# Three-year-olds' comprehension of contrastive and descriptive adjectives: Evidence for contrastive inference

Catherine DAVIES[1]*, Jamie LINGWOOD[2], Bissera IVANOVA[3], and Sudha ARUNACHALAM[4]

[1]University of Leeds, UK; [2]Liverpool Hope University, UK; [3]Aix-Marseille University, France; [4]New York University, USA

*Corresponding author: School of Languages, Cultures and Societies, University of Leeds, LS2 9JT, UK. Email: c.n.davies@leeds.ac.uk

**This is the author-accepted version of a manuscript accepted by *Cognition*, 26/03/2021**

## Acknowledgements

This work was funded by the Economic and Social Research Council (grant ES/P010296/1) and the National Science Foundation (NSF BCS-1748826). We thank all of the families for their generous participation. We gratefully acknowledge assistance from Kurt Debono and Sam Hutton at SR Research for advice on experimental programming and data preparation, Michelle Peter for guidance on the speed of processing measure, Annika van Wijk for assistance in data entry and for helpful discussions about the study, and Dan Mirman for advice on statistical modelling.

# Three-year-olds' comprehension of contrastive and descriptive adjectives: Evidence for contrastive inference.


## Abstract

Combining information from adjectives with the nouns they modify is essential for comprehension. Previous research suggests that preschoolers do not always integrate adjectives and nouns, and may instead over-rely on noun information when processing referring expressions (Fernald, Thorpe, & Marchman, 2010; Thorpe, Baumgartner, & Fernald, 2006). This disjointed processing has implications for pragmatics, apparently preventing under-fives from making contrastive inferences (Huang & Snedeker, 2013).

Using a novel experimental design that allows preschoolers time to demonstrate their abilities in adjective-noun integration and in contrastive inference, two visual world experiments investigate how English-speaking three-year-olds (N=73, $M_{age}$=44 months) process size adjectives across syntactic (prenominal; postnominal) and pragmatic (descriptive; contrastive) contexts.

We show that preschoolers are able to integrate adjectives and nouns to resolve reference accurately by the end of the referring expression, in a variety of pragmatic and syntactic contexts and in the presence of multiple distractors. We reveal for the first time that they can contrastively infer, given a slowed speed of presentation and visually salient size contrasts. Our findings provide evidence for a continuity in the development of pragmatic skills, which do not appear to be linked to children's language proficiency or speed of processing.



## Keywords

Language development; developmental pragmatics; contrastive inference; adjectives; eye tracking.

## 1. Introduction

Children's comprehension of referring expressions develops throughout the preschool years. They gradually master the referential links between nouns and their real-world referents between 6 and 24 months (e.g., Bergelson & Swingley, 2012; Fernald et al., 1998). However, it is not until their third year that children start to integrate (or combine) information from adjectives to refine their referential understanding, and it is not until a year later that they are able to do this efficiently and flexibly in naturalistic contexts (Klibanoff & Waxman, 2000). The development of adjective comprehension is therefore a protracted process relative to other open word classes (e.g., Berman, 1988; Booth & Waxman, 2009; Gentner & Boroditsky, 2001, Ninio, 1988; Ramscar, Thorpe, & Denny, 2007; Waxman & Booth, 2001).

Several reasons have been proposed for this lengthier path of acquisition, mainly relating to the conceptual and distributional aspects of adjectives. Adjectives make reference to only a property of an object, for example its texture or colour, which violates the whole-object assumption that a new word refers to a complete object (Markman, 1990; Sandhofer & Smith, 2007). They make up around 10% of tokens in child-directed speech: a lower proportion of the input relative to other open classes (Sandhofer, Smith, & Luo, 2000). Adjectives may also pose difficulties due to their semantic, syntactic, and pragmatic variability. The meaning of adjectives often depends on the noun they modify. Consider the relational relativity that is involved in interpreting "little" in relation to a mouse or an elephant; the range of meanings between "nice day", "nice meal", and "nice work"; or the colour similarity between grey clouds and black clouds, grey hair and white hair (Gentner, 1982; Medin & Shoben, 1988; Smith, Cooney, & McCord, 1986; Syrett, Kennedy, & Lidz, 2010). In languages such as English that often place the adjective before the noun, this semantic variability may be doubly hard since the adjective is presented before the noun that constrains its meaning (e.g., Arunachalam, 2016; Ninio, 2004). Pragmatically, adjectives play either a contrastive or a descriptive function in discourse, which determines the path of further inferential processing.

This study focuses on children's real-time processing of adjectives across syntactic and pragmatic contexts. It is important to investigate processing because processing efficiency mediates the association between language input and growth outside of the lab (Weisleder & Fernald, 2013). Children who are fast to process language will have better learning opportunities to acquire subsequent unfamiliar words in the speech stream (Fernald, Marchman & Hurtado, 2008; He, Kon, & Arunachalam, 2020). This is evidenced by studies showing that the faster three-year-olds process an adjective-noun referring expression, the more likely they are to acquire a novel noun later in the utterance (e.g., "The red car is on the deebo"; Fernald, et al., 2008). It is especially important to study how adjectives are processed because in natural speech, they commonly appear in combination with other words, meaning that efficient processing is necessary not just for comprehending adjectives themselves, but also for the constituents they combine with.

There has been less recent research on adjective acquisition compared to noun and verb acquisition, and it is important that older research questions are revisited using newer experimental methodologies. Many existing developmental studies on adjective comprehension have used methods that monitor children's offline performance, i.e., after an adjective or utterance has been presented (Gao, Zalazo, Sharpe, & Mashari, 2014; Hall, Waxman, & Hurwitz, 1993; Mintz & Gleitman, 2002; Mintz, 2005; Nelson & Benedict, 1974; Ninio, 2004; Taylor & Gelman, 1988; Waxman & Booth, 2001). This provides only a limited picture. Since offline studies only capture children's final referential choice (which may have been heavily influenced by the phrase-final noun, or by an earlier parsing decision that children fail to revise), they don't reflect earlier stages of processing, e.g., processes triggered by competitors, and impacts of these processes on accuracy and latency. Offline studies tell us nothing about the timecourse of comprehension, e.g., how long after a prenominal adjective is presented do children show that they have encoded it? By definition, the live record captures processing abilities, which play an important role for acquisition, as discussed above. Finally, online measures can uniquely tell us about children's predictions about elements yet to appear: this is important specifically for investigating contrastive inference (defined in section 1.2), as well as for the wider domain of processing research. For all of these reasons, and given the syntactic, lexical, and pragmatic dimensions of adjectives that need to be integrated online, it is essential that children's real-time processing is analysed.

The current study focuses on three-year-olds' online interpretation of adjectives in continuous speech. We examine this age group for both empirical and theoretical reasons. As discussed below, although adjective comprehension in two-referent displays has been evidenced in three-year-olds (Thorpe, Baumgartner, & Fernald; 2006; Weisleder & Fernald, 2009), this has been limited to paradigms in which the adjective or the noun (but not both) have to be comprehended. In contrast, our design requires an understanding of both adjective and noun in the same phrase: an ability not yet tested in this age group. Likewise, unscaffolded contrastive inference ability has not yet been tested in the under-fives. On the theoretical assumption that children need a substantial amount of language experience to tackle complex pragmatic or sentence processing, studies in this domain have rarely tested preschoolers (e.g., studies on over-fours by Snedeker & Trueswell, 2004; Trueswell et al., 1999; Weighall, 2008; Woodard, Pozzan, & Trueswell, 2016). Here we investigate whether younger children, i.e., those with less language experience can pass complex pragmatic tasks, given time (exp. 2). If they fail, this would promote the importance of language for pragmatics. Conversely if they succeed, this might suggest a role for domain-general skills in pragmatics, and that sophisticated language abilities are less important for specific pragmatic tasks, in this case contrastive inference.

Overall, we analyse comprehension across sentential and pragmatic contexts by addressing four distinct but related research questions in two experiments within a single study (thereby eliminating disparities brought about by different methods used between previous studies). In doing so, it conveys a detailed account of the development of adjective understanding.

### *1.1 Children's integration of adjectives and nouns*

Our first research question examines three-year-olds' adjective-noun integration, i.e., to what extent do they combine crucial information from the adjective with the noun to derive a composed meaning and uniquely resolve the intended reference. Meaning integration is necessary across all levels of language processing and is especially pertinent in adjective interpretation given that the primary function of adjectives is to specify the meaning of a noun, and that adjectives rarely occur as isolated words (Davies, Lingwood, & Arunachalam, 2020). Comprehension is at risk if children do not integrate and instead process the elements serially, for example when asked to pick the "second green ball", five-year-olds picked the second ball in the series which also happened to be green, but not the second of two green balls in the set (Matthei, 1982). At later stages of acquisition, adjective-noun integration is likely to act as a stepping stone for more complex referential structures, e.g., relative clauses or constructions containing a chain of adjective or adverb modifiers.

Integrating adjectives and nouns is challenging for young children, and there are several ways in which this can fail. A widely attested strategy is to use only information from the noun, which has been shown across languages that place nouns both before and after adjectives, ruling out a bias based on linear order. Ninio (2004) showed that Hebrew-speaking children (1;6 – 4;4) frequently ignored postnominal adjectives and unreliably prioritised noun information to resolve reference. In response to requests to point to e.g., a big teddy, they pointed to a small teddy in almost a quarter of trials. Interestingly, their low performance was only apparent in the presence of an adjective competitor – a different object sharing the target's size, e.g., a big clock. Performance significantly increased when the choice was restricted to the contrasting noun pair alone (a big and a small teddy), suggesting that the children could understand the adjective when the noun was not at issue. Thorpe et al. (2006) refined Ninio's study by testing discrete age groups on simple referent pairs such as a red car and a blue car. They showed that English-speaking children at 2;5 over-relied on postadjectival noun information (exp. 1) and did not integrate the noun and adjective until after the whole phrase had been heard (exp. 2). This difficulty was resolved by 3;9 when they were able to do simple adjective-noun integration online.

The offline penalty introduced by competitors that Ninio (2004) and Thorpe et al. (2006) have documented also surfaces in studies measuring online processing. In displays with nine referents including a target referent (a red butterfly), an adjective competitor (a red fox), a noun competitor (a purple butterfly), and six unrelated distractors, Russian-speaking six-year-olds did not use information from the prenominal adjectives to fixate the red referents during the adjective, and instead waited until they had heard the noun (Sekerina & Trueswell, 2012). However, in much simpler displays of two referents, e.g., a red car and a blue car, three-year-olds showed a preference for the target during the prenominal adjective region (Fernald, Thorpe, & Marchman, 2010). Likewise, Spanish-speaking 3;6 year-olds did not wait for the completion of noun-adjective expressions but successfully interpreted them at the earliest possible opportunity (Weisleder & Fernald, 2009). However, in these simple contexts, processing the adjective (or noun in the Spanish case) was sufficient for reference resolution. In the English case, three-year-olds may be treating the adjectives as referential terms in their own right, and ignoring the following noun. Therefore, although the early looking behaviour of three-year-olds in these two studies suggests that they can rapidly recruit meaning to restrict reference, it does not constitute evidence of adjective-noun integration where information from both elements is required.

In an attempt to investigate preschoolers' ability to integrate meaning from adjectives and nouns, Tribushinina and Mak (2016) tested whether three-year-olds could integrate properties of adjectives (e.g., soft) with relevant objects (e.g., pillow) by measuring whether the children looked at the target referent during the prenominal adjective. When the adjective was uninformative (e.g., new), looks to the target object unsurprisingly increased only on hearing the noun. However, when the adjective was informative (e.g., soft), three-year-olds showed a preference for the target object during the adjective, suggesting integration of adjective semantics, informativeness, and world knowledge. However, in line with Fernald et al. (2010), this task could be passed by attending to the adjective and using conceptual knowledge of the target object, i.e., knowing that a pillow is typically soft whereas a competitor (e.g., a book) is not. Adjective-noun co-occurrence statistics are also likely to have scaffolded the early looking behaviour. Thus, Tribushinina and Mak's (2016) results do not directly demonstrate adjective-noun integration.

Collectively, existing research on children's comprehension of adjectives suggests that preschoolers do not reliably integrate adjective-noun combinations online and instead over-rely on information from just one of these constituents: the noun when both adjective and noun information is required for disambiguation, and the adjective when noun information is not required. For successful comprehension however, the child must hold on to the adjective before they hear the noun, combining information from both elements. To robustly test children's integration of referring expressions when *both* the adjective and the noun are required, the current study includes conditions containing both noun and adjective competitors in the same visual display. We also use non-collocational adjectives to remove any opportunity for children to complete the task via co-occurrence statistics or from world knowledge.

The specific adjectives we use are "big" and "little". These are scalar adjectives, meaning that the noun they modify can possess the property they denote (in this case, size) to varying degrees. They are inherently comparative because they can only be interpreted with reference to something else, for example, a big mouse is big for a mouse (see e.g., Kennedy, 2012 for further details about adjective taxonomies). These adjectives are ideally suited for this investigation for several reasons. First, they are early acquired; at two years children can interpret these size terms accurately (Ebeling & Gelman, 1988). They are also highly frequent, e.g., they were the two most frequent scalar adjectives found in a recent corpus study on child-directed speech (Davies et al., 2020). Because they are scalar adjectives, their interpretation necessarily involves semantic integration of the adjective and noun (e.g., Ziegler & Pylkkanën, 2016). Finally, they were used in a comparable study with five-year-olds (Huang & Snedeker, 2013).

*1.2 The emergence of contrastive inference*

After ascertaining whether three-year-olds can integrate noun and adjective information by the end of a referring expression, our second research question investigates whether children can integrate adjectival and referential information earlier in the utterance. That is, we analyse the incidence of contrastive inference in younger children than has been documented previously. In response to modified nouns, e.g., "the tall glass", adults routinely contrastively infer, i.e., they resolve reference during the prenominal adjective, before the noun has been produced. They engage in this type of pragmatic inferencing by exploiting the relationship between the linguistic input, the nonlinguistic context, and their knowledge of referential principles. Sedivy et al. (1999) were among the first to show this effect in adults by documenting early looks to a target member of a contrast set, e.g., a tall glass alongside a short glass in the presence of a singleton object that was also tall, e.g., a tall jug (see also Grodner & Sedivy, 2011; Ryskin, Kurumada, & Brown-Schmidt, 2019). Theoretically, contrastive inference is explained using Grice's Cooperative Principle (Grice, 1975) and its second maxim of quantity, by which comprehenders reason that a speaker would not have used an adjective to refer to the singleton object because it would be overinformative to do so. As a result of this inferential processing, comprehenders fixate the tall member of the contrast set during the adjective. Contrastive inference is key to efficient sentence processing for several reasons. By implicitly signalling a focus on the contrast set, it allows listeners to quickly eliminate the singleton item and reduces the need for speakers to explicate this. Deriving meaning before the end of the referring expression means that comprehension can proceed more quickly, leaving more attentional resources for other kinds of processing. Contrastive inference is also a form of redundancy, safeguarding against potential failures in the system later on, e.g., if the noun signal becomes degraded. Finally, it reduces working memory demands by fostering a global representation rather than the serial retention of individual words (Omaki & Lidz, 2015: 162-3).

To date, adult-like contrastive inference has not been attested online in the under-fives. Seven-year-olds have been shown to fail in an offline contrastive inferencing task (Kronmüller, Morisseau & Noveck, 2014). In an indirect test of contrastive inference, five- but not three-year-olds showed delays and checking behaviour when responding to overinformative expressions (Morisseau, Davies & Matthews, 2013). Four-year-olds were able to contrastively infer in simple, two-referent displays, but only when given framing cues or training (Horowitz & Frank, 2014). It is not until children are five years of age that novel and unscaffolded adjective-noun integration has been documented online, constituting evidence for contrastive inference (Huang & Snedeker, 2013, countering earlier data by Nadig et al., 2003). In displays of four real objects, children showed a preference for the target (a big coin) in the presence of a contrast-mate (a small coin) and a competitor sharing the property of the target (a big stamp). Crucially, this preference emerged during the adjective. Thus, in simple visual contexts, five-year-olds can integrate meaning from a scalar adjective and a noun, and use their knowledge of adjective semantics and informativeness to engage in incremental processing before the onset of the noun, though at slower rates than adults. Huang and Snedeker's (2013) results suggest that this mechanism is functioning at five years of age. However, this is relatively late compared to other kinds of pragmatic inference, many of which have been found in younger children once task demands have been reduced, for example scalar implicature (Pouscoulous et al., 2007; Stiller, Goodman & Frank, 2015), relevance implicatures (Schulze, Grassmann, & Tomasello, 2013), metaphor (Pouscoulous & Tomasello, 2019), metonymy (Falkum, Recasens, Clark, 2016), and presuppositions (Berger & Höhle, 2012). In line with this recent trend in experimental pragmatics, we aim to discover whether contrastive inference is within reach of under-fives.

### *1.3 Children's comprehension of prenominal and postnominal adjectives*

Our third research question investigates whether adjective position affects children's processing of modified noun phrases. The majority of research with English-speaking children has focused on adjectives in prenominal (attributive) positions since this ordering allows researchers to measure children's online comprehension of adjectives versus reliance on the later-presented noun. However, English adjectives can also occur postnominally when used predicatively in a sentence (e.g., "that car is blue") or in a relative clause (e.g., "the car that's blue"). Reducing the hypothesis space to a subset of referents using a noun, and then narrowing it down further to the target using an adjective should intuitively facilitate identification of the referent (Ninio, 2004). This two-step process of adjective comprehension predicts that postnominal frames will result in more efficient processing than prenominal frames.

To our knowledge, only one study has compared the processing of adjectives in both positions, finding that 3-4 year-old English speakers processed referring expressions with a postnominal modifier (in a preposition phrase or relative clause) faster than those with a prenominal adjective (Arunachalam, 2016). Relatedly, training that presented colour modifiers postnominally in English resulted in better learning of these colour words by two-year-olds (Ramscar, et al., 2010).

So why should children find prenominal adjectives more demanding than those in postnominal position? Given that interpretation of the adjective is dependent on the noun it modifies (e.g., Kamp & Partee, 1995), children may fare better if they get the constraining word first. After all, children have limited memory and processing capacities, and having to process and retain the meaning of the adjective before the noun is heard may overtax them (e.g., Arunachalam, 2016). Further, noun labels may be more familiar, more robustly represented, and more accessible for children than modifier labels (e.g., Hall, Waxman, & Hurwitz, 1993). Perhaps paradoxically, corpus studies of child-directed speech show that modifying adjectives occur more frequently in prenominal positions. In an analysis of 12 common adjectives used by and to children, parents used prenominal frames in 52% of the 3,067 occurrences in a child-directed speech corpus, with colour words occurring prenominally roughly 70% of the time (Thorpe & Fernald, 2006). In a larger corpus of adjectives in child-directed speech across a range of interactive and socioeconomic contexts, adjectives were found to occur prenominally in 52% of occurrences, cf. 41% postnominally (Davies et al., 2020). These findings present a puzzle: the forms that should be more cognitively taxing for children are also more frequent in the input. Our study tackles this incongruity by ascertaining empirically whether prenominals are indeed more taxing than postnominals.

### *1.4 Drivers of contrastive inferencing ability*

Although this small research base documents young children's emerging skills in adjective-noun integration and contrastive inferencing, it is not yet clear what matures in the child to enable them to master these abilities. Thorpe et al. (2006) speculate that holding an adjective in mind while listening to a noun poses a memory demand, manifest in younger children's retention of utterance-final noun information. Vocabulary may also play a role, such that the richer a child's lexicon, the faster they are at responding to familiar words, and the better they will be at leveraging off those words to interpret word combinations (Fernald, Perfors &

Marchman, 2006). In a detailed discussion, Fernald et al. (2010, p.210ff) suggest three potential causes of developmental changes in contrastive inference: robustness of lexical knowledge, language processing speed, and semantic integration. Our final research question (addressed in Experiment 2) analyses the relationship between these skills and children's performance in our contrastive inference task in an attempt to reveal what may underpin adjective-noun integration.

## 2. Experiment 1

The first experiment analysed eye movement data to investigate three-year-olds' integration of adjectives and nouns, their contrastive inferencing ability, and their comprehension of prenominal and postnominal adjectives. It examined the nature and timecourse of each of these skills across two pragmatic and two syntactic contexts. Stimuli exploited size contrasts using "big" and "little". Four conditions were included in a fully crossed design. The two pragmatic conditions were Contrastive (i.e., there was a competitor object in the display from the same object category that contrasted in size) and Descriptive (i.e., there were no competitor objects from the same object category). The two syntactic conditions were Prenominal and Postnominal (relative clause) positioning of the adjective (see Table 1 for stimulus details). Here, we briefly outline the hypothesized computations that listeners should make for each condition. In the Contrastive conditions, (mature) listeners can use the first element in the phrase (i.e., the adjective in the Prenominal conditions, or the noun in the Postnominal conditions) to begin to narrow their search for the referent. Importantly, in the Prenominal Contrastive condition, if listeners are able to use contrastive inference, the presence of a contrast set plus a prenominal adjective enables early target fixation during the adjective (addressed by RQ2 below). In the Descriptive conditions, we would expect listeners to show a slightly different pattern. Because the Descriptive conditions did not have a noun competitor, reference can be resolved quickly in the Postnominal Descriptive condition (as soon as the noun is heard), but only later in the Prenominal Descriptive condition, due to the presence of an adjective competitor. Across all conditions, children may rely only on the first element, or only on the noun (Ninio, 2004), which would render them unable to select reliably between the target and the competitors.

The experiment addresses three research questions, formulated to give a comprehensive account of three-year-olds' adjective comprehension both offline and online. RQ1b was included to develop earlier research on the role of distractors in the computation of referential meaning (Ninio (2004).

1. *a. Do preschoolers integrate adjectives and nouns to reliably resolve reference?*
   *b. To what extent does the presence of competitors that share property or object features with the target threaten reference resolution?*

We hypothesised that three-year-olds will integrate nouns and adjectives to preferentially look at the target referent by the offset of the referring expression in all conditions, and that the presence of both noun and adjective competitors in the Contrastive displays will reduce target preference as compared to the Descriptive condition.

2. *Do preschoolers show contrastive inference?*

Since contrastive inference has not been widely tested in this age group, two hypotheses of differing strengths drive this analysis. The first, stronger hypothesis predicts a developed skill in contrastive inferencing. The second, weaker one predicts an emerging skill.

a. Children will show a stronger preference for the target during the prenominal adjective in the Contrastive condition relative to the Descriptive condition (by using the presence of a contrast set in the Contrastive condition to infer that a speaker intends their adjective to distinguish between members of that contrast set).
b. Children will show greater distraction from the adjective competitor in the Descriptive condition relative to the Contrastive condition (since in the absence of a contrast set in the Descriptive condition, the prenominal adjective could equally apply to the adjective competitor and the target).

3. *Do preschoolers process modified noun phrases more quickly when adjectives occur pre-or post-nominally?*

We hypothesised that children will show stronger and earlier target preference in response to utterances containing postnominal adjectives compared to prenominal adjectives.

## *2.1 Method*

### *2.1.1 Participants*

Child participants (*N* = 37) were recruited from a database of family volunteers at the lead author's institution. One participant was excluded as they were outside the target age range. The final sample of 36 children (21 girls, 15 boys) had a mean age of 3 years 9 months (= 45 months; range 42 – 48 months, *SD* = 2). This sample size allowed detection of a medium/large effect size with a two-sided 5% significance level and a power of 80%. All were typically developing, monolingual, native speakers of British English with normal or corrected-to-normal vision and hearing. Caregivers were asked to complete a short family questionnaire that collected demographic information. Regarding the highest level of maternal education, 19% had completed high school, 39% had a Bachelor's degree, 25% had a Master's degree, and 6% had a PhD. Three percent of participants chose not to answer. Families received £10 for their participation.

### *2.1.2 Design*

Using a 2x2 repeated measures design, two variables were manipulated within subjects. We manipulated the **pragmatic function** of adjectives. Their function was either contrastive (for disambiguating between a big cow and a little cow) or descriptive (for describing a singleton cow). We also manipulated the **syntactic frame** by presenting adjectives either in prenominal or postnominal position (e.g., "Where's the big cow?" vs. "Where's the cow that's big?"). The four conditions are exemplified in Table 1.

### *2.1.3 Materials: Visual world task*

The visual world task used grayscale stimulus images created from child-friendly drawings of familiar objects (originally created for Davies, Andrés-Roqueta & Norbury, 2016). None of the object names began with the same onset as other objects in the concurrent array, or with the same onset as "big" or "little" to avoid false anticipation of the target. All images fitted within a 234 x 247 pixel interest area. The big images fitted tightly within this frame and were 1.5 times the size of little ones. Each display contained 4 images. Twenty-six trials were created: 16 critical items (4 in each condition), 8 filler items, and 2 practice items.

As exemplified in Table 1 and Figure 1, in the two Contrastive conditions there was a target object, a noun (category) competitor that was the same object as the target but of a contrasting size, an adjective (property) competitor that was an unrelated object sharing the same attribute as the target, and an unrelated distractor. In the two Descriptive conditions there was a target object, an adjective competitor, and two unrelated distractors. No other adjectives were required to discriminate the target from its noun competitor.

**Table 1**. Example stimuli in the critical conditions plus filler items.

| **Condition** | **Utterance** ***(Where's the…)*** | **Target** | **Noun competitor / Distractor** | **Adjective competitor** | **Distractor** |
|---|---|---|---|---|---|
| Prenominal Contrastive | "big cow" | big cow | little cow | big flower | tree |
| Postnominal Contrastive | "cow that's big" | big cow | little cow | big flower | tree |
| Prenominal Descriptive | "big cow" | big cow | scissors | big flower | tree |
| Postnominal Descriptive | "cow that's big" | big cow | scissors | big flower | tree |
| Filler items | "book" | book | little melon | big melon | lorry |

Filler items always contained two noun competitors (i.e., a contrast set), and two unrelated objects. Filler targets were never a member of the contrast set and were always described using an unmodified noun. The fillers were designed this way to mask the pattern inherent in the contrastive trials (where the target was always a member of the contrast set), and in doing so reduced the predictability of the target.

Participants viewed displays while listening to pre-recorded utterances of the form "Where's the [big/little] [noun]?" or "Where's the [noun] that's [big/little]?" All trials ended with the question "Can you point to it?" Utterances were recorded by a female native speaker of English without pitch accent to prevent prosodic cues (Nadig et al., 2003). The average utterance duration was 1500 ms (*SD* = 233) for the prenominal trials and 1504 ms (*SD* = 274) for the postnominal trials. All stimuli can be found at osf.io/hp9ns.

The critical items appeared in 4 pseudorandomised lists, counterbalanced for the target size adjective and block randomised. For example, half the participants saw the little cow as the target, while the other half saw the big cow as the target. No object appeared as target more than once throughout the experiment, and the position of the target and the contrast objects was rotated around each quadrant of the displays. Between lists, critical target images appeared once as a target and once as an adjective competitor. The order of stimuli presentation was pseudorandomised such that there were at most two consecutive trials of the same condition.

The trial sequence with timings is shown in Figure 1. A colourful, jangling animation in the centre of the screen acted as an attention getter. This was gaze-contingent so that each successive trial would not begin until the participant had fixated the attention getter for 500 ms. In cases where they did not focus on it for 500 ms, the next trial automatically began after 2000 ms.

*2.1.4 Materials: Standardised tests*

Subscales of the Clinical Evaluation of Language Fundamentals Preschool 2 UK (CELF) (Wiig, Second, & Semel, 2006) were administered to the children to investigate associations between their linguistic abilities and their performance in the visual world task. However, due to uncertainties about the interpretation of the children's performance on the visual world task, we did not analyse these measures for Experiment 1. Full details of this part of the experiment can be found in section 3.1 below.

*2.1.5 Procedure*

Participants were tested individually in a purpose-designed lab. The experimenter welcomed families to the warm-up area and played with the child until they were comfortable in the setting. The procedure was then explained to the caregiver. Caregivers gave their informed consent on behalf of their child before completing the family questionnaire.

Verbal assent was secured from each child before proceeding. The children first completed an object recognition task in which the 26 target images used in the visual world task were presented one by one on PowerPoint slides. All children were able to name all of the images.

For the visual world task, participants then moved to a neighbouring room set up for eyetracking. Each child was sat in a car seat in front of a 21.5" Iiyama monitor (1920 x 1080 resolution) at a distance of approximately 90cm. An SR Research EyeLink 1000Plus eyetracker sampling at 500Hz with a 16mm lens was used to monocularly track the right ($N$ = 35) or left eye ($N$ = 1). A 5-point calibration and validation was performed. During the experiment participants viewed the visual displays while listening to pre-recorded utterances presented through external speakers. Caregivers were positioned behind the child and were asked not to talk to them outside of the scheduled breaks.

Children were told that they would see pictures on the screen and would be asked to point to one of them. They were asked to place their hands on two cut-out handprints in front of them between trials, and to point to the named object using the hand on the same side of the screen as the object. The purpose of the pointing task was to get children actively involved in the task; the data was not analysed. Trials were presented in blocks of 6, with the opportunity for breaks in between. The visual world task lasted 15 minutes.

On completing the session, the families were thanked, debriefed, and paid for their participation. Participants also received a book, a small toy, and a certificate. The whole testing session lasted approximately one hour. The study was approved by the Faculty Research Ethics Committee at the lead author's institution.

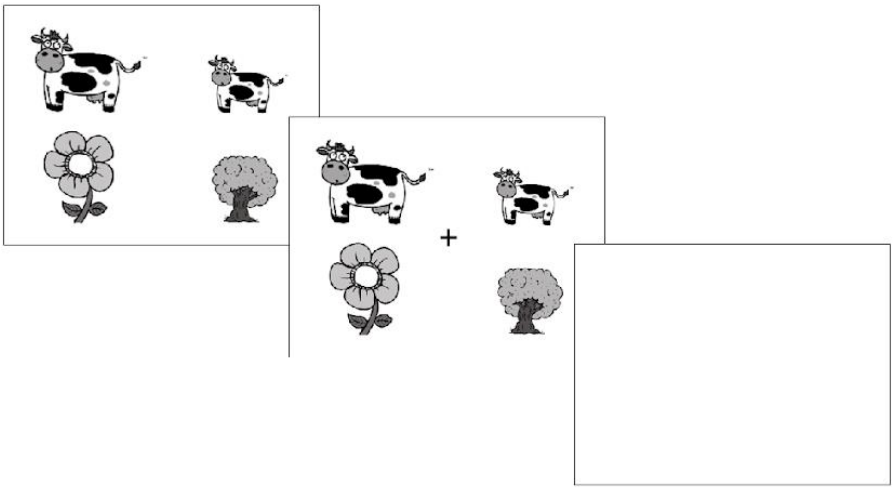

**Figure 1.** Trial sequence. 1) The display was previewed for 1000 ms. 2) A fixation cross appeared in the centre of the display for a further 1000 ms. With the display visible, the audio stimulus was played, e.g., "Where's the big cow?" (≈1500 ms in Experiment 1; ≈2800 ms in Experiment 2), followed by a 2500 ms pause and "Can you point to it?", at which point the participant pointed to image which best matched the referring expression. 3) Trials ended with a blank screen lasting 1000 ms.

*2.2 Results*

*2.2.1 Data preparation and analytical approach*

Prior to analysis we excluded trials in which the eyetracker lost track of participants' eyes on more than half of the samples per trial. This resulted in 180 exclusions out of the original 560 trials (32% of the original data set).

To address research question 1a, we calculated log gaze probability ratios for the target, which are calculated as: log(proportion of looks to the target + 1/sum of proportion of looks to competitors +1)-1 (Wienholz & Lieberman, 2019). Positive values indicate greater preference for the target; negative values indicate greater preference to the other images. We used linear mixed-effects regression with probability ratio as the dependent variable and participant as random factor. A significant intercept parameter in this model indicates that participants looked significantly more to the target than the competitors.

To address research questions 1b and 2, the effect of condition on looks to the target image was modelled for specific time windows using Growth Curve Analysis (GCA; Mirman, 2014). GCA is a multilevel regression technique designed for analysing time course data. By using polynomial models that are able to capture any data shape, it provides a way to explicitly model change in gaze preference over time. It also quantifies both group-level effects (i.e., experimental manipulations) and the effects of individual differences[1].

We used two outcome variables in the analyses: a) proportion of looks to target, and b) target advantage. The proportion of looks measure indicates the strength of preference for the target over all other sections of the array. Observations were aggregated into 100 ms bins (Barr, 2008), and the proportion of looks to target was derived by dividing the number of samples that fall in the target interest area by the number of samples that fall elsewhere, i.e., the other three predefined interest areas, onscreen looks outside of the predefined interest areas, and off-screen looks. This gives a value between 0 and 1. Target advantage then refines the first measure; it is often used in visual world studies to indicate the extent to which a specific competitor draws attention away from the target (e.g., Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Schwarz, 2014; Tian, Ferguson, & Breheny, 2016). Target advantage is derived by subtracting the proportion of looks to the most relevant distractor from the proportion of looks to the target, giving a value of 1 (solely fixating target), 0 (fixating neither target nor specified distractor) and -1 (solely fixating specified distractor). For example, in the Prenominal conditions, target advantage would indicate the degree to which

---

[1] Recently published analyses have raised concerns about GCA. For example, that it can lead to biased parameter estimates and spurious interactions when observed proportions are based on few observations or show floor/ceiling effects (Donnelly and Verkuillen, 2017), or that it is anticonservative (Huang & Snedeker, 2020). Following Huang and Snedeker's recommendation that logistic regression should be used in place of GCA, we modelled our data using both approaches and found similar results. Comparative analyses can be found in the scripts at osf.io/hp9ns.

participants were solely fixating the target (e.g., the big cow), or fixating the image that shared the same property as the target (e.g., the big flower).

Both outcome measures were transformed using an empirical logit transformation (elogit) (Barr, 2008). It is calculated as $log(Y+.5N-Y+.5)$, where Y is the number of samples within the 100 ms timebin for which the gaze fell within the bounds of the target object and N is the total number of samples within each bin. Log is an approximation of log odds. Although some researchers have argued that floor and ceiling effects can mean that elogit analysis can produce biased parameter estimates (Donnelly & Verkuilen, 2017), there were no such effects in our data.

Analyses were performed using mixed-effects regression as implemented in the package *lme4* (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2018). All pre-processing was conducted in EyeLink Data Viewer v.4.1.63 (2020). Full details of model fitting can be found in Supplementary Materials B. Data and analysis scripts are available at osf.io/hp9ns.

*Preliminary observations*

Before presenting the analyses, we note that visual inspection of Figure 2 (which depicts the proportion of looking to each interest area by condition over time) offers three preliminary insights that are supported by the analyses reported below.[2] First, looks to the target increase as the utterance unfolds (examined statistically in research question 1a). Second, target preferences emerge at different time points across conditions. This is not surprising because reference can be resolved at different points in each condition: when the adjective occurs before the noun (Prenominal conditions), the target can be uniquely identified during the adjective in the Contrastive condition (if contrastive inference occurs), but not until the noun in the Descriptive condition. Note that the emergence of target preference around the noun offset in the Prenominal Contrastive condition suggests that preschoolers are not drawing contrastive inferences (investigated further in research question 2). In the Postnominal conditions, the target can be uniquely identified during the noun in the Descriptive condition (with no pragmatic inferencing), but not until the adjective in the Contrastive condition. These disambiguation points are indicated in bold text annotations in Figure 2. Third, competition emerges from different interest areas across conditions (examined in research

[2] A sample of adult participants undergoing the same experiment show similar patterns to those shown in Figure 2. A summary of the method and results from the adult sample can be found in Supplementary Materials A.

question 2). For example, in the Postnominal Contrastive condition, the noun competitor presents strong competition for the utterance "the cow that's big" until the disambiguation point.

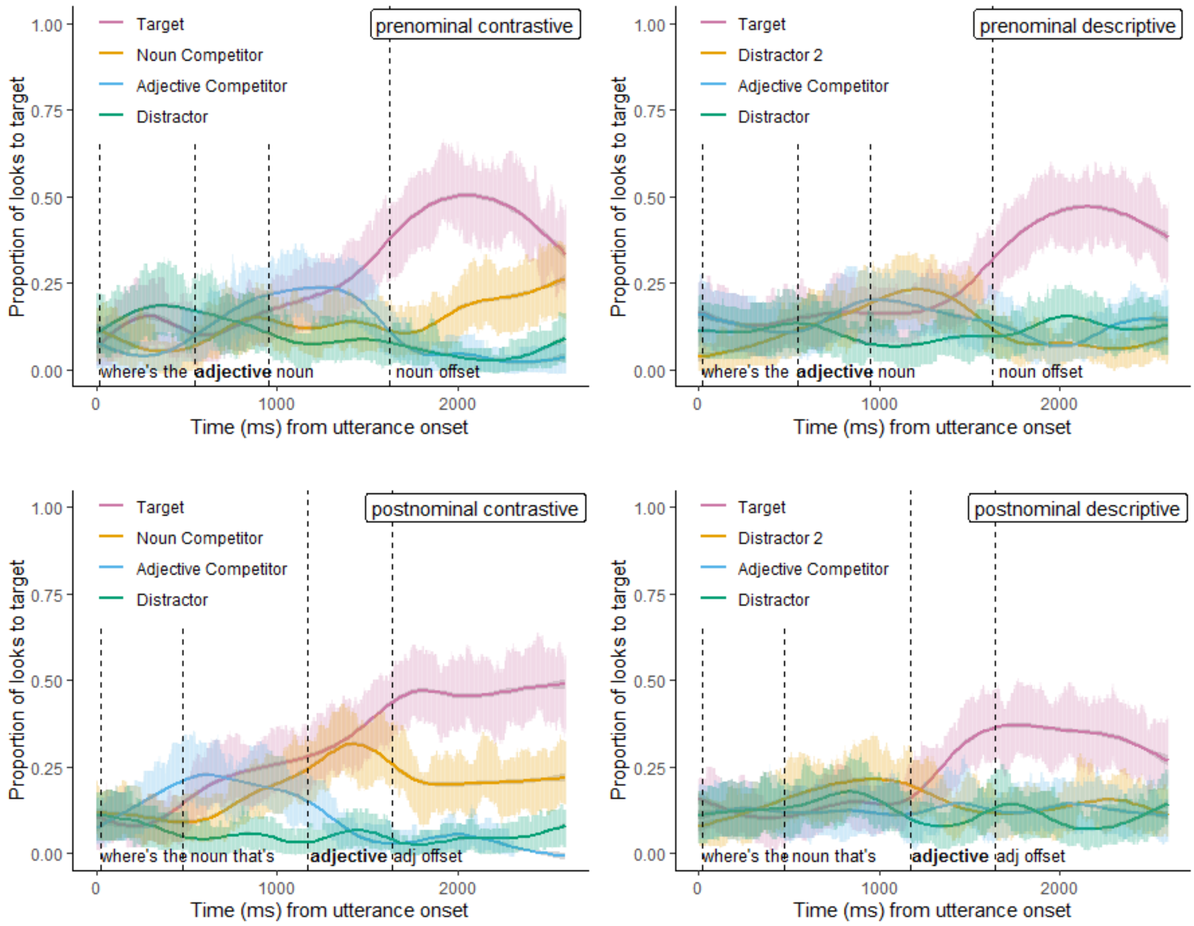


**Figure 2.** Proportion of looks (untransformed) to each interest area in each condition. Vertical dashed lines represent mean onset times Confidence bands show standard error of participant means. Bold text annotations indicate disambiguation points.

*2.2.2 RQ1a: Do preschoolers integrate adjectives and nouns to reliably resolve reference?*

Here, because our focus is on whether children ultimately resolve reference, we examine looking behaviour after the utterance has unfolded (akin to an offline measure rather than a measure of incremental processing), specifically, during a 2000 ms window from the offset of the utterance during which there was silence. We calculated log gaze probability ratios for the target relative to all other images to quantify target preference. Values averaged over participants, items, and conditions suggest a greater preference for the target ($M$ = 0.10, $SD$ =

0.24, range = -0.49 - 0.99). We then fitted a linear-mixed effects regression to compare the probability ratios to the intercept with participant as a random effect. This revealed a main effect of the intercept ($\beta$ = .04; *SE* = .02; *t* = 2.19; *p* < .05), indicating that participants looked significantly more to the target picture than the competitors, as predicted.

*2.2.3 RQ1b. Where the adjective appears before the noun, to what extent does the presence of an adjective and a noun competitor threaten reference resolution?*

Next, we asked whether the presence of an adjective and a noun competitor weakens target preference (measured as proportion of looks to the target) in the Prenominal conditions. Recall that displays in the Contrastive condition contained both a category (noun) and a property (adjective) competitor, whereas displays in the Descriptive condition contained only an adjective competitor. This difference allows us to run an analysis with condition as the predictor and strength of target preference as the outcome. If target preference is weaker in the Contrastive condition, this should reflect the additive effect of two types of competitor drawing attention away from the target. Our initial hypothesis was that this would be the case, and this pattern is reflected in Figure 2 (upper panels), in which children look more to the noun competitor (yellow line) in the Contrastive conditions than the Descriptive conditions.

Like for research question 1a, we were interested in preschoolers' final interpretation of the utterance, but here, we included the noun in the time window as well. Because the preschoolers appear not to be making a contrastive inference, their functional disambiguation point falls during the noun in both Prenominal conditions, allowing us to analyse looking behaviour during the same time window. The time window for analysis therefore runs from the onset of the noun and for the following 2000 ms. The mean duration of the noun was 675 ms (*SD* = 164, range 502 - 1116). Because the audio stimuli were identical, any differences in gaze behaviour will be due to differences in the visual stimuli.

We now present the growth curve analyses of pragmatic condition (Contrastive, Descriptive) on proportion of looks to target. Growth functions were added stepwise to the model and the overall curves were modelled with third-order orthogonal time orders (OTs) in addition to the fixed effect of condition. Table 2 shows the fixed effect parameter estimates and their standard errors along with *p*-values estimated using the normal approximation for the t-values. There was no effect of condition, indicating no differences in overall target fixation proportions (β = -0.39, *SE* = 0.49, *p* = .43). Likewise, there was no significant interaction between condition and any of the linear, quadratic or cubic terms (all *p*-values > .05), indicating no difference in the curvature between conditions. Significant main effects of those terms reflect the change (increase) in looking to the target over time. Figure 3 shows elogit-transformed mean proportions of looks with GCA cubic curves.

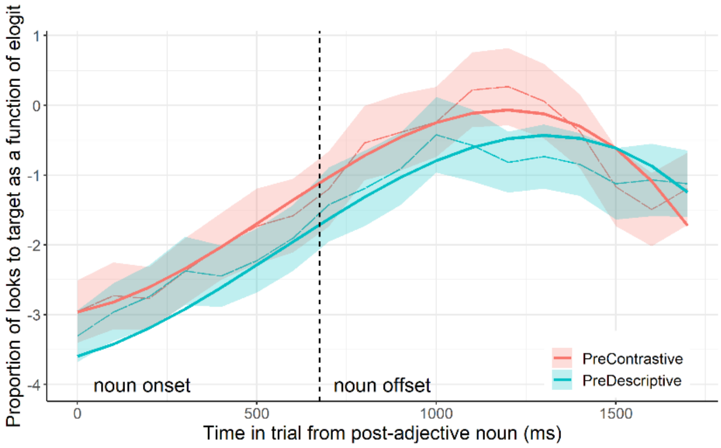


**Figure 3.** Elogit-transformed proportion of looks to target (dashed curves) in Prenominal conditions from the onset of the noun. Bold curves indicate cubic growth curves fitted to the data. Confidence bands show standard error of participant means.

**Table 2:** Model summary for effect of condition on proportion of looks to target from the onset of the adjective. PreD = Prenominal Descriptive. P-values are marked with an asterisk only if critical to the analysis.

| Term | Estimate | SE | t | *p* |
|---|---|---|---|---|
| (Intercept) | -1.23 | 0.42 | -2.92 | 0.01 |
| Linear | 3.02 | 0.43 | 7.02 | 0.00 |
| Quadratic | -2.46 | 0.42 | -5.84 | 0.00 |
| Cubic | -1.22 | 0.42 | -2.88 | 0.00 |
| PreD | -0.39 | 0.49 | -0.81 | 0.43 |
| Linear: PreD | 1.09 | 0.59 | 1.84 | 0.07 |
| Quadratic: PreD | 0.72 | 0.58 | 1.23 | 0.22 |
| Cubic: PreD | -0.72 | 0.54 | -1.33 | 0.60 |

The lack of difference between conditions indicates that contrary to our prediction, preschoolers did not show a weaker preference for the target in the Contrastive condition where there were both noun and adjective competitors.

*2.2.4 RQ2: Do preschoolers show contrastive inference?*

This analysis investigates whether during the adjective, preschoolers show a stronger preference for the target in the Contrastive condition – where they could use Gricean reasoning to exclude the singleton object as the intended target – relative to the Descriptive condition. Data supporting this pattern would evidence that preschoolers are able to make contrastive inferences. Visual inspection of the prenominal contrastive panel in Figure 2 tentatively suggests that contrastive inference is out of reach of this age group.

Here we used target advantage as the outcome variable. This measure is most suitable because it indicates preference for the target in relation to the strength of competition from the adjective competitor, thus providing a measure of how much preschoolers consider the adjective competitor (which is the only other object that fits the unfolding utterance) as a likely referent for the referring expression. If participants generate a contrastive inference, the adjective competitor should not present competition effects. By analysing looks to the same competitor object in both conditions, we can compare the extent to which that competitor is drawing attention away from the target. If preschoolers use the presence of the contrast set to infer that the adjective is likely to refer to one of its members rather than to the singleton item, and equally, use the *absence* of a contrast set to infer that the adjective is equally likely to refer to either of the images that matches the adjective, they should show lower levels of distraction from the adjective competitor in the Contrastive condition, and more distraction in the Descriptive condition.

Two hypotheses of differing strengths drive this analysis. The first, stronger hypothesis predicts a developed skill in contrastive inferencing. The second, weaker one predicts an emerging skill.

1. Preschoolers will show a stronger preference for the target in the Contrastive condition relative to the Descriptive condition.
2. Preschoolers will show greater distraction from the adjective competitor in the Descriptive condition relative to the Contrastive condition.

*Prenominal conditions*

To investigate the strength of competition away from the target, we analysed the effect of condition on proportion of looks to the target minus looks to the adjective competitor (= target advantage) during the adjective window. The mean duration of this window was 407 ms ($SD$ = 65, range 265 - 506). If children used the presence of a contrast set to infer that the adjective refers to one of its members, they should show fewer looks to the adjective competitor, and thus a stronger target advantage in the Contrastive condition. Growth functions were added stepwise to the model and the overall curves were modelled with first-order OTs in addition to our fixed effect of condition. For this analysis, observations were aggregated into 50 ms bins because of the short duration of the time window. As Table 3 shows, there was no effect of condition on the intercept term, indicating no overall differences in target advantage between Prenominal Descriptive and Prenominal Contrastive conditions (β = -0.01, $SE$ = 1.09, $p$ = .99). There was also no significant interaction between the linear term and condition (β = -0.10, $SE$ = 0.71, $p$ = .88), confirming that there was no difference in the linear slopes of target advantage scores between conditions. Figure 4 shows elogit-transformed target advantage scores with GCA linear curves.

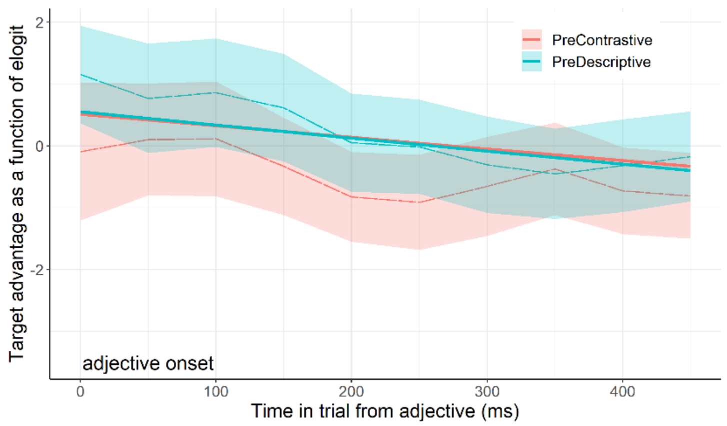


**Figure 4.** Elogit-transformed target advantage scores (dashed curves) in Prenominal conditions from the onset of the adjective. Bold curves indicate linear growth curves fitted to the data.

**Table 3.** Model summary for effect of condition on target advantage scores during the adjective. PreD = Prenominal Descriptive.

| Term | Estimate | SE | t | *P* |
|---|---|---|---|---|
| (Intercept) | 0.09 | 0.90 | 0.10 | 0.92 |
| Linear | -0.85 | 0.52 | -1.62 | 0.10 |
| PreD | -0.01 | 1.09 | -0.01 | 0.99 |
| Linear:PreD | -0.10 | 0.71 | -0.15 | 0.88 |

This analysis confirms that preschoolers did not show direct or emerging contrastive inferencing. They did not show a stronger preference for the target in the Contrastive condition relative to the Descriptive condition, nor did they show greater distraction from the adjective competitor in the Descriptive condition relative to the Contrastive condition.

*Postnominal conditions*

Although not traditionally analysed in studies of contrastive inference, we also ran a post hoc analysis on the effect of Postnominal conditions on target advantage during the noun + relative pronoun window. Here we ask whether preschoolers use the presence of the singleton object in the Descriptive condition to infer that no adjective is needed after the noun, and look at the target before they hear the adjective. If this is the case, they should show an earlier target advantage in the Descriptive than in the Contrastive condition.

Growth functions were added stepwise to the model and the overall curves were modelled with fourth-order OTs in addition to our fixed effect of condition. For this analysis, observations were aggregated into 50 ms bins, because of the small duration of this time window. As Table 4 shows, there was a significant effect of condition, though against our predictions, there was a higher overall target advantage in the Contrastive condition rather than in the Descriptive condition ($\beta = -2.66$, $SE = 0.58$, $p < .01$, $d = -2.46$). There was no significant interaction between the linear term and condition ($\beta = 0.22$, $SE = 0.75$, $p = .77$), indicating no difference in trajectories of target advantage across condition. Figure 5 shows elogit-transformed target advantage scores with GCA linear curves.

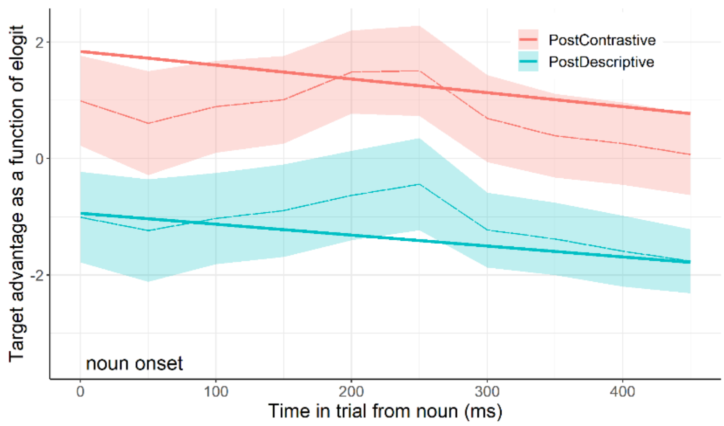


**Figure 5.** Empirical logit-transformed target advantage scores (dashed curves) in Postnominal conditions from the onset of the noun. Bold curves indicate linear growth curves fitted to the data.

**Table 4.** Model summary for effect of condition on target advantage scores during the noun. PostD = Postnominal Descriptive.

| **Term** | **Estimate** | **SE** | **t** | ***p*** |
|---|---|---|---|---|
| (Intercept) | 1.30 | 0.61 | 2.13 | 0.04 |
| Linear | -1.07 | 0.70 | -1.53 | 0.13 |
| PostD | -2.66 | 0.58 | -4.56 | 0.00* |
| Linear:PostD | 0.22 | 0.75 | 0.30 | 0.77 |

This analysis confirms that preschoolers did not use the uniqueness of the target in the Descriptive condition to resolve reference during the noun.

*2.2.5 RQ3: Do preschoolers process modified noun phrases more quickly when adjectives occur pre-or post-nominally?*

We restricted this analysis to those conditions in which the adjective was required for unique reference resolution, i.e., the two Contrastive conditions. We analysed proportion of looks to the target immediately after the earliest time window in which reference could be resolved in each condition. Because the adjective appears as the first lexical element in the Prenominal condition and as the second lexical element in the Postnominal condition, disambiguation can in principle occur at different time points in each condition, if contrastive inference occurs. However, since the analysis above revealed that preschoolers do not contrastively infer, we assume that they resolve reference during the second lexical element in both conditions. Thus, we investigate the effect of syntactic frame by comparing the proportion of looks to the target in each Contrastive condition during the post-utterance time window. Based on two-step models of adjective comprehension (Ninio, 2004), we expected to see a facilitatory effect of postnominal adjectives, which would manifest as a stronger preference for the target in the Postnominal condition.

Growth functions were added stepwise to the model and the overall curves were modelled with second-order OTs in addition to our fixed effect of condition. As Table 5 shows, there was no effect of condition, indicating no difference in overall looks to target between the Postnominal condition and the Prenominal condition ($\beta$ = -0.02, *SE* = 0.15, *p* = .88). However, there was an effect of condition on the quadratic term, indicating differences in the curvature between conditions ($\beta$ = -1.34, *SE* = 0.65, $p < .05$, *d* = -0.07). During the post-utterance time window, the proportion of looks to target follow a shallow curve in the Postnominal condition reflecting a slow increase in looks to the target at utterance offset, followed by a further rise towards the end of the time window. The proportion of looks to target follow deeper curve in the Prenominal condition, with a larger peak at utterance offset, followed by a decrease. Figure 6 shows elogit-transformed mean proportions of looks with GCA cubic curves.

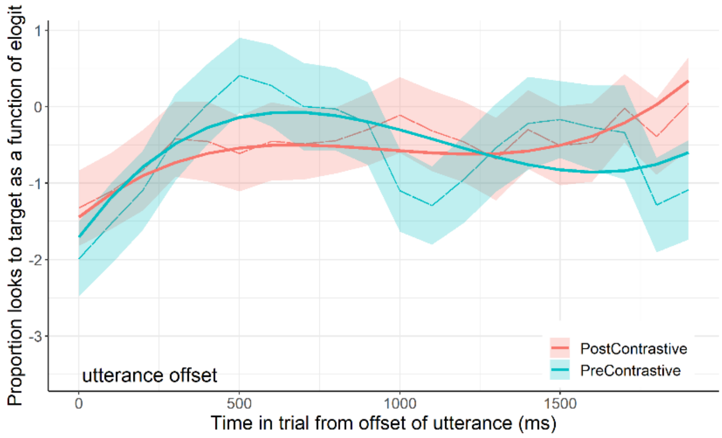


**Figure 6.** Elogit-transformed proportion of looks to target (dashed curves) in Contrastive conditions from utterance offset. Bold curves indicate cubic growth curves fitted to the data.

**Table 5.** Model summary for effect of condition on proportion of looks to target from the onset of the adjective. PreC = Prenominal Contrastive. P-values are marked with an asterisk only if critical to the analysis.

| **Term** | **Estimate** | **SE** | **t** | ***p*** |
|---|---|---|---|---|
| (Intercept) | -0.55 | 0.25 | -2.15 | 0.04 |
| Linear | 1.31 | 0.44 | 2.95 | 0.00 |
| Quadratic | 0.01 | 0.44 | 0.02 | 0.98 |
| Cubic | 0.93 | 0.44 | 2.13 | 0.03 |
| PreC | -0.02 | 0.15 | -0.15 | 0.88 |
| Linear: PreC | -1.23 | 0.66 | -1.88 | 0.06 |
| Quadratic: PreC | -1.34 | 0.65 | -2.05 | 0.04* |
| Cubic: PreC | 0.27 | 0.65 | 0.42 | 0.68 |

This analysis indicates that after hearing the entire utterance, the syntactic frame of the utterance did not influence the speed at which preschoolers processed modified noun phrases (mirroring patterns in the adult data; see Supplementary Materials). However, the significant difference in curvature suggests that despite the lack of an overall difference between conditions, there were differences in the pattern of changes in looking over time. The higher peak in the Prenominal condition suggests some integration of adjectival information from earlier in the utterance.

### *2.3 Discussion*

The results of Experiment 1 show that by the end of the referring expression, three-year-olds can integrate adjectives and nouns to resolve reference accurately in a variety of pragmatic and syntactic contexts, as hypothesised. On the whole, there is no evidence that they do so during the expressions, as the 5-year-olds and adults did in Huang and Snedeker's (2013) similar task. However, contrary to our predictions, and counter to previous research (Ninio, 2004; Sekerina & Trueswell, 2012), the presence of two types of competitor does not impose an extra processing toll relative to contexts with only one type. Instead, preschoolers only show a preference for the target at the offset of the utterance. This indicates that they do not successfully engage in incremental processing for these types of utterances and instead require all the information before settling on the target – at least when the utterance is presented at a natural speed.

On the question of contrastive inferencing, Experiment 1 shows that three-year-olds did not show an early preference for the target in the presence of a contrast set. A more subtle indication of emerging contrastive inference ability would be a consideration of the adjective competitor (the big flower) as a target in the absence of a contrast set (whereas its presence may cue them to discount the adjective competitor as a potential target). Our data did not reveal this pattern either. In line with research question 1, three-year-olds do not resolve reference until the end of the utterance, even when it is pragmatically possible to do so earlier. For utterances where the noun preceded the adjective, preschoolers did not prefer the target until they had heard the adjective, even when reference could be resolved during the noun.

However, there may be a methodological explanation for this apparent delay in processing in both syntactic frames: Because our naturalistic stimuli were presented as continuous speech, children's relatively slow reference resolution may have merely coincided with the presentation of the next lexical element. We return to this issue in Experiment 2.

On the question of whether adjectives are more helpful when presented pre- or postnominally, we find that the syntactic frame of the utterance did not influence how quickly preschoolers process modified noun phrases. This, too, though contrary to our hypothesis, is in line with the previous findings suggesting that preschoolers are not making use of the incoming information incrementally to eliminate potential targets.

Taken together, results from Experiment 1 suggest that three-year-olds adopt a wait-and-see strategy when processing modified noun phrases rather than engaging in incremental processing that recruits sophisticated pragmatic abilities. But, the naturalistic stimuli used in this experiment may not have afforded children the opportunity to show their developing incremental skills. For example, if preschoolers did in fact contrastively infer during prenominal contrastive utterances but were delayed by their generally slower processing capacity, any such ability would have been masked by the rapidly incoming noun. To detect any incremental abilities, a pause is required between the adjective and the noun. If three-year-olds can in fact contrastively infer, this should manifest during the pause. Adapting the stimuli to allow for young children's processing speed may also reveal other hidden abilities probed by the Experiment 1 analyses. Experiment 2 aimed to address these concerns.

## *3.* Experiment 2

We made multiple changes to the Experiment 1 method. Several changes were made to the audio stimuli to allow participants more time for processing. We also adjusted the visual stimuli to facilitate performance. All changes are detailed in *Materials* below. We also secured each child in a car seat during the visual world task to reduce track loss.

Experiment 2 addressed the same research questions as Experiment 1. Additionally, and due to the improved method used for the visual world task, we were able to address research question 4, which probed the skills that may drive children's emerging contrastive inferencing ability.

### *3.1 Method*

#### *3.1.1 Participants and Design*

Using the same power calculation as Experiment 1, which yielded a target sample size of 36, 40 new child participants were recruited from the same population. Three participants were excluded: two for refusing to participate and one for equipment failure. The final sample of 37 children (19 female) had a mean age of 3 years 8 months (= 44 months; range 42 – 49 months, *SD* = 2). Caregivers completed a short family questionnaire that collected information on educational background and income. Regarding the highest level of maternal[3]

[3] One caregiver in the sample was a father, so qualification refers to paternal education in this case.

education, 5% had completed high school, 54% had a Bachelor's degree, 22% had a Master's degree, and 8% had a PhD. For total household income (including all tax credits), 5% of participants were earning between £0 and £14,000, 5% between £14,001 and £24,000, 22% between £24,001 and £42,000, and 68% £42,001 or more. Overall, participants had similar demographic characteristics to the sample from Experiment 1. Participant remuneration and experimental design were the same as Experiment 1.

### *3.1.2 Materials: Visual world task*

Several adaptations were made to the stimuli used in the visual world task. To give children the opportunity to demonstrate their contrastive inferencing ability, audio manipulations were made using PRAAT (Boersma, & Weenink, 2019). New utterances were recorded by a female native speaker of English, again without pitch accent. In prenominal utterances, 500 ms of silence was inserted between the offset of the adjective and the onset of the noun, and in postnominal trials, 500 ms of silence was inserted between the offset of "that's" and the onset of the adjective. In the prenominal positions, the same token of "where's the big/where's the little" was used for every utterance, and the duration of the adjective was manually lengthened by 75%. In the postnominal positions, different tokens of "where's the NOUN that's" were used for each utterance, and the duration of the adjective (big/little) was manually lengthened by 60%[4]. Finally, the duration of all postnominal utterances were further lengthened by 10% to ensure that the prenominal and postnominal utterances were perceptually matched for speed. No manipulations were made to the fillers. In the final set of stimuli, the average utterance duration was 2621 ms ($SD$ = 110) for the prenominal trials and 3093 ms ($SD$ = 92) for the postnominal trials.

The visual stimuli were also adapted (see Figure 7 for an example). The original images were replaced with images that were more lifelike. These were grayscale drawings of familiar objects from the MultiPic repository (Duñabeitia et al., 2018). All images fitted within a 378 × 345 pixel interest area. The big images fitted tightly within this frame and were three times the size of little ones; a substantial increase in the relative size difference between the images used in Experiment 1. We also ensured that all images that shared an array were of similar real-world size, e.g., car, sofa, camel, horse (after Long et al. 2019, showing that 3-4 year-olds are slower to process images that are incongruent with their real-world size). Regarding

[4] In the original recordings, the adjectives had a shorter duration in prenominal position. Therefore, increasing the adjective duration by 60% in the postnominal utterances was enough to perceptually match its duration in the prenominal positions.

array composition, we minimised differences between contrastive and descriptive trials by only substituting targets and contingent competitors across lists and keeping all other distractors constant. For example, the first descriptive trial in list 1 contained a little fairy (target), a balloon (distractor), a crown (distractor), and a little pear (adjective competitor), and the first contrastive trial in list 2 contained a little fairy (target), a big fairy (noun competitor), a crown (distractor), and a little pear (adjective competitor).

Like Experiment 1, there were 26 trials: 16 critical items (4 in each condition), 8 filler items, and 2 practice items. All stimuli can be found at osf.io/hp9ns. Randomisation, counterbalancing, trial sequencing, and the procedure were the same as in Experiment 1.

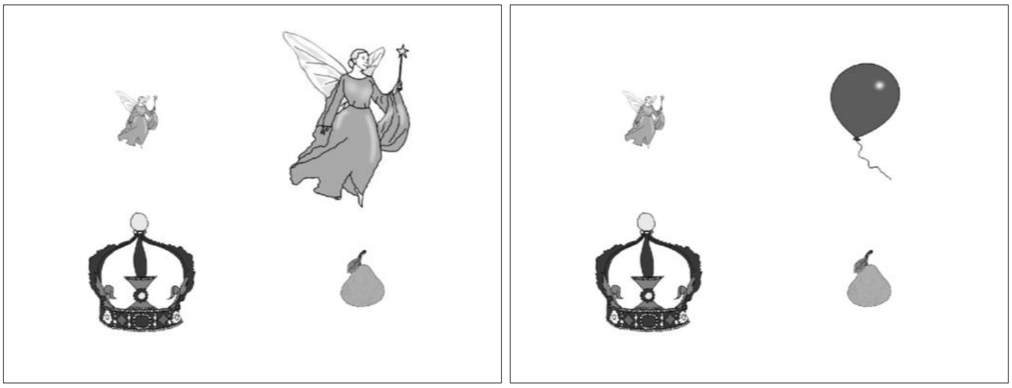

**Figure 7**. Example visual stimuli from Experiment 2 (left panel = Contrastive; right panel = Descriptive). The audio stimulus for these trials was “Click on the little fairy”.

*3.1.3 Materials: Standardised tests*

Standardised tests of language ability were administered to the participants to investigate associations between their linguistic abilities and their performance in the visual world task. We selected relevant subscales from the Clinical Evaluation of Language Fundamentals Preschool 2 UK (Wiig, Second, & Semel, 2006). First, the Language Content Index (a composite measure comprised of Expressive Vocabulary, Concepts and Following Directions, and Basic Concepts subtests) was used to measure vocabulary and semantics. Second, the Sentence Structure subtest was used to measure syntactic comprehension. We chose these measures because contrastive inferencing requires an understanding of adjective semantics as well as the ability to process multi-word utterances. The researcher administering the CELF coded the children’s responses live using the stopping rules published in the test manuals.

*3.1.4 Reliabilities*

The four CELF subtests were coded live using the protocols from the manual. A second researcher recoded participant responses from 10% of the sample using the video recording to check the reliability of the test administration and scoring. Intra-class correlation coefficients were computed along with 95% confidence intervals (CI) to assess the agreement between two raters. There was excellent absolute agreement between the two raters using the two-way mixed effects model and single rater unit, ICC = 0.96 (.954 - .969), $p < .001$.

*3.2 Results*

Prior to analysis, we excluded any trials in which the eyetracker lost track of participants' eyes on more than half of the samples per trial. This resulted in 117 exclusions out of the original 592 trials (19% of the original data set). Data preparation and analytical approach was the same as Experiment 1.

*3.2.1 RQ1a: Do preschoolers integrate adjectives and nouns to reliably resolve reference?*

As with Experiment 1, we hypothesised that three-year-olds will integrate nouns and adjectives to preferentially look at the target referent by the offset of the referring expression in all 4 conditions.

Figure 8 shows the proportion of looks to each interest area by condition. Broadly, Experiment 2 replicated the findings of Experiment 1. There is a clear preference for the target at the end of the utterance in all conditions, indicating that preschoolers integrate adjectives and nouns to correctly resolve reference in all conditions (log gaze probability ratios for the target relative to all other images from the offset of the utterance and the following 2000 ms; $M = 0.08$, $SD = 0.31$ range = -0.48 to 0.98). A linear-mixed effects regression revealed a main effect of the intercept ($\beta = .08$; $SE = .03$; $t = 2.53$; $p < .05$) indicating that preschoolers looked significantly more to the target picture than the competitors, as predicted.

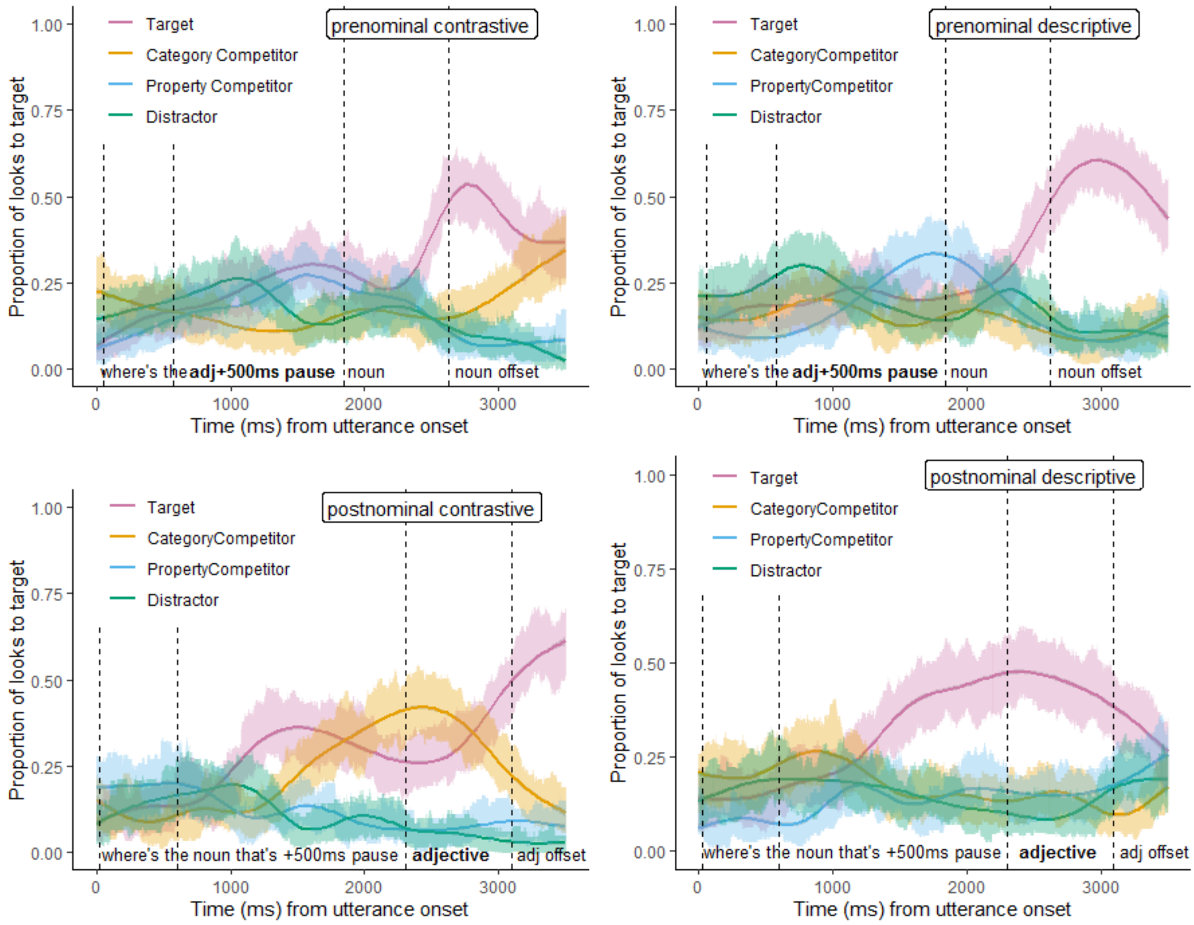


**Figure 8.** Proportion of looks (untransformed) to each interest area in each condition. Vertical dashed lines represent mean onset times. Bold text annotations indicate disambiguation points.

### *3.2.2 RQ1b. Where the adjective appears before the noun, to what extent does the presence of an adjective and a noun competitor threaten reference resolution?*

We pursued the patterns shown in Figure 8 to reveal whether the presence of an adjective and a noun competitor weakens target preference in Prenominal conditions. Recall that if target preference is weaker in the Contrastive condition, this is likely to reflect the additive effect of two types of competitor. We originally hypothesised that this would be the case, though this was not found in Experiment 1.

As in Experiment 1, the analysed time window runs from the onset of the noun and for the following 2000 ms, which captures the offset of the expression plus a period of silence. The mean duration of the noun was 779 ms (*SD* = 124, range 505 - 946). Growth functions were added stepwise to the model and the overall curves were modelled with fourth-order OTs in addition to our fixed effect of condition. Table 6 shows the fixed effect parameter estimates and their standard errors along with *p*-values estimated using the normal approximation for the t-values. There was an effect of condition indicating higher overall target fixation proportions for the Descriptive condition relative to the Contrastive condition (β = 0.41, *SE* = 0.12, $p < .01$, *d* = 0.10). This is complemented by a significant effect of condition on the linear term (β = 0.41, *SE* = 0.12, $p < .01$, *d* = 0.12), confirming a steeper linear climb in the Descriptive condition relative to the Contrastive condition. Figure 9 shows elogit-transformed mean proportions of looks with GCA cubic curves.

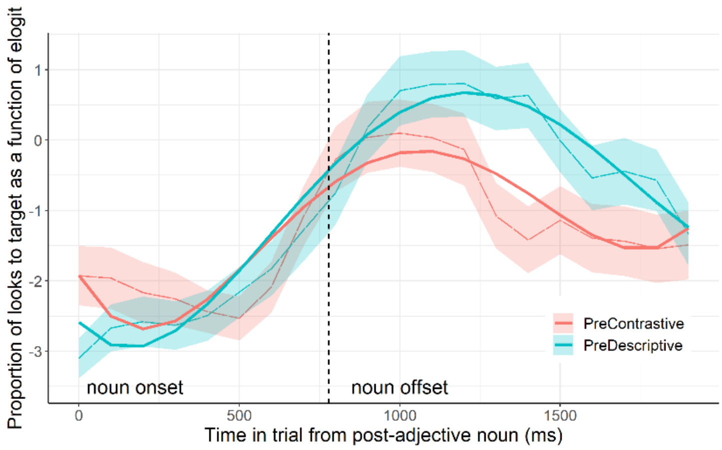


**Figure 9.** Elogit-transformed proportion of looks to target (dashed curves) in Prenominal conditions from the onset of the noun. Bold curves indicate quartic growth curves fitted to the data.

**Table 6.** Model summary for effect of condition on proportion of looks to target from the onset of the adjective. PreD = Prenominal Descriptive. P-values are marked with an asterisk only if critical to the analysis.

| Term | Estimate | SE | t | *p* |
|---|---|---|---|---|
| (Intercept) | -1.28 | 0.24 | -5.39 | 0.00 |
| Linear | 1.88 | 0.39 | 4.86 | 0.00 |
| Quadratic | -2.37 | 0.38 | -6.21 | 0.00 |
| Cubic | -0.81 | 0.38 | -2.13 | 0.03 |
| Quartic | 1.77 | 0.38 | 4.64 | 0.00 |
| PreD | 0.41 | 0.12 | 3.36 | 0.00* |
| Linear: PreD | 2.07 | 0.54 | 3.81 | 0.00* |
| Quadratic: PreD | -1.04 | 0.54 | -1.93 | 0.05 |
| Cubic: PreD | -0.97 | 0.54 | -1.81 | 0.07 |
| Quartic: PreD | -0.72 | 0.54 | -1.33 | 0.18 |

Unlike in Experiment 1, the difference between conditions indicates that preschoolers showed a weaker preference for the target when there was a noun competitor object in addition to an adjective competitor in the display, as originally hypothesised. But was this because they were additionally distracted by the adjective competitor in the Contrastive condition (which was the only other object that would fit the unfolding utterance), or did they simply find this condition more taxing and so they spent less time looking? If the former, the difference between conditions should hold if looks to the adjective competitor are factored into the dependent variable. If the latter, the difference between conditions should reduce.

To investigate the source of competition away from the target, we analysed the effect of condition on the proportion of looks to the target minus looks to the adjective competitor (target advantage) during the same time window as the preceding analysis. Since GCA requires a binary outcome variable, when modelling target advantage we included only the samples when participants were looking at the target or the relevant competitor. Growth functions were added stepwise to the model and the overall curves were modelled with fourth-order OTs in addition to our fixed effect of condition. As Table 7 shows, there was no effect of condition, indicating no difference in target advantage scores for the Descriptive and the Contrastive conditions (β = -0.27, *SE* = 0.14, *p* = .06). This suggests that preschoolers were not specifically distracted by the adjective competitor in the Contrastive condition. The significant effect of condition on the quadratic term, indicating a brief target advantage in the Contrastive condition at the very beginning of the time window (β = -2.04, *SE* = 0.62, *p* < .01, *d* = -0.12) is likely to reflect processing from earlier in the utterance. Figure 10 shows elogit-transformed target advantage scores with GCA quartic curves.

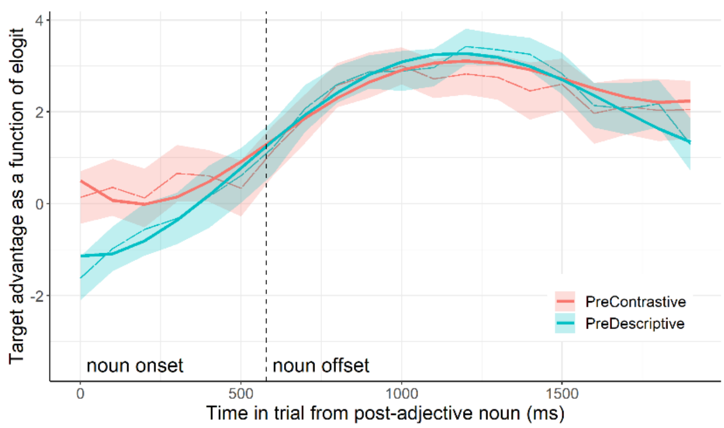

**Figure 10.** Elogit-transformed target advantage scores (dashed curves) in Prenominal conditions from the onset of the noun. Bold curves indicate quartic growth curves fitted to the data.

**Table 7.** Model summary for effect of condition on target advantage scores from the onset of the adjective. PreD = Prenominal Descriptive.

| **Term** | **Estimate** | **SE** | **t** | ***p*** |
|---|---|---|---|---|
| (Intercept) | -1.28 | 0.24 | -5.39 | 0.00 |
| Linear | 1.88 | 0.39 | 4.86 | 0.00 |
| Quadratic | -2.37 | 0.38 | -6.21 | 0.00 |
| Cubic | -0.81 | 0.38 | -2.13 | 0.03 |
| Quartic | 1.77 | 0.38 | 4.64 | 0.00 |
| PreD | 0.41 | 0.12 | 3.36 | 0.00* |
| Linear: PreD | 2.07 | 0.54 | 3.81 | 0.00* |
| Quadratic: PreD | -1.04 | 0.54 | -1.93 | 0.05 |
| Cubic: PreD | -0.97 | 0.54 | -1.81 | 0.07 |
| Quartic: PreD | -0.72 | 0.54 | -1.33 | 0.18 |

The lack of difference between conditions in this target advantage analysis (which incorporated looks to a specific competitor) indicates that the effect found in the proportion of looks analysis was not due to the inclusion of the adjective competitor. The Prenominal Contrastive panel in Figure 8 suggests that distraction instead stemmed from the noun competitor at the end of the utterance, potentially because they were scanning between the big and little contrast set to check the size difference.

*3.2.3 RQ2: Do preschoolers show contrastive inference?*

*Prenominal conditions*

As shown in Figure 8, the emergence of target preference at noun offset in the Prenominal Contrastive condition suggests that despite amending our stimuli to help children show latent ability in contrastive inference, this aspect of pragmatic development still appears to be beyond their grasp. However, to capture what are likely to be subtle effects in this age group, we ran a fine-grained analysis of by-condition differences during the prenominal adjective.

To investigate the strength of competition away from the target, we analysed the effect of condition on proportion of looks to the target minus looks to the adjective competitor (target advantage) during the adjective + pause window. The mean duration of this window was

1266 ms ($SD$ = 69, range 1194 - 1365). If children used the presence of a contrast set to infer that the adjective refers to one of its members, they should show fewer looks to the adjective competitor, and thus a stronger target advantage in the Contrastive condition. Growth functions were added stepwise to the model and the overall curves were modelled with first-order OTs in addition to our fixed effect of condition. As Table 8 shows, there was an effect of condition, indicating higher target advantage for the Contrastive condition ($\beta$ = -0.51, $SE$ = 0.22, $p$ < .05, $d$ = -0.12). There was also a significant interaction between the linear term and condition ($\beta$ = -3.02, $SE$ = 0.83, $p$ < .01, $d$ = -0.19). In the Descriptive condition, there was a linear decline in target advantage from adjective onset, whereas in the Contrastive condition target advantage remained linearly stable. Figure 10 shows elogit-transformed target advantage scores with GCA linear curves.

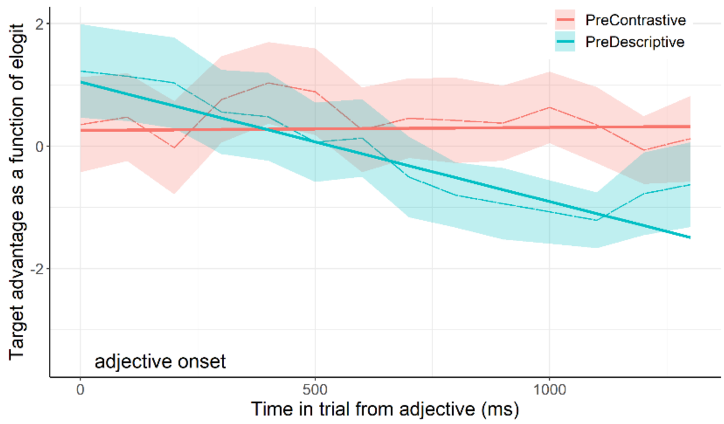


**Figure 11.** Elogit-transformed target advantage scores (dashed curves) in Prenominal conditions from the onset of the adjective. Bold curves indicate linear growth curves fitted to the data.

**Table 8.** Model summary for effect of condition on target advantage scores during the adjective. PreD = Prenominal Descriptive.

| **Term** | **Estimate** | **SE** | **T** | ***p*** | ***d*** |
|---|---|---|---|---|---|
| (Intercept) | 0.29 | 0.54 | 0.52 | 0.60 | NA |
| Linear | 0.07 | 0.57 | 0.13 | 0.90 | 0.01 |
| PreD | -0.51 | 0.22 | -2.36 | 0.02* | -0.12 |
| Linear:PreD | -3.02 | 0.83 | -3.66 | 0.00* | -0.19 |

Unlike Experiment 1, these results confirm that preschoolers showed a stronger preference for the target during the adjective in the Contrastive condition relative to the Descriptive condition. In the Descriptive condition, the decline in target advantage (and the corresponding rise in distractor advantage) evidences a greater distraction from the adjective competitor.

*Postnominal conditions*

Here we ask whether participants use the presence of the singleton object in the Descriptive condition to infer that no adjective is needed, and look at the target before the adjective is heard. If this is the case, they should show an earlier target advantage in the Descriptive than in the Contrastive condition.

Growth functions were added stepwise to the model and the overall curves were modelled with second-order OTs in addition to our fixed effect of condition. As Table 9 shows, there was a marginally significant effect of condition indicating higher overall target advantage for the Descriptive condition relative to the Contrastive condition (β = 0.34, *SE* = 0.17, *p* = .05, *d* = 0.08). There was also a significant interaction between the linear term and condition (β = 7.94, *SE* = 0.76, *p* < .01, *d* = 0.44), indicating different trajectories in preference across condition. In the Descriptive condition, target advantage followed a curved incline whereas in the Contrastive condition it showed a curved decline. Figure 12 shows elogit-transformed target advantage scores with GCA quadratic curves.

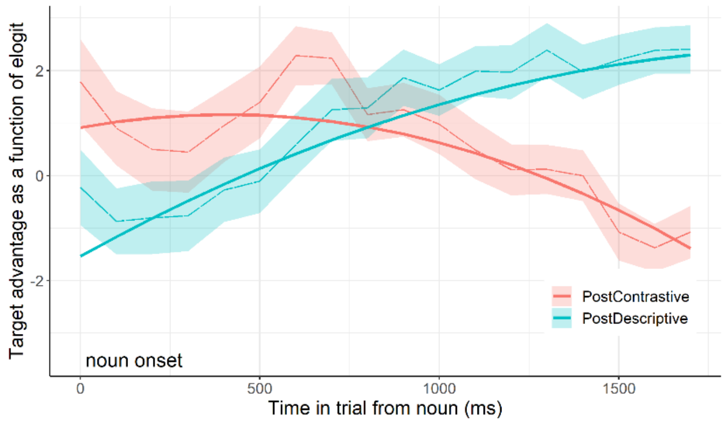


**Figure 12.** Elogit-transformed target advantage scores (dashed curves) in Postnominal conditions from the onset of the noun. Bold curves indicate quadratic growth curves fitted to the data.

**Table 9.** Model summary for effect of condition on target advantage scores during the noun. PostD = Postnominal Descriptive.

| Term | Estimate | SE | t | *p* |
|---|---|---|---|---|
| (Intercept) | 0.45 | 0.39 | 1.14 | 0.27 |
| Linear | -2.98 | 0.85 | -3.49 | 0.00 |
| Quadratic | -1.54 | 0.52 | -2.98 | 0.09 |
| PostD | 0.34 | 0.17 | 1.98 | 0.05† |
| Linear:PostD | 7.94 | 0.76 | 10.45 | 0.00* |
| Quadratic:PostD | 0.63 | 0.71 | 0.88 | 0.38 |

Unlike Experiment 1, these results indicate that preschoolers used the uniqueness of the target in the Descriptive condition to resolve reference during the noun + relative pronoun window. This is supported by the early target looks in the Postnominal Descriptive panel in Figure 8.

*3.2.4 RQ3: Do preschoolers process modified noun phrases more quickly when adjectives occur pre-or post-nominally?*

As in Experiment 1, we restricted this analysis to the two Contrastive conditions where the adjective is required for reference resolution. Thus, we investigate the effect of syntactic frame by comparing the proportion of looks to the target in each the Contrastive conditions, during and after the noun at the end of the utterance. Again, we expected to see a facilitatory effect of postnominal adjectives, manifest as a stronger preference for the target.

Growth functions were added stepwise to the model and the overall curves were modelled with second-order OTs in addition to our fixed effect of condition. As Table 10 shows, there was no effect of condition, indicating no difference in overall looks to target across condition ($\beta = -0.26$, $SE = 0.26$, $p = .32$). This indicates that after the effective disambiguation time window, the syntactic frame of the utterance did not influence whether preschoolers processed modified noun phrases more quickly. However, there was an effect of condition on the linear term, indicating differences in the slope between conditions ($\beta = 2.40$, $SE = 0.64$, $p < .01$, $d = 0.13$). This is complemented by a significant effect of condition on the quadratic term, indicating differences in curvature between conditions ($\beta = 1.44$, $SE = 0.63$, $p < .05$, $d = 0.08$). During the post-utterance time window, the proportion of looks to target follow a shallow U-shaped curve in the Prenominal condition, first declining then increasing slightly towards the end of the time window. In the Postnominal condition, proportion of looks followed a downwards linear slope (but note that the slope did not capture the brief increase in looks to target that begins around 500 ms). Figure 13 shows elogit-transformed mean proportions of looks with GCA quadratic curves.

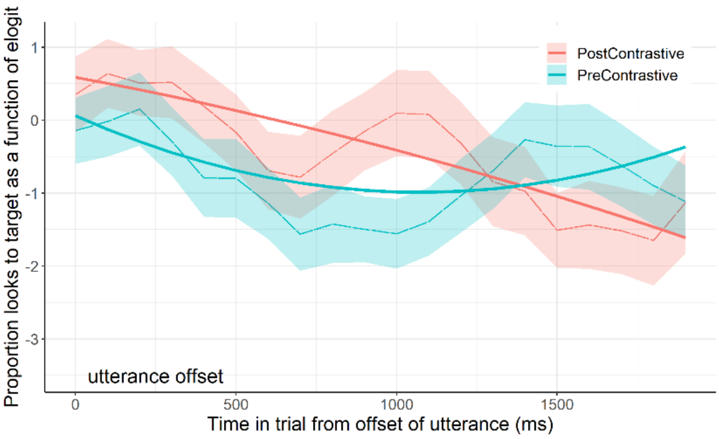


**Figure 13.** Empirical logit-transformed proportion of looks to target (dashed curves) in Contrastive conditions from offset of the utterance. Bold curves indicate quadratic growth curves fitted to the data.

**Table 10.** Model summary for effect of condition on proportion of looks to target from the onset of the adjective. PreC = Prenominal Contrastive

| Term | Estimate | SE | t | *p* |
|---|---|---|---|---|
| (Intercept) | -0.41 | 0.29 | -1.38 | 0.18 |
| Linear | -2.98 | 0.49 | -6.08 | 0.00 |
| Quadratic | -0.23 | 0.48 | -0.48 | 0.63 |
| PreC | -0.26 | 0.26 | -1.02 | 0.32 |
| Linear: PreC | 2.40 | 0.64 | 3.75 | 0.00* |
| Quadratic: PreC | 1.44 | 0.63 | 2.29 | 0.02* |

Like Experiment 1, this analysis indicates that after hearing the entire utterance, the syntactic frame of the utterance did not influence the speed at which preschoolers processed modified noun phrases. The differences in curvature in Figure 13 suggest that once reference has been resolved, children may start to look around the screen at the other items in the array, for example the noun competitor. This may be more common in the Postnominal condition where they have more recently received the disambiguating information.

#### *3.2.5 RQ4: Is there an association between preschoolers' language ability, their speed of processing, and their contrastive inferencing ability?*

To address this research question, we analysed correlations between a) measures of contrastive inferencing and language and b) measures of contrastive inferencing and speed of processing. We hypothesised that children who showed contrastive inferencing will score higher on measures of semantics and syntax, and a show a faster speed of processing.

To measure language ability, recall that we used two subscales from the Clinical Evaluation of Language Fundamentals Preschool 2 UK (Wiig et al., 2006). From the Language Content composite measure, we used the sum of scaled scores from each subtask ($M = 103$, $SD = 12$, range 61-120). From the Sentence Structure we used the scaled score ($M = 12$, $SD = 2$, range 6-16).

To measure speed of processing (SoP), we used responses to the filler trials ($N = 8$ per participant). Recall that in these trials, the target image was always a singleton object within the array, and utterances contained an unmodified noun, e.g., "where's the bus?" SoP was defined as the average latency from noun onset to first valid fixation to the target (>100 ms). The critical time window ran from 300 ms after the onset of the noun to its offset. Trials were included only if the participant was not already fixating the target prior to the time window of interest. After exclusions, 203 trials went forward for analysis (69% of the original dataset). The mean latency to fixate the target was 1003 ms ($SD = 196$, range 665-1505). Latencies were log-transformed prior to analysis to remove some of the skewness in the data (Baayen & Milin, 2010).

To measure contrastive inferencing, we used two measures of looks to target during the adjective for the Prenominal Contrastive condition ($N = 4$ per participant). First, proportion of looks to target from the onset of the adjective (+300 ms) to its offset. This indexed the strength of preference for the target. It is derived by dividing the number of samples that fall in the target interest area by the number of samples that fall elsewhere, i.e., the other three predefined interest areas, onscreen looks outside of the predefined interest areas, and off-screen looks, giving a value between 0 and 1. This was then averaged over trials, giving each participant a mean score for proportion of looks to target. The group mean proportion of looks to target was .26 ($SD = .24$, range 0-1). The second measure was the onset of the first valid fixation (>100 ms) to the target from the onset of the adjective (+300 ms) to its offset. This is equivalent to mean reaction time to look at the target. This indexed the speed of preference for the target. These reaction times were averaged over trials, giving each participant a mean RT for first fixation to target. The group mean RT to fixate the target was 1035 ms (*SD* = 110, range 793-1310).

As Figure 13 shows, neither of the language measures, nor speed of processing significantly correlated with either of the contrastive inference measures.

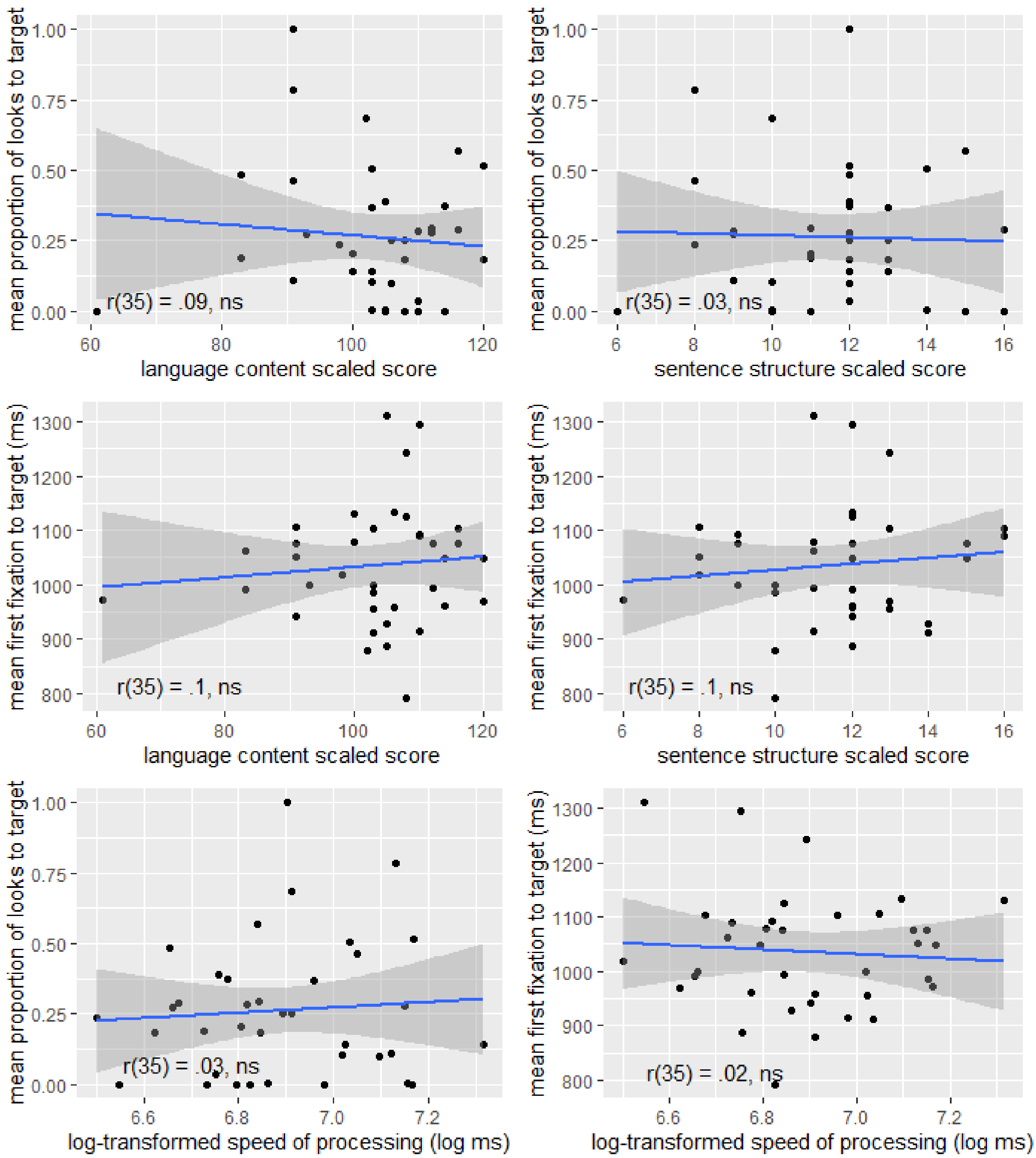


**Figure 13.** Scatterplots showing associations between contrastive inference and both language measures, and contrastive inference and speed of processing.

*3.3 Discussion*

Stimuli used in Experiment 2 were designed to enable three-year-olds to show their latent abilities in online adjective-noun comprehension that may have been masked by those used in Experiment 1. Overall, they fulfilled this aim. While some of the abilities evidenced by Experiment 1 were replicated by Experiment 2, some new abilities were evidenced, and some remained out of preschoolers' reach.

As in Experiment 1, children showed that they can resolve reference accurately by integrating adjectives and nouns by the end of the referring expression to in a variety of pragmatic and syntactic contexts. Unlike Experiment 1, they showed a weaker preference for the target when there was both an adjective and a noun competitor in the display. However, this was not due to increased competition from the object that shared a property with the target, but due to post-utterance checking of the target's contrast mate. This was afforded by the slower speed of presentation and/or the enhanced size differences in this experiment.

Experiment 2 elicited emerging evidence of contrastive inferencing in three-year-olds. Unlike Experiment 1, preschoolers showed a preference for the target during the adjective when a contrast set was present (Figure 11). Children also show greater distraction from the adjective competitor where no contrast set is present, suggesting that they use the absence of the set to infer that the adjective is likely to have a descriptive rather than contrastive function. These results support two manifestations of contrastive inference, facilitated by slower presentation of the stimuli, the pause between the adjective and the noun, and clearer size distinctions.

Where the adjective appears postnominally, preschoolers used the uniqueness of the target in the Descriptive condition to resolve reference early, i.e., during the noun + relative pronoun window. This is enabled by the slower speed of presentation, the postnominal pause, or the clearer visual size contrast, and is in stark contrast to Experiment 1 in which preschoolers waited to hear the final element of the utterance (Figures 2 and 8; postnominal descriptive panels).

On the question of whether adjectives are more helpful when presented pre- or postnominally, Experiment 2 replicated the findings of Experiment 1. After hearing the entire utterance – even at a reduced speed - the syntactic frame of the utterance did not influence the speed at which preschoolers processed modified noun phrases.

Our novel correlational analysis found no relationship between contrastive inferencing and language or speed of processing. This is despite a good range of abilities in inferencing in our sample. On the basis of this data we conclude that inferencing is not supported by language or by processing speed, as measured using these particular instruments.

Taken together, results from Experiment 2 suggest that given the opportunity through slower, clearer stimulus materials, three-year-olds can use their developing pragmatic skills to integrate visual and auditory information and incrementally process modified utterances.

## 4. General Discussion

Previous studies of three-year-olds' ability to use adjectival information in resolving reference have relied on either 'end-point' offline data, or have analysed online behaviour in response to very simple or scaffolded displays where integration with nouns has been unnecessary. Our experiments have taken a comprehensive, rigorous approach by analysing high-resolution online data in response to stimuli that demand full integration of both lexical elements, in younger children than have been tested previously. Taken together, results show that like adults, children use multiple sources of information to interpret language in real time.

### *4.1 Summary of results*

Table 11 summarises the main findings from experiments 1 and 2. Both experiments centred on a visual world task. The first used naturalistic audio stimuli; the second adapted these so that they were presented more slowly and contained pauses. This allowed children to demonstrate their latent ability in contrastive inferencing. Both experiments revealed that three-year-olds were able to integrate adjectives and nouns to resolve reference accurately by the end of the referring expression. Experiment 1 showed that the presence of both a noun and an adjective competitor did not reduce target preference. Although Experiment 2 elicited a reduction in target preference in the presence of two distractors, this was not until after reference had been resolved, so we take it to represent post-utterance checking rather than compromised online processing. Thus, we conclude that children can resolve reference accurately in a variety of pragmatic and syntactic contexts, and in the presence of multiple distractors.

Experiment 2 revealed for the first time that three-year-olds are able to contrastively infer. They showed a stronger preference for the target during the adjective in contrast displays than in non-contrast displays, and greater attention on the adjective competitor in the latter. When the adjective occurred postnominally, they were able to use the presence of a singleton object to infer during the noun that no adjective is needed, and showed early reference resolution. None of these effects were found in Experiment 1, suggesting that young children need ample time to demonstrate these sophisticated online skills.

In both experiments, children processed modified noun phrases equally quickly regardless of adjective position, like adults (see Supplementary Materials A). Finally, children's skills in language and speed of processing do not appear to be linked to their contrastive inferencing abilities.

**Table 11**. Summary of results.

| RQ | Effect | Exp. 1 | Exp. 2 |
| --- | --- | --- | --- |
| 1 | Integration of adjectives and nouns to resolve reference by the offset a referring expression. | ✓ | ✓ |
| | Reduced target preference in the presence of two types of distractors. | ✗ | ✓ |
| 2 | Contrastive inference, manifest as: - stronger preference for the target during the adjective in contrast displays; | ✗ | ✓ |
| | - greater distraction from the adjective competitor in non-contrast displays. | ✗ | ✓ |
| | In postnominal frames, use of a singleton object to infer during the noun that no adjective is needed. | ✗ | ✓ |
| 3 | Quicker reference resolution when adjectives occur post-nominally than prenominally. | ✗ | ✗ |
| 4 | Relationship between contrastive inferencing and language or speed of processing | NA | ✗ |

*4.2 Children's integration of adjectives and nouns*

We have evidenced that integration of noun and adjective information is achievable by preschoolers when faced with 4-referent displays. This finding is more robust than conclusions made on the basis of previous research with this age group, 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). Our finding that the presence of multiple competitors does not jeopardise reference resolution suggests that by three years of age, children are neither distracted by nor over-rely on information from either the noun or the adjective, reflecting integration, contrary to what has been found in slightly younger children (Ninio, 2004; Thorpe et al., 2006), as well as older children (Nadig et al., 2003, exp. 2). Further, once the noun has been presented and reference has been resolved, preschoolers look away from the target towards the noun competitor, sensibly checking their choice against the contrast mate (replicating behaviour found in 5-6 year-olds by Nadig et al., 2003). Studies examining online processing can therefore shed light on *when* children show evidence of resolving reference, even if their overall looking time at the target over a long time window, or the final location on which their gaze lands, do not necessarily suggest that the child prefers the target.

### *4.3 Contrastive inference*

Whether a real developmental limitation, or a methodological flaw conflating late contrastive inferencing with noun processing, previous research with three-year-olds (Fernald et al., 2010; Sekerina & Trueswell 2012; Thorpe et al., 2006) suggested that preschoolers listened through prenominal material and waited for the noun before they fixated the target object. In contrast, our results show that this age group can in fact deduce the informativeness of the adjective online then use it in incremental adjective interpretation.

Crucially, the insertion of a pause between the adjective and the noun in Experiment 2 allowed children to demonstrate their emerging skills in contrastive inference. Cognitively, it allowed them more time to process adjective information. Methodologically, it allowed us to separate the point at which contrastive inferences manifest from the point at which the (delayed) noun is presented. In Experiment 1, these points coincided, leaving two alternate conclusions available: a) preschoolers are unable to generate contrastive inferences, or b) they do contrastively infer, though more slowly than adults, during the noun. For this reason, it was not possible to state unequivocally that three-year-olds fail to contrastively infer when presented with naturally-paced utterances. Assuming that their apparent failure is due to cognitive rather than methodological limitations, here we explore reasons for why they can only demonstrate contrastive inference when time allows.

On the basis of the age of acquisition of the vocabulary used in our stimuli, we argue that preschoolers' lexical processing is sufficient to complete the task, but when faced with the onslaught of incoming information they are unable to deploy their developing pragmatic skills to infer the speaker's meaning. It is entirely feasible that coordinating their lexical knowledge of familiar adjectives and nouns with the pragmatic demands of incorporating referential context and inferential reasoning while processing continuous speech and building representations, is beyond the reach of this age group. However, our findings suggest that extra time and clear standards of comparison facilitate their fragile developing abilities, enabling them to coordinate lexical and referential processing, and recruit pragmatic information incrementally. Thus we provide evidence for developmental continuity in contrastive inference: three-year-olds possess the knowledge and skills required, but due to processing limitations are not typically able to demonstrate it.

Why does extra time allow young children to show contrastive inferencing? During this kind of processing, comprehenders must coordinate strong bottom-up constraints from the auditory signal with top-down, resource-heavy referential constraints (Dell, 1986). Identifying the pragmatic implications of using an adjective to refer to a singleton item then feeding this knowledge into a referential choice takes time. Without this time (as in naturalistic speech), young children are likely to resort to the simpler, unambiguous bottom-up signal from the postadjectival noun to resolve reference at the end of the utterance. Our study has highlighted that given that extra time, young children can engage in the necessary processing, and that this ability may not have been absent but merely delayed in previous findings. The facilitatory effect of slow speech has also been shown in a recent study suggesting that contrary to classic findings (Trueswell et al., 1999), five-year-olds can recruit referential information to guide syntactic parsing if given time to do so (Qi et al., in press).

Our findings on slow speech (and potentially also clearer visual differences) have important implications for both research and practice. We would expect slower speech to facilitate online referential processing as well as word learning (following Fernald et al., 2008 and He et al., 2020). In previous work (Davies et al., 2020), we endorsed therapeutic materials that emphasised contrast for children who struggle with adjectives. With the benefit of our current findings, we would also recommend slow speech to further scaffold their learning.

Robust comparison skills are central to our task. As pointed out by Huang and Snedeker (2013: 1100), most semantic theories propose that we must establish a standard of comparison to determine what counts as having a certain property in a given context (Barner et al., 2009; Barner & Snedeker, 2008; Bierwisch, 1987; Kamp & Partee, 1995; Kennedy, 1999). In our experiments, participants would have used the saliently contrasting image in the scene to do this *(…that cow is big relative to the other one).* The fact that preschoolers only showed contrastive inferencing in Experiment 2 where size contrasts were greater suggests that a clear standard of comparison is key for children still developing their pragmatic system.

Methodology clearly plays a role in the contrasting findings of Experiments 1 and 2. A recent research programme in pragmatic development has elicited latent abilities in a variety of pragmatic phenomena in younger children than initially documented, once task demands have been reduced (Berger & Höhle, 2012; Falkum et al., 2016; Pouscoulous et al., 2007; Pouscoulous & Tomasello, 2019; Schulze et al., 2013; Stiller, et al., 2015). In adjective processing, Syrett et al. (2019) showed that although 36-month-olds failed to recruit known adjectival semantics in a passive online task to select an animate referent, they were successful in a more interactive, offline version of the task without time pressure. Our study illustrates once again that when young children are given tasks that incorporate sufficient time to deploy higher level reasoning skills, they are able to engage in highly sophisticated language processing.

Considering the striking preponderance of adjectives that function descriptively relative to those that function contrastively in child-directed speech (94% descriptive vs. 6% contrastive documented in a large corpus; Davies et al., 2020), it is all the more impressive that young children could readily infer a contrast function of adjectives in our task. This casts doubt on the requirement for high-frequency models in the input. Indeed, although correlations between adjective frequencies in child speech and child-directed speech are strong for 18-month-olds, this relationship decreases over time as children develop independent adjective use (Tribushinina et al., 2014). Instead, contrastive inference may be driven by a more domain-general reasoning process.

### *4.4 Comprehension of prenominal and postnominal adjectives*

This study was partly motivated by a puzzle proposed by Davies et al. (2020). Because of the need to calibrate adjectives to the nouns they modify, it was hypothesised that prenominal adjectives are more challenging for children. However, in a corpus of child-directed speech, prenominal adjectives were found to occur more often. Why should the more challenging forms occur more frequently in the input? The current study offers an answer to that paradox: prenominals are in fact no more taxing than postnominals.

Against our hypothesis, the two-step strategy enabled by postnominal frames did not help preschoolers. Although homing in on the nominal class before using the adjective to disambiguate the target seems like an appealing strategy, prenominal and postnominal frames were processed equally quickly in both experiments and by both populations. We propose several possible explanations for this. First, any postnominal advantage may have been cancelled by emerging (in Expt. 1) or secure (Expt. 2) contrastive inferencing abilities, which would facilitate earlier reference resolution in prenominal frames. Second, adjectives in prenominal position are said to bias towards a contrastive meaning (Diesendruck, Hall, & Graham, 2006; Prasada, 1992; Prasada & Cummins, 2000). Relatedly, subsective adjectives like "big" and "little" may have slightly different interpretations in prenominal position and postnominal position (Higginbotham, 1995). That is, saying that an object is big (predicative use) may leave open the standard of comparison; the phrase, "big cow" denotes a cow that is big for a cow, while the phrase "cow that's big" may be ambiguous as to what the cow is big relative to. Given that our analysis of syntactic frame was restricted to the contrastive conditions, prenominal utterances may have facilitated reference resolution to the exact pragmatic function that the children faced. This bias may have worked against the postnominal utterances. Third, postnominal frames may only be helpful in challenging contexts, e.g., long-distance linguistic dependencies or visually complex scenes. When an array is cluttered with many competitors and distractors, focusing on the nominal class should allow comprehenders to usefully rule out a number of these on their way to resolving reference (Gatt et al., 2012; Rubio-Fernández, 2016). In our simple array of four objects, this strategy is unlikely to apply. Fourth, in our (English) stimuli, the postnominal information was part of a relative clause, a late-developing syntactic construction. Taken together, developing contrastive inference skills, a prenominal contrastive bias, the simple displays, and the more complex postnominal syntax may have masked a postnominal advantage that may manifest under different conditions. Alternatively, considering the adjective-noun integration evidenced in the experiments, perhaps preschoolers simply have no problem keeping adjectival information in mind until the noun information is presented, especially as the noun follows rapidly from the adjective. Crosslinguistic studies that capitalise on prenominal/postnominal alternates of comparable complexity have the potential to add to this body of evidence (see Rubio-Fernández, Mollica, & Jara-Ettinger, 2021).

*4.5 Drivers of contrastive inferencing ability*

In an attempt to reveal what matures in the child to enable them to integrate adjectives and nouns online, we measured semantic and syntactic comprehension and speed of processing in our sample. None of these measures significantly correlated with contrastive inferencing, casting doubt on the hypothesis that proficiency in lexical knowledge and processing speed is required. Coupled with the fact that children only needed a firm understanding of the adjectives “big” and “little” (which should be strongly represented in three-year-olds), we rule out lexical knowledge as a driver of contrastive inferencing. Our finding that preschoolers (i.e., children with relatively little language experience) can pass complex pragmatic tasks (given time) suggests that sophisticated language abilities may be less important for this kind of inferential processing. Although we did not find a significant correlation between speed of processing and contrastive inferencing, the fact that preschoolers demonstrate contrastive inferencing in Experiment 2 suggests that time for processing is a key component. Further, Fernald et al. (2009) found that processing speed differences within their 30-month-olds were associated with differential success in online interpretation of adjective–noun phrases. It may be that the way that we measured processing speed (i.e., the speed at which children shifted to the correct target in a lexical task) was too narrow. Instead, a broader conception of processing speed which encompasses the psycholinguistic and neural mechanics of pragmatic processing (as probed in our contrastive inference task) may be a more appropriate measure. Indeed, a recent study on the relationship between processing speed, vocabulary size, and subsequent vocabulary growth reveals a complex, dynamic, and variable interaction (Peter et al., 2019; see also Koenig, Arunachalam, & Saudino, 2020). Future work investigating the drivers of pragmatic inferencing should take this complexity into account.

The aspects of cognition that we analysed were of course just a subset of a wider range of skills that may be relevant for contrastive inference, separately or in combination. For example, semantic short-term memory may be implicated in the need to store and manipulate adjectival information during the processing of referring expressions (Hanten & Martin, 2000; Martin & He, 2004). This is likely to be related to other examples of combinatorial language processing. Future work could compare adjective-noun integration with the integration of linguistic units that rely less on pragmatics to reveal the relative contribution of linguistic and pragmatic skills in generating contrastive inferences.

*4.6 Future directions*

Several questions remain as a result of our chosen methodology. Our design cannot definitely answer the question of whether or not three-year-olds were contrastively inferring (albeit slowly) in Experiment 1. All we know at this point is that when time is provided to demonstrate / measure it, and key visual and audio manipulations were made to the stimuli, contrastive inference manifests in this age group. Future experiments should separate speed of presentation, and size/speed of stimuli to identify their relative contribution. To ascertain the generalisability of our findings, we would like to extend this paradigm to adjectives with different semantics, e.g., colour, height, (cf. Jincho et al., 2019), or those that are less polar, or imageable. “Big” and “small” have served as a useful starting point for testing contrastive inferencing in such young children due to their familiarity and their strong links to multiple-referent contexts in the child’s language experience (Huang & Snedeker, 2013). However, it is possible that more challenging adjectives might elicit different patterns of results with respect to inferencing, pre- vs. post-nominal performance, or correlations with language ability. Relatedly, we would welcome studies that test our findings in less controlled environments, e.g., during shared reading or free play. Because lab-based processing in preschoolers correlates well with vocabulary (Koenig et al., 2020), we would expect our results to generalise as long as the extra time affordances were retained.

*4.7 Conclusion*

Findings from two experiments provide evidence that children’s interpretation of adjective-noun combinations is integrated, and informed by multiple information sources recruited online. Unlike previous research, we provide evidence of a continuity in children’s development of sophisticated, adult-like pragmatic skills. Critically, we found that three-year-olds understand that modification is expected in the presence of multiple referents of the same class, and are able to apply this principle during referential processing, when given the time to do so.

## References


Arunachalam, S. (2016) A new experimental paradigm to study children’s processing of their parent’s unscripted language input. *Journal of Memory and Language*, 88, 104-116. doi: 10.1016/j.jml.2016.02.001

Baayen, R. H., and Milin, P (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3.2*, 12-28. doi: 10.21500/20112084.807.

Barner D. and Snedeker J. (2008). Compositionality and statistics in adjective acquisition: 4-year-olds interpret tall and short based on the size distributions of novel noun referents. *Child Development*, 79(3), 594-608. doi: 10.1111/j.1467-8624.2008.01145.x

Barner, D., Goodman, N., Schmidt, L., and Tenenbaum, J. (2009). How tall is tall? Compositionality, statistics, and gradable adjectives. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31.

Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474. doi: 10.1016/j.jml.2007.09.002

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. doi: 10.18637/jss.v067.i01.

Bergelson, E., and Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*, 3253–3258.

Berger, F., and Höhle, B. (2012). Restrictions on addition: children's interpretation of the focus particles auch (‘also’) and nur (‘only’) in German. *Journal of Child Language*. 39, 383-410.

Berman, (1988). On the ability to relate events in narrative, *Discourse Processes 11*, 469-497. doi: 10.1080/01638538809544714

Bierwisch, M. (1987). The semantics of gradation. In M. Bierwisch and E. Lang (Eds.), *Dimensional Adjectives* (pp. 71–262). Berlin, Germany: SpringerVerlag.

Boersma, P., and Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.16, retrieved 6 June 2020 from http://www.praat.org/

Booth, A. E., and Waxman, S. R. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development*, *80*(1), 15-22. doi.org/10.1111/j.1467-8624.2008.01242.x

Brown-Schmidt, S., Gunlogson, C., and Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134. doi: 10.1016/j.cognition.2007.11.005

Davies, C., Andrés-Roqueta, C., and Norbury, C. F (2016). Referring expressions and structural language abilities in children with specific language impairment: A pragmatic tolerance account. *Journal of Experimental Child Psychology, 144,* 98-113. doi: 10.1016/j.jecp.2015.11.011

Davies, C., Lingwood, J. and Arunachalam, S. (2020) Adjective forms and functions in British English child-directed speech. *Journal of Child Language, 47,* 159-185. doi: 10.1017/S0305000919000242

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321. doi: 10.1037/0033-295X.93.3.283

Diesendruck, Gil, Hall, D. Geoffrey, and Graham, Susan (2006). Children's use of syntactic and pragmatic knowledge in the interpretation of novel adjectives. *Child Development*, 77(1), 16–30. doi:10.1111/j.1467-8624.2006.00853.x

Donnelly, S., and Verkuilen, J. (2017). Empirical logit analysis is not logistic regression. Journal of Memory and Language, 94, 28-42. doi: 10.1016/j.jml.2016.10.005

Duñabeitia J. A., Crepaldi D., Meyer A. S., New B., Pliatsikas C., Smolka E., et al. (2018). MultiPic: a standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology. Exp. Psychol.* 71 808–816. doi: 10.1080/17470218.2017.1310261

Ebeling, K. S., and Gelman, S. A. (1988). Coordination of size standards by young children. *Child Development,* 59, 888–896. doi: 10.2307/1130256

EyeLink Data Viewer v.4.1.63 (2020). Mississauga, Ontario, Canada: SR Research Ltd

Falkum, I. L., Recassen, M., and Clark, E. (2017). "The moustache sits down first": on the acquisition of metonymy. *Journal of Child Language*, 44, 87-119. doi: 10.1017/S0305000915000720

Fernald, A., Marchman, V., and Hurtado, N. (2008) Input affects uptake: How early language experience influences processing efficiency and vocabulary learning. *Proceedings of the 7th IEEE International Conference on Development and Learning.* IEEE; pp. 37–42. doi: 10.1109/DEVLRN.2008.4640802

Fernald, A., Perfors, A. F., and Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the second year. *Developmental Psychology,* 42, 98-116. doi: 10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., and McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9,* 228–231. doi: 10.1111/1467-9280.00044

Fernald, A., Thorpe, K., and Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology, 60,* 190–217. doi: 10.1016/j.cogpsych.2009.12.002

Gao, H. H., Zelazo, P. D., Sharpe, D., and Mashari, A. (2014). Beyond early linguistic competence: Development of children's ability to interpret adjectives flexibly. *Cognitive Development, 32,* 86–102. doi.org/10.1016/j.cogdev.2014.08.003

Gatt, A., R. van Gompel, E. Krahmer and K. van Deemter (2012). Does domain size impact speech onset time during reference production? *Proceedings of the 34th annual meeting of the Cognitive Science Society (CogSci 2012)*. Sapporo, Japan, pp. 1584 – 1589

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Language, cognition, and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.

Gentner, D., and Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman and S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, UK: Cambridge University Press.

Grice, H. P. (1975) 'Logic and conversation'. In P. Cole and J. Morgan (eds) *Studies in Syntax and Semantics III: Speech Acts*, New York: Academic Press, pp. 183-98.

Grodner, D., and Sedivy, J. C. (2011). The effect of speaker specific information on pragmatic inferences. In N. Pearlmutter and E. Gibson (eds.). *The Processing and Acquisition of Reference* (Vol. 2327, pp. 239– 272). MIT Press.

Hall, D. G., Waxman, S. R., and Hurwitz, W. M. (1993). How two- and four-year-old children interpret adjectives and count nouns. *Child Development, 64*(6), 1651–1664. doi.org/10.2307/1131461

Hanten, G., and Martin, R. C. (2000). Contributions of phonological and semantic short-term memory to sentence processing: Evidence from two cases of closed head injury in children. *Journal of Memory and Language*, 43, 335–361. doi: 10.1006/jmla.2000.2731

He, A. X., Kon, M, and Arunachalam, S. (2020). Linguistic context in verb learning: Less is sometimes more. *Language Learning and Development*, *16*(1), 22-42. doi: 10.1080/15475441.2019.1676751

Higginbotham, J. (1985). On semantics. *Linguistic Inquiry, 16,* 547-594.

Horowitz, A. C. and Frank, M. C. (2014). Preschoolers infer contrast from adjectives if they can access lexical alternatives. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.*

Huang, Y. and Snedeker, J. (2013). The use of referential context in children's online interpretation of scalar adjectives. *Developmental Psychology, 49,* 1090-1102. doi: 10.1037/a0029477

Huang, Y., and Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition, 104251.* doi: 10.1016/j.cognition.2020.104251

Jincho, N., Oishi, H., and Mazuka, R. (2019). Developmental changes in the utilization of referential visual context during sentence comprehension: Eye movement and pupil dilation evidence from children and adults. *Language Learning and Development*, *15*(4), 350-365. doi: 10.1080/15475441.2019.1645668

Kamp, H., and Partee, B. (1995). Prototype theory and compositionality. *Cognition*, *57,* 129–191. doi:10.1016/0010-0277(94)00659-9

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison.* New York, NY: Garland.

Kennedy, C. (2012). Adjectives. In Russell, G. and D. Graff Fara (eds.), *Routledge Companion to Philosophy of Language*. Routledge.

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(3), 649–659. https://doi.org/10.1111/1467-8624.00173

Koenig, A., Arunachalam, S., and Saudino, K. (2020). Lexical processing of nouns and verbs at 36 months of age predicts concurrent and later vocabulary and school readiness. *Journal of Cognition and Development*, 21:5, 670-689, doi: 10.1080/15248372.2020.1802277

Kronmüller, E., Morisseau, T., and Noveck, I. (2014). Show me the pragmatic contribution: A developmental investigation of contrastive inference. *Journal of Child Language*, 41(5), 985-1014. doi:10.1017/S0305000913000263

Long, B., Moher, M., Carey, S., and Konkle, T. (2019). Real-world size is automatically encoded in preschoolers' object representations. *Journal of Experimental Psychology: Human Perception and Performance*. doi: 10.1037/xhp0000619

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–77. doi: 10.1207/s15516709cog1401_4

Martin, R. C., and He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language*, 89, 76–82. doi: 10.1016/S0093-934X(03)00300-6

Matthei, E. M. (1982). The acquisition of prenominal modifier sequences. *Cognition*, 2, 301-332. doi: 10.1016/0010-0277(82)90018-X

Medin, D. L., and Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20,* 158–190. doi: 10.1016/0010-0285(88)90018-7

Mintz, T. H. (2005). Linguistic and conceptual influences on adjective acquisition in 24- and 36-month-olds. *Developmental Psychology*, 41, 17–29

Mintz, T. H., and Gleitman, L. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition*, 84, 267–293

Mirman, D. (2014). *Growth curve analysis and visualization using R* (Chapman and Hall/CRC the R series). Boca Raton: CRC Press.

Morisseau, T., Davies, C. and Matthews, D. (2013). How do 3- and 5-year-olds respond to under- and overinformative utterances*? Journal of Pragmatics*, 59, 26-39 doi: 10.1016/j.pragma.2013.03.007

Nadig, A., Sedivy, J., Joshi, A., and Bortfeld, H. (2003). The development of discourse constraints on the interpretation of adjectives. In B. Beachley, A. Brown, and F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 568-579). Somervill, MA: Cascadilla Press.

Nelson, K. and Benedict, H. (1974). The comprehension of relative, absolute, and contrastive adjectives by young children. *Journal of Psycholinguistic Research,* 3, 333-42.

Ninio, A. (1988). On formal grammatical categories in early child language. In Y. Levy, I. M. Schlesinger, and M. D. S. Braine (Eds.), *Categories and processes in language acquisition* (pp. 99–119). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ninio, A. (2004). Young children's difficulty with adjectives modifying nouns. *Journal of Child Language, 31*, 255–285. doi: 10.1017/s0305000904006191.

Omaki, A. and Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge.*Language Acquisition*, 22:2, 158-192, doi: 10.1080/10489223.2014.943903

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M. and Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cognitive Psychology*, 115, 101-238. doi: 10.1016/j.cogpsych.2019.101238

Pouscoulous, N. and Tomasello, M. (2019). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, 156, 160-167. doi: 10.1016/j.pragma.2019.05.021

Pouscoulous, N., I. Noveck, G. Politzer, and A. Bastide (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, 14: 347-376. doi: 10.1080/10489220701600457

Prasada, S. (1992). *Acquisition of adjective meanings: A lexical semantic approach*. Doctoral dissertation, MIT.

Prasada, S., and Cummins, M. (2001). Structural constraints on the interpretation of novel count nouns. *Proceedings of the Boston University Conference on Language Development*, 25, 623-632.

Qi, Z., Love, J., Fisher, C., & Brown-Schmidt, S. (2020). Referential context and executive functioning influence children's resolution of syntactic ambiguity. *Journal of experimental psychology. Learning, memory, and cognition*, 46(10), 1922–1947. https://doi.org/10.1037/xlm0000886

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramscar, M., Thorpe, K., and Denny, K. (2007). Surprise in the learning of color words. Proceedings of the 29th Meeting of the Cognitive Science Society, Nashville, TN.

Ramscar, M., Yarlett, D., Dye, M., Denney, K., and Thorpe, K. (2010) The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science, 34,* 909–957. doi: 10.1111/j.1551-6709.2009.01092.x

Rubio-Fernandez, P., Mollica, F., & Jara-Ettinger, J. (2021). Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General*, 150(3), 583–594. https://doi.org/10.1037/xge0000963

Rubio-Fernández, P. (2016) How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology.* 7:153. doi: 10.3389/fpsyg.2016.00153

Ryskin, R., Kurumada, C., and Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, 43(8), e12769. doi: 10.1111/cogs.12769

Sandhofer, C. M., and Smith, L. B. (2007). Learning adjectives in the real world: how learning nouns impedes learning adjectives. *Language, Learning and Development, 3,* 233–67. doi: 10.1080/15475440701360465

Sandhofer, C. M., Smith, L. B., and Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning? *Journal of Child Language, 27,* 561–585. doi: 10.1017/S0305000900004256

Schulze, C., Grassmann, S., and Tomasello, M. (2013). 3-year-old children make relevance inferences in indirect verbal communication. *Child Development*, 84 (6), 2079-2093.

Schwarz, F. (2014). Presuppositions are fast, whether hard or soft - evidence from the visual world. In M. Wiegand, T. Snider, and S. D'Antonio (Eds.), *Semantics and Linguistic Theory* (SALT). Vol. 24, pp. 1-22). LSA and CLC Publications.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G. and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109-147. doi: 10.1016/s0010-0277(99)00025-6.

Sekerina, I. A. and Trueswell, J. C. (2012). Interactive processing of contrastive expressions by Russian children. *First Language, 32,* 63–87. doi: 10.1177/0142723711403981

Smith, L. B., Cooney, N. J., and McCord, C. (1986). What is "high"? The development of reference points for "high" and "low." *Child Development, 57*(3), 583–602. doi.org/10.2307/1130338

Snedeker, J., and Trueswell, J. C. (2004). The developing constraints on parsing decisions: the role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49(3), 238–299. doi: 10.1016/j.cogpsych.2004.03.001.

Stiller, A., Goodman, N., and Frank, M. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, 11 (2), 176-190.

Syrett, K., Kennedy, C. and Lidz, J. (2010). Meaning and Context in Children's Understanding of Gradable Adjectives, *Journal of Semantics*, 27(1), 1-35. doi.org/10.1093/jos/ffp011

Syrett, K., LaTourrette, A., Ferguson, B., and Waxman, S. R. (2019). Crying helps, but being sad doesn't: Infants constrain nominal reference online using known verbs, but not known adjectives. *Cognition, 104033*, doi: 10.1016/j.cognition.2019.104033

Taylor, M. and Gelman, S. A. (1988). Adjectives and nouns: Children's strategies for learning new words. *Child Development*, 59, 411–419

Thorpe, K., and Fernald, A. (2006). Knowing what a novel word is not: Two year-olds ''listen through" ambiguous adjectives in fluent speech. *Cognition, 100,* 389–433. doi: 10.1016/j.cognition.2005.04.009

Thorpe, K., Baumgartner, H., and 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

Tian, Y., Ferguson, H., and Breheny, R. (2016). Processing negation without context - Why and when we represent the positive argument. *Language, Cognition and Neuroscience, 31*(5), 683–698. doi: 10.1080/23273798.2016.1140214

Tribushinina, E. and 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,* 425-44. doi: 10.1017/S0305000915000173

Tribushinina, E., van den Bergh, H., Ravid, D., Aksu-Koç, A., Kilani-Schoch, M., Korecky-Kröll, K., Leibovitch-Cohen, I., Laaha, S., Nir, B., Dressler, W. U. and Gillis, S. (2014). Development of adjective frequencies across semantic classes - A growth curve analysis of child speech and child-directed speech. *Language, Interaction and Acquisition*, 5 (2), 185-226. doi: 10.1075/lia.5.2.02tri

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89–134. doi:10.1016/S0010-0277(99)00032-3

Waxman, S. R., and Booth, A. E. (2001). On the insufficiency of evidence for a domain-general account of word learning. *Cognition*, *78*, 277–9. doi: 10.1016/S0010-0277(00)00119-0

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, 99(2), 75–95. doi: 10.1016/j.jecp.2007.10.004.

Weisleder, A. and Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143-2152. doi:10.1177/0956797613488145

Weisleder, A., and Fernald, A. (2009). Real-time processing of postnominal adjectives by Latino children learning Spanish as a first language. In J. Chandlee, M. Franchini, S. Lord, and G.-M. Rheiner (Eds.), *Proceedings of the 33rd Annual Boston University Conference on Language Development* (pp. 611-621). Somerville, MA: Cascadilla Press.

Wienholz, A and Lieberman, A. M. (2019). Semantic processing of adjectives and nouns in American Sign Language: effects of reference ambiguity and word order across development. *Journal of Cultural Cognitive Science, 3,* 217–234. doi: 0.1007/s41809-019-00024-6

Wiig, E., Secord, W.A. and Semel, E. (2004). *The clinical evaluation of language fundamentals–Preschool*. (2nd edn). San Antonio, TX: Harcourt Assessment.

Woodard, K., Pozzan, L., and Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology,* 141, 187–209. doi: 10.1016/j.jecp.2015.08.005

Ziegler, J., and Pylkkänen, L. (2016). Scalar adjectives and the temporal unfolding of semantic composition: An MEG investigation. *Neuropsychologia*, *89*, 161-171. doi.org/10.1016/j.neuropsychologia.2016.06.010