Irina A. Sekerina | Antje Sauermann

# Visual attention and quantifier-spreading in heritage Russian bilinguals

Suggested citation referring to the original publication:
Second Language Research 31 (2015) 1, pp. 75–104
DOI http://dx.doi.org/10.1177/0267658314537292
ISSN (print) 0267-6583
ISSN (online) 1477-0326

*Article*

second language research

Second Language Research
2015, Vol. 31(1) 75–104

DOI: 10.1177/0267658314537292
slr.sagepub.com
SAGE

# Visual attention and quantifier-spreading in heritage Russian bilinguals

**Irina A Sekerina**
City University of New York, USA

**Antje Sauermann**
University of Potsdam, Germany

## Abstract

It is well established in language acquisition research that monolingual children and adult second language learners misinterpret sentences with the universal quantifier *every* and make quantifier-spreading errors that are attributed to a preference for a match in number between two sets of objects. The present Visual World eye-tracking study tested bilingual heritage Russian–English adults and investigated how they interpret of sentences like *Every alligator lies in a bathtub* in both languages. Participants performed a sentence–picture verification task while their eye movements were recorded. Pictures showed three pairs of alligators in bathtubs and two extra objects: elephants (Control condition), bathtubs (Overexhaustive condition), or alligators (Underexhaustive condition). Monolingual adults performed at ceiling in all conditions. Heritage language (HL) adults made 20% *q*-spreading errors, but only in the Overexhaustive condition, and when they made an error they spent more time looking at the two extra bathtubs during the Verb region. We attribute *q*-spreading in HL speakers to cognitive overload caused by the necessity to integrate conflicting sources of information, i.e. the spoken sentences in their weaker, heritage, language and attention-demanding visual context, that differed with respect to referential salience.

## Keywords

eye-tracking, heritage language, quantifier-spreading, Russian, universal quantifiers, visual attention

**Corresponding author:**
Irina Sekerina, Department of Psychology, College of Staten Island, 2800 Victory Boulevard, 4S-108, Staten Island, NY 10314, USA.
Email: irina.sekerina@csi.cuny.edu

# 1 Introduction

Since the seminal work by Inhelder and Piaget (1964), it is well established that 5-to-12-year-old monolingual children misinterpret sentences with the universal quantifier *every* (Brooks and Braine, 1996; Brooks and Sekerina, 2005/2006; Crain et al., 1996; Drozd, 2001; Freeman and Schreiner, 1988; Geurts, 2003; Minai et al., 2012; O'Grady et al., 2010; Philip, 1995). The errors children make in sentences with the universal quantifier *every* come in three flavors – overexhaustive, underexhaustive, and the so-called 'bunny' errors – collectively referred to as 'quantifier spreading' (*q*-spreading) because the domain of the quantifier seems to have spread beyond the noun it modifies. In a typical experiment of this type, children see a single picture with two sets of objects, for example, alligators and bathtubs (Figure 1), in which one of the objects is extra.

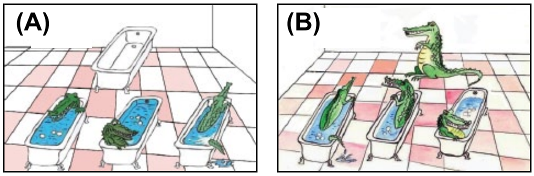
What is manipulated in such experiments is the mismatch in the number of the objects, i.e. four bathtubs but only three alligators (Figure 1A) or four alligators but only three bathtubs (Figure 1B). Simultaneously with the picture, children are asked a Yes/No question presented in isolation that contains the universal quantifier *every*:

(1) Is every alligator in a bathtub?

Logically, a universally-quantified sentence in (1) is true when the picture contains one or more extra bathtubs that are empty (Figure 1A). However, many children reject the pairing of (1) with Figure 1A and say *No*. This error is referred to in the literature as an overexhaustive error, also known as exhaustive pairing (Drozd, 2001), Type-A (Geurts, 2003), and classic spreading (Roeper et al., 2004). The same sentence (1) is false when the picture contains extra alligators (Figure 1B). Thus, when children erroneously accept the pairing of (1) with Figure 1B and say *Yes*, they make a complementary, underexhaustive error; this is also known as underexhaustive pairing (Drozd, 2001) and Type-B (Geurts, 2003). The third type, 'bunny' spreading, occurs when the picture contains an additional pair of objects depicting an event (e.g. a dog with a bone) different from the main one (e.g. three bunnies with one carrot each). An overwhelming majority of children make overexhaustive errors well into middle school years (Roeper et al., 2004), and a few make underexhaustive errors. 'Bunny'-spreading is typically found with very young children and will not be discussed any further in this article.

Children's error-prone performance on sentences with the universal quantifier *every* is typically contrasted with that of control monolingual adults who are believed to be error free. They correctly accept the pairing of (1) with the picture in Figure 1A and reject it with Figure 1B because they rely on real world knowledge to establish the domain of quantification. Recently, however, the target-like comprehension by adults has been empirically challenged. First, Brooks and Sekerina (2005/2006) found that in a timed picture-selection task with the two pictures similar to the ones in Figure 1 and two foils, college students were only 75% correct. Then Street and Dąbrowska (2010) demonstrated that less educated monolingual adults who are employed as shelf-stackers, packers, assemblers, and clerical workers also were only 78% correct in the picture selection task identical to the one usually administered to children. Finally, Japanese adults tested by Minai and colleagues (2012) in the eye-tracking study with the pictures

**OVEREXHAUSTIVE** **UNDEREXHAUSTIVE**




*Is every alligator in a bathtub?*

| Correct answer: YES | Correct answer: NO |
|---|---|
| 'NO' is an overexhaustive *q*-spreading error | 'YES' is an underexhaustive *q*-spreading error |

**Figure 1.** An example of a typical experimental trial in first language acquisition studies.

of Figure 1A type achieved an accuracy of 59%. The explanations proposed to account for *q*-spreading errors in adults include shallow processing (Brooks and Sekerina, 2005/2006), qualitatively poorer linguistic experience with less education (Street and Dąbrowska, 2010), and impact of salience of the single extra object in the picture selection task (Minai et al., 2012).

Adult second language (L2) learners also make *q*-spreading errors. DelliCarpini (2003) tested 60 adult L2 learners of English of various proficiency levels on the pairing similar to (1) – Figure 1A – and found 70% *q*-spreading errors in the lower proficiency group, but this error rate went down to 35% in the higher proficiency group. Berent et al. (2009) also employed two different proficiency groups, but their participants had to view multiple pictures (including Figures 1A and 1B) instead of the single one and answer a Yes/No question for each of them. Contrary to the findings by DelliCarpini, there was no difference in the *q*-spreading error rates between the two groups: 28% for the lower proficiency and 26.5% for the higher proficiency learners. They suggested that a deficiency in discourse pragmatic knowledge of referential restrictions on universal quantifiers was responsible for such errors. Sorace (2011) argues that interpreting such sentences involves an interface between linguistic and other cognitive domains that are less likely to be acquired completely in L2 acquisition and, therefore, are more subject to individual variation. This variation then can diverge from the differences based on group proficiency canceling the effect of proficiency.

What are the underlying mechanisms that cause non-target-like comprehension of the universal quantifiers in most monolingual children, some monolingual adults, and many adult L2 learners? Is a unified explanation across ages and diverse learner groups possible? The accounts for *q*-spreading errors proposed for adults are not the same as for children; after all, the former have appropriate linguistic knowledge of universal quantifiers of their native language and have fully developed executive control. For monolingual adults, Brooks and Sekerina (2005/2006) extended the Shallow Processing hypothesis of Ferreira et al. (2002) to processing of the quantified sentences while Minai and colleagues (2012) suggested that their participants were distracted by single extra objects (Figure 1A). Street and Dąbrowska (2010) argued that the *q*-spreading errors are attributed to qualitatively poor linguistic experience of the less educated monolingual adults in their study. Berent and colleagues (2009) hypothesized that development of pragmatic knowledge in both children and adult L2 learners may be affected by ‘other aspects of maturation including attention and memory’ (p. 279). Thus, all of them appeal to the individual differences approach popular in cognitive psychology (Bayliss et al., 2003) and psycholinguistics (Sekerina, 2012; Swets et al., 2007) and propose various cognitive and socioeconomic factors as the source of *q*-spreading errors in monolingual and adult L2 learners.

The goal of the current study is to evaluate comprehension of the universal quantifier *every* in another adult group, i.e. that of bilingual heritage speakers of Russian, whose knowledge of heritage language (HL) often shares characteristics with monolingual first-language (L1) and adult L2 acquisition, and explore whether they make *q*-spreading errors in Russian sentences with the universal quantifier *kazhdyj* ‘every’ similar to monolingual children and adult L2 learners. Bilingual HL adults can provide an important contribution to the theories of bilingualism and language acquisition: Does the knowledge of universal quantification in HL speakers look more like the one in monolingual adults? Or, are they like some sort of ‘fossilized’ monolingual children or adult L2 learners? Like monolingual children, HL speakers were exposed to their heritage language at home and in a naturalistic setting, and they often exhibit the same developmental errors observed in children (Benmamoun et al., 2014). On the other hand, HL input and use are severely reduced in heritage speakers after the onset of schooling, and their proficiency in the heritage language varies greatly, just like in L2 learners.

Participants in our study were bilingual heritage Russian–English adults who performed a classical sentence–picture verification task in both their weaker heritage language (Russian) and their dominant language (English). Spoken quantified sentences with the universal quantifier *every* were paired with pictures that included interfering visual information creating conditions conducive to overexhaustive and underexhaustive *q*-spreading. Their spoken comprehension was assessed online and compared to that of monolingual Russian-speaking control adults as their eye movements were recorded. We chose the sentence–picture verification task so that we could compare our results with the well-established pattern of *q*-spreading errors in cross-linguistic L1 acquisition (Philip, 1995), but adapted it from its offline version to be performed in conjunction with eye-tracking in the visual world paradigm (VWP; Trueswell and Tanenhaus, 2004). It allowed us to capture moment-by-moment time course of how HL Russian speakers interpreted the quantified sentences and to catch *q*-spreading errors ‘on the fly’, as they were being committed. Finally, we also compared performance of the same HL speakers in interpreting the quantified sentences in Russian to their performance on the identical sentences in English.

## 1 Quantifier-spreading in monolingual children: The role of visual attention

*Q*-spreading errors are pervasive and universal. They have been reported for a wide range of ages, i.e. from 5 to 9–12 years of age (Roeper et al., 2004), and in a wide variety of typologically diverse languages, such as Romance (Pouscoulous et al., 2007), Germanic (Hollebrandse, 2004), Asian (Kang, 2001; O'Grady et al., 2010), Kannada (Lidz and Musolino, 2005), and Russian (Kuznetsova et al., 2007). Different explanations have been entertained from cognitive and linguistic perspectives (see Rakhlin, 2007, for an overview), but there is no consensus for which one provides the best account for multiple factors that seem to affect children's *q*-spreading errors.

Two recent studies of *q*-spreading errors in Japanese-speaking children – O'Grady et al. (2010) and Minai et al. (2012) – explicitly appeal to the role of visual attention in the sentence–picture verification task. Following Rakhlin (2007), they argue that relative salience of extra objects in the pictures imposes perceptual restrictions on the domain of application of quantifiers that is too strong for children to ignore; thus, it drives *q*-spreading errors. When the task is changed in such a way as to encourage children to disengage from the additional extra object, either by switching to an act-out task (O'Grady et al., 2010) or by increasing the number of extra objects from one to three (Minai et al., 2012), children's *q*-spreading errors decrease significantly. Moreover, while O'Grady and colleagues could only speculate on the critical role of visual attention, Minai et al. directly observed its effects by recording the eye movements of 4-to-5-year-old Japanese children (and control adults) who first viewed a block of multiple extra object pictures (three turtles each holding an umbrella and three extra umbrellas), followed by a block of single extra object pictures (three turtles each holding an umbrella and one extra umbrella). Children's tendency to commit *q*-spreading errors was significantly lower (48%) in the latter block in comparison to the control group of children (86%) for whom the order of the two blocks was reversed.

The eye movements of the children (Minai et al., 2012) revealed that they fixated on the extra objects (one or three umbrellas) more than the adults, and these fixations peaked at 1.5 s before the sentence began, i.e. during the silent picture preview phase. Critically, though, the latency, magnitude, and duration of these fixations were modulated by two factors: first, whether the child was susceptible to *q*-spreading errors or not, and, second, his or her score on the dimensional change card sort (DCCS) task. The child's score on the DCCS task was operationalized to approximate attentional abilities in switching perspective. There was a correlation between the amount of *q*-spreading errors, the DCCS score, and eye movements in that the children who scored low on 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 switching from the visually salient extra objects, especially when there is only one, is one of the critical factors responsible for *q*-spreading errors in young monolingual children.

## 2 *Quantifier-spreading in adult L2 learners and bilingual HL speakers*

There are no *a priori* reasons to think that when adult L2 learners make *q*-spreading errors (Berent et al., 2009; DelliCarpini, 2003) they do so because of deficiencies in controlling visual attention hypothesized for monolingual children by Minai et al. (2012). After all, it is assumed that adults possess fully developed cognitive control, including attentional ability to switch perspectives. An alternative explanation proposed by Berent and colleagues focuses on non-target-like knowledge of pragmatic restrictions on universal quantification in adult L2 learners of English when they perform the sentence–picture verification task.

Do bilingual HL adults exhibit *q*-spreading errors and who do they resemble the most: monolingual children or adult L2 learners? On the one hand, if bilingual HL adults could be thought of as 'fossilized' monolingual children whose HL knowledge was arrested at the time they switched to the dominant language, as Polinsky (2011) reasons, they should be prone to *q*-spreading errors in their heritage Russian, just like monolingual Russian children (Kuznetsova et al., 2007). However, the types of *q*-spreading errors and error rates are an open question. On the other hand, the hallmark of HL speakers is significant variation in proficiency (Montrul, 2008), i.e. when they are closer to monolingual adults in some characteristics but to adult L2 learners in others. HL speakers are usually closer to monolingual controls in phonology, spoken language and naturalistic tasks, but resemble adult L2 learners in morphology and pragmatics, written language, and metalinguistic tasks (Laleko and Polinsky, 2013). If target-like comprehension of quantified sentences by bilingual HL adults depends on their proficiency in HL as suggested by DelliCarpini (2003) for L2 adults, we would expect the rate of *q*-spreading errors to vary as a function of age of arrival. Alternatively, it is possible that the interface hypothesis (Sorace, 2011) that ascribes difficulty with quantified sentences to the fact that their interpretation requires coordination of information from different domains such as syntax, semantics, and discourse may provide a better explanation.

To the best of our knowledge, there are no published studies of comprehension of quantified sentences in HL speakers, and the purpose of this study is to fill this gap. We chose to work with bilingual Russian–English young adults whose heritage language is Russian because they constitute a large HL community in the USA, and there are two recent L1 acquisition studies (Kuznetsova et al., 2007; Katsos, 2010) that have established that monolingual Russian children exhibit the usual pattern of *q*-spreading errors.

To provide direct comparison of our results to those from existing L1 studies with children and adult L2 learners, we measured the participants' accuracy in the sentence–picture verification task, but in order to contribute to the discussion of the unified cause of *q*-spreading errors across ages and diverse populations, we also collected reaction time and recorded eye-movement data using the VWP similar to the Minai et al.'s (2012) study. Recording eye movement was critical to establish how participants' visual attention was allocated in the visual context with extra objects. If bilingual HL speakers were to engage in *q*-spreading while processing sentences in their heritage language, eye movements would allow us to capture it as it happens, that is, at the moment when real-time processing of the quantified sentence starts to deviate from that of control group of monolingual adults. Crucially, we also conducted an identical experiment with the same participants in English, their dominant language, in which we did not expect them to make *q*-spreading errors.

## II Russian experiment

### 1 Method

*a Monolingual Russian participants.* Traditional college-age undergraduate students ($N$ = 40, 10 men; mean age 21.5) from the Department of Psychology of St. Petersburg State

University, Russia, volunteered to participate in the experiment in exchange for $3 (equivalent in rubles). All were native speakers of Russian, with beginner's knowledge of one foreign language that they were learning at the university as a part of their undergraduate curriculum.

*b Bilingual HL Russian–English participants.* A similar group of traditional college-age undergraduate students (*N* = 28, 9 men; mean age 19.4, range 18–26) from the College of Staten Island (CSI), USA, served as the experimental group. They self-identified as HL Russian–English bilinguals when signing up for the experiment. Prior to participating, each participant filled out a bilingual background questionnaire in English adapted from Fernández (2003).

All of the participants were born to Russian-speaking parents so their exposure to Russian was from birth and has continued throughout their lives. Four were born in the USA; the mean age of arrival for the others was 5.9 years (range 0–12), which was taken as the equivalent of the first contact with English (see Table 1). In addition, Table 1 displays the participants' self-reported language dominance. When asked to rate their proficiency in both languages on a five-point scale ranging from 1 = 'very good' to 5 = 'very poor', the participants' self-reported ratings as a group indicated higher proficiency for English (minus indicates higher ratings for English) in all four areas: oral comprehension ($M = -0.89$), speaking ($M = -1.14$), reading ($M = -2.25$), and writing ($M = -2.64$). Thus, the participants were typical HL Russian speakers, with English being their dominant language especially in literacy skills. Comprehension and speaking in Russian, nevertheless, were strong so that this allowed us to classify our participants as highly proficient in spoken HL Russian. This is important for our experiment as it tested processing of the quantified sentences during spoken language comprehension.

*c Design and materials.* The experiment had 12 experimental, 58 fillers, and three practice items (see Appendix 1 for the complete set of materials). Each item consisted of a picture (see Figure 2) paired with the spoken sentence (2) that contained the quantifier *kazhdyj* 'every' in the subject position (MASC = masculine, FEM = feminine, NOM = Nominative case, PREP = Prepositional case, SG = singular, PRES = present tense).

| | | | | |
|---|---|---|---|---|
| (2) | Kazhdyj | alligator | lezhit | v vanne. |
| | $\text{Every}_{\text{MASC-NOM-SG}}$ | $\text{alligator}_{\text{MASC-NOM-SG}}$ | $\text{lies}_{\text{PRES-SG}}$ | in $\text{bathtub}_{\text{FEM-PREP-SG}}$ |
| | 'Every alligator lies in a/the bathtub.' | | | |

All of the pictures were color drawings of entities, animals (e.g. alligators, dogs, birds, etc.) or inanimate objects (e.g. bananas, candles, spoons, etc.), and containers (e.g. bathtubs, baskets, nests, etc.). There were three types of pictures representing the Type of Picture (visual context) factor, a control and two experimental ones, the overexhaustive and underexhaustive *q*-spreading environment. In all three picture types, an animal (or inanimate object) was drawn in a 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. (For the ease of exposition, we will collectively refer to animals and inanimate objects as 'objects'.) The experimental pictures differed from the control one by what was depicted

**Table 1.** Difference scores in self-rated proficiency for 30 heritage language (HL) speakers in four areas, spoken comprehension, speaking, reading, and writing.

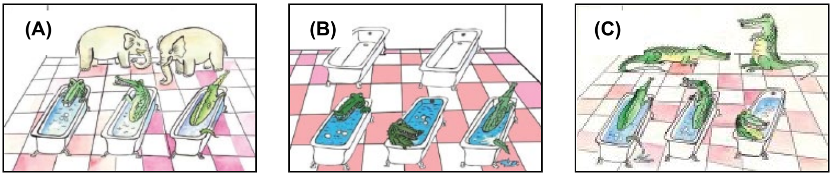
| Participant number | Arrival to the USA (years) | Self-reported difference in scores (English–Russian) | | | | |
|---|---|---|---|---|---|---|
| | | Comprehension | Speaking | Reading | Writing | Average |
| 10 | 5 | –3 | –3 | –4 | –4 | –3.50 |
| 38 | birth | –2 | –3 | –4 | –4 | –3.25 |
| 8 | birth | –2 | –2 | –3 | –4 | –2.75 |
| 18 | 6 | –2 | –1 | –4 | –4 | –2.75 |
| 14 | 5 | –1 | –2 | –4 | –4 | –2.75 |
| 28 | 6 | –1 | –1 | –4 | –4 | –2.50 |
| 35 | 4 | –1 | –1 | –4 | –4 | –2.50 |
| 16 | birth | –1 | –2 | –3 | –3 | –2.25 |
| 1 | 6 | –1 | –2 | –3 | –3 | –2.25 |
| 6 | 8 | –1 | –1 | –3 | –4 | –2.25 |
| 9 | 4 | –1 | –1 | –3 | –2 | –1.75 |
| 19 | 5 | –1 | –1 | –2 | –3 | –1.75 |
| 3 | 8 | –1 | –1 | –2 | –3 | –1.75 |
| 25 | 8 | –1 | –1 | –3 | –2 | –1.75 |
| 17 | 7 | –1 | 0 | –3 | –3 | –1.75 |
| 13 | 6 | 0 | 0 | –3 | –4 | –1.75 |
| 4 | 9 | –1 | –1 | –2 | –2 | –1.50 |
| 36 | 4 | –1 | –2 | –1 | –2 | –1.50 |
| 26 | 12 | –1 | –2 | 0 | –2 | –1.25 |
| 39 | birth | –1 | –2 | –1 | –1 | –1.25 |
| 20 | 4 | 0 | –1 | –2 | –2 | –1.25 |
| 27 | 7 | 0 | –1 | –2 | –2 | –1.25 |
| 33 | 10 | 0 | –1 | –1 | –3 | –1.25 |
| 5 | 9 | 0 | 0 | –2 | –3 | –1.25 |
| 23 | 12 | –1 | –1 | 1 | 0 | –0.25 |
| 30 | 6 | 0 | 0 | 0 | –1 | –0.25 |
| 2 | 10 | 0 | 0 | 0 | –1 | –0.25 |
| 24 | 4 | 0 | 0 | 0 | 0 | 0 |

*Notes.* 1 = very good, 5 = very poor; minus indicates higher proficiency in English, and plus indicates higher proficiency in Russian.

in the back. In the Overexhaustive experimental condition, there were two extra empty containers (e.g. two bathtubs in Figure 2B) by themselves, and in the Underexhaustive experimental condition, there were two extra animals or objects (e.g. two alligators in Figure 2C). In the Control condition, the two extra objects in the back were distractors, i.e. they were different from the ones in the front (e.g. two elephants in Figure 2A).

Sentence (2) was of the type that has been used in all of the previous studies of *q*-spreading errors in children, with the quantifier *every* appearing early, i.e. within the subject NP in the sentence-initial position. The verbs used were the present tense forms

| CONTROL | OVEREXHAUSTIVE | UNDEREXHAUSTIVE |
| --- | --- | --- |




*Kazhdyj alligator lezhit v vanne.* ‘Every alligator lies in a/the bathtub.’

Correct answer:

| YES | YES | NO |
| --- | --- | --- |

**Figure 2.** Three pictures corresponding to the three conditions in the present experiment. (A) Control, (B) Overexhaustive, (C) Underexhaustive.

of unaccusative verbs *sidit* ‘sits’, *lezhit* ‘lies’, *stoit* ‘stands’, *visit* ‘hangs’, and *spit* ‘sleeps’ repeated in several items. 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 speed.

The pairing of each of the three pictures (Figure 2) with the same sentence in (2) resulted in a design with one independent variable, Type of Picture, with three conditions, i.e. Control, Overexhaustive, and Underexhaustive. The correct answer in the Control (Figure 2A) and Overexhaustive (Figure 2B) conditions was ‘yes’, but in the Underexhaustive one (Figure 2C), it was ‘no’. Answering ‘no’ for (Figure 2B) would constitute an overexhaustive *q*-spreading error whereas answering ‘yes’ in (Figure 2C) would yield an underexhaustive *q*-spreading error.

The 58 filler items were of two types. One type was 32 reversible pictures depicting a transitive action with two animals that were paired with either canonical SVO or scrambled OVS sentences, 16 each (e.g. SVO *Belka vygulivaet utku* ‘The squirrel is walking the duck’, OVS *Utku vygulivaet belka* ‘The duck is being walked by the squirrel’). They were experimental items for a different experiment, and will not be discussed any further. The other type was 26 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-thirds of the filler items and one third of the experimental items required a ‘no’ answer; thus, the number of ‘yes’ and ‘no’ responses was balanced across the entire experiment.

Three versions of the experiment were created for the between-participants design, with experimental items rotated through the three conditions, four items per condition, in a Latin square design that allowed us to include random effects for items in our linear mixed effects statistical analysis. Participants were randomly assigned to one of the three versions. There were two main reasons for keeping the number of the experimental items low, i.e. 12 in total. First and foremost, we wanted to prevent participants from guessing the goal of the experiment and paying special attention to the quantified sentences (demand characteristics). Second, we also ran an identical experiment with 5-to-7-year-old Russian-speaking children in which it was important not to make the experiment too long. Thus, we followed the first language acquisition standards for keeping the number of items per condition small to avoid fatigue in children. For example, Roeper and colleagues (2004) had total of seven items in their experiment (two ‘Bunny’-spreading, two underexhaustive, and three overexhaustive items), and O’Grady et al. (2010) used four items per condition.

*d Procedure.* Participants were seated in front of the 17-inch HP laptop on which the pictures were presented, with auditory sentences being simultaneously played through speakers. The task was the classical sentence–picture verification in which the participants pushed the ‘yes’ button if the sentence correctly described the picture, and ‘no’ if it did not. The presentation of the pictures and sentences, recording of button pushes on a game pad, and collection of accuracy and reaction times were controlled by the DMDX software (Forster and Forster, 2003).

The stimuli laptop was connected to a remote free-viewing eye-tracking system (ETL-500) from ISCAN, Inc. that recorded the participants’ eye movements (for technical details, see Sekerina and Trueswell, 2011: 11). Eye movements were sampled at a rate of 30 times per second and were recorded on a digital SONY DSR-30 video tape-recorder. Prior to the experiment that lasted 20 minutes each participant underwent a short calibration procedure.

*e Data treatment and analysis.* We conducted two types of analyses, accuracy and reaction times, and fine-grained analysis of the eye movements. Trials that were not recorded due to the equipment malfunctioning constituted the missing data for the eye-movement analyses, 2.7% (13 trials) for the monolingual controls and 1.4% (5 trials) for the HL speakers. In addition, track losses which contained 2.54% and 4.69% of the eye-movements data for the monolingual and HL groups, respectively, were also excluded from these analyses.

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, seven categories were coded: looks to the first, second, and third of the entities in the front of the picture (identical in all three conditions; see Figure 2); looks to the first or second of the distractors in the back (elephants in the Control condition, extra bathtubs in the Overexhaustive and extra alligators in the Underexhaustive conditions); looks elsewhere in the picture, and track loss. Descriptive statistical analysis revealed that the middle pair in the front and the leftmost extra object/distractor in the back attracted the overwhelming majority of looks; thus, we collapsed the looks to the three pairs in the front into one category and the looks to the two extras/distractors in the back into another. Track loss (see above) and looks elsewhere constituted a very small proportion of total looks and were removed from the eye-movement analyses; therefore, fixations to the pairs in the front were in complimentary distribution with fixations to the extras/distractors in the back. Recall that the presence of extra objects (e.g. two alligators that are not in the bathtubs) or extra containers (e.g. two empty bathtubs) in the picture has been hypothesized in the L1 language acquisition studies (Minai et al., 2012) as responsible for children's *q*-spreading errors. Therefore, we chose to analyse the looks to the two extras/distractors in the back of the pictures.

In the fine-grained analysis, the proportions of looks were analysed in separate time windows, or *regions of interest* (ROIs). Each trial was segmented into four ROIs that were defined relative to the onset of the three phrases that comprised the spoken sentence (3):

| (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' | 'lies' | 'in a bathtub' | until the response via button push |

For each ROI on each trial, we computed the proportions of looks to the distractors (elephants) or extra objects (empty bathtubs/alligators), and all eye-tracking graphs will use the proportion of time the participants spent looking at them as the dependent variable.

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

Linear mixed effects models were calculated for both the accuracy and eye-movement data. Treatment contrasts were set for the fixed factor Type of Picture represented by two (sub-)predictor comparisons, Control vs. Overexhaustive and Control vs. Underexhaustive, and for the fixed factors Accuracy (correct vs. incorrect) and Group (monolinguals vs. HL speakers).

## 2 Results and discussion

To simplify the description of the statistics for the three types of behavioral measures – accuracy, reaction times, and fine-grained eye-movement data – as well as to provide the most straightforward comparison between monolingual Russian and HL Russian participants, we report the statistics of the fixed effects of the mixed effects models in tables. Each table consists of three panels: the left one shows the results for monolinguals, the middle one shows results for heritage speakers, and the right one shows the direct comparison of monolinguals and heritage speakers. 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 and *Pic* for Type of Picture).

Model fitting was performed in a step-wise fashion, starting with the most complex model that included the full factorial set of random effects (random slope-adjustment for the Type of Picture with random effects for participants and items in the models for accuracy and reaction times; slope-adjustment for Type of Picture, Accuracy and the

Type of Picture × Accuracy interaction for participants and items in the models for eye movements). During model fitting, the complex models were trimmed down in a stepwise fashion using log-likelihood tests for model comparisons (Baayen, 2008; Baayen et al., 2008). Slope-adjustments were kept in the models if the models fitted the data better than the less complex models. Model reduction started with the random effects for items and excluded first the random slope-adjustment for the interactions and then the adjustment for Accuracy and Type of Picture. Model reduction was then applied to the random effects for participants. For all of our statistical models, model reduction ended up with models with random intercept adjustment. Accordingly, we used the *pvals.fnc* function of the language R package (Baayen, 2008) to generate *p*-values for the models predicting reaction times and eye-movement patterns.

### *a Accuracy and reaction times*

Fillers: The monolinguals performed near ceiling on the 58 fillers, at 98.1%. The HL speakers' accuracy was also solid, 93.4%. We take these data to mean that the sentence–picture verification task was easy for both groups of participants.
Experimental items: The mean accuracy and confidence intervals as well as the reaction times for both groups of participants are presented in Figure 3. Table 2 lists the statistics of the LME models for accuracy and reactions times. It shows that, as expected, monolinguals were equally at ceiling for accuracy in all three conditions, 95.33%. In contrast, HL speakers were significantly less correct in the Overexhaustive condition (Figure 2A) than in the Control one, 80% versus 95%, but there was no difference in accuracy between the Underexhaustive and the Control conditions, 94% versus 95%. For the correctly answered trials, reaction time data revealed no differences in mean response times (RTs) among the three conditions for either the monolinguals or the HL speakers.

When comparing the accuracy between the groups (Table 2, right panel), we found a significant interaction ($b = -1.36$, $z = -.97$, $p < .05$) between the Type of Picture factor (Control vs. Overexhaustive) and Group (Monolinguals vs. HL) that was driven by the HL speakers who demonstrated 15% lower accuracy in the Overexhaustive condition compared to the Control one.

Next, we addressed the issue of whether *q*-spreading errors in the Overexhaustive condition characterize the HL group as a whole or whether there are distinct subtypes of HL speakers, with those who (a) made no errors, like monolingual controls, (b) who made an occasional error, and (c) who consistently made *q*-spreading errors. Descriptive statistics for accuracy in the Overexhaustive condition revealed that less than half of HL speakers (12) made no errors, 14 made one error (75% accuracy), and two participants made the most errors (Participants #13, 0% and #24, 25% accuracy, respectively). These data indicate that *q*-spreading in HL processing is not an exception, but rather is a relatively robust phenomenon that characterized 57% of the HL participants in the current study.

However, in contrast to monolingual children, HL speakers made *q*-spreading errors only in the Overexhaustive condition, and not in the Underexhaustive condition. The two extra empty containers in the former, but not the two extra objects in the latter, led to *q*-spreading in bilingual HL speakers. To address a possible source of these *q*-spreading

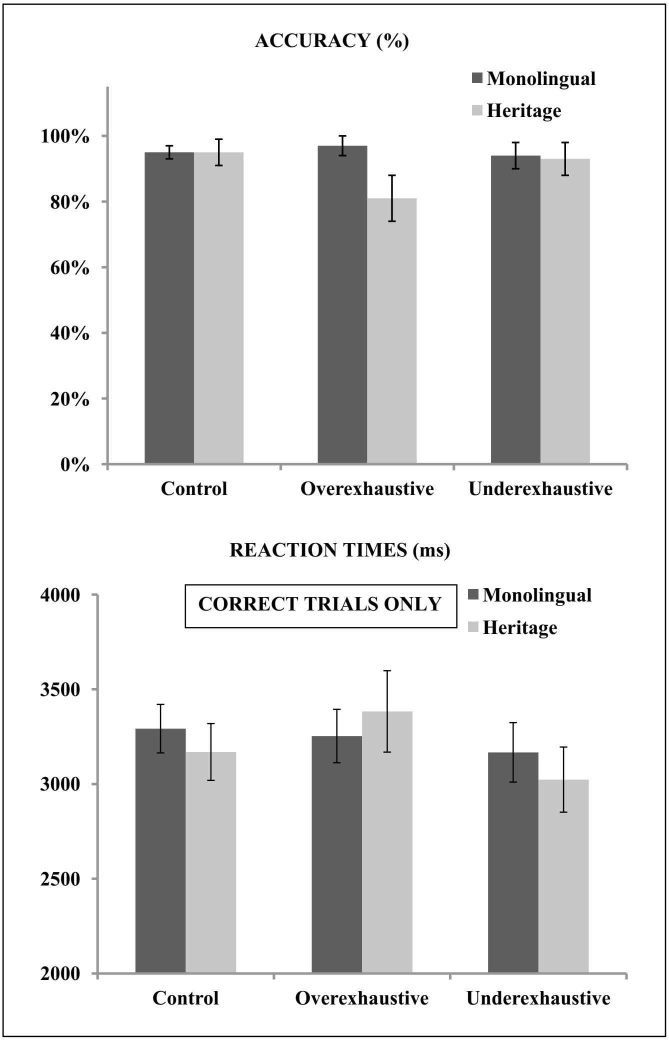



**Figure 3.** Mean accuracy (%) and reaction time data (ms) for the monolinguals (dark bars) and bilingual heritage Russian speakers (gray bars). Error bars represent 95% confidence intervals.

errors, correlational analyses were conducted to establish a relationship between accuracy and extralinguistic variables, i.e. age of arrival and the difference in self-assessed proficiency scores between Russian and English (see Table 1). None of these analyses yielded any significant correlations.

*b Fine-grained eye-movement data.* The fine-grained eye-movement data constitute the critical piece in establishing how overexhaustive *q*-spreading errors in HL processing described above happen. They allow us to uncover a relationship between visual attention and online language comprehension as spoken quantified sentences unfold

**Table 2.** Accuracy and reaction times results (mixed-effect models statistical analysis).

| | Monolingual speakers | | | Heritage speakers | | | Monolingual vs. heritage speakers | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy | *b* | SE | *z*-score | *b* | SE | *z*-score | *b* | SE | *z*-score | *p*-value |
| (Intercept) | 3.50 | 0.46 | 7.647*** | 3.21 | 0.49 | 6.510*** | 3.55 | 0.34 | 10.317 | <.001 |
| Overexhaustive | 0.53 | 0.65 | 0.817 | **–1.70** | **0.54** | **–3.162**** | –0.39 | 0.42 | –0.914 | .3607 |
| Underexhaustive | –.025 | 0.55 | –0.449 | –0.35 | 0.63 | –0.558 | –0.46 | 0.42 | –1.086 | .2775 |
| Group | | | | | | | –0.26 | 0.61 | –0.425 | .6712 |
| Group x Overexhaustive | | | | | | | **–1.36** | **0.69** | **–1.972** | **.0486*** |
| Group x Underexhaustive | | | | | | | 0.09 | 0.76 | 0.121 | .9037 |
| **Reaction times** | ***b*** | ***SE*** | ***t–score*** | ***b*** | ***SE*** | ***t–score*** | | | | |
| (Intercept) | 3295 | 100 | 32.73 | 3214 | 126 | 25.43 | | | | |
| Overexhaustive | –32 | 87 | –0.27 | 160 | 113 | 1.41 | | | | |
| Underexhaustive | –153 | 92 | –1.67 | –201 | 110 | –1.83 | | | | |

*Notes*. Left panel: monolingual speakers; middle panel: heritage language (HL) speakers; right panel: comparison of monolingual and heritage speakers. Significant effects are in bold face. * significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** highly significant at $\alpha = 0.001$, $^{+}$ trend at $0.05 < \alpha < 0.1$.

against the backdrop of a conflicting visual context. It is well established that eye movements reflect the tight coupling of visual context and referential processing; thus, upon hearing the words *alligator* (ROI 1) and *bathtub* (ROI 3) in (3), we would expect participants to focus their attention on the three object-container pairs in the front of the picture. All participants regardless of the group should be doing it in all three conditions. Predictably, there will be fewer looks to the extras/distractors in the back of the picture in all three conditions, but the extra containers (empty bathtubs) in the experimental Overexhaustive condition are expected to attract more looks than the distractors (elephants) in the Control condition because the latter are not mentioned in the sentence. In addition, increase in the looks to the extras compared to the distractors should happen as early as during the Verb region (ROI 2). This effect, however, may be delayed because HL speakers tend to be slower in their processing speed in comparison to monolingual controls.

Capitalizing on the important methodological point from previous eye-tracking studies (Dickey et al., 2007; Hanne et al., 2011), we considered the correctly and incorrectly interpreted trials in the Overexhaustive experimental condition for the HL group separately. This separation did not make sense for the monolingual control group due to a very small number of incorrect trials (4.7%). For HL speakers, a relatively large number of incorrectly interpreted trials (20% in the Overexhaustive condition) was critical in identifying an attentional 'signature' pattern of *q*-spreading in eye movements. The prediction is that more looks to the extra containers (two empty bathtubs in Figure 2B) in the Overexhaustive condition will indicate that HL bilingual speakers are distracted by interfering nonverbal information. When they dedicate disproportionally more attention to the irrelevant extra containers in the picture they make *q*-spreading errors similar to monolingual children.

We compared the time course of looks to the extra containers (the Overexhaustive condition) with the looks to the distractors (the Control condition) by focusing on each of the four consecutive ROIs in (3). The proportions of looks to the extras/distractors averaged for each ROIs are graphically represented in Figure 4 for both groups.

Monolingual participants' proportions of looks are plotted in black, while HL participants' proportions of looks are plotted in gray (dashed gray line represents incorrectly interpreted trials in the Overexhaustive condition). Statistical analyses that directly compared trials eliciting incorrect responses to the trials eliciting correct responses are unlikely to be reliable for the monolingual speakers in all three conditions, and for the HL speakers in the Control and Underexhaustive conditions, where they made very few errors (5% and 6%, respectively); thus, these are not discussed any further. Only the Overexhaustive condition yielded a sufficient amount of incorrect trials for HL speakers (20%). The results presented below focus on (1) comparisons of looks to the extras (containers/objects) vs. distractors in the correct trials for the three conditions for both groups, and (2) an additional comparison of correct vs. incorrect trials in the Overexhaustive condition for the HL speakers.

*ROI 1, 'Every alligator'*. The fixations that took place in this region are not informative because they happened too early in the sentence so they will not be discussed any further. In the remaining three regions (ROI 2, the Verb, through ROI 4, silence), both groups of participants looked equally often at the extras/distractors, with one striking

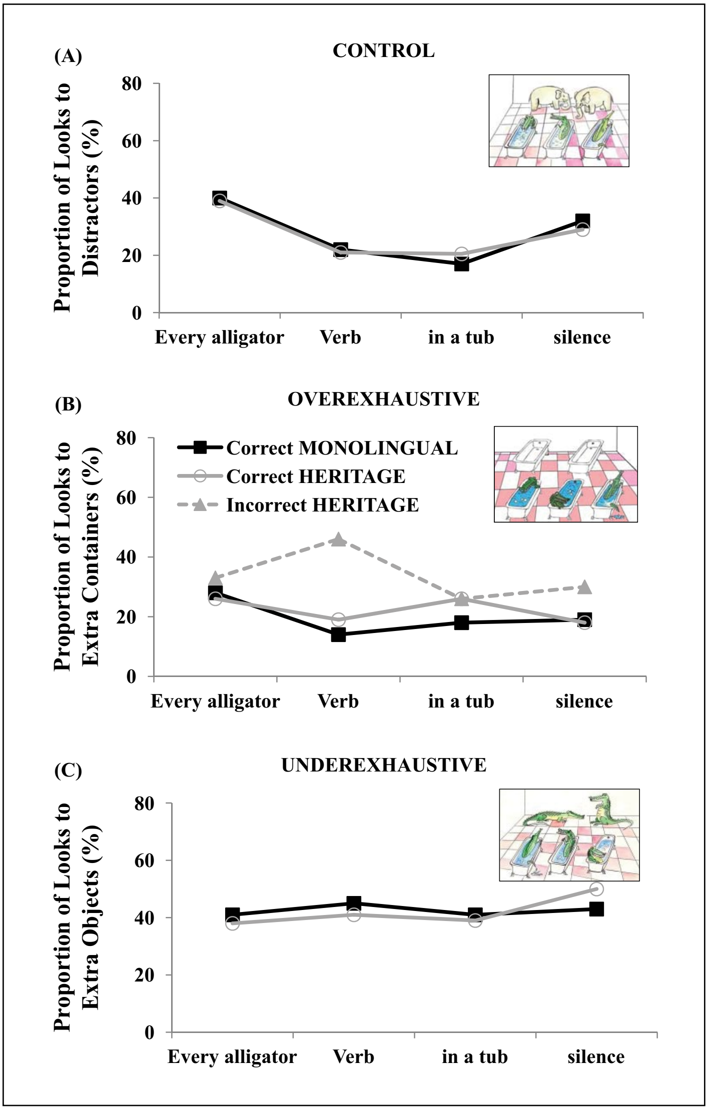



**Figure 4.** Fine-grained eye movements: Proportions of looks to the Extras/Distractors for monolinguals (black line – correct trials) and bilingual heritage speakers (solid gray line – correct trials, dashed gray line – incorrect trials) in four Regions of Interest.

exception of increased looks to the extra containers at ROI 2 in the incorrectly answered trials by HL speakers.

*ROI 2, the Verb*. In this region, there was a significant effect of accuracy for the HL speakers in the Underexhaustive condition: There were more looks to the extra containers (two empty bathtubs, Figure 4B) in the incorrectly answered trials (44%, dashed line) compared to correctly answered trials (16%; solid gray line) ($b$ = 0.28, $SE$ = 0.08, $t$ = 3.344, $p < .05$). Note that in this model, the Underexhaustive condition was the baseline because of a small amount of incorrect trials in the Control condition. Nevertheless, comparing the impact of accuracy between the Overexhaustive and Control conditions revealed that it had a similar effect on the proportions of looks in the Control condition (51% vs. 20%; $b$ = 0.02, $SE$ = 0.18, $t$ = 0.125). Accuracy did not influence the eye movements in the Underexhaustive condition (29% vs. 45%).

The eye-movement patterns also indicate a difference between monolingual and HL speakers with respect to the Overexhaustive condition, although the Group effect was not significant in the combined model. However, the separate models for each group indicate that monolingual speakers made fewer looks to the extras in the Overexhaustive condition compared to the distractors in the Control condition (14% vs. 22%, $b$ = −0.08, $SE$ = 0.04, $t$ = −2.084, $p < .05$), whereas the HL speakers looked equally often (16% vs. 20%, $b$ = −0.04, $SE$ = 0.45, $t$ = −0.755). A much larger difference was found in the incorrectly answered trials, where the increase in fixations to the extras by the HL speakers represents, in our opinion, a 'signature pattern' of *q*-spreading in eye movements. It indicates that HL speakers experienced interference from the irrelevant referents in the visual scene while the monolingual controls did not. However, this interference was transient: ROI 2 was the only region where accuracy was correlated with the eye-movement patterns; there was no effect of accuracy anywhere else later in the sentence.

In the Underexhaustive condition, the eye-movement patterns were similar for the monolingual controls and HL: There were significantly more looks to the extra objects (Figure 4C) compared to the looks to the distractors (Figure 4A) in the Control condition (monolinguals: 44% vs. 22%, $b$ = 0.22, $SE$ = 0.04, $t$ = 5.417, $p < .01$; HL speakers: 45% vs. 20%, $b$ = 0.25, $SE$ = 0.05, $t$ = 5.379, $p < .01$). This makes sense as the extras here are two additional alligators, and the sentence is about alligators. This word is the subject of the sentence paired with the quantifier *every*, and looking at all of the alligators is necessary to assess the truth of the proposition.

*ROI 3, Locative PP, and ROI 4, Silence*. This increase in the looks to the extras in the Underexhaustive condition continued to be present in the subsequent ROI 3 (locative PP) and ROI 4 (silence) (ROI 3: monolinguals: 41% vs. 15%, $b$ = 0.26, $SE$ = 0.03, $t$ = 7.649, $p < .001$; HL speakers: 42% vs. 20%, $b$ = 0.22, $SE$ = 0.04, $t$ = 5.680, $p < .001$; ROI 4: monolinguals: 44% vs. 30%, $b$ = 0.13, $SE$ = 0.04, $t$ = 3.284, $p < .01$; HL speakers: 46% vs. 30%, $b$ = 0.16, $SE$ = 0.05, $t$ = 3.220, $p < .01$).

In ROI 4, in the Overexhaustive condition, the eye-movement pattern was reversed for both groups, i.e. significantly fewer looks to the extra containers compared to the looks to the distractors (monolinguals: 18% vs. 30%, $b$ = −0.13, $SE$ = 0.04, $t$ = −3.375, $p < .001$; HL speakers: 17% vs. 30%, $b$ = −0.13, $SE$ = 0.05, $t$ = −2.656, $p < .05$). Again, this pattern makes sense if we take into the account the mismatch in the referents between the visual scene and the spoken sentence: the sentence is about the animals, namely, alligators, but there are no animals in the back of the picture for the Overexhaustive condition; in the Control conditions, the animals in the back are elephants (not mentioned in the sentence); and only in the Underexhaustive condition, additional looks are warranted by the presence of the two extra alligators.

In sum, when considering the correctly answered trials, the fixation patterns were very similar for the two groups throughout the entire sentence. Only in the incorrectly answered trials (the Overexhaustive condition for the HL speakers) did the clear difference emerged between the two groups: a significant increase in the looks to the extra containers led to *q*-spreading errors in HL speakers providing empirical evidence of failure of half of the participants in our sample to inhibit irrelevant interfering visual information in the case of the Overexhaustive condition.

Next we turn to the results for an identical experiment that was conducted in English with the HL speakers only, using the same design. In contrast to the Russian experiment, we expected that HL speakers should not experience difficulty in interpreting the quantified sentences in their dominant language (here, English), i.e. they should not make *q*-spreading errors in English.

## III English Experiment

### 1 Method

*a Participants.* The same 28 HL speakers that participated in the Russian experiment took part in the English experiment. All of them rated their knowledge of English as being superior to their knowledge of Russian, although comprehension and speaking scores were more similar in both languages than reading and writing.

*b Design and materials.* The experiment had the same 12 experimental, 58 fillers, and 3 practice items as the Russian experiment presented in English. The same three types of pictures were paired with the spoken sentence (4) creating the Control, Overexhaustive, and Underexhaustive experimental conditions:

(4) Every alligator is in a bathtub.

The sentence (4) was of the type that has been used in all of the previous studies of *q*-spreading errors in monolingual English-speaking children, with the quantifier *every* appearing as the subject in the sentence-initial position.

Three versions of the experiment were created for the between-participants design, with experimental items rotated through the three conditions, four items per condition, in a Latin square design. Participants were randomly assigned to one of the three versions.

*c Procedure.* Participants were required to take part in both Russian and English experiments that were separated by two weeks. The order of the experiment was counterbalanced: Each participant was randomly assigned to start with either the Russian or the English experiment. Thus, 14 participants were run on the Russian version followed by the English one, and for another 14 participants, the order of the experiments was reversed.

The remaining details of the procedure were exactly the same as in the Russian experiment, with the only difference being the language of the experiment (i.e. English).

*d Data treatment and analysis.* As the description of the results below will show, the participants' performance on the English experiment was superior to their performance in the Russian experiment. Thus, we only report the accuracy and RT results here and do not present the eye-movement data. There were very few incorrectly answered trials to warrant fine-grained eye-movement analysis.

### 2 Results and discussion

The accuracy of HL speakers in the English experiment was at ceiling in all three conditions, as shown in Figure 5. It presents the accuracy and RT data side-by-side for the Russian and English experiments for the HL speakers (error bars represent 95% confidence intervals). There were no differences in the accuracy among the three conditions in the English version of the experiment.

There were very few incorrect trials, and the RTs shown in the lower panel of Figure 5 are for the correctly answered trials only. Linear mixed effect models revealed that the reaction times for the HL speakers in the English version of the experiment were slower in the Control condition and the Overexhaustive experimental condition than in the Underexhaustive experimental condition ($b = -232$, $SE = 62.72$, $t = -3.712$, $p < .01$). In general, the performance of the HL speakers on the English version of the quantified sentences in terms of accuracy and RTs was much better than their performance on the Russian version although RTs may additionally reflect the fact that English sentences were shorter in duration than the Russian ones. What is important though is the fact that the Overexhaustive condition revealed no evidence of *q*-spreading errors in English.

## IV General discussion

In two experiments, monolingual Russian and bilingual HL Russian–English adults performed the standard sentence–picture verification task in which a spoken sentence containing the universal quantifier *kazhdyj* 'every' was paired with a picture depicting objects and containers that did not match in number. The task was conceptually identical to the one that has been used with monolingual children and adult L2 learners revealing that both groups of learners often make *q*-spreading errors in interpreting such quantified sentences. In our study, we tested a new adult group, i.e. bilingual HL speakers, and adapted the task to be performed in conjunction with eye-tracking that allowed us to investigate the moment-by-moment time course of processing of quantified sentences and estimate how the HL participants allocated visual attention to the picture in their heritage and dominant languages.

In the Russian experiment, in the Control and the Underexhaustive experimental conditions, both groups exhibited equally high accuracy rates and fast reaction times in interpreting the quantified sentences and showed identical eye-movement patterns of allocating visual attention to the two extra objects in the picture. Surprisingly, in the Overexhaustive condition, HL speakers' accuracy was significantly lower (80%) than in the Control condition and that of the monolingual controls (95%). Analyses of fine-grained eye-movement patterns performed separately for the correctly and incorrectly interpreted trials for the HL group revealed that the fixations diverged depending on how

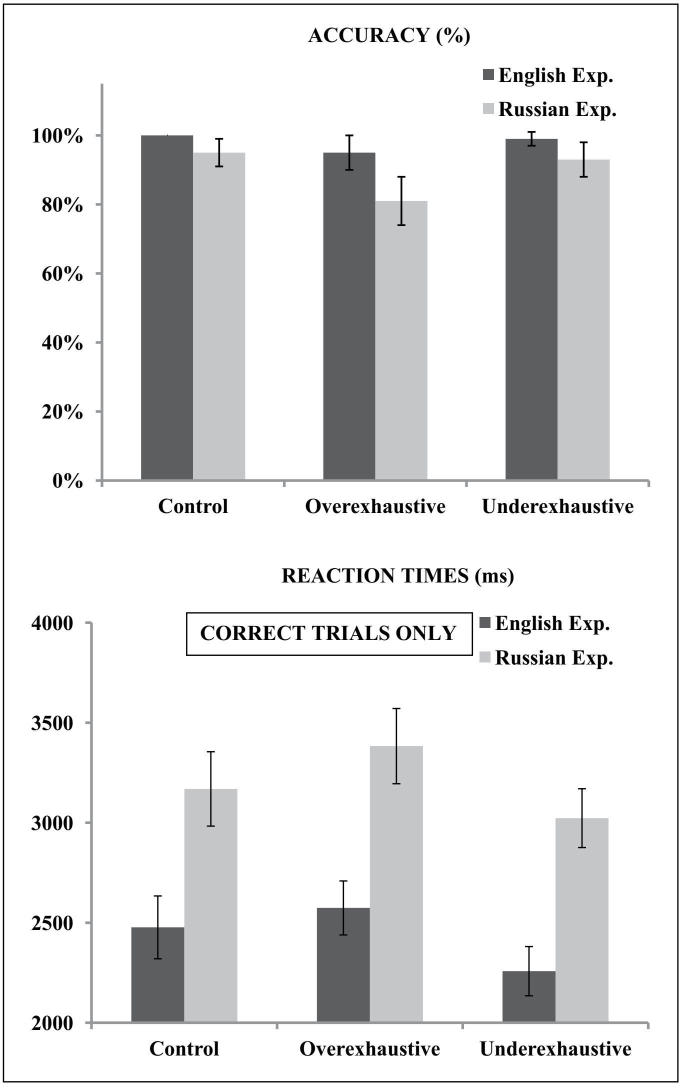



**Figure 5.** Mean accuracy (%) and reaction time data (ms) for the bilingual heritage Russian speakers in the English (dark bars) and Russian (gray bars) versions of the experiment. Error bars represent 95% confidence intervals.

the HL speakers assessed the logical truth of the quantified sentence. Those of them who misinterpreted the sentence, looked significantly more at the extra empty containers than the participants who correctly verified the sentence. These fixations took place immediately after the participants heard the quantified NP *every alligator* in the sentence, i.e. during ROI2, the Verb region, but disappeared soon thereafter. This suggests that the presence of the extra containers in the Overexhaustive condition worked as a trigger for *q*-spreading errors in HL speakers making their processing in heritage language similar to that of monolingual Russian children (Kuznetsova et al., 2007) and adult L2 learners of English (Berent et al., 2009; DelliCarpini, 2003). Slightly more than a half of the HL participants in our sample (57%) engaged in *q*-spreading indicating that bilingual HL speakers can be distracted by interfering nonverbal information when it is paired with spoken stimuli in the weaker, heritage, language. In contrast, the same HL speakers who made *q*-spreading errors in the Overexhaustive condition in their heritage Russian language were at ceiling for accuracy for exactly the same quantified sentences in their dominant language, English.

The results contribute to two issues at the intersection of bilingualism in general and HL bilingualism in particular, cognition, and language processing. The first is novel empirical evidence that in addition to adult L2 learners, another group of adults, namely, bilingual HL speakers, can be susceptible to *q*-spreading errors of a particular type. The second is implications of the bilingual HL speakers' errors in interpreting quantified sentences in real-time for the theories of HL learning and maintenance.

In this study, we have extended the findings from L1 and L2 acquisition research that has shown that monolingual children and adult L2 learners go through a stage in their language learning when they make *q*-spreading errors, to a special group of bilingual HL adults. The accuracy results confirmed that many HL speakers misinterpret quantified sentences in their heritage language, with the error rate that is significantly higher than in monolingual adults but lower than what is typically found with children and adult L2 learners. HL bilinguals also differ from monolingual children (no data available for adult L2 learners) in that their *q*-spreading errors were restricted to the overexhaustive sentences, and they made no errors in the underexhaustive items; children, on the other hand, make both types of *q*-spreading errors. This asymmetry is intriguing: from the logical point of view, the presence of the extra bathtubs in the Overexhaustive condition does not play any role in determining the truth value of the sentence whereas the presence of the extra alligators in the Underexhaustive condition does. Specifically, although the extra bathtubs are irrelevant for logical evaluation of the quantified sentence, they serve as a reliable trigger for *q*-spreading in HL processing whereas the extra alligators that are critical for the same purpose do not seem to work in the same way. We speculate that this asymmetry has to do with the mismatch in the salience of the referents between the spoken sentence and the visual context. In the Overexhaustive condition, the linguistic salience of the alligators that is prominently coded as the sentential subject is in conflict with the visual salience of the extra bathtubs in the scene. In the Underexhaustive condition, the linguistic and visual salience match; the alligators are the salient referents in both cases, and HL adults take advantage of this information.

The hypothesis that it is this mismatch in salience of referents between the spoken sentence and the visual context is confirmed by fine-grained eye-movement patterns of the HL participants in the incorrectly interpreted overexhaustive sentences. When HL speakers were confused by the mismatch of the salient referents, they dedicated disproportionally more attention to the irrelevant but visually salient referents (i.e. extra containers), and that ultimately led them to *q*-spreading errors in the sentence–picture verification task. We argue that these fixations to visually salient but irrelevant referents represent an online 'signature' pattern of *q*-spreading similar to what was found by Minai and colleagues (2012) for Japanese children.

What is the most plausible explanation of *q*-spreading errors that bilingual HL young adults made in our study while interpreting the sentences with the universal quantifier *kazhdyj* 'every' in Russian? One possibility that we entertained was the maturational account of *q*-spreading errors by adult L2 learners of DelliCarpini (2003) in that the lower proficiency participants made twice as many *q*-spreading errors, at the rate comparable to monolingual children, than the higher proficiency participants. The maturational account would predict that age of arrival and self-assessed proficiency scores in heritage Russian might correlate with the accuracy rate of the HL speakers. Our sample contained four participants who were born in the USA, 16 who arrived at or before age 6, and 8 who arrived after the age of 6 but before age 12 (mean 5.9 years). Self-assessed proficiency scores ranged from strong English-dominant (–3.5) to balanced (–0.25) between English and Russian (see Table 1). The correlations between these two sociolinguistic variables and accuracy were not significant, replicating the lack of proficiency effect in reducing the *q*-spreading error rate reported for the two adult L2 learner samples in the Berent et al.'s (2009) study. However, this lack of correlation should be interpreted with caution as the sample size and sample characteristics matter, and we leave the important question of the effect of proficiency for future work.

Returning to the issue of the underlying mechanisms that are responsible for *q*-spreading errors in comprehension of universal quantifiers in most monolingual children and adult L2 learners, some monolingual adults, and now bilingual HL speakers, we suggest that a unified explanation across ages and diverse learner groups is possible. The cause of *q*-spreading errors is of cognitive nature and has to do with cognitive (over)load. It quickly arises during online processing of the quantified sentences whose successful comprehension requires knowledge of how linguistic domains interact and real-time integration of multiple sources of information of diverse nature, i.e. the interface of syntax, semantics, pragmatics, and visual context. This cognitive feat is especially challenging for children (Grinstead, 2010; Trueswell et al., 1999), adult L2 learners (Berent et al., 2009), and even less educated monolingual adults (Street and Dąbrowska, 2010).

Bilingual HL speakers are not immune to this cognitive pressure, either. O'Grady et al. (2011) argue that processing difficulties arise when HL adults have to deal with phenomena that place a heavy demand on the HL processor, such as contrastiveness (Sekerina and Trueswell, 2012). These difficulties disproportionally affect the heritage language, and not the dominant one; recall that our HL participants did not experience cognitive overload in processing the same quantified sentences in English. If HL processing, in general, is more laborious, less efficient, and less automatic than processing in the dominant language, then integration of conflicting linguistic information in the weaker, heritage, language with attention-demanding visual context, as is the case with the overexhaustive quantified sentences, is likely to result in cognitive overload for many HL speakers. However, the fact that some of our HL participants were able to overcome this processing bottleneck strongly suggests that individual differences in cognition must be taken into consideration in the context of the participants and task demands when studying HL processing (Sekerina, 2012).

Finally, the present results have implications for the theories of HL learning and maintenance. The absence of a correlation between the age of arrival and the rate of *q*-spreading errors speaks against the incomplete acquisition account according to which HL adults are like 'fossilized' monolingual children. The HL participants who arrived between the ages of 6 and 12 were not immune to *q*-spreading errors in their heritage language; nor did the ones who arrived before the age of 6 make as many errors as monolingual children of that age. The attrition account that puts bilingual HL speakers closer to adult L2 learners on the continuum of language acquisition (Polinsky, 2011) fairs much better, but we need a more comprehensive theory of HL attrition. Issues of which linguistic and interface phenomena are susceptible to attrition in heritage language and which are resistant, as well as how they interact with cognitive and individual differences factors, clearly warrant further investigation.

## Declaration of Conflicting Interest

The authors declare that there is no conflict of interest.

## Funding

This work was partially supported by the National Science Foundation under ADVANCE Grant #0137851 and the PSC-CUNY 38 grant (#696053-00-38) to the first author. The second author was partially funded by the German Research Foundation (DFG), within the Collaborative Research Centre SFB 632 'Information structure' (Project C3). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## References

Baayen HR (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen RH, Davidson DJ, and Bates DM (2008) Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390–412.

Bates DM, Maechler M, and Dai B (2009) *lme4: Linear mixed-effect models using S4 classes: R package version 0.999375-31* [software]. Vienna: R Foundation for Statistical Computing.

Bayliss DM, Jarrold C, Gunn DM, and Baddeley AD (2003) The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General* 132: 71–92.

Benmamoun E, Albirini A, Montrul S, and Saadah E (2014) Arabic plurals and root and pattern morphology in Egyptian and Palestinian heritage speakers. *Linguistic Approaches to Bilingualism* 4: 89–123.

Berent GP, Kelly RR, and Schueler-Choukairi T (2009) Economy in the acquisition of English universal quantifier sentences: The interpretations of deaf and hearing students and second language learners at the college level. *Applied Psycholinguistics* 30: 251–90.

Brooks PJ and Braine MDS (1996) What do children know about the universal quantifiers 'all' and 'each'? *Cognition* 60: 235–68.

Brooks PJ and Sekerina IA (2005/2006) Shortcuts to quantifier interpretation in children and adults. *Language Acquisition: Special issue: Theoretical approaches to children's isomorphic interpretations* 13: 177–206.

Crain S, Thornton R, Boster C, Conway L, Lillo-Martin D, and Woodams E (1996) Quantification without qualification. *Language Acquisition* 5: 83–153.

DelliCarpini M (2003) *Developmental stages in the semantic acquisition of quantification by adult L2 learners of English: A pilot study*. In: Liceras JM, Zobl H, and Goodluck H (eds) *Proceedings of the 6th Generative Approaches to Second Language Acquisition conference (GASLA 2002): L2 Links*. Somerville, MA: Cascadilla, 55–63.

Dickey MW, Choy J-WJ, and Thompson CK (2007) Real-time comprehension of wh-movement in aphasia: Evidence from eyetracking while listening. *Brain and Language* 100: 1–22.

Drozd K (2001) Children's weak interpretations of universally quantified questions. In: Bowerman M and Levinson SC (eds) *Language acquisition and conceptual development*. Cambridge: Cambridge University Press, 340–76.

Fernández EM (2003) *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam/Philadelphia: John Benjamins.

Ferreira F, Bailey K, and Ferraro V (2002) Good enough representations in language comprehension. *Current Directions in Psychological Science* 11: 11–15.

Forster KL and Forster JC (2003) DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers* 35: 116–24.

Freeman NH and Schreiner K (1988) Complementary error patterns in collective and individuating judgements: Their semantic basis in 6-year-olds. *British Journal of Developmental Psychology* 6: 341–50.

Geurts B (2003) Quantifying kids. *Language Acquisition* 11: 197–218.

Grinstead J (2010) Linguistic interfaces and child Spanish. In: Domínguez L and Guijarro-Fuentes P (eds) *Proceedings of the romance turn 3*. Newcastle: Cambridge Scholars Publishing, 189–218.

Hanne S, Sekerina IA, Vasishth S, Burchert F, and De Bleser R (2011) Chance in agrammatic sentence comprehension: What does it really mean? Evidence from eye movements of German agrammatic aphasics. *Aphasiology* 25: 221–44.

Hollebrandse B (2004) Topichood and quantification in L1 Dutch. *IRAL: International Review of Applied Linguistics in Language Teaching* 42: 203–15.

Inhelder B and Piaget J (1964) *The early growth of logic in the child*. London: Routledge and Kegan Paul.

Kang H-K (2001) Quantifier spreading: Linguistic and pragmatic considerations. *Lingua* 111: 591–627.

Katsos N (2010) What can the acquisition of quantification tell us about assessing language development? Evidence from cross-linguistic, bilingual and SLI research. Paper presented at the Cost A33 Conference. University of Cambridge, UK.

Kuznetsova J, Babyonyshev M, Reich J, Hart L, and Grigorenko E (2007) The acquisition of universal quantifiers in Russian. In: Belikova A, Meroni L, and Umeda M (eds) *Proceedings of the 2nd conference on Generative Approaches to Language Acquisition North America (GALANA)*. Somerville, MA: Cascadilla Proceedings, 224–32.

Laleko O and Polinsky M (2013) Marking topic or marking case: A comparative investigation of heritage Japanese and heritage Korean. *Heritage Language Journal* 10: 40–64.

Lidz G and Musolino J (2002) Children's command of quantification. *Cognition* 84: 113–54.

Minai U, Jincho N, Yamane N, and Mazuka R (2012) What hinders child semantic computation: Children's universal quantification and the development of cognitive control. *Journal of Child Language* 39: 919–56.

Montrul S (2008) *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins.

O'Grady W, Kwak H-Y, Lee O-S, and Lee M (2011) An emergentist perspective on heritage language acquisition. *Studies in Second Language Acquisition* 33: 223–45.

O'Grady W, Suzuki T, and Yoshinaga N (2010) Quantifier spreading: Evidence from Japanese. *Language Learning and Development* 6: 116–25.

Philip W (1995) *Event quantification in the acquisition of universal quantification*. PhD dissertation, University of Massachusetts. Amherst: GLSA Publications.

Polinsky M (2011) Reanalysis in adult heritage language: New evidence in support of attrition. *Studies in Second Language Acquisition* 33: 305–28.

Pouscoulous N, Noveck IA, Politzer G, and Bastide A (2007) A developmental investigation of processing costs in implicature production. *Language Acquisition* 14: 347–75.

Rakhlin N (2007) A new pragmatic account of quantifier-spreading. *Nanzan Linguistics* 8: 239–82.

R Development Core Team (2009) *R: A language and environment for statistical computing*. (version 2.9.9) [software]. Vienna: R Foundation for Statistical Computing.

Roeper T, Strauss U, and Pearson BZ (2004) *The acquisition path of quantifiers: Two kinds of spreading*. Unpublished manuscript, University of Massachusetts at Amherst, MA, USA. Available at: http://people.umass.edu/roeper/online_papers/pathof%20acquisition3-15c.pdf (May 2014).

Sekerina IA (2012) We need an integrated, multiple-predictor model of native language proficiency. *Linguistic Approaches to Bilingualism* 2: 304–07.

Sekerina IA and Trueswell JC (2011) Processing of contrastiveness by heritage Russian bilinguals. *Bilingualism: Language and Cognition* 14: 280–300.

Sekerina IA and Trueswell JC (2012) Interactive processing of contrastive expressions by Russian children. *First Language* 32: 1–25.

Sorace A (2011) Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism* 1: 1–33.

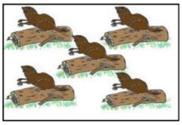
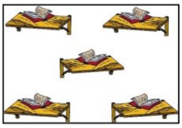
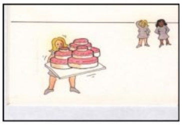
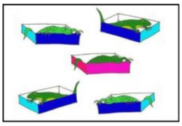
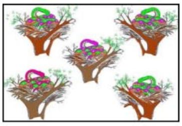
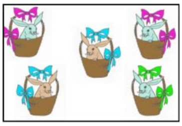
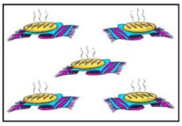
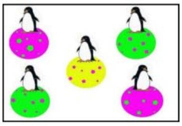
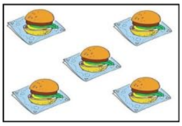
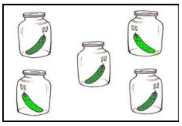
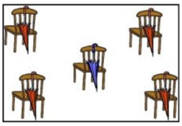
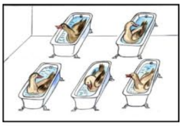
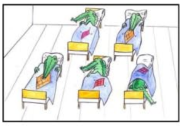
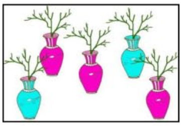
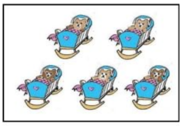
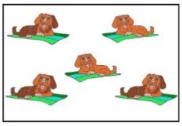
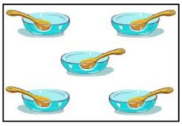
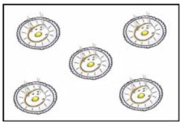
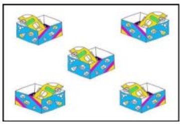
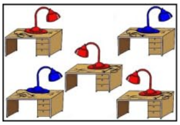
Street JA and Dąbrowska E (2010) More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua* 120: 2080–94.

Swets B, Desmet T, Hambrick DZ, and Ferreira F (2007) The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General* 136: 64–81.

Trueswell JC, Sekerina IA, Hill N, and Logrip M (1999) The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* 73: 89–134.

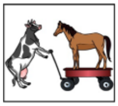
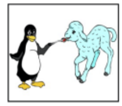
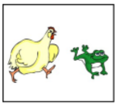
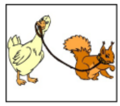
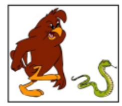
Trueswell JC and Tanenhaus MK (eds) (2004) *Approaches to studying word-situated language use*. Cambridge, MA: The MIT Press.

**Appendix 1.** Experimental and Filler Items.

| | Fillers (Type 1) | | | | |
|---|---|---|---|---|---|
| 100[a] (No) |  | 101 (No) |  | 102 (No) |  |
| 103 (No) |  | 107 (No) |  | 108 (No) |  |
| 109 (No) |  | 110 (No) |  | 112 (No) |  |
| 114 (No) |  | 115 (No) |  | 119 (No) |  |
| 121 (No) |  | 122 (No) |  | 124 (No) |  |
| 126 (No) |  | 127 (No) |  | 128 (No) |  |
| 130 (No) |  | 132 (No) |  | 133 (No) |  |
| 134 (No) |  | 136 (Yes) |  | 137 (No) |  |
| 138 (No |  | 142 (No) |  | | |

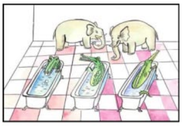
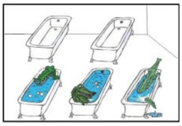
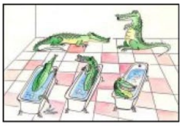
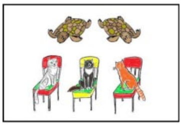
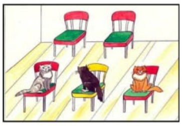
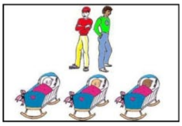
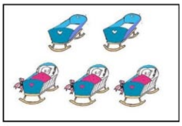
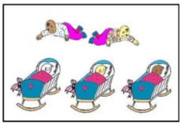
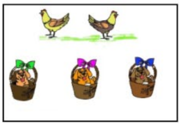
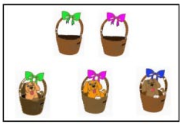
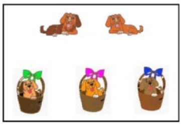
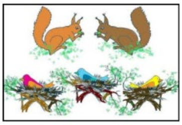
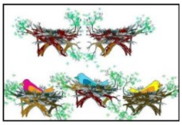
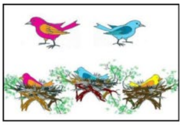
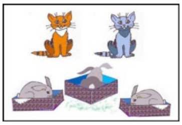
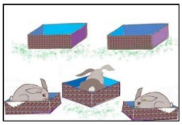
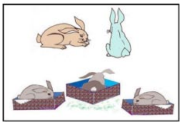
*Note.* [a]'No' and 'Yes' designate the correct answer in the sentence–picture verification task.

| Number | Fillers in Russian | English translation |
| --- | --- | --- |
| 100. | На этой картинке пять человек купают трех медведей. | In this picture 5 people are bathing 3 bears. |
| 101. | На этой картинке коровы живут в палатках. | In this picture the cows live in tents. |
| 102. | На этой картинке больше синих диванов, чем зеленых. | In this picture there are more blue sofas than green ones. |
| 103. | Бобры спятались под бревнами. | The beavers are hiding under the logs. |
| 107. | На розовых столах стоят белые книжки. | On pink desks there are white books. |
| 108. | Женщина несет один огромный торт. | A woman is carrying a giant cake. |
| 109. | Четыре ящерицы сидят в розовых коробках. | Four lizards are sitting in pink boxes. |
| 110. | У четырех змей зеленая голова. | Four snakes have green heads. |
| 112. | На этой картинке у зайчика в центре зеленый бантик. | In this picture the bunny in the center has a green bow. |
| 114. | Полотенца, на которых лежит хлеб, черно-белого цвета. | The towels on which there is bread are black and white. |
| 115. | У двух пингвинов шары желтого цвета. | Two penguins have yellow balls. |
| 119. | Салфетки под гамбургерами красного цвета. | The napkins under the hamburgers are red. |
| 121. | У четырех морских свинок розовые бантики. | Four guinea pigs have pink bows. |
| 122. | Желтая собачка сидит на красном стуле. | The yellow dog is sitting on a red chair. |
| 124. | В банках лежат шесть огурцов. | There are 6 pickles in the jars. |
| 126. | На этой картинке два зонтика синего цвета. | In this picture 2 umbrellas are blue. |
| 127. | В ваннах купаются черепахи. | The turtles are bathing in bath tubs. |
| 128. | На этой картинке люди купают маленького медведя. | In this picture the people are bathing a small bear. |
| 130. | На кроватях спит шесть крокодилов. | Six crocodiles are sleeping in beds. |
| 132. | На этой картинке больше голубых ваз, чем розовых. | In this picture there are more blue vases than red ones. |
| 133. | В кроватках спят птички. | The birds are sleeping in beds. |
| 134. | Собачки лежат на розовых ковриках. | The dogs are sitting on pink mats. |
| 136. | На этой картинке бумажные тарелки красного цвета. | In this picture the paper plates are red. |
| 137. | Деревянные ложки лежат в стеклянных мисках. | Wooden spoons are in glass bowls. |
| 138. | Коробки, в которых лежат автобусы, к оричневого цвета. | The boxes in which there are buses are brown. |
| 142. | На этой картинке больше синих ламп, чем красных. | In this picture there are more blue lamps than red ones. |

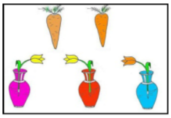
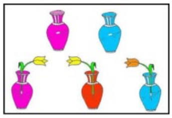
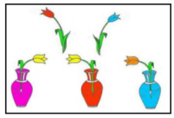
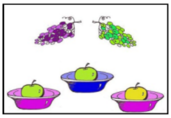
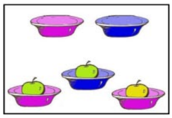
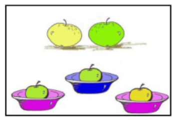
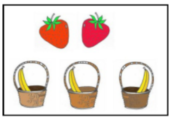
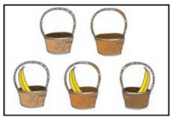
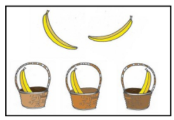
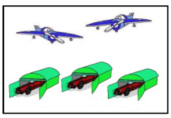
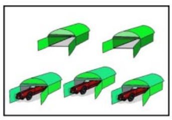
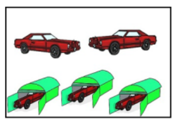
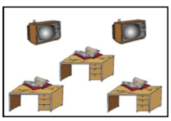
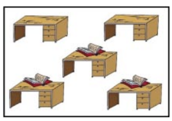
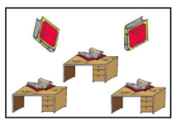
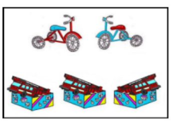
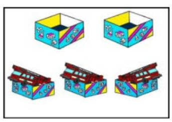
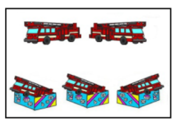
| | Fillers (Type 2) | | | | | | |
|---|---|---|---|---|---|---|---|
| 201[a] (No) |  | 203 (No) |  | 205 (Yes) |  | 207 (No) |  |
| 209 (No) |  | 211 (No) |  | 213 (Yes) |  | 215 (Yes) |  |
| 217 (Yes) |  | 219 (Yes) |  | 221 (Yes) |  | 223 (No) |  |
| 225 (No) |  | 227 (Yes) |  | 229 (Yes) |  | 231 (No) |  |

*Note.* [a] 'No' and 'Yes' designate the correct answer in the sentence–picture verification task.

| Number | Fillers in Russian | English translation |
|---|---|---|
| 201. | Зайчик держит медведя. | The bunny is holding the bear. |
| 203. | Лошадка везет корову. | The horse is pushing the cow. |
| 205. | Обезьянка причесывает льва. | The monkey is brushing the lion. |
| 207. | Собачку моет хрюшка. | The dog, the pig is washing. |
| 209. | Черепаха красит слона. | The turtle is painting the elephant. |
| 211. | Мышка щекочет крокодила. | The mouse is tickling the crocodile. |
| 213. | Девочка катает мальчика. | The girl is pushing the boy. |
| 215. | Овечку кормит пингвин. | The sheep, the penguin is feeding. |
| 217. | Лягушку преследует курица. | The frog, the chicken is chasing. |
| 219. | Кота держит динозавр. | The cat, the dinosaur is holding. |
| 221. | Утка выгуливает белку. | The duck is walking the squirrel. |
| 223. | Тигра обливает лиса. | The tiger, the fox is spraying. |
| 225. | На бабушку смотрит ребенок. | The Grandma, the baby is watching. |
| 227. | Жирафа лягает ослик. | The giraffe, the donkey is kicking. |
| 229. | Щенок рисует котенка. | The puppy is drawing the kitten. |
| 231. | Птицу преследует змея. | The bird, the snake is chasing. |

| | Experimental Items: Animate | | | |
|---|---|---|---|---|
| | Control | Overexhaustive | Underexhaustive | Sentence |
| 1. |  |  |  | Каждый аллигатор лежит в ванне. 'Every alligator is lying in a bathtub.' |
| 2. |  |  |  | Каждая кошка сидит на стуле. 'Every cat is sitting on a chair.' |
| 3. |  |  |  | Каждый ребенок спит в кроватке. 'Every baby is sleeping in a crib.' |
| 4. |  |  |  | Каждая собака сидит в корзинке. 'Every dog is sitting in a basket.' |
| 5. |  |  |  | Каждая птичка сидит в гнезде. 'Every bird is sitting on a nest.' |
| 6. |  |  |  | Каждый зайчик сидит в коробке. 'Every bunny is sitting in a box.' |

Experimental Items: Inanimate

| | Control | Overexhaustive | Underexhaustive | Sentence |
|---|---|---|---|---|
| 7. |  |  |  | Каждый цветок стоит в вазе. ‘Every flower is standing in a vase.’ |
| 8. |  |  |  | Каждое яблоко лежит в тарелке. ‘Every apple is lying in a plate.’ |
| 9. |  |  |  | Каждый банан лежит в корзинке. ‘Every banana is in a basket.’ |
| 10. |  |  |  | Каждая машина стоит в гараже. ‘Every car is in a garage.’ |
| 11. |  |  |  | Каждая книжка лежит на столе. ‘Every book is lying on a desk.’ |
| 12. |  |  |  | Каждая пожарная машина лежит в коробке. ‘Every fire truck is in a box.’ |