NELS 40

Proceedings of the
Fortieth Annual Meeting of the
North East Linguistic Society

Volume One

Massachusetts Institute of Technology

November 13-15, 2009

Edited by
Seda Kan, Claire Moore-Cantwell, and
Robert Staubs

Published by GLSA (Graduate Linguistics Student Association)
Department of Linguistics
South College
University of Massachusetts
Amherst, MA 01003-7130
U.S.A.

glsa@linguist.umass.edu
http://www.linguist.umass.edu/linguist/GLSA

ISBN-13: 978-1492268741
ISBN-10: 1492268747

Table of Contents

Volume I

Volume II

Do children know when their room counts as clean? *

Francesca Foppolo & Francesca Panzeri

University of Milano-Bicocca

1. Theoretical background

Gradable adjectives (GAs) denote properties that can be graded, measured. We adopt a distributional criterion to delimit the class: GAs are those adjectives that can enter into a comparative construction, and can be modified by degree expressions (such as *very*, *5-feet*, *absolutely*, etc.). Thus, the adjectives in (1) may be classified as gradable, and those in (2) as non-gradable:

(1) a. Leo is taller than Fred.
b. This bottle is less full than that one.
c. The table is very dirty.

(2) a. * Napoleon is more dead than Churchill.
b. * Leo is very unemployed.
c. * A cat is an absolutely four-legged animal.

An influential approach (cf. Kennedy, 1999 and references therein) assumes that gradable adjectives map the object they refer to onto a scale of fully ordered degrees. A GA like *tall*, for instance, is a function from individuals to degrees of height. Some of the structural features of the scales evoked by GAs have an impact on their linguistic behavior. In particular, scales may have, or have not, intrinsic boundaries.[1] A scale may be "open" – with neither a lower nor an upper boundary. This is the case, for instance, of

* We would like to thank Maria Teresa Guasti, Carlo Cecchetto and the audience at the BIL seminars in Bicocca. This paper is the result of a joint work between the two authors. For the purpose of Italian academy, Francesca Foppolo takes responsibility for sections 2 and 3 and Francesca Panzeri for sections 1 and 4.

[1] Besides being (un)bounded, scales have a "dimension": GAs like *wide* and *long* project onto scales that share the same dimension of height/length, that may be expressed for instance in centimeters, and thus the adjectives can be compared. And scales have a "direction of ordering": pairs of antonyms, e.g. *tall*/*short*, *big*/*small*, may be viewed as projecting onto the same scale, but with opposite directions of ordering of degrees: the "positive" forms of the pairs make reference to an ordering that goes from the left to the right – e.g., from the less tall to the tallest –, the "negative" forms to an inverse ordering – e.g., from the tallest / the less short to the less tall / the shortest.

scales evoked by adjectives like *tall*, *big*, *heavy*.[2] Or, a scale may be closed at one or both ends. Thus, the scale corresponding to an adjective like *clean* contains an upper limit: objects may be ranked into a scale of "cleanness", that goes from very low to higher degrees of neatness. Such a scale is open on its left side (there is no conceivable limit to the dirtiness, unfortunately), but closed on its right side, inasmuch as there is a standard of "absolute purity" that cannot be overcome. The antonym of *clean*, i.e., *dirty*, may be viewed as projecting on the same scale of neatness, but with the opposite direction in the ordering of the degrees. In this case, what was the upper boundary of absolute purity now counts as the lower limit for considering an object "dirty". Examples of scales closed at both ends are the pairs of antonyms *open*/*closed* and *full*/*empty*. A particular door may be more open than another, and a particular glass may be fuller than another (thus, they constitute gradable adjectives), but there is a limit both to the "openness" and to the "closedness" of an object (a limit beyond which that object cannot be "more open" or "more closed"), just like a glass cannot be filled or emptied beyond the limits of what counts as completely full and completely empty.

The presence/absence of boundaries on a scale has an impact on the way GAs are interpreted. In a nutshell, the standard according to which an individual possesses the property denoted by a GA projecting on a totally open scale needs to be contextually fixed: a particular individual counts as "tall" only relative to a standard of height that is contextually retrieved. On the other hand, when a GA evokes a bounded scale, the boundary of the scale serves as the standard: an object is clean if it possesses the maximum standard of cleanness and not if it exceeds a contextual standard of cleanness. Thus, within the class of GAs, we can draw a partition between Relative gradable adjectives (REL GAs), that project on totally open scales, and are necessarily evaluated with respect to a contextual standard,[3] and Absolute gradable adjectives (ABS GAs), that project on closed scales, and whose standard of evaluation is identified with the minimum or maximum standard (min or max std) constituted by the lower or upper boundary of a closed scale.[4]

When we consider GAs that evoke scales with only one end, that only boundary constitutes the standard of evaluation. Thus, *straight* and *bent* are antonyms that project onto the same scale of degrees of straightness, with opposite directions of ordering, and the limit of absolute straightness constitutes both the maximum standard that an object must possess in order to be judged as straight, and also the minimum standard that an object must exceed in order to be considered as bent. When we turn to GAs that project on scales that have two ends, in principle any boundary could constitute the standard: if

[2] It is natural to assume that the scale corresponding to an adjective like *tall* does not have an upper boundary. The claim that it does not have a lower boundary either (that is, that there is no limit to the "shortness") amounts to saying that there are no (conceptual or) linguistic means to pin down a minimum standard – that is, a point beneath which the adjective *short* does not apply anymore.

[3] Relative GAs are always interpreted with respect to a contextual standard. This standard can be explicitly provided, e.g. by means of a class of comparison (as in "Leo is tall *for a horse-jockey*"), or it can be implicitly supplied, making reference to a normative class (e.g., a mitten, even if presented in isolation, can be judged as "big" making reference to an implicit normative class of normal mitten size).

[4] Cf. also Paradis (2001)'s bounded / unbounded adjectives.

we consider for instance the scale of fullness and that one of openness, that have a lower limit (complete emptiness/closedness) and an upper limit (complete fullness/openness), any of these boundaries could count as the standard according to which an object is considered to be full or open. In the literature (cf. a.o. Kennedy and McNally 2005) it is claimed that *full* is interpreted as a maximum standard GA: an object is full only if it is completely filled, whereas *open* is a minimum standard GA: an object is open as long as it has a minimum degree of openness. As for their antonyms, *empty* and *closed*, they are both treated as maximum standard GAs: in order to be judged as *empty* or *closed* an object must have no amount at all of fullness or openness.

2. Experimental background

In the past years, several studies have investigated the interpretation of GAs by children and adults (a.o. Ebeling & Gelman 1988; Barner & Snedeker 2007; Syrett, 2007; Frazier et al. 2008). In particular, Syrett (2007) investigated children's interpretation of Relative and Absolute GAs, in order to establish whether children know that Abs GAs are not evaluated with respect to contextual standards, but by making reference to the maximum or minimum standard that corresponds to the upper or lower limit of the scale.[5] In a Pre Scalar Judgment Task (PSJT), subjects were asked to judge if each element in a series of 7 objects – that were identical except that they decreased with respect to a relevant dimension (e.g. length) – had the property denoted by the adjective (they were asked: "Is this *Adj*?"). She tested two positive Relative adjectives (*big* and *long*) and two Absolute adjectives, a minimum standard GA (*spotted*) and a maximum standard GA (*full*)[6]. Comparing children's and adults' responses, Syrett found that, even if both children and adults differentiate between REL and ABS Gas in general, an interesting difference is found in case of *full*: while adults judge a container "almost full" of lentils as "not full", thus interpreting *full* as a maximum standard GA, 40% of children judge it "full".[7] This finding on *full* was replicated in a second study, a Presupposition-Assessment Task (PAT). Participants were given two objects and were asked to satisfy the request of a puppet ("Give me the *Adj* one"). When subjects were presented with two containers, one almost full of lentils (corresponding to object #2 in the PSJT), and another less than halfway full (corresponding to object #4 in the PSJT), 88% of adults objected to the puppet's request "Give me the full one", saying that there was no "full" container, while 11 out of 18 children gave the puppet the "full*er*" container.[8]

---

[5] Other studies (a.o. Ebeling & Gelman 1988) demonstrated that children as young as 2 are able to interpret Relative GAs by making reference to contextual standards, e.g., perceptual standards (comparing for instance the size of an object to another physically present object) and normative standards (comparing for instance the size of an object to a class standard stored in memory).

[6] Syrett (2007) analyses both *full* and *spotted* as Absolute GAs, without taking into consideration neither the distinction between minimum standard (*spotted*) and maximum standard (*full*) GAs, nor between GAs projecting on scales closed on only one end (*spotted*), or closed on both ends (*full*).

[7] In case of *spotted*, on the other hand, there was no difference between children and adults: any object with at least one spot on it (i.e. that possesses even a minimal amount of the "spotted-property") is judged to be spotted by both groups.

[8] Children behave like adults with Relative adjectives and with *spotted*, handing the puppet "*Adj-est*" in case of "long" and "big", and, crucially, objecting to the request "Give me the spotted one" when presented with two spotted objects.

3 Our experimental study

We started from Syrett's studies. Our aim was to replicate her results concerning children's interpretation of REL and ABS GAs, by using a different experimental design and by testing children and adults in Italian, and moreover, to further investigate her unexpected findings on children's interpretation of *full*.

3.1 Participants

We tested 20 3-year-olds (3;0–3;11, MA= 3;6), 18 5-year-olds (5;0–6;1, MA 5;4), 23 adults. Children were recruited from two daycares in the Milan area and tested in a quiet room in their school. Adults were undergraduate students at the University of Milan-Bicocca who received credits for their participation. Subjects were divided in 2 lists.

3.2 Material and Procedure

Participants were tested by two experimenters: one manipulated a puppet and the other showed the experimental objects and asked questions. The experiment comprised two parts: a Truth Value Judgment Task (TVJT, Crain and Thornton, 1998) and a Scalar Judgment Task (SJT) analogous to the one used by Syrett. As for the TVJT, in each trial the experimenter put a single object on the table in front of the subject; the puppet described it by using an adjective (e.g. "This is *long*"); the participant had to evaluate the puppet's description by using three options of response: "yes, correct", "no, incorrect", "I can't tell/it depends".[9] Adults recorded their own answers on a pre compiled score-sheet, while children's sessions were audio-recorded and the responses were coded and transcribed afterwards.

Participants in each list were shown a total of 24 trials that comprised: 6 Relative GAs (3 "positive" and 3 "negative" antonyms, e.g. *big*/*small*), 6 Absolute GAs that project on a scale closed on one end (3 max std and 3 min std antonyms, e.g. *clean* – max std, and *dirty* – min std); 4 Absolute GAs projecting on a totally closed scale (i.e. *full*/*empty* and *open*/*closed*); 8 fillers and controls. Each object used was part of an "ideal" scale of seven objects decreasing along a relevant dimension. For example, for the scale of "height" we used 7 different rods ranging from the tallest one (20 cm. tall) to the shortest one (5 cm. tall) with intervals of 2.5 cm between the elements in the series. We presented one object in isolation in the TVJT and the whole series in the SJT. A sample of the material is provided in figures 1-3:

[9] To train subjects on this third option of response, in the Warm Up session a toy-boy with a big hat was presented, described by the puppet as "This is blond". Given that the big hat covered the whole head of the boy, it was impossible to see what colour his hair was. Thus, the response "I can't tell" was prompted.

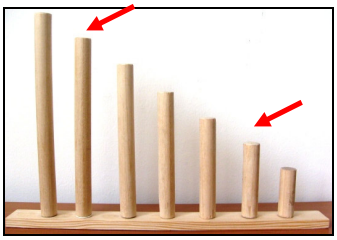

Fig. 1

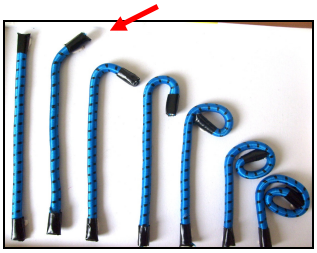

Fig. 2

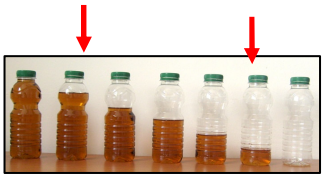

Fig. 3

Fig. 1-3. Sample of the material used to test Relative and Absolute GAs in the TVJT and in the SJT. The set in Fig. 1 was used to test *tall* and *short*, REL GAs; in the TVJT, the 2nd from the left was used for *tall* and the 2nd from the right was used for *short*. The set in Fig. 2 was used to test *straight* (max std) and *bent* (min std), i.e. ABS GAs projecting on a scale closed on one end; in the TVJT, the 2nd from the left was used for both antonyms. The set in Fig. 3 was used to test *full* and *empty*, ABS GAs projecting on totally closed scales; in the TVJT, the 2nd from the left was used for *full* and the 2nd from the right was used for *empty*.

In the case of Relative adjectives, the "positive" antonym was tested using the 2nd element in the series of seven, while the negative antonym was tested by showing the 6th element of the same series. In the case of Absolute GAs projecting on scales closed on one end, the same object in the series was used for testing couples of antonyms: for example, the second rope from the left in Fig. 2 was used to test both *straight* (max std) and *bent* (min std). Finally, in the case of the 4 Absolute GAs projecting on totally closed scales (i.e. *full*/*empty* and *open*/*closed*), two different series of seven objects were created as described above: seven bottles and seven paint-tubes, ranging from being completely full to completely empty; seven purses and a box with lid ranging from being completely open to completely closed. For example, the 2nd item from the left in Fig. 3 was used to test *full* and the 2nd item from the right was used to test *empty*.

Some notes on the changes that we introduced in our experimental design, material and procedure, compared to Syrett's, are in order here. First of all, we decided to begin with the TVJT, asking subjects to judge the appropriateness of a description ("This is *Adj*") referred to a single object in isolation for two reasons: we wanted to avoid the risk that the presentation of a series of 7 objects could induce a comparative interpretation of the GA (e.g. suggesting to interpret *full* as *fuller*); and we believe that the difference between Relative and Absolute GAs can emerge also from this task. The idea is that, since Absolute GAs are interpreted with respect to the "intrinsic" minimum or maximum standard that corresponds to the lower or upper boundary of the scale, they can be judged as having or not having the property in question even if presented in isolation. This means that we expect subjects to be able to judge whether a single object, for instance a rope, do have, or do not have, the ABS GA property, for instance whether it is straight or bent. On the other hand, Relative GAs require the retrieval of a contextual standard to judge whether a particular object has the property in question. Thus, when presented with an abstract object, for instance a wooden 17.5 cm. tall rod, and asked to judge it as *tall*, we expect subjects to split between acceptance/rejection (with 50% of the subjects accepting and the other half rejecting the description) and/or a high proportion of "I can't tell" responses. In the second place, the Absolute GAs tested by Syrett were *full* and *spotted*, and she found unexpected results with children's interpretation of *full*. However,

full and *spotted* exhibit other semantic distinctions that could be relevant: (i) *full* projects on a scale closed on both ends; *spotted* on a scale closed on only one end; (ii) *full* is interpreted as a maximum standard; *spotted* as a minimum standard. Thus, we refined the comparison within the class of Absolute adjectives, distinguishing between those that project on a scale closed on one end and those that project on a totally closed scale, and between maximum and minimum standard. In the third place, pursuing the hypothesis that children's "comparative-like" interpretation of *full* could depend on the type of object used to test it (a container of lentils, that is something "to be filled"), we tested *full* and *empty* using two different sets of objects: bottles (that can be "filled" at a different degree) and paint tubes (that can be "emptied" at a different degree).

The second part of our experiment consisted of a Scalar Judgment Task like the one used by Syrett. Participants were presented with different series of 7 objects, such as those in Fig. 1-3. For each element in the series, they were asked "Is this *Adj*?". A total of 9 scales were created, 4 for Relative GAs, 2 for Absolute GAs projecting on scales closed on one end and 3 for Absolute GAs projecting on totally closed scales. Of these, 2 series were used to test *full*/*empty* (a series of bottles and paint tubes, for the reason discussed above) and a series of purses that were open at different degrees was used to test the couple of antonyms *open*/*closed*.[10] Our general aim was twofold: firstly, we wanted to compare children's and adults' interpretation of Relative vs. Absolute adjectives by means of a different experimental design; secondly, we intended to investigate other differences within the broader class of GAs that have not been previously considered, in particular: investigate the difference between adjectives that project on scales closed on one end (e.g. *clean* and *bent*) vs. scales closed on both ends (e.g. *open* and *full*); and scales with a max std (e.g. *clean* and *full*) vs. scales with a min std (e.g. b*ent* and *open*).

3.2 Results

Our first interest was that of replicating – in a different language (Italian) and by means of a different experimental design – the results obtained by Syrett and others that children and adults differentiate between Relative and Absolute GAs. A note on data coding: in order to facilitate a direct comparison across conditions, the data were coded as follows: "I can't tell" responses were assigned a value of 0.5; in case of Relative GAs, answers "yes" were coded as 1 and "no" were coded as 0; in case of Absolute GAs, answers were coded as "correct" (i.e. assigned the value 1) and "incorrect" (i.e. assigned the value 0), depending on their standard semantics: since max std GAs can be truly predicated only of an object possessing the maximal value of the given property, and since we were testing objects that were close to, but not reaching, the max std (e.g., subjects had to judge "this is straight" referred to a minimally bent rope) the correct answer, coded as 1, was "no", and yes-answers were coded as 0. Please recall that there are max std GAs projecting on scales closed on only one end (ABS, e.g. *clean*¸ *straight*) and on scale closed on both ends (ABSC), i.e. *full*, *empty* and *closed*. On the other hand, min std GAs can be used to

[10] Again, antonyms were tested between lists, while the order of presentation was balanced across participants and lists: half participants in each list were shown a Relative first, the other half encountered *full* as first item. This manipulation was meant to control for the order of presentation, a factor that might affect participants' responses on *full*, as observed by Syrett.

correctly describe any object that possesses even a minimal amount of the property in question (e.g., a cloth with one stain, that is minimally dirty, counts as *dirty*), and thus the correct answer, coded as 1, was "yes", whereas no-answers were coded as 0. In case of ABS GAs (closed on only one end), the min std tested were the antonyms of the max std GAs; and there is only one GA projecting on a scale closed on both ends that is interpreted as minimum standard, i.e. *open*. No transformation was applied to the data of the SJT instead, that will be analyzed differentiating between maximum and minimum standard (and positive/negative polarity for REL for convenience) separately.

As far as the TVJT is concerned, given our experimental design and according to the standard semantics of the tested adjectives, we predicted: (i) correct responses (corresponding to value 1) "at ceiling" for all Absolute GAs; (ii) chance distribution for Relative GAs. For the peculiarity of our experimental design, chance distribution could arise either by a split between acceptance/rejection (i.e., a proportion of "acceptance" around 50%) or, alternatively, by a high proportion of "I can't tell" responses when evaluating, e.g., "This is tall" referred to the 17.5 cm. tall rod shown in isolation. As for the SJT, we expected a replication of Syrett's findings, namely: (i) for adults: a drop of "yes" on the 2nd item in case of max std Absolute GAs and a persistence of "yes" answers up to the 6th item in case of min std Absolute GAs[11]; a drop of "yes" answers around the 4th item in case of Relative GAs instead; (ii) for children: we expected the same trend, eventually with a less categorical drop, as attested in the literature. Overall results of the TVJT are plotted in Fig. 4 below.

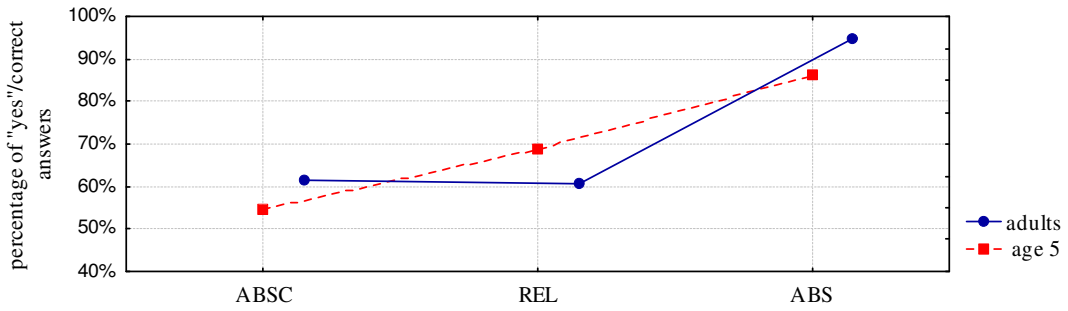

Fig. 4. The TVJT: percentage of correct answers for each condition: Relative GAs (REL), Absolute GAs projecting on totally closed scales (ABSC) and those projecting on scales with one boundary (ABS).

Three main facts are evident from the graph: (i) children at age 5 (represented by the dashed line) seem to pattern similarly to the adults (the solid line) in all conditions; (ii) participants of all age groups clearly distinguish between REL and ABS GAs, as expected. In particular, participants' performance was almost at ceiling in case of ABS (5 year olds children provided 86% correct answers and adults 95%) while it was lower for REL (5 year olds children acceptance rate was 68% and adults' was 60%); (iii) unexpectedly, though, both adults and children also distinguish within the class of

[11]When testing antonyms across lists, we always started asking question from the most representative element in the series, which is crucially reversed across conditions: e.g., we started from the only clean towel if the target question was "Is this clean?" and from the dirtiest one if it was "Is this dirty?".

Absolute GAs, between those projecting on scales closed on one end (ABS) and those projecting on totally closed scales (ABSC): as we said, performance was almost at ceiling for the ABS while the proportion of "correct" answers produced for the ABSC was much lower for both age groups, even lower than that obtained for the REL (54% correct answers for the 5 year olds children and 61% for adults). Statistical analysis confirmed what we evinced from the graph.[12]

We suspected that the bad performance obtained for the class of ABSC was effectively due only to some of the adjectives included in this class. For this reason we plotted the performance of the 5 year olds and the adults on all the ABSC GAs used, differentiating between lists:

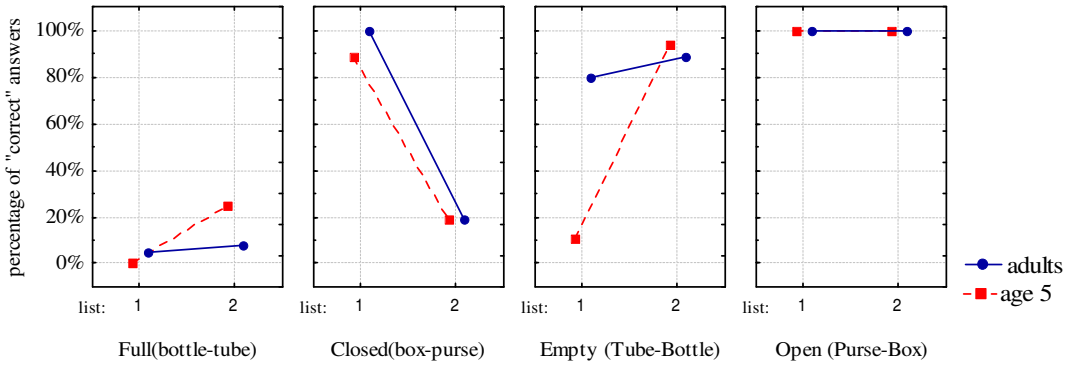

Fig. 5. Percentage of correct answers to ABSC GAs differentiated by age (solid line = adults; dashed line = 5 year old children) and by List (list 1 on the left of each quadrant, list 2 on the right). Remember that different objects were used to test for the same adjective across lists, as specified below each panel. E.g. *full* was tested with a bottle in List 1 and with a paint tube in List 2, and this is captured by the label below the first panel "Full(bottle-tube)". The same specification holds for the other quadrants

Some clear patterns emerge from a first inspection of the graph. In particular, considering the fact that the expected answer was 1 in all cases (i.e. as we said, we expected 100% of correct answers for all ABSC GAs), one immediately gathers that children's and adults' behavior on *open* (the only min std GA projecting on a totally

[12] We first submitted our data to a 3 (Age: age 3;age 5; adults) x 3 (Adjective: REL; ABS; ABSC) analysis of variance ANOVA, including the 3 year olds. A marginal main effect of Age ($F(2, 944)=2,85$, $p=.058$) and a significant effect of Adjective ($F(2, 944)=135,75$, $p<.0001$) were found, plus a significant interaction ($F(4, 944)=18,84$, $p<.0001$). Post hoc analysis by means of LSD Fisher Test showed that the marginal difference across ages was due to the performance of the 3 year olds, significantly different from the adults' ($p=.021$); no difference is revealed between age 5 and adults instead ($p=.49$, n.s.). However, the data obtained in the TVJT from the 3 year olds showed a very unstable and inconsistent pattern: we think that they deserve a closer inspection before being properly analysed and for this reason we will drop them from further analysis and only compare the results from the group of the 5 year olds and the adults in the rest of this paper. For the reason just mentioned, we re-submitted our data to a 2 (Age: age 5; adults) x 3 (Adjective: REL; ABS; ABSC) analysis of variance ANOVA and found no effect of Age ($F(1, 634)=.73$, $p=.39$) this time, but a significant effect of Adjective ($F(2, 634)=47.8$, $p<.0001$) and a significant interaction ($F(2, 634)=3.67$, $p=.03$). Post hoc analysis by means of LSD Fisher Test revealed a significant difference between ABSC and ABS and between ABS and REL for both age groups (all $ps<.001$); a significant difference between ABSC and REL in case of children only ($p<.01$); conversely, no significant difference between adults and 5 year old children was found for any class of adjective (all $ps>.05$, n.s.).

closed scale) is perfectly in line with the stated expectation (cf. 4th panel in Fig. 5), while their performance on *full* seems to pattern exactly the opposite way from what predicted by the standard semantics of this adjective, independently of the age of the participants and of the object used to test it (i.e. they unexpectedly accepted the description "This is *full*" when presented with an almost full bottle or paint tube). As for *empty* and *closed*, a difference seems to emerge across lists and participants, namely: while adults pattern as expected in case of *empty* in both lists (i.e. they rejected the description "This is *empty*" when presented with a not completely empty bottle or paint tube, cf. 3rd panel in Fig. 5), children pattern like adults only in List 2 but not in List 1: like adults, they rejected "*This is empty*" referred to an almost empty bottle but, unlike adults, they accepted it when referred to an almost empty paint tube. Conversely, no age difference is observed in case of *closed*: independently of age, all participants interpreted *closed* as expected in case of an almost closed box (i.e., they rejected the description) but they patterned in the opposite way from what is predicted by the standard semantics of this adjective in case it was evaluated with respect to a not completely closed purse (i.e. they unexpectedly accepted the description). All these observations were supported by statistical analysis.[13]

To account for the asymmetry observed for *closed* and *empty* with respect to different items we can only speculate, at this point. We would like to propose a functional interpretation that might apply to purses and paint tubes: an almost closed purse like the one shown to our participants might be considered "closed" as long as it leaves no room for coins to fall out of it; analogously, a paint tube might be considered "empty" if not enough paint (for a drawing, for example) is left, even if a last shot of paint is in fact left. Interestingly, this functional interpretation for paint tubes was only recorded for children, not for adults, who, presumably, do not preserve the same familiarity with paints, paint tubes and drawings that children have.

Let's now turn to the analysis of the second part of our experiment. Comparable findings seemed to emerge in the Scalar Judgment Task, as shown in Fig. 7: here the percentage of "yes" answers is plotted for each of the seven items in the series and for each age group separately, including the 3 year olds, differentiating between max std/positive and min std/negative antonyms.

[13] Statistical analysis revealed: an overall marginal effect of age ($F(1, 144)=3.80$, $p=.053$); a significant effect of adjective ($F(3, 144)=86.83$, $p<.00001$); no overall effect of list ($F(1, 144)=.99$, $p=.32$); significant interactions between all combinations of these factors (all $ps<.01$). More in detail, post hoc analysis showed that: (i) the only condition in which children differ from adults is *empty*, ($p<.0001$); (ii) adults' performance across lists is different only on *closed* (box vs. purse: $p<.0001$; all other comparisons between lists: $p>.05$, n.s.); (iii) children' performance across lists is different on *closed* in the same way as adults' (box vs. purse in children: $p<.0001$), but it is also different in case of *empty* (bottle vs. paint tube: $p<.0001$); (iv) children's performance across lists is different from adults' only on *closed* evaluated with respect to the paint tube (*closed*-List 1: children vs. adults: $p<.0001$; all other comparisons between ages: all $ps>.05$, n.s.).

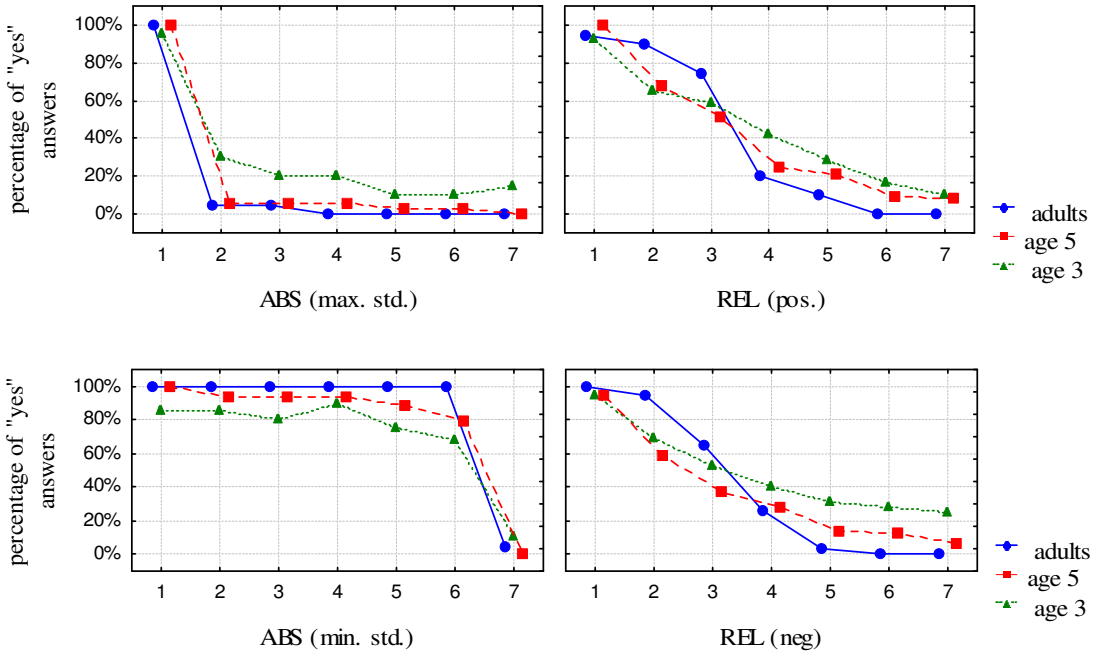

Fig. 6. Percentage of "yes" by adults (solid line), 5 year old children (dashed line) and 3 year old children (dotted line) in the SJT, differentiated between max std ABS/positive REL (top) and min std ABS/negative REL antonyms (bottom). Note that ABSC GAs are not included (cf. Fig. 7).

As it is evident from the graphs, the pattern of response differs between the two conditions (Relative vs. Absolute) for all age groups: participants are much more categorical in the Absolute GAs condition in general (left panels),[14] independently of age. In particular, both adults and children drastically drop their "yes" answers on item #2 in case of ABS max std but not in case of REL GAs, as expected and as already found by Syrett: in case of Relative GAs (right panels), adults' "yes" responses constantly decrease and drop below 50% acceptance only on the 4th item and the same pattern is observable in children, even if the curve is smoother for the 5 year old children, and even smoother for the 3 year olds. These observations were supported by statistical analysis.[15]

[14] They correctly answered "no" to the question "Is this clean?" to all items with at least one spot on it (items #2-7) and "yes" only to the completely clean item (item #1) – cf. Fig. 6, top-left panel; conversely, they correctly answered "yes" to the question "Is this dirty?" to all items #1-6 (ranging from having 1 to 6 spots) and answer "no" only to the 7th item with no spots on it – cf. Fig. 6, bottom-left panel.

[15] The analysis of variance ANOVA showed: (i) a significant main effect of class of Adjective (ABS vs REL; $F(1, 2468)=3.09$, $p<.0001$); a significant main effect of Age (3 vs. 5 vs. adults; $F(2, 2468)=4.66$, $p<.01$); a significant interaction between Age and class of Adjective ($F(2, 2468)=3.05$, $p<.05$. Post hoc analysis revealed that children (of both ages) and adults do not differ from each other with respect to the class of Absolute GAs (all ps>.05, n.s.); a difference is instead found for the class of Relative GAs between 3 year olds and 5 year olds and 3 year olds and adults (both ps<.01), but not between 5 year olds and adults (p=.09, n.s.). Including Item number (1 to 7) as an additional within factor to our analysis, we found a significant interaction of this with Age and class of Adjective ($F(12, 2468)=4.45$, $p<.0001$), but no interaction if age is also added ($F(12,2468)=.65$, $p=.80$, n.s.). All comparisons revealed a significant difference between REL and ABS on the proportion of "yes" answers to item #2 for each age group (all ps <.001).

As for the class of Absolute GAs that project on totally closed scales (ABSC GAs), we will limit our analysis to the pair *full*/*empty*, given that only in this case we tested subjects with different items (bottle vs. paint tubes) across lists. The percentage of "yes" answers for all age groups are plotted below, distinguishing between adjective (*full* in the two panels on the left, *empty* in the two on the right) and type of object used (bottles vs. paint tubes):

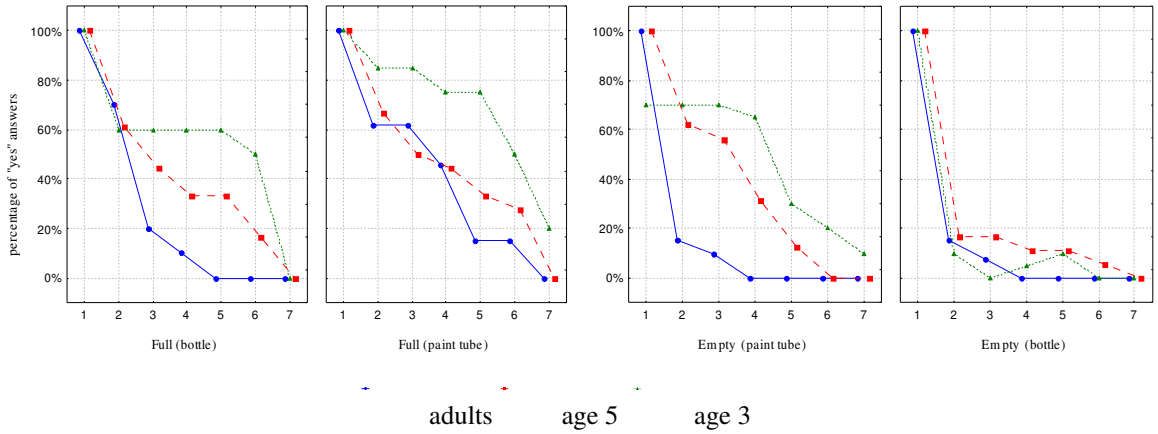

Fig. 7. The SJT: percentage of "yes" for items 1-7 by adults (solid line), 5 year old children (dotted line) and 3 year old children (dashed line), differentiated between "full" (left panels) and "empty" (right panels) and between lists/object used, as specified below each panel.

Assuming standard semantics, being both *full* and *empty* max std GAs, a shift to "no" was expected already on item #2, as found before for ABS max std like *clean*. As one can evince from the graph, in case of "empty": (i) adults pattern as expected in both conditions (i.e. they correctly answer "no" to the question "Is this *empty*?" to all items except the 1st one, with a drop of "yes" below 20% on item #2, as expected); (ii) children at age 3 and 5 pattern like adults only when *empty* is evaluated with respect to the set of bottles (cf. 4th panel), but not when it is evaluated with respect to the paint tubes, for which the trend resembles that observed for Relatives (more than 60% of children judged "empty" the almost empty paint tube, and only on the 4th item the rejection rate dropped below 40%, cf. 3rd panel). Unexpected results are also found in case of *full*, both across lists and ages (cf. left panels). In particular, 70% of the adults and about 60% of the children kept judging the almost full bottle in the series of 7 as "full" (cf. 1st panel from the left).[16] Another interesting observation is worth mentioning: 50% of the children at age 3 kept considering "full" all bottles or paint tubes up to item #6, converging on the answer "no" only on the completely empty object (item #7). We will come back to this result in the last section of this paper.

[16] Considering item #2 for max std/positive only, a 3 (Adjective: ABS vs. ABSC vs. REL) X 3 (Age) analysis of variance ANOVA revealed a significant main effect of Adjective ($F(2, 327)=45.95$, $p<.0001$) but no significant effect of Age ($F(2, 327)=.98$, $p=.38$, n.s.). The effect is due to the difference between ABSC and ABS, statistically significant for all age groups, as revealed by post hoc analysis (all $ps<.01$); more precisely, to the difference between the acceptance rate of "This is *full*" in case of item #2 (compared to #1) for all age groups (all $ps<.05$).

4. Conclusive Remarks

We conducted a two-part experiment in order to investigate whether Italian children are able to interpret correctly Relative GAs (that need to be contextually interpreted), and Absolute GAs, that are evaluated with respect to the minimum or maximum standard identified with the lower or upper boundary of the closed scale evoked by the adjective. Our overall results confirm Syrett (2007)'s findings that children are sensitive to this semantic distinction. Nevertheless, children's performance was perfectly in line with adults' behaviour for Relative GAs and for Absolute GAs projecting on scales with a single boundary (irrespectively of the fact that that boundary constitutes the minimum or the maximum standard for the adjective), whereas we obtained unexpected results within the class of Absolute GAs projecting on totally closed scales, not only for children, but also for adults. The only ABSC GA that patterns like the other ABS GAs is *open*, that is interpreted as a minimum standard GA. On the other hand, both adults and children seem to interpret differently the adjective *closed* when it describes the lid of a box (rejecting the description "This is *closed*" if the lid is minimally open, thus interpreting *closed* as a maximum standard GA) or the zip of a purse (accepting the same description referred to a not completely closed zip). In the case of *empty*, on the other hand, adults interpret it in accordance with the semantic predictions, always as a maximum standard adjective, both in the TVJT and in the SJT, whereas children differentiate between bottles and paint tubes: a minimally filled bottle is considered to be *not* empty, while paint tubes are judged to be empty even if there is some paint left – and nearly 30% of children continue to consider empty a half-way filled paint tube. Since these unexpected findings are obtained only when these adjectives are applied to particular items (purses, and not boxes, paint tubes for children, and not bottles), we propose to interpret these results as the application of a functional interpretation of the meaning of the adjective: the description "This is empty/closed" is probably interpreted as "This purse is closed enough for the coins not to fall out", and "This paint tube counts as empty enough for using the paint to draw something". Recall that the items we tested for the interpretation of the other adjectives were the most "abstract" we could find (especially for avoiding the reference to a normative class for Relative GAs), whereas in this case the purse and the paint tubes are more prone to a functional, "real life" interpretation. Subsequent work needs to be done to explore this possibility further.

In the case of *full*, on the other hand, the results were somehow more surprising. Summing up, we found that nearly all adults and children accepted the description "This is *full*" referred to a not completely filled bottle or paint tube in the TVJT. Moreover, in the SJT, when the targeted item was placed immediately after a completely full bottle, 60% of adults kept judging it full. Children's behaviour in this task was even more remarkable, with 30% of the 5 year olds judging a less than half-way full bottle as *full* (cf. the 5th bottle from the left in Fig. 3) and with 50% of the 3 year olds considering all the 6 bottles, up to the minimally filled one, to be *full*. Recall that Syrett obtained similar results with children's interpretation of *full* referred to a container filled with lentils, whereas the adults' performance was in line with what the semantics predicts. We think that there are in principle two different explanations that could be proposed to account for these facts. We could maintain that the semantics attributed to *full* is in fact correct: *full* is

a maximum standard adjective, i.e, an object can be judged as full only if it possesses the fullness property to a maximum degree. Nevertheless, pragmatic considerations could explain why an almost full bottle or paint tube counts as full: even if strictly speaking they should be regarded as not full, they are pragmatically viewed as being full. We are not convinced by this line of reasoning for mainly two reasons. It has been noticed (cf. Rotstein & Winter 2004) that other Absolute GAs give rise to a pragmatic/functional interpretation of what counts as the maximum standard: e.g., what counts as *clean* for a kitchen knife is a much looser standard than the one invoked for a surgery scalpel to count as clean. Nevertheless, all other Absolute GAs are interpreted as "real" maximum standard GAs, e.g. the standard of cleanness or straightness is never "loosened". Moreover, there is a clear developmental trend that can be recognized: as highlighted above, the younger children (aged 3) continue to judge the less than half-way filled bottles or paint tubes as being full, and older children (aged 5) start being more categorical, even if they still apply the description *full* to more items than adults. In other words, it seems that younger children interpret *full* not as a maximum standard adjective, but as a minimum standard GA: something counts as *full* as long as it has a minimal amount of fullness.

In fact, there is something peculiar about the Absolute GAs that project on totally closed scales: if – as in the case of *full*/*empty* – the pairs of antonyms are both interpreted as maximum standard GAs, the antonyms are not complementary anymore and a large portion of the scale of fullness/emptiness lacks a label and becomes "inexpressible": if *full* means "completely full" and *empty* means "completely empty", how should we describe any object that lies in between? We therefore propose that children start out by assigning to *full* a semantics analogous to that assigned to *open*, namely, by interpreting *full* as a minimum standard GA, and that only at a later stage in the development they re-adjust the semantics and converge on the correct one. Again, further research is needed in order to test this hypothesis.

References

Barner, David, and Jesse Snedeker. 2007. Compositionality and statistics in adjective acquisition: 4-year-olds interpret *tall* and *short* according to the typical sizes of novel noun referents. In Proceedings of the 31st Annual Boston University Conference on Language Development. Somerville, MA:Cascadilla Press.

Bartsch, Renate, and Theo Vennemann. 1972. The grammar of relative adjectives and comparison. Linguistische Berichte 20:19–32.

Bierwisch, Manfred. 1989. The Semantics of gradation. In Dimensional adjectives, eds. M. Bierwisch, & E. Lang. Berlin: Springer-Verlag.

Crain, Stephen, and Rosalind Thornton. 1998. Investigations in Universal Grammar: A Guide to Experiments in the Acquisition of Syntax and Semantics. Cambridge, MA: MIT Press.

Cresswell, Maxwell John. 1977. The semantics of degree. In Montague grammar, ed. Barbara Partee. New York: Academic Press.

Ebeling, Karen S., and Susan A. Gelman. 1988. Coordination of Size Standards by Young Children. Child Development 59.4:888-896.

Ebeling, Karen S., and Susan A. Gelman. 1994. Children's Use of Context in Interpreting "Big" and "Little". Child Development 65.4: 1178-1192.

Frazier, Lyn, Charles Jr. Clifton, and Britta Stolterfoht. 2008. Scale structure: Processing minimum standard and maximum standard scalar adjectives. Cognition 106.1:299-324.

Kennedy, Chris. 1999. Projecting the adjective: The syntax and semantics of gradability and comparison. New York: Garland. (1997 UCSC Ph.D thesis).

Kennedy, Chris, and Louise McNally. 2005. Scale structure and the semantic typology of gradable predicates. Language 81.2:345–381.

Gelman Susan A. and Karen S. Ebeling. 1989. Children's Use of Nonegocentric Standards in Judgments of Functional Size. Child Development 60.4: 920-932.

Paradis, Carita. 2001. Adjectives and boundedness. Cognitive Linguistics 12.1:47-65.

Rotstein, Carmen, and Yoad Winter. 2004. Total adjectives vs. partial adjectives: scale structure and higher order modifiers. Natural Language Semantics 12:259–288.

Stechow, Arnim von. 1984. Comparing semantic theories of comparison. Journal of Semantics 3:1–77.

Syrett, Kristen. 2007. Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives. Ph.D. thesis, Northwestern University, Evanston, IL.

Syrett, Kristen, Evan Bradley, Chris Kennedy, and Jeffrey Lidz. 2006. Shifting Standards: Children's Understanding of Gradable Adjectives. In Proceedings of the Inaugural GALANA. UCONN/MIT Working Papers in Linguistics, Cambridge (MA).

Department of Psychology
University of Milano-Bicocca
Piazza dell'Ateneo Nuovo, 1
20126 Milano
Italy

francesca.foppolo@unimib.it
francesca.panzeri@unimib.it