

Running head: INTERACTIVE PROCESSING OF CONTRASTIVE EXPRESSIONS

**INTERACTIVE PROCESSING OF CONTRASTIVE EXPRESSIONS BY RUSSIAN CHILDREN**

Irina A. Sekerina

College of Staten Island and Graduate Center of the City University of New York

John C. Trueswell

University of Pennsylvania

In press, *First Language*.

Address for correspondence:

Dr. Irina Sekerina, Department of Psychology, 4S-108

College of Staten Island

2800 Victory Boulevard

Staten Island, NY 10314, USA

Email: [Irina.Sekerina@csi.cuny.edu](mailto:Irina.Sekerina@csi.cuny.edu)

Telephone: (718) 982-3760

**Abstract**

Children's ability to interpret color adjective noun phrases (e.g., red butterfly) as contrastive was examined in an eyetracking study with 6-year-old Russian children. Pitch accent placement (on the adjective *red*, or on the noun *butterfly*) was compared within a visual context containing two red referents (a butterfly and a fox) when only one of them had a contrast member (a purple butterfly) or when both had a contrast member (a purple butterfly and a grey fox). Contrastiveness was enhanced by the Russian-specific 'split constituent' construction (e.g., *Red put butterfly...*) in which a contrastive interpretation of the color term requires pitch accent on the adjective, with the NonSplit sentences serving as control. Regardless of the experimental manipulations, children had to wait until hearing the noun (butterfly) to identify the referent, even in splits. This occurred even under conditions for which the prosody and the visual context allows adult listeners to infer the relevant contrast set and anticipate the referent prior to hearing the noun (accent on the adjective in 1-Contrast scenes). Pitch accent on the adjective did facilitate children's referential processing, but only for the NonSplit constituents. Moreover, visual contexts that encouraged the correct contrast set (1-Contrast) only facilitated referential processing after hearing the noun, even in splits. Further analyses showed that children can anticipate the reference like adults but only when the contrast set is made salient by the preceding supportive discourse, that is, when the inference about the intended contrast set is provided by the preceding utterance.

**Keywords:** children; context; eyetracking; prosody; Russian contrastiveness

## 1. Introduction

### 1.1 *Interpretation of Contrastive Focus*

The interpretation of contrastive focus is an especially interesting topic for psycholinguistic theorizing because successful interpretation of contrastively focused expressions requires the integration of multiple levels of linguistic knowledge: prosodic, syntactic, semantic and discourse-pragmatic. Such expressions therefore pose considerable processing challenges for both the language learner and the proficient adult user of the language. In this paper, we present an experimental investigation of how children interpret contrastively focused expressions, concentrating especially on the real-time processes that comprise their referential interpretation.

Consider first the mental computations an English listener must go through, and the sources of linguistic knowledge he/she must consult, to correctly identify and interpret expressions like those in (1).

- (1) a. *Give me the YELLOW carnations.* (As opposed to, e.g., the red carnations.)  
 b. *Give me the yellow CARNATIONS.* (As opposed to, e.g., the yellow roses.)

Acoustically, the listener must identify those syllables that are receiving greater than expected articulatory and tonal emphasis (which is often referred to as a *pitch accent* within autosegmental accounts of prosody, e.g., Bruce, 1977; Goldsmith, 1976; Pierrehumbert, 1980). And, although debated, many psycholinguistic theories posit that listeners rapidly categorize pitch accents based on their tonal shape, and use these categories (in combination with the syntactic and semantic analyses of the sentence) to systematically alter referential and

pragmatic implications (e.g., Pierrehumbert, 1980). For instance, speakers often signal the intention to contrastively focus a semantic element via the use of an L+H\* pitch accent<sup>1</sup>, which is a low tonal target containing a steep rise to the pitch maximum (Pierrehumbert & Hirschberg, 1990). Interpretation of the pitch accent is however highly context dependent; if L+H\* appears on an adjective (like 1a), then a contrastive focus is the most likely interpretation (the yellow carnations, as opposed to the red ones). In contrast, L+H\* on a noun can signal contrastive focus (as in 1b), but it can have other focus meanings (e.g., broad focus, as in response to “Can I help you?”, see Ladd, 1980 as discussed in Weber, Braun & Crocker, 2006; also Krahmer & Swarts, 2001). Finally, accurate interpretation of the contrastive focus constituent requires computation, or implicit recognition, of the contrast set (i.e., non-yellow carnations, or yellow ‘non-carnations’). Contrast set information is typically present either in the discourse context or the immediate visual context of the conversants; if not present, the listener must infer this information (i.e., it becomes a discourse presupposition of the contrastively focused expression).

Given the striking complexities associated with the interpretation of contrastively focused elements, it is rather surprising how quickly and effortlessly adult native speakers interpret them. For instance, under supportive referential contexts, listeners can often identify the intended referent of contrastive adjective noun phrases (like 1a) prior to fully processing the head noun, e.g., prior to fully hearing *carnation* (Ito & Speer, 2008; Weber, Braun, and Crocker, 2006). For instance, Ito and Speer (2008) recorded listeners’ eye movements as they responded to spoken instructions containing contrastive and noncontrastive expressions. Spoken instructions were all related to a task of decorating a Christmas tree with ornaments of

different colors. Critical sentences were two sentence sequences embedded in the longer discourse each containing an adjective noun phrase (*First hang the green ball...now hang the BLUE ball.*). These occurred in the presence of multiple objects that had the relevant attribute (i.e., there were other blue objects present including blue stars, blue drums, etc.). It was found that providing the appropriate L+H\* pitch accent on the second adjective resulted in listeners making eye movements to the intended referent (the blue ball) while hearing the noun (ball). A sizable proportion of these eye movements occurred during the first 200 ms of hearing the noun, which, given the 150 ms delay expected from eye movement programming (Matin, Shao, & Boff, 1993), means that many listeners identified the blue ball in the visual scene based on the word *blue* alone, despite the presence of other blue objects. Anticipatory processing of this sort was reduced (but not completely eliminated) when the adjective no longer had an L+H\* pitch accent. Weber et al. (2006) reported very similar effects for the interpretation of contrastive adjectives in German. Taken together, these findings suggest that adjectives tend to be interpreted as contrastive when the preceding discourse explicitly evokes a contrast set (*First hang the green ball...now hang the blue ball*) resulting in anticipatory referential processing at the adjective (*blue*). Moreover, this anticipatory processing can be enhanced by the presence of an appropriate pitch accent on the second adjective (e.g., *the BLUE ball*).

Interestingly, other eyetracking studies by Sedivy and colleagues suggest that certain visual contexts can result in anticipatory processing of adjectives even in the absence of a discourse context that explicitly evokes the relevant contrast set (Sedivy, Tanenhaus, Chambers & Tanenhaus, 1999). Hearing *Pick up the tall glass* in the presence of one tall glass and one short glass results in anticipatory looks at *tall* whereas no such anticipatory looks are found

when the visual context does not contain the contrast object (the short glass) and instead contains an unrelated object. Intersective adjectives (e.g., color adjectives) appear to be less likely to evoke relevant contrast sets from the visual world alone (*Pick up the blue comb*, in the presence of a blue comb and a yellow comb) perhaps because such adjectives are sometimes used redundantly by listeners to aid object identification (see Sedivy, 2003).

Taken together, the experimental evidence suggests that multiple sources of evidence contribute to evoking contrastive interpretations in the minds of listeners, including the structure of the visual world, discourse structure, and prosodic, syntactic and semantic evidence. Less is understood about how the listener weighs these sources of evidence, and even less is understood about how children build the relevant evidentiary sources to achieve adult-like processing of contrastive expressions.

Here we will ask how children respond to contrastive expressions when we parametrically manipulate these sources of evidence (prosodic, visual world, discourse) for the child. In this way, we can get a quick handle on what kinds of evidence are viewed by the child as more informative, and whether such evidence can trump other sources of evidence when they conflict. Moreover, we will conduct this study in a language where linguistic evidence to contrastive focus is arguably more clear-cut: Russian, whose grammar offers constructions that explicitly encode contrastive focus, which when combined with prosody offer opportunities for unambiguous contrastive focus expressions. English offers no such opportunity, in that the accurate interpretation of pitch accent, as discussed above, is often ambiguous and almost always highly dependent on context.

## 1.2 *Interpretation of Contrastive Expressions in Russian*

Russian is an especially interesting case for studying the interpretation of contrastive expressions because in addition to having prosodic means for signaling contrastive focus, it is also a language whose sentences can overtly encode this information in the syntax. According to the functional approach of the Prague Linguistic School adopted by Russian grammarians (Švedova, 1982), the canonical structure of a simple Russian sentence is SVO but it can have different surface word order realizations. For example, with other things being equal, all six permutations of the subject, verb and object are possible because of the rich case morphology of Russian, which greatly aids in thematic Role assignment. The different word orders of Russian are believed to be driven by ‘communicative intention’ (Švedova, 1982, p. 91) that divides a sentence into two categories, *theme* (given information) and *rheme* (new information), and the theme typically precedes the rheme. This, however, does not diminish the importance of prosody. Russian sentential prosody exists not only as an additional linguistic means to express the Information Structure of a sentence but can also play a critical role in distinguishing communicative intentions, as is the case, for example, for declarative sentences vs. yes-no questions.<sup>2</sup>

Prosody in Russian can act independently of or in consort with order variation which creates numerous information structure possibilities in spoken sentences. Of particular interest in the present investigation are contrastive focus constructions like those appearing in (2). Here, imagine that the speaker is instructing an individual to move clipart images around on a computer screen.

- (2) a. **KRAS**nuju položite babočku v paket.  
**RED**<sub>ACC-FEM</sub> put butterfly<sub>ACC-FEM</sub> in paper bag.
- b. Krasnuju položite **BA**bočku v paket.  
Red<sub>ACC-FEM</sub> put **BUTTERFLY**<sub>ACC-FEM</sub> in paper bag.
- c. **KRAS**nuju babočku položite v paket.  
**RED**<sub>ACC-FEM</sub> butterfly<sub>ACC-FEM</sub> put in paper bag.
- d. Krasnuju **BA**bočku položite v paket.  
Red<sub>ACC-FEM</sub> **BUTTERFLY**<sub>ACC-FEM</sub> put in paper bag.

'Put the red butterfly in the paper bag.'

Instructions (2a) and (2b) utilize a 'split constituent' construction, a grammatically marked option in spoken Russian that encodes contrastive focus by moving the adjective to the sentence-initial position while keeping the head noun in its canonical position after the verb. The split constituent construction is very frequent in colloquial Russian, poetry and folk literature, and is particularly productive with 'light' verbs and simple three-word sentences of the type *Gorjačego xočeš' čaju?* (lit. 'Hot you want tea?') 'Do you want hot tea?', in which the split components are at the periphery of the entire clause, not just the VP. The split constituent construction is formally characterized by (a) obligatory movement of the split-constituent adjective and the split-constituent noun to the leftmost and rightmost periphery of the VP, respectively (Sekerina, 1997; see Pereltsvaig, 2008, for syntactic analysis), and (b) an obligatory contrastive pitch accent on one of the split-components, combined with lengthening of the stressed syllable and a boundary L% tone. These special prosodic features must be assigned to only one of the two components, either the split-constituent adjective (2a) or the split-

constituent noun (2b). Thus, the contrastive pitch accent on the split-constituent adjective in (2a) is realized as a 'hat' contour combination of H and L accents H+L\* while in (2b), the H and L accents on the split-constituent noun are ordered in reverse, as L+H\* (Meyer & Mleinek, 2006; Odé, 2008). Impressionistically, assignment of contrastive pitch accent to the split-constituent adjective *KRASnuju* 'red' in (2a) is a more frequent case than to the split-constituent noun *BAbočku* 'butterfly' in (2b). Interestingly, while (2a) unambiguously implies contrastive focus on the adjective (e.g., the red butterfly as opposed to the purple one), a pitch accent on the noun in (2b) is ambiguous: it can mean either wide focus on the VP or contrastive focus on the noun (e.g., the red butterfly as opposed to a red fox). In the absence of a formal spoken corpus count, we speculate that a pitch accent on the adjective is exclusively contrastive, which is consistent with the first author's (native speaker) intuitions. Nonsplit constituents can also be scrambled to the left periphery as in (2c) and (2d), and it is also possible to place a pitch accent on the adjective or the noun in such constructions to imply contrastive focus of the word in question, but again, intuitively, it is a less common option than the split constituent variants (2a) and (2b). Such an interpretive pattern, if correct, suggests that the split constituent becomes an unambiguous signal of contrastive focus in Russian when the pitch accent is placed on the left-most split component.

In two eyetracking experiments with Russian speaking adults (Sekerina and Trueswell, submitted), we explored the real-time processing of the Russian-specific 'split constituent' construction, whose sole function seems to be to convey contrastiveness. The second of these two experiments was identical in design to the present study with children and is thus of central interest. In that experiment, native Russian-speaking adults heard instructions that

contained split constituents and nonsplit scrambled constituents, as illustrated in (2a through 2d) above. Simultaneously, they looked at a laptop computer screen that contained a referentially ambiguous configuration of colored clipart images such that the visual context always contained two objects of the same color (e.g., a red butterfly and a red fox, as in Figure 1 below). We carefully selected these pairs of shapes on the basis of the gender of their common Russian names: both had to be either feminine or masculine to avoid the use of gender as an additional disambiguating cue. In the 1-Contrast condition (Figure 1a), only the Target, the red butterfly, belonged to a possible contrast set (i.e., there was a purple butterfly but no additional fox). In the 2-Contrast condition, both the Target and the Competitor, the red fox, were part of contrast sets because an unrelated distractor object was replaced with a fox of another color (e.g., a gray fox). For each of these visual context types, the location of the contrastive pitch accent was manipulated in the instruction. Early Accent sentences had contrastive pitch accent H+L\* on the adjective (e.g., *KRASnuju položite babočku...* ‘RED put butterfly...’) whereas Late Accent sentences had L+H\* on the head noun (e.g., *Krasnuju položite BAbočku...* ‘Red put BUTTERFLY...’). Thus, the design consisted of eight conditions that resulted from fully crossing the type of construction (Split NP vs. NonSplit NP), the placement of pitch accent (Early vs. Late) and visual context (1-Contrast vs. 2-Contrast).

-----  
 Insert Figure 1 here  
 -----

Two critical features of our design distinguish it from the studies by Sedivy and colleagues (Sedivy et al., 1999; Sedivy, 2003), Weber et al. (2006), and Ito and Speer (2008)

described above. First, the Split constituent construction ADJ-V-N allowed us to explore the time course of interpretation of contrastiveness by increasing the distance between the color adjective and the noun. The verb *položite* ‘put’ created a 450-ms interval during which Russian listeners could already reliably identify the target referent, i.e., at least 200 ms prior to appearance of the noun in the instruction. Second, the study parametrically manipulated all known contributors to the contrastive focus interpretation, allowing for a clearer understanding of cue integration in contrastively focused adjectives.

The results were straightforward and in line with the expectations derived above about the syntactic encoding of contrastive focus in Russian. For Split constituents, evidence for anticipatory referential processing was observed prior to hearing the noun (during the verb *put*) but only when the prosody and the visual referent world supported a contrastive interpretation. Specifically, when a pitch accent appeared on the adjective (*RED put butterfly...*), adults showed anticipatory looks to the red butterfly but only in the 1-Contrast condition (where the red butterfly had a color contrast object but the red fox did not). Such a finding suggests that the Russian split constituent becomes an unambiguous signal of contrastive focus for adults when the pitch accent is placed on the left-most split component. Thus, even intersective adjectives (color adjectives) in Russian can be interpreted contrastively based on the visual context alone provided that contrastive focus on the adjective is unambiguously signaled. The NonSplit sentences showed essentially no signs of anticipatory processing regardless of prosody or visual scene, suggesting that leftward scrambling of the entire constituent is not a reliable signal for contrastive focus of the adjective.

Importantly, a subset of the items in this study were of the sort that there was also discourse support of the referential contrast, like in some previous studies. I.e., the target instruction was the second instruction, and was preceded by reference to the color contrast object (*Put the purple butterfly next to the bowl. RED put butterfly in paper bag.*). Here too anticipatory processing was observed for split constituents. But unlike the earlier English and German studies (of Ito & Speer, 2008; Weber et al., 2006), a pitch accent on the adjective was still required for listeners to show anticipatory processing. In those previous experiments, evoking the contrast set via the discourse was sufficient for generating anticipatory referential processing of the immediately following adjective (with or without a pitch accent). Thus, Russian speakers appear to place special importance on the split constituent construction as a signal for contrastive focus and also expect prosodic markings to align with the pragmatic function of the structure.

### **1.3 Children's Interpretation of Contrastive Focus**

The combination of the experimental method of visual world eyetracking and the language-specific paradigm of Russian contrastive focus offers an important opportunity to explore how young children process contrastive expressions whose adult-like interpretations require computationally intensive processing at linguistic and nonlinguistic interfaces: i.e., the integration of visual, discourse, syntactic and prosodic evidence for referential assignment. How might Russian-speaking children weigh these cues to interpretation and reference?

Predictions can be derived by looking at other studies that have examined the coordination of multiple sources of information by children, specifically in the area of syntactic

ambiguity resolution rather than referential ambiguity resolution (e.g., Trueswell, Sekerina, Hill & Logrip, 1999; see also Trueswell & Gleitman, 2004, 2007 and references therein). Without going into the details here, these studies suggest that young children (ages 4-6 years) have great difficulty evoking the discourse presuppositions associated with particular structures from the configuration of a visual referent world alone. For instance, five-year-old children fail to realize that in response to an ambiguous prepositional phrase (*Now tap the frog with the feather*) a visual referent world of two frogs, one holding a feather and one holding a leaf, evokes the discourse presuppositions associated with a modifier interpretation of *with the feather* (see Trueswell et al., 1999; Snedeker & Trueswell, 2003). Instead, they appear to be slaves to the lexico-syntactic evidence in the sentence itself, pursuing a modifier or instrument interpretation of *with the feather* based on the syntactic/semantic preferences of the verb. So, in this case, children, regardless of whether there are two frogs or just one frog in the visual referent world, show a high proportion of instrument actions (picking up a feather and using it to tap a frog). Adults in contrast show sensitivity to the visual context manipulation, producing fewer instrument actions and more modifier actions in two-frog scenes as compared to one-frog scenes. Interestingly, children appear to only utilize the structure of the visual context to resolve syntactic ambiguity when the preceding verbal discourse explicitly highlights and structures the presence of the two frogs (Hurewitz, 2001). Furthermore, developmental changes in the use of particular linguistic cues to interpretation appear to be due (in part) to their reliability in the input, and the ease to which these linguistic cues can be detected in the utterance (see Trueswell & Gleitman, 2004; Bates and MacWhinney, 1987).

With these facts from syntactic ambiguity resolution in mind, the following expectations can be derived about children's processing of contrastively focused expressions. First and foremost, we might expect children to require discourse support of the contrast set to see contrastive (anticipatory) processing in Split constructions. That is, children should not be able to use the structure of the visual world to evoke the discourse presuppositions associated with the split constituent (i.e., that it needs a contrast set to be felicitous). This makes the strong prediction that children will show absolutely no anticipatory referential processing in response to Split constructions in 1-Contrast conditions; they will have to delay establishing reference until hearing the head noun. However, when the discourse highlights the relevant contrast (via explicit mention of the contrast member), then we expect children to show anticipatory processing in line with adults.

It is less clear what role prosody will play in Russian children's processing of contrastive focus. Although previous studies have found prosody to aid syntactic ambiguity resolution in English speaking children of this age range (Snedeker & Yuan, 2008), the presence of a strong and clear syntactic cue for contrastive focus (split scrambling) may make children less sensitive to prosodic markings. This would be in line with the notion that cue reliability plays an important role in the ordering of the acquisition of linguistic cues to interpretation: an obvious syntactic marking ought to trump a more probabilistic cue to the same interpretation.

Thus far, experimental investigation of children's interpretation of contrastive focus of adjectives has been limited to offline act-out and picture selection studies. The only on-line study of children has been an eyetracking study by Ito and colleagues in Japanese (Ito, Jincho, Minai, Yamane, & Mazuka, 2008). Although this study finds some evidence for children's

sensitivity of prosody in Japanese, they find no signs of anticipatory processing. Other studies have examined children's understanding of focus operators in other languages (e.g., Paterson, Liversedge, Rowland & Filik, 2003; Gualmini, Maciukaite, & Crain, 2003) but have used offline measures of interpretation with the exception of one study on German-speaking children (Höhle, Berger, Müller, Schmitz, & Weissenborn, 2009). We will return to this evidence in the general discussion after presenting the current findings from Russian.

## **2. Method**

### **2.1 Participants**

Thirty-two monolingual Russian-speaking children (20 girls, 12 boys; mean age 6;2, range 5;1-6;11) were recruited and tested at a large preschool center in Moscow, Russia. The parents of all participating children signed an informed consent form in Russian. All of the children were screened for language disorders by a Russian speech-language pathologist as a routine procedure in determining children's preparedness for elementary school, which begins at the age of 7. Children received a toy for their participation in the experiment.

### **2.2 Design and Materials**

On each experimental trial, participants viewed one of two types of visual displays, as illustrated in Figure 1 above. Each display was accompanied by one of four possible variants of the same spoken instruction, an example of which appears in (2a-d) above. Each visual stimulus contained nine pictures arranged in the cells of a 3x3 grid: a smiley face in Position 5, five

colored pictures of different objects, and three black-and-white pictures of containers taken from a set of nine (e.g., skillet, barrel, paper bag, suitcase, bowl, jar, box, basket). Every experimental item, regardless of condition, contained two objects of the same color as the color term mentioned in the instruction. One was the intended referent (e.g., the Target object, the red butterfly) and the other was a referentially competing object (e.g., the Competitor object, the red fox).

The Type of Visual Context represented the first experimental manipulation: 1-Contrast (Fig. 1A) or 2-Contrast (Fig. 1B). In the 1-Contrast context, only the Target had a contrasting object of different color, i.e., there was a purple butterfly in the display but no additional foxes. The two remaining objects for the display were of the same shape (fish) but of different color (blue and gray) and were never the same color as the Target. In the 2-Contrast context, both the Target and the Competitor were members of contrast sets (Fig. 1B). This was achieved by, e.g., replacing the gray fish with a gray fox. As discussed above, only the 1-Contrast context is expected to elicit a contrastive interpretation of the color adjective *red* early on and point to the correct Target, the red butterfly, even before the head noun becomes available in speech. The initial locations of the Target and the color Competitor, as well as their destination locations, were counterbalanced across trials. We selected 130 (24 experimental and 2 practice trials x 5 objects) concrete objects that were easily rendered in line drawings. The drawings were taken from the picture set created by Cycowicz, Friedman, Rothstein, and Snodgrass (1997) that has been extensively used in many previous picture-naming experiments. We selected the Target–Competitor pairs of objects so that their names were of the same grammatical gender in Russian preventing participants from using gender agreement on the

adjective as a disambiguating cue. Hence, all things being equal, the Russian adjective *krasnuju* ‘red<sub>FEM</sub>’ could be used to refer to either the Target *babočku* ‘butterfly<sub>FEM</sub>’ or the Competitor *lisicu* ‘fox<sub>FEM</sub>’.

The second experimental manipulation, Prosody, varied the placement of contrastive pitch accent. In sentences with Early Accent, it was placed on the adjective (e.g., *KRASNuju položite babočku...* ‘RED put butterfly...’) whereas in sentences with Late Accent, it was on the head noun (e.g., *Krasnuju položite BAbočku...* ‘Red put BUTTERfly...’). An acoustic analysis of the stimuli revealed highly marked prosodic differences between the sentences with different pitch accent placements. In the Early Accent condition, the stressed syllable *KRAS-* ‘RED’ had a longer duration, greater amplitude and a sharply falling pitch accent H+L\* which are typical characteristics of contrastive pitch accents in Russian (Bryzgunova, 1977; Makarova, 2007; Mehlhorn, 2004; Meyer & Mleinek, 2006; Odé, 2008). In the Late Accent condition, the stressed syllable of the noun *BA-* ‘BUTTERfly’ had a greater amplitude and sharply rising pitch accent L+H\*.

In addition to pitch accent placement, word order was also manipulated to create two types of spoken sentences, the Split and NonSplit conditions (see Example 2 above). Spoken instructions consisted of adjective-noun pairs, the verb *put*, and a location phrase. Twelve different color adjectives were used repeatedly (white, yellow, red, orange, pink, gray, brown, green, blue, navy, purple, and black) combined with singular nouns in the Accusative case. All adjectives and nouns were common and highly frequent Russian words, and no special efforts were taken to match them in frequency of occurrence in the language.

Instructions were recorded by a female native speaker of Russian in a soundproof booth, sampling at 22,050 Hz. Durations of the adjective, the verb, and the noun were measured (Table 1). Not surprisingly, the words with contrastive pitch accent were on average 150-300 ms longer than the same words without such accent.

-----  
 Insert Table 1 here  
 -----

Each trial consisted of two spoken instructions presented along with the visual display with five colored objects and the happy face. For half of the trials, the experimental instruction was presented first followed by a filler (3a-b), and for the other half, a filler instruction preceded the experimental one (4a-b) (only the Early Accent, Split conditions are used for illustration).

- (3) a. KRASnuju      položite      babočku      v paket.  
          RED            put            butterfly      in the paper bag.
- b. A teper'        položite      lilovuju babočku    sleva ot sinej ryby.  
          And now        put            purple butterfly    to the left of the blue fish.

'Put the RED butterfly in the paper bag.

And now put the purple butterfly to the left of the blue fish.'

- (4) a. Položite      lilovuju babočku    sleva            ot sinej ryby.  
          put            purple butterfly    to the left      from the blue fish
- b. A KRASnuju      položite      babočku      v paket.  
          And RED        put            butterfly      in the paper bag.

‘Put the purple butterfly to the left of the blue fish.

And [now] put the RED butterfly in the paper bag.’

Note that the presentation order of the experimental instruction was not formally manipulated as a factor to avoid an overly complex design; however, it resulted in a replication of the ‘discourse-second’ design used in the previous eyetracking studies of contrastiveness (Sedivy, et al., 1999; Sedivy, 2003; Weber et al., 2006; Ito and Speer, 2008; Ito et al., 2008). This allowed us to have a number of experimental items which had discourse support, as in (4). For six of them, the filler instruction (4a) set up an explicit contrast set for the Target *red butterfly* because the Competitor *purple butterfly* was verbally mentioned. For the remaining discourse-second items, discourse support was absent because the filler instruction did not mention the Competitor (e.g., *Put the gray fish to the left of the blue fish*) effectively rendering them like the discourse-first items in (3).

A total of eight experiment lists were created, each containing two practice trials and 24 experimental trials. The three experimental factors—Context (1-Contrast, 2-Contrast), Prosody (Early Accent, Late Accent), and Sentence Type (NonSplit, Split)—were crossed in a 2x2x2 within-subject factorial design with eight conditions, resulting in three experimental items per condition per list. Each item was rotated to a different condition across each list. Order of items was randomized. Children were randomly assigned to one of the eight lists, with four participants run on each list.

### **2.3 Procedure**

Children were tested individually in a quiet room of their preschool. Prior to conducting the experiment, each child was familiarized with the equipment, the task requirements and was given an opportunity to play with the computer mouse. After obtaining the child's oral consent, the child was seated comfortably in front of the stimuli laptop to which the remote eyetracking camera was attached, and an eyetracking calibration was performed. A small subset of participants (six children) did not have a computer at home and, thus, lacked fluent mouse-clicking skill. They participated in a short practice session to learn how to use the mouse prior to taking part in the experiment. Other children reported a regular exposure to computer video games at home and demonstrated adult-like motor fluency with the mouse.

We used a remote free-viewing eye-tracker (ETL-500) from ISCAN Inc., to record children's eye movements during the experiment. Eye movements were sampled at a rate of 30 frames per second, and were recorded on a digital SONY DSR-30 video tape recorder. The visual stimuli were presented on a 19-inch HP laptop computer while the spoken instructions were simultaneously played through the two speakers located on either side of the laptop. The materials were programmed into an interactive presentation using Macromedia Flash 8 software that allowed for smooth integration of visual displays and spoken instructions, and was response-contingent. Each trial started with the presentation of a three-by-three matrix containing a yellow smiley face in the center (see Figure 1) against light turquoise background. Initially, only the smiley face was visible that blinked three times at 1 s intervals. We expected that the rhythmical blinking of the smiley face would capture children's attention, and they would hold their gaze until the onset of the first spoken instruction. After three seconds, in

addition to the smiley face, the other objects appeared in different locations in the matrix. The first instruction, either experimental (3a) or filler (4a), was played simultaneously with the appearance of the pictures.

Children used a child-size mouse attached to the laptop to click and drag the colored objects to the containers. After the child dragged and dropped the correct object into the correct container (e.g., the red butterfly into the paper bag), the feedback instruction was played (*Pravil'no* 'correct'). In the case of an error (a wrong object or container was selected), the incorrect feedback was played (*Net, podumaj ešče* 'No, think again'), and the program returned the object into its original cell which prompted the child to revise her action. The procedure was repeated for the second instruction, and the same feedback was provided. After both instructions were correctly performed, a prominent red right-pointing arrow appeared to the right of the matrix which allowed the child to move to the next trial at her own pace. The experiment lasted about 20 minutes.

#### **2.4 Data Treatment**

Out of the possible 768 trials (32 participants x 24 trials), thirteen trials (1.7%) constituted missing data: they were not recorded because of equipment malfunctioning or other external circumstances. All of the trials in which the child did not click on the Target in the experimental instruction on the first attempt were coded as errors (24 trials, 3.2%) and excluded from the eye-movement analyses reported below.

Eye-movement data were extracted from videotape using a SONY DSR-30 VCR with frame-by-frame control and synchronized video and audio. Nine fixation categories were coded: the

smiley face, the Target (e.g., the red butterfly), the Competitor (the red fox), the contrast object for the Target (the purple butterfly), the contrast object for the Competitor (the gray fox), distractors (fish), looks in between objects, and track loss. All trials with track losses during the adjective or verb were dropped from analyses.

Eye-movement data from target trials were binned into time windows that were defined relative to the onset of words in the speech. Each word had two time windows, corresponding to the first and second half of a word. For instance, w1a and w1b refer to the first and second half of the first word, which was always the adjective of the scrambled direct object. Windows w2a and w2b corresponded to the second word, which for NonSplit sentences was the head noun of the direct object and for Split sentences was the matrix Verb (put). Windows w3a and w3b corresponded to the matrix verb for NonSplit sentences and to the head noun of the direct object for Split sentences. To take into account the approximate 200 ms delay in eye movements in response to speech (e.g., Allopenna et al., 1998), all time windows were offset positively by 200 ms.

For each time window on each trial, we computed the proportion of time the child spent looking at two critical regions: the Target region (e.g., the red butterfly) and the Competitor region (e.g., the red fox). We also computed Target advantage scores, which are defined as the proportion of time spent looking at the Target minus the proportion of time spent looking at the Competitor. All graphs will use these dependent variables to illustrate the timing of referential processing.

### **3. Results**

#### **3.1 Accuracy and reaction time data**

The experimental instruction in the trial (e.g., *KRASnuju položite babočku v paket* 'RED put butterfly in the paper bag') required the child to click on the picture that depicted the target object and drag it to a designated container. Children's accuracy in this task was almost at ceiling, with 96.8% correct overall. No further analysis was performed on the few incorrect trials because they were approximately equally distributed across the eight experimental conditions, with none standing out as particularly difficult. We took this finding to indicate that the task was easy for the children. Children's reaction times, i.e., the time it took them to click on the target object, were obtained from the eye-movement protocol. The time stamp of the participant's mouse click on the target object was coded as the reaction time. The average RT was 4344 ms (SD=1571 ms) including several outliers, ranging from 440 ms to 16 s depending on the child's previous experience with the mouse. This large variability in mouse-clicking skills prevented us from formally analyzing the obtained RT data.

#### **3.2 Eye-movement data**

We conducted fine-grain analyses of eye movements so as to investigate when during the utterance the three experimental manipulations (Visual Context, Prosody, and Sentence Type) contributed to the interpretation of contrastive expressions and to assess whether children can rapidly coordinate these sources of information to identify a target referent. These analyses revealed that, of all the factors manipulated in this study, only the Sentence Type (Split

vs. NonSplit) had a strong and systematic effect on the timing of referential processing. This is illustrated in Figure 2, which plots the data as a function of Sentence Type collapsing across all other factors.

As this figure illustrates, children rapidly discriminated the Target (e.g., the red butterfly) from the color-matched Competitor (the red fox) only upon hearing the head noun (e.g., *butterfly*). For NonSplit structures, the head noun corresponded to time windows w2a and w2b, whereas for Split constituents it corresponded to w3a and w3b. This is best seen in Figure 2B, which plots Target Advantage scores. Here the preference to look at the Target over the Competitor rises immediately in response to the head noun in both conditions.

-----  
 Insert Figure 2 here  
 -----

This delay in establishing reference until hearing the head noun held in all experimental conditions, even the condition for which adults showed anticipatory processing of the Target. This is illustrated in Figures 3 and 4, which divide up the results from Figure 2 into the corresponding eight experimental conditions.

-----  
 Insert Figures 3 and 4 here  
 -----

As can be seen in the figures, sharp rises in the Target advantage scores occurred only upon hearing the head noun in all conditions, even for those in which the visual context and sentence type support making the correct inference about the contrast set based on the adjective alone

(i.e., 1-Contrast, Early Accent, see Figure 3b). Here, the Split constituent, with stress on the adjective, signals that the Target object is being contrasted with its color contrast member (the RED butterfly, as compared to the purple one). Thus, if children were making the correct inference, hearing in Russian *RED put...* should be just as informative as “RED butterfly” in conveying that the red butterfly is the intended referent. For this condition, we instead see the same pattern that held for all Split constituent conditions, namely that children need to hear the head noun to establish reference. As discussed above in Section 1.2, this is precisely the condition that Russian adults showed anticipatory referential processing by looking to the Target rather than the Competitor as the verb was being processed (Sekerina & Trueswell, in prep.).

Early contrastive pitch accent (on the adjective) nevertheless did facilitate children’s referential processing. But this facilitation occurred only for the NonSplit constituents. Specifically, stressing the adjective sped establishment of the Target as the referent in 1-Contrast context, but only for NonSplit sentences, i.e., the circles rise faster in Figure 3b (NonSplit, Early Accent) as compared to Figure 3d (NonSplit, Late Accent). And, if anything, stressing the adjective of a Split constituent hindered establishing the referent (i.e., compare the square symbols across Figure 3b and 3d).

It is also the case that referentially unambiguous visual contexts (1-Contrast conditions as compared to 2-Contrast conditions) facilitated the establishment of the Target as the referent. This occurs for both Sentence Types (NonSplit and Split) but is relegated to events occurring during and after the processing of the head noun. This is illustrated when comparing

Figures 3 (1-Contrast) to Figure 4 (2-Contrast); the rises in Target advantage scores are generally more pronounced for the 1-Contrast context.

### 3.3 *Statistics*

The patterns discussed above were supported by statistical modeling that takes into account the variability that occurred across participants and across items. Separately for critical time windows (w2a, w2b, w3a and w3b), we developed hierarchical (a.k.a., multilevel) linear models which were designed to predict Target advantage scores based on the experimental factors. All models were provided with (crossed) random intercepts for Subject and Items (see Baayen, Davidson & Bates, 2008). Prior to entering the data into the models, the proportion of time spent looking at the Target and the proportion of time spent looking at the Competitor were transformed using an Empirical Logit function (see Barr, 2008) and then the difference was taken to compute E-logit Target Advantage scores.<sup>3</sup>

Table 2 presents the results of the mixed models containing all possible experimental factors. These models were not necessarily the best fitting models (as determined by Restricted Log Likelihood scores). However, the best fitting models for each region, which necessarily dropped some effects, still included those effects reported as significant below. As seen in the table, all critical time windows contained a reliable effect of Sentence Type, reflecting the fact that participants differentiated the Target from the Competitor more readily in NonSplit constructions as compared to Split constructions. Some other reliable effects (indicated in bold) were also observed. In w2a, we see that Sentence Type interacted with Prosody. As discussed above, this reflects the fact that helpful prosody (Early contrastive pitch accent on the

adjective) facilitated children's referential processing, but only for the NonSplit constituents. In w2b, a similar pattern resulted only in an effect of Prosody, suggesting that pitch accent on the adjective may have benefited the Split constituents as well, but numerically most of the benefit occurs for NonSplit constituents (see Figures 3 and 4). Finally, in w3b, Sentence Type interacted with Visual Context. As discussed above, this reflects the relatively late benefit of the 1-Contrast context, which appears to benefit especially NonSplit constituents.

-----  
 Insert Table 2 here  
 -----

In summary, the results discussed thus far indicate two things about children's processing of contrastiveness in Russian. First, hearing a color adjective in the presence of two color-matching objects for which only one has a contrastive pair (1-Contrast context) is not sufficient to allow children to infer the relevant referent. They do benefit from such scenes, but only *after* hearing the head noun. Specifically, in response to a linguistic structure that is marked for contrastiveness (Split constituents with stress on the adjective) children did not identify the Target until hearing the head noun. Russian-speaking adults, in contrast, can make such an inference and anticipate the correct referent (Sekerina & Trueswell, in prep.). And second, contrastive pitch accent on the adjective is beneficial to establishing the relevant referent and contrast set. However, this benefit is especially strong for NonSplit constituents while Split ones show a delayed and less enhanced benefit.

### 3.4 *Effects of Discourse*

Here we ask whether children can anticipate the Target when the contrast set is made salient by the preceding discourse, that is, when the inference about the intended contrast set is provided by the preceding utterance as illustrated in Example (4) above. On a subset of the items (six in total), the first (filler) instruction referred to the other member of the color contrast set just prior to giving the Target instruction: *Put the purple butterfly to the left of the blue fish. And RED put butterfly in the paper bag.* The question of interest here is: Will making explicit verbal reference to the color contrast member facilitate processing of the Target generally, or only for the linguistically marked (Split) structure? If the latter, it suggests that children understand the contrastive function of the split-constituent (and can anticipate the referent) but simply cannot infer the contrast set from the visual scene alone – they need support from the discourse structure.

Figure 5 (Panels A and B) presents the subset of trials for which the discourse explicitly established the color contrast set in 1-Contrast context. For comparison, the data from the 1-Contrast context without discourse support are plotted in Figure 5 (Panels C and D).

-----  
 Insert Figure 5 here  
 -----

Here we see clear signs that children can anticipate the intended referent of Split constituents, but only when the discourse provides the relevant contrast. As shown in Figure 5b, rises in the Target advantage are similar for Split and NonSplit constituents, suggesting that hearing the

verb *put* in Split constituents had a similar effect as hearing the actual head noun (e.g., *butterfly*) in NonSplit constituents (see w2a and w2b).

It is important to note that this analysis collapses across the two Prosody conditions. Similar findings were found regardless of whether the contrastive pitch accent was on the adjective or on the noun for Split constituents, suggesting that children are not yet fully attuned to the contrastive pitch accent as a prosodic cue to contrast (Wells et al., 2004).

Interestingly, we see similar discourse benefits for the 2-Contrast context. Figure 6 below includes both the 1-Contrast and 2-Contrast contexts (collapsed together) for the supportive discourse-second items. With more trials, the data become cleaner and more stable. Such a pattern makes sense because although two color contrast sets are present in 2-Contrast context (e.g., one for the butterflies, the other for the foxes), only the Target contrast set (the one for the butterflies) is established by the discourse as being relevant. Hence children anticipate the Target even in the 2-Contrast context.

-----  
 Insert Figure 6 here  
 -----

In both Figures 5 and 6, it is clear that NonSplit constituents ‘harm’ the establishment of the Target as the referent when a supportive discourse context is provided (compare the rise of the circles in Panels 5b and 6b to the same rise in the circles in 5d). Such a pattern suggests that children understand that the Split constituent, not the NonSplit constituent, is best suited for making referential contrasts.

### 3.5 *Statistics*

To examine statistically the effects of Discourse, we simply added Discourse Support (Present vs. Absent) as a fixed effect to the models discussed above. Multilevel mixed models are especially well suited for examining unbalanced designs such as this, where a small subset of the items had Discourse Support. For w2a, no effects or interactions with Discourse support were observed. However, consistent with the description above, the effect of Sentence Type was found to interact reliably with Discourse Support in w2b, and w3a, such that Discourse Support had a stronger influence on Split rather than NonSplit constituents (from the best fitting models, w2b:  $F(2,39) = 4.76$ ,  $p < 0.05$ ; w3a:  $F(2,40) = 3.63$ ,  $p < 0.05$ ). In w3b, Discourse Support showed no main effect or interaction with Sentence Type. In both these word positions, consistent with what was discussed above, Prosody did not interact with these effects. Entering Discourse Support into the models for these word positions did not eliminate the reliable effects of Prosody and Split seen in w2b, nor did it eliminate the effect of Split seen in w3a.

## 4. Discussion

We found that Russian-speaking children, in contrast to adults, did not compute contrast sets during the referentially ambiguous portion of the utterance, at least when they had to infer the set from the visual context alone. Instead they had to hear the head noun to confidently pick out the Target referent, effectively resorting to a ‘wait-and-see’ strategy. Interestingly, children were also not sensitive to the contrastive pitch accent manipulation:

early accent on the adjective in the split constituent construction had only a small facilitatory effect in identification of the Target referent.

This, however, does not mean that Russian-speaking children in this age range haven't yet mastered the function of the split constituent construction. Specifically, adding an explicit verbal reference to the other member of the color contrast set, which creates discourse support, is necessary for children's understanding that split constituents encode contrastiveness. This understanding then allows them to anticipate the Target referent even prior to hearing the head noun in discourse supported situations. This discourse support was especially helpful for split constituents rather than nonsplit constituents, suggesting fairly specific knowledge about the function of the split structure.

#### **4.1 *The Role of Context in Interpreting Contrastive Focus***

Our experiment provides a powerful demonstration that context does matter for children but this effect critically depends on the type of context. In line with the findings from previous child eyetracking experiments (Trueswell et al., 1999), the structure of the visual context alone is too subtle to exert its influence in the absence of explicit, definitive linguistic evidence in the form of referent names. Indeed, as mentioned in the introduction, a similar phenomenon has been observed in children's resolution of temporary syntactic ambiguity; a spoken discourse that highlights the relevant aspects of the visual referent world altered children's interpretation of temporary syntactic ambiguities (Hurewitz, 2001).

This observation is also in line with previous offline measures of children's understanding of non-canonical constructions. For instance, Otsu (1992) found that 3-4-year-

old Japanese children's poor (50%) accuracy in understanding scrambled sentences like *Sono ahiru-san-o kame-san-ga osimasita* ('It was the duck the turtle pushed') improved dramatically to 90% simply by preceding the sentence with verbal material that mentioned the fronted constituent (i.e., preceding it by a simple sentence like 'A duck was in the park', see also Murasagi & Kawamura, 2004). Costa and Szendrői (2006) also showed that given a supportive discourse context, Portuguese-speaking 3-to-5-year-old children could correctly interpret (79%) sentences with focus-related word order variation, with the indirect object scrambled over the direct one. Our own findings show that discourse structure elevates performance on contrastive expressions to such a level in children that it allows for adult-like anticipatory processing from just partial linguistic input.

#### **4.2 The Role of Prosody in Interpreting Contrastive Focus**

The present results suggest that Russian-speaking 5-to-6-year-old children have not fully mastered the interpretation of pitch accent. As mentioned in the introduction, such a finding would be expected under a cue-reliability and/or cue-salience account of acquiring the relevant linguistic properties associated with interpretation of contrastive focus in Russian. The split constituent construction in Russian is an obvious syntactic marking which appears to be obligatory for the contrastive focusing of a modifier such as an adjective. The prosodic evidence however is more ambiguous, as pitch accents frequently serve multiple interpretive roles, and are heavily context-dependent (e.g., interpretation depends on the syntactic environment of the pitch accent).

However, some of the existing literature suggests that the developmental delay of understanding contrastive pitch accents may extend beyond Russian and be a more general phenomenon (e.g., Baauw, Ruigendijk, & Cuetos, 2004; Cruttenden, 1974, 1984; Hüttner, Drenhaus, van der Vijver, & Weissenborn, 2004; McDaniel and Maxfield, 1992; Solan, 1980; Szendrői, 2003; Wells, Peppé, & Goulandris, 2004; Zuckerman, Vasić, & Avrutin, 2002). Moreover, a striking convergence can be seen when comparing the results of the present study with a series of experimental studies investigating children's contrastive interpretation of sentences with focus operators, such as *only*. Paterson and colleagues (Paterson et al., 2003) showed that English-speaking children in a wide range of ages from 4 to 12 years failed to construct a discourse model that included a contrast set and, therefore, were unable to correctly interpret sentences with the pre-verbal (e.g., *The fireman is only holding a hose*) and pre-subject (e.g., *Only the fireman is holding a hose*) focus operator *only*. Simply adding contrastive pitch accent to the associate of the focus operator does not help either, as Gualmini et al. (2003) showed using the truth value judgment task. They presented 4-to-5-year-old English-speaking children with sentences like *The farmer only sold BANANAS/bananas to Snow White/SNOW WHITE* and found that the children's interpretation of the sentences with *only* was adult-like when the indirect object (e.g., Snow White) but not when the direct object (e.g., bananas) was contrastively stressed. Like the present study, only by adding a supportive discourse context did children show correct interpretation with the narrowly focused direct object. Gualmini et al.'s conclusion that adult-like interpretation of sentences with *only* develops late in children was replicated for German children with *auch* (Hüttner, Drenhaus, van der Vijver, & Weissenborn, 2004) and Dutch children with *alleen* (Szendrői, 2003). In a language

where scope-related ambiguities with focus operators can be expressed both with scrambling (noncanonical word orders) and contrastive pitch accent shift, as in European Portuguese, 3-to-5-year-old children demonstrated adult-like comprehension of the sentences with the scrambled indirect object and *só* 'only' (e.g., *O Tigre só deu ao Piglet o jogo* 'It was only the game that Tigger gave to Piglet') but not of the canonical sentences with just pitch accent shift (e.g., *O Tigre só deu O JOGO ao Piglet*) (Costa & Szendrői, 2006). All of these studies suggest that a convergence of linguistic and discourse information is needed for children to make the appropriate interpretive commitments associated with contrastive focus.

Finally, despite the lack of an early effect of prosody, our findings do not contradict the child eyetracking experiment of Ito et al. (2008) in which Japanese-speaking children were found to show a facilitative effect of contrastive pitch expansion on reference resolution with color ADJ+N phrases. They, just like the Russian children, waited until they heard the noun to make a commitment and used prosody only at the offset of the referent name. Russian children only showed this delayed facilitatory effect of prosody for the nonsplit constructions, and of course, the Japanese study used only nonsplit constructions because split constructions are ungrammatical in the language. We suspect however, given the past difficulties in observing effects of prosody on children's interpretation of contrastiveness, that its use is quite delayed in comprehension and may even be subject to individual differences, perhaps into adulthood. Adults are after all not immune to variation in their linguistic abilities (and experience) that affect their language comprehension (Swets, Desmet, Hambrick, & Ferriera, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2008).

### **4.3 Closing Remarks**

As argued above, the accurate interpretation of Russian contrastive focus requires the integration of multiple linguistic and nonlinguistic sources of evidence; internal representations of prosodic, syntactic, semantic and discourse structure can sometimes align in the mind of the adult listener to the point that it allows for anticipatory referential processing. Here we have taken advantage of this phenomenon to study how children integrate the relevant sources of evidence. We have shown a pattern that we and others have observed in other domains of child language comprehension: Young children (ages 4-6 years) have considerable difficulty evoking the discourse presuppositions associated with specific syntactic structures, especially when this process must rely on their own perceptual organization of a visual referent world. Such real-time inferences appear to be beyond the scope of the developing sentence processing system. Rather, accurate performance requires the shaping of the child's mental organization of the world via the verbal discourse. By letting the speaker convey to the child how he is currently carving up the world along its multiple joints, the child can then make the appropriate interpretive leaps, sometimes even before all is said and done.

### **Acknowledgement**

We thank the three anonymous reviewers for their valuable comments and suggestions on the earlier version of this paper.

## Notes

1. In this paper we will use the autosegmental tone annotation system ToBI, Tones and Break Indices (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005). This annotation system claims there are five types of pitch accents categories in American English: H\*, L\*, L\*+ H, L + H\*, and H + !H\*. The letters H and L correspond to High and Low tone; the asterisk (\*) implies a connection between the tone and a stressed syllable; and the exclamation point (!) indicates a downstep.
2. A fuller description of the Russian prosodic system goes well beyond the scope of this article and dates back to Bryzgunova (1977) and the Russian Academic Grammar (Švedova, 1982). Recently, Odé (Odé, 2008) has created an adaptation of the ToBI system for Russian with an inventory of the tonal contrasts and unified notation called ToRI, Transcription of Russian Intonation. Other recent experimental investigations of Russian prosody (Makarova, 2007; Mehlhorn, 2004; Meyer & Mleinek, 2006) have provided detailed acoustic characteristics of different types of Russian sentences including contrastive constituents.
3. We used the MIXED procedure within the generalized linear model package of SPSS. For simplicity, we did not include control factors such as List or Target position in the models.

## References

- Baauw, S., Ruigendijk, E., & Cuetos, F. (2004). The interpretation of contrastive stress in Spanish-speaking children. In J. van Kampen & S. Baauw (Eds.), *Proceedings of GALA 2003*. Vol. 1. (pp. 103-114). The Netherlands: LOT.
- Baayen, R.H., Davidson, D. J. & Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412.
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457-474.
- Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*. (pp. 157-193). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Beckman, M., Hirschberg, J. & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. (pp. 9-54). Oxford – New York: Oxford University Press.
- Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective*. Lund: CWK Gleerup, LiberLaromedel.
- Bryzgunova, E. A. (1977). *Zvuki i intonatsiia russkoi rechi*. Moskva: Nauka.
- Costa, J., & Szendrői, K. (2006). Acquisition of focus marking in European Portuguese. Evidence for a unified approach. In V. Torrens & L. Escobar (Eds.), *The Acquisition of Syntax in Romance Languages*. (pp. 319-329). Amsterdam: John Benjamins.
- Cruttenden, A. (1974) An experiment involving comprehension of intonation in children from 7 to 10. *Journal of Child Language*, 1, 221-231.

- Cruttenden, A. (1984) Intonation comprehension in ten-year-olds. *Journal of Child Language*, 12, 643-661.
- Cycowicz, Y., Friedman, D., Rothstein, M., & Snodgrass, J. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, 65, 171-237.
- Goldsmith, J. (1976). *Autosegmental Phonology*. Ph.D. Dissertation, MIT.
- Gualmini, A., Maciukaite, S., & Crain, S. (2003). Children's insensitivity to contrastive stress in sentences with *only*. In S. Arunachalam, E. Kaiser, & A. Williams (Eds.), *Penn Working Papers in Linguistics*, Vol. 8(1).
- Höhle, B., Berger, F., Müller, A., Schmitz, M., & Weissenborn, J. (2009). Focus particles in children's language: Production and comprehension of *auch* in German learners from 1 year to 4 years of age. *Language Acquisition*, 16, 36-66.
- Hurewitz, F. (2001). *Developing the ability to resolve syntactic ambiguity*. Ph.D. dissertation. University of Pennsylvania.
- Hüttner, T., Drenhaus, H., van der Vijver, R., & Weissenborn, J. (2004). *The acquisition of the German Focus particle auch 'too': Comprehension does not always precede production*. Poster presented at the 28th annual Boston University Conference on Language Development.
- Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2008). *Use of emphatic intonation for contrast resolution in Japanese: Adults versus 6-year-olds*. Poster presented at the 21<sup>st</sup> Annual CUNY Conference on Human Sentence Processing. University of North Carolina, Chapel Hill.

- Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*, 541-573.
- Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, *34*, 391-405.
- Ladd, R. D. (1980). *The Structure of Intonational Meaning*. Bloomington, IN: Indiana University Press.
- Makarova, V. (2007). The effect of pitch peak alignment on sentence type identification in Russian. *Language and Speech*, *50*(3), 385-422.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372-380.
- McDaniel, D., & Maxfield, T. L. (1992). Principle B and contrastive stress. *Language Acquisition*, *2*(4), 337-358.
- Mehlhorh, G. (2004). The prosodic pattern of contrastive topic in Russian. In A. Steube (Ed.), *Information Structure: Theoretical and Empirical Aspects*. (pp. 241-258). Berlin – New York: Walter de Gruyter.
- Meyer, R., & Mleinek, I. (2006). How prosody signals force and focus — A study of pitch accents in Russian yes–no questions. *Journal of Pragmatics*, *38*, 1615-1635.
- Murasugi, K., and Kawamura, T. (2005). On the acquisition of scrambling in Japanese. In J. Sabel and M. Saito (Eds.), *The Free Word Order Phenomenon: Its Syntactic Sources and Diversity*. (pp. 221-242). Berlin – New York: Mouton de Gruyter.
- Odé, C. (2008). ToRI, A Transcription of Russian Intonation. An interactive research tool and learning module on the Internet. In P. Houtzagers, J. Kalsbeek, & J. Schaeken (Eds.),

- Dutch Contributions to the Fourteenth International Congress of Slavists, Ohrid: Linguistics*. Studies in Slavic and General Linguistics. Vol.34. (pp. 431-449). Amsterdam–New York: Rodopi.
- Otsu, Y. (1994). Case-Marking Particles and Phrase Structure in Early Japanese Acquisition. In B. Lust, M. Suner, & J. Whitman (Eds.), *Syntactic Theory and First Language Acquisition: Cross-linguistic Perspectives*, Vol. 1: *Heads, Projections and Learnability* (pp. 159-169). Hillsdale, NJ: Lawrence Erlbaum.
- Paterson, K. B., Liversedge, S. P., Rowland, C., & Filik, R. (2003). Children's comprehension of sentences with focus particles. *Cognition*, 89, 263-294.
- Pereltsvaig, A. (2008). Split phrases in colloquial Russian. *Studia Linguistica*, 62(1), 5-38.
- Pierrehumbert, J. (1980). *The Phonetics and Phonology of English Intonation*. Ph.D. Dissertation, MIT.
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. L. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication*. (pp. 271-312). Cambridge, MA: MIT Press.
- Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, 32(1), 3-23.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2), 109-147.
- Sekerina, I. A. (1997). The Scrambling Complexity Hypothesis and processing of split scrambling constructions in Russian. *Journal of Slavic Linguistics*, 7(2), 218-265.
- Sekerina, I. A., & Trueswell, J. C. (in prep.) *Processing of contrastiveness in Russian*.

- Sekerina, I. A., & Trueswell, J. C. (submitted). Processing of contrastiveness by heritage Russian-English bilinguals. *Bilingualism: Language and Cognition*.
- Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103-130.
- Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, 58, 574-608.
- Solan, L. (1980). Contrastive stress and children's interpretation of pronouns. *Journal of Speech & Hearing Research*, 23(3), 688-698.
- Švedova, N. Yu. (Ed.) (1982). *Russkaja Grammatika. II.* (in Russian. *Russian Grammar. II*). The USSR Academy of Sciences. Moscow: Nauka.
- Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136(1), 64-81.
- Szendrői, K. (2003). Acquisition evidence for an interface theory of focus. In J. van Kampen & S. Baauw (Eds.), *Proceedings of GALA 2003*. Vol. 2. (pp. 457-468). The Netherlands: LOT.
- Trueswell, J. C., Sekerina, I. A., Hill, N. and Logrip, M. (1999). The Kindergarten-Path effect: Studying on-line sentence processing in young children. *Cognition*, 73, 89-134.
- Trueswell, J. C., & Gleitman, L. R. (2004). Children's eye movements during listening: evidence for a constraint-based theory of parsing and word learning. In J. M. Henderson & F. Ferreira (Eds.), *Interface of Language, Vision, and Action: Eye Movements and the Visual World*. (pp. 319-346). New York, NY: Psychology Press.

- Trueswell, J. C., & Gleitman, L. R. (2004). Learning to parse and its implications for language acquisition. In G. M. Gaskell (Ed.), *Oxford Book of Psycholinguistics*. (pp. 635-656). Oxford: Oxford University Press.
- Trueswell, J. C., & Tanenhaus, M. K. (Eds.) (2004). *Approaches to Studying Word-Situated Language Use*. Cambridge, MA: MIT Press.
- Weber, A., Braun, B., and Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49(3)*, 367-392.
- Wells, B., Peppé, S, & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31(4)*, 749-778.
- Wells, J B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58(2)*, 250-271.
- Zuckerman, S., Vasić, N., & Avrutin, A. (2002). The syntax-discourse interface and the interpretation of pronominals by Dutch-speaking children. In S. Fish, & A. H.-J Do (Eds.), *Proceedings of the 26<sup>th</sup> Annual Boston University Conference on Language Development*. (pp. 781-792). Somerville, MA: Cascadilla Press.

## Tables

Table 1. *The average duration of the words in the experimental instructions (ms).*

Conditions		Example			
Sentence Type	Prosody	w1: ADJ	w2: N or V	w3: V or N	w4: LOC
NonSplit					
	Early Accent	KRASnuju	babočku	položite	v paket.
		873	708	509	
	Late Accent	Krasnuju	BAbočku	položite	v paket.
		556	907	515	
Split					
	Early Accent	KRASnuju	položite	babočku	v paket.
		859	531	632	
	Late Accent	Krasnuju	položite	BAbočku	v paket.
		550	534	895	

Table 2. *Effects of Structure, Visual Context and Prosody on E-Logit Target Advantage Scores.*

WordPos	Effect	Numerator Denominator		F	Sig.
		<i>df</i>	<i>df</i>		
w2a	<b>Sentence (S)</b>	<b>1</b>	<b>672</b>	<b>6.85</b>	<b>&lt;0.01 **</b>
	Visual Context (VC)	1	673	3.62	0.06
	Prosody (P)	1	674	0.01	n.s.
	S x VC	1	673	1.63	n.s.
	<b>S x P</b>	<b>1</b>	<b>674</b>	<b>5.63</b>	<b>&lt;0.05 *</b>
	S x VC x P	2	673	0.38	n.s.
w2b	<b>Sentence (S)</b>	<b>1</b>	<b>646</b>	<b>30.6</b>	<b>&lt;0.001 ***</b>
	Visual Context (VC)	1	645	0.02	n.s.
	<b>Prosody (P)</b>	<b>1</b>	<b>646</b>	<b>7.36</b>	<b>&lt;0.01 **</b>
	S x VC	1	646	0.74	n.s.
	S x P	1	646	0.27	n.s.
	S x VC x P	2	646	0.07	n.s.
w3a	<b>Sentence (S)</b>	<b>1</b>	<b>671</b>	<b>38.0</b>	<b>&lt;0.001 ***</b>
	Visual Context (VC)	1	672	0.02	n.s.
	Prosody (P)	1	673	1.41	n.s.
	S x VC	1	672	0.00	n.s.
	S x P	1	672	0.92	n.s.
	S x VC x P	2	673	0.12	n.s.

w3b	<b>Sentence (S)</b>	<b>1</b>	<b>644</b>	<b>11.95</b>	<b>&lt;0.001 ***</b>
	Visual Context (VC)	1	643	0.86	n.s.
	Prosody (P)	1	645	0.00	n.s.
	<b>S x VC</b>	<b>1</b>	<b>644</b>	<b>6.19</b>	<b>&lt;0.05 *</b>
	S x P	1	645	1.89	n.s.
	S x VC x P	2	644	0.73	n.s.

---

## Figures

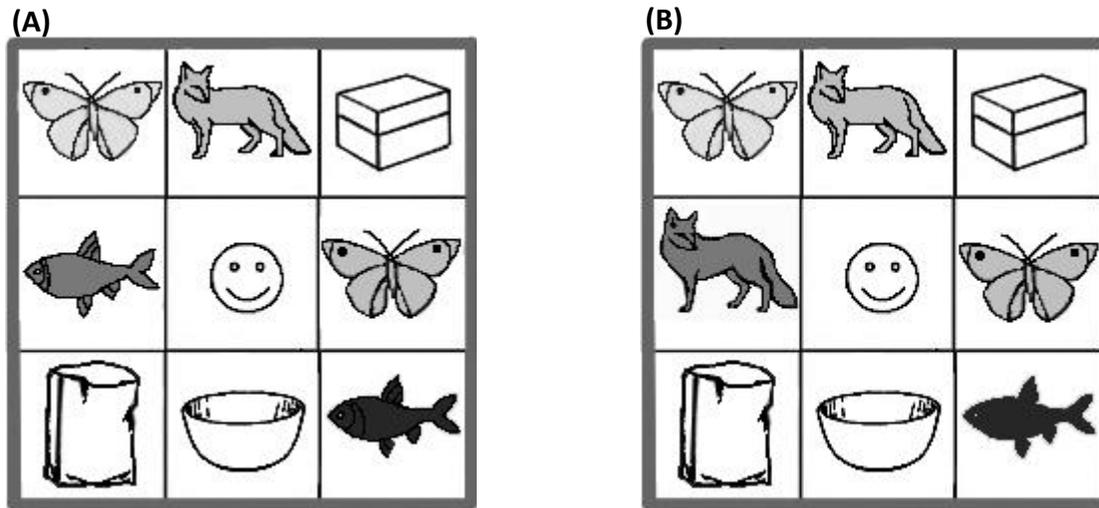
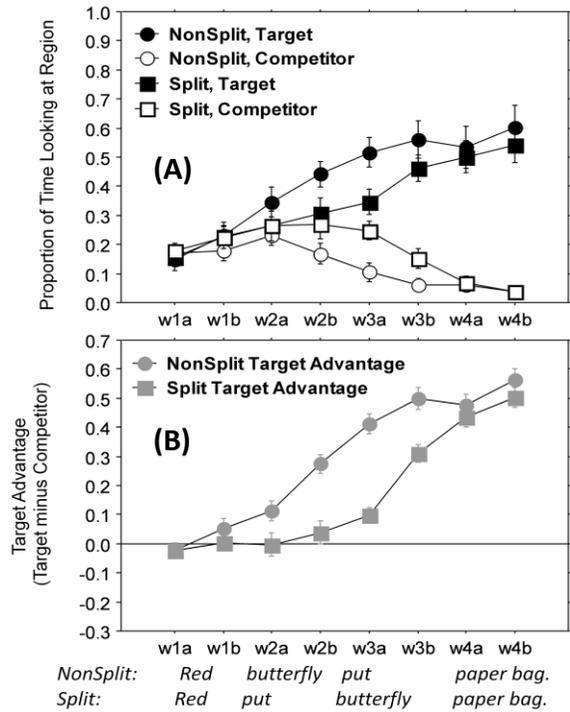


Figure 1. The two types of visual context. Panel A: 1-Contrast condition: Target *red butterfly* (position 6), Color Competitor *red fox* (position 2), target's contrast object *purple butterfly* (position 1), two distracters, *gray fish* and *blue fish* (positions 4 and 9, respectively). Panel B: 2-Contrast condition is the same except that the gray fish (position 4) has been replaced with a gray fox that serves as the contrast object for the Color Competitor.



*Figure 2.* Panel A: The proportion of time spent looking at the Target (filled symbols) and the Competitor (open symbols) in response to NonSplit (circles) and Split (squares) constituents. Panel B: Target advantage scores, i.e., the proportion of time spent looking at the Target minus the proportion of time spent looking at the Competitor. (Plotted are subject means, with 95% confidence intervals.)

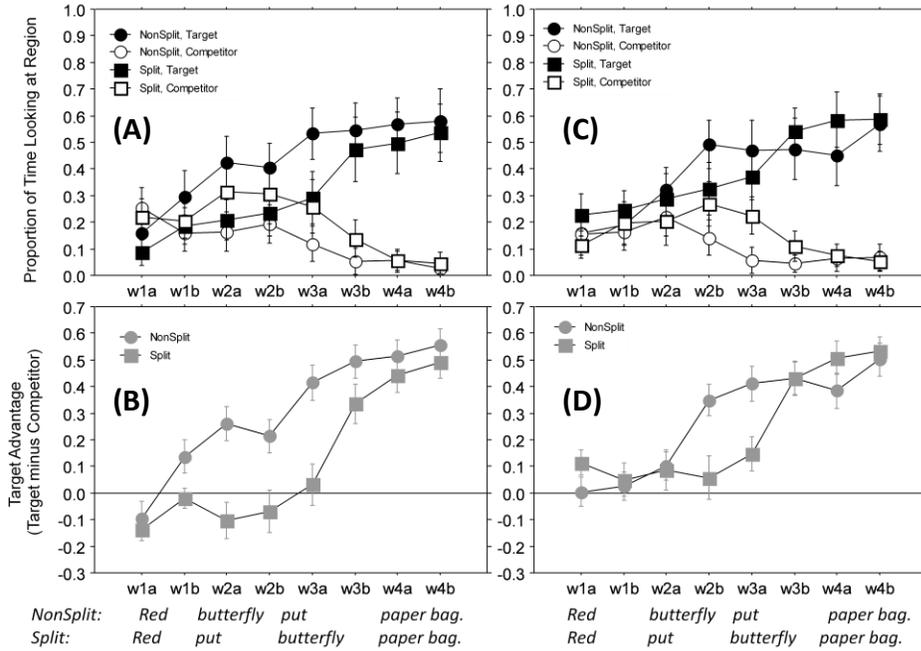


Figure 3. Results from (the referentially unambiguous) 1-Contrast visual contexts, when pitch accent was placed on the adjective (Panels A & B) and when pitch accent was instead placed on the head noun (Panels C & D). (Plotted are subject means, with 95% confidence intervals.)

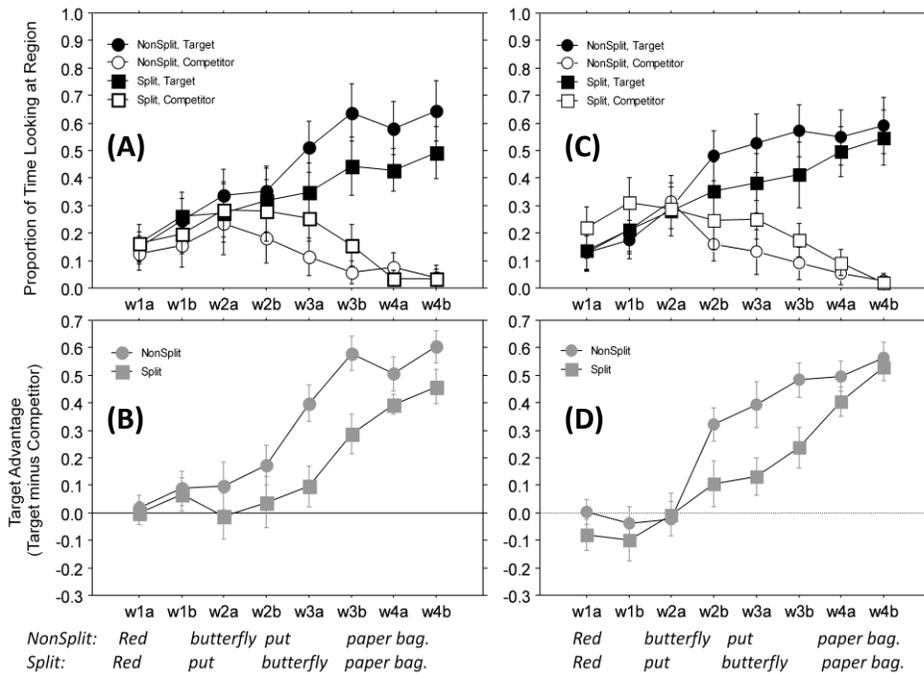


Figure 4. Results from (the referentially ambiguous) 2-Contrast visual contexts, when pitch accent was placed on the adjective (Panels A & B) and when pitch accent was instead placed on the head noun (Panels C & D). (Plotted are subject means, with 95% confidence intervals.)

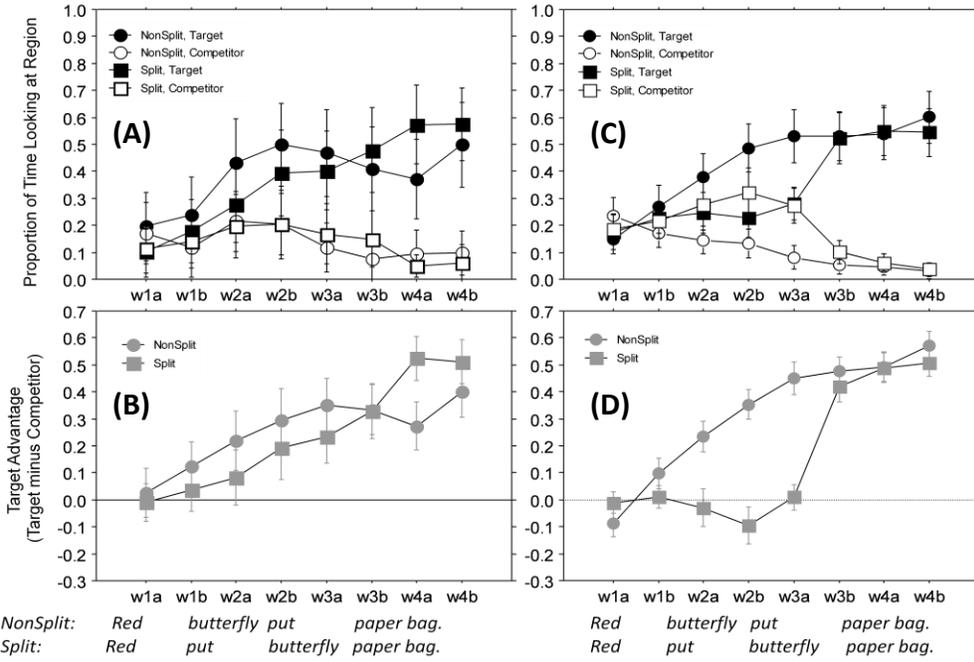


Figure 5. Results from (the supportive) 1-Contrast contexts, when the discourse provided the relevant color contrast (Panels A & B) and when it did not (Panels C & D). (Plotted are subject means, with 95% confidence intervals.)

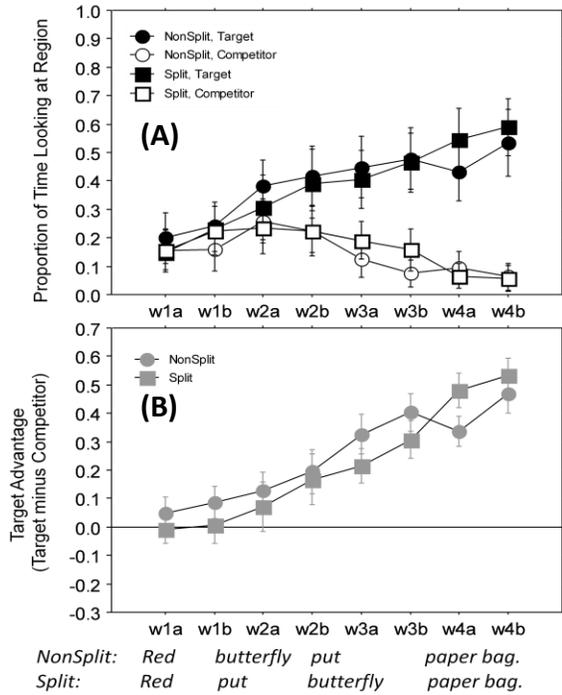


Figure 6. Results from when the discourse provided the relevant color contrast, collapsing across 1-Contrast and 2-Contrast contexts. (Plotted are subject means, with 95% confidence intervals.)