RESEARCH

Quantifier spreading in child eye movements: A case of the Russian quantifier *kazhdyj* 'every'

Irina A. Sekerina[1] and Antje Sauermann[2]

[1] College of Staten Island and The Graduate Center, City University of New York, US

[2] Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, DE

Corresponding author: Irina A. Sekerina (irina.sekerina@csi.cuny.edu)

Extensive cross-linguistic work has documented that children up to the age of 9–10 make errors when performing a sentence-picture verification task that pairs spoken sentences with the universal quantifier *every* and pictures with entities in partial one-to-one correspondence. These errors stem from children's difficulties in restricting the domain of a universal quantifier to the appropriate noun phrase and are referred in the literature as *quantifier-spreading* (*q*-spreading). We adapted the task to be performed in conjunction with eye-movement recordings using the Visual World Paradigm. Russian-speaking 5-to-6-year-old children (*N* = 31) listened to sentences like *Kazhdyj alligator lezhit v vanne* 'Every alligator is lying in a bathtub' and viewed pictures with three alligators, each in a bathtub, and two extra empty bathtubs. Non-spreader children (*N* = 12) were adult-like in their accuracy whereas *q*-spreading ones (*N* = 19) were only 43% correct in interpreting such sentences compared to the control sentences. Eye movements of *q*-spreading children revealed that more looks to the extra containers (two empty bathtubs) correlated with higher error rates reflecting the processing pattern of *q*-spreading. In contrast, more looks to the distractors in control sentences did not lead to errors in interpretation. We argue that *q*-spreading errors are caused by interference from the extra entities in the visual context, and our results support the processing difficulty account of acquisition of quantification. Interference results in cognitive overload as children have to integrate multiple sources of information, i.e., visual context with salient extra entities and the spoken sentence in which these entities are mentioned in real-time processing.

Keywords: language development; attention; eye movements; Russian; quantifier spreading

1 Introduction

How children understand sentences with quantifiers, i.e., words like *all, every, some, most*, and the like, is one of the most prominent areas in language acquisition, and rightly so: although quantifiers constitute a small set of closed-class elements of grammar, they are critical in conveying quantities of things and individuals and their grouping into sets. Inhelder and Piaget (1964) were the first to report persistent errors in children's understanding of sentences with the quantifier *all*. When presented with a series of three red squares, two blue squares and two blue circles and asked a question in (1), 5-to-9-year-old children were accurate only in 63% to 91% of the cases (Inhelder & Piaget 1964: 60–63):

(1)   *Are all the circles blue?*   Child's answer: *No, there are only two.*

Philip's dissertation (1995) revived an interest in this topic, and for the past 20 years, acquisition of universal quantifiers has grown into a subfield that spans across multiple

dimensions. Since then the extensive cross-linguistic work has documented that regardless of the language, children up to the age of 9–10 make three types of errors in sentences with the universal quantifier *every*—overexhaustive (classic), underexhaustive, and *bunny*-type, collectively referred to as *quantifier spreading* or *q-spreading* (Brooks & Braine 1996; Rakhlin 2007). In a typical sentence-picture verification experiment, children see a picture with two sets of entities, for example, alligators lying in bathtubs. Critically, these entities (i.e., alligators and bathtubs) are in a partial one-to-one correspondence (Figure 1B). Children are asked a yes/no question presented out of context that contains the quantifier *every,* e.g., *Is every alligator in a bathtub?* in the visual context when there are three alligators in bathtubs and an extra empty bathtub in the picture, many children erroneously make the widely attested *q*-spreading error referred to the *overexhaustive (or classic) pairing* by Roeper and colleagues (Roeper, Strauss & Pearson 2004): they say 'No' as they fail to consider the extra empty bathtub(s). In the context when there are extra alligators, some children say 'Yes', thus, making an *underexhaustive pairing q*-spreading error (Drozd 2001). Finally, the *bunny*-spreading errors (Roeper et al. 2004) are committed by very young children when they hear a question *Is every bunny eating a carrot?* that is paired with the picture containing two different events (e.g., three bunnies eating a carrot each and one dog eating a bone). The present study focuses on the overexhaustive *q*-spreading errors that Russian-speaking children make in a standard sentence-picture verification task while their eye movements are recorded.

1.1 Theoretical accounts of q-spreading errors

Age-related errors children make in interpreting the quantifier *every* have been explained by three different hypotheses: faulty logical reasoning, partial linguistic competence, and full linguistic competence coupled with processing difficulty. Inhelder and Piaget (1964) accounted for *q*-spreading errors as resulting from their general theory of gradual development of logical reasoning. Within the functional language acquisition work, it was proposed that for distributive quantifiers such as *each* and *every,* children expect two entities in partial one-to-one correspondence to be a perfect (exhaustive) one-to-one pairing because they prefer to associate the quantifier with the pairing of the two entities, and not with the NP that is modified by the quantifier (Brooks & Braine 1996). Within the generative theory of language acquisition (see Rakhlin 2007 for an overview), Philip (1995) argued that an explanation should be sought in children's non-adult-like grammar rather than in their faulty logical reasoning. His Event Quantification account was further developed and modified by subsequent linguistic hypotheses, e.g., weak quantification (Drozd & van Loosbroek 1998), noncanonical mapping from syntax to semantics (Geurts 2003), and syntactic restructuring (Roeper et al. 2004; Roeper 2007). The faulty logical reasoning and partial competence accounts, however, face two major problems: they need to explain how children move from non-adult-like to adult-like interpretation of universal quantification, and why there are large discrepancies in error rates across different studies that use different materials and designs.

In contrast to the partial competence hypothesis, advocates of the full competence hypothesis (Poeppel & Wexler 1993) argue that children's knowledge of universal quantifiers is completely adult-like, but their *q*-spreading errors are caused by processing difficulties children experience when dealing with pragmatic infelicity and other task demands of the design. Crain and colleagues (Crain et al. 1996; Meroni, Gualmini & Crain 2007) showed that children make errors only when the experimental design does not satisfy the felicity conditions of asking yes/no questions, e.g., when the pictures are paired with the sentences without felicitous contextual support. Rakhlin (2007) proposed that in the absence of a felicitous linguistic context, sentences like *Every alligator is in a bathtub*

are ambiguous with respect to how to interpret the indefinite NP *a bathtub*. It can be interpreted as either a non-singleton {3 pairings of alligator-bathtub and a single bathtub} or a singleton {a single bathtub}. In the absence of a fairly rich linguistic context and in presence of a misleading visual context where a single empty bathtub is perceptually salient, children often choose the singleton interpretation of *a bathtub* and arrive at the erroneous semantic interpretation in which the indefinite *a* takes a wide scope over the quantifier *every* (e.g., "There is a bathtub y in the set such that every contextually relevant alligator x lies in y.") (Rakhlin 2007: 252).

O'Grady and colleagues (O'Grady, Suguzi & Yoshinaga 2010) demonstrated that a simple methodological change, i.e., removing the perceptually salient visual context by getting rid of the picture and allowing children to perform an act-out task, can eliminate *q*-spreading errors altogether. Instead of pictures, the authors used cutouts (e.g., 3 cats and 4 fish) and asked the Japanese 5-year-old children (20 participated, 11 included in the analysis) to act out the sentence *Every cat is biting a fish*. This manipulation resulted in a near-perfect performance (0.82 wrong out of 4 items) as the children exhaustively paired each of the three cats with one of the three fish and did not touch the extra fish. Three days earlier, the same children made many *q*-spreading errors (3.73 out of 4) in the standard sentence-picture verification task. However, while these results clearly demonstrate that children can be easily distracted by extra available entities, the act-out task may actually mask *q*-spreading. As children are strongly biased toward an exact one-to-one correspondence, the individual cutouts of cats and fish in the study by O'Grady and colleagues lent themselves perfectly to an exhaustive pairing.

The fact that the act-out task may mask *q*-spreading emphasizes the necessity to investigate the role that visual context (i.e., pictures) plays in a novel way. Minai and colleagues (Minai et al. 2012) did just that by employing the Visual World Paradigm (VWP; Trueswell & Tanenhaus 2004) to study what eye movements of Japanese-speaking children tell us of how they process quantified sentences in the presence of a visual context with partial one-to-one correspondence between two entities. They argue that the relative salience of extra objects in the pictures imposes perceptual restrictions on the domain of application of quantifiers, and children's attention is attracted to them ultimately leading to *q*-spreading errors. In their experiment, children in the target group first viewed a block that included 4 pictures with multiple extra objects (3 turtles each holding an umbrella and 3 extra umbrellas) followed by a block that included 4 pictures with a single extra object (3 turtles each holding an umbrella and 1 extra umbrella). The latter is the standard sentence-picture verification set-up in which *q*-spreading errors usually manifest themselves whereas the former helps reduce such errors by manipulating visual salience of multiple extra objects in contrast to just one (Rakhlin 2007: 244). In the control group, the order of the blocks was reversed. The children in the control group made 84% *q*-spreading errors whereas those in the target group only 48%. Thus, the initial exposure to three extra objects in the first block helped them to disengage from the additional extra object presented in the second block, and their rate of *q*-spreading errors was significantly lower compared to the control group.

Nevertheless, eye movements of the children in both groups revealed that they fixated the extra objects (1 or 3 umbrellas) more than the control adults. These fixations peaked at 1.5 s *before* the sentence began, i.e., during the silent picture preview phase. The latency, magnitude, and duration of these fixations were modulated by two factors: first, whether the child made *q*-spreading errors or not, and, second, his or her score on the dimensional change card sort task (DCST). The latter was operationalized to approximate children's abilities in switching attention. A correlation was found among the amount of *q*-spreading errors, the DCST score, and eye movements in that the children who scored

low on attention switching abilities, made more *q*-spreading errors and demonstrated increased fixations to the extra objects. Minai and colleagues concluded that not quite developed ability to control attention, i.e., difficulties with disengaging from the visually salient extra objects, is a critical factor responsible for *q*-spreading errors in young monolingual children.

There are two important issues that weaken the conclusions of Minai et al. (2012). First, they employed the between-participants design (control vs. target group) and then divided 29 children in the control group into 5 who made a few *q*-spreading errors (0.75 accuracy) and 24 who made many *q*-spreading errors (0.02 accuracy) (Table 2: 933). Their eye movements (i.e., the proportion of looks to the extra objects) were averaged despite the fact that these two groups were so imbalanced, and it makes the interpretation of the results difficult. Second, the peak of looks to the extra objects happened in the preview phase, i.e., 1.5 s before the sentence began; after that, during the actual sentence presentation, the children looked at them at chance level (except in the silence region). This lack of tight coupling of the eye movements to the extra objects and the appearance of the quantifier in speech that is the hallmark of the VWP in the study of Minai and colleagues does not allow us to establish the causal relationship between allocation of visual attention and *q*-spreading.

1.2 Purpose and predictions of the current study

The current study strives to overcome the shortcomings of Minai et al. (2012) by contrasting eye movements of Russian-speaking adults and 5-to-6-year-old children who are adult-like, i.e., non-spreaders ($N = 12$), and who are not, i.e., *q*-spreading children ($N = 19$). We employed the standard sentence-picture verification task coupled with the VWP and paired the same spoken quantified sentence with two types of pictures: a control picture (alligators and bathtubs in exact one-to-one correspondence plus two distractor elephants: Figure 1A) or an experimental picture (alligators and bathtubs in partial one-to-one correspondence with two extra empty bathtubs: Figure 1B). In contrast to Minai et al., our design was within-participants as we exposed adult and child participants to both conditions. Our goal was to investigate how adults (control group) and two groups of children (non-spreaders and *q*-spreading children) allocate visual attention to the two distractor objects or extra containers in the pictures in a moment-by-moment fashion as the spoken sentence unfolded making it possible to capture *q*-spreading in real-time. We also did not average across all of experimental sentences, but analyzed the correctly and incorrectly interpreted sentences separately.

The predictions were based on two measures, i.e., accuracy of performance on the sentence-picture verification task and patterns of eye movements. First, following the previous behavioral studied with Russian children (Kuznetsova et al. 2007; Katsos et al. 2012) (see Section 2 below), we expected our *q*-spreading children to demonstrate a large number of errors in the experimental condition, but not in the control one. Second, we hypothesized that being distracted by the extra empty containers in the form of more looks to them in the experimental sentences would lead to *q*-spreading errors. Moreover, the critical point in the sentence where *q*-spreading should manifest itself in eye movements would be the verb region (e.g., *lezhit* 'is lying') because it immediately follows the appearance of the quantified NP *kazhdyj alligator* 'every alligator' in speech. As children hear *Kazhdyj alligator lezhit …* 'Every alligator is lying …' they should begin anticipating the mention of containers and look at the extra bathtubs because the verb semantics in combination with the picture presupposes that the alligators are placed into (or paired with) the bathtubs. In the control sentences where the visual context contains

distractor objects, adult and children participants should look less at them, if at all; after all, elephants are not mentioned in the sentence. Thus, we predicted that increased looks to the extra containers in comparison to the distractor objects are the cause of *q*-spreading errors in interpretation.

In general, both non-spreaders and *q*-spreading children would look more than adults at the *both* distractors and extras (Nation, Marshall & Altmann 2003), but visual context with extra containers should lead to *q*-spreading errors *only* when it is combined with the spoken sentence that mentions these extra containers in the scope of the quantifier *every* (e.g., preceded by *every alligator*), i.e., in the overexhaustive trials. The extra containers that are explicitly mentioned in the spoken sentence create interference by attracting visual attention subsequently leading to an incorrect interpretation of the entire sentence. Critically, this cognitively taxing interference should arise only when children have to integrate complex visual context with the spoken quantified sentences explaining why *q*-spreading is found in the sentence-picture verification task, but not in act-out (Rakhlin 2007; O'Grady et al. 2010; Minai et al. 2012). The current study tests the hypothesis that correct interpretation of quantifiers depends on how successfully a child is able to prioritize these different sources of information, inhibit interference from irrelevant visual details and revise an interpretation in real-time. Its novelty lies in pinpointing where in the sentence children commit to erroneous interpretation of the quantified sentences and how such errors reflected in eye movements lead to *q*-spreading.

2   The Russian quantifier *kazhdyj* 'every' and its acquisition

The Russian universal quantifier *kazhdyj* 'every-NOM-SG-MASC' is inflected for case, number, and gender so for the purposes of grammatical agreement it behaves like an adjective that agrees with the head noun it modifies. It assigns a standard distributive function to the noun it occurs with. Note that Russian does not have articles, making the definite-indefinite distinction as well as NP/DP distinction complicated (*cf.* Rakhlin 2007). In the experimental sentences used in the current study, the quantified NP *every* 'kazhdyj' + N (e.g., *kazhdyj alligator* 'every.NOM-SG-MASC alligator.NOM-SG-MASC') was always in the subject position, and the word order of the sentences themselves was canonical, as illustrated in (2).

(2)   Kazhdyj   alligator   lezhit   v vanne.
      every.NOM-SG-MASC alligator.NOM-SG-MASC is lying.SG-PRES in bathtub.PREP
      'Every alligator is lying in a/the bathtub.'

Two previous experimental studies (Kuznetsova et al. 2007; Katsos et al. 2012) have established the pattern of *q*-spreading errors with the quantifier *kazhdyj* 'every' in Russian-speaking children. Kuznetsova and colleagues tested 42 children, 10 of them with specific language impairment (SLI), in a wide age range of 4 to 12 years. In their modified sentence-picture verification task, they presented the children with sentence-picture sets in three conditions, two of which were with the quantifier *kazhdyj* 'every' in the subject position (e.g., *Kazhdaja devochka est morozhenoe* 'Every girl is eating an ice-cream') and in the object position (e.g., *Mal'chik neset kazhdyj chemodan* 'The boy is carrying every suitcase'). The three pictures included the "symmetrical" one (3 girls eating one ice cream each), the "asymmetrical" one (4 girls but only three are eating ice cream), and a distractor. They reported no differences between the typically-developing and SLI children and found the expected pattern of errors: the children provided adult-like responses in 92% of bare plurals, 79% of the quantified subject NPs, and 60% of the quantified object NPs (Kuznetsova et al. 2007: 230). Thus, Russian-speaking 4-to-12-year-

old children are not an exception to the general cross-linguistic trend: as expected, they make *q*-spreading errors in the standard sentence-picture verification task.

3 The experiment

The experiment described below is a part of a larger study in which three groups of participants took part—monolingual Russian adults (control), bilingual heritage Russian-English-speaking adults, and monolingual Russian children. Due to the developmental limitations, the children's experiment was simpler (i.e., only two conditions and 48 fillers) than the adults' (three conditions and 66 fillers). The adult experiment in its entire form is reported in Sekerina and Sauermann (2015). Here, for the control monolingual adult group, we provide only the necessary details for a meaningful comparison with the child data.

3.1 Method
3.1.1 Participants

Monolingual Russian adults: young college-age undergraduate students ($N = 40$, 10 men; mean age 21.5) from the St.-Petersburg State University (Russia) volunteered to participate in the experiment in exchange for $3 (equivalent in rubles).

   Monolingual Russian children: thirty-one children (12 boys; mean age 6;0, range 5;1–6;11) were recruited and tested at a large preschool center in Moscow, Russia. 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 ($3 value) for their participation in the experiment. The experiment lasted 20 minutes on average.

   This study was carried out in accordance with the ethical principles of psychologists and code of conduct of the American Psychological Association and was approved by the Institutional Review Board of the College of Staten Island (#01-N-08). All participants, adults and children's parents, gave written informed consent in Russian in accordance with the Declaration of Helsinki.

3.1.2 Design and materials

The experiment included eight experimental items, 48 fillers of two types, and three practice items. The full set of materials, pictures and spoken sentences is available for download at IRIS, a digital repository of data collection instruments for research in second language learning.[1]

   Each item consisted of a picture paired with a spoken sentence (see Figure 1). All of the pictures were color drawings of entities, animals (e.g., alligators, dogs, birds) or inanimate objects (e.g., bananas, candles, spoons), and containers (e.g., bathtubs, baskets, nests). There were two types of pictures representing the Type of Picture factor, a control one and an experimental one. In both types, an animal (or inanimate object) was drawn in an appropriate container (e.g., an alligator lying in the bathtub; a banana in the basket), and this pairing was repeated three times in the front of the picture. The experimental pictures differed from the control ones by what was depicted in the back, as they contained two extra empty containers (e.g., two bathtubs in Figure 1B). In the control pictures, the two extra objects in the back were distractors different from the ones in the front (e.g., two elephants in Figure 1A). No *bunny*-spreading errors were expected with the control pictures because the two distractor elephants depicted did not lend themselves to an event interpretation parallel to the alligators lying in the bathtubs. Thus, the experimental pictures showed a partial one-to-one correspondence of the entities (three alligators in

---

[1] http://www.iris-database.org/iris/app/home/search?query = sekerina.
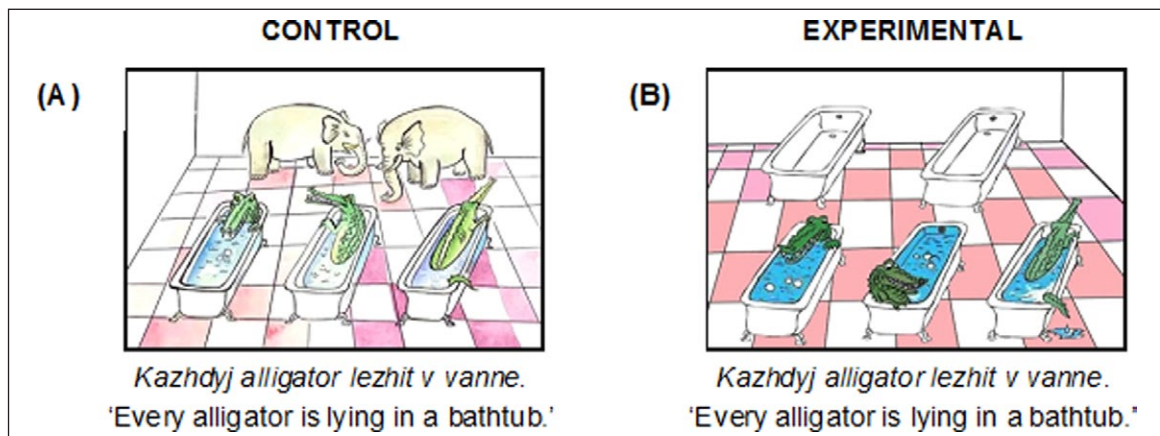
Figure 1: Two pictures corresponding to the two conditions in the present experiment. **(A)** Control, **(B)** Overexhaustive.

three bathtubs and two more bathtubs) whereas the control pictures depicted them in strict one-to-one correspondence (three alligators in bathtubs) with distractor objects (two elephants).

The experimental and control versions of the pictures were always paired with the same spoken sentence in (2). The quantifier *every* was the first word in the sentence as the quantified NP was always the subject. The verbs used were the present tense forms of unaccusative verbs in the imperfective aspect, e.g., *sidit* 'is sitting', *lezhit* 'is lying', *stoit* 'is standing', *visit* 'is hanging', and *spit* 'is sleeping' repeated in several items. The English translation of (2) is in the present progressive tense because in Russian, it is not distinguishable from the simple present tense. The spoken sentences were recorded individually by a female native speaker of Russian (the first author), using mono-mode sampling at 22,050 Hz and were pronounced with normal adult speed.

The 48 filler items were of two types. One type ($N = 24$) was reversible pictures depicting a transitive action with two animals that were paired with either SVO or OVS sentences, 12 each (e.g., 'The squirrel is walking the duck' or 'The duck, the squirrel is walking'). The other type ($N = 24$) was pictures very similar to the experimental ones, depicting either the same animals but in different containers (e.g., five alligators sleeping in beds) or different animals in the same containers (e.g., five ducks in bathtubs). The spoken sentences referred to either the number of the pairings, their color, or included a comparison (e.g., 'Six alligators are sleeping in beds', 'There are more pink flowers than blue ones').

Two versions of the experiment were created for the within-participants design, with eight experimental items rotated through the two conditions in a Latin-square design.

3.1.3 Procedure

The task that children performed was sentence-picture verification. The stimuli materials (the picture and the spoken sentence) were programmed into a script run by DMDX, a free Windows-based software for language processing experiments (Forster & Forster 2003), and were presented on a 19-inch HP laptop computer to which a remote eye-tracking camera was attached. On each trial, the picture appeared on the screen as soon as the spoken sentence started playing. The children were asked to provide a 'yes' answer if they thought that the sentence correctly described the picture, and a 'no' answer otherwise. To record accuracy, this sentence-picture verification task was performed using a gamepad attached to the stimuli laptop computer. The gamepad had three buttons, 'yes', 'no', and 'next'. Only a few children, mostly the oldest ones, were able to manipulate the gamepad

successfully, as was determined during the practice trials. The rest of the child participants provided their answers by saying 'yes' or 'no' out loud while the experimenter used the gamepad to record their answers and to advance the experiment to the next trial. Thus, only the accuracy data and not the RTs were collected for the children.

Children's eye movements were recorded using the ISCAN remote portable eye-tracking system. Eye movements were sampled at a rate of 30 times per second and were recorded on a digital SONY DSR-30 video tape-recorder. Spoken sentences were played through speakers connected to the stimuli computer and were recorded simultaneously with eye movements. Each child underwent a short calibration procedure prior to the experiment.

3.1.4 Data treatment and analyses

We conducted two types of analyses: the analyses of accuracy and fine-grain analyses of the eye movements. Trials that were not recorded due to the equipment malfunctioning constituted the missing data for the eye-movement analyses, 2.3% (13 trials).

Eye movements were extracted from videotape using a SONY DSR-30 video tape-recorder with frame-by-frame control and synchronized video and audio. For each trial, four categories were coded: looks to the three pairs of entities in the front of the picture (identical in both conditions, see Figure 1); looks to the two objects or containers in the back (elephants in the control pictures, extra bathtubs in the experimental pictures); looks elsewhere in the picture, and track loss. The two empty extra containers (i.e., bathtubs) in the experimental condition will be referred to as *Extras* (see Figure 1B) and the two elephants in the control conditions will be referred to as *Distractors* (see Figure 1A). Track loss (9.2%) and looks elsewhere (1.4%) constituted a small proportion of total looks and were removed from the eye-movement analyses; thus, fixations to the three pairs in the front were in complimentary distribution with fixations to the Extras/Distractors in the back. As we hypothesized that the allocation of visual attention to irrelevant Extras in the experimental pictures was responsible for children's *q*-spreading errors, we chose to analyze the looks to the two Extras/Distractors in the back of the pictures.

In the fine-grain analyses, the proportions of looks were analyzed in separate time windows or *regions of interests* (ROIs). Each spoken sentence was segmented into four ROIs that were defined relative to the onset of the three phrases (3). For each ROI on each trial, we computed the proportions of looks to the Distractors (two elephants) or Extras (two empty bathtubs), and all eye-tracking graphs will use the proportion of time the participants spent looking at them as the dependent variable.

(3)

ROI 1	ROI 2	ROI 3	ROI 4
Subject '*Every* N'	Verb	Locative PP	Silence
Kazhdyj alligator	*lezhit*	*v vanne.*	After the offset of the last word
'every alligator'	'is lying'	'in a bathtub'	until the response via button push

Linear mixed-effects models (LME models) were used to conduct inferential statistical analyses of the accuracy and eye-movements data using the *lmer* function of the *lme4* package (Bates, Maechler & Dai 2009) in the R environment (R Development Core Team 2009). They correspond to multiple regression analyses that take into account variation due to participants and items. LME models were particularly appropriate for our experimental data because they do not require a balanced design.

For both accuracy and the eye-movements we report the results of the LME models for the child data as well as the conjoined models that assessed the differences between

control adults and children. The LME models estimated the fixed effects of the independent variables (see below) and the random effects of participants and items, but we report only the results for the fixed effects.

The LME models for the child data estimated the fixed effects of Type of Picture (Control versus Experimental) on the probability of correct answers (accuracy data) and the fixed effects of Accuracy, Type of Picture and the interaction between Accuracy and Type of Picture on the arcsin-transformed proportions of looks to the Extra/Distractors (eye-movement data). In the models assessing differences between adults and children, Age and the interaction between Age and Type of Picture (and, for eye-movements, Accuracy) were included as fixed factors. We used the *pvals.fnc* function of the language R package (Baayen 2008) to generate *p*-values for the eye-movement data. Sum contrasts were set for the fixed factors in each model, resembling the contrast coding used in traditional ANOVAs.

The reported LME models only considered a restricted set of random effects, i.e., intercept adjustment for participants and items because only for these models *p*-values could be generated. We conducted a model fitting procedure in which the LME models with a full factorial set of random effects (random slope-adjustments) were trimmed down in a step-wise fashion using log-likelihood tests for model comparisons (see Baayen 2008). This fitting procedure almost always led to the simpler models that included only intercept adjustment. In the remaining cases, the significance of the fixed effects ($\alpha = .05$) did not differ between the models reported and the final models of the fitting procedure.

3.2 Results

To simplify the description of the statistics for the two types of behavioral measures—accuracy and fine-grain eye-movement data—as well as to provide the most straightforward comparison between adults and children, we report the statistics of the fixed effects of the mixed-effect models in tables. Each table consists of two panels, the left one shows the results for children, and the right one, the direct comparison of adults and children. The statistical components included are estimates (*b*), standard errors (*SE*), *z*- or *t*-scores, and *p*-values. Significant effects and interactions are shown in bold (*Acc* stands for accuracy, *Type of Pic* for Type of Picture).

3.2.1 Accuracy

Fillers: the children's accuracy was solid, 92.3% (range 69.2–100%), and there were 10 children who made no mistakes at all. We take these data to mean that the children were able to successfully perform our sentence-picture verification task.

Experimental Items: the mean accuracy for adults and both groups of children are represented in Figure 2, and their statistical analysis in Table 1, respectively.

Adults' accuracy in the sentence-picture verification task (black bars, Figure 2) did not differ as a factor of the Type of Picture (Control: 95%, Overexhaustive: 97%; $b = -0.55$, $z = -0.867$, $p = .386$). For all of the children combined ($N = 31$), we found the expected pattern of *q*-spreading errors: They were significantly more accurate with the control (90%) than with the overexhaustive pictures (65%), as revealed by a main effect of Type of Picture.

When we separated the children into two groups based on their accuracy for the Overexhaustive condition, there were 12 non-spreaders, i.e., children who performed at ceiling (accuracy: 100%, dark gray bars) and 19 who made one or more errors out of the 4 relevant trials (accuracy: 43%, light gray bars, Figure 2). We will refer to them as *q*-spreading children and analyzed them separately because they allowed us to establish a pattern of eye movements reflective of *q*-spreading.
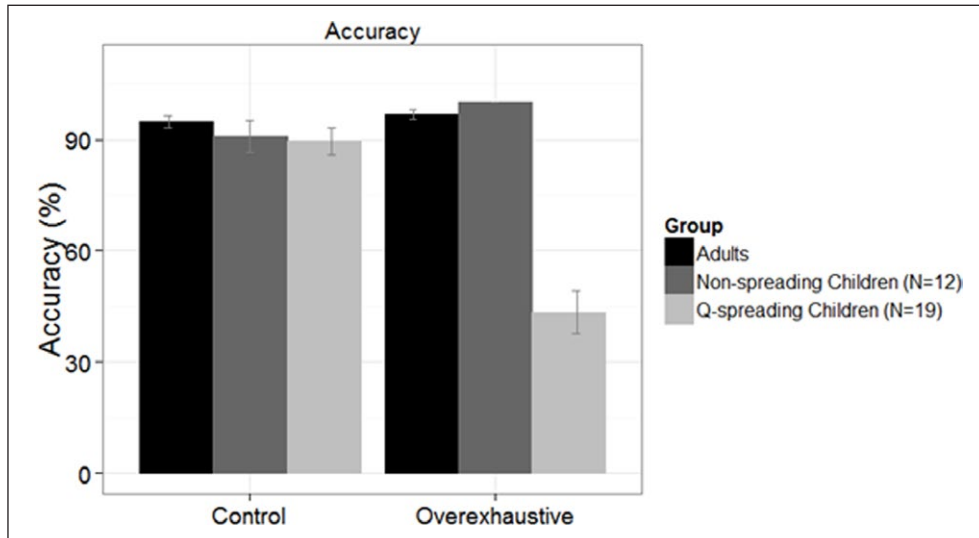
Figure 2: Mean accuracy (%) for the control adults (dark bars), all children taken as a whole (*N* = 31; dark gray bars), and *q*-spreading children (*N* = 19, light gray bars). Error bars represent 95% confidence intervals.

Table 1: Accuracy (mixed-effect models statistical analysis). Left panel: all children (*N* = 31) and children who made *q*-spreading errors (*N* = 19); right panel: comparison of adults and children.

Accuracy	All children (*N* = 31)				Adults vs. all children (*N* = 31)			
	b	*SE*	*z*-score	*p*-value	*b*	*SE*	*z*-score	*p*-value
(Intercept)	1.72	0.29	5.964	<.001	2.69	0.25	10.671	<.001
Type of Picture	**1.88**	**0.39**	**4.816**	**<.001**	0.68	0.38	1.770	.0768
Group					−1.90	0.45	−4.202	<.001
Group × Type of Picture					2.42	0.77	3.159	<.01
Accuracy	*q*-spreading children (*N* = 19)				Adults vs. *q*-spreading children (*N* = 19)			
	b	*SE*	*z*-score	*p*-value	*b*	*SE*	*z*-score	*p*-value
(Intercept)	0.96	0.24	4.017	<.001	2.09	0.20	10.704	<.001
Type of Picture	**2.47**	**0.45**	**5.533**	**<.001**	0.96	0.37	2.590	<.01
Group					−2.26	0.38	−5.931	<.001
Group × Type of Picture					2.98	0.74	4.017	<.001

The right panel in Table 1 shows two statistical comparisons for the accuracy data; the first compares the results between adults and all children (all, *N* = 31) and the second, between adults and the *q*-spreading children (*N* = 19). We found an expected main effect of Group, with children making significantly more errors than adults, and a significant interaction for Group × Type of Picture. Both children taken as a whole (*N* = 31) and *q*-spreading children (*N* = 19) taken separately were significantly less accurate in the Overexhaustive than Control condition (65% vs. 90% and 43% vs. 89%, respectively) compared to adults (97% vs. 95%); these interactions were driven by the error rate in the Overexhaustive condition that were made by the *q*-spreading children.

3.2.2 Eye-movement data

Fine-grain analyses: we investigated the time course of looks to the Extras/Distractors by focusing on each of the four consecutive ROIs in the sentences represented in (3). For the adults, the results are averaged across the entire group because they were at ceiling for

accuracy. The children's eye movement data are presented separately for non-spreaders and *q*-spreading children. Figure 3 represents the averaged proportions of looks to the Distractors in the control condition for each ROI. We are only reporting the model comparing the eye movements between children and adults because the adult results by themselves are not informative with respect to *q*-spreading.

Figure 4 represents the averaged proportions of looks to the Extras in the experimental (overexhaustive) condition for each ROI. Statistical analyses were calculated on the arcsin-transformed proportions of looks Extras/Distractors because they fitted the data better than models for the non-transformed data. There were few significant effects for adults and the children taken as a whole.

All children (N = 31): the fixations which took place in ROI 1, 'every alligator', are not informative because they happened too early in the sentence to influence processing so main effects and interactions in this region will not be discussed. In ROI 2 (Verb) and ROI 3 (PP), there were no main effects of either Type of Picture (ROI 2: $b = 0.06$, $SE = 0.12$, $t < 1$, $p = .6487$; ROI 3: $b = 0.00$, $SE = 0.09$, $t < 1$, $p = .9755$; Figure 3 vs. Figure 4), Accuracy (ROI 2: $b = 0.08$, $SE = 0.06$, $t = 1.319$, $p = .1887$; ROI 3: $b = 0.05$, $SE = 0.05$, $t = 1.075$, $p = .2834$; dark gray vs. light gray lines), or interaction between Type of Picture and Accuracy (ROI 2: $b = -0.05$, $SE = 0.12$, $t < 1$, $p = .6687$; ROI3: $b = 0.07$, $SE = 0.08$, $t < 1$, $p = .4304$). The proportions of looks to the Extras (Overexhaustive condition) and the Distractors (Control condition) were statistically the same in the incorrect and correct trials (40% vs. 32%). In ROI 4 (silence), there was a main effect of the Type of Picture ($b = 0.14$, $SE = 0.06$, $t = 2.284$, $p < .05$), as there were more looks to the Distractors (the two elephants) than the Extras (the two empty containers).
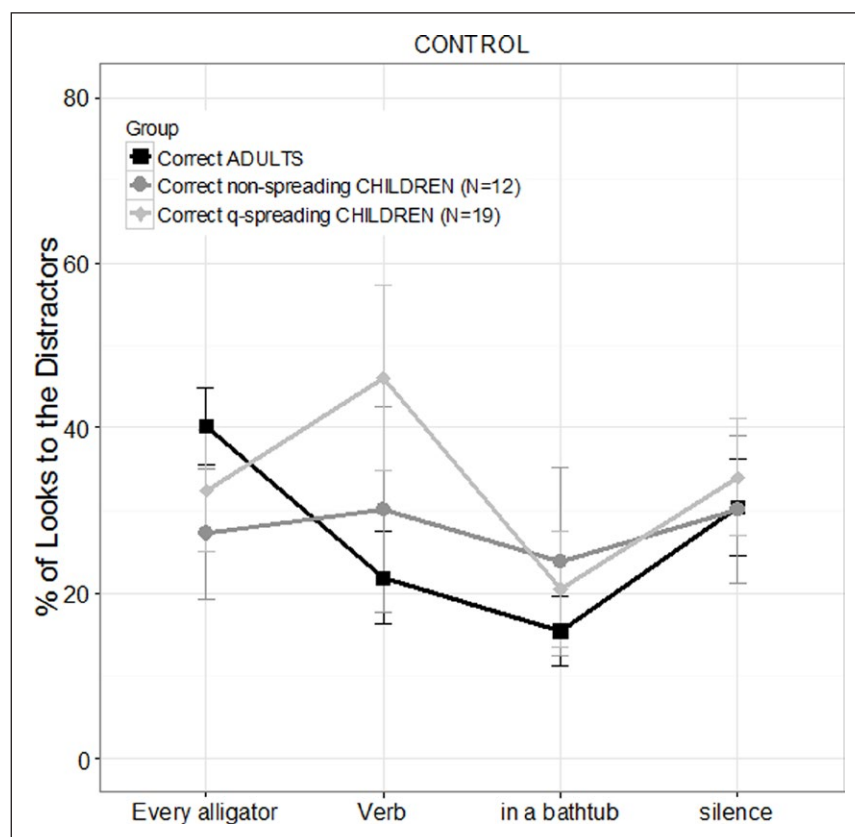

Figure 3: Fine-grain eye movements: Correct trials. Proportions of looks to the Extras/Distractors for control adults (black lines), all children taken as a whole (*N* = 31; dark gray lines), and *q*-spreading children (*N* = 19, solid light gray lines) in four Regions of Interest.

Note, however, that in analyzing the looks to the Extras/Distractors for the children as a whole ($N = 31$) as described above, we actually are comparing almost the entire dataset (90% of the trials with correct interpretation) with a smaller subset of data (57% of the trials interpreted incorrectly) contributed by the 19 *q*-spreading children. The data from the adult-like 12 children who never make *q*-spreading errors may be "washing out" the eye-movement patterns characteristic of *q*-spreading. However, the small number of adult-like children does not provide a subset of control data robust enough for comparison; to avoid this problem, we compared the eye-movement patterns of 19 *q*-spreading children for the correct and incorrect trials with those of adults (Figure 4).

Quantifier-spreading children (N = 19): Table 2 gives the results for the statistical models that were calculated for correct and incorrect trials, with Accuracy included as a factor. The comparison between children and adults (right panel) is based only on the trials answered correctly because adults made very few errors. This analysis revealed a different pattern of eye movements than the one described above that took children as a whole. Here, in ROI 2 (Verb), three main effects emerged. First, there was a significant main effect of Accuracy: when the children interpreted the Overexhaustive trials incorrectly they looked more at the Extras than when they made no *q*-spreading errors (48% vs. 28%, respectively; Figure 4). (This effect also extended into the next ROI 3, the locative PP). Second, there were significant main effects of Group and Type of Picture, as even in the correctly answered trials, the *q*-spreading children still looked more to the Extras than the adults. No other effects or interactions reached significance in ROI 2–3.
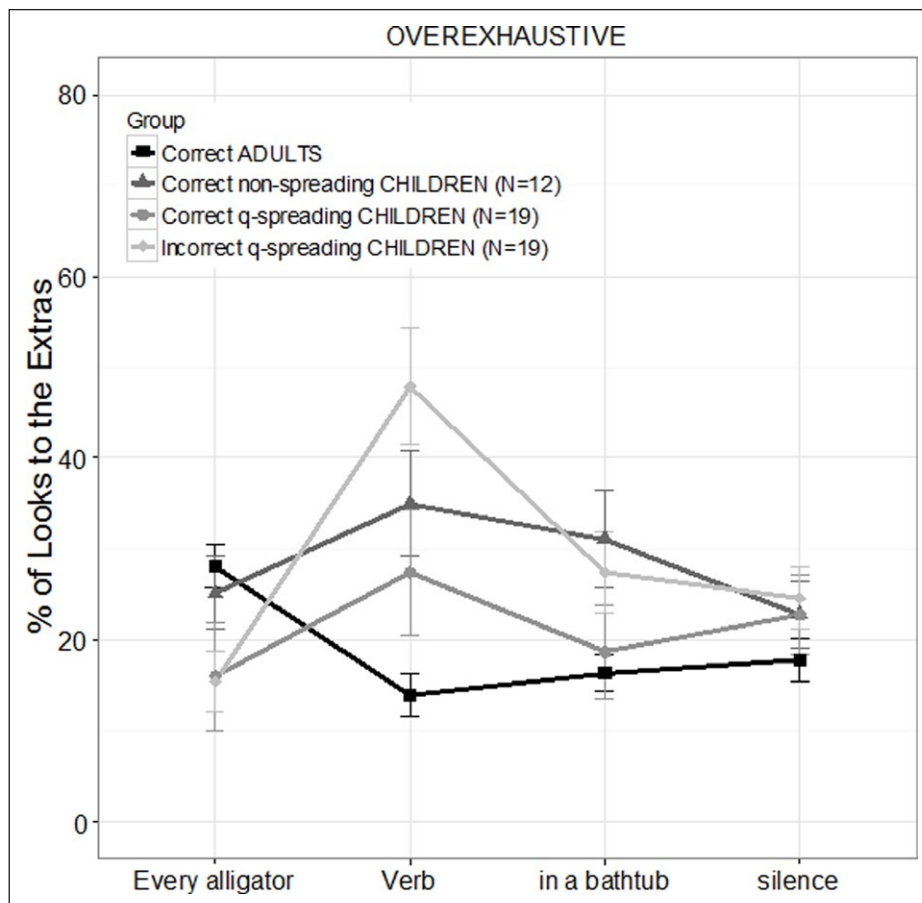

Figure 4: Fine-grain eye movements: All trials. Proportions of looks to the Extras/Distractors for control adults (black lines), non-spreading children (*N* = 12; dark gray lines), and *q*-spreading children (*N* = 19) , correct trials (gray line) versus incorrect trials (light gray line) in four Regions of Interest.

Table 2: Linear mixed-effect models on the arcsin-transformed mean proportions of looks to the Extra/Distractors in the four regions of interest. Left panel: *q*-spreading children (*N* = 19, correct and incorrect trials); right panel: comparison of adults and children.

	q-spreading children (*N* = 19)				Adults vs. *q*-spreading children (*N* = 19)			
	b	*SE*	*t*-score	*p*-value	*b*	*SE*	*t*-score	*p*-value
1. Every alligator								
(Intercept)	0.38	0.07	5.077	<.001	0.50	0.05	10.313	<.001
Type of Picture	**0.19**	**0.08**	**2.276**	**<.05**	**0.23**	**0.05**	**4.872**	**<.001**
Accuracy	−0.03	0.04	−0.598	.5510				
Type of Pic × Acc	−0.08	0.09	−0.851	.3961				
Group					−0.12	0.07	−1.753	.0803
Type of Pic × Group					0.10	0.09	1.028	.3045
2. Verb								
(Intercept)	0.71	0.09	7.574	<.001	0.43	0.04	10.591	<.001
Type of Picture	0.26	0.15	1.759	.0807	**0.20**	**0.06**	**3.187**	**<.01**
Accuracy	**0.15**	**0.07**	**2.025**	**<.05**				
Type of Pic × Acc	−0.00	0.15	−0.017	.9866				
Group					**0.29**	**0.07**	**3.927**	**<.001**
Type of Pic × Group					0.14	0.13	1.101	.2716
3. In a bathtub								
(Intercept)	0.44	0.05	8.887	<.001	0.31	0.03	9.602	<.001
Type of Picture	0.12	0.10	1.195	.2341	0.00	0.05	0.089	.9295
Accuracy	**0.10**	**0.05**	**2.089**	**<.05**				
Type of Pic × Acc	0.10	0.10	0.974	.3318				
Group					0.08	0.05	1.442	.1500
Type of Pic × Group					0.03	0.10	0.277	.7820
4. Silence								
(Intercept)	0.46	0.05	8.842	<.001	0.44	0.04	10.424	<.001
Type of Picture	0.08	0.08	0.922	.3578	**0.16**	**0.06**	**2.891**	**<.01**
Accuracy	−0.02	0.04	−0.473	.6371				
Type of Pic × Acc	−0.05	0.09	−0.633	.5275				
Group					0.09	0.08	1.174	.2411
Type of Pic × Group					−0.09	0.11	−0.850	.3956

We speculate that the pattern of children's eye movements in the incorrectly answered overexhaustive trials is reflective of *q*-spreading: the more children looked at the extra empty containers (i.e., two bathtubs) immediately after the quantified NP *every alligator* was explicitly mentioned in the sentence (i.e., during the verb), the more likely they were to make an error in the interpretation at the end of the sentence. This effect extended into ROI 3 (PP), but disappeared in ROI 4 (silence) indicating that once the children made an early commitment to pursue an incorrect interpretation they were unwilling to revise it.

4 Discussion

In the Visual World eye-tracking experiment, we recorded eye movements of the Russian-speaking adults and 5-to-6-year-old children as they performed the standard sentence-picture verification task with spoken sentences containing the quantifier *kazhdyj* 'every'.

The task was conceptually identical to the one used in previous behavioral experiments on acquisition of quantified sentences cross-linguistically. As expected, we replicated the well-established finding: the children taken as a whole made many *q*-spreading errors in the experimental overexhaustive condition in which the objects and the containers were in a partial one-to-one correspondence. However, not all of the children engaged in *q*-spreading errors; 12 of out 31 were non-spreaders and demonstrated adult-like knowledge of the quantifier *kazhdyj* 'every'. The majority nonetheless were *q*-spreading children (*N* = 19) who exhibited only 43% accuracy.

The novelty of our study was an examination of moment-by-moment processing of the quantified sentences as they unfolded in speech and how it was reflected in the participants' eye movements. We focused on two comparisons: (1) the control adults, non-spreaders, and *q*-spreading children, and (2) the correctly versus incorrectly interpreted sentences (*q*-spreading children only). First, we found that non-spreaders and *q*-spreading children looked more than adults at *both* distractor objects (control condition) and extra containers (experimental condition). This general difference between the adults and children as a whole is not surprising as it is a well-established fact that children tend to look around more regardless of experimental manipulations (Nation et al. 2003). Comparison of eye-movement patterns of the adults with the *q*-spreading children only revealed that the proportions of looks differed between these two groups: the *q*-spreading children, but not the adults, looked significantly more at the extra containers than the distractor objects. Moreover, the looks to the extra containers were modulated by accuracy: more looks led to *q*-spreading errors in the sentence-picture verification task at the end.

The increase in children's looks to the extra containers started at the verb allowing us to pinpoint where *q*-spreading happens during processing of the spoken quantified sentence. As children hear *Kazhdyj alligator lezhit …* 'Every alligator is lying …' in the experimental condition, they begin anticipating the mention of containers and look more at the extra bathtubs because the verb semantics in combination with the picture presupposes that the alligators are placed into (or paired with) the bathtubs. Such anticipation is quickly disconfirmed in the control condition where the visual context contains distractor objects; after all, elephants are not mentioned in the sentence. Thus, it is not the case that simply looking more at the extra containers leads to *q*-spreading errors. Visual context with extra containers leads to *q*-spreading errors *only* because it is combined with the spoken sentence that mentions these extra containers in the scope of the quantifier *every* (e.g., *every alligator* precedes *a bathtub*), i.e., in the experimental (overexhaustive) condition.

We argue that it is the *combination* of multiple sources of information, i.e., visual context with extra containers and the spoken sentence with the quantifier *every* in which they are explicitly mentioned, that causes *q*-spreading errors. It also makes an additional specific prediction. If the control condition is altered in such a way that the visual context contains distractor containers (e.g., beds) instead of distractor objects (e.g., elephants), children should demonstrate anticipatory looks to the distractor containers on a par with the looks to the extra containers. Unfortunately, the current study did not have such a control condition; however, we also conducted a similar study with English-speaking 5-to-12 – year-old children reported in the chapter submitted to the edited book on quantification (Sekerina et al. forthcoming) that had such a control condition. In it, the spoken sentence *Every alligator is in a bathtub* was paired with the picture of three alligators in the bathtubs and two empty beds. In the region of interest that combined the verb (e.g., *is*), the location (the PP *in a bathtub*), and silence, the average proportion of looks to the distractor containers was 26.42% and to the extra containers, 31.28%. Similar to the Russian study, the increased looks to the extra containers also correlated with *q*-spreading errors in the English-speaking children.

Why should the combination of the visual context with the extra containers and the spoken quantified sentence lead to *q*-spreading errors in children's processing? We argue that the extra containers that are explicitly mentioned in the sentence in the scope of the quantifier distract children and create interference in processing and thus increase cognitive load on the still developing processing system of children. Similar interference in the form of retrieval interference has been proposed as an explanation of difficulties in adult processing of anaphor resolution and *wh*-dependencies and agreement attraction errors (see Felser, Phillips & Wagers 2017 for an editorial overview of the research topic *Encoding and Navigating Linguistic Representations in Memory*). In our case, the interference for children is caused by visual rather than linguistic referents. We leave additional empirical testing of the role of the interference theory (Van Dyke & McElree 2006) in development of processing in children to the future research.

Our results contribute to cross-linguistic debate concerning acquisition of quantifiers and causes of errors children make in interpreting them. First, the sentence-picture verification task employed in the previous studies including Minai et al.'s (2012) and ours critically depends on simultaneous combination of two sources of information, i.e., linguistic (the spoken sentence that contains the quantified NP) and visual context creating a situation in which the two sources of information can compete with each other. While we corroborate Minai et al.'s (2012) findings that allocation of visual attention as reflected in eye movements to the extra entities is a critical factor that is responsible for *q*-spreading errors in young monolingual children, we attribute children's difficulty not to undeveloped attention-switching mechanism, but to interference from the extra entities that results in cognitive overload.

For some children more than others (19 out of 31 in our case) the interference from the extra containers in the visual context is too strong to resist: they started to anticipate the containers as soon as they heard *every alligator*, and we managed to capture them in the process of *q*-spreading as it occurred. Moreover, children's eye-movement patterns showed no signs of recovery from *q*-spreading demonstrating their deterministic parsing and inability to revise their initial incorrect commitment. This conclusion is supported by the findings from the previous developmental eye-tracking studies on syntactic ambiguity resolution in sentences with PP-attachment ambiguity (Trueswell et al. 1999) and on referential ambiguity resolution with short-distance pronouns (Sekerina, Stromswold & Hestvik 2004). What allows some children to move to the adult-like mastery in combining multiple sources of information in processing of quantified sentences earlier than others awaits future research as a part of a broader question of how different cognitive systems, notably, memory and executive function interact with sentence processing in development.

Thus, recording of eye movements provided us with an empirical justification of the hypothesized act of looking for a one-to-one correspondence between the two entities that leads to *q*-spreading errors. This is novel experimental evidence in support of the full competence hypothesis in the acquisition of quantification: children's knowledge of universal quantifiers is completely adult-like, and their *q*-spreading errors are of processing nature. They are caused by cognitive overload that, as we have demonstrated, can stem from not only pragmatic infelicity and task demands of the design (Meroni et al. 2007; O'Grady et al. 2010), but also from interference when multiple sources of information compete with each other in real-time. The insight into the nature of the *q*-spreading phenomenon in development of sentence processing that was made possible by our use of the Visual World Paradigm puts the eye-tracking methodology at the forefront of language acquisition studies (see also Bergmann, Paulus & Fikkert 2012). It reveals a tight connection between allocation of visual attention and the spoken input that is essential in

pinpointing the moment in real-time comprehension when children stumble. When faced with the necessity to combine different sources of information as is the case with the sentence-picture verification task, children tend to rely more on visual information; and if it is distracting, children find it hard to resist.

Abbreviations

NOM = nominative, LME = linear mixed-effects models, MASC = masculine, PRES = present, PREP = prepositional, SG = singular, Q = quantifier

Appendix

The full set of materials, pictures and spoken sentences is available at (http://www.iris-database.org/iris/app/home/search?query = sekerina).

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Baayen, Harald R. 2008. *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511801686

Bates, Douglas M., Martin Maechler & Ben Dai. 2009. *lme4: Linear mixed-effect models using S4 classes*. R package version 0.999375-31. [Software]. Vienna, Austria: R Foundation for Statistical Computing.

Bergmann, Cristina, Markus Paulus & Paula Fikkert. 2012. Preschoolers' comprehension of pronouns and reflexives: The impact of the task. *Journal of Child Language* 39(4). 777–803. DOI: https://doi.org/10.1017/S0305000911000298

Brooks, Patricia J. & Martin D.S. Braine. 1996. What do children know about the universal quantifiers *all* and *each*? *Cognition* 60. 236–268. DOI: https://doi.org/10.1016/0010-0277(96)00712-3

Crain, Stephen, Rosalind Thornton, Carole Boster, Laura Conway, Diana Lillo-Martin, Elaine & Woodams. 1996. Quantification without qualification. *Language Acquisition* 5(2). 83–153. DOI: https://doi.org/10.1207/s15327817la0502_2

Drozd, Ken 2001. Children's weak interpretations of universally quantified questions. In Melissa Bowerman & Stephen C. Levinson (eds.), *Language acquisition and conceptual development*, 340–376. Cambridge, MA: Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511620669.014

Drozd, Ken & Erik van Loosbroek. 1998. Weak quantification, plausible dissent, and the development of children's pragmatic competence. In Annabel Greenhill, Heather

Littlefield & Cheryl Tano (eds.), *Proceedings of the 23rd Boston University Conference on Language Development*, 184–195. Somerville, MA: Cascadilla Press.

Felser, Claudia, Colin Phillips & Matthew Wagers. 2017. Editorial: Encoding and navigating linguistic representations in memory. *Frontiers in Psychology* 8. 164. DOI: https://doi.org/10.3389/fpsyg.2017.00164

Forster, Kenneth I. & Jonothan C. Forster. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35(1). 116–124. DOI: https://doi.org/10.3758/BF03195503

Geurts, Bart. 2003. Quantifying kids. *Language Acquisition* 11(4). 197–218. DOI: https://doi.org/10.1207/s15327817la1104_1

Inhelder, Barbel & Jean Piaget. 1964. *The early growth of logic in the child.* London: Routledge & Kegan Paul.

Katsos, Napoleon et al. 2012. The acquisition of quantification across languages: Some predictions. In Alia K. Biller, Esther Y. Chung & Amelia E. Kimball (eds.), *BUCLD 36: Proceedings of the 36th Annual Boston University Conference on Language Development*, 258–268. Somerville, MA: Cascadilla Press.

Kuznetsova, Julia, Maria Babyonyshev, Jodi Reich, Lesley Hart & Elena Grigorenko. 2007. The acquisition of universal quantifiers in Russian. In Alyona Belikova, Luisa Meroni & Mari Umeda (eds.), *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, 224–232. Somerville, MA: Cascadilla Press.

Meroni, Luisa, Andrea Gualmini & Stephen Crain. 2007. The strength of the universal quantifier in child language. In Alyona Belikova, Luisa Meroni & Mari Umeda (eds.), *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, 277–284. Somerville, MA: Cascadilla Press.

Minai, Utako, Nobuyuki Jincho, Naoto Yamane & Reiko Mazuka. 2012. What hinders child semantic computation: Children's universal quantification and the development of cognitive control. *Journal of Child Language* 39. 919–956. DOI: https://doi.org/10.1017/S0305000911000316

Nation, Kate, Catherine M. Marshall & Gerry T. M. Altmann. 2003. Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology* 86. 314–329. DOI: https://doi.org/10.1016/j.jecp.2003.09.001

O'Grady, W., Suguzi, T. & Yoshinaga, N. 2010. Quantifier spreading: Evidence from Japanese. *Language Learning and Development* 6(2). 116–125. DOI: https://doi.org/10.1080/15475440903352799

Philip, William. 1995. *Event quantification in the acquisition of universal quantification.* Amherst, MA: University of Massachusetts dissertation.

Poeppel, David & Kenneth Wexler. 1993. The full competence hypothesis of clause structure in early German. *Language* 69(1). 1–33. DOI: https://doi.org/10.2307/416414

Rakhlin, Natalia. 2007. A new pragmatic account of quantifier-spreading. *Nanzan Linguistics* 8. 239–282.

Roeper, Thomas. 2007. *The prism of grammar*. Cambridge, MA: MIT Press.

Roeper, Thomas, Uri Strauss & Barbara Z. Pearson. 2004. The acquisition path of quantifiers: Two kinds of spreading. Ms., University of Massachusetts at Amherst. (http://people.umass.edu/roeper/online_papers/pathof%20acquisition3-15c.pdf) (Accessed 2011-1-25).

Sekerina, Irina A. & Antje Sauermann. 2015. Visual attention and quantifier-spreading in heritage Russian bilinguals. *Second Language Research* 31(1). 75–104. DOI: https://doi.org/10.1177/0267658314537292

Sekerina, Irina A., Patricia J. Brooks, Luca Campanelli & Anna Schwartz. Forthcoming. Quantifier-spreading in school-age children: An eye-tracking study. In Katalin È. Kiss (ed.), *Linguistic and cognitive aspects of quantification.* Springer.

Trueswell, John C. & Michael K. Tanenhaus (eds.). 2004. *Approaches to studying word-situated language use.* Cambridge, MA: MIT Press.

van Dyke, Julie A. & Brian McElree. 2006. Retrieval interference in sentence comprehension. *Journal of Memory and Language* 55. 157–166. DOI: https://doi.org/10.1016/j.jml.2006.03.007