

ARTICLE

# Processing adjectives in development: Evidence from eye-tracking

Michela REDOLFI  and Chiara MELLONI

University of Verona

**Corresponding author:** Michela Redolfi; Email: [michela.redolfi@univr.it](mailto:michela.redolfi@univr.it)

(Received 04 October 2022; revised 25 September 2023; accepted 25 November 2023)

## Abstract

Combining adjective meaning with the modified noun is particularly challenging for children under three years. Previous research suggests that in processing noun-adjective phrases children may over-rely on noun information, delaying or omitting adjective interpretation. However, the question of whether this difficulty is modulated by semantic differences among (subsective) adjectives is underinvestigated.

A visual-world experiment explores how Italian-learning children (N=38, 2;4–5;3) process noun-adjective phrases and whether their processing strategies adapt based on the adjective class. Our investigation substantiates the proficient integration of noun and adjective semantics by children. Nevertheless, aligning with previous research, a notable asymmetry is evident in the interpretation of nouns and adjectives, the latter being integrated more slowly. Remarkably, by testing toddlers across a wide age range, we observe a developmental trajectory in processing, supporting a continuity approach to children's development. Moreover, we reveal that children exhibit sensitivity to the distinct interpretations associated with each subsective adjective.

**Keywords:** adjectives; eye-tracking; Italian; processing; language development

## Introduction

The acquisition of adjectives is particularly challenging for young children for several reasons and has been observed to emerge later compared to other word classes, particularly nouns. Many explanations have been proposed to account for this slower path of acquisition, most of which concern the conceptual and distributional properties of adjectives. First, understanding the meanings of adjectives depends on linguistic knowledge to a greater extent than is the case for concrete nouns, the latter mostly being conceptually basic, perceptually coherent and, hence, easily individuated even by infants (Gentner, 1982; Gentner & Boroditsky, 2001; Markman, 1989; Syrett *et al.*, 2010). In contrast, the denotation of property words like *big* and *good* can vary considerably depending on the different nouns they modify. According to this view, the so-called RELATIONAL RELATIVITY, adjectives, verbs, prepositions, and other relational terms are, with respect to concrete nouns, cross-linguistically more variable in the way they map

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

from concept to word. Second, while nouns are used to label categories of objects sharing many correlated features, adjectives have far fewer shared correlated features, as they shift semantically in different linguistic contexts (Gentner, 1982). In addition, adjectives make up around 10% of tokens in child-directed speech, resulting in a lower proportion of the input compared to nouns and verbs (Sandhofer *et al.*, 2000).

Many developmental studies have investigated at what age and under what conditions children can correctly identify the property indicated by the adjective, monitoring their offline performance, *i.e.*, after the adjective has been presented. These studies found that at 21 months children extend adjectives to other members of the basic category they are referred to (Waxman & Booth, 2003), but it is not until 36 months of age that the ability to extend adjectives to objects of other categories emerges (*e.g.*, Klibanoff & Waxman, 2000; Waxman & Klibanoff, 2000). These studies, however, could only provide a limited picture, as the processes involved in learning novel adjectives and mapping a single word to a relevant property may significantly differ from the processes involved in the rapid interpretation of familiar noun-adjective phrases. Crucially, children's difficulty in the interpretation of noun-adjective combinations has rarely been framed in terms of processing costs and challenges at the level of conceptual integration. This study specifically addresses this issue, focusing on children's real-time processing of adjectives combined with nouns in continuous speech, and examining the interpretation of different semantic classes of adjectives.

### Processing attributive adjectives

The interpretation of attributive adjectives (*i.e.*, adjectives that are immediately adjacent to the noun they modify) seems to be particularly problematic for young children. Ninio (2004) proposed that the difficulty children show in processing noun-adjective combinations depends on the computational complexity of the integration of noun and adjective meanings. Using a picture-pointing task, Ninio (2004) asked Hebrew-learning toddlers to point to a *big teddy* in a set of four items that crossed the relevant property with the object kind (*e.g.*, big and small fishes and teddies). Although the children were fairly accurate in identifying the correct picture, a more detailed analysis of first responses and self-corrections showed that, when making a mistake, children always selected the other object (*e.g.*, the other teddy) and not the other property (*e.g.*, the other big object). To account for these results, Ninio proposed that interpreting noun-adjective combinations requires a two-step process: first, the identification of the set of objects labeled by the noun, and second, the subcategorization of that set containing those objects that also possess the property labeled by the adjective. In other words, interpreting the phrase *big teddy* implies identifying the set of teddies in the discourse context, and then the subset of the teddies that are big. While adults perform this two-step procedure instantly and without awareness, the same process seems to be particularly challenging for toddlers younger than 36 months. Before this age, children adopt a NOUN ANCHORING<sup>1</sup> mechanism by focusing on the identification of a suitable referent for the noun and omitting or delaying the second step of adjectival meaning integration, performing an OVERT two-step procedure.

Although Ninio's (2004) experimental findings were very clear, her research could not account for cross-linguistic variation concerning the adjective position in

<sup>1</sup>Throughout the present paper, we adopt the term NOUN ANCHOR(-ING) from Weisleder and Fernald (2009) to refer to the hypothesis on N-A combination processing proposed by Ninio (2004).

attributive use, as the tested language was Hebrew, which has post-nominal adjectives. The sequential steps of interpretation she proposes are aligned with the order of nouns and adjectives that are actually heard in the tested language. To explore this issue, Thorpe *et al.* (2006) conducted a replication of Ninio (2004)'s study on English, exploiting the same off-line methodology. Their results were in line with those found in Hebrew as the youngest children showed the same noun-biased error pattern of the original study. Both studies, however, used a picture-pointing task, which only allows measuring accuracy but cannot provide any information about the real-time processing of noun-adjective combinations.

Tribushinina and Mak (2016) examined the incremental interpretation of adjective-noun phrases in an online eye-tracking study. 3-year-old Dutch-speaking children were presented with two pictures on the screen (e.g., a grey stone and a grey butterfly) while hearing an adjective-noun combination where the adjective was either informative or uninformative about the following noun. For instance, in a visual context depicting a stone and a butterfly, the adjective 'heavy' is informative, i.e., it can be used to 'predict' the following noun (as in the heavy stone). On the other hand, if both pictures are grey, the adjective grey is uninformative in the same context (as in the grey stone). All informative adjectives were stereotypical properties of objects and were not depicted on the screen. Unsurprisingly, when the adjective was uninformative (e.g., *grey*), looks to the target object increased only upon hearing the noun. By contrast, when the adjective was informative (e.g., *heavy*), children showed a preference for the referent object while hearing the adjective, suggesting early integration of adjective semantics with world knowledge. To resolve the task, however, children could have attended to the adjective using conceptual knowledge of the target object, i.e., knowing that a stone is typically heavy whereas a competitor (e.g., a butterfly) is not. Adjective-noun co-occurrence statistics are also likely to have influenced the early-looking behavior. Although Tribushinina and Mak (2016) concluded that their results discard the hypothesis that 3-year-olds adopt a NOUN ANCHORING mechanism in interpreting such combinations, we believe that they did not directly demonstrate adjective-noun integration and, crucially, they only tested children older than 36 months, which is claimed to be the cut-off age for mastering attribution.

Leaving aside the noun-first / adjective-later pattern expected by the two-step process proposed by Ninio (2004), a further prediction deriving from the NOUN ANCHOR HYPOTHESIS mechanism is that postnominal frames result in more efficient processing than prenominal frames. For speakers of a language with adjective-noun phrases, the process could involve additional processing costs, requiring holding the prenominal adjective in memory, waiting until the noun before beginning the interpretation of the whole combination, and then retrieving the adjective for the identification of the proper referent (Weisleder & Fernald, 2009).

Fernald *et al.* (2010) partially addressed this issue by designing a looking-while-listening task. English-learning children were presented with two pictures while listening to familiar adjective-noun phrases (e.g., *Where is the blue car?*) in a visual context where the color-adjective was either informative (e.g., a blue car paired with a red car) or uninformative (e.g., a blue car paired with a blue house) about the referent object. The analysis of participants' gaze pattern revealed that, while 30-month-old children failed to take advantage of the prenominal color-word to identify the target even if it was informative, 36-month-olds were more successful and showed an impressive gain in the ability to integrate nouns and adjectives in combination in real time. The authors concluded that children develop the skill to process color adjectives and nouns in

combination over the third year and that it is possible that, returning to Ninio (2004)'s proposal, 30-month-olds waited for the noun before interpreting the property term. Since the noun could refer to either of the two objects on the screen, children looked back and forth between the pictures until they remembered the adjective, which finally allowed them to identify the target.

The same experiment was conducted by Weisleder and Fernald (2009) with Spanish-speaking children, aiming at investigating whether children's processing of adjective-noun phrases (in English) vs. noun-adjective phrases (in Spanish) is affected by the order in which the relevant information is heard and by cross-linguistic differences. They found that 36- to 40-month-old Spanish-learning children interpreted the speech stimuli incrementally. When the noun was informative (e.g., *carro*, 'car', in a blue-car/blue-house visual condition), children were able to rapidly identify the target upon hearing it. When only the adjective was informative (e.g., *azul*, 'blue', in a blue-car/red-car condition), children listened through the whole combination and shifted their gaze to the target after the adjective word. Thus, in comparing these results with those found for English learners, the authors concluded that children's processing of phrases with color adjectives is facilitated when the noun is heard before the adjective.

To test children's integration of referring expressions in pre-nominal vs. post-nominal adjective conditions, Davies *et al.* (2021) designed an eye-tracking study with English-speaking 3-year-olds. Children were presented with four black-and-white pictures on the screen (e.g., a big cow, a small cow, a big flower and a small tree) and eye movements were recorded while listening to a question in two syntactic conditions. In the pre-nominal condition, children heard an adjective-noun combination (e.g., *Where is the big cow?*), whereas in the post-nominal condition, the adjective was uttered in a relative clause (e.g., *Where is the cow that is big?*). Interestingly, they found that, like adults, children processed modified noun phrases equally quickly regardless of adjective position. We believe this finding is more robust than previous research in which experimental paradigms meant that the referential task could be passed using adjective information alone (Fernald *et al.*, 2010; Thorpe *et al.*, 2006) or world knowledge (Tribushinina & Mak, 2016). However, as Davies *et al.* (2021) underline, the use of relative adjectives only (such as *big* and *small*) limited the outcome of the experiment. Moreover, they only tested 3-year-olds and, consequently, nothing can be said about younger children's online processing of such combinations.

Collectively, existing research on children's processing of attributive adjectives is scarce, the methods are inconsistent, and further investigation is needed. To robustly test children's integration of nouns and adjectives, the current study requires participants to interpret both words in the combination, including both noun and adjective competitors in the same visual display. We are interested in children's looking behavior during the integration of noun and adjective meanings, in terms of over-reliance on noun information, developmental patterns of interpretation and computational differences among three classes of substantive adjectives.

### The interpretation of substantive adjectives

Formal semantic accounts involving adjectives point to the difficulty of providing a characterization of adjectival meaning. Here, we adopt Kamp and Partee's (1995) classification of adjectives into two broad categories (i.e., substantive and non-substantive adjectives), modeled in set-theoretic terms and based on the inferences that a noun-

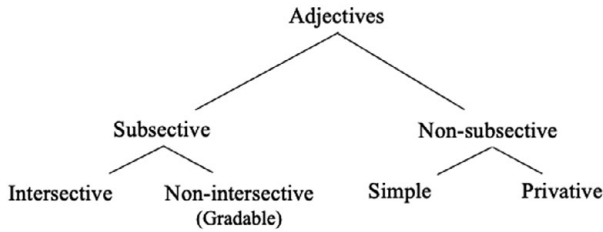


Figure 1. Kamp and Partee (1995)'s classification of adjectives.

adjective combination (or intersection) triggers.<sup>2</sup> Figure 1 reports a summary of this classification<sup>3</sup>.

Among the so-called subjective adjectives, the simplest category includes adjectives with the most stable core meanings, referred to as the class of INTERSECTIVE ADJECTIVES. Under a classical model-theoretic view of semantic representation, the meaning of such adjectives can be identified with the set of entities that bear the property encoded by the adjective. Thus, the meaning of the word *red* corresponds to the entities that have the property of being red. When combined with nouns, it is easy to see how simple expressions (e.g., *red sweaters*) can compose into more complex ones, and the meaning of this combination is simply the intersection between the set of objects labeled by the noun (e.g., objects that are sweaters) and the set of properties labeled by the adjective (e.g., objects that are red). Although various studies showed that color adjectives can be context-dependent to different degrees (see, e.g., Kennedy & McNally, 2010), this type of adjectives is usually interpretable independently of the context and we assume does not evoke a contextually salient comparison class in their interpretation.

Many adjectives, however, do not conform to the simple compositional analysis described above. Treating noun-adjective combinations as an intersection of predicates only works for those adjectives with an independent and stable core meaning. Various adjectives appear to lack an invariant meaning, as they depend upon the noun they modify and can only be evaluated with respect to the set denoted by the head noun. Thus, the set denoted by the noun-adjective combination is necessarily a subset of the set referred to by the noun. Non-intersective adjectives (or gradable adjectives; Kennedy,

<sup>2</sup>This work investigates the interpretation of subjective adjectives only. For clarity and completeness, non-subjective adjectives include:

(i) “simple” non-subjective adjectives, whose combination with a noun implies neither the adjective nor the noun:

X is Adj N  $\not\Rightarrow$  X is a N    X is an alleged murderer  $\not\Rightarrow$  X is a murderer

X is Adj N  $\not\Rightarrow$  X is Adj    X is an alleged murderer  $\not\Rightarrow$  \*X is alleged

(ii) “privative” non-subjective adjectives, which, in a noun phrase, imply a negative inference for the noun; that is, this combination is never an instance of the noun alone:

X is Adj N  $\Rightarrow$  X is not a N    X is a fake diamond  $\Rightarrow$  X is not a diamond

X is Adj N  $\not\Rightarrow$  X is Adj    X is a fake diamond  $\not\Rightarrow$  X is fake

<sup>3</sup>Note that the adjectives in Kamp and Partee’s (1995) classification can be used in attributive and predicative structures, except for simple non-subjective adjectives, which cannot appear in predicative construction (e.g., *alleged*, \**The murder is alleged*).

1999) have context-sensitive meanings (e.g., Kennedy, 2007; Kennedy & McNally, 2005; McNally, 2011) and map the object they refer to onto a scale of fully ordered degrees (Kennedy, 1999, 2007; Kennedy & McNally, 2005). According to how context influences their meaning and according to the structural features of the scales, a further distinction needs to be made between RELATIVE ADJECTIVES and ABSOLUTE ADJECTIVES.

RELATIVE ADJECTIVES (e.g., *big*, *long* or *old*) are context-sensitive in the sense that the context determines what specific value is required to count as e.g., big, long or old and the standard of comparison is located around the midpoint of the scale. Moreover, relative adjectives are mapped onto “open” scales, i.e., with neither a lower nor an upper boundary. By contrast, for absolute adjectives like *clean* or *empty*, the context determines how much deviation from e.g., total cleanness or emptiness is allowed to count as clean or empty. They evoke a scale that is “closed” at one end (e.g., *clean/dirty*) or both ends (*empty/full*). For example, a particular cloth can be dirty with no conceivable limit of dirtiness, but there is a standard of cleanness that cannot be overcome; however, a particular glass can be “fuller” than another, but there is a limit in both directions because a glass cannot overcome what counts as completely full or completely empty. Both the context and the presence/absence of boundaries on a scale determine how gradable adjectives are interpreted. Hence, relative adjectives, such as *big*, need to be interpreted according to a standard of comparison that is contextually fixed: an object counts as *big* only relative to a standard of size that can be contextually retrieved. ABSOLUTE ADJECTIVES, on the other hand, evoke a bounded scale, and that boundary serves as the standard: an object counts as clean only if it possesses the maximum standard of cleanness. If the gradable adjective has two boundaries (e.g., *empty*), both ends could constitute the standard.

Various experimental studies investigated the interpretation of these adjective classes by children and adults, mainly using offline methodologies (e.g., Foppolo & Panzeri, 2013; Syrett, 2007; Syrett *et al.*, 2006) and focusing on adult processing (e.g., Aparicio *et al.*, 2016; Sedivy *et al.*, 1999). Most of the studies conducted with children made use of the Scalar Judgment Task, presenting 3- to 5-year-old participants with a series of seven objects displaying different degrees of the same property. Children were asked to judge whether the property labeled by the adjective in question was true for each of the seven objects (‘Is this ADJ.?’; e.g., *Is this full?*). Results from different languages confirmed that relative and absolute adjectives are interpreted in different ways, i.e., the cut-off point was identified in the middle of the scale for relative adjectives (Syrett, 2007 – for English; Foppolo & Panzeri, 2013 – for Italian; Tribushinina, 2013 – for Dutch; Weicker & Schulz, 2018 – for German) and on one boundary for absolute adjectives (Foppolo & Panzeri, 2013; Syrett, 2007; Weicker & Schulz, 2018). Although offline tasks rely on ‘end-point’ data and do not examine real-time processing, the findings reported in these studies interestingly suggested that absolute adjectives share properties with both relative and intersective adjectives. Like relative adjectives, absolute adjectives are, to different degrees, context-sensitive. On the other hand, similarly to intersective adjectives, objects described by absolute adjectives are always judged as either possessing the property or not and, unlike objects described by a relative adjective which are in the middle of a scale, are never vague.

A limited number of studies have investigated the impact of semantic distinctions on the online interpretation of these adjective types, and, to the best of our knowledge, none has examined children’s processing in this regard. Sedivy *et al.* (1999) conducted an eye-tracking study to investigate how the presence of a comparison class in the visual display influences the adult processing of sentences containing relative adjectives. Participants

heard a sentence such as *Pick up the tall glass* while presented with four objects simultaneously in two visual conditions. In the Contrast condition, the visual scene displayed the target (e.g., a tall glass), an object-competitor (e.g., a short glass), a property-competitor (i.e., another tall object) and a distractor (i.e., an object that could be described neither by the noun nor by the adjective). In the No-contrast condition, they did not include the object-competitor but added another distractor. The main finding was that participants' fixations converged on the target faster in the Contrast condition than in the No-Contrast condition. Crucially, participants zoomed into the target object during the adjective-window when the head noun had not been processed yet, suggesting that the information about the contrasting object was used very quickly, already at the point in which the linguistic instruction was still compatible with both the target and the property-competitor. Although in the same article Sedivy *et al.* (1999) reported a similar study with intersective adjectives (e.g., colors and shapes), they did not provide any comparison between adjective classes.

In another eye-tracking study, Aparicio *et al.* (2016) investigated adults' interpretation of intersective, relative and absolute adjectives, used as restrictive modifiers in pre-nominal position. In an experimental design similar to Sedivy *et al.* (1999), they exploited intersective adjectives as a baseline and found that when the visual context supports a restrictive interpretation of the adjective (i.e., in the Contrast condition), target identification is faster for both relative and absolute adjectives. However, for absolute adjectives, this effect is significantly delayed. They argued that target identification for relative adjectives was faster because the contrasting object provided a contextually salient comparison class and facilitated lexical-semantic processing. However, they suggested that this effect of contrast appeared later for absolute adjectives because participants were asked to commit to a precise interpretation of predicates when presented with endpoint-oriented absolute adjectives (e.g., *clean*). They argued that committing to this precise interpretation might be costlier than committing to the imprecise interpretation of absolute adjectives involving one open boundary (e.g., *dirty*) and, hence, of relative adjectives (which have two open boundaries by definition).

To date, studies on adjective interpretation have addressed various issues concerning incrementality, semantic differences and processing costs, but none has explored these aspects from the developmental point of view and in real time. Building upon these results, the present study takes on a new approach by investigating the early development of children's ability to interpret familiar adjectives in online sentence comprehension using eye-tracking. Children and adult participants were tested in Italian, a language with post-nominal adjectives that allowed us to test a) children's use of the NOUN ANCHORING mechanism; b) potential differences in the interpretation of different semantic classes of adjectives (i.e., intersective vs. relative vs. absolute adjectives) in real time; c) differences between children and adults at the level of speed, accuracy and processing costs; and d) the emergence of adult-like patterns of interpretation in development.

### The current study

The goal of the present study is to investigate how young toddlers between two and five years of age process Italian noun-adjective combinations online and how they distinguish between different semantic types of adjectives. To this end, we created a Visual World task with eye-recording to examine the time course of sentence comprehension during the processing of questions containing noun-adjective phrases. Participants' eye movements



were recorded while looking at four pictures on the screen of two objects crossed by two properties (e.g., a black shoe, a white shoe, a black sock and a white sock), while each stimulus sentence unfolded (i.e., “Where is the Noun-Adj?”, e.g., *Dov’è la scarpa nera?*, lit. ‘Where is the shoe black?’). Participants were tested in three adjective-conditions – namely, intersective (e.g., *black*), relative (e.g., *big*) and absolute (e.g., *closed*).

The experiment addresses three research questions, formulated to give a comprehensive account of noun-adjective integration and adjective semantics interpretation in young children and adults.

RQ1. Is there evidence of a NOUN ANCHORING mechanism in children’s online processing of Italian noun-adjective combinations? Do children focus on the interpretation of the noun-word while omitting or delaying the integration of the adjective meaning? Do children and adults differ in processing noun-adjective combinations?

We predicted that all children would interpret the noun-word as soon as they hear it, by rapidly and accurately identifying the object referent. However, upon hearing the adjective, we expected the youngest children to be slower in the interpretation, showing target preference only after sentence offset or to even leave out reference resolution, i.e., to not be able to discard the object competitor and to keep looking back and forth between the two category competitors.

In general, we expected significant differences between children and adults regarding speed and accuracy as a result of additional processing costs for children. We predicted adults to be overall faster than children in object identification first (i.e., at the noun), and in shifts to the target later (i.e., at the adjective), both within the time-windows in which the disambiguating word was spoken (i.e., the noun-window and the adjective-window respectively). Furthermore, since this experiment was specifically designed for young children, we predicted adults’ accuracy to be at ceiling in all trials, unlike for children, whom we expected to be overall less accurate.

RQ2. Are there differences in the way children interpret different semantic classes of adjectives in real time? Are children sensitive to contextual cues affecting the interpretation of each adjective class? Does adults’ processing vary across the different classes of adjectives?

We expected children to be faster in target identification with intersective adjectives (e.g., *black*). Once identified, e.g., the black shoe, participants did not need to look for the other shoe to verify its blackness. However, we expected this to happen with relative adjectives (e.g., *big*). Since relative adjectives are context-dependent, the identification of the target (e.g., the big teddy) was only possible after a comparison between the members of the object class identified by the noun (e.g., the big teddy and the small teddy). Hence, we predicted this comparison to take time, resulting in a delayed gaze shift to the described object. As for absolute adjectives (e.g., *closed*), two hypotheses drove this analysis. As for intersective adjectives, objects described with an absolute adjective could be judged as either possessing the property or not (e.g., participants can say if a shirt is dirty or not, without the need to compare it with the clean shirt). However, like relative adjectives, absolute adjectives were to some degree vague and context-sensitive (e.g., a shirt can be judged as dirty, but less dirty than another one). Any difference in looking behavior between absolute adjectives and the other two classes would suggest that children are sensitive to the differences in their semantic interpretation, thus providing experimental evidence that this knowledge is acquired very early in development. Finally, since the task was designed in such a way that a (young) participant has sufficient time to look at the four objects on the screen prior to the start of the auditory stimulus, we did not



expect adults to perform a comparison between members of the noun category in the interpretation of relative (nor absolute) adjectives.

RQ3. Does children's performance vary as a function of age?

We hypothesized older children show more accurate and faster integration of adjective meaning in Noun-Adjective combinations. Since the age range here considered is wide, and spans from 2;4 years to 5;4 years, we expected age to be a relevant factor.

## Methods

### Materials

All items of the visual-world task consisted of a visual stimulus paired with an auditory stimulus.

The visual stimuli were four digitalized colored drawings on a white background simultaneously presented on the screen. The four pictures consisted of two objects or animals crossed with two attributes in a two-by-two design, e.g., a black shoe, a white shoe, a black sock and a white sock (see Figure 2a). The position of the objects on the screen was pseudo-randomized and counterbalanced across trials.

The auditory stimuli were recorded by an Italian female native speaker reading with pragmatically neutral intonation (processed in Praat version 6.0.49; Boersma & Weenink, 2016). All stimuli were comparable in duration ( $M = 1660\text{ms}$ ). The stimuli consisted of a question about one of the pictures, e.g., *Dov'è la scarpa nera?* ('Where is the black shoe?'), containing the carrier phrase ('Where is the'), a noun, and a post-nominal adjective. The visual display contained: 1) the **target**, identified by the mentioned object and property (e.g., a black shoe); 2) an **object-competitor** that belonged to the same object class as the target but could not be described by the mentioned adjective (e.g., a white shoe); 3) a **property-competitor** that shared the target property, but belonged to a different object class (e.g., a black sock); and 4) a **distractor** that did not belong to the target object class, nor could be described by the property in the sentence (e.g., a white sock). Items in the visual display were matched for grammatical gender. Half of the items were feminine, and half were masculine. In addition, according to the adjective type, stimuli were divided into 3 adjective-conditions – namely, intersective adjectives (INT), relative adjectives (REL) and absolute adjectives (ABS) (see Figure 1). Each participant was presented with 4 trials for each adjective-condition in a randomized order.

Visual and auditory stimuli were combined to form 48 trials divided into 16 lists. All stimuli are listed in Table 1.

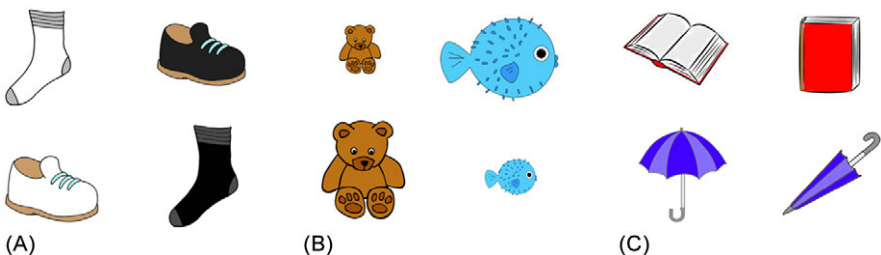


Figure 2. Example of visual stimuli: intersective (A), relative (B) and absolute (C) adjective-condition.

**Table 1.** List of items used in the experiment divided by adjective type

Adjective type		Nouns	Adjectives		
1	INT	<i>birillo, cappello</i>	'pin', 'hat'	<i>rosso, verde</i>	'red', 'green'
2		<i>calza, scarpa</i>	'sock', 'shoe'	<i>bianca, nera</i>	'white', 'black'
3		<i>cucchiaio, fiore</i>	'spoon', 'flower'	<i>giallo, viola</i>	'yellow', 'purple'
4		<i>macchina, palla</i>	'car', 'ball'	<i>blu, rosa</i>	'blue', 'pink'
5	REL	<i>albero, tavolo</i>	'tree', 'table'	<i>alto, basso</i>	'high', 'short'
6		<i>bambina, gallina</i>	'girl', 'hen'	<i>grassa, magra</i>	'fat', 'thin'
7		<i>matita, sciarpa</i>	'pencil', 'scarf'	<i>corto, lungo</i>	'short', 'long'
8		<i>pesce, orso</i>	'fish', 'bear'	<i>grande, piccolo</i>	'big', 'small'
9	ABS	<i>bicchiere, piatto</i>	'glass', 'plate'	<i>pieno, vuoto</i>	'full', 'empty'
10		<i>lampadina, torcia</i>	'lightbulb', 'flashlight'	<i>acceso, spento</i>	'bright', 'dim'
11		<i>libro, ombrello</i>	'book', 'umbrella'	<i>aperto, chiuso</i>	'open', 'closed'
12		<i>maglietta, mano</i>	't-shirt', 'hand'	<i>pulito, sporco</i>	'clean', 'dirty'

### Participants

Thirty-eight typically developing Italian children (2;4-5;3, mean age = 3;6, S.D. = 0;7) participated in the experiment; 20 were female. Children's age was evenly distributed. 11 children were younger than 3 years of age; 20 toddlers span between 3;0 and 4;0 years; 7 children were older than 4 years. Three additional participants were excluded due to fussiness or inattentiveness during testing (i.e., failure to look at the four pictures on more than half of the trials). Caregivers gave complete written and informed consent and were asked to complete a questionnaire indicating whether the child knew and understood the words in the experiment. The checklist included a total of 24 nouns and 24 adjectives. According to the parents' responses, child participants understood an average of 99.3% of the words. Families received 5 euros for participation, while children received a book of their choice. Twenty-four Italian adults (19;1-29;9, mean age = 25;4, S.D. = 2;6) served as controls; 17 were female. They had normal or corrected to normal vision by means of glasses or soft contact lenses. None of the adult participants had reported history of speech, hearing or language disorders. They gave written and informed consent and were paid 3 euros for participation.

The study was approved by the local Ethics Committee of the University of Verona, and was conducted in accordance with the standards specified in the 2013 Declaration of Helsinki.

### Procedure

Participants were tested individually in a dimly lit and soundproof testing room at the University of (removed for anonymization). Participants' eye movements were recorded

using an SR Research EyeLink 1000 Plus eye-tracker, which was in head-free remote mode and sampled monocularly at 500 Hz with a 16mm lens. The experiment was run on a computer connected to a 24" colour BenQ monitor for visual stimulus presentation. Speech stimuli were played over two loudspeakers on both sides of the screen. The experimental procedures were implemented in Experiment Builder and eye-movement data were extracted through Data Viewer. A small sticker on the participant's forehead tracked head movements. Calibration and validation procedures were carried out using a five-point display at the beginning of the experiment and a drift correction was repeated once every three trials.

The adult participants were seated on a chair in front of the screen and were told that they would participate in an experiment created for children and asked to follow the visual and auditory stimuli. The child participants sat on the caregiver's lap in front of the screen. They were told that they would play a game with a cartoon character (i.e., Peppa Pig), whose voice would tell them what to do and whose picture would appear on the screen every now and then, i.e., for calibration, validation and drift correction. Before the experiment began, a 5-point calibration and validation were performed, followed by a familiarization phase. Participants were shown 8 "warm-ups" in which a single image was labeled by a sentence played over speakers (e.g., "Look! A butterfly!") and appeared on one of the four quadrants of the screen, familiarizing the child with the four positions of the pictures on the screen. After the familiarization phase, the experimental session began. A drift correction was performed every three trials. The fixation dot for the drift correction appeared in the shape of Peppa Pig and was paired with one of the three filler sentences (e.g., "Are you having fun?"), recorded in a child-friendly intonation. The testing session lasted approximately 5 minutes. As exemplified in Figure 3, in each trial the four pictures appeared on the screen simultaneously. The auditory stimulus started once the child fixated all four pictures on the screen (i.e., it was gaze-contingent). If the child did not fixate all four pictures, the auditory stimulus started 5 seconds after picture display. From

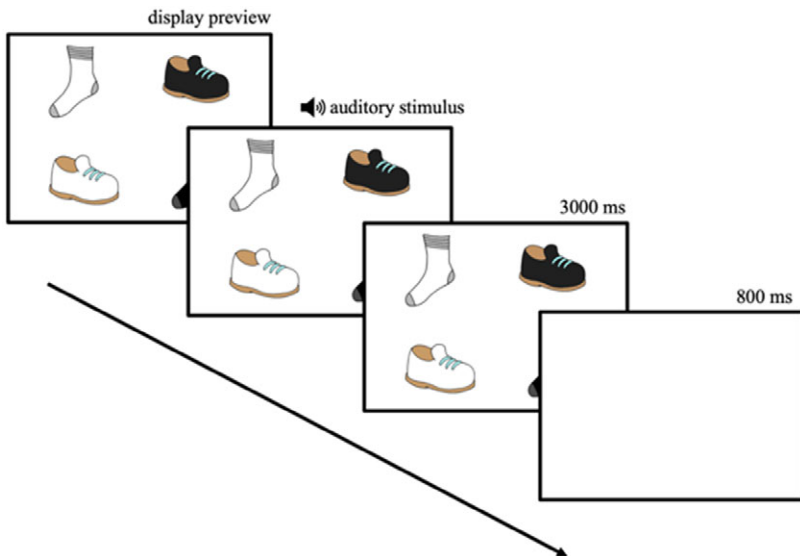


Figure 3. The trial sequence.

sentence onset, pictures remained displayed on the screen for 3000ms, followed by an 800ms blank screen that ended the trial.

## Results

Eye-movement data were prepared for the statistical analysis in R (R Core Team, 2018). Data were analyzed using the packages *itsadug* and *gamm4* (bam, mgcv, compareML, fvisgam functions, Wood, 2006, 2011). The packages *ggplot2* (Wickham, 2016) and *tidymv* (plot\_smooths and plot\_difference functions, Coretta, 2020) were employed for data visualization. Prior to analysis, we excluded the trials in which the target picture was never fixated, neither before nor after the auditory stimulus was displayed. This led to the removal of 51 trials, corresponding to 6.8% of the data. In addition, for each participant, we removed all trials in which none of the four pictures was fixated for more than the 50% of the duration, leading to the removal of another 52 trials corresponding to 6.9% of the data.

From the eye-tracking record, we determined gaze position in 50ms steps. For each 50ms-bin, we aggregated raw proportions of looks to the four objects on the screen (target, object-competitor, property-competitor and distractor), filtering out blinks and looks outside the interest areas. Proportion of looking time at the named target picture was accessed over three time intervals in the speech stimulus. First, the noun-window corresponded to the mean duration of the noun, starting from the noun onset (1000ms-1400ms). Second, the adjective-window corresponded to the mean duration of the adjective, starting from the average adjective onset (1450ms-1800ms). Finally, the post noun-phrase-window captured fixations during a 700ms window after the average offset of the adjectives (1850ms-2550ms).

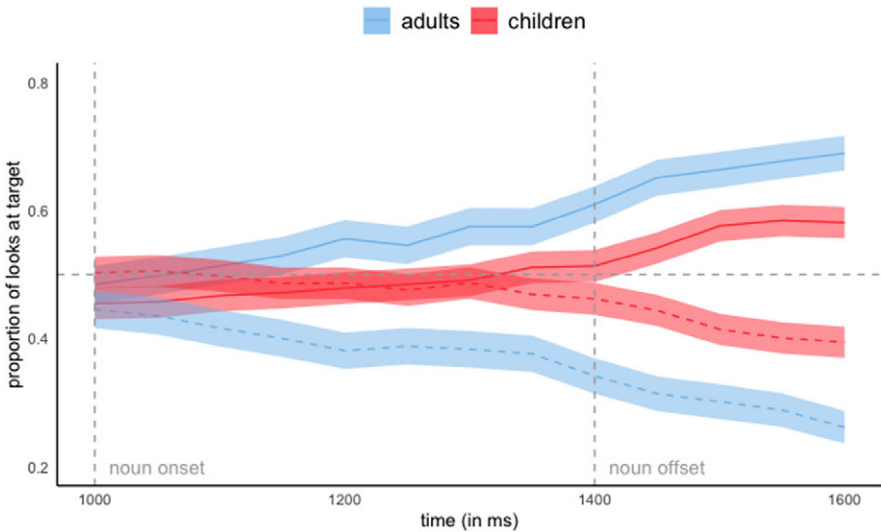
### *Integrating noun and adjective meanings*

To answer RQ1, we performed a statistical analysis to investigate a) when participants start to interpret the noun and when they integrate its meaning with that of the adjective (to assess potential effects of noun anchoring) and b) group differences in this respect. Specifically, this analysis examines fixations to the object-referents (e.g., the shoes) compared to the object-competitors (e.g., the socks) over the noun-window, and to the target (e.g., the shoe that is black) compared to the competitor (e.g., the shoe that is not black) during the adjective-window. The rationale behind this analysis is that, if a NOUN ANCHORING mechanism is in place, an asymmetry in processing time between noun and adjective interpretation would emerge, with adjectives taking more time.

We started by examining how quickly participants started to fixate significantly more on those objects that could be labeled by the noun (i.e., the target and the object-competitor), as opposed to those objects that could not (i.e., the property-competitor and the distractor). Following Aparicio *et al.* (2016), we collapsed fixations to *target & object-competitor* on the one hand (e.g., looks at the black shoe and looks at the other shoe), and fixations to the *property-competitor & distractor* on the other (e.g., looks at the other black object and looks at the distractor).

Before presenting the analyses, we note that visual inspection of Figure 4 offers some preliminary insights.

During the noun-window, the visual identification of the noun category (looks at target and object-competitor) is faster for adults than for children, the latter decreasing



**Figure 4.** Proportions of fixations to target/object-competitor (straight line) divided by property-competitor/distractor (dotted line) over time for children (in red) and adults (in blue). The first vertical line indicates noun onset, the second vertical line indicates the average noun offset. Confidence bands show  $\pm 1$  standard error of participant means. The horizontal line indicates the chance level.

their looks to the property-competitor/distractor only by the noun offset. To investigate when in time the two groups started to significantly differ in fixations to the labeled objects, we used a general additive model in R (GAM), whose visual representation indicates when in time the effect of group becomes significant on a response variable (in this case, looks to the labeled objects). We ran two models<sup>4</sup>. The baseline model included time as smooth term and event (i.e., the combination of participant and item as a unique identifier) as random effect (see e.g., Porretta *et al.*, 2016; Zahner *et al.*, 2019). The second model also included group as a categorical variable transformed into an ordered factor. Moreover, treatment contrast coding was set for group. A model comparison was run with the compareML function to assess the fit and differences between the two generalized additive models. The model comparison results indicated a significant improvement in fit for the full model compared to the null model ( $p < .001$ ), see Figure 5 (the output of GAM is only meaningful when visualized).

The graph in Figure 6 shows that adults' looks at the target & object-competitor are significantly higher than children's from about 150ms after sentence offset (when the two confidence intervals stop overlapping). However, while adults at that point look at the labeled objects above chance level, children do so only upon the end of the noun-window (i.e., at around 1400ms).

To investigate when participants start interpreting the adjective-word and integrate its meaning with the noun-word, a second analysis was performed on the adjective-window, when disambiguating the target was possible by discarding the object-competitor.

We ran a GAM analysis, specifying looks to the target as dependent variable, *group* as predictor, and *time* and its interaction with *group* as smooth terms. Treatment contrast

<sup>4</sup>All model specifications are reported in the Appendix.

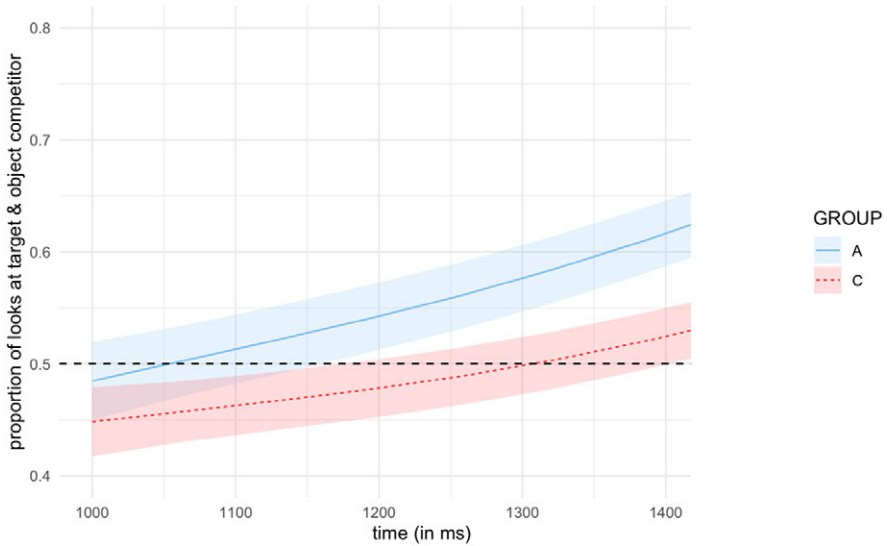


Figure 5. Smoothed lines depicting looks at labeled objects by adults and children in the noun-window (generated using the `plot_smooths()` function). Colored bands indicate the 95% confidence interval (CI). The dashed black line represents chance level.

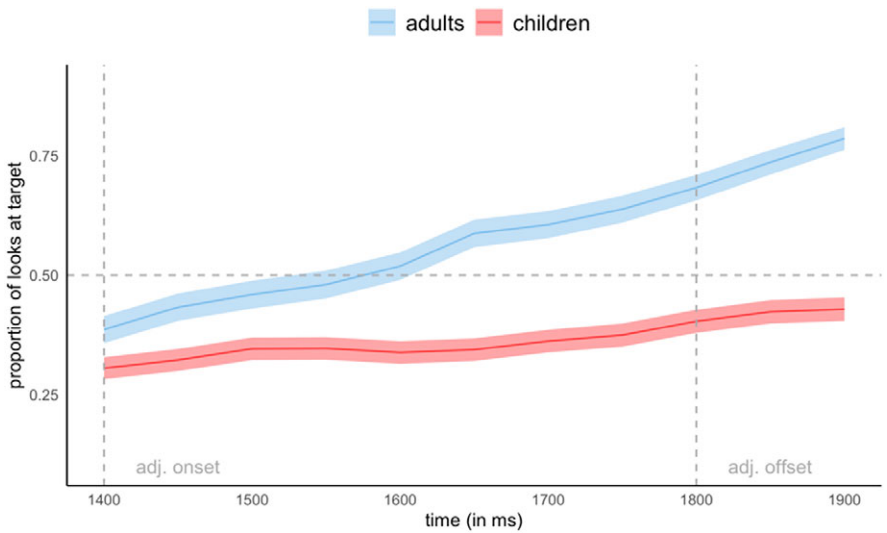
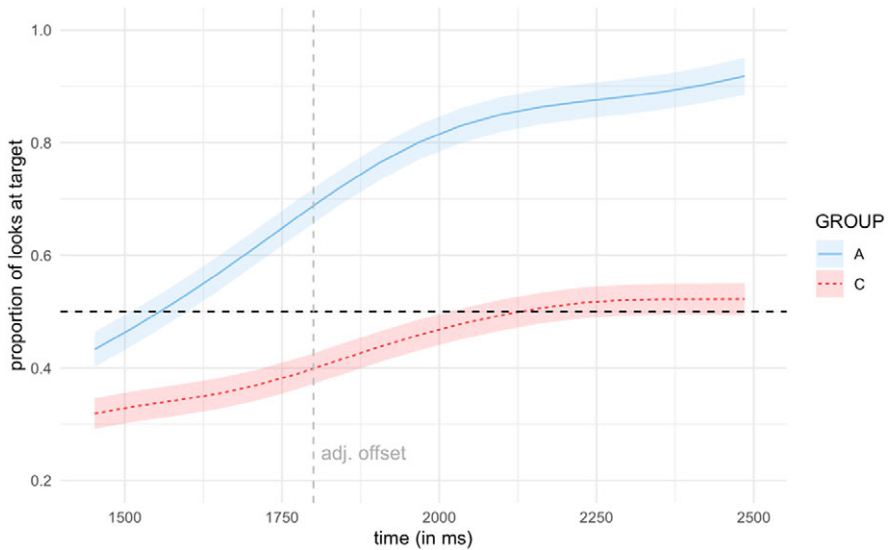


Figure 6. Proportions of fixations to the target for children (plotted in red) and adults (plotted in blue) during the adjective-window. The first vertical line indicates the average onset of the adjective, the second vertical line indicates the average offset of the adjective. Confidence bands show  $\pm 1$  standard error of participant means. The horizontal line indicates the chance level.





**Figure 7.** Smoothed lines depicting looks at labeled objects by adults and children in the adjective-window (1450–1800ms) and over the post noun-phrase-window (1850ms–) (generated using the `plot_smooths()` function). Colored bands indicate the 95% confidence interval (CI). The dashed black line represents chance level.

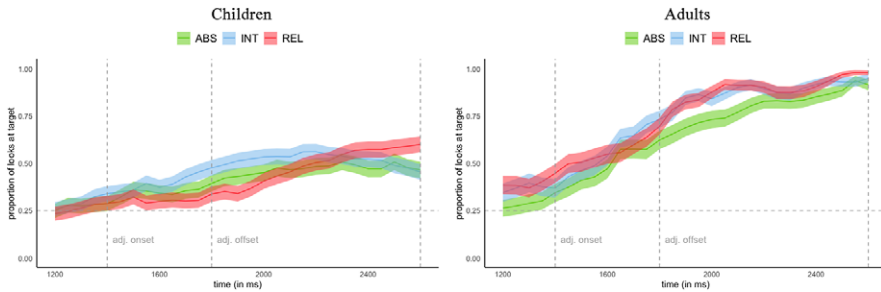
coding was set for *group*. Further, we added *event* as random effect. A null model was also run, excluding the variable *group*. A model comparison revealed that group significantly increased the fit of the model ( $p < .001$ ) (see Figure 7).

During the adjective-window adults fixate on the target object significantly more than children during the whole window. While adults' looks overcome chance level at around 1600ms (i.e., 150ms after adjective onset), children are remarkably slower and do not significantly overcome chance level. The statistical analysis confirmed that adults are fast and accurate in reference resolution and that, already while hearing the adjective-word, they are able to integrate its meaning with that of the noun. By contrast, during the adjective-window children do not integrate the adjective. About 550ms after adjective offset, their looks are on average above chance level, although never significantly, showing an asymmetry between noun and adjective interpretation. In line with the prediction that adjectives are harder and slower to interpret, children seemingly adopt an effortful two-step process in interpreting the adjective, though not completely omitting its integration with noun meaning.

### *The online interpretation of substantive adjectives*

To answer RQ2, we performed a second analysis aimed at comparing looks to the target across adjective-conditions and groups to determine whether there is a difference in the online processing of different semantic types of adjectives across children and adults. Figure 8 contains the plotted proportions of looks to the target for each adjective class for children and adults.

From these graphs, some preliminary observations can be made. First, in all three adjective-conditions, adults are overall faster and more accurate in target identification.



**Figure 8.** Proportion of looks at the target picture throughout the trial for children (left) and adults (right) for each adjective-condition. The first vertical line indicates the average onset of the adjective, the second vertical line indicates the average offset of the adjective. Confidence bands show  $\pm 1$  standard error of participant means. The horizontal line indicates the chance level.

Furthermore, children's visual identification of the target object is faster for color adjectives than for gradable adjectives (relative and absolute adjectives).

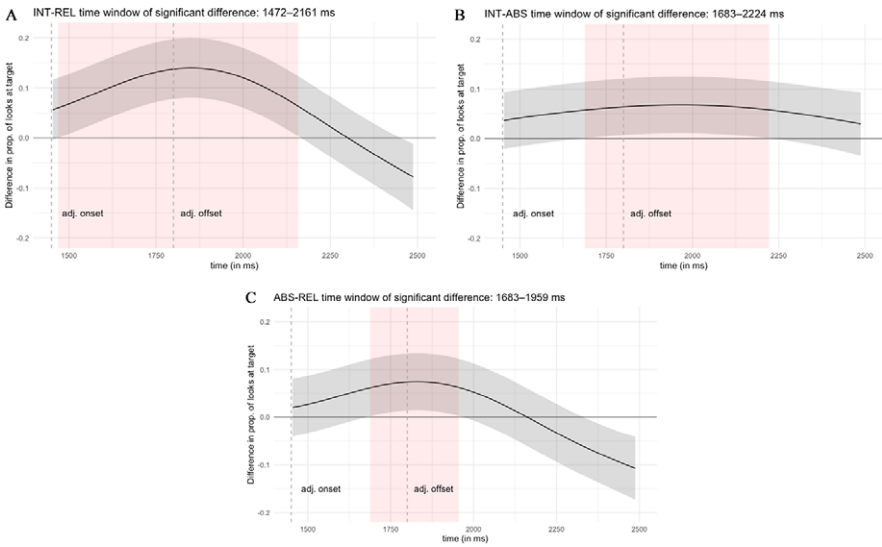
Data analysis was conducted on the adjective- and post noun-phrase-windows. We performed a model comparison using the *compareML()* function in R to assess the fit and differences between two generalized additive models. The null (baseline) model included the proportion of looks at the target as the response variable, *time* as smooth term and *event* as random effect. The full model also included *group-condition* (i.e., the combination of *group* and *condition* as a unique identifier) as categorical variable, transformed into an ordered factor. Additionally, we set treatment contrast for *group-condition*. The model comparison results indicated a significant improvement in fit for the full model compared to the null model ( $p < 0.001$ ).

Figure 9 reports plotted differences between conditions for children.

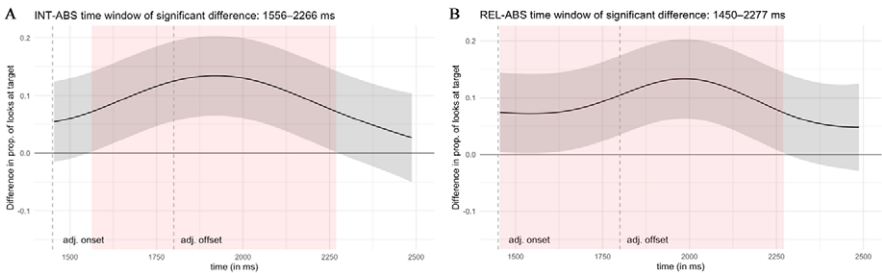
Very early during the adjective-window, children's proportion of looks at the target object in the intersective condition is significantly higher than in the relative adjective condition (Figure 9a). This difference persists until about 300ms after adjective offset, indicating that relative adjectives require more processing time than intersective, which are also less demanding than absolute adjectives. Indeed, a significant difference between intersective and absolute adjectives emerges from around 150ms before adjective offset until 400ms after (Figure 9b). However, absolute adjectives are interpreted faster than relative adjectives. The proportion of looks at the target in the absolute condition is significantly higher than in the relative condition from the end of the adjective-window until about 150ms after sentences offset (Figure 9c).

As for adults, no significant difference emerged between the proportion of looks at the target in the intersective vs. relative condition. However, the proportion of looks at the target in the absolute adjective condition was found to be significantly lower compared to the intersective and relative conditions (see Figure 10).

To summarize, the statistical analysis confirmed that children and adults differ significantly with respect to both speed and accuracy of target reference resolution. Moreover, significant differences were found in the processing patterns depending on adjective-condition. Specifically, children integrate the meaning of nouns and intersective adjectives (e.g., *black*) faster than in the other two adjective-conditions. Relative adjectives (e.g., *big*) are significantly more challenging than intersective and absolute adjectives (e.g., *closed*), with intersective adjectives being processed the fastest. Adults, by contrast, are fast



**Figure 9.** Difference curve in children’s looks to target from adjective onset (plotted with the plot\_difference()) function. The grey band indicates the 95% confidence interval (CI) of the mean difference. Values above zero indicate more target looks in the first mentioned condition. Values below zero indicate more target looks for the last mentioned condition. The difference is significant if the 95% CI does not include zero.



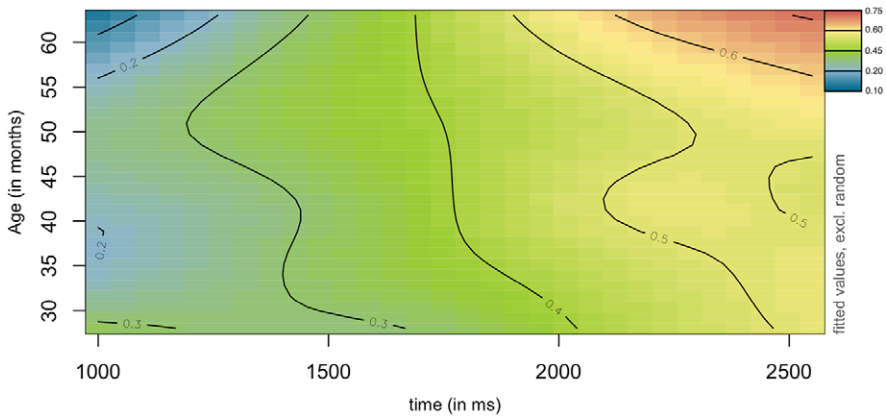
**Figure 10.** Difference curve in adults’ looks to target from adjective onset. The grey band indicates the 95% confidence interval (CI) of the mean difference. Values above zero indicate more target looks in the first-mentioned condition. Values below zero indicate more target looks for the last-mentioned condition. The difference is significant if the 95% CI does not include zero.

and accurate in all adjective-conditions. Surprisingly, in interpreting absolute adjectives the statistical analysis revealed that adults are slower than in the other two adjective-conditions.

*Interpreting noun-adjective combinations in development*

The third analysis aims to answer RQ3, i.e., to determine the possible effects of age in the interpretation of noun-adjective combinations.

To examine the effect of age on children’s looks at the target picture over time we conducted a generalized additive model (GAM) analysis. Two models were run and then



**Figure 11.** Heatmap visualization of the GAM analysis. The x-axis represents time, the y-axis represents children's age (in months). The color scale on the right indicates the density of fixations, with red indicating areas of high fixation density and blue indicating areas of low fixation density.

compared using the `compareML()` function. The first model included age and time as smooth terms, a tensor product interaction term between time and age, and an event term as random effect. The null model was a simplified version of the full model, excluding age and time-age interaction terms. The results of the comparison showed a significant difference between the full model and the null model ( $p < .001$ ), indicating that the inclusion of age and time-age terms improved the fit of the GAM model.

A heatmap visualization was employed to depict the output of the GAM model. The heatmap displays the intensity of looks at the target using a color scale, where hotter colors represent higher fixation density (Figure 11).

Upon inspection, a distinct diagonal pattern is not readily apparent in the heatmap except when looking at older children, for whom a noticeable positive correlation between the proportion of looks at target and age emerges. The cluster of high values in the upper-right corner suggests that children above 55 months of age (i.e., 4;7 years) are particularly accurate and overcome 0.6 proportion of looks at the target upon the end of the post noun-phrase-window. Despite the absence of red hues, the yellow shades effectively convey the directionality of the relationship between target fixation and age. Regarding younger children, although they successfully integrate noun and adjective meanings, no gradual developmental progression becomes apparent over time. This suggests that adjective integration remains deficient until a significant improvement in this demanding cognitive process occurs at the onset of the fourth year.

## General discussion

Previous studies on children's interpretation of nouns and adjectives in combination have rarely used refined online technologies and have never been conducted to investigate languages with noun-adjective order, such as Italian. Furthermore, the interpretation of subsecutive adjectives has only been tested offline and never using the visual-world paradigm with children, much less in comparison to adults. Our experiment took a comprehensive approach and analyzed children's eye-movement behavior during the

simultaneous presentation of visual and linguistic stimuli that demanded the full integration of nouns and adjectives belonging to three different semantic classes.

### *Interpreting nouns and adjectives in combination*

The first research question (RQ1) asked whether, in interpreting Italian attributive adjectives in post-nominal position, young children show a NOUN ANCHORING mechanism, i.e., if they interpret noun-adjective combinations starting from the noun-word but delaying or omitting altogether the integration of the adjectival meaning. How would the NOUN ANCHOR HYPOTHESIS translate in terms of eye movements? In a language like Italian, where the noun precedes the adjective, the eye-movement pattern would entail a fast and accurate gaze towards the target object upon hearing the noun, with a comparatively slower and less accurate gaze towards the object possessing the designated property upon hearing the adjective. This expected pattern mirrors the empirical evidence found in our experiment.

More specifically, we have shown that children between two and five years of age are able to integrate noun and adjective meanings to resolve reference when faced with 4-referent displays. In processing Italian questions of the type *Where is the Noun-Adjective?* (e.g., *Where is the shoe black?*), while looking at two objects crossed by two properties on the screen (e.g., a black shoe, a white shoe, a black sock, and a white sock), children start the interpretation of the noun by focusing their looks on the two labeled objects (e.g., the two shoes) and, AFTER hearing the adjective, they also shift their gaze to the target object (e.g., the black shoe). However, significant asymmetries emerged between adults' and children's processing speed and accuracy, and between noun interpretation and adjectival meaning integration. Other asymmetries emerged as a function of age, which will be discussed separately (see *infra*).

First, the comparison of children's looking patterns with adults' eye movements revealed that children are overall slower and less accurate, as hypothesized. Our data showed that whereas adults are able to interpret each word of the combination during the time-window in which the word is heard, children need to wait for word offsets, and are overall slower in reference resolution. Thus, although children possess the language skills required, processing limitations emerge and result in a significant difference between children's and adults' speed and accuracy.

Moreover, the gap between children and adults becomes especially relevant when considering the interpretation of nouns and adjectives separately. In fact, while adults interpret nouns and adjectives at each word onset, children show a different processing pattern with nouns on the one hand, and adjectives on the other. Nouns get interpreted once they have been heard (i.e., at noun-word offset), but adjective meaning computation is slower: children needed half a second after hearing the adjective to integrate its meaning and solve the task. Therefore, our results are fresh and robust evidence that adjective meaning integration is taxing for children, in line with what has been found in previous studies with children equal in age (Ninio, 2004; Thorpe *et al.*, 2006).

Our results are in compliance with the NOUN ANCHOR HYPOTHESIS, according to which attribution is especially demanding in terms of processing resources and, after having identified the noun category, children might fail in adjective integration. In particular, the processing difficulties with attribution would stem from the complexity of the logical operation required for the integration of adjectival meaning. As hypothesized by Ninio (2004, p. 256), the comprehension of attribution requires a two-step

process whereby, first, the noun must be interpreted, yielding the identification of an object category; then, the adjective must be interpreted relative to the noun, yielding a sub-categorization within the same object category (for instance, the shoes that are black). Whereas adults manage this multistep process effortlessly, for children the integration of information involved in generating a subset of objects is taxing and results in slower and less accurate processing.

The (alternative) hypothesis that children's challenges in adjectival comprehension depend solely or mostly on vocabulary limitations appears tenuous, given that the adjectives deployed in the experiment were drawn from a lexicon characterized by a notably high frequency of usage. Additionally, children's familiarity with these adjectives has been ascertained through the administration of a parental questionnaire. Consequently, we posit that the challenges posed by the integration of adjectival semantics stem from the cognitive processing demands intrinsic to the logical operation of attribution. Crucially, the burden of this cognitive operation is contingent upon the category of adjectives being considered: relative adjectives, as opposed to intersective adjectives, introduce an additional layer of complexity, mandating a comparative assessment with another object along the relevant dimension. Remarkably, our empirical evidence has shown that relative adjectives pose the highest challenge to children across various age cohorts, further underlining their cognitive demands. Hence, the stratification we observed in children's comprehension of different adjectival categories substantiates the notion that the attribution operation, owing to its inherent logical intricacy, appears particularly demanding for children, and especially, for those in the early stages of development. While previous studies proposed this interpretation of facts based on offline behavioral data like error patterns in picture identification and self-corrections, the current study supports the NOUN ANCHOR HYPOTHESIS with an eye-movement analysis, detailing processing strategies throughout the unfolding of noun-adjective combinations and corroborating the challenging nature of attribution in language development.

### *The role of semantic classes in the online interpretation of adjectives*

The second research question (RQ2) aimed to investigate whether there are differences in children's interpretation of different semantic classes of adjectives and how children differ from adults.

Our results show that children can integrate nouns and adjectives to identify the target object in all adjective-conditions. However, as hypothesized, significant differences between semantic classes of adjectives emerged. While the interpretation of intersective adjectives (e.g., *black*) happened shortly after the adjective offset, integrating nouns with relative adjectives (e.g., *big*) was significantly more challenging and required more time. This is in line with the prediction that, in interpreting relative adjectives, once the noun has been presented and reference has been resolved, children look away from the target towards the noun competitor, sensibly checking their choice against the contrast object. Indeed, most semantic theories propose that establishing a standard of comparison is necessary to determine what counts as having a certain property in a given context (e.g., Kamp & Partee, 1995; Kennedy, 1999). Upon hearing *the teddy big*, children first identified the two objects "teddy" on the screen, and later confronted them to identify the one that was big. As for absolute adjectives, our results match the theoretical observation that they share properties with both intersective and relative adjectives



(e.g., Kennedy, 1999, 2007; Kennedy & McNally, 2005). These findings provide evidence that children are sensitive to the inherent semantic differences among subjective adjectives and, consequently, more challenging adjectives (i.e., relative adjectives) require more processing time.

Interestingly, the analysis of adults' eye movements showed that they were fast and accurate in all adjective-conditions, but that absolute adjectives were interpreted more slowly than the other two adjective-conditions. We propose two possible interpretations for this. First, this finding may have a theoretical explanation recalling a similar result found by Aparicio *et al.* (2016). In interpreting prenominal absolute adjectives, adult participants showed a delayed effect of contrast with respect to color (i.e., intersective) and relative adjectives. They argued that this result might be a consequence of the precise interpretation required in processing two-closed-boundaries absolute adjectives (e.g., *open/closed*) in comparison to one-closed-boundary absolute adjectives (e.g., *clean/dirty*) and open-boundaries relative adjectives (e.g., *big/small*). While in interpreting *closed book* no comparison among books is required, to identify the *dirty* object, participants may need to check the object-competitor to make sure that the target object was the dirtier on the screen. Second, there may also be a methodological explanation for this apparent delay in adults' processing of absolute adjectives. Since the task was specifically designed for children and no difficulties were expected in adults' performance, this finding might result from the manipulation of the visual stimuli in this adjective-condition. Indeed, the visual representation of the properties labeled by absolute adjectives required the manipulation of the drawing on multiple dimensions. To put it more clearly, representing intersective and relative adjectives required changing only one aspect of the drawing, e.g., the black/white shoes were identical except for their color and the big/small fishes were identical while differing in size. Absolute adjectives, by contrast, required the objects to be represented by two different pictures. For example, the 'open book' and the 'closed book' were drawings differing both in the predominant color (red of the cover when closed, white of the pages when open) and size (see Figure 2c). Thus, a more attentive analysis of the multidimensional differences among the objects within the same category might have delayed participants' shifts to the target object. We maintain that this second explanation better accounts for our findings in this experiment. Multidimensional differences within the object-category are indeed relevant in this task. We believe that our results cannot be accounted for by assuming that adults have processing difficulties with absolute adjectives in general, nor that these limitations only affect the interpretation of this adjective class. Hence, we argue that this finding is (mainly) the result of the peculiar visual representation of the properties labeled by absolute adjectives.

### *Two- to four-year-olds' processing of noun-adjective combinations*

The third research question (RQ3) investigated whether children's looking behavior reflected differences in development in relation to the online processing of noun-adjective combinations and to the semantic differences among subjective adjectives.

Although contradictory results were found as to whether children interpret noun-adjective and adjective-noun combinations incrementally, all relevant studies in the previous literature showed how, at around 36 months of age, toddlers show a twist in development. At this age, toddlers become more accurate (Ninio, 2004; Thorpe *et al.*, 2006) and manage to

interpret adjective-noun combinations incrementally (Thorpe *et al.*, 2006; Tribushinina & Mak, 2016), unlike 30-month-old toddlers (Fernald *et al.*, 2010). Our results revealed that children do manage to integrate noun and adjective meanings, although less fast and less accurately than children above 4 years and 7 months of age. Their still-developing cognitive resources make this task more challenging, resulting in a significantly slower process of interpretation. This is revealing considering that utterances were presented at natural speed. As already discussed, previous results with children as young as 30 months of age (Fernald *et al.*, 2010; Ninio, 2004; Thorpe *et al.*, 2006) provided evidence in favor of the NOUN ANCHOR HYPOTHESIS. In line with this literature, our findings revealed a robust asymmetry between noun and adjective interpretation, the latter being more demanding and requiring significantly more time. However, testing children in a wide age range led to the compelling observation of a developmental trajectory of adjective integration that is gradually mastered during the third and fourth years of age.

## Conclusion

Our experiment has taken a rigorous approach by analyzing high-resolution online eye-tracking data in response to stimuli that demand full integration of nouns and substantive adjectives in children between two and five years of age. Findings from the current study provide evidence of continuity in children's development of sophisticated, adult-like processing skills. Children as young as two years and four months show a remarkable slowness in the interpretation of noun-adjective combinations with respect to experienced adults, significantly delaying adjective integration. This process, however, gets faster and more accurate with age. Moreover, we have demonstrated that young toddlers are sensitive to the different manners in which each class of substantive adjectives is interpreted. Crucially, examining online processing with eye-tracking has shed light on aspects concerning noun-adjective integration and patterns of interpretation elicited by adjective semantics that never emerged in previous literature.

We conclude by pointing out the limits of the current study and by proposing avenues for future research. It is evident that the findings presented herein necessitate validation through a larger cohort of children spanning the broad age spectrum examined in this study. A larger sample size would also enable the delineation of a more finely tuned developmental trajectory. Moreover, to gain a deeper understanding of attribution processing, it would be paramount to explore the effects of the NOUN ANCHOR HYPOTHESIS on eye-movement patterns in languages with prenominal adjectives. Such a study would shed light on the intricate interplay between the processing demands of attribution and other cognitive variables such as working memory and attention. The pursuit of this research avenue is reserved for future inquiry.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/S0305000923000703>.

**Competing interest.** The authors declare that they have no competing interests.

**Author statement.** All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. In particular, MR and CM contributed to conceptualization, design, writing and revision of the manuscript. MR is responsible for data collection and statistical analysis.

## References

- Aparicio, H., Xiang, M., & Kennedy, C. (2016). Processing gradable adjectives in context: A visual world study. In *Semantics and Linguistic Theory* (Vol. 25, pp. 413–432).
- Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.19, retrieved 13 June 2016 from <http://www.praat.org/>.
- Coretta, S. (2020). tidyv: Tidy model visualisation for generalised additive models. R package version 3.0.0.
- Davies, C., Lingwood, J., Ivanova, B., & Arunachalam, S. (2021). Three-year-olds' comprehension of contrastive and descriptive adjectives: Evidence for contrastive inference. *Cognition*, 212, 104707.
- Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive psychology*, 60(3), 190–217.
- Foppolo, F., & Panzeri, F. (2013). Do children know when their room counts as clean? In S. Kan, C. M. Cantwell, & R. Staubs (Eds.), *NELS 40: Proceedings of the Fortieth Annual Meeting of the North East Linguistic Society: Volume I* (p. 205–218). Amherst, MA: GLSA.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj, *Language development: Language, thought, and culture* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. *Language acquisition and conceptual development*, 3, 215–256.
- Kamp, H., & Partee, B. H. (1995). Prototype theory and compositionality. *Cognition*, 57(2), 129–191.
- Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Routledge.
- Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy*, 30(1), 1–45.
- Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81(2), 345–381.
- Kennedy, C., & McNally, L. (2010). Color, context, and compositionality. *Synthese*, 174, 79–98.
- Klibanoff, R. S., & Waxman, S. R. (2000). Basic level object categories support the acquisition of novel adjectives: Evidence from preschool-aged children. *Child Development*, 71, 649–659.
- Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: The MIT Press.
- McNally, L. (2011). The relative role of property type and scale structure in explaining the behavior of gradable adjectives. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *ViC 2009* (Papers from the ESSLI 2009 Workshop on Vagueness in Communication) (p. 151–168). Berlin: Springer.
- Ninio, A. (2004). Young children's difficulty with adjectives modifying nouns. *Journal of Child Language*, 31, 255–285.
- Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, 58, 1–21.
- R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Sandhofer, C. M., Smith, L. B., & Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning? *Journal of Child Language*, 27, 561–585.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2), 109–147.
- Syrett, K. (2007). Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives. (Doctoral dissertation, Northwestern University).
- Syrett, K., Bradley, E., Kennedy, C., & Lidz, J. (2006). Shifting Standards: Children's Understanding of Gradable Adjectives. In K. U. Deen, J. Nomura, B. Schulz, & B. Schwartz (Eds.), *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition - North America (GALANA)* (Vol. 4, p. 353–364). Cambridge, MA: UConn Occasional Papers in Linguistic.
- Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and Context in Children's Understanding of Gradable Adjectives. *Journal of Semantics*, 27, 1–35.
- Thorpe, K., Baumgartner, H., & Fernald, A. (2006). Children's developing ability to interpret adjective–noun combinations. In *Proceedings of the 30th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

- 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 Psycholinguistic Research*, **42**, 205–225.
- Tribushinina, E., & Mak, W. M.** (2016). Three-year-olds can predict a noun based on an attributive adjective: evidence from eye-tracking. *Journal of child language*, **43**(2), 425.
- Waxman, S., & Booth, A.** (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science*, **6**(2), 128–135.
- Waxman, S., & Klibanoff, R. S.** (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology*, **36**, 571–581.
- Weicker, M., & Schulz, P.** (2018). Is ‘clean’ the same as ‘not dirty’? On the understanding of absolute gradable adjectives. In *Proceedings of the 42th Annual Boston University Conference on Language Development* (pp. 790–802).
- Weisleder, A., & Fernald, A.** (2009). Real-time processing of postnominal adjectives by Latino children learning Spanish as a first language. In *Proceedings of the 24th Annual Boston University Conference on Language Development*. 611–621.
- Wickham, H.** (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Wood, S.** (2006). Generalized additive models: an introduction with R. Chapman and Hall/CRC.
- Wood, S.** (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(1), 3–36.
- Zahner, K., Kutscheid, S., & Braun, B.** (2019). Alignment of f0 peak in different pitch accent types affects perception of metrical stress. *Journal of Phonetics*, **74**, 75–95.