlined in Section 5. Finally, Section 6 discusses the findings and concludes the paper.

2 The Semantics of Relative Gradable Dimensional Adjectives

Gradable adjectives can be informally defined as predicative expressions whose domains can be partially ordered according to some property that allows for grading (e.g., 'height') (Kennedy 1999). Relative gradable adjectives are thus predicates mapping an individual to a degree on a scale whose nature depends on the property expressed by the adjective (e.g., height, weight, etc.) (Wheeler, 1972; Cresswell, 1976; Kennedy, 1999; Kennedy & McNally, 2005; Kennedy, 2007; Morzycki, 2016; but see Klein, 1980 for a non-degree analysis). Degrees are abstract representations of measurement standing as spatial or temporal distances, while a scale is a set of degrees ordered with respect to a specific dimension. Degrees can be overtly marked by comparative morphology (e.g. long-*er*, bigg-*er*) or by a measure phrase (e.g. Michael Jordan is *1.98 m* tall), or the positive form of an adjective can fulfill the degree requirement by combining with a covert positive morpheme. The latter provides the standard of comparison that is necessary to evaluate whether the adjective can be applied to a specific object. For relative gradable adjectives, the standard is given by a context-dependent comparison class which in attributive constructions is usually the modified noun. According to most accounts (Wheeler, 1972; Kennedy, 1999, 2012; Barner & Snedeker, 2008), gradable adjectives need to first set a standard of comparison, e.g., in the case of 'long', a standard of length. This standard of comparison needs to be computed on the basis of a comparison class (the relevant set of objects) in order to denote a property, i.e. to establish whether it is true that the object under evaluation can count as 'long'. For instance, to assign a truth value to an expression such as 'big mouse', one needs to set a standard of comparison which is appropriate to the members of the comparison class under evaluation, i.e. 'mice'. The standard of comparison set for the class of mice would necessarily be lower than the one set to assess the truth value of an expression such as 'big elephant', as in the latter case 'big' would mean 'big for an elephant'. Thus, the interpretation of these adjectives depends on contextual factors that might set different

standards of comparisons. The truth value of a sentence like "this paper is long" depends on the situation in which we evaluate the paper's length: if the paper is eight pages long, whether it can be defined as a long paper or not will depend on the standards of the journals we consider.

Another typical feature of relative gradable adjectives is vagueness (Kennedy, 1999, 2012; van Rooij, 2011, a.o.): the evaluation of a relative gradable adjective admits a grey area, i.e. borderline cases. Although the law of the excluded middle in classical logic does not allow for an entity to be both *P* and not *P* (or neither *P* nor not *P*) at the same time, we usually face this condition when dealing vague predicates like 'long' and 'big' (Russell, 1923; Wright, 1975; Castroviejo et al., 2018). Vagueness is what makes us unable or unwilling to determine the truth-value of some points along an ordered scale created by the meaning of a particular term. If we compare five tables whose size varies linearly from 1 meter to 5 meters and we have to assign to each table the property of being 'short' or 'long', it is usually the case that we are undecided as to whether, for instance, the third table measuring 3 meters should count as a long or a short table. Gradable adjectives represent an ideal empirical ground for the investigation of vagueness, as their basic forms introduce properties of objects that are true or false according to the position of the objects on a scale (Kennedy, 2012). According to Keefe (2000), vague relative gradable adjectives seem to also lack sharp boundaries:[1] they lack well-defined extensions such that we cannot easily draw a neat boundary separating the objects to which the predicate applies from those to which it does not apply (Williamson, 1994).

A final distinctive feature of relative gradable adjectives which must be taken into account when interpreting an adjective-noun combination is dimensionality. Kamp (1975) claims that an adjective is unidimensional if it is possible to associate with it a unique measurable aspect on which membership depends. Such aspect is, for example, 'temperature' for *hot* (van Rooij, 2011). Multidimensional adjectives like *big*, instead, order individuals along more than one dimension (Kamp, 1975; Klein, 1980). There is no unique way to determine whether an object is 'big', as it is not clear whether it is its volume, its height, its surface, or a combination of these which determines whether an object is big (Williamson, 1994).[2] Since more than one criterion is applied to order the individuals that have the property encoded by multidimensional adjectives, multidimensional adjectives may be considered semantically more complex than unidimensional ones (Wolfsdorf, 2019: 219).

After having illustrated the properties of relative gradable adjectives recognized in the theoretical literature, we now summarize the state of the art on the acquisition of these adjectives.

3 Previous Studies

Relative gradable adjectives have been quite extensively investigated from a developmental perspective since the Seventies (Clark, 1970; Sera & Smith, 1987; Ebeling & Gelman, 1988; Barner & Snedeker, 2008; Tribushinina & Gillis, 2012; Hohaus et al., 2014, a.o.). Two robust findings have been reported: (i) the acquisition of adjectives is marked by gradual achievements and (ii) children up to age 3 understand that adjectives apply to objects of different categories only if the linguistic or extralinguistic context is supportive. At the age of 2, children start producing relative adjectives and, by the age of 3, they productively use adjectives such as 'big'/'small', 'tall'/'short', 'long'/'short', 'high'/'low', 'heavy'/'light' (Clark, 1970; Clark, 1972).

[1] Vague predicates also give rise to sorites paradoxes (Williamson, 1994; Fara, 2000; see also van Rooij, 2011 for a detailed discussion on possible solutions of these paradoxes).

[2] Various questions regarding multidimensionality are still open: it is not clear the way adjectival semantics is affected by each scalar dimension and the way a set of contextually relevant dimensions may be construed (see Castroviejo et al., 2018). In addition, it remains open how information regarding different dimensions in such a set may be incorporated within context to create a single uniform interpretation for a given multidimensional adjective (Lasersohn, 2005; Sassoon, 2013; Bylinina, 2016; McNally and Stojanovic, 2017; Solt, 2018).

As already clarified, the interpretation of these adjectives necessarily depends on the noun with which they occur. To yield a meaningful interpretation, the concepts of the adjective and of the noun must be combined to generate a composed meaning (Barner & Snedeker, 2008). In order to obtain a composed meaning, children must be aware of the existence of a standard of comparison and must be able to determine its value on the basis of the noun. Accordingly, children and adults need to evaluate whether a given adjective can apply to an object by comparing its size or length to the standard of comparison (Barner & Snedeker, 2008). Various studies have shown that children between 2 and 4 years old have knowledge of the typical sizes of things such as buttons, mittens, plates, and shoes, and can therefore judge whether these objects can count as big for the set to which they belong (Ebeling & Gelman, 1988; Sera et al., 1988). In addition, Ebeling & Gelman (1988) showed that children as young as 2 can differentiate between normative and perceptual judgements of size. For instance, they classify a depicted mitten as 'big' or 'small' for being a mitten and they make use of perceptual cues as well, judging a mitten as 'big' or 'small' compared to another physically present mitten. Furthermore, children know that what counts as 'big' or 'small' also changes depending on the size of the entity with which the object is compared: a shot glass is small for a person, but big for a doll (Gelman & Ebeling, 1989; Ebeling & Gelman, 1994; Carey, 1978).

At the same time, though, children are reported to make a consistent series of errors. In the first place, they seem reluctant to switch the contextual standard of comparison. In addition, different studies also showed that children make substitution errors (Clark, 1972, 1973; Sera & Smith, 1987). For example, they interpret the negative antonym of a pair as if it were the positive (e.g., they use big for little or wide for narrow) and use more general terms instead of specific ones (e.g., they use big instead of long or tall). Children exhibit extreme labeling: when they are asked to judge a series of objects decreasing along a relevant dimension, younger children apply the relative term only to the extremes of that series. For example, they consider only the first object as tall and only the last one as short (Smith et al., 1986; Smith et al., 1988; Syrett, 2007; Panzeri et al., 2013).

Barner & Snedeker (2008) investigated children's knowledge of the compositional semantics of adjective-noun combinations and their ability to compute standards of comparison by means of statistical operations over sets. They made use of sets of novel objects with novel names (so as to ensure the absence of previous ostension and test whether children can compute standard values for novel objects), whose sizes shifted across conditions. They found that children shifted their judgements on the value of the standard accordingly, and that when some of the objects differed from the others both in name and physical features, these did not make children shift their judgements, because such objects were considered as a different set. However, when the differing objects had the same name of the other objects, children considered them as part of the same set, and shifted their judgement on the standard value accordingly: they restricted the comparison class for *tall* by relying on linguistic kind information. Hence, Barner & Snedeker (2008) showed that English-speaking children seem to be able to deploy a compositional semantics for gradable adjective-novel noun combinations that is sensitive to shifts in the statistics of sets, and partly mediated by linguistic labels. Their study showed that 4-year-old children are able to derive composed meaning and can integrate subtle changes in the statistical properties of object arrays to shift their standard. Hohaus et al. (2014), however, showed that mastery of the full semantics of relative gradable adjectives, including comparison constructions, seems to be achieved much later than age 4.

From this brief overview, a variegated and in some respect contradictory picture emerges. Whereas some studies found that children already at age 2 master the semantics of gradable adjectives, others demonstrated that certain properties, such as setting or switching a standard of comparison, is a later development. No studies we are aware of have focused on vagueness or the scale of degree used by children. In addition, we know of no previous study that has tested what defines a comparison class for Italian children (e.g., the class of things considered when evaluating whether something is long). For adults, comparison classes are defined at least

in part by nouns. For instance, a 'long table' is long relative to all things named by the noun table. No previous study on Italian has investigated whether applying the same name to an array of novel and physically distinct objects is sufficient for grouping these objects in a single comparison class for adjective interpretation.

4 Experiment

In order to investigate how Italian children and adults interpret subsective gradable adjectives in novel word combinations, we designed a card selection experiment, which was inspired by the task used in Barner & Snedeker's (2008) experiment 1. Children were tested on the interpretation of two dichotomic pairs of subsective gradable adjectives, i.e. *lungo*/*corto* 'long/short' and *grande*/*piccolo* 'big/small'. Four research questions were formulated: (Q1) What do children consider a single comparison class for adjective interpretation?, (Q2) Do children set a standard of comparison? If so is it different from that of adults?, (Q3) Does (multi)dimensionality have an effect in adjective interpretation?, and (Q4) Are gradable adjectives vague for children as well as they are for adults?

4.1 Participants

Participants were 16 monolingual Italian-speaking children between 38 months and 71 months (Mean age: 4;6; 9 girls, 7 boys): five children were three-year-olds (Mean age 3;6); four were four-year-olds (Mean age 4;3), and seven children were five-year-olds (Mean age 5;7).

Children were recruited and tested in the kindergarten "Maria SS. Consolatrice" in Friola, municipality of Pozzoleone (VI), Italy. All parents gave written consent for their children's participation in this study. A parental questionnaire ensured that none of the child participants had signs of language impairment, language delay, or hearing problems. A control group of 20 adults, divided into four age groups (20-25; 25-30; 30-40; 40-60; M: 33;6; 12 women, 8 men), was also administered the same test. All of the subjects gave written consent for their participation in this study. A questionnaire ensured that none of them had language impairment problems or any other problem which would invalidate their performance for the present study.

4.2 Procedures and Stimuli

We tested children's compositional knowledge of noun-subsective gradable adjective combinations by using nine objects, which only differed with respect to their sizes.

The objects referred to non-existing objects and creatures and were assigned invented names obeying the phonotactic and morphological constraints of Italian. Using non-existing nouns and objects assured the absence of semantic influences, previous ostensive learning, or world knowledge. The invented object-names were eight: four object-names per each pair of adjectives. In the *lungo/corto* 'long/short' adjectival pair, the objects were called *cruffi*, *bruldi*, *divuchi*, *loniterpi*, while in the *grande/piccolo* 'big/small' pair they were called *mavorbi*, *catabombi*, *niodòli*, *falafurfe.*

For each test item we drew nine objects differing in size. The objects' sizes were kept constant across all sets of items and varied gradually from object 1 (the biggest/longest/with longest x, where x stands for an object) to object 9 (the shortest/smallest/with shortest x). The size measures adopted for each object are listed below. They correspond to objects' height in the *big/small* cases and to their overall length in the *long/short* cases: obj. 1 = 14.7 cm; obj. 2 = 14.3 cm; obj. 3 = 14 cm; obj. 4 = 12.5 cm; obj. 5 = 11.5 cm; obj. 6 = 10 cm; obj. 7 = 8.2 cm; obj. 8 = 6.5 cm; obj. 9 = 4.8 cm.

Each object was printed on a card and the cards were arranged pseudo-randomly on a sheet of white cardboard of 100 x 70 cm. Thus, differently from Barner & Snedeker's experiment 1, the objects were not arranged in a row but spread across the sheet. In our design the child had to look at all the objects in the width of the sheet's space and build a mental order by comparing

them and calculating their sizes in relation to each other. This aspect was thus made more challenging than in Barner & Snedeker's experiment, as the relations between objects' sizes are even less transparent. The cards were attached to the cardboard with hook and loop fasteners. The order in which the cards were arranged was kept for all participants and across conditions for all the test items. Figure 1 provides an example of a test item.

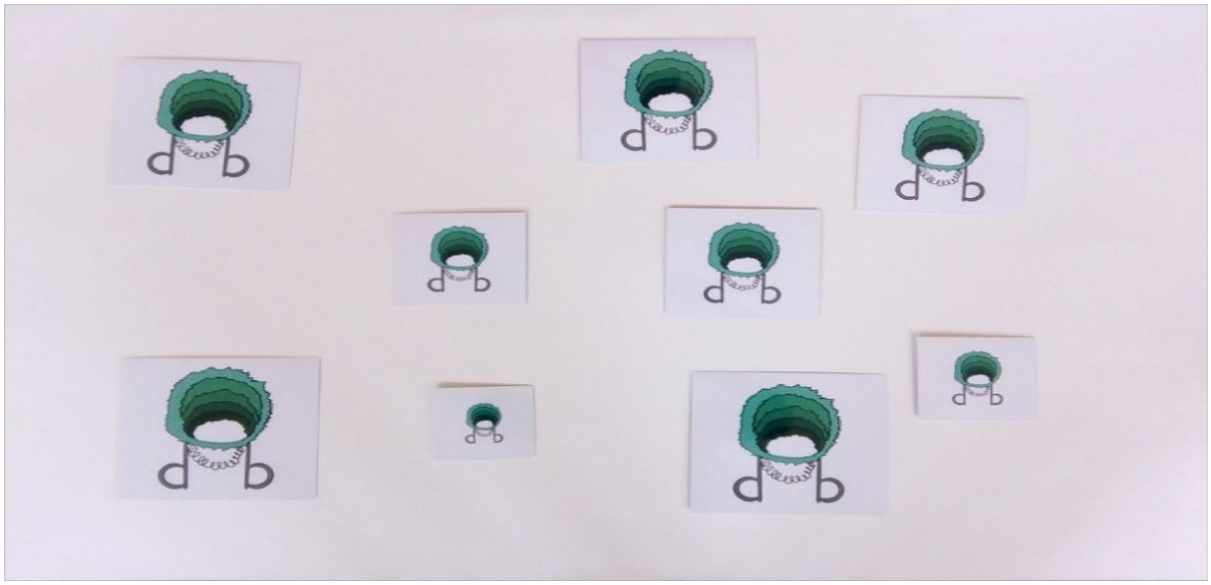


Figure 1. Example of test-item: nine objects on the white cardboard (test item: *catabombi*)

At the beginning of the experimental trials, the child was told s/he was about to enter Fantasyland, where there are creatures and objects that do not exist in the real world, and that s/he had to select some of them according to what the examiner would tell her/him. For each set (that is, for each sheet of pseudo-randomly arranged objects of the same type) two trials were comprised: one 'lungo' and one 'corto' trial, or one 'grande' and one 'piccolo' trial, depending on the item. Adjectives were always tested in pairs: *lungo* vs. *corto* and *grande* vs. *piccolo*, presenting the linguistic contrast to the participant. The cardboard sheets were piled one upon the other and placed in front of the child. After each set had been tested the sheet was removed, revealing the next one with its new set of objects. On the back of each card there was a number corresponding to the size of the object in the 1-9 scale, so that the examiner could register the child's answer.

When the child was shown a new set of objects (as in Figure 1), s/he was asked to select the cards which according to her/him represented the *big/small/long/short* N, where N stands for the plural name of one of the relevant items. For example, s/he was told: "Guarda! Questi sono dei *catabombi*. Scegli i *catabombi grandi*" ("Look, these are some *catabombi*. Select the *big catabombi*").[3] The child would then remove all of the selected cards and put them in a box to indicate her/his choice. Between trials, cards were returned to their original positions on the cardboard, therefore children could select the same object(s) in both trials, i.e. an object could be labelled, for instance, as both *big* and *small* (or neither *big* nor *small*). After the objects had been relocated on the cardboard, the child was asked to look again at all of the N and select the cards which according to him/her represented the *small* N: "Adesso guarda di nuovo tutti i *catabombi* e scegli i *catabombi piccoli*" ("Now look again at all of the *catabombi* and choose the *small catabombi*"). Once the child had made her/his choice, the examiner moved to the next sheet (and, thus, set). After each couple of trials, if the child had not chosen one or more objects in neither trial, *viz*. if s/he did not apply neither the long nor the short label to the object(s), the examiner asked the child, for instance: "E questi *catabombi*? Sono *catabombi grandi* o sono

[3] In the test's questions, subsective adjectives were located in postnominal position, as this allows for a restrictive interpretation of the adjective, which is required in order to choose the proper set of objects (Cinque, 2010).

catabombi piccoli?" ('What about these *catabombi*? Are they *big catabombi* or are they *small catabombi*?), and registered the participant's answer. The methodological reason behind the choice of asking participants to classify non-selected objects is that non-choice in itself, especially when it comes to children's (non)answers, can be due to a number of different and otherwise unidentified reasons. By asking participants to provide a reason for not having classified non-chosen objects, we gain clearer data quality in terms of non-choice. It should also be pointed out that eliciting answers did not always produce value adaptation, in other words, participants did not always force non-chosen objects into either one or the other possible category (e.g. either big or small; see the Discussion section).

In total, each participant was tested on 16 items: four object-names presented twice for two adjectival pairs.

Before beginning the test, each child was given a pre-test to familiarize them with the task and to test their lexical knowledge of the adjectives used in the experiment. The pre-test consisted of two cardboard sheets of the same kind of those used in the experiment conditions. Two figures were placed one next to the other in the middle of each sheet: two triangles in one sheet, and two horizontally oriented rectangles in the other one. For each couple, one of the two figures had the same height (in the case of triangles) or length (in the case of the rectangles) of object 1 above, the other one the same as object 9 (triangle and rectangle). Therefore, one figure was unambiguously smaller/shorter than the other. The child was first asked to select the small triangle by detaching it from the sheet and putting it into a box. The examiner then repositioned it and showed the child the following sheet - the one with the two rectangles. The child was asked to select the short one in the same way as s/he did before. Therefore, for instance, the child was told: "*Questi sono due triangoli. Scegli il triangolo piccolo*" ('These are two triangles. Select the small triangle'). Once children had successfully completed the pre-test, they moved to the experimental trials. All of the children performed correctly in the pre-test.

5 Results

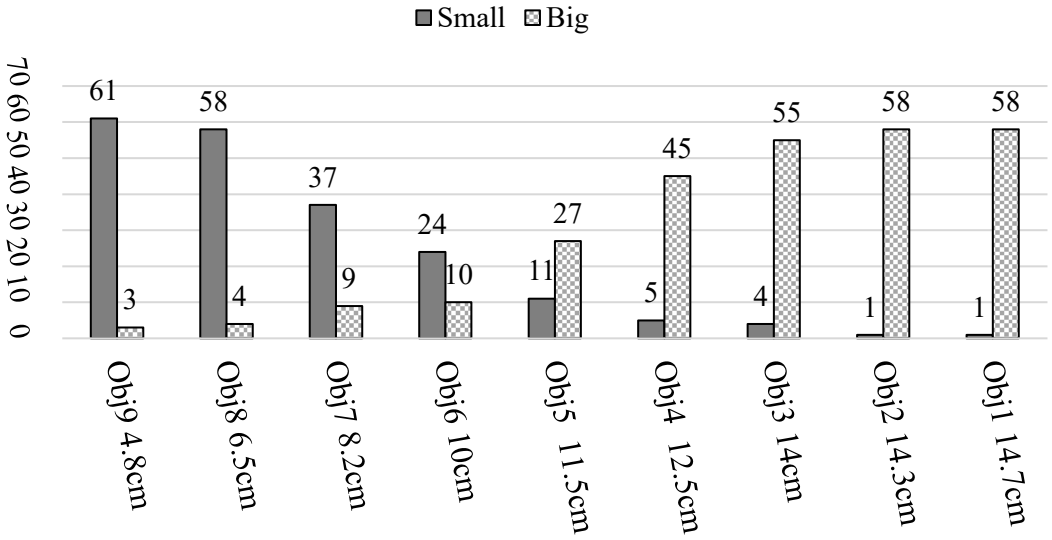
In order to investigate how Italian children and adults interpret subsective gradable adjectives modifying novel nouns, and to address our four specific research questions, we performed two analyses. We first analyzed which objects on the cardboard were considered by children and adults as being long or short and which ones were assigned the value big or small (Section 5.1). This enabled us to determine whether children considered the depicted objects on the card a single comparison class for adjective interpretation (Q1), whether they set a standard of comparison and if so, whether this differs from that of adults (Q2), and finally, whether (multi)dimensionality has an effect in adjective interpretation (Q3). Next, we performed an individual analysis of the non-selected objects as well as of the objects which received both values, e.g., long and short, by the same participant (Section 5.2). In so doing, we aimed to establish whether the vagueness of subsective gradable adjectives was manifested in our participants' responses and whether both adults and children identified similar objects, and thus sizes, as being part of a grey area (Q4).

In the children's group we collected 256 datapoints: 128 for each adjectival-pair condition. Of these datapoints, we only analyzed those responses in which a linear selection was performed, namely, a selection that followed the linear increase of the objects' size. Hence, our analysis was based on 227 datapoints out of the 256 responses.

In the adults' group we collected 320 datapoints: 160 for each adjectival-pair condition. As we did for children, we further analyzed the responses with a linear selection. Hence, our analysis considered 317 responses out of the 320 datapoints.

5.1 What Counts as Long/Big or Short/Small?

We first investigated how frequently each object was selected by children in the small/big condition. Graph 1 shows the amount of times each object was chosen by children in the small/big condition. The total of test items in which a linear selection was performed with the adjective 'small' was 61, whereas it was 58 with the adjective 'big'. Recall that we manipulated our objects' size from object 9 being the smallest object to object 1 being the biggest one.

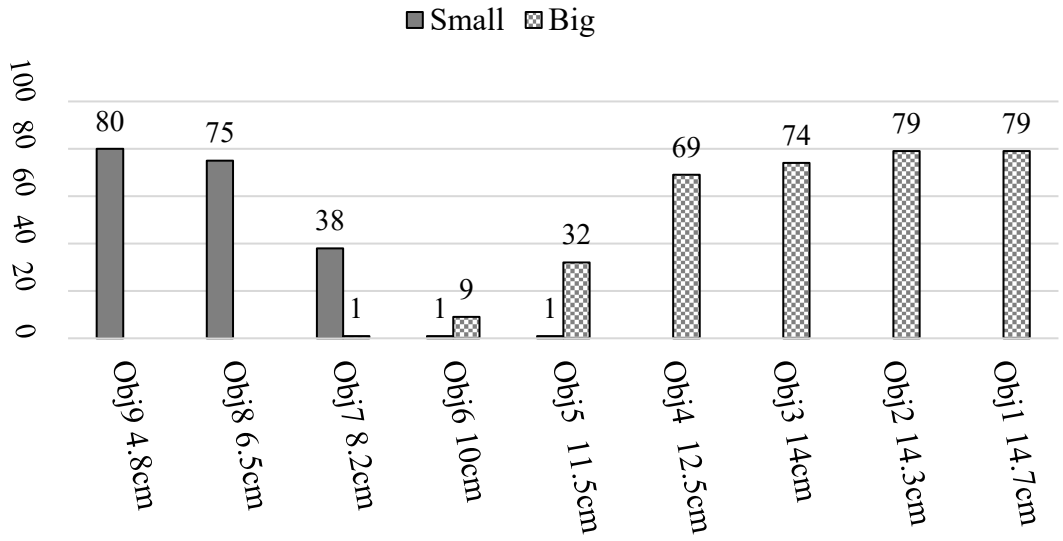


Graph 1. Number of test items each object was selected by children in the small/big condition

Graph 1 demonstrates that children properly performed the task selecting the objects on the cardboard. Moreover, our data show that in the majority of items, the objects selected and, thus, considered by children as 'big' were objects 1, 2, 3, and 4. Conversely, the objects selected and thus considered as 'small' objects were objects 9, 8, and 7.

Graph 1 also reveals that children chose all objects on the cardboard three times when asked to select the 'big' N, and once when asked to select the 'small' N. Only two children did so. Child GB, 44 months old, exhausted all objects on the cardboard twice in the 'big' condition. Child TZ, 71 months old, selected all objects once in the 'big' condition and once in the 'small' condition. No child showed the extreme label, i.e. the choice of only the biggest object, object 1, in the 'big' condition. In contrast, the extreme object, object 9, was uniquely selected three times by two children, NR (69 months) and TZ (71 months) in the 'small' condition.

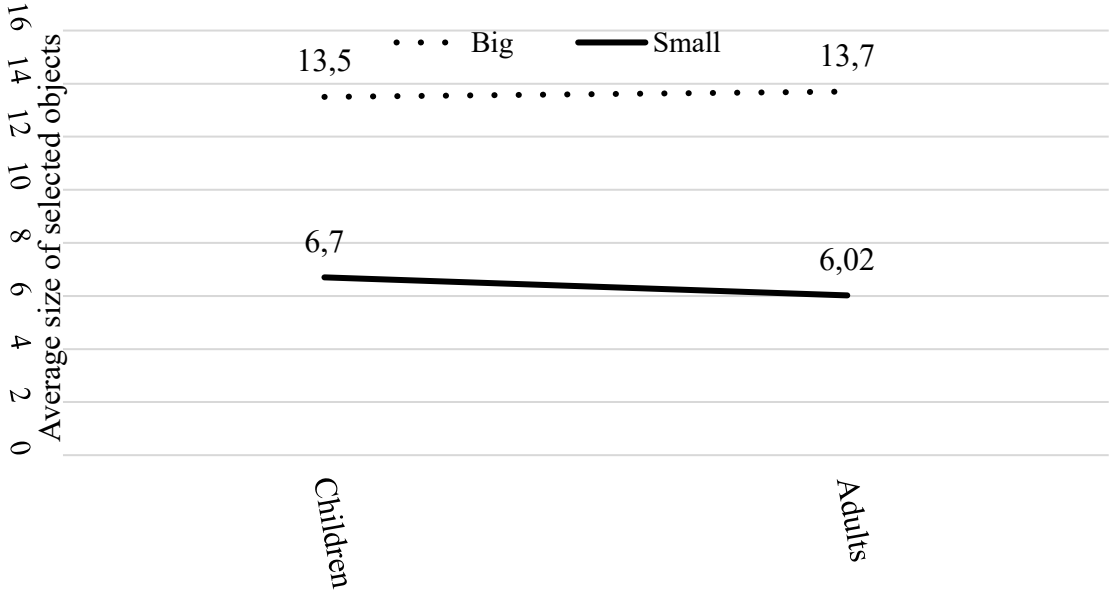
We next performed the same analyses on the adults' responses. As for children, we only considered those responses in which a linear selection was performed. Hence, the total of test items in which a linear selection was performed with the adjective 'small' was 80, whereas it was 79 with the adjective 'big'. Graph 2 shows the amount of times each object was chosen by adults in the small/big condition.



Graph 2. Number of test items each object was selected by adults in the small/big condition

Graph 2 shows that, in addition to the extreme objects, in the majority of test items adults selected objects 8 in the 'small' condition and objects 2, 3, and 4 in the 'big' condition. Differently from children, no adult selected all the objects on the cardboard in either conditions. Similar to children, three adults showed the extreme label pattern: they selected only the extreme object and they did so five times in the 'small' condition.

As in Barner & Snedeker (2008), we calculated the average size of the objects chosen by children and adults in the big/small conditions. Graph 3 depicts this information.

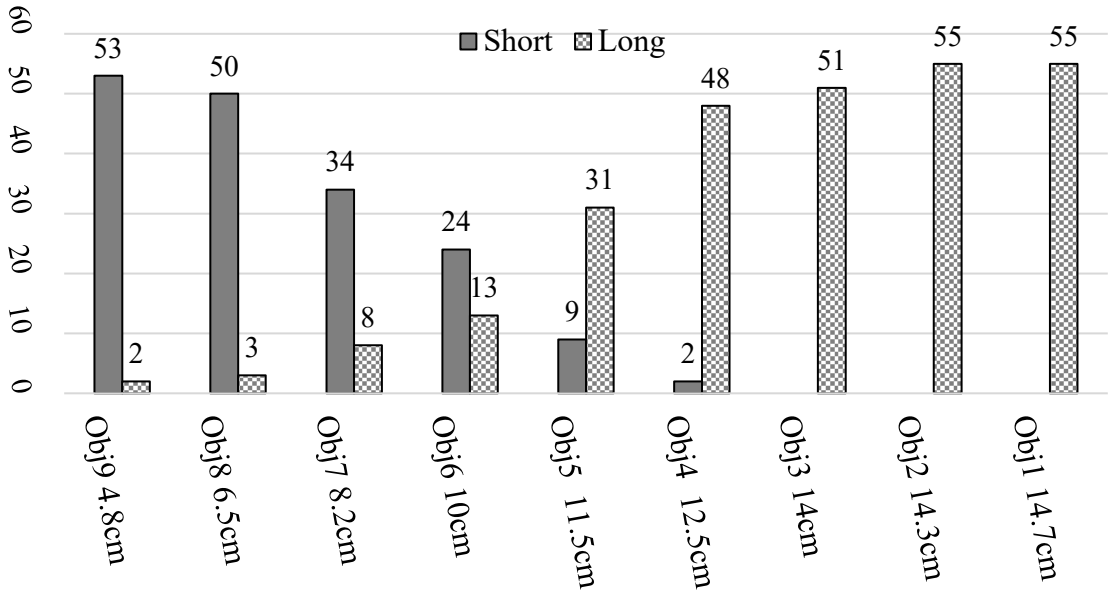


Graph 3. Average size of the objects selected in the small/big condition across age groups

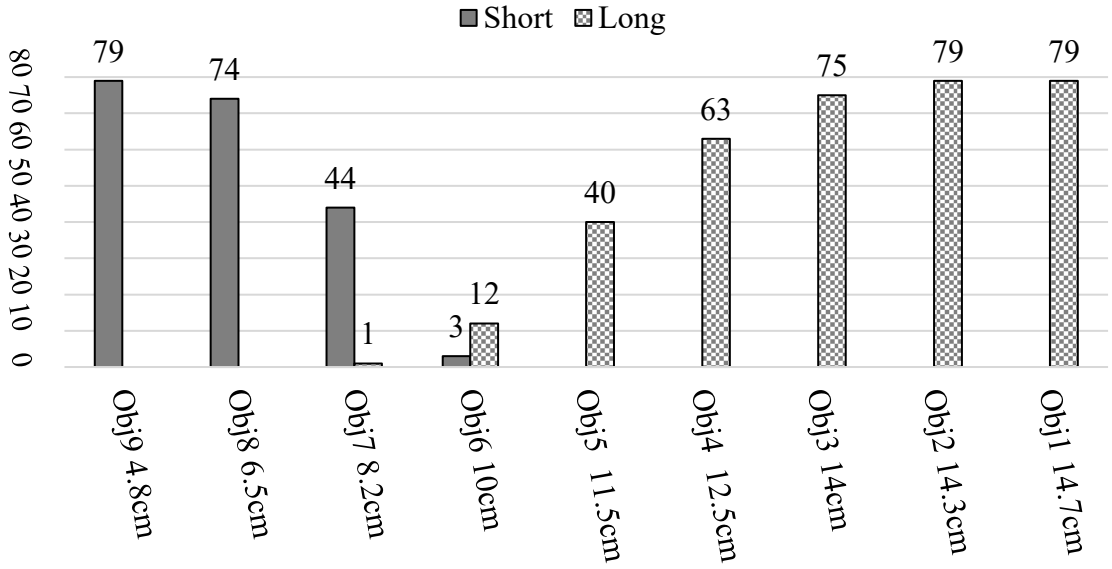
Data were submitted to a two-way analysis of variance (ANOVA) with average size as our dependent variable, size (big vs. small) and age (children vs. adults) as our fixed factors. There was a significant effect of size ($F(1,1) = 2689.5$; $p < .001$), but there was not a significant effect

of age (F (1,1)= 3.639; p=.061). The model also revealed a significant interaction between the factors age and size (F (1,1)= 11.321, p=.001). Post hoc comparisons using the Tukey HSD test on the age by size interaction indicated that children's average of the chosen object significantly differed from that of adults in the 'small' condition (p<.001) but not in the 'big' condition (p=0.7).

Next, we investigated how frequently each object was selected by children and adults in the long/short conditions. Graph 4 shows the number of test items in which each object was selected by children, whereas Graph 5 depicts the same information for adults.



Graph 4. Number of test items each object was selected by children in the short/long condition

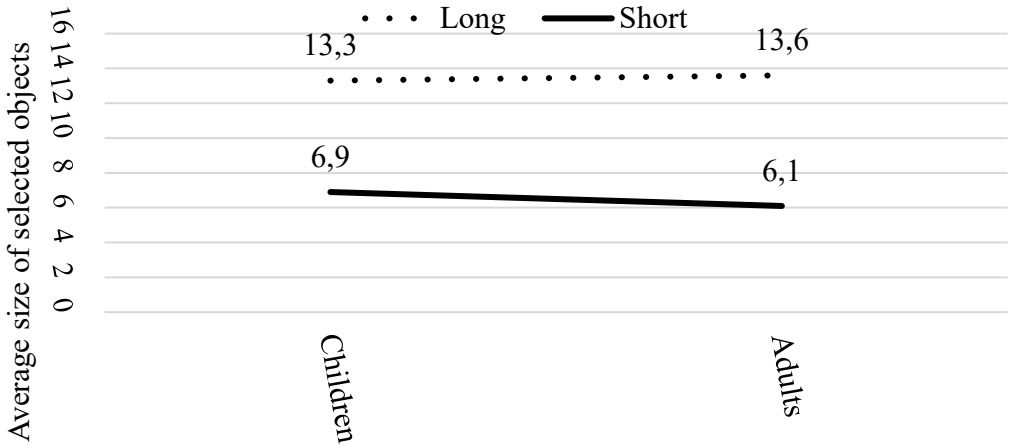


Graph 5. Number of test items each object was selected by adults in the short/long condition

Similar to what we saw with Graph 1 for the small/big condition, children properly performed the task selecting the objects on the cardboard. Graphs 4 and 5 show that in at least the majority

of test items the objects which were assigned the value 'long' by both children and adults were objects 1, 2, 3, 4, and 5, while objects 9, 8, 7 were selected as 'short' objects. Our data also revealed that two children, GB (44 months) and TZ (71 months), considered all objects as being 'long' twice in the 'long' condition. In contrast, no adults exhausted all the objects on the cardboard in either the long or the short condition. In addition, three children, NR (69 months), ST (39 months), and TZ (71 months) exhibited the extreme label pattern: they labeled only object 9 as 'short' for a total of three times in the 'short' condition. This pattern is also attested in the adults' responses: three adults did so in the 'short' condition. The extreme object labeling did not occur in the 'long' condition in either children or adults.

As in Barner & Snedeker (2008), we calculated the average size of the objects children and adults chose in the short/long condition. This information is reported in Graph 6.



Graph 6. Average size of the objects selected in the short/long condition across age groups

A two-way ANOVA was performed with the average of the size of the chosen objects as the dependent variable. Age (children vs. adults) and length (long vs. short) were set as fixed factors. The model revealed a significant effect of the factor length ($F(1,1) = 2757.9$; $p<.001$) but no effect of age ($F(1,1) = 2.22$; $p=0.14$). In addition, a significant interaction between the factors length by age was found ($F(1,1) = 14.8$; $p<.001$). We then performed post hoc comparisons using the Tukey HSD test, which revealed that the average size of the objects children chose significantly differed from that of the objects selected by adults in the 'short' condition ($p=.002$) but not in the 'long' condition ($p=.35$).

Finally, we investigated whether there were differences in the average sizes our participants selected across the two adjectival pairs according to dimensionality. Whereas long/short are defined as unidimensional adjectives, big/small are claimed to be multidimensional. We compared the average sizes children and adults selected in the 'big' vs. 'long' conditions and in the 'small' vs. 'short' conditions. The ANOVA with the average size as the dependent variable and dimensionality (big vs. long) as fixed factor indicated no significant effect of dimensionality ($F=0.6$; $p=0.45$) in the big/long condition as well as in the small/short condition ($F=0.6$; $p=0.43$). Hence, the average size of the objects our participants selected did not differ depending on whether the adjective was multidimensional or unidimensional.

5.2 Non-selected and Doubly Selected Objects

As our final analysis, we investigated whether all objects were chosen in each adjectival pair, whether some were not selected or whether some were considered as both big/small (or long/short). The non-selection of an object in both the big/long prompts and the small/short ones as well as the selection of the same object in both antinomic pairs can be interpreted as the impossibility or difficulty of assigning a positive or a negative value to a given object. Hence, for each novel noun in each adjectival pair, e.g., *catabombi* in the big/small adjectival pair, we controlled whether the selection in both prompts exhausted all the objects on the cardboard, and if not, whether some objects were not selected, or alternatively, whether some objects were doubly selected.

In the big/small condition, two children exhausted all the objects on the cardboard in the two prompts. One child in one item treated object 6 as being both big and small. Fourteen children did not select at least one object in at least one item. The non-selected objects range from object 4 to object 7. The average size of the non-selected objects was 10.4, a size which situates between objects 5 and 6. In the long/short condition, two children exhausted all the objects on the cardboard: one child was one of the two that did so in the big/small condition. In two items each, four children assigned objects 5 and 6 both the long and the short values. In this condition fourteen children did not select at least one object in at least one item. The non-selected objects range from object 4 to object 7. The average size of the non-selected objects was 10.5, a size which situates between objects 5 and 6.

We performed the same investigation for the adult participants. In the big/small condition, only one adult exhausted all the objects on the cardboards with the two prompts. No adults assigned an object to both values. Nineteen adults did not choose at least one object in at least one item. As found for the children's group, the non-selected objects range from object 4 to object 7. The average size of the non-selected objects was 10.1, a size which situates around object 6. In the long/short condition, two adults exhausted all the objects on the cardboards with the two prompts. No adults assigned an object as being both 'long' and 'short'. Eighteen adults did not choose at least one object in at least one item. The average size of the non-selected objects was 10.02, a size which situates around object 6.

We then calculated whether the average size of the non-selected objects differed across age groups and across adjectival pairs. We performed a two-way ANOVA which revealed a significant effect of age ($F(1,1)=5.7$; $p=.02$), but no significant effect of the adjectival pair type ($F(1,1)=6.1$; $p=0.9$). In addition no significant age by adjectival pair interaction was found ($F(1,1)=0.3$; $p=.6$).

6 Discussion

This paper investigated how Italian children and adults interpret subsective gradable adjectives modifying novel nouns. The aim of our study was to verify whether children deploy the compositional knowledge needed to interpret such adjectives and have the ability to extract size information about an identified comparison class. In particular, we were interested in understanding (Q1) whether children considered the depicted objects on the card a single comparison class for adjective interpretation, (Q2) whether they set a standard of comparison and if so, whether this differs from that of adults. In addition, we wanted to investigate (Q3) whether (multi)dimensionality had an effect in adjective interpretation, and (Q4) whether gradable adjectives are vague for children as well as for adults.

To address these questions, a card selection task was designed testing two adjectival pairs, 'long/short' and 'big/small'. Participants were shown a cardboard depicting nine identical objects which only differed with respect to their size: their size gradually diminished from object 1 being the longest/biggest to object 9 being the shortest/smallest. We asked participants to select from the nine objects on the cardboard which object(s) they considered big or small, and similarly long or short.

As for (Q1), our results showed that children properly performed the task. Since objects with different length or size, but same name, were included in determining the application of an adjective-noun pair, we can conclude that children considered the objects on the cardboard as a comparison class. This suggests that applying the same name to an array of novel and physically distinct objects is sufficient for grouping these objects in a single comparison class for adjective interpretation. Our finding is in line with the results on English-speaking children in Barner & Snedeker (2008).

As for (Q2), i.e. whether children set a standard of comparison, and if so, whether it is different from that of adults, our analyses demonstrate that children clearly distinguished between big/long objects and small/short objects from a novel array. The average size of the objects labeled big/long significantly differed from the average size of the objects labeled small/short. Given that we used novel nouns, previous ostensive learning could not account for children's behavior. Likewise, no world knowledge about the range of typical size/length of our objects could play a role in adjective interpretation. Since novel objects can only be labeled long or short relative to other objects of the same class and children succeeded in distinguishing big/long vs. small/short objects, we can conclude that children correctly interpret gradable adjectives, thereby mapping the objects on a scale of size/length, and that they compute a standard of comparison by which they evaluated each object. In line with the findings reported in Barner & Snedeker (2008) on English-speaking children, our study reveals that Italian children can deploy compositional knowledge: children already at the age of 3 derive composed meanings from novel adjective-noun combinations. Both children and adults set a similar average of the size of the objects considered as big/long. In contrast, a difference in the average of the size of the objects considered small/short emerged between children and adults.

This difference resembles the asymmetry found in children's interpretation of positive and negative terms reported in various studies (Clark, 1972, 1973; Smith et al., 1986), which suggest that children generally master "positive" terms like big/long in an adult-like way earlier than "negative" terms like small/short.

The individual analysis we performed revealed two interesting aspects. Whereas the majority of our participants selected some of the objects on the cardboard, two further response patterns emerged: (i) an exhaustive response pattern, and (ii) an extreme labeling response pattern. As for (i), our data showed that children and adults differ. Whereas no adult exhibited this response pattern, two children exhausted all objects in one adjective condition: they considered all the depicted objects as being, for instance, 'long'. One child, GB, did so only when the adjective was the positive term, long and big. The other child, TZ, did so in the long, big, and small conditions. This suggests that the exhaustive response pattern may be connected with development. However, further research is needed to fully understand this response pattern. Indeed, it should be stressed that both children did not exhibit the exhaustive response pattern consistently: they selected all the objects in few test items only. As for the extreme labeling response pattern (ii), previous acquisition literature has reported that children exhibit "extreme labeling": when presented with a series of items that decrease along a relevant dimension, children usually label only the extreme item of the series as big or small, while adults put the cross-over point around the mid of the series (Smith et al., 1986; Syrett, 2007; Panzeri et al., 2013). Our data show that both children and adults exclusively selected the extreme object in the small and short conditions. Three children, NR, ST, and TZ, exhibited this pattern: NR and TZ selected only the extreme object in both small and short conditions, whereas ST did so only in the short one. In addition, three adults showed the extreme object labeling in the small and short condition. Differently from what reported in previous studies, our study demonstrates that this response patter is not proper of children only, but it can be found in adults as well. In addition, no child or adult exclusively exhibited the "extreme labeling" across all the test items. Therefore, it seems plausible to conclude that what is responsible for the "extreme labeling" may have a processing nature, more than a linguistic or developmental one and it may suggest an explanation in terms of performance or task effect.

Next, we investigated (Q3) whether (multi)dimensionality had an effect on our participants' computation of the standard of comparison. The theoretical literature has distinguished gradable adjectives between unidimensional and multidimensional adjectives. Whereas it is possible to associate with unidimensional adjectives a unique measurable aspect on which membership depends, multidimensional adjectives like 'big', instead, order individuals along more than one ordering dimension (Kamp, 1975; Klein, 1980). Previous acquisition studies have claimed that the order of acquisition of dimensional adjectives goes from the more general concept of dimensionality (e.g., big) to more dimension-specific ones (e.g., long) (Clark 1970; Bartlett 1976). We may rephrase these findings in terms of (multi)dimensionality and conclude that multidimensional adjectives seem to be acquired prior to unidimensional ones. No significant difference was found in our study in children's responses depending on (multi)dimensionality. Further research with a greater sample size is needed to fully address this aspect.

Finally, turning to (Q4), our study demonstrates that children interpret relative gradable adjectives as being vague. Our last analysis in Section 5.2 on the non-selected objects as well as on the objects which received both values by the same participant revealed that both children and adults seem to define a common 'grey area', i.e. to share a similar, though not identical, perception of what counts as being 'vague' along the *gradatum* formed by the objects. Few children selected the same object(s) in both the positive, long/big, and negative, short/small prompt, thereby assigning both values to the object(s). In addition, some objects were not assigned either the positive or the negative value in both the children's and the adults' groups. The average size of the non-chosen objects, although descriptively similar, significantly differed between children and adults. Whereas adults considered the size instantiated by object 6 as neither big/long or small/short, children's grey area was wider, including the size range between object 5 and 6. Although our results are based on a very small dataset and thus, further research is needed, this difference may indicate that in children's perception more borderline cases might be tolerated than in adults' perception (see Fiser & Aslin, 2002, on cognitive and visual perception development). Interestingly, in addition to the indirect measure of non-selection or double selection, we also collected direct evidence for concluding that gradable adjectives are vague for children. Our conclusion is further supported by the comments made by the participants who left some object(s) without a value at the end of the selection task. When they were asked to clarify why a given object had not been selected in either prompts, both children and adults replied using the following expressions: the object is/ the object are *medio* 'average', *metà via/via di mezzo/nel mezzo* 'midway', *normale* 'normal'. Gradable adjectives are vague, as they give rise to borderline cases, *viz*. cases in which it cannot be clearly stated whether the adjective applies or not to the object it refers to (Kennedy, 1999, 2012; van Rooij, 2011). Vagueness blurs boundaries between what can be certainly judged as being, for instance, long, and what can be certainly judged as not being long. This is a long-lasting complex issue in philosophy of language which we, of course, have no pretension of solving here. What we can appreciate from our data is that there is a common agreement on which are the 'vague' objects forming the fuzzy center of our sequence, between the two extremes. Likely, these vague objects have both qualities and neither quality at the same time: when compared with the undoubtedly long objects, they are not that long, but if compared with the undoubtedly small, they are not that small either: they are borderline.

In conclusion, our study demonstrated that children have the ability to extract size information about an identified comparison class. By means of both indirect and direct evidence, we also proved that children, like adults, interpret these adjectives as being vague. In addition, since children assigned the objects the labels 'long/short/big/small' and they evaluated lengths/sizes on the basis of a standard of comparison established within the set of depicted objects, we conclude that children also have the compositional knowledge needed to interpret subsective gradable adjectives: they map objects on a scale of size/length. Our study represents a first step in the acquisition of subsective gradable adjectives in Italian, but many further questions remain open. Further experiments may be needed to make this conclusion more robust

and to clearly disentangle whether the nominal information is indeed interpreted or whether it is only a perceptual difference in size/length independent of the noun. We would need an experimental setting in which among the nine depicted objects with different sizes there are two types of Ns, e.g., *gumpi* and *catabombi*. In addition, other properties of subsective gradable adjectives, e.g., shift of the standard of comparison or contextual variability, remain to be investigated in order to fully determine the acquisition of their semantics.

References

Bartlett, E. (1976). Sizing things up: the acquisition of the meaning of dimensional adjectives. *Journal of Child Language, 3*, 205-219.

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

Bylinina, L. (2016). Judge-dependence in degree constructions. *Journal of Semantics, 34*, 291-331.

Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller, & M. Halle (Eds.), *Linguistic Theory and Psychological Reality* (pp. 264-293). Cambridge, MA: MIT Press.

Castroviejo, E., McNally, L., & Weideman-Sassoon, G. (2018). *The Semantics of Gradability, Vagueness, and Scale Structure*. *Experimental Perspectives*. US: Springer.

Cinque, G. (2010). *The Syntax of Adjectives: A Comparative Study* (Vol. 57). Cambridge, MA: MIT Press.

Clark, E. V. (1972). On the child's acquisition of antonyms in two semantic fields. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 750-758.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 65-110). New York: Academic Press.

Clark, H. H. (1970). The primitive nature of children's relational concepts. *Cognition and the Development of Language*, 269-278.

Cresswell, M. J. (1976). The semantics of degree. In B. Partee (Eds.), *Montague Grammar* (pp. 261-292). New York, NY: Academic Press.

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

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

Fara, D.G. (2000), Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics, 28*(1), 45-81.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*(24), 15822-15826.

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

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. *Cognition, 17*(2), 85-136.

Hohaus, V., Tiemann, S., & Beck, S. (2014). Acquisition of comparison constructions. *Language Acquisition, 21*(3), 215-249.

Kamp, H. (1975), Two theories of adjectives. In E. Keenan (Ed.), *Formal Semantics of Natural Language* (pp. 123-155). Cambridge: Cambridge University Press.

Keefe, R. (2000). *Theories of Vagueness*. Cambridge: Cambridge University Press.

Kennedy, C. (1999). *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. New York & London: Routledge.

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

Kennedy, C. (2012). Adjectives. In D. G. Fara & G. Russell (Eds.), *The Routledge Companion to Philosophy of Language* (pp. 328-341). New York & London: Routledge.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language, 81*(2), 345-381.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy, 4*(1), 1-45.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy*, *28*(6), 643-686.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*(2), 333-367.

Matthei, E. H. (1982). The acquisition of prenominal modifier sequences. *Cognition, 11*(3), 301-332.

McNally, L., & Stojanovic, I. (2017). *Aesthetic adjectives*. In J. Young (Ed.), *The Semantics of Aesthetic Judgment* (pp. 17-37). Oxford, Oxford University Press.

Morzycki, M. (2016). The lexical semantics of adjectives: More than just scales. In M. Morzycki (Ed.), *Modification. Key Topics in Semantics and Pragmatics* (pp. 13-87). Cambridge: Cambridge University Press.

Panzeri, F., Foppolo, F., & Guasti, M. T. (2013). 10 Acquisition meets comparison: an investigation of gradable adjectives. In I. Caponigro & C. Cecchetto (Eds.), *From Grammar to Meaning the Spontaneous Logicality of Language* (pp. 266-293). Cambridge: Cambridge University Press.

Partee, B. H. (1995). Lexical semantics and compositionality. In L. Gleitman & M. Liberman (Eds.), *Language. An Invitation to Cognitive Science* (pp. 311-360). Oxford University Press.

Russell, B. (1923). Vagueness. *The Australasian Journal of Psychology and Philosophy, 1*(2), 84-92.

Sassoon, G. W. (2013). *Vagueness, Gradability and Typicality: A Comprehensive Semantic Analysis*. Leiden: Brill.

Schmidt, L. A., Goodman, N. D., Barner, D., & Tenenbaum, J. B. (2009). How tall is tall? Compositionality, statistics, and gradable adjectives. In *Proceedings of the 31st annual conference of the cognitive science society*, *Vol. 31* (pp. 2759-2764).

Sera, M., & Smith, L. B. (1987). Categorical and relative interpretations of ‘‘big’’ and ‘‘little’’. *Cognitive Development, 2*, 89-112.

Sera, M., Troyer, D. E., & Smith, L. B. (1988). What do two-year-olds know about the sizes of things? *Child Development, 59*, 1489-1496.

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

Smith, L. B., Rattermann, M. J., & Sera, M. (1988). “Higher” and “lower”: Comparative and categorical interpretations by children. *Cognitive Development, 3*(4), 341-357.

Solt, S. (2018). *Multidimensionality, subjectivity and scales: Experimental evidence*. In E. Castroviejo, L. McNally & G. Weideman-Sassoon (Eds.), *The Semantics of Gradability, Vagueness, and Scale Structure* (pp. 59-91). US: Springer.

Syrett, K. (2007). *Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives*. (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.

Tribushinina, E. (2013). Adjective semantics, world knowledge and visual context: comprehension of size terms by 2- to 7-year-old Dutch-speaking children. *Journal of Psycholinguist Research, 42*, 205-225.

Tribushinina, E., & Gillis, S. (2012). The acquisition of scalar structures: production of adjectives and degree markers by Dutch-speaking children and their caregivers. *Linguistics, 50*(2), 241-268.

van Rooij, R. (2011). Vagueness and linguistics. In G. Ronzitti (Ed.), *Vagueness: A Guide* (pp. 123-170). Dordrecht: Springer.

Weicker, M., & Schulz, P. (2019). Red train, big train, broken train. In T. Ionin & M. Rispoli (Eds.), *Three Streams of Generative Language Acquisition Research: Selected Papers from the 7th Meeting of Generative Approaches to Language Acquisition - North America, University of Illinois at Urbana-Champaign*, *Vol. 63* (pp. 203-221). Amsterdam: John Benjamins Publishing Company.

Wheeler, S. C. (1972). Attributives and their modifiers. *Noûs, 6*(4), 310-334.

Williamson, T. (1994). *Vagueness*. New York & London: Routledge.

Wright, C. (1975). On the coherence of vague predicates. *Synthese, 30*(3), 325-365.

Wolfsdorf, D. C. (2019). *On Goodness*. Oxford: Oxford University Press.