Studies in Theoretical Psycholinguistics 47

Katalin É. Kiss
Tamás Zétényi *Editors*

Linguistic and Cognitive Aspects of Quantification

Springer

Studies in Theoretical Psycholinguistics

Volume 47

The goal of this series is to bring evidence from many psychological domains to the classic questions of linguistic theory. The fundamental question from which the others flow is: What is the mental representation of grammar? Evidence from all aspects of language are relevant. How is the grammar acquired? How is language produced and comprehended? How is the grammar instantiated in the brain and how does language breakdown occur in cases of brain damage? How does second language acquisition and processing differ from first language acquisition and processing? A satisfactory theory of language calls for articulated connections or interfaces between grammar and other psychological domains. The series presents volumes that both develop theoretical proposals in each of these areas and present the empirical evidence needed to evaluate them.

Book proposals for this series may be submitted to the Publishing Editor:

Jolanda Voogd, Senior Publishing Editor, Springer, Van Godewijckstraat 30, 3300 AA Dordrecht, The Netherlands

Jolanda.Voogd@springer.com

More information about this series at http://www.springer.com/series/6555

Katalin É. Kiss · Tamás Zétényi
Editors

Linguistic and Cognitive Aspects of Quantification

Springer

Editors
Katalin É. Kiss
Research Institute for Linguistics
Hungarian Academy of Sciences
Budapest
Hungary

and

Pázmány Péter Catholic University
Budapest
Hungary

Tamás Zétényi
Department of Ergonomics and Psychology
Budapest University of Technology
and Economics
Budapest
Hungary

ISSN 1873-0043 ISSN 2215-1788 (electronic)
Studies in Theoretical Psycholinguistics
ISBN 978-3-319-91565-4 ISBN 978-3-319-91566-1 (eBook)
https://doi.org/10.1007/978-3-319-91566-1

Library of Congress Control Number: 2018941533

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Contributors

Asya Achimova Linguistics Program, Wayne State University, Detroit, USA

Oliver Bott University of Tübingen Project B1, Collaborative Research Centre SFB 833 & Project CiC, DFG Priority Program XPrag.de, Tübingen, Germany

Patricia J. Brooks Department of Psychology, The Graduate Center and College of Staten Island, City University of New York, New York, NY, USA

Luca Campanelli The Graduate Center, City University of New York, New York, NY, USA

Hamida Demirdache LLING UMR 6310 CNRS/Université de Nantes, Nantes, France

Viviane Déprez Department of Linguistics, Rutgers University, New Brunswick, USA; L2C2, CNRS, French academy of sciences, Lyon, France

Katalin É. Kiss Research Institute for Linguistics of the Hungarian Academy of Sciences; Pázmány Péter Catholic University, Budapest, Hungary

Martin Hackl Department of Linguistics and Philosophy, MIT, Cambridge, MA, USA

Nataša Knežević LLING UMR 6310 CNRS/Université de Nantes, Nantes, France

Thomas Hun-tak Lee Department of Linguistics and Modern Languages, Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Margaret Ka-yan Lei Department of Linguistics and Modern Languages, Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Levente Madarász Central European University, Budapest, Hungary

Julien Musolino Department of Psychology, Rutgers University, New Brunswick, USA

Irina Onoprienko Department of Linguistics and Philosophy, MIT, Cambridge, MA, USA

Fabian Schlotterbeck University of Tübingen Project B1, Collaborative Research Centre SFB 833 & Project CiC, DFG Priority Program XPrag.de, Tübingen, Germany

Anna M. Schwartz The Graduate Center, City University of New York, New York, NY, USA

Irina A. Sekerina Department of Psychology, The Graduate Center and College of Staten Island, City University of New York, New York, NY, USA

Ayaka Sugawara Department of Humanities, Mie University, Tsu, Mie, Japan

Balázs Surányi Research Institute for Linguistics of the Hungarian Academy of Sciences; Pázmány Péter Catholic University, Budapest, Hungary

Ken Wexler Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA; Department of Linguistics and Philosophy, MIT, Cambridge, MA, USA

Tamás Zétényi Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Budapest, Hungary

Introduction

Katalin É. Kiss

Abstract The Introduction briefly discusses some of the issues that quantification raises in syntax, semantics, prosody, and psycholinguistics. It highlights the aspects of quantification that invite experimental testing: the ambiguity of quantificational constructions, the virtual movement rules assumed in their derivation, differences in children's and adults' grammars of quantification, competing semantic and pragmatic accounts of certain interpretations, etc. Psycholinguistic studies testing the role of language in mathematical cognition are also mentioned. The Introduction also summarizes each chapter, surveying the types of quantifiers analyzed, the languages involved, the theories tested and compared, and the experimental methods employed.

Keywords Quantification · Mathematical cognition · Acquisition · Ambiguity
Quantifier scope · Quantificational domain · Scalar implicature · Distributivity
Eye-tracking · Quantifier spreading

Quantification has been in the focus of interest of generative linguistic theory since the nineteen seventies (see Chomsky 1976; May 1977, 1985; Huang 1982; Reinhart 1983, etc.). The principle of compositionality states that the meaning of a quantified sentence is derived from the meanings of its constituents and the rules used to combine them. Quantified sentences, however, are often ambiguous, sometimes in multiple ways, which is reconcilable with the principle of compositionality only if they are assigned multiple structures. Some (or in certain theoretical frameworks, all) of the structures assigned to a quantified sentence are derived from surface structure representations by virtual movement rules not affecting spellout. The assumption of operations not connected to spellout directly has been a challenge for psycholinguistics, as well.

K. É. Kiss (✉)
Research Institute for Linguistics of the Hungarian Academy of Sciences,
Benczúr utca 33, 1068 Budapest, Hungary
e-mail: e.kiss.katalin@nytud.mta.hu

K. É. Kiss
Pázmány Péter Catholic University, Szentkirályi u. 28, 1088 Budapest, Hungary

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_1

A line of psycholinguistic investigations has been testing whether children can access the multiple meanings of quantified sentences; whether the virtual movement rules deriving the logical forms that are subjected to semantic interpretation are parts of child grammar. Results showing that children cannot access certain interpretations, or cannot access the different interpretations with equal ease, have provided arguments for various hypotheses on whether the non-adult-like analysis and interpretation of quantificational structures is the manifestation of immature grammar (Philip 1995; Musolino 1998; Musolino et al. 2000; Roeper et al. 2004), or is due to processing difficulties related to memory limitations (cf. Musolino and Lidz 2003, 2006; Syrett and Lidz 2011), or is the consequence of pragmatic inexperience (see, e.g., Crain and Thornton 1998; Gualmini 2004, 2008; Philip 2011, among many others).

The interpretation of quantified sentences sometimes also causes problems for adults, and the investigation of their difficulties may shed light on how, by what mechanisms quantification is processed by the mature mental grammar (see, e.g., Bott and Schlotterbeck this volume). Psycholinguistic experiments can help us to choose between competing linguistic or psycholinguistic models of the given phenomenon. We can test their predictions on large populations in order to tell which of them matches speakers' behaviour more closely.

It has also been a productive research question how children and adults resolve the ambiguities of quantified sentences; which are their preferred interpretations, and how various pragmatic conditions affect the preferences (see cf. Brooks and Braine 1996; Musolino 2009; Pagliarini et al. 2012; Syrett and Musolino 2015; É. Kiss and Zétényi 2017, among many others). The results of these studies can contribute both to grammatical theory—e.g., by helping to distinguish default and derived structures (Papafragou and Musolino 2003), and to pragmatics—by providing experimental data in sufficiently large numbers to draw reliable generalizations (Surányi and Madarász this volume).

Whereas most psycholinguistic research into quantification has been motivated by questions of linguistic theory, there have also been psycholinguistic studies aiming to understand the role of language in numerical/mathematical cognition. So far these studies have mainly been concerned with the form and structure of number words, pointing out, e.g., that their compositionality in Chinese speeds up arithmetic cognition (Zhang and Simon 1985), or that the specific marking of dual number in Slovenian and Saudi Arabic speeds up the acquisition of the notion of 'two' (Almoammer et al. 2013; Marušič et al. 2016). Dechamps et al. (2015) found differences in the processing of the linguistic expressions *fewer, more* and the symbols <, >. This area of study is still mostly unexplored, providing many untapped research possibilities.

Another line of research investigates the mathematical cognition of speakers of languages with no exact numbers beyond 3 or 4, like Pirahã (Gordon 2004), Mundurukú (Pica et al. 2004), Australian aboriginal languages (Butterworth et al. 2008), or a Nicaraguan sign language (Spaepen et al. 2011), aiming to find out whether language and mathematical cognition interact in a deterministic way. Apparently speakers of languages with no exact numbers lack exact arithmetic, but have approximate arithmetic (Carey 2001; Spelke 2003; Dehaene (1997); Izard et al.

2008). Exact arithmetic is acquired in a language-specific format, as pointed out first by the behavioral and brain-imaging experiments of Dehaene et al. (1999), and confirmed by a large number of studies involving bilingual speakers, e.g., Spelke and Tsivkin (2001).

The majority of the chapters of this volume give account of experiments that were motivated by competing linguistic theories, e.g., theories of quantifier interpretation—concerning the determination of quantifier scope, the determination of quantificational domain, the conditions of distributive versus collective interpretation, etc. The experimental approach of psycholinguistics is particularly suitable to test pragmatic theories, or to confront syntactically or semantically based theories with pragmatic explanations, because the predictions of pragmatic theories are often preferences, the correctness of which can only be proven by statistically evaluated experimental results. The experiments presented involve various types of quantifiers (universals, existentials, numerals), and various languages (English, German, Serbian, Chinese, and Hungarian).

Notwithstanding the linguistic motivation, the results of these studies also bear on basic issues of psycholinguistics, sometimes even of psychology. Most studies have a developmental aspect, testing both children and adults, and some of them also investigate the potential correlation of linguistic achievement with intelligence and attention. Whereas in theoretical linguistics the question of the psychological reality of models rarely emerges, the experiments presented here, especially those involving eye-tracking and reaction time measurements, aim to reveal the mental procedure of quantification, and of sentence processing, in general.

The chapter entitled 'Structural asymmetry in question/quantifier interactions' by Asya Achimova, Viviane Déprez, and Julien Musolino helps to answer a question that has been present in the generative literature since the 1970s (see May 1977). The question is why sentence pairs like (1) and (2) have different scope possibilities; why only the former question elicits a pair-lists answer.

(1) *Which assignment did every student complete?*
(2) *Which student completed every assignment?*

The problem has actually turned out to be even more complex (see Kuno 1991): the pair-list answer becomes possible also in the latter sentence if the universal quantifier *every* is replaced by *each*:

(3) *Which student completed each assignment?*

The structural difference between the minimal pair in (1)–(2) suggests that the scopal difference is the manifestation of a subject-object asymmetry, which early analyses from May (1985) to Chierchia (1993) derived from various structural constraints. The minimal pair in (2)–(3), however, is structurally parallel; what (2) and (3) differ in is the specificity/distributivity of the universal quantifier. Incorporating this observation, more recent proposals (from Szabolcsi 1997 to Agüero-Bautista 2001) argue that quantifier type, too, affects scope possibilities. Achimova et al. tested experimentally whether a pair-list answer (i.e., wide scope) for a quantifier in a wh-question is licensed by structural position or by quantifier type, or whether the two conditions interact in some way. The experimental data show that both conditions have a role: wide scope assignment is easier to a subject quantifier than to an object quantifier whether the subject quantifier is an *every* or *each* phrase, and, wide scope assignment is easier to an *each* phrase whether it is in subject position or object position. It is proposed that a quantifier can be assigned wide scope if it can be construed as a topic. Subjecthood, and the strong distributivity characterizing *each* facilitate wide scope by evoking topic interpretation.

The chapter 'Children know the prosody-semantic/pragmatic link: Experimental evidence from Rise-Fall-Rise and scope' by Ayaka Sugawara, Martin Hackl, Irina Onoprienko and Ken Wexler investigates what determines the scope interpretation of sentences containing a universal quantifier in subject position and sentential negation (e.g., *All of the apples didn't fall*) by children aged 5;2–5;3 and by adults. Such sentences are, in principle, ambiguous scopally, however, they tend to be disambiguated—by the context and/or by prosody. According to Roberts (1996) and Büring (2003), the inverse scope reading is elicited by the 'contextually given' feature of the universal quantifier functioning as a contrastive topic. Büring (1997) and Krifka (1998), on the contrary, emphasize the role of contrastive prosody (the rise-fall-rise contour) in inverse scope interpretation. Sugawara and her colleagues carried out an experiment testing the role of both factors. The test sentences occurred in two different contexts; in one of them, the universal quantifier was new information, whereas in the other one it was given; and it occurred in both contexts with two different contours. It has turned out that the intonation contour does, the context does not, significantly influence scope interpretation. Somewhat surprisingly, children and adults were sensitive to the role of prosody in similar proportions (about 70% of both groups associated inverse scope with the rise-fall-rise intonation contour). This result also bears on a more basic issue, namely, whether the logical form of sentences subjected to semantic interpretation should include, or should have a direct access to, prosodic information. The finding that prosody plays a crucial role in quantifier scope interpretation is not compatible in a straightforward manner with the currently assumed architecture of grammar, where semantic interpretation has no access to prosodic information and phonological interpretation has no information about the movement rules carried out in logical form.

The question how children interpret the relative scope of negation and a universal quantifier quantifying over the subject also emerges in the chapter entitled *Differentiating universal quantification from perfectivity: Cantonese-speaking children's command of the affixal quantifier* saai3 by Margaret Lei and Thomas Hun-tak Lee. The chapter gives account of a study testing whether Cantonese children can distinguish the quantificational effects of a perfectivity-marking morpheme and a universal quantifier. In incremental-theme contexts a homomorphic mapping takes place between the noun phrase and the verbal predicate, i.e., quantification performed over subparts of an individual or a set is equivalent to quantification over sub-events denoted by a verbal predicate. Consequently, the 'totality of event(s)' meaning conveyed by the perfective aspect marker *zo2* is not distinct from the 'totality of object(s)' reading evoked by the universal quantifier suffix *saai3*. However, the two readings differ under negation. Negated perfective sentences denote the non-realization of the event, resulting in a ‘none’ reading. Negated universal quantifiers yield a partial, ‘not all’ reading—corresponding to the surface prominence of the negative auxiliary over the universal quantifier attached to the verb. The question whether children can access both perfectivity and universal quantification and whether they can distinguish them was tested in the context of negation. The experiment showed that children as young as 3;6–4;6 were able to tell the two readings apart, although a subject–object asymmetry was observed; children had problems with associating the quantifier with subject nominals, which Lei and Lee attribute to the intervention of the negator between the subject and the postverbal universal quantifier. In intransitive sentences denoting motion events, children interpreted quantification on the extent of the path traversed, within the scope of negation. In intransitive sentences containing no potential target of quantification other than the subject, children tended to assign to the universal quantifier scope over negation, i.e., they tended to opt for the ‘none’ reading.

The paper entitled *Scalar implicature or domain restriction: How children determine the domain of numerical quantifiers* by Katalin É. Kiss and Tamás Zétényi gives account of a series of experiments testing why Hungarian children have difficulties in test situations with accessing the ‘at least *n*’ interpretation in sentences like ‘If a boy has three hits [on the dartboard], he should get a candy’; why they think that boys with four or five hits should get none. Their experiments show that the ‘at least *n*’ reading of the numeral is only blocked if the domain of quantification is represented as a predetermined, fixed set. If the domain appears to be flexible, manipulatable, especially if it is not clearly demarcated, the majority of children realize that they can perform domain restriction. These results are hard to explain in the framework of the so-called neo-Gricean theory of numeral interpretation, where the basic meaning of a numeral *n* is the ‘at least *n*’ interpretation, and the ‘exactly *n*’ reading is a scalar implicature, derived by Grice’s maxims of quantity. An alternative theory of numerals, according to which the basic meaning of a numeral is the ‘exactly *n*’ reading, and the ‘at least *n*’ interpretation is due to pragmatic inferencing, is discarded on the basis of linguistic evidence (it cannot account for the fact that in the Hungarian sentence, the ‘at least *n*’ reading is the generally available interpretation; the ‘exactly *n*’ interpretation arises in the structural focus position, presumably as a consequence of the [+exhaustive]/[+maximal] feature associated with structural focus.) Instead of these two explanations, Stanley and Szabó’s (2000) semantic theory is adopted, according to which neither interpretation is derived from the other; the interpretation of a quantifier expression is always contextually determined, and what is flexible and is subject to change is the domain of quantification. The experiments of É. Kiss and Zétényi have shown that children are capable of domain widening and domain restriction depending on relevance, unless the quantificational domain is presented by the experimenter as a predetermined, fixed entity. Their results suggest that children may not follow the complex procedures of logical–semantic models in deriving quantificational domains; they may simply interpret contextual cues and manipulate sets.

Two chapters of the book deal with the phenomenon of distributivity. In the chapter *Universal quantification and distributive marking in Serbian,* Nataša Knežević and

Hamida Demirdache point out interpretive differences between three versions of distributive constructions. Languages may encode distributivity by marking the distributive key, i.e., the event participants over which the distribution takes place (e.g., ***Each boy*** *received an apple*), or by marking the distributed share, i.e., the entity that is being distributed (*The boys received* ***an apple apiece***), or by marking both. A question is if encoding distributivity by a distributive key marker (a universal), and encoding it by a distributed share marker yield equivalent interpretations. It has been argued that, whereas distributive key markers can imply either strong or weak distributivity, distributed share markers enforce strong distributivity. According to Balusu (2006), the strong distributivity of distributed share markers seemingly occurring without a distributive key is due to a covert universal quantifier ranging over spatiotemporal units. Knežević (2015) argued that the distributive-share marker *po* in Serbian is a pluractionality marker; it denotes a plurality of events, enforcing distributivity over spatiotemporal locations. It blocks collective readings, but it does not enforce exhaustivity. The distributive-key marker *svaki* 'every', on the contrary, enforces exhaustivity and atomicity—without blocking collective readings. In their present study, Knežević and Demirdache tested the acquisition of sentences involving both a distributive-key-marking *svaki* in subject position and a distributive-share marking *po* in object position. The acquisition path of distributivity indicates the independence of universal quantification and *po*. *Po* is acquired earlier; children at the age of 9 reject collective interpretations in the presence of *po*, but accept non-atomic interpretations in the presence of *svaki*. This suggests that, in languages that have both pluractional markers and universal quantifiers, such as Serbian, children acquire pluractionals before universal quantifiers.

The distributive–collective ambiguity and Information Structure by Balázs Surányi and Levente Madarász gives account of an experiment testing whether the discourse role of the subject affects the resolution of the ambiguity of sentences having both a collective and a distributive reading. The authors tested sentences with three different types of indefinite subjects, one of which (bare numeral indefinites like *five students*) only has a cardinal reading, while the other two (upward entailing comparative numeral phrases like *more than three students*, and *many* phrases, e.g., *many students*) have both a cardinal and a quantificational interpretation. The experiment was preceded by a series of pretests comparing the acceptance of the collective and distributive interpretations of 162 neutral sentences with subjects of the above three types. Only sentences in the case of which the pretest showed no significant pragmatically motivated bias towards either the collective or the distributive reading were included in the main test. The sentences of the main test occurred in three versions: the subject QNP was either topicalized, or focused, or was left in its base-generated vP-internal position, where it had no special discourse role. (The test was performed with Hungarian speakers in Hungarian, where the topic and focus functions are associated with distinct, easily recognizable structural positions.) It has been found that focusing significantly enhances the likelihood of distributive interpretation for all three subject types. This is only true of topicalization in the case of quantificational subjects (those of the type *more than three students* and *many students*), the quantificational reading of which is inherently distributive. In the case of focusing, the distributive interpretation is more optimal both for cardinal indefinite subjects, and for *more than* n and *many* subjects under their cardinal indefinite interpretation because it activates a smaller set of focus alternatives than the collective reading, thereby incurring smaller processing costs. Topicalization is argued to strengthen the distributivity of *more than* n and *many* subjects by supporting their quantificational interpretation. The quantification interpretation prevails in topic position because these quantificational expressions are associated with an existential presupposition, which meets the presuppositionality requirement of topic. In the case of bare indefinites, which lack a quantificational reading, topichood does not significantly affect interpretive preferences. The research reported in this paper is also interesting methodologically. The data were collected by crowd sourcing, and a program was developed to exclude spammers and careless respondents. The filtering of participants was partially criteria-dependent and was partially data-driven.

Two further chapters investigating the phenomenon of quantifier spreading, focusing on the processing of sentences containing a universally quantified subject binding an indefinite, also pertain to the issue of distributivity indirectly. The authors of *Quantifier spreading in school-age children*: *An eye-tracking study,* Irina Sekerina, Patricia Brooks, Luca Campanelli, and Anna Schwartz, investigated among children aged 5–12 why a sentence like *Every bunny is in a box* is often rejected in a situation involving, say, three bunnies, each in a box, and an empty box. The experiment involved sentence–picture verification, in the course of which the authors performed eye-tracking, and measured reaction times. They also tested the verbal and non-verbal intelligence of the subjects. They have found that errors involve greater numbers of fixations to the extra objects, which occurred right after the utterance of the quantified noun phrase. This suggests that quantifier spreading cannot be a consequence of children's lack of control of attention, contra the proposal of Minai et al. (2012). Correct responses required longer reaction times, indicating that additional processing is needed for children to correctly restrict the universal quantifier to the appropriate noun phrase. Children's achievement did not correlate with intelligence, and only weakly correlated with their age. The fact that quantifier spreading mistakes only mildly decrease by maturation is hard to accommodate in frameworks that attribute children's errors to immature grammar. It is concluded that the theory which can account for the full range of the facts attested is the theory that attributes errors to the superficial processing of sentence structure (Brooks and Braine 1996; Brooks and Sekerina 2005/2006). Shallow sentence processing generates 'good enough' (underspecified) representations of sentence structures that under most circumstances are sufficient for comprehension. When relying on shallow processing, children (and also adults) use canonical collective and distributive representations as defaults.

The chapter *Turning adults into children: Evidence for resource-based accounts of errors with universal quantification* by Oliver Bott and Fabian Schlotterbeck shows that not only children but adults, too, are prone to commit quantifier spreading errors in circumstances that make great demands on their cognitive resources. Some adults commit the extra object error observed in the case of children, i.e., they reject a sentence meaning 'each pupil was praised by exactly one teacher' in a situation where there is an extra teacher not praising anyone. However, even more adults commit a so-called branching error not observed before. Namely, they reject sentences meaning 'each pupil was praised by exactly one teacher' in a situation where each pupil was praised by exactly one teacher, but some teacher praised two pupils. The authors were interested in whether current competing theories of quantifier spreading can account for these facts. They tested adults in two conditions. In the first experiment, the picture (a set diagram) and the corresponding sentence were shown incrementally, i.e., the participants saw the sentence word by word after the picture had disappeared. They had to judge at each step whether the unfolding sentence still matches the diagram. The occurrence rate of the branching error was 44%. The experiment was also repeated in an offline version when the picture and the full sentence were shown simultaneously. In this condition, adults made practically no mistakes. No mistakes were attested in a third online version of the experiment, either, where the universal quantifier was replaced by an indefinite. This test excluded the possibility that the errors in the first experiment were retention failures. Bott and Schlotterbeck assume that speakers automatically assign to sentences containing a universal and an indefinite a default symmetrical interpretation, which is the cognitively simplest state of affairs making the sentence true. In the case of an extra object, it is relatively easy to recognize that the default model is a proper part of the actual picture. In the case of a branching line, however, the matching procedure breaks down, they have to verify whether each pupil is connected to exactly one teacher, which is a demanding process affected by resource limitations. None of the current theories of quantifier spreading—which derive spreading errors from a grammatical deficit, or processing problems, or infelicitous pragmatic conditions—can fully account for these findings. The proposed account is a resource-based processing explanation, claiming that the more complex a verification procedure is, the more exposed it is to resource limitations.

In sum: the chapters of this volume have something to offer to linguists, psycholinguists, and psychologists alike. Old puzzles of scope interpretation concerning the relative scope of wh-phrases and universal quantifiers, and universal quantifiers and negation have been resolved. The semantics of distributivity has been completed with further details—concerning the contribution of the different ingredients of distributive constructions, the distributive force of different universal quantifiers, and potential overlaps between cumulativity and distributivity. Theories aiming at psychological plausibility have been supported experimentally—e.g., Stanley and Szabó's (2000) theory of the context dependence of quantifier domain. The papers focusing on sentence processing offer new insights into cognitive processes, among them the interaction of attention, cognitive load, and linguistic analysis. (As expected, intelligence was not among the factors found to correlate with processing achievement.) The existence of shallow sentence processing, generating underspecified or default representations, has been confirmed.

The book also illustrates the great variety of methodological solutions that can be applied in the study of quantification. The experiments employed various versions of truth-value judgement and forced choice tasks involving the verification of sentence–picture and sentence–video pairs. In some cases, acceptibility judgements were supplemented by elicitation tasks. The preponderance of intuitional data reflects the fact that quantification tends to result in ambiguities, the resolution of which involves a great extent of intuitional uncertainty, deriving from the interaction of structural, semantic, contextual and pragmatic factors. However, some papers also present ways of increasing the objectivity of intuitional data, for example, by a meticulous pretest screening the stimuli so as to exclude all examples with any potentially distracting idiosyncratic features, and by a meticulous post-test screening the seriousness and the concentration of the informants. In some of the experiments described, intuitional results are supplement by biological data—for example, by visual-world eye-tracking in the case of children and self-paced reading and reaction time measurement in the case of adult subjects. The diversity of approaches is in part a consequence of the fact that the experiments range over several age groups from preschoolers to adults, thereby outlining the acquisition path of various quantificational constructions.

References

Agüero-Bautista, C. 2001. Cyclicity and the Scope of wh-Phrases. PhD dissertation, MIT.

Almoammer, A., J. Sullivan, C. Donlan, F. Marušič, R. Žaucer, T. O'Donnell, and D. Barner. 2013. Grammatical Morphology as a Source of Early Number Word Meanings. *PNAS* 110 (46): 18448–18453. https://doi.org/10.1073/pnas.1313652110.

Balusu, R. 2006. Distributive Reduplication in Telugu. *Proceedings of NELS* 36: 39–53.

Brooks, P.J., and M.D.S. Braine. 1996. What Do Children Know About the Universal Quantifiers *All* and *Each*? *Cognition* 60: 235–268.

Brooks, P.J., and I.A. Sekerina. 2005/2006. Shortcuts to Quantifier Interpretation in Children and Adults. *Language Acquisition* 13(2): 177–206.

Büring, D. 1997. The Great Scope Inversion Conspiracy. *Linguistics and Philosophy* 20: 175–194.

Büring, D. 2003. On D-Trees, Beans, and B-Accents. *Linguistics and Philosophy* 26: 511–545.

Butterworth, B., R. Reeve, F. Reynolds, and D. Lloyd. 2008. Numerical Thought With and Without Words: Evidence from Indigenous Australian Children. *PNAS* 105 (35): 13179–13184. https://doi.org/10.1073/pnas.1313652110.

Carey, S. 2001. Cognitive Foundations of Arithmetic: Evolution and Ontogenesis. *Mind and Language* 16: 37–55.

Chierchia, G. 1993. Questions with Quantifiers. *Natural Language Semantics* 1: 181–234.

Chomsky, N. 1976. Conditions on Rules of Grammar. *Linguistic Analysis* 2: 303–351.

Crain, S., and R. Thornton. 1998. *Investigations into Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge: MIT Press.

Dehaene, S. 1997. *The Number Sense*. Oxford, New York: Oxford University Press.

Dehaene, S., E. Spelke, P. Pinel, R. Stanescu, and S. Tsivkin. 1999. Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence. *Science* 284:970–974.

Deschamps, I., G. Agmon, V. Loewenstein, and Y. Grodzinsky. 2015. The Processing of Polar Quantifiers, and Numerosity Perception. *Cognition* 143: 115–128.

É. Kiss, Katalin, and T. Zétényi. 2017. Why is Children's Interpretation of Doubly Quantified Sentences Non-isomorphic? *Linguistics* 55(6).

Gordon, P. 2004. Numerical Cognition Without Words: Evidence from Amazonia. *Science* 306: 496–499.

Gualmini, A. 2004. Some Knowledge Children Don't Lack. *Linguistics* 42 (5): 957–982.

Gualmini, A. 2008. The Rise and Fall of Isomorphism. *Lingua* 118: 1158–1176.

Huang, C.-T. J. 1982. Logical Relations in Chinese and the Theory of Grammar. PhD dissertation, MIT.

Izard, V., P. Pica, E. Spelke, and S. Dehaene. 2008. Exact Equality and Successor Function: Two Key Concepts on the Path Towards Understanding Exact Numbers. *Philosophical Psychology* 21 (4): 491–505.

Knežević, N. 2015. Numerals and Distributivity in Serbian: At the Syntax-Semantics-Acquisition Interface. PhD dissertation, University of Nantes.

Krifka, M. 1998. Scope Inversion Under the Rise-fall Contour in German. *Linguistic Inquiry* 29: 75–112.

Kuno, Susumu. 1991. Remarks on Quantifier Scope. In *Current English Linguistics in Japan*, ed. H. Nakajima, 261–287. Berlin: Mouton de Gruyter.

Marušič, F.,R. Žaucer, V. Plesničar, T. Razboršek, J. Sullivan, and D. Barner. 2016. Does Grammatical Structure Accelerate Number Word Learning? Evidence from Learners of Dual and Non-dual Dialects of Slovenian. *PLOS* 2016. http://dx.doi.org/10.1371/journal.pone.0159208.

May, R. 1977. The Grammar of Quantification. PhD dissertation, MIT.

May, R. 1985. *Logical Form: Its Structure and Derivation*. Cambridge: MIT Press.

Minai, U., N. Jincho, N. Yamane, and N.R. Mazuka. 2012. What Hinders Child Semantic Computation: Children's Universal Quantification and the Development of Cognitive Control. *Journal of Child Language* 39: 919–956.

Musolino, J. 1998. Universal Grammar and the Acquisition of Semantic Knowledge: An Experimental Investigation into the Acquisition of Quantifier-Negation Interaction in English. PhD dissertation, University of Maryland, College Park.

Musolino, J. 2009. The Logical Syntax of Number Words: Theory, Acquisition and Processing. *Cognition* 111: 24–45.

Musolino, J., S. Crain, and R. Thornton. 2000. Navigating Negative Quantificational Space. *Linguistics* 38 (1): 1–32.

Musolino, J., and J. Lidz. 2003. The Scope of Isomorphism: Turning Adults into Children. *Language Acquistion* 11 (4): 277–291.

Musolino, J., and J. Lidz. 2006. Why Children Aren't Universally Successful with Quantification. *Linguistics* 44 (4): 817–852.

Pagliarini, E., G. Fiorin, and J. Dotlacil. 2012. The Acquisition of Distributivity in Pluralities. In *36th Annual Boston University Conference on Language Development*, 387–399. Cascadilla Press.

Papafragou, A., and J. Musolino. 2003. Scalar Implicatures: Experiments at the Syntax Semantics Interface. *Cognition* 78: 165–188.

Philip, W. 1995. Event Quantification in the Acquisition of Universal Quantification. PhD dissertation, University of Massachusetts at Amherst, GLSA Publications.

Philip, W. 2011. Acquiring Knowledge of Universal Quantification. In *Handbook of Generative Approaches to Language Acquisition*, ed. J. de Villiers and T. Roeper, 351–394. Dordrecht: Springer.

Pica, P., C. Lemer, V. Izard, and S. Dehaene. 2004. Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science* 306: 499–502.

Reinhart, T. 1983. *Anaphora and Semantic Interpretation*. London: Croom Helm.

Roberts, C. 1996. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Working Papers in Linguistics, Ohio State University Department of Linguistics,* 91–136. Republished in *Semantics & Pragmatics* 5: 1–69.

Roeper, T., U. Strauss, and B.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.

Spaepen, E., M. Coppola, E. Spelke, S. Carey, and S. Goldin-Meadow. 2011. Number Without a Language Model. *PNAS* Feb. 7. https://doi.org/10.1073/pnas.1015975108.

Spelke, E. 2003. What Makes Us Smart? Core Knowledge and Natural Language. In *Language and Mind. Advances in the Study of Language and Tought,* ed. D. Gentner and S. Goldin-Meadow, 277–311. Cambridge, MIT Press.

Spelke, S., and S. Tsivkin. 2001. Language and Number: A Bilingual Training Study. *Cognition* 78: 45–88.
Stanley, J., and Z.G. Szabó. 2000. On Quantifier Domain Restriction. *Mind and Language* 15: 219–261.
Syrett, K., and J. Lidz. 2011. Competence, Performance, and the Locality of Quantifier Raising: Evidence from 4-Year-Old Children. *Linguistic Inquiry* 42: 305–337.
Syrett, K., and J. Musolino. 2015. Collectivity, Distributivity, and the Interpretation of Plural Numerical Expressions in Child and Adult Language. *Language Acquisition* 20: 259–291.
Szabolcsi, A. 1997. Strategies for Scope Taking. In *Ways of Scope Taking*, ed. A. Szabolcsi, 109–154. Dordecht: Kluwer.
Zhang, G., and H. Simon. 1985. STM Capacity for Chinese Words and Idioms: Chunking and Acoustical Loop Hypotheses. *Memory and Cognition* 13: 193–201.

Structural Asymmetry in Question/Quantifier Interactions

Asya Achimova, Viviane Déprez and Julien Musolino

Abstract The interaction of universal quantifiers and *wh*-phrases in questions, such as *Which class did every student take?*, gives rise to structural ambiguities. The availability of pair-list answers (*Mary took Syntax, and Jane took Semantics*) to such questions reveals whether the quantifier can take wide scope over the *wh*. In this paper, we use an acceptability judgment task to test whether, as some theoretical accounts suggest (e.g. May 1985), the quantifier position affects the likelihood of an inverse scope reading for distributive quantifiers, such as *every* and *each*. We show that pair-list answers remain less available for questions with object quantifiers than for questions with subject quantifiers even when the quantifier is *each* (*contra* Beghelli 1997). At the same time, speakers find pair-list answers to questions with *each* more acceptable than to questions with *every*, confirming that the distributivity force of a quantifier also plays a role. We discuss how these findings fit into the existing analyses of quantifier scope in relation to quantifier semantics and discourse structure.

A. Achimova (✉)
Linguistics Program, Wayne State University, Detroit, USA
e-mail: asya.achimova@gmail.com

V. Déprez
Department of Linguistics, Rutgers University, New Brunswick, USA
e-mail: deprez@rci.rutgers.edu

V. Déprez
L2C2, CNRS, French Academy of Sciences, Lyon, France

J. Musolino
Department of Psychology, Rutgers University, New Brunswick, USA
e-mail: julienm@ruccs.rutgers.edu

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_2

1 Introduction

Questions with universal quantifiers may be structurally ambiguous and allow multiple readings. The question in (1) can be understood as (1a) where there is a single assignment that every student completed, or as (1b), where there are pairings of students and their individual assignments. Finally, we could specify the pairings of students and assignments not extensively by listing them, but rather by naming a function, in this case, the hardest assignment (1c), which is presumably different for every student.

(1) Which assignment did every student complete?
   a. The semantics assignment. *Single answer*
   b. John completed the semantics assignment, Jane completed the syntax assignment, and Mary completed the phonology assignment. *Pair-list answer*
   c. The hardest assignment. *Functional answer*

May (1985) was one of the first to observe that the position of the quantifier determines the range of possible answers. He argued that pair-list answers (PLA) are lacking for questions with object quantifiers, such as (2).[1]

(2) Which student completed every assignment?
   a. Mary.
   b. *John completed the semantics assignment, Jane completed the syntax assignment, and Mary completed the phonology assignment.

However, this structural restriction on PLA availability does not hold for all universal quantifiers equally. Beghelli (1997) reported that PLAs to questions with *each* (3) in object position freely allow pair-list readings (3b), indicating that the wide scope of the quantifier is possible. Single answers are available as well (3a).

(3) Which student completed each assignment?
   a. John did.
   b. John completed the semantics assignment, Jane completed the syntax assignment, and Mary completed the phonology assignment.

In this paper, we show using experimental tools that the structural position of the quantifier in fact affects the accessibility of a PLA regardless of the lexical differences between universal quantifiers, such as *every* and *each*. The rest of the paper is structured as follows: we first review the theoretical background explaining the role of structure and quantifier semantics. We follow with the results of our acceptability judgment experiments. We conclude with a discussion of the subject-object asymmetry and speculate about the possible sources of this effect.

[1] Since functional answers are not the focus of this paper they will not be discussed further here.

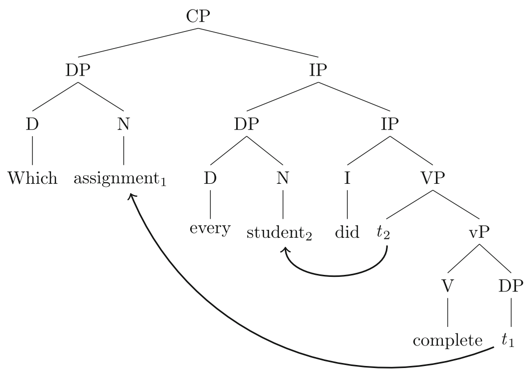
2 Theoretical Background

2.1 Structural Limits on the Wide Scope Reading of Quantifiers

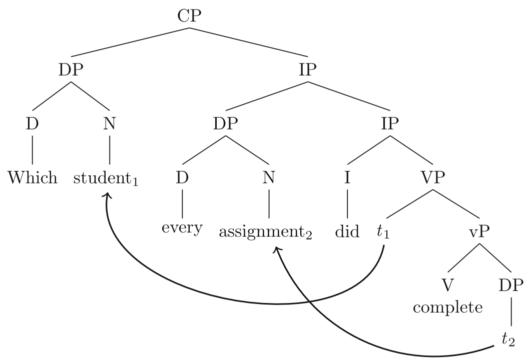
The observation that certain questions with quantifiers in object position lack pair-list readings led to the development of several analyses to account for this fact. We will first review the accounts that attribute the inability of object quantifiers to take wide scope over a *wh*-phrase to syntactic effects.

May (1985) argues that object quantifiers fail to scope over a *wh*-phrase due to a violation of constraints on movement. In May's view, the inverse scope of a quantifier phrase and a *wh*-phrase is possible if they can form a special $\sum$-sequence. Members of the $\sum$-sequence are governed by the same maximal projection. If such a formation is possible at the level of LF, members of the sequence can freely interact and scope over each other giving rise either to a single answer or to a PLA. While subject quantifiers can raise to a position close enough to the *wh*-phrase (4) to form a $\sum$-sequence, the movement path of an object quantifier must cross the movement path of a subject *wh*-phrase in (5).

(4)



(5)



This path crossing in (5) then violates the Path Containment Condition (PCC) proposed by Pesetsky (1982) who argues that multiple A' movement paths must nest rather than cross. What we have in the end is the inability of an object quantifier phrase and a *wh*-phrase to form an appropriate $\sum$-sequence that could license an inverse scope reading of the quantifier phrase. Hence, there can be no PLAs for questions, such as (2). May's structural account of the subject-object asymmetry in the availability of PLAs is related to other subject-object asymmetries known as Comp-trace effects (Pesetsky (1982), among others). The view that Comp-trace effects result only from characteristic structural asymmetry has been questioned in works starting with Déprez (1991, 1994).

Aoun and Li (1993) developed an alternative proposal that explains the inability of questions with object quantifiers to give rise to PLAs. Relying on evidence from Chinese and Spanish, as well as English, the authors argue against the PCC-based analysis of May, and propose a new analysis that relies on the Minimal Binding Requirement (MBR) (6) and the newly defined Scope Principle (7) (Aoun and Li 1993: 11).

(6) *The Minimal Binding Requirement (MBR)*
Variables must be bound by the most local potential antecedent (*A*'-binder).

(7) *The Scope Principle*
A quantifier *A* may have scope over a quantifier *B* iff *A* c-commands a member of the chain containing *B*.

In (8) the QP *everyone* is the most local binder for x_i—this satisfies the MBR. At the same time, the QP *everyone* does not qualify as a potential binder for the subject-*wh* trace x_j, since assignment of the index of *everyone* would result in a Principle C violation. There is no other potential intervening antecedent between *what* and the object trace x_j—the variable is then properly bound and the MBR is satisfied.

(8) What did everyone buy (for Max)? (Aoun and Li 1993: 58)

a. $[_{\text{what}_j}$ $[_{\text{IP}}$ $everyone_i$ $[_{\text{IP}}$ $[_{\text{NP}}$ x_i $]$ $[_{\text{I}'}$ $[_{\text{VP}_1}$ t_i $[_{\text{VP}_2}$ $[$ buy x_j $]]]]]]]$

According to the Scope Principle (7), both scopal readings are possible. Since *what* c-commands *everyone* and its variable, *what* has scope over the QP—a necessary configuration for a single answer (e.g. *Everyone bought coffee for Max*). For a PLA, we need a configuration where *everyone* c-commands the variable x_j within VP_2. We have this configuration for (8), so *everyone* can take scope over *what* and a pair-list reading is possible, making the question in (8) ambiguous.

Let us now see how the principles defined in (6) and (7) account for the lack of a PLA to a question with an object quantifier, such as (9) (Aoun and Li 1993: 61–62). Aoun and Li show that the quantifier *everything* can adjoin either to VP_2 (9a) or VP_1 (9b).

(9) Who bought everything?

a. [who$_i$ [x_i [$_{VP_1}$ t$_i$ [$_{VP_2}$ everything$_j$ [$_{VP_2}$ V x_j]]]]]

b. [who$_i$ [x_i [$_{VP_1}$ everything$_j$ [$_{VP_1}$ t$_i$ [$_{VP_2}$ V x_j]]]]]

The Scope Principle (7) predicts that the question in (9) should be ambiguous as well, since *everything* c-commands the object trace x_j. However, according to Aoun and Li (1993), the question in (9) is in fact non-ambiguous and only allows a configuration where *who* takes scope over *everything*. Hence, the authors stipulate that only the operators and intermediate traces (elements in *A*'-positions and not in θ-positions) are relevant for the determination of relative scope. Since the trace t_i in (9) does not count for the determination of scope, the only available configuration is the one in which *who* takes scope over *everything*. The wide scope reading of *everything* is lacking. Hence, a PLA is ruled out for questions, such as (9) with the quantifier in object position.

Chierchia (1993) also challenges May's analysis proposing a different mechanism to account for the absense of pair-list readings for questions with object quantifiers. Chierchia uses constraints on pronominal binding to explain why an object quantifier cannot take wide scope over the *wh-phrase*. *Wh*-phrases, he argues, are associated with two traces: a functional trace and an argument trace. The functional trace is bound by the *wh*-phrase that appears in Spec CP. The argument trace, co-indexed with an NP, acts like a pronominal element, and may be bound by the quantifier. If the binding is possible, the question has a pair-list reading. In (10a), for example, the binding allows extracting the information about the domain of a function which, in turn, provides pairings of people and those who love them (10b).[2]

(10) a. Who$_i$ does everyone$_j$ love t_i^j?

b. Mary loves John, and Sue loves Peter.

While binding is possible for questions with subject quantifiers, and thus a PLA is available, object quantifiers trying to bind the pronominal trace give rise to a Weak Crossover (WCO) effect—a general constraint on pronominal binding. WCO emerges when the movement of an element, here the *wh*-term, crosses over a pronominal trace, like in (11).

(11) Who$_i$ does his$_i$ mother love t_i?
His$_i$ mother loves every boy$_i$.

In questions, the quantifier fails to bind the pronominal trace left by *wh*-movement, like in (12). Thus, no PLA is possible for such questions.

---

[2]Chierchia derives pair-list answers from functional answers (see also Engdahl 1986; Déprez 1994b) In his view, when the QP binds the argument index on the trace, the bindings provides the domain of a function. It is then possible to spell out the function extensionally by listing its members, and eventually provide pairings of people from the domain and the range of the function $love\{x, y\}$. Such pairings constitute the pair-list answer. While some authors derive PLAs from functional answers, others argue for a separate treatment(Groenendijk and Stokhof 1984).We do not independently test whether these answers are available in the same syntactic environments.

(12) Who$_i$ t_i^j loves everyone$_j$?

Agüero-Bautista (2001) develops yet another account to capture the same subject-object asymmetry in the availability of PLAs. He appeals to the notion of reconstruction and argues that reconstruction of a *wh*-phrase below the quantifier is necessary for an inverse scope reading to obtain. While subject quantifiers can always scope over some reconstructed position of an object question, object quantifiers, which only rise as high as the *vP* domain, will only be able to scope over the lowest position of a reconstructed subject *wh*, i.e. its original θ-position. Agüero-Bautista further stipulates that only non-presuppositional *wh*-phrases, such as *who* but not *which NP* can reconstruct into their θ-position inside the *vP*. Hence, he predicts that PLAs are possible for questions with *who* interacting with an object *everything/everyone*, but not for *which*.

Moreover, Agüero-Bautista (2001) shows that PLAs are possible for questions with object quantifiers and degree-denoting *wh*-phrases, such as *How many NP* (13).

(13) How many students took every candidate out for dinner?
2 students, Danny Fox; 4 students, Norvin Richards;…

(Agüero-Bautista 2001:52)

In sum, Agüero-Bautista (2001) takes into account not only the structural position of the quantifier but also the semantics of the *wh*-phrase as well as the quantifier phrase.

While the analyses differ in the precise mechanisms they use to explain the inability of questions with object quantifiers to give rise to a PLA, they all appeal to a difference in structural position to predict that PLAs should be ruled out with a quantifier in object position. Yet these accounts start to diverge in their predictions once we pay attention to the type of the quantifier and the *wh*-phrase involved. For May (1985) and Chierchia (1993) both universal quantifiers, *every* and *each* are predicted to pattern alike since both are assumed to obey constraints on movement (PCC) and pronominal binding (WCO) in the same way. Agüero-Bautista (2001) goes further by proposing that different quantifiers can raise to different structural positions depending on the force of their distributivity, and consequently can have different scopal behavior. In the next section, the relations between quantifier scope and distributivity are discussed, with a focus on how distributivity ultimately affects the availability of a PLA.

2.2 *Quantifier Distributivity and Scope*

Beghelli (1997) proposed that the distributive properties of universal quantifiers can affect their syntactic behavior. To be more precise, he argued that strongly distributive quantifiers, such as *each*, are able to take wide scope even when they occur in object position since they target a position in the syntactic tree higher than where *every* can raise.

There are several syntactic environments where the differences between *every* and *each* in their ability to take inverse scope can be observed. One of them concerns the interaction with negation. In (14) both scopal readings are available: the first corresponds to the inverse scope (14a) and the second to the surface scope (14b).

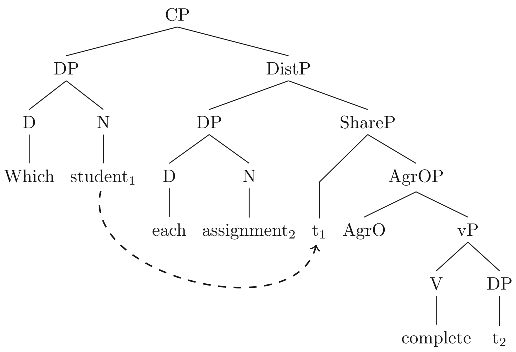
(14) Every horse didn't jump over the fence. (Musolino 1998)
  a. It is not the case that every horse jumped over the fence ('some reading').
  b. For every horse it is true that it did not jump over the fence ('none reading').

However, the pattern is different for *each*: only the surface scope (15b) is available and the inverse scope (15a) is not possible for (15). Beghelli and Stowell (1997) argue that *each* occupies a position higher than negation (*NegP*), so the inverse scope 'some reading' (15a) where negation takes scope over the quantifier phrase {negation » *each*} is unavailable.

(15) Each horse didn't jump over the fence.
  a. *It is not the case that each horse jumped over the fence. 'some reading'
  b. For each horse it is true that it did not jump over the fence. 'none reading'

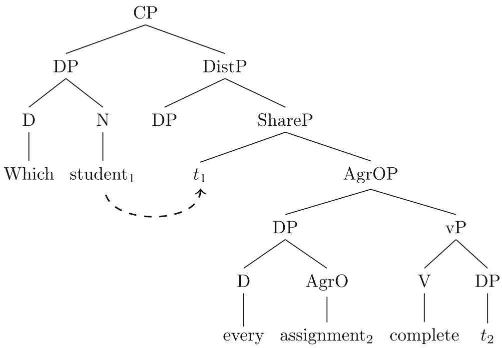
Beghelli and Stowell propose a hierarchy of syntactic positions where *DistP*—a position where strong distributive quantifiers raise – dominates *ShareP*—a projection where subjects get agreement. Let us now see what predictions this hierarchy makes for questions with object quantifiers. In (16) *each* is able to take scope over the reconstructed subject *wh*-phrase. Hence, a PLA is possible for (16).

(16) Which student completed each assignment?



In contrast, for Beghelli and Stowell (1997) the quantifier *every* cannot raise as high as *each* from an object position. Its highest possible landing site then is the *AgrOP*—a site for object agreement. As a consequence, the subject *wh*-phrase has no place to reconstruct below the quantifier, as in (17), and a PLA is not possible for (17).

(17) Which student completed every assignment?



In Beghelli's terms, the quantifier *each* is strongly distributive: it exhibits its ability to target the *DistP* as a raising cite in any syntactic environment. On the other hand, *every* needs special syntactic circumstances (being bound by existential closure) to raise as high, and hence it is pseudo-distributive. In that sense, we talk about the distributivity force of a quantifier. Tunstall (1998) formulates the difference between the two quantifiers by establishing requirements on the event structure that these quantifiers have. She proposes that *each* has a stricter requirement on the event distributivity structure—the differentiation requirement: the events must be differentiated on some dimension, such as time or space, for example.

Additional evidence for the role of distributivity force of quantifiers in determining the possibility of inverse scope comes from experimental studies. Brasoveanu and Dotlačil (2015) tested the ability of *each* and *every* to give rise to an inverse scope reading in declarative sentences. Using a binary choice and a self-paced reading tasks, the authors show that the quantifiers have distinct scopal behavior. They found that *each* increases the probability of inverse scope by approximately 17% in a binary choice task.[3] In a self-paced reading experiment, Brasoveanu and Dotlačil found that greater processing difficulty is associated with the inverse scope reading of the quantifier *every* as compared to *each* when a resultative predicate was present. The authors suggest that this behavior of *each* follows from its differentiation requirement (Tunstall 1998).

Taken together, the theoretical proposals of Beghelli (1997) and Tunstall (1998), as well as the experimental evidence presented in Brasoveanu and Dotlačil (2015) confirm that *each* has a higher ability to give rise to an inverse scope reading. In terms

[3] The authors use a binomial logit-mixed model to test the effect of quantifier type on the probability of inverse scope. They report a β coefficient of 0.7 on the logit scale. Using the algorithm described in Gelman and Hill (2007), Brasoveanu and Dotlačil (2015) suggest that in order to get an estimate on a more intuitive probability scale we need to divide the coefficient by 4, giving us around 17% increase in probability of inverse scope for *each* compared to *every*.

of *wh*-questions with quantifiers, we expect questions with *each* to allow PLAs more easily than questions with *every* regardless of the syntactic position of the quantifier.

2.3 Summary

Accounts that rely purely on syntactic constraints, such as May (1985), Aoun and Li (1993), and to some extent Chierchia (1993), differ from proposals that take into consideration the semantics of quantifiers (Beghelli 1997) or *wh*-phrases (Agüero-Bautista 2001) in the predictions they make for PLA availability. In this paper, we directly tested the predictions of these two families of accounts using experimental tools. We studied the acceptability of PLAs in different conditions with the following questions in mind:

1. Do questions with subject quantifiers allow PLAs more readily than questions with object quantifiers regardless of the quantifier type?
2. Does the quantifier type affect the availability of a PLA?
3. Can the quantifier type supersede differences due to quantifier position? In other words, does the subject-object asymmetry affect only questions with *every* but not with *each*?

Let us quickly review the predictions. If the distributivity force of quantifiers affects their scopal behavior, we expect that participants should find PLAs more acceptable in questions with *each* than in questions with *every*. Concerning the role of quantifier position in the acceptability of PLAs, the picture is more complex. If both quantifiers obey the same structural restrictions, as some earlier accounts suggested (May 1985; Aoun and Li 1989), a higher acceptability level for questions with subject quantifiers as compared to object quantifiers should be observed. However, if the quantifier semantics also matters, we should see an effect of quantifier position with *every* but not with *each* (Beghelli 1997; Agüero-Bautista 2001).

3 Experimental Data

We tested the acceptability of PLAs as a response to questions with quantifiers using a Likert-scale to assess the relative weight of structure and quantifier semantics. In a $2 \times 2 \times 2$ design we manipulated quantifier position in questions (subject or object quantifier), the type of the quantifier used (*every* versus *each*), and the type of answer (a single answer versus a PLA). If all universal quantifiers obey the structural constraints on movement (either in May's (1985) or Chierchia's (1993) perspective), we expect to see a lower acceptability for PLAs to questions with object quantifiers than to questions with subject quantifiers across the board. However, the semantic accounts (Beghelli 1997) entail that PLAs should be possible and acceptable for questions with *each* regardless of the quantifier position, while questions with *every* should show the subject-object asymmetry, and only allow PLAs when *every* is in subject position.

3.1 Method

Materials and procedure. Participants were randomly assigned to one of the four randomized item-lists. The experiment started with the presentation of three practice questions. The main test lasted approximately 15–20 min. Participants could take as long as they wanted to give their answers, but they were not allowed to return to a previous question and change their responses. Each trial consisted of a question and an answer to that question. The task was to determine whether that answer was a possible answer to the relevant question on a 1–7 scale (where 1 was 'definitely no' and 7 'definitely yes', other values not labeled). A sample question is given in (18).

(18) Which driver took everybody home last night?
Tom took Ms. Franko, Bob took Ms. Dombovski, and Jack took Mr. Perkins.

Participants rated 32 critical items (8 conditions, 4 items per condition) and 60 control/filler statements, which included answers to questions with *wh*-words only (19), quantifiers only (20), questions with clearly acceptable (21) or unacceptable answers (22), questions allowing PLAs (23), as well as questions with pragmatically odd answers (24).

(19) Which countries share a border with the US?
Canada and Mexico.

(20) Did each doctor get a license?
No, only 2 of them did.

(21) Which animal in this zoo is the tallest one?
The giraffe.

(22) Did you read every book on the list?
Yes, I read 3 out of 8.

(23) Who bought what?
Mary bought the cheese, Sue bought the milk, and Jim bought the potatoes.

(24) Which girls ate the cake?
Mary did.

The experiments were run using the Survey Monkey software (SurveyMonkey.com, LLC).

Participants. 29 undergraduate students, all native speakers of English, participated in this experiment. They received course credit for their participation.

3.2 Results and Discussion

We fitted cumulative link mixed models to compare the acceptability ratings for PLAs in different conditions. Our dependent measure was a rating on a 7-point scale, indicating how acceptable the subjects found a given answer. The fixed effects included quantifier position and quantifier type. We built in the maximal random effect structure that still converged (as per Barr et al. 2013). In our case these were random intercepts for subjects and items, as well as random slopes for subjects and quantifier position and quantifier type. The analysis revealed that overall, PLAs to questions with *each* were more acceptable than those to questions with *every* ($\beta = -1.385$, $SE = 0.334$, $p < 0.01$), confirming the predictions of Beghelli (1997), Tunstall (1998), and the experimental findings of Brasoveanu and Dotlačil (2015) (Fig. 1, error bars represent 95% confidence intervals).

We observed an unexpected result with respect to the structural factor: it is significant for both quantifiers ($\beta = 2.699$, $SE = 0.769$, $p < 0.01$), and the interaction between quantifier type and quantifier position is insignificant ($\beta = -0.78$, $SE = 0.867$, $p = 0.368$). We can directly compare the magnitude of the asymmetry for the two quantifiers using a Bayesian t-test. It returns a Bayes Factor of 5 in favor of a hypothesis that the difference between the ratings for PLAs to questions with subject versus object quantifiers is the same for *every* and *each*. A Bayes Factor of 5 corresponds to substantial evidence on Jeffreys (1961) scale, suggesting that the structural position of the quantifier plays a similar role in determining PLA availability for the two quantifiers types studied here.

Let us now consider what these results mean. Figure 1 reveals that there is an asymmetry with the quantifier *every*, as all the theoretical accounts predict. PLAs to questions with subject quantifier *every* are more acceptable than PLAs to questions with object-*every*. Second, overall PLAs to questions with *each* are more easily

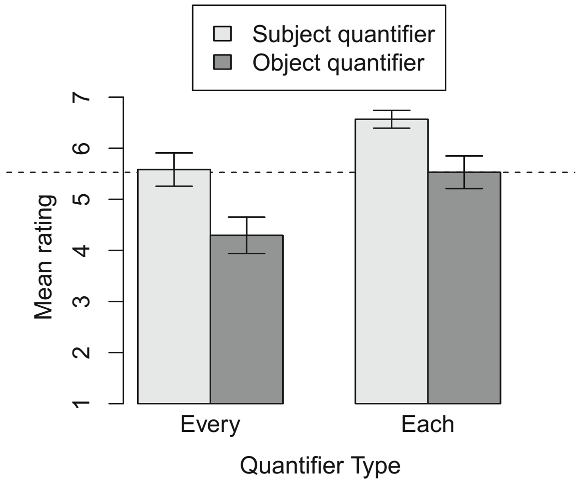


Fig. 1 Acceptability ratings for PLAs depending on quantifier position and quantifier type

accessible than PLAs to questions with *every*, as suggested by the analyses taking into account the distributive force of a quantifier (Beghelli 1997; Tunstall 1998).

We also see in Fig. 1 that PLAs to questions with object-*each* are less acceptable than PLAs to questions with subject-*each*. While such an asymmetry has several straightforward explanations when the quantifier is *every*, things are a lot more complicated for *each*. We cannot argue that PLAs to questions with object-*each* are ruled out due to a violation of some syntactic constraint. Otherwise, we would be forced to argue that such PLAs are not acceptable, and pair-list readings of questions with object-*each* unavailable. Both of these claims are problematic, since (a) we observe a rather high average rating for PLAs to questions with object-*each*, and (b) PLAs to questions with *subject-every* fall in the same category: the mean rating for PLAs to questions with object-*each* is 5.53 and subject-*every*—5.58 on a 7-point scale. In other words, if we declare anything below the 5.53 line as being ungrammatical, we would have to assume that the PLAs to questions with *every* are unavailable altogether regardless of the quantifier position. This is clearly wrong: both the naïve speakers' judgments and the theoretical predictions converge here: questions with at least subject-*every* can have pair-list readings and are indeed ambiguous between allowing a single or a pair-list answer. A question then arises: what is driving this difference between subjects and objects, if it is not a violation of a grammatical constraint?

In order to account for the subject-object asymmetry effect with *each*, we need an analysis that would (1) predict that PLAs to questions with object quantifiers are less acceptable than PLAs to questions with subject quantifiers; (2) would not completely rule out the possibility of a pair-list reading for a question with an object quantifier, since PLAs to questions with object-*each* are acceptable.

4 Nature of the Subject-Object Asymmetry in Scopal Interactions

The diverging scopal behavior of different universal quantifiers has long been known. However, it has remained unclear to what extent quantifier semantics can override structural considerations on the availability of PLAs. We tested the ability of *each* and *every* to give rise to pair-list readings in questions. While as predicted by Beghelli (1997), PLAs to questions with *each* are more acceptable than analogous answers to questions with *every*, we found an unexpected effect—structure still plays as much of a role for questions with the quantifier *each* as it does for questions with the quantifier *every*. To be more precise, we observe an asymmetry between the acceptability of PLAs depending on the structural position of the quantifier even for the quantifier *each*.

What is puzzling about the asymmetry observed with *each* is the possible source of such an asymmetry. Recall that the asymmetry with *every* has been attributed to a number of factors, including the fact that object-*every* is unable to form a proper syntactic configuration where it could take wide scope over an intermediate trace (Aoun and Li 1993), a reconstructed copy (Agüero-Bautista 2001), or form a $\sum$-sequence with the *wh*-phrase (May 1985). According to these analyses, the grammar rules out the {object-*every*≫*wh*-phrase} scopal configuration.

However, for *each*, a PLA is permitted from a structural perspective both when *each* occurs in subject and in object position. We can therefore argue that wherever the difference in acceptability comes from, it cannot stem from one configuration being grammatical and the other not. Reinhart (2006) offers a computational perspective to account for the difference between subjects and objects in their capacities to take inverse scope, she refers to this process as the 'scope-shift'. She assumes that the ease of a potential scope-shift depends on contextual and structural factors. Reinhart (2006) makes the following descriptive generalization: if a subject can be interpreted distributively "scope-shift of the object is not allowed" (Reinhart 2006:115). Since the scope-shift, or an inverse scope of the QP and the *wh*-phrase are required for a PLA to be available, the absense of a scope shift in the case of a subject *wh*-phrase and an object QP yields a low acceptability of such a PLA.

It is crucial, however, that other properties of the context can affect the likelyhood of the scope-shift. Reinhart views the topic-focus structure of a sentence as one of such factors. We find a similar idea in Krifka (2001). He suggests that subjects and objects differ in their ability to act as topics. Since subjects tend to take wide scope more easily than objects, PLAs are easier to obtain for subject-quantifier questions than for object-quantifier questions. Endriss (2009) develops Krifka's ideas and argues that only topic-phrases are able to take wide scope. In the case of questions with quantifiers, we would then say that a quantifier phrase has to be the topic of a question in order to take scope over the *wh*-phrase, and give rise to a PLA.

It is extremely difficult to define what precisely a topic is. Endriss (2009) distinguishes between two main components of topichood discussed in the literature: aboutness and familiarity. Endriss argues against including the familiarity part in the notion of topichood and follows Reinhart's (1981) definition of topichood as pragmatic aboutness. Reinhart (1981) discusses what it actually means to be about something. Since Reinhart focuses on declarative sentences, we will first lay out her view for declarative sentences, and later we will show how Eilam (2011) applies her theory to questions. Reinhart (1981) appeals to the notion of a context set to introduce the notion of aboutness. Following Stalnaker (1978), she defines a context set as a set of propositions that "we accept to be true at this point" (Reinhart 1981:78). During a conversation, interlocutors add new propositions to the context set. The crucial part of her analysis lies in the internal structure of the context set. For practical reasons, Reinhart suggests, it is unlikely that the context set is organized as a list of all the propositions in it. Rather, the context set is centered around some topic, just like a library catalogue, using her metaphor. We could then think of NP-topics as referential entries under which we organize propositions in the context set.

We now need to define what makes the topic and the focus of a *wh*-question. Following Lambrecht (1996) and Lambrecht and Michaelis (1998), Eilam (2011) suggests that questions have the information structure focus (normally the *wh*-phrase) and the ground (the rest of the question), just like declarative sentences do. Yet, unlike the focus in a declarative sentence, the *wh*-phrase does not contribute new information. Rather, the focus status of a *wh*-phrase is a biproduct of the pragmatic status of a question. Eilam suggests that the *wh*-phrase is the only candidate for question focus, since the rest of the question is typically given. He further explains that the topic of a question is an address where the information contributed by the focus of a question will be stored.

We could then hypothesize that the quantifier phrase can act as a question topic. Quantifiers are probably not ideal candidates for being a topic even in a declarative sentence, since they are not referential (Endriss 2009). However, Reinhart (1981) specifies that we could think of universally quantified *NP*s as denoting sets. In that sense, sentences containing such phrases can be understood as asserting something about the sets and their members. Topichood tests also confirm that universally quantified *NP*s can in principle be topics. For example, Endriss (2009) shows that these phrases can occur in German in the middle field position—a position where topics normally occur (Frey 2004).

Let us now turn to *wh*-questions with universal quantifiers, and see how the information structure account would explain the observed subject-object asymmetry. PLAs are normally available for questions with subject quantifiers, since subjects tend to be topics. Evidence for that generalization comes from the works of Li and Thompson (1976), Reinhart (1981), Lambrecht (1996), Erteschik-Shir (1997), and Krifka (2001).[4] Objects, on the other hand, are not prototypical topics (Krifka 2001), therefore questions with object quantifier phrases do not easily allow PLAs.

Krifka (2001) further observes that questions with focused quantifiers even in subject position cannot yield PLAs (25).

(25) a. Which dish did EVERYONE make? Krifka (2001, 24)
b. *Al the pasta, Bill the salad, and Carl the pudding.

In sum, subject and object quantifier phrases differ in their ability to act as topics—and consequently in their likelihood of taking wide scope over a *wh*-phrase and give rise to a PLA. We expect the effect of information structure to be the same for the quantifiers *every* and *each*—PLAs to questions with object quantifiers are expected to be less available than PLAs to questions with subject quantifiers.

The example in (25) also shows that the topichood status of a question is a more flexible notion compared to such a structural dichotomy as subject versus object. While a tendency exists for subjects but not objects to be topics, it seems possible to alter the default information structure of a question, making a PLA for an object-quantifier question available. In order to arrive at a pair-list interpretation of such a question, a speaker has to access a non-default topic/focus configuration—the one where an object phrase acts as a topic, and a *wh*-phrase as the focus of a question. Our data suggest that speakers differ in their ability to construct a context that could allow them to access such a non-default configuration.

[4] Lambrecht (1996) mentions that even though subjects are often found to be topics, the notions of topic and subject need not be conflated, as they do not always refer to the same individual in a sentence.

There is one more aspect of the data that requires an explanation, namely, the difference between *each* and *every*. It is possible that phrases headed by *each* and *every* differ in their capacity to act as topics due to their D-linking status. According to Gil (1991), *each* is anaphoric, while *every* is not required to be such. Gil also noticed that *every* but not *each* can occur in generic sentences. It appears that for *each* to be felicitous, it has to be D-linked, and in Reinhart's terms that means having its common noun set already in the context set.

Being accessible in discourse is what relates D-linking and topichood. The QP-phrase headed by *each* has a common noun set already present in discourse, therefore a possibility exists for it to act as the topic and take wide scope. At the same time, the QP headed by *every* is not necessarily D-linked, and the likelyhood of such a phrase to act as a topic if therefore lower (even though the latter is not impossible), and a PLA is harder to derive.

We can now turn to Reinhart (2006) and her account of phonosyntactic interface strategies. She discusses the connection between an element being de-stressed and it being D-linked (Pesetsky 1982). Recall, that Krifka (2001) observed that Focus-marked constituents cannot take inverse scope. De-accenting a QP, he showed, made the inverse scope more likely, supposedly because a de-accented constituent is given. Yet in another line of research being introduced in prior discourse, or rather being salient in prior discourse is an essential characteristics of a constituent that is given, rather than new (Schwarzschild 1999). Such constituents, Schwarzschild submits, are de-accented, and cannot be Focus-marked. As such, they seem to be candidates for being the topic of an utterance.

The requirement of being introduced in prior discourse, or to be D-linked, has also been discussed under the term 'specificity' in a number of papers including É. Kiss (1993); Kagan (2006), among others. É. Kiss (1993) looks at multiple *wh*-questions and argues that in order to take inverse scope, a *wh* phrase has to be specific. She further shows that her account extends to universal quantifiers as well. We could then say that QPs headed by *each* are inherently specific, while *every*-phrases are not necessarily such.

5 Conclusion

In this paper we looked at the interaction of *wh*-phrases and quantifiers, focusing on their ability to give rise to different scopal readings. We used experimental data to test whether the strong distributivity of *each* can make PLAs equally available regardless of its structural position in a question. Furthermore, we wanted to compare the behavior of *every* and *each* in their ability to give rise to PLAs depending on the structural position they occupy. The data revealed that while *each* facilitates the access to a pair-list reading compared to *every*, both quantifiers exhibit a subject-object asymmetry. Following Krifka (2001), Endriss (2009), and Eilam (2011), we proposed that it is the ability of a quantifier phrase to be construed as a topic that defines the likelihood of a PLA. Such an account simultaneously covers two facts: (1) higher acceptability of PLAs to questions with *each*; and (2) the subject-object asymmetry observed for both quantifiers. What is crucial, an information structure analysis of the subject-object asymmetry does not completely rule out PLAs to questions with object quantifiers. Rather, such PLAs may appear less likely given the difficulties constructing a context where an object quantifier phrase would act as a topic.

Acknowledgements We would like to thank the anonymous reviewers for their insightful comments and suggestions. We are also grateful for the valuable feedback we received at the Budapest Workshop on Linguistic and Cognitive Aspects of Quantification.

References

Agüero-Bautista, Calixto. 2001. *Cyclicity and the Scope of wh-Phrases*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Aoun, Joseph, and Yen-hui Audrey Li. 1989. Scope and Constituency. *Linguistic Inquiry* 20 (2): 141–172.

Aoun, Joseph, and Yen-hui Audrey Li. 1993. *Syntax of Scope*, vol. 21. Cambridge: MIT Press.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal. *Journal of Memory and Language* 68 (3): 255–278.

Beghelli, Filippo. 1997. The Syntax of Distributivity and Pair-List Readings. In *Ways of Scope Taking*, 349–408, ed. Anna Szabolcsi. Springer.

Beghelli, Filippo, and Tim Stowell. 1997. Distributivity and Negation: The Syntax of *Each* and *Every*. In *Ways of Scope Taking*, ed. Anna Szabolcsi, 71–107. Springer.

Brasoveanu, Adrian, and Jakub Dotlačil. 2015. Strategies for Scope Taking. *Natural Language Semantics* 23 (1): 1–19.

Chierchia, Gennaro. 1993. Questions with Quantifiers. *Natural Language Semantics* 1: 181–234.

Déprez, Viviane. 1991. Economy and the That-t Effect. In *Proceedings of the Western Conference on Linguistics*, vol. 4, 74–87. California State University.

Déprez, Viviane. 1994a. A Minimal Account of the That-t effect. In *Paths Toward Universal Grammar: Studies in Honor of Richard S. Kayne*, ed. G. Cinque, J. Koster, J.-Y. Pollock and R. Zanuttini, 121–135. Georgetown University Press.

Déprez, Viviane. 1994b. Questions with Floated Quantifiers. In *Proceedings of the VIth Conference on Semantics and Linguistic Theory (SALT)*, ed. Mandy Harvey and Lynn Santelmann, vol. 4, 96–113.

Eilam, Aviad. 2011. *Explorations in the Informational Component*. Philadelphia: University of Pennsylvania Dissertation.

Endriss, Cornelia. 2009. *Quantificational Topics: A Scopal Treatement of Exceptional Wide Scope Phenomena*, Studies in Linguistics and Philosophy, vol. 86. Springer.

Engdahl, Elisabet. 1986. *Constituent Questions*. Dordrecht: D. Riedel Publishing Company.

Erteschik-Shir, Nomi. 1997. *The Dynamics of Focus Structure*. Cambridge University Press.

É. Kiss, Katalin. 1993. Wh-movement and Specificity. *Natural Language & Linguistic Theory* 11 (1): 85–120.

Frey, Werner. 2004. A Medial Topic Position for German. *Linguistische Berichte* 198 (565): 154–190.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Hierarchical/Multilevel Models*. Cambridge: Cambridge University Press.

Gil, David. 1991. *Universal Quantifiers: A Typological Study. EUROTYP Working Papers*, vol. 7 12. Berlin: European Science Foundation, EUROTYP Programme.

Groenendijk, Jeroen A.G. and Martin J.B. Stokhof. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Amsterdam: University of Amsterdam dissertation.

Jeffreys, Harold. 1961. *Theory of Probability*. Oxford: Clarendon Press.

Kagan, Olga. 2006. Specificity as Speaker Identifiability. In *Proceedings of the Ninth Symposium on Logic and Language*, ed. Beáta Gyuris, László Kálmán, Chris Piñoń and Károly Varasdy, vol. 9, 82–89. Budapest.

Krifka, Manfred. 2001. Quantifying into Question Acts. *Natural Language Semantics* 9 (1): 1–40.

Lambrecht, Knud. 1996. *Information Structure and Sentence Form: A Theory of Topic, Focus, and the Mental Representations of Discourse Referents*, vol. 71. Cambridge: Cambridge University Press.

Lambrecht, Knud, and Laura A. Michaelis. 1998. Sentence Accent in Information Questions: Default and Projection. *Linguistics and Philosophy* 21 (5): 477–544.

Li, Charles N. & Sandra A. Thompson. 1976. Subject and Topic: A New Typology of Language. In *Subject and Topic*, ed. Charles N Li , 457–589. Academic Press.

May, Robert. 1985. *Logical Form: Its Structure and Derivation*. Cambridge, MA: MIT Press.

Musolino, Julien. 1998. *Universal Grammar and the Acquisition of Semantic Knowledge: An Experimental Investigation into the Acquisition of Quantifier-Negation Interaction in English*. College Park: University of Maryland dissertation.

Pesetsky, David. 1982. *Paths and Categories*. Cambridge: Massachusetts Institute of Technology dissertation.

Reinhart, Tanya. 1981. Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27 (1): 53–94.

Reinhart, Tanya. 2006. *Interface Strategies*, vol. 45. Linguistic Inquiry Monographs. Cambridge: MIT Press.

Schwarzschild, Roger. 1999. Givenness, Avoid f and Other Constraints on the Placement of Accent. *Natural Language Semantics* 7 (2): 141–177.

Stalnaker, Robert. 1978. Assertion. In *Pragmatics: Syntax and Semantics*, vol. 9, ed. P. Cole. New York: Academic Press.

Tunstall, Susanne. 1998. *The Interpretation of Quantifiers: Semantics and Processing*. Amherst: University of Massachusetts Amherst dissertation.

Children Know the Prosody-Semantic/Pragmatic Link: Experimental Evidence from Rise-Fall-Rise and Scope

Ayaka Sugawara, Martin Hackl, Irina Onoprienko and Ken Wexler

Abstract Children's comprehension of scope interaction has received much attention especially since Musolino's (Universal grammar and the acquisition of semantic knowledge: An experimental investigation into the acquisition of quantifier-negation interaction in English, 1998) observation of isomorphism. Many studies report that English-speaking children prefer to assign the surface scope interpretations/isomorphic readings to sentences with multiple quantifiers or a quantifier and a logical operator, especially when there is no contextual support prior to a target sentence which would make one of the readings felicitous. Our set of experiments investigates whether children are sensitive to prosodic cues for scope assignment comparing a falling contour (i.e. neutral contour) and the Rise-Fall-Rise (RFR) contour, which in the literature has been argued to lead to the inverse scope interpretation (Jackendoff in Semantic interpretation in generative grammar. MIT Press, Cambridge, 1972; Büring in Linguist Philos 20: 175–194, 1997; Constant in Linguist Philos 35: 407–442, 2012; Contrastive topics: Meanings and realizations, 2014; a.o.). The results show that children are keenly sensitive to the difference between the two intonational patterns, and that they strongly associate the RFR contour with the inverse scope interpretation just like adults do.

Keywords L1 acquisition · Quantifier · Scope · Prosody · Intonation Rise-Fall-Rise · Contrastive topic

A. Sugawara (✉)
Department of Humanities, Mie University, 1577 Kurimamachiya-cho, Tsu-city, Mie, Japan
e-mail: ayakasug@human.mie-u.ac.jp

M. Hackl · I. Onoprienko · K. Wexler
Department of Linguistics and Philosophy, MIT, 77 Massachusetts Avenue, 32-D808, Cambridge, MA 02139, USA
e-mail: hackl@mit.edu

K. Wexler
Department of Brain and Cognitive Sciences, MIT, 77 Massachusetts Avenue, 32-D808, Cambridge, MA 02139, USA
e-mail: wexler@mit.edu

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_3

1 Introduction

For about twenty years, it has been noted in the acquisition literature that English-speaking children have a strong preference for the surface scope reading/isomorphic interpretation for sentences which contain a universal quantifier in the subject position and negation, as in (1).

(1) Every horse didn't jump over the fence.

a. No horse jumped over the fence. (∀>¬, surface scope reading)
b. It is not the case that all horses jumped over the fence. (¬>∀, inverse scope reading)

The interpretation of the sentence in (1) is ambiguous between the two readings described in (1a) and (1b). While English-speaking adults do not have difficulty interpreting (1) to mean (1b), English-speaking children tend to interpret (1) to mean (1a). The literature reports that the preference is robust; children reject the sentence (1) as a description of a situation where a proper subset of the horses (e.g. two out of three) jumped over the fence, 75–93% of the time (Musolino 1998; see also comparable conditions in Musolino et al. 2000; Musolino and Lidz 2006; and in Viau et al. 2010; a.o.). On the other hand, it has also been noted that children at the same age do not have difficulty accessing the LF of "Not > All" in sentences such as (2). They successfully accept the sentence as a description of a situation where the Smurf bought a proper subset of the oranges (e.g. one out of three), 85% of the time (Musolino et al. 2000).

(2) The Smurf didn't buy every orange.

This fact clearly indicates that what children have difficulty with is not the interpretation of "¬>∀" itself, but some mechanism to derive the interpretation from a sentence such as in (1). What would the mechanism be? In English, a standard analysis given to explain why the "Not > All" LF is available for (1) is that the universal quantifier in the subject position can reconstruct/undergo Quantifier Lowering (QL) to a position lower than the negation at LF (In the reconstruction view, it is generally assumed to be [Spec, vP]). In cases such as (2), the "Not > All" LF is less complex to derive; negation c-commands the object universal quantifier in the overt syntax, and thus there is no reconstruction/QL required to get the "Not > All" interpretation.[1] Musolino and colleagues argue that the phenomenon that children have a strong preference for the surface scope reading can be characterized as the Isomorphism-by-default (IBD) hypothesis, i.e., children prefer interpretations that are isomorphic to surface syntactic relations between the operators.[2]

[1] On the assumption that the lowest interpretable position for a quantifier in object position is below sentential negation.

[2] One of the other hypotheses that is discussed in the literature is the Question-Answer-Requirement hypothesis discussed by Gualmini et al. (2008). See Bajaj et al. (2014) for discussion on how much and when Question-Answer Requirement affects children's interpretation. It is also possible to propose that children have difficulty with reconstruction/QL (Sugawara and Wexler 2014).

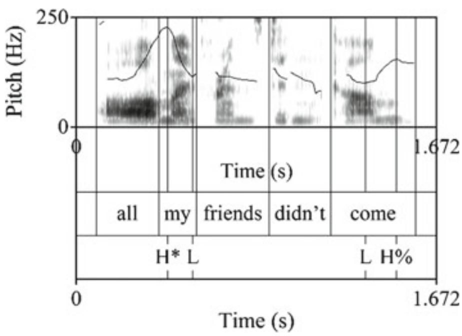


Fig. 1 Pitch track generated by Praat using the sound file in Constant (2012 (46))

More recently, research has revealed that children do access the inverse scope reading under certain circumstances. Specifically, when children's processing load is alleviated by experimental manipulations, they seem to be more readily capable of accessing the "Not > All" LF (Musolino and Lidz 2006; Viau et al. 2010 to be reviewed in Sect. 3.1). In the current work, we take a different approach. We ask whether other manipulations can help children access the inverse scope reading. We study the effect of prosody on scope interactions, specifically a certain prosodic contour dubbed as Rise-Fall-Rise (RFR, most clearly described in Constant 2012; see Fig. 1). Sentences with this contour have the "Not > All" reading as a preferred reading.

The mechanism describing how the specific contour affects the interpretation will be reviewed in Sect. 2. The purpose of the experiment reported in this chapter is to test whether children are sensitive to this prosodic manipulation and if they are, whether they will interpret the sentences like adults do.

The structure of this chapter is as follows. In Sect. 2, the mechanism of RFR is summarized. In Sect. 3, we will review some of the recent acquisition work that shows that children can access the inverse scope interpretation to a much higher extent than previously thought.[3] We will also review recent experiments that look into an effect of certain prosody on scope interaction with adult participants. Our current study is reported in Sect. 4, and the discussion and conclusions are in Sect. 5.

2 Rise-Fall-Rise (RFR) Prosody

Let us closely look at what it takes to be the prosody to convey the "Not > All" interpretaion. Jackendoff (1972) proposes that the pitch accent on the universal quantifier and falling at the end together correspond to the "All > Not" reading, and that the pitch accent on the universal quantifier and rising at the end correspond to the "Not > All" reading.

[3]For more extensive review of the literature, see Musolino (2011).

(3) a. ALL the men didn't go.$_{L\%}$[4] (∀>¬)
b. ALL the men didn't go…$_{L\text{-}H\%}$ (¬>∀)

The contour that is associated with (3b) and Fig. 1 has been variously referred to as 'B-accent', 'Fall-Rise' etc. in the literature. However, following Constant, we would like to call it a Rise-Fall-Rise (RFR) contour, since the name is theory-neutral and it transparently describes the pitch tracks of the contour. Figure 1 shows the pitch track of a sentence "All my friends didn't come" with the RFR contour.[5] H* indicates a high tone with a pitch accent, and H% means high phrasal boundary tone. As we can see in the pitch track, there is a sharp rise with a stress (L+H*) on the universal quantifier and a boundary tone that is rising (L-H%) at the end of the (intonational) phrase (InP), which in this case (and in all of our experimental sentences) is sentence-final. As we will see shortly, the high tone with a pitch accent aligns with the Contrastive Topic of the sentence. In order to rise at the end, the pitch must "fall" after the initial rise. The final rise is at the phrase boundary—the RFR occurs entirely (stretched, if necessary) within one intonational phrase.

Büring (1997, 2003) discusses the effect of the Contrastive Topic contour on scope relations in German and English.[6] Based on his analysis, Constant (2012, 2014) analyzes the semantic consequences of the RFR contour.[7] Following Constant (2012), we assume that the RFR contour is not compatible with alternative dispelling foci and disambiguates away from interpretations that do not have an alternative proposition which remains unresolved. That is, the characteristics of RFR are decomposed into these steps—(i) The Rise-Fall on the Contrastive Topic activates its relevant alternatives, (ii) Alternative propositions are generated based on the alternatives activated in (i), (iii) For surface-ambiguous sentences multiple possible LFs and their alternatives are computed, and (iv) An LF is deemed viable if it satisfies the condition, attributed to the final rise, that there is at least one proposition in the set of its alternatives that is not yet resolved/dispelled after accepting the proposition expressed by that LF.

To illustrate, the example in (4–5) shows why the sentence in (4-B) under the RFR-contour is infelicitous: when pronounced with RFR, the alternative set in (5) is generated. RFR further requires that there should be unresolved or not-dispelled LFs remaining post assertion. However, the proposition expressed by "All my friends

---

[4]L and H stand for low and high tones, respectively. Boundary tones are indicated using the symbol %, e.g., low boundary tone as L%.

[5]The audio recordings of the examples in Constant (2012) are available at http://semanticsarchive.net/Archive/jhmYTI5M/.

[6]See Bobaljik and Wurmbrand (2012) for a different account concerning the effect of intonation, which makes use of the notion that "the special intonation represents a TOPIC >> FOCUS accent (Bobaljik and Wurmbrand 2012: 401)." Their account explains how the Information Structure, together with the special intonation that affects the IS, makes QR or reconstruction possible in an environment where QR or reconstruction is otherwise disallowed. We thank a reviewer for directing our attention to this point.

[7]In Constant (2012), RFR and CT are treated separately though he suggests that proposing a unified account of RFR and CT might be possible. Constant (2014) does not use the term RFR and instead calls it "lone CT," which is not accompanied by a Focus. In the current study, let us keep using the term RFR to cover the union of the phenomena.

liked it" entails the truth of all the other LFs in the alternative set. Hence no alternative proposition is still open for discussion and this violates the condition imposed by the final rise of the RFR-contour.

(4) A: Did your friends like the movie?
B: *[ALL]$_{CT}$ my friends liked it… (w/RFR contour)

(5) {all my friends liked it, some of my friends liked it, …}

Consider, by contrast, the example in (6–7). The utterance in (6-B) with RFR generates the alternative set described in (7). Note that the other LFs in the alternative set remain unresolved after asserting "John liked it." As a result, B in (6) is felicitous with RFR. Moreover, the utterance of B with RFR implicates (based on Gricean reasoning) that the speaker does not know the information about the other friends or that (s)he deliberately avoids mentioning the information about the others at this moment.

(6) A: Did your friends like the movie?
B: [JOHN]$_{CT}$ liked it… (w/RFR contour)

(7) {John liked it, Mary liked it, Fred liked it, …}

Finally let us turn to the case of the scope interactions between the universal quantifier and negation. The sentence "All my friends didn't come" is surface-ambiguous between the "All > Not" reading and the "Not > All" reading. If one entertains the interpretation of "All > Not," the set of alternatives to the LF that expresses that reading looks like the one in (9). However, once the sentence is asserted, these alternatives are dispelled since they are entailed by the assertion. This does not meet the condition of RFR that at least one alternative must remain unresolved after assertion. On the other hand, the set of alternatives for the "Not > All" interpretation of the sentence will be the one in (10). These alternative propositions are not entailed by the assertion and therefore they remain unresolved, which satisfies the condition of RFR. In other words, since only the inverse scope LF for (8-B) satisfies the condition of having unresolved alternatives, the sentence is disambiguated in favor of the "Not > All" reading when pronounced with RFR.[8]

(8) A: Did your friends like the movie?
B: [ALL]$_{CT}$ my friends didn't like it… (w/RFR contour)

(9) $\forall > \neg$:	Assertion	For all friends of mine, they did not like it.
	Alternatives	{For all friends of mine, they did not like it., For some friends of mine, they didn't like it ..}
(10) $\neg > \forall$:	Assertion	It is not that all friends of mine liked it.
	Alternatives	{It is not that all friends of mine liked it., It is not that some friends of mine liked it, ..}

[8]Constant's account does not require there to be Focus in addition to Contrastive Topic in such sentences as B in (8). Büring's (1997) analysis on scope inversion in German can be understood in a similar manner but his analysis requires both Focus (on negation) and CT. For another approach where Focus does not have to be taken into account, see É Kiss and Gyuris (2003) for Hungarian scope inversion caused by a (fall-)rise intonation on CT. What attracts the (fall-)rise intonation in Hungarian is in the topic position, guaranteed by the discourse-configurational nature of Hungarian.

While the (Rise-)Fall-Rise contour and its impact on interpretation are fairly well understood, it is less widely discussed under what circumstances the contour is felicitously used; Given the nature of Contrastive Topic being a type of Topic, RFR is felicitous when the F-marked element (*ALL* in our examples) is already available, e.g. *given* in the Question under Discussion (QUD, see Roberts 1996/2012 and Buring 2003; a.o.). That is, the use of RFR is not felicitous in contexts like (11). Similarly, dialogues such as (12) where the CT is not *given* in the context, are less natural, while the conversation in (13) is perfectly natural, illustrating why RFR is most naturally uttered as an answer to a question raised by an interlocutor that contains the target of the CT accent in the RFR-answer.

(11) I am telling you how the party last night went. #ALL my friends didn't come… (with RFR)

(12) A: What happened at the party?
B: # ALL my friends didn't come… (with RFR)

(13) A: Hey, did (all) your friends come to the party?
B: ALL my friends didn't come… (with RFR)

To sum up the section, RFR is a general phenomenon in English that is not only found with quantified expressions (even though the interactions between RFR and scope relations have attracted most of the attention), and that the mechanism of interpreting sentences with a RFR contour requires that there be alternative proposition that remain unresolved at the post-assertion stage.

The point of the current study is to see whether children can access the inverse scope "Not > All" interpretation when read with the RFR contour, in a context where RFR is felicitously licensed.

3 Previous Studies

3.1 Children Can Access the Inverse Scope Reading

Musolino and Lidz (2006) hypothesize that children do not interpret the sentence in (1) to mean "Not > All" because the children are somehow not ready for processing negative sentences under the "Not > All" reading out of the blue. Their Experiment 1 tested the set of conditions exemplified in (14). Condition 1 replicates the results reported in the previous literature and serves as the baseline for Condition 2. In Condition 2, an affirmative sentence precedes the target sentence, "in order to familiarize children with the intended domain of quantification (ibid: 825)."

(14) a. **Condition1**: Every horse didn't jump over the fence.
b. **Condition2**: Every horse jumped over the log, but every horse didn't jump over the fence.[9]

[9]They also tested the sentences that employed *and* instead of *but*. The results did not differ from each other.

They tested 10 five-year-old children with Condition 1 (mean age 5;7) and another 10 five-year-olds with Condition 2 (mean 5;2). The experiment is carried out as a Truth-Value Judgment task (Crain and Thornton 1998). All of the target sentences in the experiment had a universal quantifier *every NP* in the subject position followed by negation. The stories were constructed so that a proper subset of the objects have completed the action, i.e., "$\neg > \forall \wedge \exists$" is made true in the story. The authors take the response "yes" to mean that the children access the "Not > All" interpretation, and the response "no" to mean that they access the "All > Not" interpretation. In Condition 1, children's "yes" rate was 15%, whereas in Condition 2, children's "yes" rate was 60%, which is significantly higher than in Condition 1. These results show that the children are, in fact, able to access the "Not > All" interpretation even when it is not isomorphic. They argue that, together with the results from their Experiment 2, what is lacking in children's linguistic system regarding this phenomenon is pragmatic abilities, and that under certain contextual manipulations such as using an affirmative sentence that would make the following negative sentence more felicitous, children's ability to accommodate the pragmatic considerations could be boosted. The idea is that "not p" is felicitous in a context where the possibility of p has been raised in the previous context.

Viau et al.'s (2010) Experiment 3 tests whether children could be "primed" by a certain interpretation. The two conditions in the experiment are illustrated in (15). Condition 1 serves as the baseline condition, where the six target sentences in the session have the same structure; a universal quantifier precedes the negation. In Condition 2, the target sentences in the first half of the session have "not" preceding the universal quantifier, and the target sentences in the last half of the session have the same structure as the sentences in the other condition. Since the isomorphic interpretation of the sentences in the first half is "Not > All," the first interpretation people will get with such a sentence is "Not > All."[10] That is, the subjects will be primed to access the LF of "Not > All" in the second half by the interpretation of the sentences in the first half of the session.

(15) a. **Condition 1**: [trials 1 through 6] Every horse didn't jump over the pig.
b. **Condition 2**: [trials 1 through 3] Not every horse jumped over the pig.
[trials 4 through 6] Every horse didn't jump over the pig.[11]

[10]The "All > Not" interpretation is not contradictory to the sentence, although Viau et al. refers to this type of sentences as unambiguous test sentence. The "All > Not" interpretation is generally excluded because of the implicature computations (often called "indirect" scalar implicature).

[11]The authors do not discuss the issue of prosody. One wonders whether it is possible that in certain conditions the readers used RFR prosody, which is natural for a "Not > All" interpretation. Were the equivalent sentences in Condition 1 and Condition 2 (e.g. *Every horse didn't jump over the pig*) presented as exactly the same auditory taper, for example splicing in the tape from Condition 1 into Condition 2, to insure that prosody remained constant? Is it possible that some of the readers unconsciously used more RFR prosody in trials 4 through 6 in Condition 2 than in the trials in Condition 1? This would be natural given that the first 3 trials of Condition 2 involved *not every*, and this is quite natural to read with special prosody related to RFR. It would be really good to control prosodic effects in these studies, as we do in our experiment in this paper.

They tested 12 four-year-old children with Condition 1 and another 12 four-year-olds with Condition 2 on a Truth-Value Judgment Task (mean age of the 24 children = 4;6). The relevant stories support the situation where "$\neg > \forall \wedge \exists$" is made true, similarly to the experiment in Musolino and Lidz (2006). Therefore, children's "yes" responses indicate that they access the "Not > All" interpretation, and their "no" responses indicate that they access the "All > Not" interpretation. Viau et al. found in the baseline condition the "yes" rate for the first half to be 20% and for the second half to be 40%. On the other hand, the priming condition shows a higher rate of "yes" responses, over 80% both in the first half and in the last half of the session. Viau et al. report a significant difference between "yes" rates in the baseline condition and in the primed condition for the second half of items and conclude from this that children are capable of accessing the inverse scope "Not > All" interpretation, but that accessing it requires a bigger processing effort, which can, as shown in their experiment, be alleviated via priming of the relevant reading. In short, Viau et al. (2010) argue that children can access the non-isomorphic interpretation when the processing load is lessened.[12]

There is an alternative explanation, however, taking account of our observation in footnote 10 that the "All > Not" interpretation is actually not contradictory to the first sentence used in Condition 2: *Not every horse jumped over the pig*. It could be that no horse jumped over the pig. However this interpretation is in a typical contextual situation ruled out by a scalar implicature. One analysis might be that *not every* has *no* as an alternative. If *no* replaces *not every*, then the derived alternative *no horse jumped over the pig* implies the original sentence *not every horse jumped over the pig*. By the usual rules for deriving scalar implicatures, the implying (stronger) sentence is negated and conjoined to the original sentence. Thus *Not (No horse jumped over the pig)* is conjoined to *not every horse jumped over the pig*. This conjunction implies that there exists a horse that jumped over the pig. Under this implicature, "yes" is the only possible answer to the first 3 sentences of Condition 2. Note that this explanation assumes that children can compute implicatures, in conflict with much literature.

If the children do not compute the implicature in some cases, then *every horse didn't jump over the pig* (that is, the standard sentence type tested) is consistent with both a *yes* and a *no* response. The implicit assumption is that when a sentence is ambiguous for a child, that is, when either *yes* or *no* gives a true answer (one for each possible reading), then the child prefers to say *yes*.[13] However, this assumption

[12] It seems reasonable to us to explain the results of Musolino and Lidz (2006) along similar lines—rather than "introducing the relevant domain of quantification" their manipulation of having an affirmative sentence of the same form prior to the target sentence may make a Yes/No question of the sort *Did every horse VP*? salient. If, furthermore, questions are understood in terms of their answers, the negative answer to such a question *It is not the case that every horse VPed* could serve as prime for the non-isomorphic reading of the target sentence thus removing some of the processing load attached to inverse scope LF.

[13] See Meyer and Sauerland (2009) for discussion of the Principle of Charity, based on a careful examination of German sentences with focus particles. They propose the constraint called Truth Dominance to account for the bias to judge an ambiguous sentence as true.

seems to be tenuous as it holds only in some cases of the TVJT. See Crain and Wexler (1999) for discussion.

A related methodological issue is that we do not know how adults perform in these tasks. Viau et al. did not do any studies of adults. Since the results are probably affected by scalar implicature calculation, as we have just discussed, we do not really know what to expect by a fully competent speaker; calculation of implicatures is quite variable in a quantitative sense. Do speakers treat the different sentences (*Not every* vs. the "Not>All" reading of *every NP didn't*) differently when it comes to calculating the implicature?

The methodological conclusion is obvious: we learn more from experiments on these issues in which adults are tested as a separate group. We need a quantitative assessment of how adults behave in order to evaluate results from children. This is a lesson that has been learned in the study of scalar implicatures in general, and clearly applies to studies of quantifier scope. In the experiment reported in this paper, we include a separate adult group so that we can compare the children to such a group.

3.2 Adults Do Not Employ But Do Hear Different Contours

Since Jespersen (1933), it has been noted that different prosodic contours (can) indicate different scope readings.[14] For such sentences where a universal quantifier and negation interact, Jackendoff (1972) points out that distinct prosodic patterns correspond to different scope relations. As we have seen in detail in Sect. 2, when a sentence is read with a sharp rise and fall on the universal quantifier and a sentence-final rise, the most salient interpretation of the sentence is "Not>All." Although there is a well-established literature on this phenomenon using off-line native speaker intuitions (Ladd 1980; Horn 1989, 2005; Krifka 1998; Büring 1997, 2003; Constant 2012, 2014 among many others; Gussenhoven 1983 and Ward and Hirschberg 1985 for counterarguments), there is little experimental work investigating under what conditions speakers use the RFR-contour and what effects it has on comprehension.

One study that looks into parents' child-directed speech is by McMahon et al. (2004). They report that parents do not differentiate prosodic patterns when reading to children. They also report that their adult participants showed no effect of intended interpretation [with distinct prosodic cues] in a comprehension study.[15]

(i) Truth Dominance: Whenever an ambiguous sentence S is true in a situation on its most accessible reading, we must judge sentence S to be true in that situation. (Meyer and Sauerland 2009: 140)

[14] Jespersen pointed out that sentences with negation and a *because*-clause can have different scope relations depending on the prosody; *call* in (i) has a rising tone and *call* in (ii) has a falling tone.

(i) I didn't call because I wanted to see her (but I called her for some other reason).
(ii) I didn't call because I wanted to avoid her. (Jespersen 1933: 299)

[15] For more experimental studies to investigate the relations between contrastive prosody and interpretations, (see McDaniel and Maxfield (1992); Baauw et al. (2004); Braun (2006); Calhoun (2006).

Syrett et al. (2014b) report an adult comprehension study where their participants could access somewhat different scope interpretations when they are tested with a comprehension experiment. Their experimental items include 48 sentences. Among them 24 sentences are target items and 24 items are control items.[16]

The experimental procedure is as follows. A participant first sees the target sentence on a computer screen (e.g., "All the moms didn't allow eyeliner"). Then an auditory version of the sentence is played in one of two conditions, which were recorded as a part of a separate production experiment (Syrett et al. 2014a). In the production experiment, the target sentences were embedded in a story, and the intended meaning (e.g., whether "All > Not" or "Not > All") was cued by context. The comprehension experiment employed these two types of recordings, i.e., (i) recordings of naïve native speakers intended to convey "All > Not" and (ii) recordings intended to convey "Not > All." Then the participant sees a set of continuations which exemplify distinct scope relations such as in (16). Finally the participant is prompted to decide on which continuation fits better as an interpretation of the target sentence.

(16) All the moms didn't allow eyeliner
Continuation 1: They were all in agreement.
Continuation 2: Only the moms of older girls let their daughters wear it.

Syrett et al. report that when the sentence was read with the prosody that was uttered to convey the "All > Not" reading, the rate of choosing the intended continuation was 63.9%. When the sentence was read with the prosody to convey the "Not > All" reading, the rate of choosing the intended continuation was 66.2%.[17] These response rates were significantly above chance, though clearly did not indicate full disambiguation by contour. Based on these results, the study argues that comprehenders can exploit the difference in prosodic contours that are produced with the intention to convey different scopal interpretations.

This is a promising result suggesting that naïve adult English speakers arrive at the interpretations specified solely by intonation as the theoretical literature has maintained. However, given that the average correct response rates were only barely

---

Part of McMahon et al. (2004) study looks into whether parents use contrastive prosody with ambiguous pronouns, and they show that the parents use suprasegmental cues to disambiguate the referents of pronouns.

[16]The target items consist of two kinds of sentences with *all … not* read with two different prosodic contours ($2*2=4$ items) read by four different speakers ($4*4=16$ items), and of one kind of sentence with *many/most … not* read with two different prosodic contours ($1*2=2$ items) read by four different speakers ($2*4=8$ items). This adds up to ($16+8=$) 24 items. The auditory stimulus was presented three times to the participants. It follows that one participant hears the same sentence 24 times in one session, three repeated by four different speakers and two different prosodic patterns for each speaker.

[17]There was variation in the rates of participants' correct responses. For example, when speaker 1 utters the sentences with the prosody to convey the "Not > All" reading, the participants access the correct interpretation 83.0% of the time, while speaker 4 utters the sentences with the prosody to convey the "Not > All" reading, the correct response rate was 52.3%. These differences appear to be larger than the differences between interpretations for the two different prosodies overall.

above 60%, the effect appears weak and one might wonder whether experiment-specific and/or task-related factors prevented the connection between prosody and meaning from coming out more strongly. As we will see in Sect. 4, our experimental results show that the prosody-meaning relationship is indeed stronger than indicated by Syrett et al.'s results even for naïve speakers of English.

As mentioned above, the auditory stimuli were recorded as a part of a production experiment (Syrett et al. 2014a). That is, the speakers were not trained for the task nor were they instructed to produce those sentences for others to judge. Though the intended meaning (e.g., whether "All > Not" or "Not > All") was clearly cued by context, there is no guarantee that the speakers were actually attempting to produce a coherent utterance within a trial. In other words, they were not expected to read the sentences aloud so that others who are blind to the context governing the interpretations would get the intended interpretations. It is quite possible that this could have resulted in noisier, less crisp productions of intonation contours making them less reliable cues for comprehenders to disambiguate. Stimuli produced by a trained speaker, on the other hand, who are given specific instructions about the relevant contours might yield cleaner signals and thus may make it easier to isolate the consequences of a particular prosodic pattern for naïve comprehenders.[18]

Besides ensuring crisp production of the RFR contour there is also second factor, felicity in context, that should be considered when designing an experimental environment for investigating the extent to which naïve speakers are sensitive to prosodic contours when producing and comprehending scopally ambiguous sentences. Specifically, the prosody most strongly associated with the "Not > All" interpretation—"sharp rise and fall on the universal quantifier and rise at the end"—is most felicitous in a context where the number of individuals who VP*ed* is the larger question under discussion (e.g. How many moms allowed eyeliner?), as discussed in Sect. 2.[19] Recall, more specifically, from Büring (1997, 2003), Constant (2012, 2014) and also from the examples in Jackendoff (1972), that the "sharp rise and fall" part is taken to be an indicator of Contrastive Topic, which, in turn, is felicitous when it is *given* in the context and contrasted with some previous utterance, typically made by another interlocutor. The effect of the contrastive topic accent can be described as indicating that the larger question under discussion (*How many moms allowed eyeliner*?) is narrowed to a yes/no QUD, *Did all the moms allow eyeliner*? This is illustrated in (17), which has the two possible answers *Yes, all the moms allowed eyeliner* and *No, it is not the case that all the moms allowed eyeliner*. For a context that fails to provide this type of environment consider (18) taken from Syrett et al. in their production experiment and intended to elicit the RFR contour (Syrett et al. 2014a).

[18] There are two points of departure regarding this issue: trained versus naïve speakers, and instruction to pronounce the sentence with the intent to disambiguate via contour versus no such instruction. Which factor is more important is an important question, but we will leave it open.

[19] It is so if it concerns sentences where a quantified phrase is the Contrastive Topic of a sentence (Noah Constant, p.c.), just like the sentences we are entertaining. This is not necessarily the case with all sentences with Contrastive Topic.

(17) Q: So, what happened to the girls at the school dance party after all? Did their moms allow them to wear eyeliner?
A: Well, ALL the moms didn't allow eyeliner… Only the moms of the older girls did.

(18) *An example of their experimental item for* "*Not>All*":
Several moms were helping their daughters get ready for the upcoming school dance. This is a progressive school, and moms are usually lenient about certain things, so even the younger girls thought their moms would approve of eyeliner. But at the dance only the older girls were wearing it. All the moms didn't allow eyeliner. Only the moms of the older girls let their daughters wear it.

In the scenario in (18), it seems to us that using RFR is actually not felicitous because the target sentence describes an entailment of the previous sentences (*Only the older girls were wearing eyeliner.)* and is not easily understood as contrasting with different options. Since the option of "All > Not" is already ruled out, the current QUD for the target sentence *All the moms didn't allow eyeliner* is thus not a yes/no question. Hence, the context does not fit the circumstances where RFR is most naturally licensed and so may have lead to speakers producing their utterances with less crisp intonation contours. Mindful of these considerations, our experiment will employ a dialogue format that will provide an appropriate QUD similar to the one in (17). This is a crucial difference in our experiment.

3.3 Testing Whether Children Know Prosody Interacts with Scope

One of the earliest studies of children's command of sentence level intonation and in fact, to our knowledge, the only study that attempts to directly assess children's understanding of the impact of the RFR contour on interpretation is Ianucci and Dodd (1980). They conducted a picture-selection experiment with children ranged from K, Grades 2, 4, to 7 and with adults to test whether the difference in prosody leads to different interpretations.[20] Their experimental stimuli consisted of five sentences, each exemplifying different kinds of sentence types with quantifiers. The conditions that are relevant for our purpose are as follows.

(19) a. **Condition 1**: All the rabbits aren't in the cages. (Stress on *all*, rise at the end—i.e., "Rise-Fall-Rise")
b. **Condition 2**: All the rabbits aren't in the cages. (Fall at the end)
c. **Condition 3**: Not all the rabbits are in the cages.

The participants were told to help the experimenter create a picture book by selecting a picture that would go well with the stories. Their choice was between two pictures, one of which illustrates the "All > Not" situation, and the other the

[20]Thanks to Thomas Hun-tak Lee for referring us to this very relevant paper.

Table 1 Part of results from Ianucci and Dodd (1980)

Groups	Condition 1	Condition 2	Condition 3
	Rates of Not > All (%)	Rates of Not > All (%)	Rates of Not > All (%)
K	38	27	55
Grade 2	47	19	79
Grade 4	46	17	83
Grade 7	58	18	100
Adults	96	18	100

"Not > All (& Some)" situation. They tested between 15 and 22 subjects per group. Their results are summarized in Table 1.

Children (and of course, adults) behave differently to Condition 1 and Condition 2. When the participants heard the prosody with falling tone at the end (Condition 2), they chose the "All > Not" pictures most of the time. On the other hand, when the participants heard the Rise-Fall-Rise prosody (Condition 1), adults chose the "Not > All" pictures most of the time, and the rates by children are between 38 and 58%. Given that the choices were among two options, the children chose the "All > Not" pictures 62–42% of the time. Though Ianucci and Dodd (1980) do not provide statistical analyses, the rates of choosing the "All > Not" pictures on Condition 1 seem to be consistently lower than that of Condition 2. It follows that the children regard the two kinds of prosody differently, although the connection between the Rise-Fall-Rise contour and the "Not > All" interpretation does not seem to be reliably built yet. It is also worth noting that in their experiment the rates of choosing the "Not > All" pictures in Condition 3 for preschoolers were as low as 55%. This shows that even when the isomorphic interpretation is "Not > All", the preschoolers in this experiment did not particularly prefer to choose "Not > All" pictures. As we have seen in Sect. 3.1, the children in Viau et al.'s experiment accepted the "Not > All" interpretation for sentences like in (8c) more than 80% of the time. Compared to that, it is possible that the children in Ianucci and Dodd's experiment had a stronger preference for the "All > Not" interpretation or that some experimental confound had prevented the children from accessing the "Not > All" interpretation when adults do so.

The study by Ianucci and Dodd is suggestive in that it shows that children regard the two types of prosody differently, and the preferences in interpretation are in the same direction as adults; with the Rise-Fall-Rise prosody, more "Not > All" interpretations are elicited. Moreover, the study obtains clear results for adults, strongly confirming linguists' intuitions regarding the effects of prosody. Our current study aims to follow-up on their study, with a more standardized procedure such as a wider variety of items, using recorded sound files by a trained speaker, and employing a dialogue which makes the prosody felicitous.

4 The Current Study

This study asks whether children know the effect of RFR on the scope assignment with universal quantifier and negation. This question is decomposed into two hypotheses.

(20) **Hypothesis 1**: If children are sensitive to the difference in contour, we expect different kinds of reactions to different contours (i.e., RFR contour and a neutral contour, to be discussed).
Hypothesis2: If children compute the effect of RFR in the same way as adults do, the rates of the "Not > All" reading will increase with the RFR contour.

The hypotheses might sound redundant, but hypothesis 1 is worth asking, given that the acquisition literature on contrastive stress has seen mixed results and it is a controversial topic whether children are sensitive to the difference in prosodic information (whether suprasegmental or not). It seems that results from previous studies using offline measurements support insensitivity to contrastive stress, while the results from studies using online method support sensitivity. For example, the pre-test experiments in McDaniel and Maxfield (1992) show that English-speaking children do not behave adult-like in an act-out task with sentences with contrastive stress.[21] Baauw et al. (2004) tested Spanish-speaking children and found that the rate of children's adult-like performance on the comprehension of the sentences with contrastive stress was around 44%, which was significantly lower than non-stressed sentences. On the other hand, using an eye-tracking method, Arnold (2008) report that four- and five-year-old English-speaking children showed the bias toward *given* objects with deaccentuation. Ito et al. (2012) also use an eye-tracking method and report that six-year-old Japanese-speaking children were able to make use of the contrastive stress to find the designated picture.

Though hypothesis 2 mentions adult reactions to RFR, there is still relatively little experimental evidence supporting the argument that (naïve) adult speakers will compute the effect of RFR and thus get the "Not > All" reading. The current study aims to add to the existing body of evidence (reviewed in previous sections) to assess the validity of claims in the theoretical literature about the effect of RFR.

[21] To be more specific, McDaniel and Maxfield (1992) tested children ranging from three to six years old, and examined children's sensitivity to contrastive stress with 10-scale scores. The average scores for each age group were the following: 4.7 for three year olds, 4.7 for four year olds, 5.8 for five year olds, and 6.7 for six year olds. McDaniel and Maxfield do not report how the individual data distributed, but it seems reasonable that three- and four-year-old children exhibit the chance performance, and the children in their experiment acquire some sensitivity to contrastive stress sometime during the time of five years old.

4.1 Method

We conducted an elaborated version of the picture-selection task, where the picture-selecting phase is preceded by a short conversation between two people. We employed a dialogue in the stories since the RFR contour is most naturally elicited as an answer to a question. Figure 2 illustrates the experimental procedure for an item. The experiment was conducted using a PowerPoint slides on a laptop computer. The sound files were pre-recorded by a trained speaker and played to the subjects as the experimenter clicks to advance the story.[22] The first phase is an introduction of the story. The girl mentions objects in the story, and the boy says he is wondering what is going to happen to the objects. The second phase of the story shows four pictures on the computer screen. The four pictures describe situations where (i) All > Not (None of the objects VP-ed), (ii) Not > All (Some but not all of the objects VP-ed), (iii) All did (All of the objects VP-ed), and (iv) irrelevant situations are depicted. The positions of the pictures were randomized across trials. Shortly after the second phase is shown, the boy asks the girl a question to prompt the answer, and the girl utters a target sentence with either of the prosodic contours. The subjects were asked to point to the picture that they think she is in or she is talking about.[23] The timing of presenting stimuli was controlled manually so we could accommodate a possible request (especially from children) to repeat a trial.[24]

We crossed question type (Baseline-Question vs. Did-All-Question) and contour (Falling vs. RFR) with contour a between-subjects factor.[25] As we discussed in Sect. 2, we assume that the appropriate QUD (Did-All-Question) would make the utterance with RFR more felicitous than the question that does not license the RFR utterance (Baseline-Question). This assumption led to the two levels of the question-

[22]The female voice was recorded by the third author, an undergraduate research assistant familiar with the literature. The male voice was recorded by another undergraduate research assistant. Both of the speakers are native speakers of English. The recording took place in a sound proof booth.

[23]We asked participants to make choices, rather than leaving an alternative to answer "I don't know." A reviewer points out that it might have been interesting to provide this option as well. However, we suspect that while the choice "I don't know" could tell a lot if adults were tested, leaving the option available with child experiments would raise more complications because interpreting the "I don't know" answer is less straightforward (e.g., it could be due to parsing difficulty, the lack of grammar, being unwilling to answer, suggesting ambiguity, etc.).

[24]Since the experiment was run using an offline method such as a picture-selection task, we did not obtain reaction time data, which would have been interesting to analyze, as one of the reviewers pointed out. It is possible to manually collect timing data from the audio recordings of the sessions, but this is left for future research.

[25]We made contour a between-subjects factor after a pilot experiment that varied contour within subjects, produced an order effect suggesting that the effect of prosody carried over between sentences with different contours. Since this pilot study employed a different method from the current study (TVJT), it is not transparently comparable with the current study. The summary of the 6 children we tested is as follows: 3 were isomorphic for the most part, 2 were flexible (i.e., accepting both "Not > All" and "All > Not" situations most of the time), and 1 accepted "Not > All" and rejected "All > Not" for the RFR condition while accepted "All > Not" and rejected "Not > All" for the Falling condition.

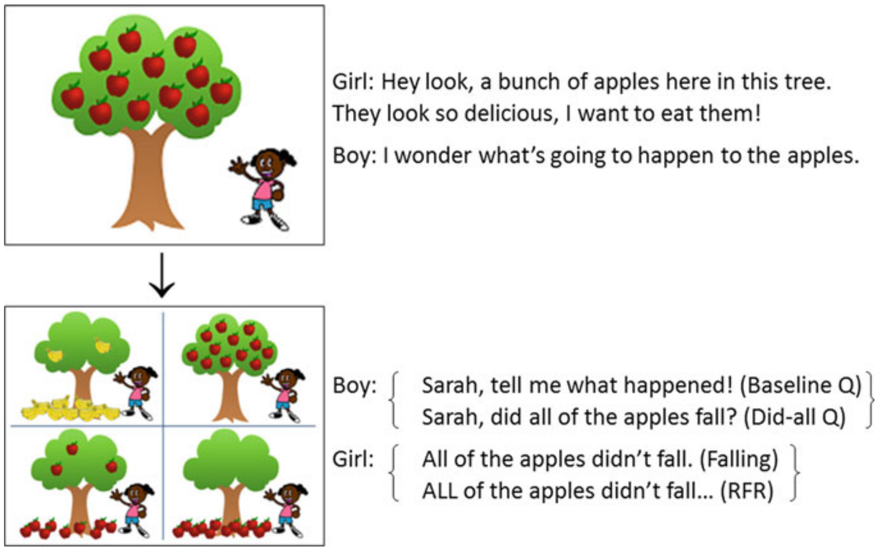


Fig. 2 An example of the experimental procedure

type factor, and the expectations were that with the Did-All-Question condition the implicature calculation with RFR is more easily done, which would elicit more "Not > All" responses compared to the Baseline-Question condition. Figure 3 shows the minimal pair (Falling contour and RFR contour) of the pitch tracks from one of the experimental items. The experiment consists of 8 target items (4 with Baseline-Q and 4 with Did-All-Q) as well as 6 filler items. The list of the predicates used in the target items can be found in (21). The verbs we used are all intransitive and frequent in child corpus. The sentences in the filler trials were non-negated sentences, and two of the examples can be found in (22).

(21) The list of the predicates used in target sentences
break, dry, fall, fly, grow, open

(22) Two examples of filler items

a. All of the glasses broke. (basic simple past)
b. All of the helicopters did fly. (emphatic past)

Since the filler items are affirmative sentences, the participants have to pay attention to positive pictures as well as the "Not > All" and/or the "All > Not" pictures during the session. Moreover, three of the six filler items had "did" as in (22b). Because of this manipulation, the participants could not jump to an early conclusion upon hearing *did*, and instead they had to pay attention to the end of the sentence, to see whether a negation followed or not. The presentation order of the items was pseudo-randomized. We created two different sets of orders, and each participant was assigned randomly to either order of the items. When a child wanted to listen to

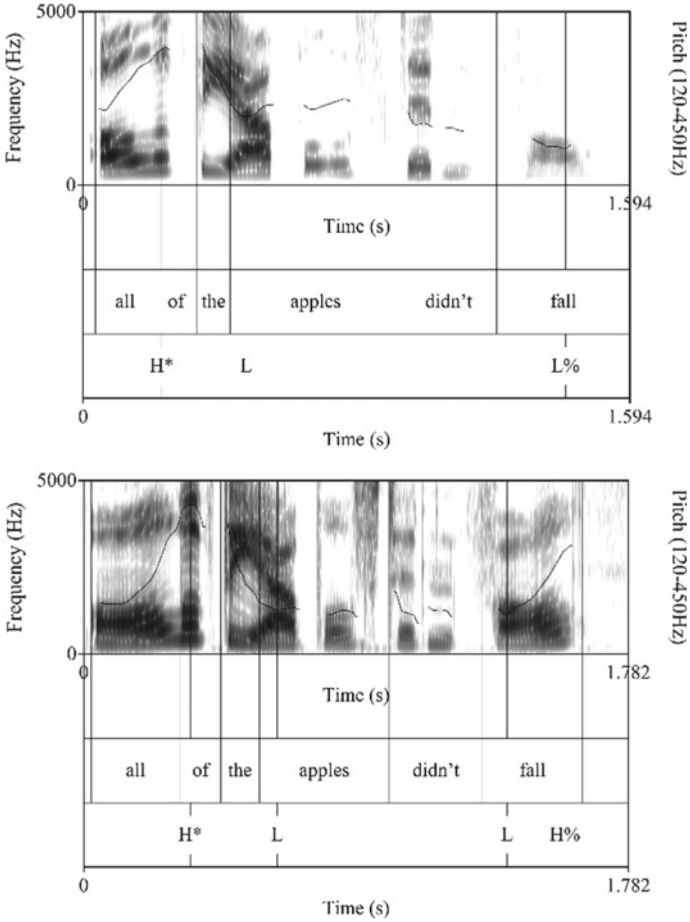


Fig. 3 Pitch tracks for "All of the apples didn't fall" in the Falling condition (above) and the RFR condition (below)

the sentence again, the pair of the question and the answer got played. The session for each participant typically took about 10 min to complete. For adult participants, an answer sheet with 4 cells for each item was handed to them and they were asked to mark the answer.

We recruited 24 adult participants (12 on the Falling condition and 12 on the RFR condition) with no or little linguistics background. The participants were mostly undergraduate students either at MIT or at Wellesley College. For the experiment with children, we recruited 32 children. Among them, 16 were tested on the Falling condition (ranging from 4;4 to 6;10, mean age = 5;3) and 16 were tested on the RFR condition(ranging from 4;5 to 6;7, mean age = 5;2). The sessions took place at local daycare centers in Boston/Cambridge area and at the Boston Children's

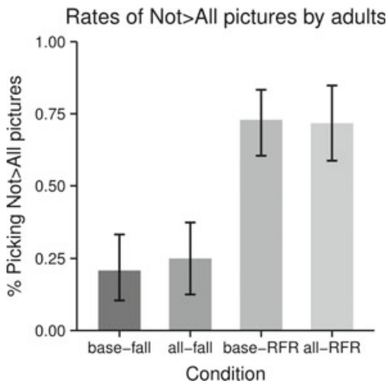


Fig. 4 Results from adults—Error bars indicate 95% confidence interval

Museum across all socioeconomic and ethnic backgrounds. Additional 4 children were excluded from the analysis because they skipped two or more items (N = 3) or answered three or more fillers wrong (N = 1).

4.2 Results—Adults

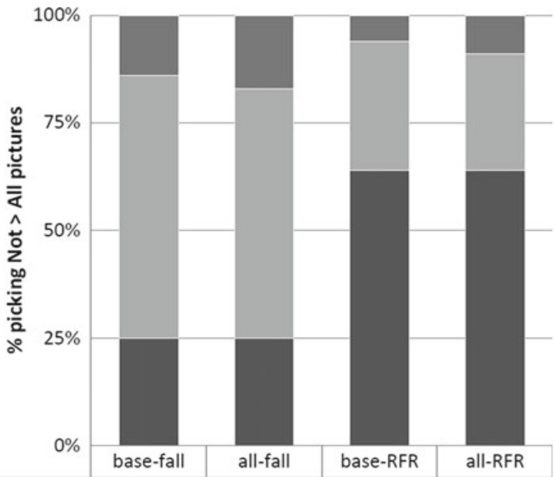
Of 192 relevant data points, 2 were excluded from the analysis since they contained responses that chose a positive picture. That is, the 190 data points are analyzed as values of a binomial variable. Figure 4 shows the rates of choosing the "Not > All" pictures. The rates of choosing the "Not > All" pictures for the Baseline-Q-Falling condition was 20.8%, for the All-Q-Falling condition was 25%, for the Baseline-Q-RFR condition was 72.9%, and for the All-Q-RFR condition was 71.7%. Using logit-LMEM,[26] statistical analysis of these rates reveals a main effect of contour ($p = .036$).[27]

[26]Since the maximally specified model did not converge, the order of the presentation was not considered as a possible factor here.

[27]The summary of the analysis of the fixed effects is as follows:

	Estimate Std.	Error	z value	Pr(>|z|)
(Intercept)	−3.4466	2.0767	−1.660	0.0970 .
contourRFR	7.7889	3.7209	2.093	0.0363 *
qTypebase	0.9689	1.5518	0.624	0.5324
contourRFR:qTypebase	−2.3461	2.8214	−0.832	0.4057

Fig. 5 Grand results from children



	base-fall	all-fall	base-RFR	all-RFR
Incorrect	0.14	0.17	0.06	0.09
All>Not	0.61	0.58	0.3	0.27
Not>All	0.25	0.25	0.64	0.64

4.3 Results—Children

Figure 5 displays all the responses from the child participants, including the errors (choosing positive pictures or irrelevant pictures). The errors account for 30 data points out of 256 data points, and the most frequent error was to choose the positive picture. The error rates were statistically not different across the four conditions.[28] The 30 error responses are excluded from the analysis hereafter. This allows us to treat the responses as binomial answers since the remaining data points only include responses either for the "Not > All" pictures or for the "All > Not" pictures.

The relative rates that only contain either "Not > All" or "All > Not" responses displayed in Fig. 6. It shows the rates of choosing the "Not > All" pictures by the conditions. The rates for the Baseline-Q-Falling condition was 29.1%, for the All-Q-Falling condition was 30.2%, for the Baseline-Q-RFR condition was 68.3%, and for the All-Q-RFR condition was 70.7%. Statistical analysis of these rates reveal a main effect of contour ($p = 0.038$).[29]

[28]The examination of the error rates across conditions using LMEM, where choices for positive ("All did") pictures and for irrelevant pictures were coded as errors and all other types of responses ("All > Not" and "Not > All") as relevant and hence non-errors, revealed no significant effect of condition.

[29]For the same reason as given in fn. 26, the order of the presentation was not considered as a possible factor. The question-type did not have an effect. The summary of the analysis of the fixed effects is as follows:

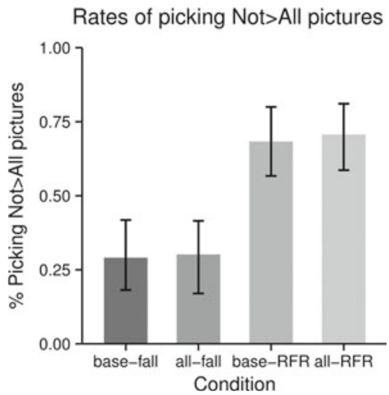


Fig. 6 Normalized results from children—Error bars indicate 95% confidence interval

5 Discussion

5.1 Discussion—Adults

The results from adults show that naïve speakers accessed the "Not > All" reading significantly more often when the sentence is read with the RFR contour. The finding supports the hypothesis that the "Not > All" interpretation is preferred with RFR. Another important thing to note is that in our experimental setting, the preceding context was always the same and neutral for the two conditions. Participants were not able to tell which picture to choose, until they heard the target sentence. The participants could arrive at the respective interpretations using only prosodic cues.

One might wonder why some of the adults did not get the "Not > All" reading even with RFR. It is possible that such participants were simply ignoring the specific contour and so failed to compute alternatives in the RFR condition. This may have been invited by the design because contour was a between-subjects factor hence exhibited no variation within a given participant to create a parallel experiment to the child study. An individual subject analysis shows that the "All > Not" responses under the RFR condition came from a small subset of the participants; one subject consistently (8 out of 8 items) picked "All > Not" pictures, and three subjects strongly preferred "All > Not" interpretations (6 or 7 out of 8). The other 8 subjects always (100%)

	Estimate Std.	Error	z value	Pr(> |z|)
(Intercept)	−2.021	1.232	−1.640	0.1009
contourRFR	3.480	1.678	2.074	0.0381 *
qTypebase	−1.814	1.622	−1.119	0.2633
contourRFR:qTypebase	2.432	2.195	1.108	0.2679

picked "Not > All" pictures. The fact that some subjects kept choosing "All > Not" pictures even when the sentence was read with RFR suggests that those subjects have a strong preference when it comes to the sentences which contain a universal quantifier and negation, and it is possible that because of the strong preference they did not pay much attention to the prosody.

In sum, the results provide strong support for the idea that the prosody, specifically the RFR contour, leads a hearer to arrive at the "Not > All" interpretations in surface-ambiguous sentences with sentence level negation and a universal quantifier in subject position. Thus, building on Ianucci and Dodd's as well as Syrett et al. studies but employing a new methodology, our study provides further and stronger support for the claim that the choice of intonation (neutral or RFR) can disambiguate scopally ambiguous sentences.

5.2 Discussion—Children

Results from child participants conform closely with those of adults. We have raised two questions to motivate this study—(i) Are children sensitive to the difference in prosodic contour? And (ii) if so, do children compute the effect of RFR in the same way as adults do? The results of this study suggest positive answers to both of these questions.

First, we observed a main effect of contour indicating that children are in fact sensitive to the difference in contours. This might come as a surprise since the literature is controversial as to whether children around the age of 4 and 6 are sensitive to sentential (suprasegmental) intonation, and studies of adult speech report that adults do not reliably produce distinctive contours to convey distinct scope relations (McMahon et al. 2004). Note that our set-up makes it difficult to explain this sensitivity as an experimental artifact. I.e., as shown in Fig. 2, the preceding stories up until the target sentences were the identical (using the same pictures and same sound files) across the conditions. Thus, the preceding stories did not favor one interpretation to other, and the only difference was the prosody of the target sentences. This suggests that children arrive at different interpretations solely depending on the difference in prosody. This seems to us to be quite strong evidence for children's knowledge of the scopal interpretation of prosody.

Second, the main effect of contour we observed was in the same direction as that for adults. This suggests that children compute the effect of RFR on sentence meaning in the way adults do. As we have seen in the Sect. 2, the computations involved determining the meaning of a sentence with RFR are rather complex involving comparing multiple LFs and their alternatives in terms of whether they satisfy the condition that after accepting the asserting under a particular LF at least one alternative to that LF must still be debatable. In spite of this complexity, children do seem to get the effect. This significantly contrasts with previous studies on the scope assignment by children. More specifically, we have reviewed two studies in Sect. 3.1 suggesting that children's inability to access the non-isomorphic interpretation in run of the mill contexts is due to a processing limitation. That is, when the processing load is alleviated, e.g. by placing an affirmative sentence beforehand or by priming children with a certain LF, inverse scope readings become more accessible. Our results contrast with these studies in that children's ability to access the non-isomorphic reading increases without any manipulation to lessen the processing load. Rather, computing the effect of RFR might have put more processing load on children, since RFR involves alternative generation and implicature computation. One way to resolve this apparent conflict might be to suggest that our children were successful in computing the RFR effects because RFR induces strong enough pressures (e.g. by effectively eliminating the otherwise preferred reading prior to assessing whether it is true in the situation) in favor of inverse scope reading while previous studies leave the choice between the two readings open until situational fit (truth-value judgment or picture selection) has to be determined.[30]

Let us have a closer look at the RFR results of adults and those of children. At first sight, the figures look very similar—72.9 and 71.7% (mean by collapsing contours = 72.3%) with adults and 68.3 and 70.7% (mean = 69.5%) with children. As we discussed in Sect. 5.1, most of the "All > Not" responses from adults (about 28%) come from a subset of the participants. That is, there are several "isomorphism" participants whose responses contribute much of the "All > Not" responses in the RFR condition. When looking into the individual results of children, it turns out the 30% of the "All > Not" responses in the RFR condition are also coming from a subset of children, who could be dubbed an "isomorphic population." What is interesting is that the isomorphic children are found virtually only in the population younger than 5½ years of age. Figure 7 shows the correlation between the % of choosing "Not > All" pictures (out of 8 trials) on the RFR condition and children's age in days. In fact, the correlation was only marginal ($t = 1.78$, $df = 14$, $p = 0.097$, $cor = 0.429$) but we can observe that there is a trend with a medium-to-large effect size. As we can see in the graph, children can be grouped into two subgroups: the group who (almost) always get the effect of RFR, and the group who behave (almost) always isomorphic. Just as for adults, we observed two populations. The children look extremely like the adults in this regard; the percentages of isomorphic participants in fact is quite close for the child and adults populations. Children who almost always get the effect range from four years old to six, but the children who are almost always isomorphic can only be found in the population younger than 5½ years of age. One possibility is that the isomorphic group might undergo some kind of maturation in sensitivity to the prosodic contour that calls for computation over scopal alternatives. The other possibility, as mentioned previously, is that the maturation involves the

[30]An alternative approach might be to reduce the facilitation effects observed in previous studies to one of the factors we manipulated in our experiment as well. The natural candidate for such a proposal would be the salience of an appropriate question under discussion, as we suggested in footnote 12. Of course, details of such a proposal would need to be worked out.

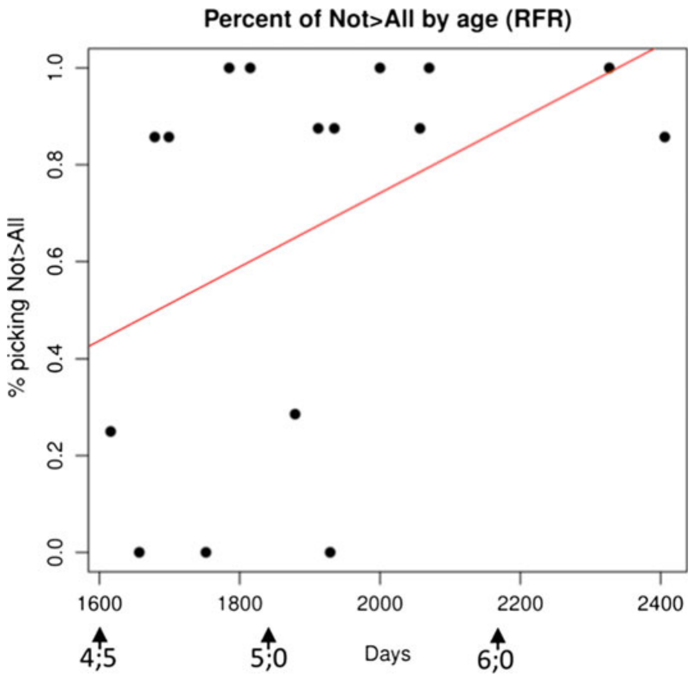


Fig. 7 Correlation between % choosing "Not > All" pictures on the RFR condition and age

ability to compute scalar implicatures, or possibly to learn/construct the relevant set of alternatives for the scalar computation that forces the existence implicature.[31,32]

6 Concluding Remarks

In this chapter, we have presented a new set of data that shows that preschool children do access the non-isomorphic reading in sentences where a universal quantifier in the subject position precedes negation. Investigation on whether children can get the non-isomorphic reading in such sentences has gotten much attention since Musolino (1998), and several more recent studies have revealed that children have the competence in computing the interpretation, when the processing load is alleviated. The results of the current study add an important new data point to the literature showing that children do access the interpretation even when processing demands are likely

[31] See e.g. Reinhart (2006) for discussion of ideas along these lines.

[32] As one of the reviewers suggests, it would be interesting to look into a potential correlation between the sensitivity to the RFR contour (which affects the truth condition of a sentence, as in the current study) and the perception of prosody in general, which does not affect the truth condition of a sentence.

higher than in out of the blue cases. We have tested whether children are sensitive to the difference in prosody and compute the effects of the Rise-Fall-Rise contour. Our results show that children arrive at different interpretations solely depending on the difference in prosody, and that they interpret the sentences in the same direction as adults do, even roughly to the same quantitative extent as adults do and with roughly the same variability across subjects as for adults. This is a remarkable demonstration of child knowledge (and processing) being similar in many of the defining features as adult knowledge and processing. Children indeed know (and use) the prosody/alternative calculation link in a similar manner to adults.

Acknowledgements We are grateful to the children and the adult participants who participated in the study, the children's parents and teachers at daycare centers. I would also like to thank Sammy Floyd and Laura Schulz at the MIT Early Childhood Cognition Lab at Boston Children's Museum for their courtesy. We would especially like to thank Noah Constant, Danny Fox, Irene Heim, Shigeto Kawahara, and two anonymous reviewers for helpful and fruitful discussions and suggestions. Finally, we are thankful to the audiences at the LF Reading Group at MIT, Mie Workshop on Linguistic Theory and Language Acquisition at Mie University, the GALA 2015 in Nantes, and the Workshop on Linguistic and Cognitive Aspects of Quantification 2015 in Budapest for helpful comments and suggestions.

References

Arnold, J.E. 2008. THE BACON not *the bacon*: How Children and Adults Understand Accented and Unaccented Noun Phrases. *Cognition* 108: 69–99.

Baauw, S., E. Ruigendijk, and F. Cuetos. 2004. The Interpretation of Contrastive Stress in Spanish-speaking Children. In *Proceedings of Generative Approaches to Language Acquisition 2003 (GALA2003)*, 103–114.

Bajaj, V., V. Déprez, and J. Musolino. 2014. The Question Under Discussion and its Role in Scopal Ambiguity Resolution. In *Proceedings of the 48th Annual Meeting of the Chicago Linguistic Society*, 33–44.

Bobaljik, J.D., and S. Wurmbrand. 2012. Word Order and Scope: Transparent Interfaces and the 3/4 Signature. *Linguistic Inquiry* 43: 371–421.

Braun, B. 2006. Phonetics and Phonology of Thematic Contrast in German. *Language and Speech* 49: 451–493.

Büring, D. 1997. The Great Scope Inversion Conspiracy. *Linguistics and Philosophy* 20: 175–194.

Büring, D. 2003. On D-trees, Beans, and B-accents. *Linguistics and Philosophy* 26: 511–545.

Calhoun, S. 2006. *Information Structure and the Prosodic Structure of English: A Probabilistic Relationship*. PhD dissertation, University of Edinburgh.

Constant, N. 2012. English Rise-Fall-Rise: A Study in the Semantics and Pragmatics of Intonation. *Linguistics and Philosophy* 35: 407–442.

Constant, N. 2014. *Contrastive Topics: Meanings and Realizations*. PhD dissertation, University of Massachusetts Amherst.

Crain, S., and R. Thornton. 1998. *Investigations in Universal Grammar: A Guide to Research on the Acquisition of Syntax and Semantics*. Cambridge: MIT Press.

Crain, S., and K. Wexler. 1999. Methodology in the Study of Language Acquisition: A Modular Approach. In *Handbook of Language Acquisition*, ed. W.C. Ritchie, and T.K. Bhatia, 387–425. San Diego: Academic Press.

É. Kiss, Katalin, and B. Gyuris. 2003. Apparent Scope Inversion Under the Rise Fall Contour. *Acta Linguistica Hungarica* 50: 371–404.

Gualmini, A., S. Hulsey, V. Hacquard, and D. Fox. 2008. The Question-Answer Requirement for Scope Assignment. *Natural Language Semantics* 16: 205–237.
Gussenhoven, C. 1983. *A Semantic Analysis of the Nuclear Tones of English*. Bloomington: Indiana University Linguistics Club (IULC).
Horn, L. 1989. *A Natural History of Negation*. Chicago: University of Chicago Press.
Horn, L. 2005. Airport '86 Revisited: Toward a Unified Indefinite Any. In *Reference and Quantification: The Partee Effect*, ed. G. Carlson, and F.J. Pelletier, 179–205. Stanford: CSLI.
Iannucci, D., and D. Dodd. 1980. The Development of Some Aspects of Quantifier Negation in Children. In *Papers and Reports on Child Language Development (PRCLD)* #19, 88–94. Stanford: Stanford University Department of Linguistics.
Ito, K., N. Jincho, U. Minai, N. Yamane, and R. Mazuka. 2012. Intonation Facilitates Contrast Resolution: Evidence from Japanese Adults and 6-Year Olds. *Journal of Memory and Language* 66: 265–284.
Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.
Jespersen, O. 1933/1964. *Essentials of English Grammar*. (Eighth printing, 1994). Tuscaloosa: The University of Alabama Press.
Krifka, M. 1998. Scope Inversion Under the Rise-Fall Contour in German. *Linguistic Inquiry* 29: 75–112.
Ladd, D.R. 1980. *The Structure of Intonational Meaning*. Bloomington: Indiana University Press.
McDaniel, D., and T.L. Maxfield. 1992. Principle B and Contrastive Stress. *Language Acquisition* 2 (4): 337–358.
McMahon E., J. Lidz, and J. Pierrehumbert. 2004. Suprasegmental Cues to Meaning in Child-directed Speech. In *Handbook of the 17th CUNY Sentence Processing Conference*. University of Maryland.
Meyer, M.-C., and U. Sauerland. (2009). A Pragmatic Constraint on Ambiguity Detection: A Rejoinder to Büring and Hartmann and to Reis. *Natural Language and Linguistic Theory* 27: 139–150.
Musolino, J. 1998. *Universal Grammar and the Acquisition of Semantic Knowledge: An Experimental Investigation into the Acquisition of Quantifier-negation Interaction in English*. PhD dissertation, University of Maryland.
Musolino, J. 2011. Studying Language Acquisition through the Prism of Isomorphism. In *Handbook of Generative Approaches to Language Acquisition: Studies in Theoretical Psycholinguistics*, vol. 41, ed. J. de Villiers, and T. Roeper, 319–349. Dordrecht: Springer.
Musolino, J., S. Crain, and R. Thornton. 2000. Navigating Negative Quantificational Space. *Linguistics* 38 (1): 1–32.
Musolino, J., and J. Lidz. 2006. Why Children Aren't Universally Successful with Quantification. *Linguistics* 44–4: 817–852.
Reinhart, T. 2006. *Interface Strategies: Optimal and Costly Computations*. Cambridge: MIT Press.
Roberts, C. 1996/2012. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics and Pragmatics* 5: 1–69.
Sugawara, A., and K. Wexler. 2014. Children do not Accept Unambiguous Inverse-Scope Readings: Experimental Evidence from Prosody and Scrambling in Japanese. *Proceedings of Formal Approaches to Japanese Linguistics 7 (FAJL7)*, 215–226.
Syrett, K., G. Simon, and K. Nisula. 2014a. Prosodic Disambiguation of Scopally Ambiguous Sentences. In *Proceedings of 43rd Annual Meeting of the North East Linguistic Society* (NELS), 141–152.
Syrett, K., G. Simon, and K. Nisula. 2014b. Prosodic Disambiguation of Scopally Ambiguous Quantificational Sentences in a Discourse Context. *Journal of Linguistics* 50: 453–493.
Viau, J., J. Lidz, and J. Musolino. 2010. Priming of Abstract Logical Representations in 4-Year-Olds. *Language Acquisition* 17 (1–2): 26–50.
Ward, G., and J. Hirschberg. 1985. Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation. *Language* 61: 747–776.

Differentiating Universal Quantification from Perfectivity: Cantonese-Speaking Children's Command of the Affixal Quantifier *saai3*

Margaret Ka-yan Lei and Thomas Hun-tak Lee

Abstract This study investigates whether Cantonese-speaking preschoolers are sensitive to the semantic differences between universal quantification and perfectivity under the differentiating context of negation. In Cantonese, the negation of a perfective predicate in the form of [NEG V] denotes the non-existence or non-realization of an event ("none" reading), while the negation of a predicate suffixed by the universal quantifier *saai3* in the form of [NEG V *saai3*] denotes the partial realization of an event ("partial" reading). Using the two-choice picture/video selection task, we tested 34 children aged between 3;6 and 4;6 (mean age = 3;10) and 72 adults in a between-subject design on sentences of the form [NEG V] (negation of perfectivity) or [NEG V *saai3*] (negation of universal quantification), paired with a none reading (non-existence or non-realization or an event) and a partial reading (partial realization of an event). Our findings reveal that children are able to differentiate universal quantification and perfectivity in the negation context. While children can understand *saai3* quantifying an object under the scope of negation, a blocking effect is observed in subject quantification with the negator intervening between *saai3* and its associated nominal.

In this paper, Cantonese is transcribed using *The Linguistic Society of Hong Kong Cantonese Romanization Scheme* (*Jyutping*) (http://www.lshk.org/). The digits after the romanization indicate the tone category, which is only provided for morphemes under discussion. Abbreviations: CL = classifier; CL_{PL} = plural fuzzy classifier *di1*; COG.OBJ = cognate object; DEM = demonstrative; DUR = durative aspect marker; PERF = perfective aspect marker; N = noun; NEG = negation; SFP = sentence final particle; V = verb. The following convention is used to indicate the age of the child: yy;mm;dd or yy;mm, in which the numbers corresponding to 'yy', 'mm' and 'dd' indicate year, month and day respectively.

M. K. Lei (✉) · T. H. Lee
Department of Linguistics and Modern Languages, Chinese University of Hong Kong,
Shatin, New Territories, Hong Kong
e-mail: margaret.lei@link.cuhk.edu.hk

T. H. Lee
e-mail: huntaklee@cuhk.edu.hk

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_4

Keywords Semantic acquisition · Child Cantonese · Universal quantification A-quantification · Domain of quantification · Negation of perfectivity · Blocking effect of negation

1 Introduction

It has been observed that the quantificational meaning expressed by verbal affixes as an A-quantifier (Partee 1995) may conflate with aspectual distinctions in incremental-theme contexts (Krifka 1992; Filip 1997, 2001). In Slavic languages such as Czech, perfectivity marked by the prefix *vy-* in (1a) denoting 'drank (up) the tea', or the prefix *s-* in (1b) denoting 'ate (up) the apples', is indistinguishable from an universally quantified interpretation on the incremental-theme object, be it the bare mass object *čaj* 'tea' in (1a), meaning 'all the tea' or 'the whole portion of tea', or the bare plural object *jablka* 'apple' in (1b), meaning 'all the apples' (Filip 2001).

(1) Czech (Filip 2001: 463)

a. Ivan vy-pil^{P} čaj.
Ivan COMPL-drink.PAST tea.SG.ACC
'Ivan drank (up) (all) the tea/the (whole portion of) tea.'

b. Ivan s-nědl^{P} jablka.
Ivan COMPL-eat.PAST apple.PL.ACC
'Ivan ate (up) (all) the apples.'

Such an alignment of perfectivity with universal quantification is similarly observed in other Slavic languages. For the Polish example in (2), perfectivity marked by the prefix *z-* on the verbal predicate can be viewed as equivalent to a universal quantification interpretation assigned on the bare mass object *kaszę* 'porridge', which bears an incremental-theme role.

(2) Polish (Wierzbicka 1967: 2237, with additional glosses from Filip 2008: 25)

On z.jadł^{P} kaszę.
he.NOM PREF.ate porridge.SG.ACC
'He ate (up) (all) the porridge.'

Another piece of evidence for the close affinity between aspect and quantification comes from the dual meanings that can be represented by the morphological marker -*aʔal* in the Papuan language Bunaq. When -*aʔal* is used with stative predicates taking a semantically plural inanimate nominal, it can either denote the final boundary of an event, as given in (3i), or express universal quantification over the nominal domain, as given in (3ii) (Huber and Schapper 2014).

(3) Bunaq (Schapper 2010: 457)

Zo	baʔa	za	h-aʔal.
mango	DEM.INAN	ripe	3INAN-ALL

(i) 'The mango(es) had finished ripening.'
(ii) 'All the mangoes were ripe.'

A highly similar interaction between universal quantification and perfectivity can be found in Cantonese—a member of the Chinese languages (or dialects) that is known for having a rich inventory of verbal suffixes that can denote aspectual and quantificational notions (Cheung 1972 [2007]; Lee 1995; Tang 2015). This chapter focuses on examining the postverbal affix *saai3*, which is suffixed to a verb or a preposition and is flexible in its quantificational domain in associating with various constituents in a sentence to denote universality 'all' (Cheung 1972 [2007]; Lee 1994, 1995, 2001; Tang 1996a, b; Au Yeung 1998; Pan and Man 1998; Lee 2004, 2012; Wong 2008; among others). This affix is reported to be the most frequently used universal quantifier among both adult and child speakers of the language (Lei 2017).

To illustrate: in an incremental-theme context, the totality reading of the object noun phrase conveyed by the suffix *saai3* 'all', as in (4b), is not distinct from an event completion meaning due to the perfective suffix *zo2* 'PERF' applied to the verbal predicate, as in (4a). There is a natural link between the extent of the bun being consumed, as expressed by the universal quantifier *saai3* (4b), and the finishing of the event of consuming the bun, as expressed by the perfective aspect marker *zo2* (4a). Likewise, for the subject quantification sentences in (5–6), when all of the people have left, the event of the group of people departing will have occurred (5); when each of the babies has fallen asleep, the event of the group of babies falling asleep will have happened (6).

(4) a.

Anna	sik	zo2	go	minbaau.
Anna	eat	PERF	CL	bun

'Anna ate the bun.'

b.

Anna	sik	saai3	go	minbaau.
Anna	eat	all	CL	bun

'Anna ate the whole bun.'

(5) a.

Keoidei	zau	zo2.
they	leave	PERF

'They have left.'

b.

Keoidei	zau	saai3.
they	leave	all

'All of them have left.'

(6) a. Di bibi fan zo2 gaau.
CL_{PL} baby sleep PERF COG.OBJ
'The babies have fallen asleep.'

b. Di bibi fan saai3 gaau.
CL_{PL} baby sleep all COG.OBJ
'All of the babies have fallen asleep.'

This correspondence in meaning is understandable, since performing an action on each member of a set of objects or each part of a mass of substance can be understood in the same way as carrying out an action on the entire set of objects or the whole quantity of substance. *Saai3*-quantification over an incremental-theme argument can thus be analyzed as exhibiting a homomorphic mapping between the noun phrase and the verbal predicate, such that quantification performed over subparts of an individual is equivalent to quantification over sub-events denoted by the verbal predicate (cf. Krifka 1992; Filip 1999).

All these cross-linguistic similarities invite us to examine children's understanding of the link between aspect and quantification in the acquisition of A-quantification. In Cantonese, the ambiguity between universal quantification marked by *saai3* and perfectivity marked by *zo2* poses a learning puzzle: as the meanings conveyed by *saai3* 'all' and *zo2* 'PERF' are truth-conditionally indistinguishable in incremental-theme contexts, children may confuse aspect with quantification, and thus, fail to assign distinctive interpretations to each of the two markers. Specifically, if young children correctly interpret the (b) sentences encoded by *saai3* in (4–6) as denoting totality of objects, do they achieve this understanding because the reading is not distinct from a perfective reading of the corresponding (a) sentences encoded by the perfective aspect marker *zo2*? Or do children already have a good command of the A-quantifier *saai3*, in addition to a basic understanding of the meaning of the perfective aspect marker *zo2*?

To address this question, one would need to look into environments which would tell the two forms apart. One such environment is the use of the universal quantifier *saai3* 'all' and the perfective aspect marker *zo2* 'PERF' under the context of negation. When the negator *mou5* negates a perfective predicate, it denotes the non-existence or non-realization of an event ("none" reading) and excludes the perfective aspect marker *zo2*, as illustrated in (7a) and (8a); when the same negator is used in conjunction with *saai3*, it yields a "not all" or "partial" reading, for the object quantification sentence in (7b) as well as the subject quantification sentences in (8b).

(7) a. Anna **mou5** sik go minbaau.
Anna not.have eat CL bun
'Anna did not eat the bun.'

b. Anna **mou5** sik **saai3** go minbaau.
Anna not.have eat all CL bun
'Anna did not eat the whole bun.'

(8) a. Di joengzai **mou5** fan gaau.
CL$_{PL}$ lamb not.have sleep COG.OBJ
'The lambs did not sleep.'

b. Di joengzai **mou5** fan **saai3** gaau.
CL$_{PL}$ lamb not.have sleep all COG.OBJ
'The lambs have not all slept.' = 'Not all of the lambs slept.'

The examples in (7-8) point to two further learning challenges for Cantonese children: first, children cannot simply rely on the occurrence of the perfective aspect marker for signaling perfectivity, due to its absence in negative contexts; second, for subject quantification of *saai3*, children need to know that the link to its domain is intervened by a negator, which is located between *saai3* and the subject noun phrase.

This chapter presents an experimental study examining children's sensitivity to the distinction between universal quantification and perfectivity, addressing the following research questions:

(i) Can children differentiate between universal quantification and perfectivity under the context of negation, in a language in which aspectual and quantificational notions are both expressed by postverbal affixes?
(ii) Specifically, are children capable of distinguishing negation of universal quantification from negation of perfective predicate, assigning a "partial" (NOT > ALL) interpretation to the former, and a "none" (ALL > NOT) interpretation to the latter?
(iii) Do children interpret the scope of the universal quantifier *saai3* and negation differently depending on whether *saai3* is linked to the object nominal or the subject nominal? Do children find it more difficult to process subject quantification due to the intervention of the negator?

Our paper will be organized as follows: first, we discuss the semantic and distributional differences between the universal affixal quantifier *saai3* 'all' and the perfective aspect marker *zo2* 'PERF'. This analysis will provide the background for the design of our experiment, which makes use of one of the differentiating contexts. Second, we highlight some basic facts about the acquisition of *saai3* and *zo2* based on our earlier acquisition research. In the third section, we report on the design and findings of our experiment, and discuss their implications for our understanding of the acquisition of A-quantification.

2 Semantic and Distributional Differences between the Affixal Universal Quantifier *saai3* and the Perfective Aspect Marker *zo2*

In spite of the apparent meaning overlap between the affixal universal quantifier *saai3* 'all' and the perfective aspect marker *zo2* 'PERF' in incremental-theme contexts that we have just discussed, the two forms exhibit a number of semantic and distributional differences in other contexts.

2.1 *The Perfective Aspect Marker* zo2

In Cantonese, perfectivity expressed by the postverbal *zo2* 'PERF' functions to denote the realization of an event (Cheung [1972] 2007; Tang 2015), as illustrated in (9). When *zo2* appears after a stative verb or an adjective, it yields an inchoative reading, as exemplified in (10).

(9)	Anna	tai	**zo2**	jat	bun	syu.
	Anna	read	PERF	one	CL	book

'Anna read a book.'

(10)	Anna	beng	**zo2**.
	Anna	sick	PERF

'Anna has fallen sick.'

When the negator *mou5* 'not-have' negates a perfective predicate, it denotes the non-existence or non-realization of an event, analogous to *méi* 'not-have' in Mandarin Chinese (Cheng et al. 1997; Lin 2003; Law 2014), as exemplified in (11) and (12).

(11)	Anna	**mou5**	sik	go	minbaau.
	Anna	not.have	eat	CL	bun

'Anna did not eat the bun.'

(12)	Di	aapzai	**mou5**	tiu	lok	seoi.
	CL_{PL}	duckling	not.have	jump	into	water

'The ducklings did not jump into (the) water.'

On some analysis (Law 2014), the negator *mou5* is composed of two underlying grammatical morphemes, that is, a negator coalesced with a perfective auxiliary *jau5* 'have'. It may be for this reason that the negative form *mou5* is mutually exclusive with the perfective aspect marker *zo2* 'PERF', which is banned from a *mou5* negative sentence.[1]

2.2 *The Universal Quantifier Suffix* saai3

The verbal suffix *saai3* functions as a universal A-quantifier, which can be suffixed to a verb or a preposition to quantify over various constituents in a sentence. When regarded as a nominal quantifier, *saai3* exerts quantificational force on either DP1 or DP2 in a structure [DP1 V-*saai3* DP2] (Lee 1994). Which nominal in a sentence is quantified by *saai3* depends on whether the nominal can be interpreted as plural and

---

[1] A similar phenomenon has been observed in Mandarin: the preverbal existential verb *yǒu* 'have' and the postverbal perfective aspect marker *le* are mutually exclusive, leading to the analysis of *yǒu* as an allomorph of the perfective aspect marker (see Wang 1965).

definite: thus, it is the object nominal in (13) and the subject nominal in (14–15) that is selected by *saai3*, as the latter contains a bare noun object that cannot denote definite reference in Cantonese (Leung 1980; Cheung 1989; Cheng and Sybesma 1999). As noted by earlier scholars, the universal quantifier suffix and aspectual suffixes are generally mutually exclusive. In particular, *saai3* is prohibited from co-occurring with an aspect marker such as the perfective *zo2* or the progressive *gan2* (Cheung 1972 [2007]: 170).

(13)	Anna	maai	**saai3**	di	daangou.
	Anna	buy	all	CL_{PL}	cake

‘Anna bought all of the cakes.’

(14)	Keoidei	zau	**saai3**.
	they	leave	all

‘All of them have left.’

(15)	Di	hoksaang	maai	**saai3**	daangou.
	CL_{PL}	student	buy	all	cake

‘All of the students bought cakes.’

As noted earlier, when *saai3* falls under the scope of negation, the negator *mou5* ‘not have’ enters into scopal relation with *saai3* such that the interpretation of the complex verbal predicate [NEG V *saai3*] denotes a “not-all” (or “partial”) reading, be it quantifying the object nominal (7b, repeated below as 16) or the subject nominal (8b, repeated as 17). It should be noted that when *saai3* quantifies the subject nominal (as in (17–18)), an added complexity may arise in cases where *saai3* occurs in sentences containing a predicate that allows a degree reading (as in (18)). In such cases, two potential readings are compatible with a NOT > ALL interpretation: (i) partitivity imposed on the set of individuals denoted by the subject nominal (i.e. object partitivity) or (ii) partitivity imposed on the predicate such that a partial state is attained (i.e. event partitivity). Despite the discontinuity of the universal quantifier *saai3* with its associated nominal, the negator remains the wide scope operator over *saai3* to yield a NOT > ALL reading, whether *saai3* is associated with the subject nominal or with the predicate.

(16)	Anna	**mou5**	sik	**saai3**	go	minbaau.
	Anna	not.have	eat	all	CL	bun

‘Anna did not eat the whole bun.’

(17)	Di	joengzai	**mou5**	fan	**saai3**	gaau.
	CL_{PL}	lamb	not.have	sleep	all	COG.OBJ

‘The lambs have not all slept.’ = ‘Not all of the lambs slept.’

(18)	Di	aapzai	**mou5**	tiu	**saai3**	lok	seoi.
	CL_{PL}	duckling	not.have	jump	all	into	water

‘The ducklings have not all jumped into (the) water.’ =
(i) ‘Not all of the ducklings jumped into (the) water.’ or
(ii) ‘The ducklings have only jumped part of the way into (the) water.’

In brief, the difference between negation of universal quantification and negation of perfectivity can be summarized as follows: when the negator *mou5* negates a perfective predicate, it denotes a “none” reading; when it is used with *saai3*, it yields a “partial” reading. This difference in interpretation will be explored in the present experimental study.

2.3 *Other Cases in Which* zo2 *and* saai3 *Cannot Be Used Interchangeably*

Due to the plurality requirement imposed on *saai3*, this suffixal quantifier is generally incompatible with singular referential arguments, such as in transitive sentences in which both the subject nominal and the object nominal are singular, as in (19b) and (20b). No such restrictions are imposed on the perfective aspect marker *zo2*, as in (19c) and (20c).[2]

[2]A reviewer asked whether a path reading can be applied to the predicate *heoi saai3 Bakging* [go **saai3** Beijing] in (19b) to mean ‘S/he went all the way to Beijing’. We note the contrast between (19b) and the sentence (i) below is a sharp one.

(i)	Keoi	tiu	saai3	lok	seoi.
	s/he	jump	all	into	water

‘S/he jumped all the way into (the) water.’

(19b) cannot receive a path reading as achievement verbs do not admit duration, as shown in the contrast in meaning between (ii) and (iii).

(ii)	Keoi	heoi	zo2	saam	go	zungtau.
	s/he	go	PERF	three	CL	hour

‘S/he has left for three hours.’

(iii)	Keoi	tiu	zo2	saam	go	zungtau.
	s/he	jump	PERF	three	CL	hour

‘S/he jumped for three hours.’

(19)	a.	Keoidei	heoi	**saai3**	Bakging.
		they	go	all	Beijing
		'All of them went to Beijing.'			
	b.	??Keoi	heoi	**saai3**	Bakging.
		s/he	go	all	Beijing
	c.	Keoi	heoi	**zo2**	Bakging.
		s/he	go	PERF	Beijing
		'S/he went to Beijing.'			

(20)	a.	Keoidei	faan	**saai3**	(keoi)	ukkei.
		they	return	all	(his/her)	home
		'All of them went home.'				
	b.	??Keoi	faan	**saai3**	(keoi)	ukkei.
		s/he	return	all	(his/her)	home
	c.	Keoi	faan	**zo2**	(keoi)	ukkei.
		s/he	return	PERF	(his/her)	home
		'S/he went home.'				

Similarly, in the case of intransitive sentences, as there is no nominal object, *saai3* will quantify the subject nominal, while *zo2*, being a perfective aspect marker, will modify the predicate. Due to this difference in their scope of application, *saai3* is incompatible with a singular subject, while *zo2* can co-occur with the latter, as illustrated in (21–22).

(21)	a.	Keoidei	sei	**saai3**.
		they	die	all
		'All of them died.'		
	b.	*Keoi	sei	**saai3**.
		s/he	die	all
	c.	Keoi	sei	**zo2**.
		s/he	die	PERF
		'S/he died.'		

(22)	a.	Di	syuzi	tyun	**saai3**.
		CL_{PL}	tree.branch	break	all
		'All of the tree branches broke.'			
	b.	*Zi	syuzi	tyun	**saai3**.
		CL	tree.branch	break	all
	c.	Zi	syuzi	tyun	**zo2**.
		CL	tree.branch	break	PERF
		'The tree branch broke.'			

3 Cantonese-Speaking Children's Early Knowledge of *saai3* and *zo2*

Before we move on to report our present experimental study, we first highlight some basic facts about the acquisition of *saai3* and *zo2* based on our earlier acquisition research.

3.1 Children's Early Use of saai3 *and* zo2 *in Spontaneous Speech*

Our analysis of the speech of ten children from two naturalistic longitudinal corpora on child Cantonese: CANCORP (Lee and Wong 1998) and HKCELA (Lee 2010) shows that Cantonese-speaking children are able to use the universal quantifier affix *saai3* 'all' and the perfective aspect marker *zo2* 'PERF' early on in their spontaneous productions (Lei 2017). The perfective aspect marker *zo2* is found to have emerged as early as 1;6. It is productively used in the correct postverbal position, in utterances containing an object noun phrase, as in (24), and also in utterances without taking an object noun phrase, as in (23).

(23) CKL (HKCELA) at 02;03;03.

Heoi	**zo2**	Dungwuising	waan	aa3.
go	PERF	City.Gate	play	SFP

'(I) went to City Gate to have fun.'

(24) CKL (HKCELA) at 02;02;08.

Sik	**zo2**	keoi.
eat	PERF	it

'(I) ate it.'

The universal quantifier suffix *saai3* 'all' first appeared at around the age of two, with the earliest token occurred at the age of 1;10. Children were able to suffix *saai3* correctly to a simple verb or a verb-compound, in utterances containing a semantically plural object, as in (25–26), and in utterances without taking an object, as in (27). Only one instance of the co-occurrence of *saai3* and *zo2* was found in the whole dataset, suggesting that children were well aware of the restriction against the co-occurrence of the two forms.

(25) CGK (CANCORP) at 02;08;08.

Tai	**saai3**	nei	di	aa3.
read	all	this	CL_{PL}	SFP

'(I) read all of these.'

(26) LLY (CANCORP) at 03;03;26.

Cyunbou	jam	**saai3**	keoi	aa1	he1.
whole.lot	drink	all	it	SFP	SFP

‘Drink all of it.’

(27) LTF (CANCORP) at 02;04;27.

Taan	**saai3.**
play	all

‘(I) play(ed) all of (it/these).’

To better understand children’s awareness of the differences between *saai3* and *zo2*, we further analyze the contexts in which verbs carrying the two affixes were used by the same child. Below we identify two contexts in which universal quantification may be indistinguishable from perfectivity.

First, in incremental-theme contexts, while an exhaustivity interpretation seemed to be assigned to *saai3*, the perfectivity meaning can hardly be excluded. When *saai3* is suffixed to a verb of consumption such as *sik6* ‘eat’, as in (28), which takes an incremental-theme object, the consumption of all subparts of the object would naturally lead to the completion of the object-eating event, thus corresponding to the meaning conveyed by *zo2* in (29), which is suffixed to the same verb *sik6* ‘eat’.

(28) LTF (CANCORP) at 02;06;01.

Ngo	sik	**saai3**	keoi	jiu	aa3.
I	eat	all	it	want	SFP

‘I must eat all of it.’

(29) CKL (HKCELA) at 02;02;08.

Sik	**zo2**	keoi.
eat	PERF	it

‘(I) ate it.’

In incremental-theme utterances containing a null object noun phrase, the same meaning overlap between aspect and quantification will arise if a plural interpretation is assumed for the underlying object. The pair of utterances illustrated in (30–31), with the former suffixed by *saai3* and the latter suffixed by *zo2*, can both be felicitously used to describe a situation involving the consumption of a plural set of contextually defined objects.

(30) LTF (CANCORP) at 02;06;01.

Sik	**saai3.**
eat	all

‘(I) ate all.’

(31) LTF (CANCORP) at 02;08;02.

Sik	**zo2.**
eat	PERF

'(I) ate.'

Second, the meaning overlap between *saai3* and *zo2* is also evidenced in sentences containing unaccusative verbs describing a change of state occurring on a null plural theme object. In (32–33), in which an extent or path-quantification reading is permitted, the two suffixes can be used interchangeably.

(32) LTF (CANCORP) at 02;06;01.

Dit	**saai3**	lok	heoi.
fall	all	down	go

'All fell down.'

(33) LTF (CANCORP) at 02;04;27.

Dit	**zo2,**	dit	**zo2,**	lok	heoi.
fall	PERF	fall	PERF	down	go

'(It/they) fell down.'

Taken together, as there is no transparent way to tease apart the two meanings based on surface form, in incremental-theme contexts as well as in certain sentences containing unaccusative verbs predicated of null plural arguments, it is possible that the intended meaning assigned by children to their *saai3*-utterances does not necessarily correspond to the exhaustive interpretation assigned by adults.

3.2 Children's Understanding of saai3 *in Exhaustive Contexts*

In a number of experimental studies (Lei 2017), we have established that children are able to comprehend the exhaustivity meaning of *saai3* in many of the contexts. Five-year-olds were able to use *saai3* to express the meaning of exhaustivity when prompted in appropriate contexts in an elicited production task.[3, 4] Some of the *saai3*-utterances produced by children are shown in (34–35):

(34)

Winnie	the	Pooh	sik	**saai3**	di	tong.
Winnie	the	Pooh	eat	all	CL_{PL}	candy

'Winnie the Pooh ate all of the candies.'

[3] In the elicitation production task, children were asked to explain why the protagonist in each of the stories was disappointed upon realizing that his/her possession had been taken away or consumed at the end of that story.

[4] A few of the utterances produced by children in the task aiming to elicit *saai3* contained *zo2* instead of *saai3*, which were often accompanied by a numeral phrase object, with the cardinality of the numeral matched with the cardinality of the object denotation depicted in the test scenarios.

(35)	Waineihung	ne	jam	**saai3**	keoi	di	caangzap.
	Winnie the Pooh	SFP	drink	all	his/her	CL_{PL}	orange.juice

‘Winnie the Pooh drank all of his/her orange juice.’

In terms of comprehension, children exhibited sensitivity to the totality meaning of *saai3* involving object quantification (as in (36)) as early as two years of age in an eye-tracking preferential looking task, as reflected in a significantly longer time to the match videos depicting “universal” events than to the non-match videos depicting “partial” events.

Children understood the universal quantificational meaning of *saai3* virtually perfectly by three-and-a-half years of age for object quantification (with sentences such as the ones in (36–37)) in a picture selection task, and by four years of age for subject quantification (with sentences such as the one in (38)) in the same task. Children comprehended the universal quantificational meaning of *saai3* equally well when *saai3* quantifies either a singular object nominal (depicting part-whole relationship) (as in (36)) or a plural set of objects (depicting set-subset relationship) (as in (37)).

(36)	Gogo	jam	**saai3**	di	caangzap.
	elder.brother	drink	all	CL_{PL}	orange.juice

‘Elder brother drank all of the orange juice.’

(37)	Gogo	dam	**saai3**	di	hungmaau.
	elder.brother	throw.away	all	CL_{PL}	panda

‘Elder brother threw away all of the pandas.’

(38)	Di	joengzai	fan	**saai3**	gaau.
	CL_{PL}	lamb	sleep	all	COG.OBJ

‘The lambs have all fallen asleep.’

However, as the contexts presented in the above experimental paradigms are largely incremental-theme context compatible with a perfectivity reading, whether children are truly capable of comprehending the exhaustivitiy meaning of *saai3* requires further empirical verification.

3.3 Children’s Knowledge of the Relative Scope of saai3 *and Negation*

With respect to the relative scope of *saai3* and negation, Cantonese-speaking five-year-olds were able to use *saai3* in the correct word order to express the “not all” interpretation in an elicited production task, reflecting a sensitivity to the syntax of scope.[5] Some of children's productions are shown in (39–40):

(39) Mou sik **saai3** di faan.
not.have eat all CL_{PL} rice
'(S/he) did not eat all the rice.'

(40) Mou zou **saai3** di gungfo.
not.have do all CL_{PL} homework
'(S/he) did not do all of the homework.'

Children were also able to use *saai3* in interaction with negation in their spontaneous speech, with a total of seventeen tokens found in the dataset (the same dataset reported in Sect. 3.1). The negator always precedes *saai3* in all of the tokens, which include *m4* 'not' (ten tokens), *mei6* 'not yet' (five tokens), *mou5* 'not have' (one token) and *m4hai6* 'not be' (one token). Some of the tokens are given in (41–42).[6]

(41) MHZ (CANCORP) at 02;01;01. The child is drinking something.
Mei jam **saai3.**
not.yet drink all
'(I) haven't finished drinking yet.'

(42) LLY (CANCORP) at 03;04;22. There is a cot in the picture with enclosed panels whereas the child's cot has bars around it.
Ngo go go dou mou zo zyu **saai3.**
my that CL also not.have block DUR all
'Mine is also not fully blocked.'

As the production data only shows children's awareness of the partial (NOT > ALL) reading in *saai3*'s interaction with negation, further investigations are needed to examine whether children differentiate between the negation of *saai3* and the negation of perfective predicate.

4 The Present Study

In this study, we tested 34 Cantonese-speaking children (age range: 3;6–4;6) (and 72 adult controls) to ascertain whether children may be relying on perfectivity in

[5]In the elicitation production task, children were asked to explain why the protagonist in each of the stories was penalized, whose actions did not cover all the items of a set or a quantity as required by the task.

[6]It should be noted that the co-occurrence of the universal quantifier suffix *saai3* with the durative aspectual marker *zyu6* in (42) is an exception to the general mutual exclusivity constraint observed in Sect. 2.2.

comprehending the universal quantifier *saai3*, making use of the differentiating context of negation using a two-choice picture/video selection task, addressing the three issues (i–iii) introduced in Sect. 1.

4.1 Participants

Thirty-four Cantonese-speaking children (age range: 3;6–4;6, mean age: 3;10) participated in this experiment. They were all recruited from a local kindergarten in Hong Kong. All of the participants were born in Hong Kong or the Guangdong province and spoke Cantonese as their mother tongue with either or both of their parents being native speakers of Cantonese. We also included 72 college-age adult native speakers of Cantonese as controls.

4.2 Test Materials and Experiment Design

A two-choice video/picture selection task was used. We adopted a between-subject design, with half of the participants tested on sentences of the form [NEG V] (negation of perfectivity) (the (a) sentences in (43–46)), and the other half of the participants tested on sentences of the form [NEG V *saai3*] (negation of universal quantification) (the (b) sentences in (43–46)).

Each participant was presented with a total of eight test trials, with four sentences involving quantification over the object nominal and another four sentences involving quantification over the subject nominal. Among the four object trials, two depicted part-whole relationship in which the object nominal in the form of [CL–N] denotes a specific singular object whose parts can be quantified, and the other two depicted set-subset relationship in which the object nominal in the form of [CL_{PL}–N] denotes a specific plural set of objects whose members can be quantified, as exemplified in (43) and (44) respectively.[7]

(43) a. Zeze **mou5** sik go minbaau.
elder.sister not.have eat CL bun
'Elder sister did not eat the bun.' ("none" reading)

b. Zeze **mou5** sik **saai3** go minbaau.
elder.sister not.have eat all CL bun
'Elder sister did not eat the whole bun.' ("not-all" reading)

[7]In general, the default interpretation of a Cantonese classifier is singular, except for the plural fuzzy classifier *di1*.

(44) a.

Gogo	**mou5**	dam	di	hungmaau.
elder.brother	not.have	throw.away	CL_{PL}	panda

'Elder brother did not throw away the pandas.' ("none" reading)

b.

Gogo	**mou5**	dam	**saai3**	di	hungmaau.
elder.brother	not.have	throw.away	all	CL_{PL}	panda

'Elder brother did not throw away all of the pandas.' ("not-all" reading)

As for the four subject trials, two consisted of stative predicates, as shown in (45), and the other two eventive predicates denoting motion toward a locative endpoint (a directional verb plus a locative noun), as shown in (46).

(45) a.

Di	joengzai	**mou5**	fan	gaau.
CL_{PL}	lamb	not.have	sleep	COG.OBJ

'The lambs did not sleep.' ("none" reading)

b.

Di	joengzai	**mou5**	fan	**saai3**	gaau.
CL_{PL}	lamb	not.have	sleep	all	COG.OBJ

'Not all of the lambs slept.' ("not-all" reading)

(46) a.

Di	zoekzai	**mou5**	fei	soeng	syu.
CL_{PL}	bird.little	not.have	fly	up	tree

'The little birds did not fly onto (the) tree.' ("none" reading)

b.

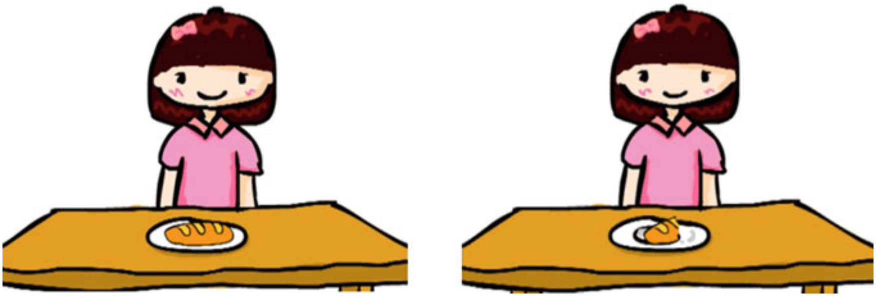
Di	zoekzai	**mou5**	fei	**saai3**	soeng	syu.
CL_{PL}	bird.little	not.have	fly	all	up	tree

'The little birds have not all flown onto (the) tree(s).' ("not-all" reading") =
(i) 'Not all of the little birds flew onto (the) tree(s).'
(ii) 'The little birds have only flown part of the way onto (the) tree(s).'

On each trial, participants were presented with a test sentence paired with two video clips or pictures that were displayed simultaneously side-by-side, one depicting a non-existence or non-realization of an event ("none" reading) (Figs. 1a, 2a, 3a, 4a) and the other depicting a partial realization of an event ("partial/not-all" reading) (Figs. 1b, 2b, 3b and 4b).[8] Situations paired with the two test sentences on subject quantification with stative predicates were presented in static pictures, and situations paired with the other six test sentences, including the four sentences on object quantification (part-whole and set-subset) and the two sentences on subject quantification with eventive predicates were presented in animated video clips.

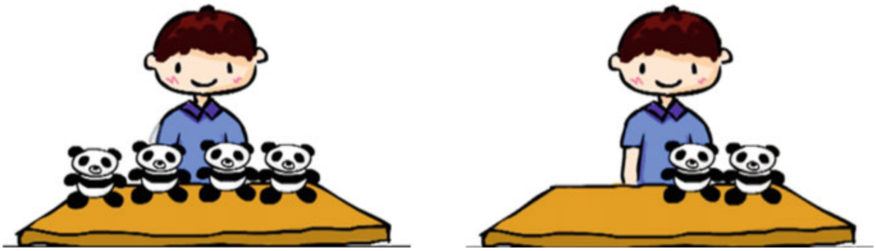
The "none" reading depicts a situation in which no member of the set denoted by the object nominal or the set denoted by the subject nominal is involved in or possesses the property denoted by the verbal predicate. The "partial" reading depicts a situation in which some members of the set denoted by the object nominal or the set denoted by the subject nominal are involved in the action described by the verbal predicate or possess the property denoted by it.

[8]The current experiment examines children's readiness to accept the "not-all" reading. See Lei (2017) for experimental studies on Cantonese-speaking children's scope interpretation of various universal quantifiers interacting with negation.



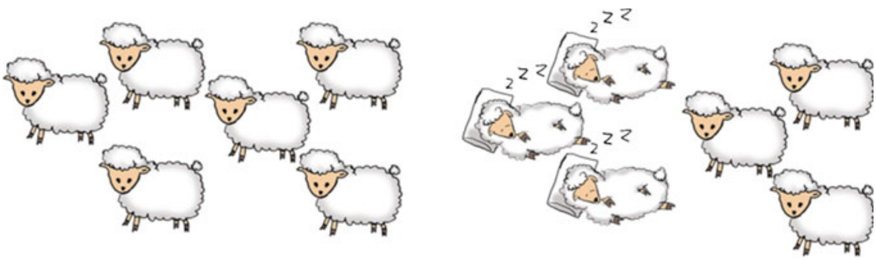
(a) “none” reading **(b)** “partial” reading

Fig. 1 Last frame of videos depicting the two readings paired with sentence (43a) or (43b) on object quantification involving part-whole relationship with [CL–N] object nominal (singular object)



(a) “none” reading **(b)** “partial” reading

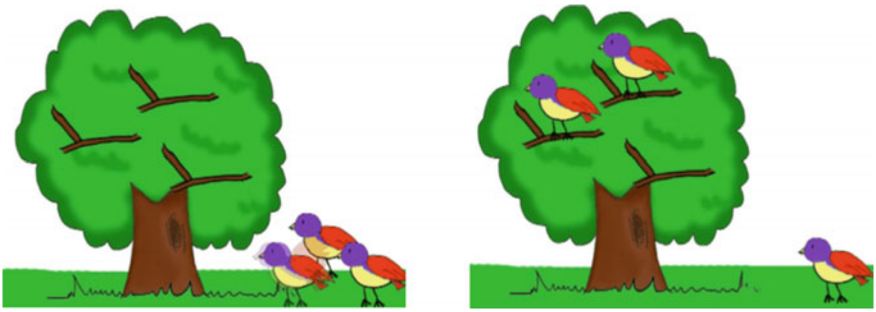
Fig. 2 Last frame of videos depicting the two readings paired with sentence (44a) or (44b) on object quantification involving set-subset relationship with [CL_{PL}–N] object nominal (plural set)



(a) “none” reading **(b)** “partial” reading

Fig. 3 Pictures depicting the two readings paired with sentence (45a) or (45b) on subject quantification consisting of a stative predicate

Participants also received four training trials, four filler trials, and eight control trials testing their understanding of the negator *mou5* ‘not have’ and the postverbal universal quantifier *saai3* (four trials each). The total of 24 test trials were presented in a pseudo-randomized order.



(a) "none" reading **(b)** "partial" reading

Fig. 4 Last frame of videos depicting the two readings paired with sentence (46a) or (46b) on subject quantification consisting of an eventive predicate with a locative complement

4.3 Procedure

In the task, a frog puppet (played by an assistant experimenter) watched the video clips and pictures together with the child. The child was told that the puppet would utter something about what happened in one of the two videos/pictures. The child was asked to indicate which of the videos/pictures best matched with the puppet's description (i.e. the test sentences) by pointing to the corresponding videos/pictures. The left/right position of the videos/pictures was counter-balanced across each sub-type of test sentences. The child participants were tested individually in a classroom at a kindergarten. The whole experiment for each child lasted around 15 min. The adult participants were tested in small groups with the test materials presented in pre-recorded clips. They were asked to write down their answers to the test sentences on an answer sheet.

4.4 Predictions

Since the interpretation of *saai3* under the scope of negation denotes a "not-all" reading, children tested with sentences involving negation of universal quantification (negative sentences with *saai3* in the form of [NEG V *saai3*]) were expected to choose video clips/pictures depicting "partial" events. On the contrary, children tested with sentences involving negation of perfectivity (negative sentences without *saai3* in the form of [NEG V]) were expected to choose the video clips/pictures depicting "none" events.

As reviewed earlier, on picture selection tasks children were found to comprehend the universal quantificational meaning of *saai3* equally well when *saai3* quantifies either a singular object nominal or a plural set of objects. Children were therefore expected to show no difference in their comprehension of the negation of univer-

Table 1 Percentage of group who selected the "none" reading for the [NEG V] sentences (negation of perfectivity)

Group	Sentences involving object-quantification			Sentences involving subject-quantification		
	Part-whole (2 trials)	Set-subset (2 trials)	Overall (4 trials)	Stative (2 trials)	Eventive (2 trials)	Overall (4 trials)
Children (N = 17)	82.4%	85.3%	83.8%	85.3%	76.5%	80.9%
Adults (N = 36)	100%	100%	100%	100%	100%	100%

sal quantification ([NEG V *saai3*]) between the object trials depicting part-whole relationship and the ones depicting set-subset relationship.

Among the two types of subject trials, children were expected to select the "not all" reading of [NEG V *saai3*] sentences with stative predicates (which denote achievement when suffixed by *saai3*) more readily than with eventive predicates (which denote accomplishment when co-occurring with *saai3*), as the interpretation of the latter would involve a potential ambiguity between two possible readings, i.e. object partitivity vs. event partitivity (see (18) and (46)).

In addition, children were expected to show a subject-object asymmetry in their understanding of the relative scope of *saai3* and negation, exhibiting a higher selection rate of the "partial" reading with object trials than with subject trials since subject quantification would involve a discontinuous link between the subject and the quantifier.

4.5 Results

4.5.1 Results on Negation of Perfectivity

For negation of perfectivity, children behaved like the adults in favoring the "none" reading over 80% of the time, both on sentences involving subject quantification (children: 80.9% vs. adults: 100%) as well as those involving object quantification (children: 83.8% vs. adults: 100%), as shown in Table 1.[9] Children did not show any difference in their responses on the two types of object quantification trials (part-whole vs. set-subset). As for trials involving subject quantification, they showed a lower acceptance of the "none" reading with eventive predicates than with stative predicates.

[9]No statistically significant effect on domain of quantification [object quantification vs. subject quantification] was found in the children (Related-Samples Wilcoxon Signed Rank Test, $p = 0.623$). A significant effect was observed on age group [children vs. adults] (Independent-Samples Mann-Whitney U Test, $p < 0.000$).

Table 2 Number (percentage) of participants who consistently selected the "none" reading for the [NEG V] sentences (negation of perfectivity)

Group	Sentences involving object-quantification (four trials)	Sentences involving subject-quantification (four trials)
Children (N = 17)	14 (82.4%)	12 (70.6%)
Adults (N = 36)	36 (100%)	36 (100%)

Table 3 Percentage of group who selected the "partial" reading on [NEG V *saai3*] sentences (negation of universal quantification)

Group	Sentences involving object-quantification			Sentences involving subject-quantification		
	Part-whole (2 trials)	Set-subset (2 trials)	Overall (4 trials)	Stative (2 trials) (%)	Eventive (2 trials)	Overall (4 trials)
Children (N = 17)	76.5%	70.6%	73.5%	35.5%	76.5%	55.9%
Adults (N = 36)	95.8%	100%	97.9%	98.6%	97.2%	97.9%

In terms of consistent individual responses (same acceptance/rejection response in at least three of four trials), 82.4% of the children consistently opted for the "none" reading on sentences involving object quantification but only 70.6% of them did so on sentences involving subject quantification, as shown in Table 2.[10]

4.5.2 Results on Negation of Universal Quantification

For negation of universal quantification, children resembled the adults in favoring the "partial" reading over 70% of the time on *saai3*-sentences involving object quantification (part-whole and set-subset) and those involving subject quantification (eventive predicates only), as shown in Table 3.

Unlike the adults, who indiscriminately favored the "partial" reading across all types of [NEG V *saai3*] sentences testing subject quantification, children gave varying responses according to predicate type. Whereas they chose the "partial" reading 76.5% of the time for eventive predicates, they selected the "partial" reading only 35.5% of the time with stative predicates, favoring the "none" reading instead.[11]

[10] A participant's selection of the "none" reading is considered to be consistent if s/he selected it on at least 3 of the 4 test trials.

[11] Significant effect was found on age group (Independent-Samples Mann-Whitney U Test, $p < 0.000$). No effect on domain of quantification [object quantification versus subject quantification] was observed in children (Related-Samples Wilcoxon Signed Rank Test, $p = 0.073$). Significant effect on predicate type [stative vs. eventive] for subject quantification trials was found in children (Related-Samples Wilcoxon Signed Rank Test, $p = 0.04$).

Table 4 Number (percentage) of participants who consistently selected the "partial" reading for the [NEG V *saai3*] sentences (negation of universal quantification)

Group	Sentences involving object-quantification (four trials)	Sentences involving subject-quantification (four trials)
Children (N = 17)	12 (70.6%)	7 (41.2%)
Adults (N = 36)	36 (100%)	36 (100%)

In terms of consistent individual responses (same acceptance/rejection response in at least three of four trials), a child-adult difference can be observed, as shown in Table 4.[12] While all of the adults opted for the "partial" reading for sentences with *saai3* quantifying the object, only 70.6% of the children (12 out of the 17 children) did so.[13] In addition, while all of the adults consistently selected the "partial" reading when *saai3* is linked to the subject nominal, only 41.2% of the children (i.e. 7 out of the 17 children) gave such a response consistently. Further item-analysis revealed that among the ten children who did not show consistent scope patterns, half of them failed to choose the "partial" reading only on the two trials involving stative predicates; on the other hand, the same children experienced no difficulty in assigning a partial reading on the other two trials involving eventive (motion) predicates, which would permit a degree interpretation.

4.5.3 Children's Differentiation between Negation of Universal Quantification and Negation of Perfectivity

The overall response patterns show that Cantonese-speaking children are able to differentiate between universal quantification and perfectivity under the context of negation. They predominantly favored a "none" reading to negation of perfective aspect, accepting this reading 83.8% of the time for object quantification and 80.9% of the time for subject quantification; but favored a "partial" reading for negation of *saai3*, accepting this reading 73.5% of the time for object quantification and 55.9% for subject quantification. A significant difference is found between children's interpretation of negation of universal quantification and that of perfectivity (Independent-Samples Mann-Whitney U Test, $p < 0.000$ for object-quantification and $p = 0.001$ for subject-quantification).

[12] A participant's selection of the "partial" reading is considered to be consistent if s/he selected it on at least three of the four test trials.

[13] Near-marginal significant effect on domain of quantification [object-quantification vs. subject-quantification] was found in children (Chi-square test of independence, $p = 0.251$).

4.6 Overall Discussion

First, in interpreting sentences involving the negation of perfectivity, children predominantly favored the "none" reading across subject quantification trials and object quantification trials, resembling their adult counterparts (Table 1). The results indicate that children understood the meaning of the negator *mou5* 'not have' in Cantonese as negating a perfective predicate to denote the non-existence or non-realization of an event by the age of three and a half, in spite of the absence of the perfective aspect marker in negative contexts.

A slight subject-object difference can be discerned in these negative sentences (without the presence of *saai3*), with the "none" reading being more accessible when the plural nominal occupies the object position than when it appears as subject (Table 2). This subject-object asymmetry might be accounted for as follows. Children consistently interpret the scope of the negator as covering the whole VP. In sentences involving object quantification, the object nominal falls within the scope of negation, with [NEG VP] signifying non-realization of the event involving the referent of the object as theme or locative goal, resulting naturally in a "none" reading. On the other hand, in sentences involving plural subjects, the non-existence or non-realization of an event denoted by the [NEG VP] predicate may not always be understood as applying exhaustively to all individuals of the set denoted by the definite plural subject (cf. Link 1983, 1998; Gillon 1987). For example, in the sentence *The students have left for their fieldtrip*, one might consider the sentence true even if a couple of the students have not left, as the predicate *have left for their fieldtrip* can still be accepted as being true of the group of students. A "partial" reading is therefore not logically incompatible with sentences involving subject quantification, due to the ambiguity of plural subjects.

Next, in interpreting sentences involving the negation of universal quantification (with the universal quantifier *saai3*), a salient subject-object asymmetry can be observed (Table 3). Children were able to handle the relative scope of the two operators well when the object nominal is quantified, correctly assigning the "partial" reading, whether the quantified elements be parts of a substance or members of a set, even though their adherence to the "partial" reading was not as rigid as the adults (Tables 3 and 4). Since a "partial" reading corresponds to isomorphic scope (NOT>ALL), the response pattern of our children seems to reflect a weaker adherence to the principle of isomorphism in scope interpretation than adults, or a greater readiness to assign inverse scope, a tendency observed in many acquisition studies (Mandarin Chinese: Chien and Wexler 1989; Lee 1991, 2002; Lee and Wu 2013; Lee and Lei 2014, 2015; Hungarian: É. Kiss and Zétényi 2017a, b; Japanese: Goro 2007; among others).

Children, however, find it more difficult to link *saai3* to the subject nominal in the presence of an intervening negator: when subject quantification is concerned in the [NEG V *saai3*] sentences, in which the negator intervenes between the subject nominal and *saai3*, children were only able to select the "partial" reading with the eventive predicates but not the stative predicates (Table 3). To account for the salient difference in children's performance between the [NEG V *saai3*] sentences with stative predicates and those with eventive predicates, one must bear in mind that in our test design, the eventive predicates denote motion toward a locative goal, involving a path being traversed. The negation of the eventive predicates denoting, for example, the act of jumping into the water or flying onto a tree, can be understood in terms of partial coverage of the path. On the other hand, the stative predicates denoting the state of being asleep or being in a couched position are not readily amenable to such a "partial" reading. If one assumes that the child cannot associate the quantifier *saai3* with the subject when an operator intervenes, the child is left with the option of associating the quantifier with a constituent inside the VP. For the predicates denoting motion events, the extent of the path traversed can represent a scale for quantification; for predicates designating states, however, no element within the VP can be quantified by *saai3*, with the result that the child will opt for the "none" reading rather than the "partial" reading. The failure of children to associate the quantifier across the negator echoes the blocking effect of negation observed in Zhou and Crain (2010), thus providing cross-linguistic evidence on children's observance of relativized minimality (Rizzi 1990). It is possible that children's "partial" reading responses for subject quantification in [NEG V *saai3*] sentences containing eventive predicates stemmed in part from a degree interpretation of the path of motion.

5 Conclusions

To conclude, our findings demonstrate that in a language in which universal quantification and perfectivity appear to be indistinguishable in the context of incremental-theme, children do not conflate the two types of linguistic phenomena, and acquire them separately. It is clear that children interpreted the negation of the affixal universal quantifier *saai3* significantly differently from the negation of perfective aspect, showing that they are not interpreting the universal affixal quantifier in terms of perfective aspect. Our findings also show that children are sensitive to the aspect-quantification distinction early on in a language in which aspectual and quantificational notions are both expressed by verbal suffixes.

Children are well aware of the role of the negator *mou5* 'not have' in negating a perfective predicate to denote the non-existence or non-realization of an event, even in the absence of the perfective aspect marker in negative contexts. A slight subject-object difference can be discerned, with the "none" reading being more accessible when the plural nominal occupies the object position than when it appears as subject, reflecting the ambiguity of plural subjects.

In interpreting [NEG V *saai3*] sentences involving object quantification, children show a weaker adherence to the isomorphic "partial" reading than the adults, a phenomenon that has been widely observed in children's interpretation of scope.

As for [NEG V *saai3*] sentences involving subject quantification, children fail to connect the quantifier *saai3* with the subject nominal, with the result that a legitimate interpretation is possible only if an element within the VP is available for *saai3* to quantify, as is the case with motion predicates that allow a degree or extent interpretation. Children's acute sensitivity to the blocking effect of the negator points to the default status of a minimality condition on quantification in language acquisition (Zhou and Crain 2010). The clear subject-object asymmetry as revealed in children's interpretation of the relative scope of negation and *saai3* points to the relevance of verb semantics to A-quantification of the suffixal type.

Acknowledgements We wish to thank the participants of the *Workshop on Linguistics and Cognitive Aspects of Quantification*, the editor and two anonymous reviewers for their suggestions and critical comments. Special thanks are due to members of the Language Acquisition Lab at the Chinese University of Hong Kong (CUHK) for their support on the experimental work. We would also like to thank the principal, teachers, parents and children at the Po Leung Kuk Mrs Tam Wah Ching Kindergarten for their support on our child language experiments, and to CUHK students who took part in the control experiments. The support of a GRF grant "The interpretation of numeral phrases by Chinese-speaking children" (CUHK#447008) to the second author as Principal Investigator is hereby gratefully acknowledged.

References

Au Yeung, Wai Hoo 歐陽偉豪. 1998. Ye Tan Yueyu 'Saai' De Lianghua Biaoxian Tezheng. 也談粵語「晒」的量化表現特徵. *Fangyan* 方言 [Dialect] 1: 58–62.

Cheng, Lisa Lai-Shen, C.-T. James Huang, and C.-C. Jane Tang. 1997. *Negative Particle Questions: A Dialectal Perspective*. Journal of Chinese Linguistics Monograph Series 10.

Cheng, Lisa Lai-Shen, and Rint Sybesma. 1999. Bare and Not-So-Bare Nouns and the Structure of NP. *Linguistic Inquiry* 30 (4): 509–542.

Cheung, Hung-nin Samuel 張洪年. 1989. Yueyu Liangci Yongfa De Yanjiu 粵語量詞用法的研究 [A Study of the Uses of Cantonese Classifiers]. In *Dier Jie Guoji Hanxue Huiyi Lunwenji: Yuyan Yu Wenzi Zu (Xiace)* 第二屆國際漢學會議論文集:語言與文字組(下冊) [Proceedings of the Second International Conference on Chinese Studies: Language and Script Section (2)], ed by Academia Sinica, 753–774. Taipei: Academic Sinica.

Cheung, Hung-nin Samuel 張洪年. (1972) 2007. *Xianggang Yueyu Yufa De Yanjiu* (*Zengding Ban*) 香港粵語語法的研究(增訂版) [Studies on the Grammar of Hong Kong Cantonese (Extended and Updated Edition)]. Hong Kong: The Chinese University Press.

Chien, Yu-Chin, and Kenneth Wexler. 1989. Children's Knowledge of Relative Scope in Chinese. *Papers and Report in Child Language Development* 28: 72–80.

É. Kiss, Katalin, and Tamás Zétényi. 2017a. Quantifier Spreading: Children Misled by Ostensive Cues. Glossa: *A Journal of General Linguistics* 2 (1): 38.

É. Kiss, Katalin, and Tamás Zétényi. 2017b. Why is Children's Interpretation of Doubly Quantified Sentences Non-Isomorphic? *Linguistics* 55 (6): 1337–1382.

Filip, Hana. 1997. Integrating Telicity, Aspect and NP Semantics: The Role of Thematic Structure. In *Formal Approaches to Slavic Linguistics, Volume 3: The College Park Meeting 1994*, ed. Jindrich Toman, 61–96. Ann Arbor, MI: Michigan Slavic Publications.

Filip, Hana. 1999. *Aspect, Eventuality Types and Nominal Reference*. New York: Routledge, Taylor & Francis Group (Garland).

Filip, Hana. 2001. Nominal and Verbal Semantic Structures: Analogies and Interactions. *Language Sciences* 23: 453–501.

Filip, Hana. 2008. Events and Maximalization. In *Theoretical and Cross-Linguistic Approaches to the Semantics of Aspect*, ed. Susan Rothstein, 217–256. Amsterdam: John Benjamins.

Gillon, Brendan S. 1987. The Readings of Plural Noun Phrases in English. *Linguistics and Philosophy* 10: 199–219.

Goro, Takuya. 2007. Language-Specific Constraints on Scope Interpretation in First Language Acquisition. Doctoral Dissertation, University of Maryland, College Park.
Huber, Juliette, and Antoinette Schapper. 2014. The Relationship between Aspect and Universal Quantification: Evidence from Three Papuan Languages from Timor and Alor. In *Number and Quantity in East Nusantara*, ed. Marian Klamer and František Kratochvíl, 152–169. Canberra: Asia-Pacific Linguistics.
Krifka, Manfred. 1992. Thematic Relations as Links between Nominal Reference and Temporal Constitution. In *Lexical Matters*, ed. Ivan Sag and Anna Szabolcsi, 29–53. Chicago: Chicago University Press.
Law, Paul. 2014. The Negation *Mou5* in Guangdong Yue. *Journal of East Asian Linguistics* 23 (3): 267–305.
Lee, Peppina Po-lun. 2004. Affixal Quantification: A Syntax-Semantics Mapping Approach to Cantonese Suffixal Quantifiers. Doctoral Dissertation, City University of Hong Kong.
Lee, Peppina Po-lun. 2012. *Cantonese Particles and Affixal Quantification*. Dordrecht: Springer.
Lee, Thomas Hun-tak. 1991. Linearity as a Scope Principle for Chinese: The Evidence from First Language Acquisition. In *Bridges Between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman*, ed. Donna Jo Napoli and Judy Anne Kegl, 183–206. Hillsdale, NJ: Lawrence Erlbaum.
Lee, Thomas Hun-tak 李行德. 1994. Yueyu *Saai3* De Luoji Tedian 粵語"晒"的邏輯特點 [The Logical Properties of *Saai3* in Cantonese]. In *Proceedings of the First International Conference on Yue Dialects* 第一屆國際粵方言研討會論文集, ed. Chow Yiu Sin, 131–138. Hong Kong: Modern Educational Publishers.
Lee, Thomas Hun-tak. 1995. Postverbal Quantifiers in Cantonese. Paper Presented at the 10th Workshop on Asian Oriental Languages, Centre de Recherches Linguistiques L'Asie Orientale, Paris, May 16–17.
Lee, Thomas Hun-tak. 2001. The Scope of Postverbal Quantifiers: Further Remarks on 'Saai3'. Paper Presented at LSHK Workshop on Cantonese Verbal Complements, Hong Kong Polytechnic University, August 28.
Lee, Thomas Hun-tak. 2002. Two Types of Logical Structure in Child Language. *Journal of Cognitive Science* 3 (2): 155–182.
Lee, Thomas Hun-tak. 2010. Issues of Continuity in Early Lexical and Syntactic Development. Completion Report of General Research Fund (GRF) 2007–2009, Research Grants Council of Hong Kong.
Lee, Thomas Hun-tak, and Margaret Ka-yan Lei. 2014. Scope (Dis)Ambiguity in Chinese Datives: The View from Language Acquisition. In *Proceedings of the 2014 Korean Society for Language and Information Workshop on Meaning and Cognition*. Seoul National University.
Lee, Thomas Hun-tak, and Margaret Ka-yan Lei. 2015. Scope (Dis)Ambiguity in Chinese Datives: The View from Language Acquisition. Invited Lecture Given *at the East Asian Language Acquisition Workshop*, Held in Conjunction with the *2015 International Conference on Language Form and Function*, Soochow University, Suzhou, China, March 27.
Lee, Thomas Hun-tak, and Colleen Wong. 1998. CANCORP—The Hong Kong Cantonese Child Language Corpus. *Cahiers de Linguistique Asie Orientale* 27 (2): 211–228.
Lee, Thomas Hun-tak, and Zhuang Wu. 2013. The Scope of Bare Nouns and Numeral Phrases: An Experimental Study of Child Mandarin. In *Proceedings of the Fourteenth Tokyo Conference on Psycholinguistics*, ed. Yukio Otsu, 137–158. Tokyo: Hituzi Syobo.
Lei, Margaret Ka-yan. 2017. The Acquisition of A-Quantification in Cantonese. Doctoral Dissertation, The Chinese University of Hong Kong.
Leung, Chung-sum 梁仲森. 1980. Guoyueyu Bijiao Yufa Zhaji – Dingzhi Zhishici 國粵語比較語法扎記–定指指示詞 [A Note on Comparative Grammar of Chinese and Cantonese – Definite determiner]. *Zhongying Yuwen Xueyan* 中英語文學研 [Language Learning & Research] 1: 15–28.
Lin, Jo-Wang. 2003. Aspectual Selection and Negation in Mandarin Chinese. *Linguistics* 41 (3): 425–459.

Link, Godehard. 1983. The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretical Approach. In *Meaning, Use and Interpretation of Language*, ed. Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow, 302–323. Berlin: De Gruyer.

Link, Godehard. 1998. Ten Years of Research on Plurals: Where Do We Stand? In *Plurality and Quantification*, ed. Fritz Hamm and Erhard Hinrichs, 19–54. Dordrecht: Kluwer.

Pan, Haihua, and Patricia Man. 1998. A Unified Account of Cantonese 'Saai'. Paper Presented at the 7th Annual Meeting of the International Chinese Linguistics Association, Stanford University, June 26–28.

Partee, Barbara H. 1995. Quantificational Structures and Compositionality. In *Quantification in Natural Languages*, Vol. 1, ed. Emmon Bach, Eloise Jelinek, Angelika Kratzer, and Barbara H. Partee, 541–602. Studies in Linguistics and Philosophy 54. Dordrecht: Kluwer.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.

Schapper, Antoinette. 2010. Bunaq, a Papuan Language of Central Timor. Doctoral Dissertation, Australian National University.

Tang, Sze-Wing. 1996a. On Lexical Quantification. In *UCI Working Papers in Linguistics 1*, ed. Brian Agbayani, Kazue Takeda, and Sze-wing Tang, 119–140. Department of Linguistics, University of California at Irvine.

Tang, Sze-Wing. 1996b. A Role of Lexical Quantifier. *Studies in the Linguistic Sciences* 26 (1/2): 307–323.

Tang, Sze-wing 鄧思穎. 2015. *Yueyu Yufa Jiangyi* 粵語語法講義. Hong Kong: The Commercial Press (Hong Kong) Ltd.

Wang, William S.Y. 1965. Two Aspect Markers in Mandarin. *Language* 41 (3): 457–470.

Wierzbicka, Anna. 1967. On the Semantics of the Verbal Aspect in Polish. In *To Honor Roman Jakobson: Essays on the Occasion of His Seventieth Birthday*, 2231–2249. Den Haag: Mouton.

Wong, Nai-fai. 2008. Scale, Maximality and the Cantonese Particle *Saai3* 晒 'All'. In *Chinese Linguistics in Leipzig*, ed. Redouane Djamouri, Barbara Meisterernst, and Rint Sybesma, 147–162. Paris: Centre de Rrecherches Linguistiques sur l'Asie Orientale, EHESS.

Zhou, Peng, and Stephen Crain. 2010. Focus Identification in Child Mandarin. *Journal of Child Language* 37: 965–1005.

Scalar Implicature or Domain Restriction: How Children Determine the Domain of Numerical Quantifiers

Katalin É. Kiss and Tamás Zétényi

Abstract A sentence containing a numerically modified noun phrase, e.g., *John has read three books for the exam,* can be ambiguous; its numeral can mean 'exactly three' or 'at least three'—depending on whether we talk about John's maximal achievement or on his satisfying some relevant requirement. In the framework of Stanley and Szabó (2000), the 'exactly *n*' and 'at least *n*' readings can be related by domain restriction–domain widening. Children have been found to access the 'at least *n*' interpretation in sentence-picture matching tasks with varying success. The present study tested the assumption that children's success depends on whether they notice the possibility of domain manipulation, which depends on how rigidly fixed the domain of quantification appears to them. We hypothesized that the more flexible, the less clearly demarcated the domain of quantification appears, the easier it will be for a preschooler to relate the 'exactly *n*' and 'at least *n*' readings. In Experiment 1, the quantifier domain was a fixed set represented on a card. In Experiment 2, it consisted of mobile disks, which facilitated the exclusion of the irrelevant elements. In Experiment 3, the quantifier domain was not clearly demarcated in space; it consisted of real objects mixed with objects of other types in a toy box. Children's success rate was below 10% in Experiment 1, it was 36% in Experiment 2, and it raised to 87% in Experiment 3. Experiment 4 tested children's ability to carry out domain restriction/domain widening by an interpretation task involving two seemingly contradictory statements about one and the same visual stimulus.

Keywords Numerically modified NP · 'exactly *n*' reading of numerals
'at least *n*' reading of numerals · Scalar implicature · Domain of quantification
Domain restriction

K. É. Kiss (✉)
Pázmány Péter Catholic University and Research Institute for Linguistics
of the Hungarian Academy of Sciences, Benczúr utca 33, Budapest 1068, Hungary
e-mail: e.kiss.katalin@nytud.mta.hu; ekiss@nytud.hu

T. Zétényi
Department of Ergonomics and Psychology, Budapest University of Technology
and Economics, Magyar Tudósok körútja 2, Budapest 1117, Hungary
e-mail: zetenyi@erg.bme.hu

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_5

1 Introduction

Sentences containing a numerically modified noun phrase, e.g., the sentence *I have read three books for the exam,* can be ambiguous; their numeral *n* can mean 'exactly *n*' or 'at least *n*'. (If, for example, three books represent the minimum requirement, a speaker may say *three books* also when he read more than three.) Children have been found to have problems with accessing the 'at least *n*' interpretation of numerical modifiers. It has been observed that the pragmatic conditions of the experiments play a crucial role in facilitating or blocking the 'at least *n*' reading for children, but generalizations with predictive force have not been formulated.

Our experiments are based on the assumption that the two interpretations depend on how widely the domain of quantification is determined. The 'exactly *n*' and the 'at least *n*' readings of a numerical quantifier can be related by domain restriction—domain widening. The hypothesis that we have tested is that the more rigidly fixed the domain of quantification appears, and the more ostensively this fixed domain is presented to the child, the more difficult it will be for him or her to manipulate it, and to derive the alternative interpretation. The more flexible, the less clearly demarcated the domain of quantification appears, the easier it will be for the child to relate the 'exactly *n*' and 'at least *n*' readings of the quantificational expression.

After briefly summarizing the linguistic and psycholinguistic background of our hypothesis in Sects. 2 and 3, we present four experiments in Sects. 4–7. In the first three experiments, children had to evaluate numerically quantified noun phrases in *if*-clauses against visually represented quantifier domains. We tested whether they judge the *if*-clauses to be true when the cardinality of the visually represented domain is larger than the exact value of the numeral. The linguistic stimulus was of the following type: *If a boy has three hits on the dartboard, he shall get a candy*. In the test cases, the visual stimulus showed a boy with more than three hits. What we were interested in was whether our subjects knew that a boy with four or five hits also has three hits, and deserves a candy. The three experiments differed in the rigidity/flexibility of the domain of quantification with respect to which the numerically modified expression had to be evaluated. In Experiment 1, the quantifier domain was a fixed set represented on a card (e.g., the picture of a dartboard with four hits). In Experiment 2, the quantifier domain to be restricted to a relevant subset consisted of mobile disks, which facilitated their manipulation, and, thereby, made the exclusion of the irrelevant elements easier. In Experiment 3, the quantifier domain to be restricted was not clearly demarcated in space; it consisted of real objects (toys of a certain kind), which were mixed with objects of other types in a toy box. As hypothesized, the less rigidly the domain of quantification was presented to the children, the more easily they could restrict it to the relevant subset that matched the numeral in the linguistic stimulus, i.e., the more easily they could derive the 'at least' interpretation of the numeral.

Experiment 4 aimed to test the psychological plausibility of the assumption that the operation that children perform when determining the relevant domain for the interpretation of a numerical quantifier is domain restriction/domain widening. We presented to the children pictures with two seemingly contradictory sentences, both of which could be true if the quantifier domain was determined appropriately, and we asked them to explain about each sentence why it could be truthfully said about the given picture. E.g., a bear and a giraffe looked at a picture showing four teaspoons and three cups of coffee. The picture induced the bear to say *There are three spoons on the table*, and it induced the giraffe to say *There are four spoons on the table*. We asked the children *Why did the bear say three spoons? And why did the giraffe say four spoons?* The great majority of children answered appropriately, i.e., they realized that four was the number of the spoons shown in the picture, and three was the number of the spoons needed for the three cups of coffee.

2 Linguistic Background

The interpretation of numerical quantifiers, or, more generally, the intepretation of scalar elements, has been the topic of intense semantic and syntactic discussion in the past decades. The issues include, among others, the relation of sentences of type (1a) and (1b):

(1) a. John read five books.
    b. John read three books.

(1a) entails (1b): if John has read five books, he has also read three, because you cannot read five books without reading three books. If (1a) entails (1b), then (1b) can be true in situations which are truthfully described by (1a). In fact, (1b) is true in every situation where John read more than three books. Consequently, the numerical quantifier in (1b) is ambiguous between the 'exactly 3' and 'at least 3' readings (and, naturally, the numerical quantifier of (1a), too, is ambiguous between 'exactly 5' and 'at least 5').

A question raised by the ambiguity of numerically quantified expressions is which of their two meanings is primary and which one is derived, and by what mechanism the derived meaning is obtained.

A long tradition, originating with Horn (1972) and Levinson (1983), analyzes the 'at least *n*' interpretation as the basic meaning. In this approach, the 'exactly *n*' interpretation is inferred; it is a so-called scalar implicature, derived from the basic meaning by means of the Quantity Maxims of Grice (1975)—see Horn (1972), Levinson (1983), Kadmon (1993, 2001), etc.

(2) Maxims of Quantity:

i. Make your contribution as informative as required.
ii. Don't make your contribution more informative than is required.

This approach predicts that in the default case, e.g., out of context, the Maxims of Quantity elicit a scalar implicature, i.e., when (1b) is uttered in isolation, it is assumed to specify the maximum number of books read by John. However, the context can easily block the scalar implicature, resulting in the lower bounded, 'at least *n*', interpretation of the numeral. This is what happens when, for example, the threshold of obtaining a certain benefit is given:

(3) If you have 40 years of service, you can retire with full benefits.

It has been noticed that the 'at least *n*' versus 'exactly *n*' interpretation of a numeral is related to the discourse function of the numerically modified expression—cf. Fretheim (1992), van Kuppevelt (1996), Wedgwood (2005), and É. Kiss (2010). The correlation is clearest in Hungarian, where the discourse function 'focus' is associated with a distinct structural position. In Hungarian, a numerically modified expression is interpreted as 'exactly *n*' in focus position, and as 'at least *n*' out of focus. It is the 'at least *n*' reading of non-focussed expressions that can be restricted to 'exactly *n*' by scalar implicatures in appropriate pragmatic conditions. In the case of a focussed numerically modified expression, the 'exactly *n*' reading is not an option but is obligatory. Compare:

(4) a. János el-olvasott öt könyv-et a vizsgá-ra.
John PRT-read five book-ACC the exam-for
'John read five books for the exam.'

b. János ÖT KÖNYVET olvasott el a vizsgára.
'It was five books that John read for the exam.'

The focussed object in (4b) can only mean 'exactly five books', whereas the object in situ in (4a) tends to be understood as 'at least five books'. The scalar implicature resulting in an 'exactly five books' reading tends not to be activated because the listener infers that if the speaker had 'exactly five books' in mind, she would have used the unambiguous construction in (4b). The meaning difference of the two sentences is particularly clear under negation:

(5) a. János nem olvasott el öt könyvet a vizsgára.
John not read PRT five book-ACC the exam-for
'John did not read five books for the exam.' [='John read less than five books for the exam.']

b. János nem ÖT KÖNYVET olvasott el a vizsgára.
'It wasn't five books that John read for the exam.' [='It was either more or less than five books that John read for the exam.']

In (5a), *five books* means 'five or more books', and its negation is understood as 'less than five books', i.e., the number of books read by John is four, three, two, one, or zero. The focussed *five books* in (5b) means 'exactly five books', and its negation is understood as 'more than five books' or 'less than five books', i.e., four, three, or two books or one book. (The option of 'no book' is excluded by the fact that the focus construction conveys the presupposition that John has read something for the exam.)

These facts of Hungarian have played a role in the discussion of which of the two meanings of a numerically modified expression is basic. They support the view that the basic meaning is the ‘at least *n*’ reading. If *n* means ‘*n*, or $n+1$, or $n+2$, …’, then the ‘exactly *n*’ reading in focus position can be derived from the interaction of ‘*n*, or $n+1$, or $n+2$, …’ and the [+exhaustive/+maximal] feature of structural focus, which excludes the alternatives other than *n* (É. Kiss 2010). (In our experiments, however, we did not exploit the interpretive difference of structural variants—partly because preschoolers had been shown not to be sensitive to the exhaustivity of structural focus (Pintér 2016). In our test sentences, the numerically quantified expressions are not focussed; they occupy postverbal argument positions.)

Whereas the (neo-) Gricean line of reasoning aimed to account for the ‘at least *n*’ and ‘exactly *n*’ readings of numerical modifiers, Carston (1998) also observed a third, ‘at most *n*’ reading, showing up in decreasing contexts, e.g.:

(6) If you have three children, you do not qualify for tax exemptions.

In (6), *three children* is interpreted as ‘three children or fewer’. Carston (1998) accounted for the three potential interpretations of numerical indefinites by assuming that they are underspecified for the ‘at least *n*’, ‘exactly *n*’, and ‘at most *n*’ distinction. The interpretation of a particular occurrence of a numeral is determined by a pragmatic principle, *the presumption of optimal relevance* (Sperber and Wilson 1995, 270). The ‘at least’ and ‘at most’ interpretations tend to arise when the number term is interacting with a modal predicate. When the issue is what is permitted/allowed, the upper limit (i.e., the ‘at most’ interpretation) is relevant, and when the issue is what is required/necessary, the lower limit (i.e., the ‘at least’ interpretation) is relevant.

Criticizing the underspecification view, Geurts (2006) argued for the ambiguity of numerals. He distinguishes quantifier and predicate senses of number words, the former associated with an ‘exactly’ interpretation, and the latter associated with an ‘at least’ interpretation. He claims that the primary meaning is that of an ‘exact’ quantifier, from which the ‘exact’ predicate meaning and the ‘at least’ quantifier meaning are derivable by type-shifting rules. The ‘at most’ reading of numerals is always pragmatically derived.

For Breheny (2008), the only basic meaning of a numeral *n* is the ‘exactly *n*’ meaning; he derives the ‘at least *n*’ reading from the ‘exactly *n*’ reading by pragmatic reasoning. (He argues that there is no compelling evidence for a genuine ‘at most *n*’ reading.) In the case of sentences like (7), the ‘at least’ reading arises via an implication or presupposition based on the background knowledge that tax laws determine the threshold of eligibility for tax benefits. Hence if a person with two dependants is eligible for tax benefits, a person with three or four dependants is also eligible for them.

(7) Everyone who has two children receives tax benefits,

Example (8), uttered in a situation where a person is looking for four chairs (needed for a meeting for instance), is claimed to represent a different case:

(8) There are four chairs in the seminar room.

(8) can be truthfully uttered also when there are more than four chairs in the seminar room. Breheny argues that in the case of (8), the speaker does not mean 'There are four or more chairs in the seminar room'. Rather, what she means is something like 'There are four chairs *for your purpose* in the seminar room'. That is, (8) involves a kind of domain restriction. The interpretation 'four chairs for your purpose' is derived by a pragmatic reflexivization process, as a result of which the NP is understood to refer to a specific collection of four chairs, which may or may not represent a subset of a larger set of chairs. Breheny argues that contexts evoking this type of interpretation are implicitly modal. (In fact, cases like (7) could also be interpreted similarly: what makes a person eligible for tax benefits is having two children. A person with three or more dependants does have the collection of two dependants needed for tax benefits.) Breheny is aware that his pragmatic theory may not work in languages like Hungarian, where the structural position of the numerically modified noun phrase also affects interpretation. He emphasizes that his theory only makes claims about English.

Chierchia et al. (2012), and Spector (2013) show that the 'exactly *n*' interpretation can also appear in embedded contexts immune to pragmatic inferencing, hence it cannot be relegated to pragmatics; it must be derived by grammatical means. The basic meaning of numeral indefinites is the 'at least *n* NP' reading, and the 'exactly *n* NP' is due to a covert exhaustivity operator that numeral indefinites are automatically associated with. Different readings are obtained depending on where the exhaustivity operator is inserted.

Most of the above proposals discuss other types of scalar of quantifiers, as well. The minimal pair in (9), involving the quantifiers *all* and *some,* displays a relation similar to that observed between (1a, b):

(9) a. John read all of the books.
  b. John read some of the books.

(9a) entails (9b), because if John has read all of the books, he has also read some of them; hence (9b) can be true in a situation that is truthfully described by (9a). The (neo-)Gricean approach predicts that out of context, speakers interpret *some n* as 'not every *n*', but judgements can be manipulated by creating relevant contexts. Horn (1992), however, also pointed out differences in the possible implications of sentences with numerical and non-numerical quantifiers.

The hypothesis that we tested experimentally is based on the approach of Stanley and Szabó (2000). Stanley and Szabó frame the context dependence of quantifier interpretation in terms of quantifier domain restriction. They argue that the proposition conveyed by an utterance containing a quantified expression is determined by context as well as the permanent linguistic features of quantified sentences. Stanley and Szabó discuss the pair of examples in (10a, b):

(10) a. Some bottles are empty.
  b. Every bottle is empty.

Both sentences can be uttered in a situation when you look in the cupboard and see that all the three oil bottles in there are empty, although the two vinegar bottles are full. Which of the two sentences is going to be uttered depends on whether the oil bottles have a privileged status for the speaker and the listener, or both oil and vinegar bottles are relevant for them. That is, the difference between the two sentences lies in how the speaker determines the quantifier domain with respect to which the quantifier is evaluated. Stanley and Szabó claim that expressions have values relative to contexts, and utterances communicate propositions relative to contexts. Context is represented in syntax; common nouns like ‘bottle’ occur with a contextual variable (<bottle, f(i)>). The domains that contexts provide for quantifiers are to be treated as intensional entities, represented as functions from worlds and times to sets. However, the set corresponding to the quantifier domain may vary only relative to worlds and times; relative to a context, contextual variables rigidly designate their value. This explains the possibility of cross-sentential anaphora. If example (11) is uttered in the context of preparations for frying some fish in oil, the quantificational domain is restricted to the oil bottles of the cupboard, and the pronoun *they* refers to them[1]:

(11) Every bottle is empty. They should have been thrown out.

We assume that Stanley and Szabó’s approach can be adopted to numerical quantifiers, as well. Consider our examples in (1a, b). Suppose there are three books that must be read for the exam that John is about to take, but there are also further books recommended by the examiner. Suppose John has read the three compulsory books, and two of the recommended ones. In this situation either (1a) or (1b) could be true, depending on how we determine the domain of quantification; whether we only regard the compulsory books representing the threshold requirement relevant, or we want to describe John’s maximal achievement.

Elements of Stanley and Szabó’s (2000) approach reappear in Breheny’s (2008) interpretation of numerically modified noun phrases, as well. However, Breheny relegates the process of domain restriction to pragmatics. Stanley and Szabó’s theory appears to be more appropriate for Hungarian because in their framework, the domain variable of a quantified noun phrase is present in syntax, hence the [+maximal] feature of structural focus provides the domain variable of a focussed quantifier phrase with an upper boundary prior to semantic interpretation; consequently, the obligatory ‘exactly’ interpretation of the quantified NP is predicted.

Following Stanley and Szabó (2000), we hypothesize that the alternative interpretations of numerically quantified noun phrases arise in differently delimited quantificational domains. From this perspective, neither the ‘at least *n*’, nor the ‘exactly *n*’ reading is derived from the other; they are alternatives which arise depending on how the domain of quantification is determined with respect to which the quantifier

[1]The interpretation of the pronoun is derived by the following principle of Neale (1990):

(i) If x is a pronoun that is anaphoric on, but not c-commanded by a non-maximal quantifier ‘[Dx:Fx]’ that occurs in an antecedent clause ‘[Dx:Fx]G(x)’, then x is interpreted as ‘[the x: FX&Gx]’.

is evaluated. Our experiments test whether children are capable of domain widening/domain restriction, and examine what factors facilitate or hinder this operation.

3 Psycholinguistic Antecedents

The question whether speakers (both adults and children) are aware of the implicational relations between sentences like (1a) and (1b), or (9a) and (9b), has been examined in several psycholinguistic experiments. Noveck (2001) studied whether children (from 5 to 10 years of age) and adults employ the Gricean implicatures when interpreting the scalar modals *must/have to* and *might*, and (the French equivalents of) the scalar quantifiers *all* and *some*. He tested whether speakers accept the weak scaler element in contexts which can be adequately described by the stronger alternative, e.g., whether they accept the sentence *Some giraffes have long neck*s when, in fact, all giraffes have long necks. Noveck found that children accept a weak scaler element in contexts allowing a stronger alternative in a much higher proportion than adults. His conclusion is summarized in the title of his paper: he found that "children are more logical than adults", i.e., they are less prone to calculate pragmatic scalar implicatures. He claims that in cognitive development, the Gricean implicatures occur only after logical interpretations have been well established, i.e., logical competence develops before pragmatic competence.

Papafragou and Musolino (2003) investigated whether young children display a genuine inability to derive scalar implicatures, or their difficulties are due to demands imposed by the experimental task, and whether they treat all scalar terms in the same way. Papafragou and Musolino tested five-year-old Greek children and Greek adults on three different scales: *all–some*, *three–two*, and *finish–start*. Their subjects were presented with contexts which satisfied the semantic content of the stronger term on each scale but were described using the weaker term. Nearly 100% of the adults rejected the weaker terms; children's rejection rate, however, was merely 65% in the case of the numerical scale, and it was only around 10% in the case of the other two scales. A second version of the experiment was preceded by a training phase, in the course of which children were taught to correct the weak, imprecise statements of a puppet by more accurate ones. Then the children heard modified versions of the stories of the first experiment, which invited scalar inferences. (The main character in each scenario was involved in a contest, where he was expected to achieve the higher scalar value. He did so, but his achievement was described by the weaker term.) As a result of these manipulations, the rejection rate increased to 93% in the case of the numerical scale, and to around 50% in the case of the other two scales. Papafragou and Musolino found that children's performance with scalar implicatures is determined by the interaction of semantics, contextual effects and task characteristics. They concluded that, in child language at least, cardinals do not have an 'at least' interpretation; children assign to the numerals either an 'exact' or an underspecified reading.

Whereas Noveck (2001) and Papafragou and Musolino (2003) focused on whether children can carry out the scalar implicatures allegedly deriving the 'exactly' reading of scalar elements from their 'at least' interpretation, Musolino (2004)—motivated by Carston (1998)—tested whether children can access the 'at least' and 'at most' interpretations of cardinal numerals. In his first experiment, children could access the 'at least' interpretation of numerals only in 35% of the cases. (This result is comparable to the result of Papafragou and Musolino's Experiment 1, where 65% of the children chose the 'exactly' interpretation of cardinal numerals.) Interestingly, the 'at most' reading proved to be much easier than the 'at least' interpretation; it was accepted in 82.5% of the cases. In a follow-up experiment, the test sentences all appeared in a modal context, where the number of items needed by a participant differed from the actual number of items. The alternative world (different from the actual world) was introduced by the modal verbs *need* and *can* in the linguistic stimulus. For example, the child saw a picture of Goofy with three cookies, while the experimenter said: *Let's see if Goofy* ***can*** *help the Troll. The Troll* ***needs*** *two cookies. Does Goofy have two cookies?* In this condition, the acceptance rate of cardinal numerals in 'at least' contexts increased to 80%.

Hurewitz et al. (2006) compared the acquisition of numerical expressions (*two, four*) and quantificational expressions (*some*, *all*) by 3-year-olds. They found that children interpret the quantifier *some* as 'some or perhaps all'—as predicted by the neo-Gricean view, but they map numerals onto precise numerosities, i.e., they understand *two* as 'exactly two'—as predicted by Breheny (2008) and others. These findings are consistent with experimental results (e.g., Sarnecka and Gelman 2003) that suggest that very young children acquire numbers and quantifiers using different mechanisms.

Huang et al. (2013) made a similar point. They investigated eye movements in the visual-world paradigm to point out the presence of scalar inferencing in the processing of sentences containing *two, three, some* and *all*. The test sentences contained an initial period of semantic ambiguity, which was rapidly resolved by children in the case of *two, three,* and *all*, but involved a temporal lag in the case *some*, which was presumably spent on pragmatic reasoning. The authors concluded that number words are disambiguated on the basis of their lexical semantics (i.e., they mean 'exactly *n*'). In the case of *some*, however, initial processing is limited to the lower-bounded, 'at least' lexical semantics of the quantifier. The temporal lag involved in the upper-bounded interpretation of *some* is evidence of the presence of a scalar implicature.

Gerőcs and Pintér (2014) tested how Hungarian preschool children interpret numerically modified noun phrases. Since Hungarian preschoolers are known to have difficulties with accessing the [+exhaustive/+maximal] feature of structural focus, Gerőcs and Pintér expected them also to have problems with accessing the 'exactly' interpretation of numerals, which is a concomitant of structural focus in Hungarian. Contrary to expectations, they found in their first experiment that children invariably interpreted the numerals as 'exactly *n*', irrespective of the structural position of the numerically modified phrase. In two subsequent experiments, Gerőcs and Pintér tried to elicit the 'at least *n*' interpretation by manipulating the pragmatic environment. Although they created contexts that supported the ‘at least’ reading, and made children personally interested in making pragmatic inferences, they could not significantly reduce the rejection rate of the ‘at least’ interpretation. In their 2nd experiment, the children played a card game with the experimenter. The child had to collect 5 identical cards to win, but she ended up with 6 identical cards. 72% of the children assumed that they were not entitled to the winner’s prize. In a 3rd experiment, the children had to find out whether the main character of each test scenario had the required number of a certain item so as to help a friend in need. Crucially, she had a higher number of items than needed by her friend; e.g., while the friend needed four apples to bake a pie, she had five. 65% of the children concluded that she did not have the number of items needed by the friend. Gerőcs and Pintér interpreted the results as evidence that the default meaning of numerals is ‘exactly *n*’–noting that this assumption also raises questions, e.g., how the ‘at least’ reading is derived by adults in positions other than focus.

The theory of Stanley and Szabó puts these results into a different perspective. In the experiments of Papafragou and Musolino (2003), children’s achievement became more adult-like after the experimenters taught them that they could make more precise statements if they evaluated a scalar term against a more restricted quantifier domain. The confronting of the actual situation with a possible world (that needed by a participant) in the experiment of Musolino (2004) served the same purpose: it emphasized that a certain scalar value can be evaluated with respect to two different quantifier domains. The pragmatic means of Gerőcs and Pintér (motivating the children by personal rewards, and emphasizing the aspect of helping friends), which gave to the ‘at least’ interpretation emotional and moral support, were apparently less successful in evoking the ‘at least’ interpretation.

4 Experiment 1

Based on the theory of Stanley and Szabó (2000), we hypothesized that children can appropriately interpret a numerically quantified expression if they are able to determine the relevant domain against which the numeral has to be evaluated. The more flexible the domain of quantification appears, the more easily children can exclude irrelevant elements, or include relevant ones, i.e., the more easily they can perform domain restriction or domain widening. In previous experiments where children had difficulties with restricting the domain provided visually, e.g., in Experiment 1 of Musolino (2004), or Experiment 1 of Gerőcs and Pintér (2014), the domain of quantification was presented as fixed, and children did not realize that it could be manipulated.

We carried out three experiments with Hungarian children in which the linguistic stimulus had the same syntactic structure; what was varied was the rigidity/flexibility of the quantifier domain. In our first experiment, the visually represented quantifier domain was same type of rigidly fixed domain that was used in Experiment 1 of Gerőcs and Pintér (2014).

4.1 Participants

The experiment was carried out in the big kids' groups of two Budapest kindergartens with 39 children, 22 girls, 17 boys. Their mean age was 6; 4, SD = 4.12 months, age range 65–85 months. The experiment was also repeated with an adult control group, consisting of 50 students of the Budapest University of Technology and Economics (mean age 23 years).

4.2 Procedure

The children were tested individually. The child, the experimenter, and a helper were seated at a table in a quiet room of the kindergarten. The experimenter told the child that they would look at pictures together, and would talk about them. Then he set up a scenario involving 6 actors, who achieved different results in a type of activity. The pictures of the 6 actors were put into a groove. Their achievements were represented on separate pictures, each one placed next to its actor. When the pictures were in place, the helper uttered the linguistic stimulus, which instructed the child to give a reward to the actors if they achieved a certain result. The rewards to be distributed in the different scenarios were candies, stickers, toy ladybirds, and pictures of posies, respectively. Figure 1 shows a scenario of Experiment 1, with a reward given to one of the actors.

When the child handed out her reward(s), the experimenter asked, *Have you finished?* or *What about the rest of the actors?* When the child indicated that she was done, the experimenter put the pictures and rewards to the side of the table, and set up the next scenario. The helper marked the answers on a sheet. The experiment was videorecorded.

4.3 Materials

The experiment involved 8 scenarios, 4 test cases and 4 fillers. In the test cases, the linguistic stimuli contained an introductory sentence giving contextual information, and a complex clause consisting of an *if*-clause with a numerically modified noun phrase, and a main clause with an imperative predicate. The (English translation of the) linguistic stimulus accompanying the set of pictures in Fig. 1 is presented in (12). The other three linguistic stimuli and the visually represented quantifier domains with respect to which they had to be evaluated are shown in (13)–(15) (Figs. 2, 3 and 4).

(12) The kids have been playing darts. If a kid has three hits, he shall get a candy.

(13) The girls are painting eggs for Easter. If a girl has painted 4 eggs, give her a ladybird!

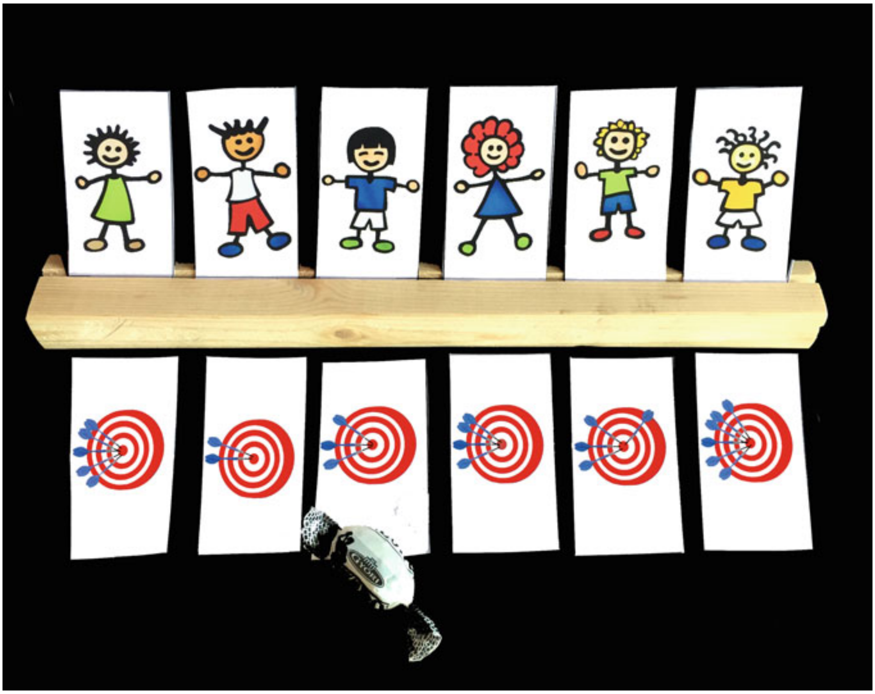


Fig. 1 Picture for example (12)

(14) The children are cleaning the forest. If somebody has collected 2 plastic bottles, she shall get a sticker.

(15) It is Mothers' Day in the kindergarten; mothers have come to the celebration together with their children. If a mom has 3 children, give her a bunch of flowers[2]!

[2] The stimuli in Hungarian:

(i) A gyerekek célbadobást játszottak. Ha egy gyereknek van 3 találata, akkor kapjon egy cukorkát!

(ii) A kislányok tojást festenek Húsvétra. Ha egy kislány megfestett 4 tojást, adj neki egy katicát!

(iii) A gyerekek takarítják az erdőt. Ha valaki összegyűjt 2 eldobott vizespalackot, kapjon egy matricát!

(iv) Anyák Napja van az óvodában; az anyukák eljöttek a gyerekeikkel az ünnepélyre. Ha egy anyukának van 3 gyereke, adj neki egy csokor virágot!

Fig. 2 Picture for example (13)

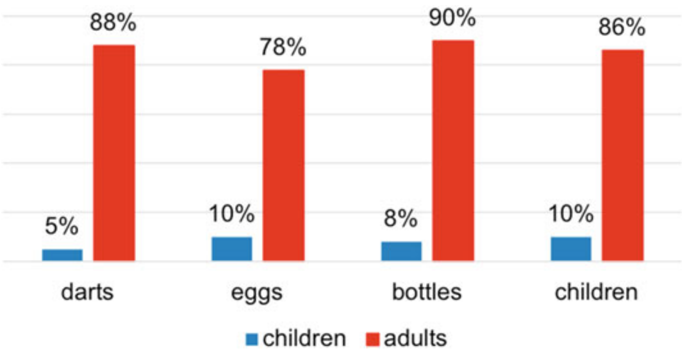
Fig. 3 Picture for example (14)

Fig. 4 Picture for example (15)

In each scenario, the achievement of one actor was of the same cardinality as specified by the linguistic stimulus. I.e., in the case of (12), one child had 3 hits, in the case of (13), one girl painted 4 eggs, in the case of (14), one child collected 2 bottles, and in the case of (15), one mother had three children. Another card in each scenario showed a smaller number of elements than specified by the linguistic stimulus. On the remaining 4 cards, the number of elements was larger than the number in the linguistic stimulus. These were the test cases. The children who interpreted the numeral in the linguistic stimulus as 'exactly *n*', handed out a single reward in each situation. The children who could access the 'at least *n*' interpretation, also rewarded the actors who achieved higher numbers than required, i.e., they handed out 5 rewards per scenario.

The filler cases were similar sentence-picture combinations, involving a group of 6 bears who collected sets of berries, or mushrooms, or both. Children were asked which of the bears collected the most berries; which of the bears collected the least mushrooms; which bear collected as many berries as mushrooms, and whether the berries collected by a certain bear were more or less than the mushrooms collected by the same bear.

Table 1 'At least n' interpretations in Experiment 1



4.4 Results

Of the 39 participants, 2 children interpreted the numeral of the linguistic stimulus as 'at least *n*' (i.e., 2 children handed out 5 rewards) in all the four test scenarios. A third child interpreted the numeral as 'exactly *n*' in the darts scenario (handing out 1 reward), and as 'at least *n*' in the other three scenarios (where he handed out 5 rewards). The proportion of the 'exactly *n*' interpretations was 95% in the darts scenario, and 92% in the other three scenarios. By Spearman's rho, the responses in the four tasks significantly correlated, hence we combined them into one score. We assigned score 1 to responses corresponding to the 'at least *n*' interpretation, involving the distribution of 5 rewards, and we assigned 0 to responses corresponding to the 'exactly *n*' interpretation. The mean score was 0.08 ($SD = 0.27$). The Kruskal-Wallis test showed that those scoring 1 did not differ significantly in age from those scoring 0 (Chi-Square = 0.55, df = 1, $p = 0.457$).

In the adult control group, the proportion of the 'at least *n*' interpretations was 88% in the darts scenario, 78% in the egg-painting scenario, 90% in the bottle-collecting scenario, and 86% in the Mothers' Day scenario. If we assign score 1 to responses corresponding to the 'at least *n*' interpretation, and score 0 to responses corresponding to the 'exactly *n*' interpretation, the adults' mean score was 0.86, SD = 0.27. Only 10% of the adults interpreted the numerals as 'exactly *n*' consistently in every task. The Kruskal-Wallis Test confirmed that the children's scores were highly different from those of the adults (Chi-Square = 55.4, df 1, $p < 0.001$) (Table 1).

4.5 Discussion

In 93% of all the test cases, the children interpreted the numerical modifier in the linguistic stimulus as 'exactly *n*', i.e., they could not access the 'at least *n*' interpretation. From a different perspective: the children evaluated the numerically quantified noun phrase against the quantifier domains shown on the cards. They did not realize that when a certain level of achievement is rewarded, values above that threshold are irrelevant and can be ignored, i.e., the visually shown quantifier domain can be restricted to the threshold value specified in the linguistic stimulus. We attributed their insistence on evaluating the numeral in the linguistic input against the entire visually presented domain to two factors: first, the quantifier domains shown visually had fixed boundaries; second, they were presented ostensively, in a manner suggesting that their boundaries were determined in the given way on purpose; their specific cardinality was somehow relevant for the experiment. The 14% of adults choosing the 'exactly *n*' interpretation must also have thought that the quantifier domains provided by the experimenter were relevant for the purpose of the experiment, and they were, therefore, to be observed.

5 Experiment 2

In Experiment 2 we wanted to test whether it makes the accessing of the 'at least' interpretation of numerically quantified expressions easier for children if the quantifier domain with respect to which the numeral has to be evaluated is manipulatable, i.e., if it consists of loose paper disks collected one by one. We also wanted to decrease the ostensivity of the presentation of the given domain to the child. Hence we embedded the task in a game, where the paper disks were rewards for good answers, i.e., they were collected by the child in quasi-natural circumstances, and their number did not appear to be predetermined by the experimenter.

5.1 Participants

Experiment 2 was carried out with the same children in the same session as Experiment 1. That is, there were 39 participants, mean age 6; 4, SD = 4.12 months.

5.2 Material and Procedure

After completing Experiment 1, the experimenter told the child that they would play riddles. The experimenter would ask the child and the helper in turns, and he would give a red disk for each good answer. (A red dot or red disk is given for good answers and good behavior in Hungarian kindergartens and elementary schools.) The riddles were easy; children could answer most of them. They were of the type *What has a single eye but cannot see? (A needle.)* The experimenter kept on asking a child until she has collected four red disks—as shown in Fig. 5.

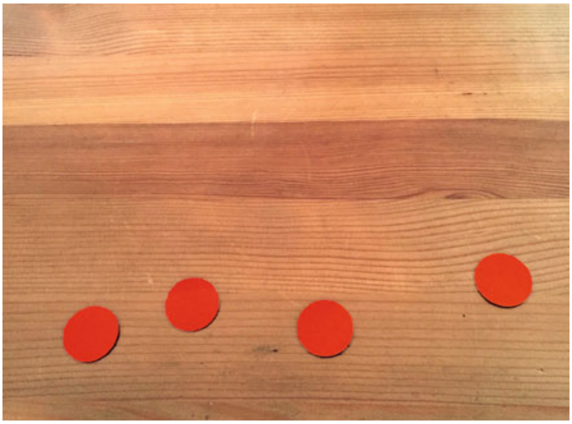


Fig. 5 Stimulus for example (16)

The helper's riddles were somewhat more difficult. She pretended to be able to answer only two, i.e., she collected two red disks. With the child having 4 red disks, and the helper having 2 red disks, the experimenter declared that the game was over. He said that the rule of rewarding was the following:

(16) If somebody has collected three red disks, he or she can take a balloon.[3]

The experimenter put a box of balloons on the table, and asked the helper: "*Do you have three disks?*" "*No, unfortunately, I don't*" she answered. Then he turned to the child: "*Do you have three disks?*" If the child said "*Yes*", she could take a balloon. (Those not taking a balloon were given a sticker.) The helper marked the child's answer on a sheet. The experiment was videorecorded.

5.3 Results

Of the 39 participants, 14 children (36%) said that they had three red disks, and could take a balloon. A frequent answer of the children not taking a balloon was: "*I cannot take one; I only have 4 red disks.*"

Those taking a balloon scored 1, whereas those not taking one scored 0. By the Kruskal-Wallis test, we found a significant difference between the boys and the girls: among the 24 children scoring 1, there were 12 girls and 2 boys. Among the 25 children not taking a balloon, there were 15 boys and 10 girls (Kruskal-Wallis Test Chi-Square 7.43, $df = 1$, $p = 0.006$). There was no significant age difference between those taking a balloon and those not taking one (Kruskal-Wallis test Chi-Square 0.54, $df = 1$, $p = 0.462$).

[3] In Hungarian:

(i) Ha valaki összegyűjtött 3 pirospontot, vehet egy lufit.

5.4 Discussion

The intention of Experiment 2 was to create a situation where the domain of quantification was not so rigid as in Experiment 1. We loosened its rigidity in two respects. Firstly, the elements of the quantifier domain were mobile, which allowed the child to hide or remove the irrelevant fourth element. Secondly, the quantificational domain was not determined by the experimenter but was collected by the child in the course of a game, in quasi-natural circumstances. This decreased its ostensivity, i.e., it decreased the impression that it was relevant that the domain of red disks with respect to which the numeral communicated in the linguistic stimulus was to be evaluated should consist of exactly 4 elements. The children may have thought that the number of elements in the quantifier domain was accidental, or it was under their own control, hence they may have been less inclined to regard it as fixed, unchangeable. It is less clear whether the 'at least' interpretation was helped by the fact that the children were personally motivated (in Experiment 3, where no personal motivation was involved, the acceptance rate of the 'at least' interpretation was much higher).

6 Experiment 3

Experiment 3 completed our series of experiments testing the hypothesis that the less fixed, less rigid the boundaries of a quantifier domain are, the more easily children can restrict it to the domain relevant in the given situation. In Experiment 3, the domain presented by the experimenter, to be restricted so as to match the numerical quantifier occurring in the linguistic stimulus, did not have clear boundaries; its elements were mixed with elements of different kinds. This created a situation where the children were forced to identify the elements of the relevant domain themselves. The elements of the domain were props, which decreased the test-like nature of the situation and helped the athmosphere of playing.

6.1 Participants

The participants of the experiment were the same children that were tested in Experiments 1 and 2. Experiment 3 was performed in a separate session. The results of three children who were not present in both sessions were omitted from the analysis. We tested 39 children, mean age: 6; 4, SD = 4.12 months. Our adult control group was also the same as in Experiment 1; it consisted of 50 university students, mean age 23.

Fig. 6 Stimulus for example (17)

6.2 *Materials and Procedure*

The children were tested individually. The test took place in a quiet room of the kindergarten, in the presence of the experimenter and a helper. The experimenter told the child that they would play with toy animals, who live in the forest. The handyman of the forest is the elephant. If the inhabitants of the forest have a problem, it is the elephant that they call.

Then the experimenter put forward the elephant, and a monkey, who was wearing a shirt with no buttons, and said (in Hungarian):

(17) The monkey was fighting and lost four buttons. If the elephant had four buttons, he could sew on the missing buttons. Does he have four buttons?

The experimenter handed the child the sewing kit of the elephant, which contained 5 identical buttons, 2 spools of thread and 2 thimbles, as shown in Fig. 6. The helper helped the experimenter in setting up the situations, and marked the children's answers (Figs. 7, 8 and 9).

The experiment included the following test scenarios, as well:

(18) The bear drove his car on the curb, and burst both front tyres. If the elephant had two tyres, he could fix the car. Does he have two tyres?

(19) This is the house of the giraffe. It is dark because the lightbulbs burnt out. If the elephant had three lightbulbs, he could replace them. Does he have three lightbulbs?

Fig. 7 Stimulus for example (18)

Fig. 8 Stimulus for example (19)

(20) The bear would like to bake a cake, but he would need three eggs. Here is the elephant coming from the market. If he had three eggs, he could help out the bear. Does he have three eggs?

The experiment was videorecorded.

Fig. 9 Stimulus for example (20)

6.3 Results

The children answered very consistently. The number of *No* answers was 6 in the first task, and 5 in the 2nd, 3rd and 4th tasks.

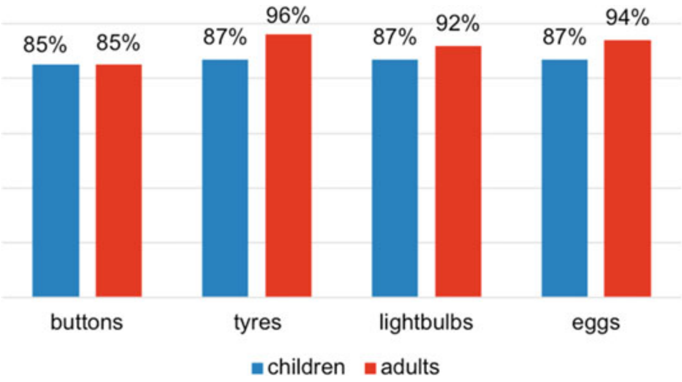
The 33 children who answered *Yes* in all the four tasks, and the child who answered *Yes* in three tasks were all assigned score 1. Hence the number of children scoring 1 was 34 (87%), and the number of children scoring 0 was 5 (13%). The mean score of the children was 0.87, SD = 0.34. The Kruskar-Wallis test did not show any significant difference between those scoring 1 and those scoring 0 either in sex (Chi-Square 0.029, df = 1, p = 0.864), or in age (Chi-Square 0.456, df = 1, p = 0.500) (Table 2).

In the adult control group, the proportion of *Yes* answers, corresponding to the 'at least *n*' interpretation, was 86% in the button sewing scenario, 84% in the tyre replacing scenario, 92% in the bulb replacing scenario, and 94% in the baking-with-eggs scenario. The mean score of the adults was 0.88, SD = 0.33, i.e., the children's and the adults' mean scores did not differ significantly (Kruskal-Wallis Test Chi-Square 0.002, df = 1, p = 0.965).

6.4 Discussion

In the tasks of Experiment 3, 87% of the children were successful in accessing the 'at least' interpretation of the numeral. The tasks were similar to the tasks of Experiment

Table 2 'At least n' interpretations in Experiment 3



1: the children had to interpret situations which required the presence of a certain number of elements, and the available set contained a higher number of elements than the number asked for by the linguistic stimulus. What the children had to decide was whether the requirements of the situation could be satisfied with the available set. On a more abstract level: what they were expected to realize was that the numeral in the linguistic stimulus specified the minimum requirement, i.e., it had to be interpreted as 'at least *n*'. Or from a different perspective: what they had to recognize was that the domain with respect to which the numeral had to be evaluated did not need to coincide with that provided by the visual stimulus; the domain offered by the visual stimulus could be restricted by omitting the irrelevant elements.

The new conditions helped children to act in an adult-like manner. The fact that the quantifier domain was not presented in an ostensively demarcated manner apparently made it clear to them that it is up to them to designate its boundaries. Thus in the presence of a set of identical elements, it is not necessary that all of them be included in the quantifier domain; the irrelevant ones can be omitted.

7 Experiment 4

The aim of Experiment 4 was to test directly the plausibility of the assumption that the operation that we perform when deriving the 'exactly *n*' and 'at least *n*' interpretations of numerical modifiers is domain restriction–domain widening. Apart from Carston (1998), according to whom numerical indefinites are underspecified for the 'at least *n*', 'exactly *n*', and 'at most *n*' distinction, theories of numeral interpretation relate the two readings by complex semantic or pragmatic inferences. According to Horn (1972) and the so-called neo-Gricean approach, the 'at least *n*' reading is primary, and the 'exactly *n*' reading is the result of a pragmatically elicited implicature. Breheny (2008), on the contrary, derives the 'at least *n*' reading from the 'exactly *n*' reading by complex pragmatic reasoning. Geurts (2006) relates the two interpretations, identified by him as the quantifier and predicate senses, by type shifting rules. We assume, instead, that the variable interpretation of a numerically modified expression is the consequence of the variable determination of the quantifier domain with respect to which the given quantified expression is to be evaluated. What we intended to test in Experiment 4 was whether children are able to change perspective; if they can designate different domains to a numerical quantifier in one and the same situation depending on relevance.

7.1 Participants

We performed this experiment with the same children in the same kindergartens as Experiments 1–3. I.e., we tested 39 children, mean age: 6; 4, SD = 4.12 months. The participants of the adult control group were also the same 50 university students (mean age 23) as in Experiments 1 and 3. Experiment 4 was carried out in the same session as Experiment 3; it immediately preceded Experiment 3.

7.2 Materials, Procedure

The children were tested individually in a quiet room of the kindergarten. The experimenter explained that they would look at pictures on the laptop in front of them. Two puppets, a bear and a giraffe, would also look at the pictures, and would say what they saw in them. It can happen that they say contradictory sentences about the same picture, and nevertheless, both of them can be right because they pay attention to different things. Then an example followed: a picture showing a row of houses, with flowers in the windows of one of the houses. The bear said (in the voice of the experimenter): "*There are flowers in every window*." The giraffe contradicted him (in the voice of the helper): "*There are flowers in some windows.*" "*Can it be the case that both are telling the truth?*"—asked the experimenter. "*Yes*," answered the helper. "*The bear was talking about his own house, which has flowers in every window. The giraffe was talking about all the houses in the street.*"

The experimenter explained that they would see further pictures, with different comments by the bear and the giraffe, and the child is expected to explain why they said what they said. The experiment consisted of three test cases involving numerically quantified expressions, listed in (21)–(23), and three fillers involving quantified expressions containing *some* and *all*.[4] The scenarios are listed in the order of their presentation, however, the order of the (a) sentences, requiring the 'exactly'

[4] A 4th test example had to be omitted. The picture contained empty and filled glasses on a silver tray. Because of reflections on the silver and on the glasses, it turned out to be very hard to interpret visually.

Fig. 10 Stimulus for example (21)

interpretations, and the (b) sentences, requiring the 'at least' interpretation, was varied (Figs. 10, 11 and 12).

(21) Giraffe: There are five eggs on the table.
Bear: There are three eggs on the table.
Experimenter: Why did the giraffe say *five eggs*? Why did the bear say *three eggs*?

(22) Giraffe: The boy has five socks.
Bear: The boy has two socks.
Experimenter: Why did the giraffe say *five*? Why did the bear say *two*?

(23) Bear: There are four spoons on the table.
Giraffe: There are three spoons on the table.
Experimenter: Why did the bear say *four*? Why did the giraffe say *three*?

The fillers tested the children's ability to adjust the quantifier domains of a different pair of quantifiers, the Hungarian equivalents of *some* and *all*. They were of the following type:

(24) Picture: a table with a vase containing white tulips, and a flower box containing red geraniums.
Bear: All flowers are white.
Giraffe: Some flowers are white.
Experimenter: What was the bear talking about? What was the giraffe talking about?

The filler tasks served as a pilot for a later, more detailed investigation of how children interpret the quantifiers *minden* 'all' and *néhány* 'some', and how they relate them to each other. What is clear at this point is that the interpretation of *some* requires further testing because many children seem to understand it as 'few'.

Fig. 11 Stimulus for example (22)

Fig. 12 Stimulus for example (23)

7.3 Results

Answers to the questions "Why did the bear say number x?" and "Why did the giraffe say number y?" (i.e., answers involving the maximal and the restricted interpreta-

Table 3 Correct interpretations of the restricted and the maximal domains in Experiment 4

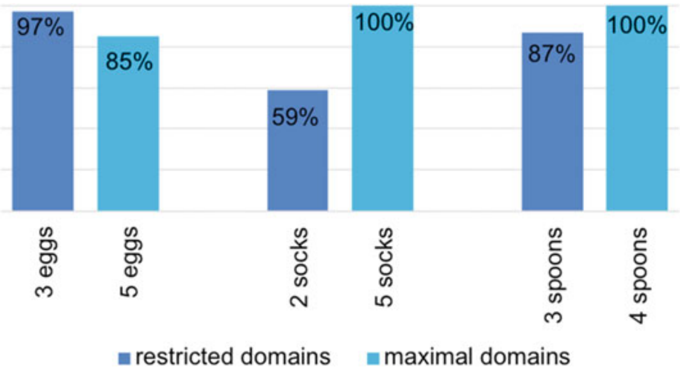
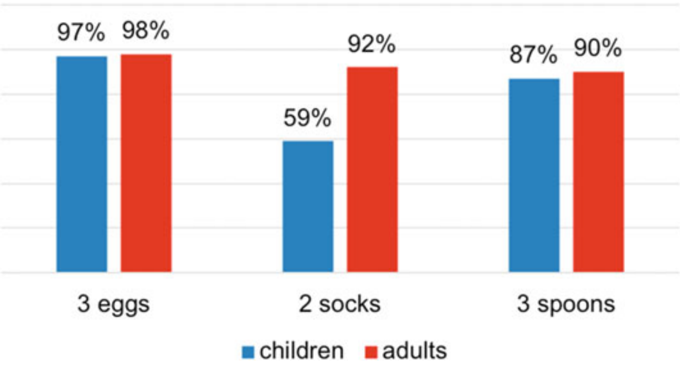


Table 4 Participants accessing the restricted quantificational domains in Experiment 4



tions of the quantificational domain in one and the same stimulus) were evaluated separately. If the child could explain why the given number of items could be relevant for the puppet making the comment, she scored 1. Scores, i.e., correct answers about picture (21), for example, were of the following type: "The giraffe said *five eggs* because he counted all the eggs on the table." "The bear said *three eggs* because he was talking about the red eggs." If the child answered "I don't know", or said that the puppet making the comment was wrong, she scored 0. The percentages of correct answers are shown in Table 3:

The adults had no problem with accessing either one of the two interpretations. In the Socks task, where 41% of the children could not access the 'at least *n*' interpretation, they were significantly more successful than the children (Kruskal-Wallis test Chi-Square 12.84, df = 1, $p < 0.001$) (Table 4).

7.4 Discussion

Interestingly, in two cases (22), (23), the 'exactly' interpretations proved to be easier, whereas in one case (21), more children could explain the 'at least' interpretation. This fact supports the assumption that neither of the two interpretations is inherently primary, with the other representing the derived reading, but, as suggested by the theory of Stanley and Szabó (2000), both interpretations can be accessed with the same mechanism, and the choice between them is determined by contextual, pragmatic factors.

The ease of accessing the different interpretations clearly depended on the salience of the relevant quantifier domains. In the case of the scenario represented in (21), the red subset of eggs was more salient than the totality of eggs on the table, that was why more children found true the sentence referring to the three red eggs than the sentence referring to the totality of five eggs. In the scenarios represented in (22) and (23), the 'at least' interpretation, i.e., the narrower quantifier domain, proved to be harder to access because the relevant subset with respect to which the numeral was to be evaluated had no distinguishing physical properties; it could only be identified on the basis of relevance. Domain restriction was easier in the case of (23), where the three salient cups of coffe on saucers with lumps of sugar strongly invoked the absence of three teaspoons. Example (22) proved to be the most difficult; it was not immediately obvious to the majority of children why the bear says that the boy has 2 socks when there are clearly 5 socks on the chair. The experimenter helped the children thinking long without answering with the following question: *Look, what is the boy in the picture doing?* The recognition that the boy was dressing made 59% of the children realize that for the purpose of dressing up, only 2 socks were relevant. These children answered the question *Why did the bear say two socks?* with versions of this sentence: *Because the boy has two feet*.

In sum: Children are capable of associating one and the same situation with two sentences containing different numerical quantifiers, and evaluating the quantifiers with respect to two different quantifier domains. This experiment has shown that neither the 'at last *n*', nor the 'exactly *n*' interpretation of numerical quantifiers is inherently primary with respect to the other. The interpretation that children first hit upon is that which is visually more salient; it is the less salient interpretation that is derived by domain widening or domain restriction. The ease of deriving the secondary interpretation depends on how obvious its relevance to the listener is, and how salient the cues to its relevance are. For our subjects, three cups and saucers occupying most of the picture strongly suggested that only 3 of the 4 teaspoons were necessary. In the picture of the dressing boy, however, the two little naked feet were partly hidden by the trousers, so the need for two socks was less salient.

8 General Discussion

The psyholinguistic testing of how children (and adults) interpret numerical quantifiers has largely been motivated by linguistic debates on the semantics of numerals; whether the 'at least' or the 'exactly' interpretation is primary, and how the secondary reading is derived. The results of the psycholinguistic experiments of Papafragou and Musolino (2003) and Musolino (2004) had a significant role in the fact that the neo-Gricean theory has lost ground (see, e.g., the discussion of the issue in Szabolcsi's monograph Szabolcsi 2010: 148). In view of the result that nearly 100% of children could process the 'exactly *n*' reading of numerals, but only 65% of them could access the 'at least *n*' reading (Papafragou and Musolino 2003), it seemed to be a legitimate conclusion that the 'exactly' reading is the default reading accessible to everyone, and the 'at least' interpretation is the result of pragmatic reasoning. The fact that among modified experimental conditions involving training, the acceptance rate of the 'at least' reading could be raised to 90% appeared to be compatible with this conclusion; it seemed to indicate that the 'exactly' reading is accessible without any context, whereas the 'at least' interpretation can only be evoked by appropriate pragmatic conditions.

However, the emerging alternative theories predicting the primacy of the 'exactly' reading cannot account for all the facts attested, either. If the primary meaning of numerical quantifiers is the 'exactly *n*' reading, and the 'at least *n*' interpretation is the result of pragmatic reasoning, as assumed by Horn (1992, 1996), Breheny (2008), and others, it is unclear how the two readings can be associated with different structural positions in the Hungarian sentence. Breheny (2008:136) has come to the implausible conclusion that "it may be that there is a different system in Hungarian and other languages for realizing the *at least* and *exactly* readings". The approach of Chierchia et al. (2012), which derives the two readings in grammar instead of pragmatics, would presumably not have any problem with the Hungarian facts; however, by assuming an invisible operator inserted randomly at random positions in the sentence, this model may put too heavy a load on young children interpreting numerically modified noun phrases. A further problem is that it does not explain how the experimental conditions affect children's choice of interpretation.

We have hypothesized that the theory which can account for both the linguistic and the psycholinguistic aspects of the interpretation of numerical quantifiers is the theory of Stanley and Szabó (2000). Whereas this theory was originally formulated in order to explain the relation of quantifier expressions containing *some* and *all*, it appears to be applicable to other scalar elements, as well. In this approach, a numerical quantifier has no primary reading. It is the domain with respect to which the numerical quantifier is evaluated that can be determined in various ways, depending on what is relevant for the participants of the discourse. Suppose a student has scored 90 points in a 100-point test where you have to score 80 to get an A, and 60 to pass. His achievement can be described in different ways depending on how we identify the quantifier domain, which, in turn, depends on what we regard as relevant. If the relevant domain is the number of the scores he made, we say *He scored 90* (which yields the *exactly* interpretation of the numeral). If we are interested in whether he has got an A, we can say *He scored 80*, and if we are interested in whether he has passed, we can say *He scored 60* (both of which represent *at least* interpretations). In this approach, the different readings of a numerical quantifier can be related by domain restriction—domain widening. As our Experiment 4 has shown, this operation is also accessible to children. The ease with which children can identify the quantifier domain assumed by the speaker depends on how salient the domain is. If one of the potential quantifier domains is visible in the stimulus, whereas the other one is an 'intensional' domain, relevant for an intended, future action, the visible domain is primary for children. It also matters how salient the cues forcasting the intended future action are. (The presence of three cups and saucers proved to be a fairly obvious indicator of the relevance of three teaspoons. A dressing scenario evoked the relevance of two socks for considerably fewer children). If both the narrower and the wider domain are visible, i.e., when they are in a subset-superset relation, either one of them can be primary, depending on their salience. (In the egg scenario of our Experiment 4, the subset of painted eggs was a more easily accessible quantifier domain than the totality of eggs.)

The framework proposed by Stanley and Szabó (2000) is compatible with the correlation between structural position and interpretation attested in Hungarian (see É. Kiss 2010). The Hungarian structural focus is interpreted as the unique maximal individual for which the predicate holds (Kenesei 1986, Szabolcsi 1994, Bende-Farkas 2008). This property of focus excludes the possibility of alternative domains where the focus referent is non-maximal, i.e., it excludes the possibility of an 'at least *n*' reading. The fact that a non-focussed numerical quantifier can elicit the 'at least' interpretation without any context may be related to the fact that the 'exactly'/'maximal' reading has grammaticalized as an obligatory concomitant of the preverbal focus position of the Hungarian sentence. If the 'maximal' reading has a grammaticalized syntactic form, speakers tend to conlcude that the non-use of this construction is motivated by the lack of maximality. An aswer to a wh-question is also interpreted as non-maximal if the constituent corrresponding to the wh-phrase is not focussed. Compare:

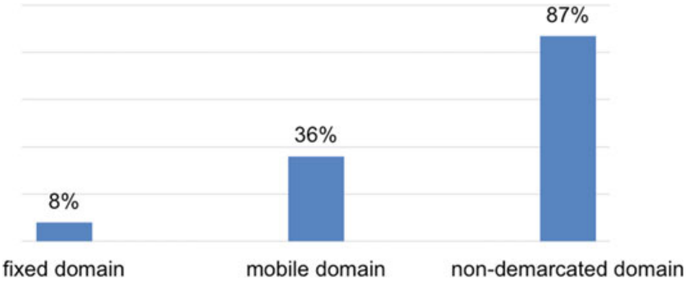
(25) Kit hívtál meg?
who.ACC invited.2SG PRT
'Who did you invite?'

a. Pétert hívtam meg.
Peter.ACC invited.1SG PRT
'I invited Peter [and noone else].'

b. Meg-hívtam Pétert.
'I invited Peter [among others].'

Whereas our Experiment 4 demonstrated that children are capable of interpreting a numerical quantifier on variable domains, Experiments 1–3 tested what blocks domain manipulation in various test situations, including situations where they would

Table 5 Children's rate of 'at least n' interpretations in Experiments 1, 2, and 3



personally benefit from it. Our hypothesis has been that if a quantifier domain is presented to the child as a fixed artificial entity created by the experimenter, the child assumes that she has to evaluate the test sentence with respect to the given domain. She believes that the domain is relevant for the purposes of the test as it is; i.e., she does not interpret relevance from the point of view of the scenario that the stimulus intends to simulate. Our Experiments 2 and 3 have demonstrated that the less rigidly fixed, the less artificial-looking the quantifier domain appears, the more children will realize that they are free to manipulate it; that they can restrict it to a subset according to the needs of the situation. By representing the quantifier domain as a set of mobile elements, the rate of 'at least *n*' interpretations could be raised from 10% to 36%, and by using a domain not clearly demarcated in space, it could be further raised to 87% (Table 5):

In view of these results, data indicating that children are incapable of accessing the 'at least' interpretation of numerals—e.g. the results of our Experiment 1, Experiment 1 of Musolino (2004), and the experiments of Gerőcs and Pintér (2014)—are consequences of misleading experimental conditions, This conclusion relates our work to the line of research showing that the experimental conditions may produce results in language acquisition experiments that are not attested in everyday circumstances, first of all to the work of Crain and Thornton (1998). The results of our Experiments 1–3 also converge with the findings of É. Kiss and Zétényi (2017), suggesting that the misleading effect of experimental conditions derives from the increased ostensivity of experimental stimuli, and from children's sensitivity to ostension (Csibra and Gergely 2009, 2011). As argued by Csibra and Gergely, young children are predisposed to attribute high relevance and great importance to information presented to them by adults in an ostensive manner. This fact has a crucial role in the transmittance of information from generation to generation. It is this communication system, called 'Natural Pedagogy', that makes possible the fast and efficient social learning of cultural knowledge that would be hard to acquire based on observational learning mechanisms alone (Csibra and Gergely 2011: 1149). A quantifier domain presented by the experimenter as a fixed set painted on a card is an ostensive cue which children feel obliged to observe. The more natural conditions we can create in acquisition experiments, the more reliable our results will be.

Acknowledgements This research was carried ont in the framework of project 108951 of OTKA, the Hungarian National Scientific Research Foundation. We owe thanks to the Budapest kindergartens Betlehem, Bükköny, and Táltos. We are grateful to the anonymous reviewers of our paper for their useful advice.

References

Bende-Farkas, Á. 2008. Adverbs of Quantification, *It*-clefts and Hungarian Focus. In *Adverbs and Adverbial Adjuncts at the Interfaces,* ed. K.É. Kiss. Berlin: Mouton de Gruyter.

Breheny, R. 2008. A New Look at the Semantics and Pragmatics of Numerically Modified Noun Phrases. *Journal of Semantics* 25: 93–139.

Carston, R. 1998. Informativeness, Relevance and Scalar Implicature. In *Relevance Theory: Applications and Implications*, ed. R. Carston, and S. Uchida. Amsterdam: John Benjamins.

Chierchia, G., D. Fox, and B. Spector. 2012. The Grammatical View of Scalar Implicatures and the Relationship Between Semantics and Pragmatics. In *Semantics: An International Handbook of Natural Language Meaning*, vol. 2, ed. C. Maienborn, K. von Heusinger, and P. Portner, 2297–2332. Berlin: Mouton de Gruyter.

Crain, S., and R. Thornton. 1998. *Investigations in Universal Grammar: A Guide to Research on the Acquisition of Syntax and Semantics*. Cambridge, MA: MIT Press.

Csibra, G., and Gy Gergely. 2009. Natural Pedagogy. *Trends in Cognitive Sciences* 13: 148–153.

Csibra, G., and Gy Gergely. 2011. Natural Pedagogy as Evolutionary Adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366: 1149–1157.

É. Kiss, Katalin. 2010. Structural Focus and Exhaustivity. In *Information Structure. Theoretical, Typological and Experimental Perspectives*, ed. M. Zimmermann, and C. Féry, 64–88. Oxford: Oxford University Press.

É. Kiss, Katalin, and T. Zétényi. 2017. Quantifier Spreading: Children Misled by Ostensive Clues. *Linguistics* 55 (6): 1337–1381.

Fretheim, T. 1992. The Effect of Intonation on a Type of Scalar Implicature. *Journal of Pragmatics* 18: 1–30.

Gerőcs, M., and L. Pintér. 2014. How do Hungarian Preschoolers Interpret Number Words? In *ConSOLE XXI: Proceedings of the 21st Conference of the Student Organization of Linguistics in Europe* (8–10 January 2013, Potsdam), ed. M. Kohlberger, K. Bellamy, and E. Dutton, 104–122. Leiden: Leiden University Centre for Linguistics.

Geurts, B. 2006. Take 'five'. The Meaning and Use of a Number Word. In *Indefiniteness and Plurality*, ed. L. Tasmowski, and S. Vogeleer, 311–329. Amsterdam: John Benjamins.

Grice, H.P. 1975. Logic and Conversation. In *Syntax and Semantics 3: Speech Acts*, ed. P. Cole, and J.L. Morgan, 41–58. New York: Academic Press.

Horn, L.R. 1972. On the Semantic Properties of Logical Operators in English, PhD Dissertation, UCLA.

Horn, L.R. 1992. The Said and the Unsaid. In *SALT II Proceedings, Ohio State University Working Papers in Linguistics 40*, ed. C. Barker, and D. Dowty, 163–192.

Horn, L.R. 1996. Presupposition and Implicature. In *Handbook of Contemporary Linguistic Theory*, ed. S. Lappin, 299–319. Oxford: Blackwel.

Huang, Y.T., E. Spelke, and J. Snedeker. 2013. What Exactly do Numbers Mean? *Language Learning and Development* 9 (2): 105–129.

Hurewitz, F., A. Papafragou, and L. Gleitman. 2006. Asymmetries in the Acquisition of Numbers and Auantifiers. *Language Learning and Development* 2: 77–96.

Kadmon, N. 1993. *On Unique and Non-unique Reference and Asymmetric Quantification*. New York: Garland.

Kadmon, N. 2001. *Formal Pragmatics*. Oxford: Blackwell.

Kenesei, I. 1986. On the Logic of Word Order in Hungarian. In *Topic, Focus, and Configurationality*, ed. W. Abraham, and S. de Meij, 143–159. Amsterdam: John Benjamins.

Levinson, S.C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Musolino, J. 2004. The Semantics and Acquisition of Number Words: Integrating Linguistic and Developmental Perspectives. *Cognition* 93: 1–41.

Neale, S. 1990. *Descriptions*. Cambridge, MA: MIT Press.

Noveck, I. 2001. When Children are More Logical than Adults: Experimental Investigations of Scalar Implicature. *Cognition* 78: 165–188.

Papafragou, A., and J. Musolino. 2003. Scalar Implicatures: Experiments at the Syntax Semantics Interface. *Cognition* 86: 253–282.

Pintér, L. 2016. Preschoolers' Interpretation of the Focus Particle *Csak* 'only' in Hungarian. In *ConSOLE XXIII: Proceedings of the 23rd Conference of the Student Organization of Linguistics in Europe*, ed. K. Bellamy, E. Karvovskaya, M. Kohlberger, and G. Saad, 200–220. Leiden: Leiden University Centre for Linguistics.

Sarnecka, B., and S. Gelman. 2003. Six Does not Just Mean a Lot: Preschoolers see Number Words as Specific. *Cognition* 92: 329–352.

Spector, B. 2013. Bare Numerals and Scalar Implicatures. *Language and Linguistics Compass* 7: 273–294.

Sperber, D., and D. Wilson. 1995. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stanley, J., and Z.G. Szabó. 2000. On Quantifier Domain Restriction. *Mind and Language* 15: 219–261.

Szabolcsi, A. 1994. All Quantifiers are not Equal: The Case of Focus. *Acta Linguistica Hungarica* 42: 171–187.

Szabolcsi, A. 2010. *Quantification*. Cambridge: Cambridge University Press.

van Kuppevelt, J. 1996. Inferring from Topics. *Linguistics and Philosophy* 19: 393–443.

Wedgwood, D. 2005. *Shifting the Focus. From Static Structures to the Dynamics of Interpretation*. Amsterdam: Elsevier.

Universal Quantification and Distributive Marking in Serbian

Nataša Knežević and Hamida Demirdache

Abstract This paper discusses experimental evidence bearing on the theoretical analysis of so-called distributive-share markers, such as Serbian *po*, as opposed to distributive-key markers, such as Serbian *svaki* or English *every*. Based on the analysis of distributive *po* as a marker of event plurality, which forces distributivity, but does not involve universal quantification and therefore atomic and exhaustive distribution, contrary to *svaki (every)* (Knežević 2015), we tested the following hypothesis: in Serbian, only sentences with both *svaki* and *po* will yield obligatory atomic-distributive and exhaustively distributive readings. A total of 98 children were tested, separated in three age groups, between the ages 4;3 and 6;9 (N = 22, MA = 5;0, SD = 0.8), between the ages 6;11 and 8;1 (N = 38, MA = 7;2, SD = 0.5), and between the ages 8;7 and 11;0 (N = 37, MA = 9;4, SD = 0.7), as well as the control group of Serbian monolingual adults (N = 31, f = 21, m = 10, MA = 27;3, SD = 8.6). Our results suggest the following. First, that the exhaustivity requirement on universal quantification is acquired before the atomicity requirement. Second, that children acquire the truth conditions of the distributive-share marker *po* prior to those of the universal quantifier *svaki* (*every*). On Knežević's (2015) proposal, this would mean that in languages that have both pluractional markers and universal quantifiers, such as Serbian, children acquire pluractionals before universal quantifiers.

Keywords Distributive-key · Distributive-share · Event-distribution
Participant-distribution · (non)-atomic distributive · Collective
Exhaustivity requirement · Atomicity requirement · Pluractionality · Serbian

N. Knežević (✉) · H. Demirdache
LLING UMR 6310 CNRS/Université de Nantes, Chemin La Censive Du Tertre BP 81227, 44312 Nantes Cedex 3, France
e-mail: Natasa.Knezevic@univ-nantes.fr

H. Demirdache
e-mail: Hamida.Demirdache@univ-nantes.fr

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_6

1 Introduction

This paper discusses experimental evidence bearing on the theoretical analysis of so-called distributive-share markers, such as Serbian *po*, as opposed to distributive-key markers, such as Serbian *svaki* or English *every*. The experimental findings reported here ultimately bear on two correlated fundamental questions. What are the core differences between distributive-key versus distributive-share marking? To what extent can distributive-share markers be analyzed as instances of universal quantification?

The context for our discussion is Knežević's (2015) proposal that the Serbian distributive-share marker *po* is not a universal quantifier (quantifying either over participants or events/spatiotemporal locations), but an event plurality marker (a pluractional), enforcing distributivity over spatiotemporal locations, but without imposing either atomic or exhaustive distribution. On this proposal, distributive-share markers crucially differ from universal quantifiers (e.g. *svaki* in Serbian), which enforce atomic and exhaustive distribution.

We start by reviewing the properties of distributive-share markers as opposed to distributive-key markers (Sect. 2), as well as the previous experimental findings bearing on these issues, from Knežević (2015), to our knowledge, the only experimental study of distributive-share markers in the literature (Sect. 3.1).[1] The adult results confirm that *po* is strongly distributive (blocking collective readings), and yields both atomic and non-atomic distributive construals. In contrast, for children (aged 7), *po* is not strongly, but weakly distributive. That is, although children accept *po*-sentences in both atomic and non-atomic distributive contexts, they crucially, unlike adults, accept it also in collective contexts. In order to further probe adult versus children's understanding of atomic versus non-atomic distribution, we designed an experimental study investigating comprehension of Serbian sentences with the universal quantifier *svaki* (*every*) in subject position and the distributive marker *po* in object position (Sect. 3.2). One of our expectations was that at least adults would no longer accept non-atomic distributive construals. This prediction was indeed verified, but the experiment also revealed two important acquisition findings (Sects. 3.3. and 3.4). First, that the exhaustivity requirement on universal quantification is acquired before the atomicity requirement. Second, that children acquire the truth conditions of the distributive-share marker *po* prior to those of the universal quantifier *svaki* (*every*). On Knežević's (2015) proposal, this would mean that in languages that have both pluractional markers and universal quantifiers, such as Serbian, children acquire pluractionals before universal quantifiers.

[1]For ongoing experimental research on Serbian *po,* see Bosnić et al. (2016) and Knežević et al. (2016) and references therein. For research on the interaction of *po* with a universal quantifier, see Sekerina and Sauermann (2017) for Russian and Dotlačil (to appear) for Czech (Sect. 3.4.4).

2 Theoretical Background

2.1 *Distributive-Key Versus Distributive-Share Markers*

In Choe (1987) and subsequent relational distributivity approaches (Oh (2001), Zimmermann (2002)), distributivity is a relation between a distributive-key and a distributive-share. The distributive-key denotes the event participant over which the distribution takes place, while the distributive-share denotes the entity that is being distributed (over the distributive-key). To illustrate, consider (1):

(1) a.	Devojke	farbaju	kutiju.
	girl-NOM.F.PL	paint-3.PL	box-ACC.F.SG
b.	The girls are	painting a box.	

The Serbian sentence in (1a), just like its English counterpart in (1b), is ambiguous between a collective and a distributive reading. On the collective reading of (1), the girls are painting a box together. On the distributive reading of (1), where each girl paints a (different) box, the NP 'girls' is said to serve as the distributive-key and the NP 'box' as the distributive-share, since a box is distributed over the members of the group of girls.

Following Choe (1987), languages vary according to whether they morphologically mark the distributive-key or the distributive-share. Typically, adnominal English quantifiers, such as *every* and *each*, are taken to mark the distributive-key. By contrast, distributive markers, such as *-ssik* in Korean (Choe (1987), Oh (2001)), mark the distributive-share. While distributive-key markers may imply either strong (i.e. obligatory) distributivity (e.g. *each*) or weak (optional or pseudo) distributivity (e.g. *every*), distributive-share markers enforce strong distributivity (i.e. they are false under collective readings, allowing only distributive readings).[2]

Knežević (2015) argues that, in Serbian, the universal quantifier *svaki* (*every*) is a distributive-key marker, while distributive *po* is a distributive-share marker. Morphosyntactic evidence for this claim is provided in (2) versus (3). *Svaki* in (2), like other distributive quantifiers (e.g. *every* or *each* in English), combines with the NP that serves as the distributive-key, that is, the NP denoting the set over which the distribution takes place (here, *girls*). In contrast, as shown in (3), *po* combines with the NP that denotes what is distributed (here, *box*).

[2] So, for instance, the Serbian sentence in (3) (below in the text) with the distributive-share marker *po* would be false in a context where say three different girls are painting the very same box. In contrast, Serbian sentence in (2a), just like its English counterpart in (2b) ('Every girl is painting the box') would remain true in that context.

(2) a.

[$_{\text{Distributive key}}$	*Svaka*[1]	devojka]	farba	kutiju.
	every-NOM.F.SG	girl-NOM.F.SG	paint-3.SG	box-ACC.F.SG

b. [$_{\text{Distributive key}}$ Every girl] is painting the same/a different box.'[4]

(3)

Devojke	farbaju	[$_{\text{Distributive share}}$	*po*	kutiju].
girl-NOM.F.PL	paint-3.PL		DIST	box-ACC.F.SG

'The girls are painting a box each.'

'The girls are painting a box separately (at different locations/time intervals).'

Note, moreover, that *po* only[3] combines with[4] indefinite, or better, non-specific, cardinal expressions,[5] as well as weak quantifiers (e.g. *mnogo* 'many' or *malo* 'few'). We thus assume, following Choe (1987), that distributive shares are dependent indefinites denoting an explicit quantity.

As shown in (3)/(4a) and (4b), *po* occupies a fixed position in the sentence, immediately preceding the NP serving as the distributive-share. In contrast, *svaki* in (2), repeated as (5a), unlike *po* in (4), can be separated from its restriction with which it agrees morphosyntactically. (5a) and (5b) illustrate (respectively) SVO and VOS order in Serbian, and (5c–e) that *svaki* and its restriction—that is, the NP serving as the distributive-key—can split off each other.

(4) a.

Devojke	farbaju	[$_{\text{Distributive share}}$	po	kutiju].
girl-NOM.F.PL	paint-3.PL		DIST	box-ACC.F.SG

b.

(#Po)	devojke	(#po)	farbaju	kutiju	(#po).
DIST	girl-NOM.F.PL	DIST	paint-3.PL	box-ACC.F.SG	DIST

[3]*Svaki* is a so-called adjectival quantifier—that is, it morphosyntactically agrees with its combining noun in case, gender and number. Thus, *svaka* in (2a) is the nominative, feminine and singular form of the quantifier *svaki*, agreeing with the noun *devojka* (*girl*).

[4]Serbian is a determinerless language. A bare noun such as *kutiju* in (2) can thus be construed as either a definite or (non) specific indefinite (depending on context). Importantly, bare nouns combining with *po,* such as *po kutiju* in (3), only allow a non-specific indefinite reading (see directly below in the text and footnote 5 for further discussion).

[5]Note, in particular, that under Knežević's (2015) analysis, since the NP serving as a distributive-share must always be a quantity denoting expression, an NP combining with *po* that appears on the surface to be bare, is analyzed in fact as combining with the silent numeral 'one'—e.g. *po kutiju* is analyzed as [$_{\text{Distributive share}}$ *po* [ONE box]].

(5) a. [$_{\text{Distributive key}}$ *Svaka* devojka] farba kutiju.
every-NOM.F.SG girl-NOM.F.SG paint-3.SG box-ACC.F.SG

b. Farba kutiju [$_{\text{Distributive key}}$ *svaka* devojka].
paint-3.SG box-ACC.F.SG every-NOM.F.SG girl-NOM.F.SG

c. [$_{\text{Distributive key}}$ Devojka] farba kutiju *svaka*.
girl-NOM.F.SG paint-3.SG box-ACC.F.SG every-NOM.F.SG

d. Farba [$_{\text{Distributive key}}$ devojka] kutiju *svaka*.
paint-3.SG girl-NOM.F.SG box-ACC.F.SG every-NOM.F.SG

e. [$_{\text{Distributive key}}$ Devojka] *svaka* farba kutiju.
girl-NOM.F.SG every-NOM.F.SG paint-3.SG box-ACC.SG

2.2 *Event Versus Participant Distribution*

A further core difference between distributive-key and distributive-share markers is that the first only yield so-called participant-distributive readings, while the second yield both participant-distributive and event-distributive readings (Oh (2001), Zimmermann (2002), Balusu (2006), Cable (2014), Knežević (2015)). For illustration, let us return to (2a/2b) and (3), repeated below as (6a/b) and (7) respectively, in the two contexts given in (8).

(6) a. [$_{\text{Distributive key}}$ Svaka devojka] farba kutiju.
every-NOM.F.SG girl-NOM.F.SG paint-3.SG box-ACC.F.SG

b. [$_{\text{Distributive key}}$ Every girl] is painting a (different) box.

i. √Participant-distributive: 'Every/each girl is painting a (different) box.'

ii. #Event-distributive: 'The girls are painting a box at different locations/time intervals.'

(7) Devojke farbaju [$_{\text{Distributive share}}$ po kutiju].
girl-NOM.F.PL paint-3.PL DIST box-ACC.F.SG

i. √Participant-distributive: 'Every/each girl is painting a (different) box.'

ii. √Event-distributive: 'The girls are painting a box at different locations/time intervals.'

(8) a. ***Participant****-distributive context*

This Wednesday afternoon, Mary, Jane and Rosa are each painting a different box at the same time.

b. ***Event****-distributive context:*

Every day of the week, the same three girls (Mary, Jane and Rosa) together paint a (different) box.

Both (6a)/(6b) with the universal quantifier *svaki/every* can be used felicitously under the *participant*-distributive context in (8a), where each of the three girls, Mary, Jane and Rosa, is painting a different box at the same time/temporal location (a given Wednesday afternoon). The difference in meaning between *po* and *svaki* comes out, however, under the *event*-distributive context given in (8b) where, at different times/temporal locations (that is, each day of a given week), there is an event of painting a (different) box involving the same three girls as agent. While it is not permissible to use either (6a) with *svaki* or (6b) with *every* to describe this scenario, it is perfectly natural to use (7) with *po* on the interpretation in (7ii).

To sum up, sentences with *svaki/every* versus sentences with *po* do not have the same truth conditions. (6a)/(6b) with *svaki/every* allow the participant-distributive reading in (6i), where each girl is painting a different box at the same time, but not the event-distributive reading in (6ii). The distribution in (6a)/(6b) is thus taken to be over the agent argument of the verb, here girls: events of painting a box on a given Wednesday are distributed over different girls. In contrast, (7) with *po*, unlike (6a)/(6b), also allows the event-distributive reading in (7ii), where the same girls are painting a box together at different temporal locations. The distribution is taken to be over the (implicit) spatiotemporal argument of the verb, here the days of the week: events of the three girls (collectively) painting a box are distributed over different locations or/and time intervals.

The controversial issue in the literature is how to account for this apparent ambiguity in (7). Is it that distributive-share markers allow distribution over both agent participants and spatiotemporal locations, as Choe (1987), Oh (2001) and Zimmermann (2002) argue? Or is it that they allow only distribution over spatiotemporal locations, with participant-based readings derived as a species of spatiotemporal distribution, as Balusu (2006), Cable (2014) and Knežević (2015) contend?

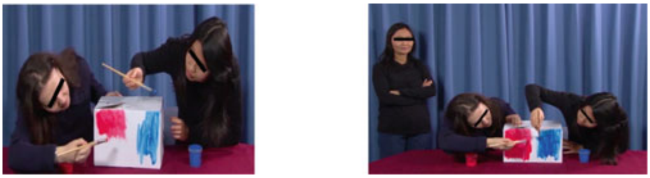
On the latter approach, both so-called *participant*-distributive and *event*-distributive readings are uniformly derived from distribution over spaciotemporal locations. The idea, in a nutshell, is that spatiotemporal distribution requires a plurality of times and/or a plurality of spatial locations. Distribution over a plurality of spatial locations, but at the same temporal location, enforces distribution over a plurality of agents for the simple reason that the same agent cannot be in different spaces at the same time. Let us go back to the readings of (7) in the contexts in (8). Crucially, event distribution over different spatial/temporal locations can, in principle, involve distribution either over *different spatial* (but not temporal) locations, as is the case on the scenario in (8a), where each of the girls is painting her own box on the same Wednesday afternoon, or over *different temporal* locations, as is the case on the scenario in (8b), where the same three girls acting as a group, collectively paint the same box on different days of the week. Spatiotemporal distribution thus yields both the so-called participant-distributive and event-distributive readings.[6]

2.3 *Strong Versus Weak Distributivity*

Knežević (2015) further argues that the distributive marker *po* is strongly distributive, while the universal quantifier *svaki*, just like its English counterpart *every*, is pseudo-distributive.[7] A strongly distributive quantifier enforces distributive readings, while a pseudo-distributive quantifier does not enforce distributive readings—that is, allows **collective** readings.

Take, for instance, the *po*-sentence in (7). Since *po* is strongly distributive, (7) will not be true under either of the *collective* contexts depicted under (9), where two girls are collectively painting the same box.

(9) **Collective contexts**



a. Exhaustive b. Non-Exhaustive

[6]For extensive discussion and, in particular, an explicit formal analysis of the semantics of distributivity with *po*, the reader is referred to Knežević (2015). Note that Balusu (2006) and Knežević (2015) both agree that distribution with distributive-share markers is over spatiotemporal units. There is, however, an essential difference between the two approaches, discussed in Sect. 3. See also Bosnić et al. (2016) for further discussion of spatial distribution and of how to experimentally tease apart these two approaches.

[7]The issue, however, of whether universal quantifiers such as *every*/*svaki* do indeed allow collective readings is a matter of contention in the literature. As Katalin É. Kiss points out (p.c.), with perfective punctual accomplishment verbs, the collective reading becomes unavailable, as the following contrast from Hungarian shows:

(i) Minden fiú emelte a zongorát.
every boy lifted the piano.
'√All the boys were lifting the (single) piano together'.
(ii) Minden fiú **fel**-emelte a zongorát.
every boy **up**-lifted the piano
'√ Every boy lifted the piano separately, on his own'.

While (i) allows a collective reading of the universal quantifier, this reading is blocked in (ii) with the perfective particle on the verb. The same contrast is also found with Serbian *svaki*. Importantly, with strong distributive markers, such as *po*, the collective reading is never available, be it with perfective or non-perfective verbs. This is what we take to be the relevant difference here for distinguishing pseudo-distributive quantifiers (i.e. *svaki*) from strongly distributive markers (i.e. *po*).

In contrast, since *svaki/every* is pseudo-distributive, (10) allows both the (participant) distributive reading in (10i) and the collective reading in (10ii). As such, it will be true under the collective (exhaustive) context in (9a), unlike (7), though crucially not under the collective (non-exhaustive) context in (9b).

(10) [$_{\text{Distributive key}}$	Svaka	devojka]	farba	kutiju.
	every-NOM.F.SG	girl-NOM.F.SG	paint-3.SG	box-ACC.F.SG

i. √Participant-distributive: ‘Every girl is painting a (different) box.’

ii. √Collective: ‘Every girl is painting the same/ a specific box.’

This is so because universal quantifiers enforce exhaustivity: (10) cannot be used to describe situations involving girls not painting a box. The restriction of the quantifier (i.e. “girl”) has to be exhausted—that is, all members of the set of girls in the context are required to participate in the described painting event. This is why (10) with *svaka*, just like its English counterpart with *every,* is true under the e*xhaustive* collective context depicted under (9a), where there are only two girls in the context, both of which are painting the same box, but false under the *non-exhaustive* collective context in (9b), where there is a third girl in the context not painting the box.

2.4 Atomicity

Lastly, Knežević (2015) argues that the distributive marker *po*, unlike the universal quantifiers *svaki* and *every*, does not enforce atomic distribution. To see why consider the sentences in (12), under the contexts depicted in (11a–b). In (11a), distribution is *atomic*: the set of two girls is partitioned atomically and each girl atom of this set is the agent of a box painting event. In (11b), the set of four girls is partitioned into *non-atomic* sets of two girls, and each of these subsets of girls is the cumulative agent of a box painting event. While the Serbian sentence in (12a) with *po* is judged true in either of these atomic contexts (irrespective of whether distributivity is atomic as in (11a), or non-atomic as in (11b)), the Serbian sentence in (12b) with *svaka*, just like its English counterpart in (12c) with *every* is judged false under the non-atomic distributive context in (11b).

(11) **Atomic vs. Non-Atomic Distributive contexts**

(a) Atomic distributive

(b) Non-Atomic distributive

(c) Non-Exhaustive Atomic distributive

(d) Non-Exhaustive & Non-Atomic distributive

(12) a. Devojke farbaju [$_{\text{Distributive share}}$ po kutiju].
girl-NOM.F.PL paint-3.PL DIST box-ACC.F.SG

b. [$_{\text{Distributive key}}$ Svaka devojka] farba kutiju.
every-NOM.F.SG girl-NOM.F.SG paint-3.SG box-ACC.F.SG

c. Every girl is painting a (different) box.

Recall, moreover, that *svaka/every* enforce exhaustivity (as discussed in Sect. 2.3 above). (12b/c) with *svaka/every* will thus be also rejected under the non-exhaustive distributive contexts depicted under (11c–d). In sum, irrespective of whether distribution is atomic (11c) or not (11d), (12b/c) will be false under these two contexts since there is a girl not participating in the box painting events.

We thus have a striking contrast between the truth conditions of (12a) with *po* versus (12b–c) with *svaka/every*. Since the later enforce exhaustivity as well as atomicity, both (12b) and (12c) are acceptable only under the atomic and exhaustive distributive contexts in (11a). In contrast, since *po* enforces neither atomicity, as we have just seen, nor exhaustivity,[8] the *po*-sentence in (12a) is acceptable under all of the distributive contexts in (11), whether distributivity is atomic or not, or whether it is exhaustive or not.

[8]For arguments that distributive-share quantifiers do not enforce exhaustivity, see Knežević (2015). Note, in particular, that the fact that (12a), unlike (12b–c), is true under either of the non-exhaustive distributive contexts in (11c–d) does not by itself show that *po* does not enforce exhaustive distribution since distribution is therefore exhaustive relative to a distributive-key. For Balusu (2006), the distributive-key (with distributive-share markers) is the set of contextually determined (relevant) spatiotemporal locations. It is thus not clear whether Balusu predicts (12a) to be true under (11c–d). For an experimental study testing exhaustivity (of spatiotemporal locations) with *po*, see Bosnić et al. (2016).

3 Experimental Investigations

3.1 Previous Experimental Results

We now briefly report the (relevant) results of previous experiments investigating the comprehension of sentences with *po*-numerals in subject or/and object positions with both Serbian adults and children. We start with the adult results, which, as we shall see, provide nice experimental evidence for two of the claims made by Knežević (2015). Namely, that *po* enforces strong distributivity (that is, does not yield collective readings), but not necessarily atomic distributive readings.

3.1.1 *'Po'* Enforces Strong Distributivity (Does not Allow Collective Readings)

If *po* is strongly distributive, then we expect that neither a sentence such as (13a) with a *po*-numeral in object position, nor a sentence such as (13b) with a *po*-numeral in subject position, should be accepted under the collective context depicted in (13c). This was indeed the case since, in this context, adults accepted (13a) only 2.4% of the time, and (13b) only 3.2% of the time (see (13d)).

(13) a. Dve devojke farbaju [**po** kutiju]
two girl-NOM.F.PL paint-3.PL DIST box-ACC.F.SG

b. [**Po** dve devojke] farbaju kutiju.
DIST two girl-NOM.F.PL paint-3.PL box-ACC.F.SG

c. Collective context

d. 'Yes' responses for (13a): 2.4%
'Yes' responses for (13b): 3.2%

3.1.2 *'Po'* Does not Enforce Atomicity

If distribution with *po* need not be atomic, then we expect *po*-sentences to be accepted in both atomic and non-atomic distributive contexts. Take the sentence in (14a), where the *po*-numeral denoting the distributive-share (that is, *po two balloons*) is in object position. We expect (14a) to be judged true, be it under the context in (14b), where two balloons are distributed over each atom of the set of three girls, or under the context in (14c), where two balloons are distributed over non-atomic sets of three girls.

(14) *po*-numeral in object position

a.

Tri	devojke	drže	[**po**	dva	balona].
three	girl-NOM.F.PL	hold-3.PL	DIST	two	balloon-PAUCAL

b. Atomic-distributive context**:**

c. Non Atomic-distributive context**:**

Likewise for the sentence in (15a), where the *po*-numeral is in subject position and what is thus being distributed are quantities of three girls. We expect (15a) to be judged true, be it in a context where three girls are distributed over each atom of the set of two balloons (15b), or in a context where three girls are distributed over non-atomic sets of two balloons (15c).[9]

[9]There are obviously other non-atomic partitioning of the set of girls in (14a)/balloons in (15a) that make these sentences respectively true—e.g.: (14a) should also be judged true in say a context where there is a total of three girls partitioned into two subsets, and quantities of two balloons are distributed over each of these subsets (that is, where one girl on her own is holding two balloons while two girls are together holding two balloons). Although (14a) is indeed judged true in such contexts (see Knežević 2015) for extensive discussion and adult judgments), we have not as yet gathered psycholinguistic evidence bearing on non-atomic, cumulative distribution with *po*.

(15) *po*-numeral in subject position

a. [**Po** tri devojke] drže dva balona.

DIST three girl-NOM.F.PL hold-3.PL two balloon-PAUCAL

b. Atomic-distributive context**:**

c. Non Atomic-distributive context**:**

The experimental results are given under Table 1. We see that *po* is accepted in both atomic and non-atomic distributive contexts, although there is a marked preference for *po*-object in the atomic distributive context versus *po*-subject in the non-atomic distributive context. This is an intriguing preference. Knežević and Demirdache (to appear) offer the following conjecture. Recall that Serbian is a determiner-less language and that a bare noun such as *tri devojke* ("three girls") in (14a) can thus, in principle, be construed as either a definite or (non) specific indefinite, depending on context. However, as the adults' volunteered comments reveal, some speakers preferred to interpret the bare numeral *tri devojke* in subject position as referring to three specific girls, while the scenario depicted in (14c) involved six girls. The question is then why the speakers dispreferred the *po*-subject sentence in (15a), where the object numeral-NP *dva balona* ("two balloons") is *po*-less, in the context in (15b). That is, since adults prefer the specific interpretation of the bare (*po*-less) subject numeral-NP *tri devojke* ("three girls") in (14a), given in (14b), we would expect the same preference for the specific interpretation of the bare (*po*-less) object numeral-NP *dva balona* ("two balloons") in (15a), given in (15b). This was however not the case. The explanation for this is given in terms of information structure in Knežević and Demirdache (to appear). Note that Serbian is a language displaying relatively free word order and where information structure plays an important role. Speakers rejecting the *po*-subject sentence in (15a), under the atomic context (15b), volunteered instead, in this very same context, (16) where the *po*-less object has been scrambled to the sentence-initial (topic) position.

(16) **Dva balona** [po tri devojke] drže.

two balloon-PAUCAL DIST three girl-NOM.F.PL hold-3.PL

Intended: 'Three girls are holding each of two balloons.'

Table 1 % of 'yes' responses

Adult results	*po*-object	*po*-subject
Atomic distributive	99	47.1
Non-atomic distributive	34.3	68.6

If this was the reason of lower acceptance of (15a) under (15b), the strong prediction would be that (16), where the bare numeral is scrambled to the sentence-initial (topic) position, would ameliorate the acceptance of (15b).[10]

Conversely, we might then expect the percentage of acceptance for *po*-object sentence in (14a), in the non-atomic context (14c), to be higher if the *po*-numeral is in sentence-initial position (on a par with (15a)), as in (17)[11]:

(17) [**Po dva balona**] tri devojke drže.
DIST two balloon-PAUCAL three girl-NOM.F.PL hold-3.PL
Intended: "Groups of three girls are each holding groups of two balloons."

3.1.3 Children

The results for 19 Serbian children ranging in age between 6;8 and 7;6 (MA = 6;5, SD = 0.5) are given under Table 2. There are two important differences between the children and the adults. First, unlike adults, children, allow collective readings with *po*. That is, *po* appears to be weakly distributive for children, and not strongly distributive as is the case for adults (recall that the latter accepted *po*-object/*po*-subject only 2.4/ 3.2% of the time in collective contexts).[12]

[10]If this conjecture is right, the strong prediction would then also be that (16) should be accepted on a **temporal** event-distributive context, such as the one suggested below (replacing the balloons by the kites):

(i) *Temporal NAD context for the po-subject sentence (16)*

We are at a summer camp during the month of August attended by girls. Every Monday, the same two kites are flown by three different girls.

[11](17) would be assigned the LF in (i) where the implicit spatiotemporal argument serves as the distributive-key:

(i) [TOPIC spatiotemporal argument] [*po* two balloons] [three girls] hold

[12]A reviewer suggests, children might be accepting *po* under the collective condition on an implicit temporally distributive reading involving collective agents, (that is, "imagining a scenario where *po* is collective over agents, but distributive over time-intervals"). Notice that this is precisely the context suggested in footnote (10) for (15a)/(16): *We are at a summer camp during the month of August attended by girls. Every Monday, the same two kites are flown by three different girls*. We have no evidence to bear on this issue since we restricted our investigation to spatial distributive (non-atomic) contexts merely for reasons of design simplicity. An obvious next step would be to test the acquisition of spatial versus temporal (atomic vs. non-atomic) distribution.

Table 2 7-year olds: % of 'yes' responses

Children results	*po*-object	*po*-subject
Collective	48.6	64.7
Atomic distributive	93.9	20.2
Non-Atomic distributive	91.2	94.7

Second, children accept *po*-object and *po*-subject indiscriminately in non-atomic distributive contexts at very high rates. The question is how to interpret this result. A 'yes' answer is the expected response in this context, given that *po* does not enforce atomic distribution. But can we, however, safely conclude, on the basis of these high percentages of 'yes' answers, that 7-year-old children know that distribution with *po* can be non-atomic, when the results for the collective context show that they do not know that *po* is strongly distributive?

To answer this question, we would need to block the non-atomic distributive reading in *po*-sentences. To this end, we can introduce the universal quantifier *svaki* (*every*) since the latter enforces atomicity (Sect. 2.4). Since *po* marks the distributive-share and *svaki* marks the distributive-key, they cannot combine directly (that is, modify the same NP), but they can co-occur in the same sentence, as in (19) below. While *po* signals the distributive-share, *svaki* in (19) forces atomic partitioning of the set (denoted by the NP it combines with), and for this reason non-atomic distribution will not longer be possible.

3.2 *Enforcing Atomicity and Exhaustivity in* **po-***Sentences by Adding* **svaki**

The goal of the experiment reported here is to investigate the comprehension of *po*-sentences with the universal quantifier *svaki*.

As we have just seen, the previous psycholinguistic research carried out provided experimental evidence for two major theoretical claims defended in Knežević (2015): *po*, unlike *svaki*, enforces strong distributivity (blocks collective readings), but does not enforce atomicity. Recall, moreover, that *svaki,* unlike *po*, allows collective readings, but enforces exhaustivity (Sect. 2.3). These claims are recapitulated below.

(18) Assumptions:

- (a) *po*, unlike *svaki*, blocks collective readings
- (b) *svaki*, unlike *po,* enforces exhaustivity
- (c) *svaki*, unlike *po,* enforces atomicity

Our hypothesis is thus that *po*-sentences with the universal quantifier *svaki* should be accepted only in contexts that are strongly distributive (satisfying assumption (18a)), exhaustive (satisfying assumption (18b)) and atomic (satisfying assumption (18c)). We created six experimental conditions illustrated in (20), to test this hypothesis with the test sentence in (19). Three conditions, atomic-distributive (AD), non-atomic-distributive (NAD), and collective (COLL), tested whether the atomic partitioning of the group of girls is required. The other three conditions tested the exhaustivity requirement with *svaki*. They were identical to the first three conditions, except that they involved an extra subject (girl), hence we call them AD+, NAD+ and COLL+.

(19) Test sentence

Svaka	devojka	farba	**po**	kutiju.
every-NOM.F.SG	girl-NOM.F.SG	paint-3.SG.	DIST	box-ACC.F.SG

'Each girl is painting a (different) box.'

(20) Experimental conditions

a. AD

c. NAD

e. COLL

b. AD+

d. NAD+

f. COLL+

Recall from Sect. 2.2., that event-distributive readings can, but need not be, either atomic or exhaustive. Atomic exhaustive readings are thus a subset of event-distributive readings, but not conversely. Our test hypothesis is that (19) should not be accepted in the event-distributive contexts in (20b–d) with non-atomic (NAD) and non-exhausted participants (AD+, NAD+), but only in the atomic participant-distributive context (AD) in (20a).

3.2.1 Participants

Adults
The monolingual Serbian speakers were recruited in Novi Sad, Serbia (N = 31, f = 21, m = 10, MA = 27;3, SD = 8.6).

Children
A total of 98 children were tested, separated in three age groups. The first group consisted of one child aged 3;0 and 22 children between the ages 4;3 and 6;9 (N = 22, MA = 5;0, SD = 0.8, 9 girls, 14 boys). These children were recruited in the kindergarten 'Povratak prirodi' in Belgrade (Zemun), Serbia. The second group were children between the ages 6;11 and 8;1 (N = 38, MA = 7;2, SD = 0.5, 24 girls, 14 boys), from the primary school 'Đura Daničić' in Novi Sad, Serbia. The third group consisted of children between the ages 8;7 and 11;0 (N = 37, MA = 9;4, SD = 0.7, 19 girls, 18 boys), from the primary school 'Petar Kočić' in Temerin, Serbia. We will refer to the three groups by using the approximate mean age per group: 5-year olds, 7-year olds and 9-year olds.

3.2.2 Design and Materials

Each of the six experimental conditions in (20), AD, NAD, COLL, AD+, NAD+ and COLL+, was matched with the test sentence in (19). These six configurations were presented with four different predicates (verb-object combinations), yielding a total of 24 test trials. The experiment also included 12 distracter items of type 'N-pl V-intransitive', such as 'Devojke kijaju' ('The girls are sneezing'). Half of the distracter items were 'true' (describing the given situation) and half of them were 'false' (not describing the given situation). The 36 items (24 test items + 12 distracters) were compilations of short scenes and audio descriptions pronounced by a pre-recorded female voice. (See Appendix for the complete list of items).

3.2.3 Procedure

A Truth Value Judgment Task was used. A program in Python was created to screen the items in a random order and collect the answers automatically. Each child was presented short videos on the computer screen, in a quiet room away from the class. The children were instructed to answer if they agreed or not with the description that the voice provided for a video. The youngest children (5-year olds) were asked to answer and the experimenter entered the responses for them. Older children (7- and 9-year olds) clicked on their own the yes/no button that appeared after each video. The comments that children provided were recorded. The adult participants did the experiment in the same way, in the presence of the experimenter who did not intervene.

3.2.4 Hypothesis and Predictions

(21) Test Hypothesis (H)
*Svaki*NP-*po*NP sentences yield only a strongly distributive atomic exhaustive reading (AD).

We expect *po* to block collective readings (thus excluding both COLL and COLL+), *svaki* to enforce atomic partitioning of the group of girls (thus excluding both NAD and NAD+), as well the exhaustive participation of all members of the set girls (thus excluding AD+).

Since *svaki* enforces atomic partitioning of the group of girls and exhaustive participation of all group members, and *po* enforces distributivity, the reading of the sentence is that each girl is individually painting a different box. That is, only atomic distributive contexts should be accepted (AD).

3.3 Results

Figure 1 reports percentages of ‘yes’ responses for each condition and group. A series of binomial tests were run to compare these results to the percentage corresponding to the null hypothesis for random answers, i.e. 50%. The tests results indicate that the null hypothesis was rejected at a level of 5%, except for 5-year olds under conditions COLL and COLL+, 7-year olds under condition COLL+ and 9-year olds under condition NAD.

Adults. The data show that the predicted interpretation AD is accepted almost at ceiling (98.4%). For the three conditions AD, NAD and COLL, the scores were significantly different ($\chi^2(2) = 223.53$, $p < 0.001$). We refined the analysis by comparing NAD (20.2%) versus COLL (13.7%). The Chi-Square test did not reveal a significant difference ($\chi^2(1) = 1.4$, $p = 0.2$). The acceptance rates of AD+ (around 25%) are significantly higher than those for NAD+/COLL+ (around 7% and 5% respectively), ($\chi^2(2) = 27.73$, $p < 0.001$).

5-year old children accept the following conditions above chance: AD, NAD, AD+ and NAD+. For COLL and COLL+, percentages of ‘yes’ responses are not significantly different from 50%. No significant difference is found between the scores on AD (81.5%) versus NAD (73.9%), (χ^2 $(1) = 1.13$, $p = 0.28$) nor between AD (81.5%) versus COLL (59.8%), ($\chi^2(1) = 9.46$, $p = 0.002$), or between NAD (73.9%) and COLL (59.8%), ($\chi^2(1) = 3.53$, $p = 0.06$). No difference is found between AD+ (63%) and COLL+ (54%), ($\chi^2(2) = 1.47$, $p = 0.48$). The difference between 5-year olds and adults on the target AD condition was found to be significant (81.5% vs. 98.4% respectively) ($\chi^2(1) = 16.7$, $p < 0.001$).

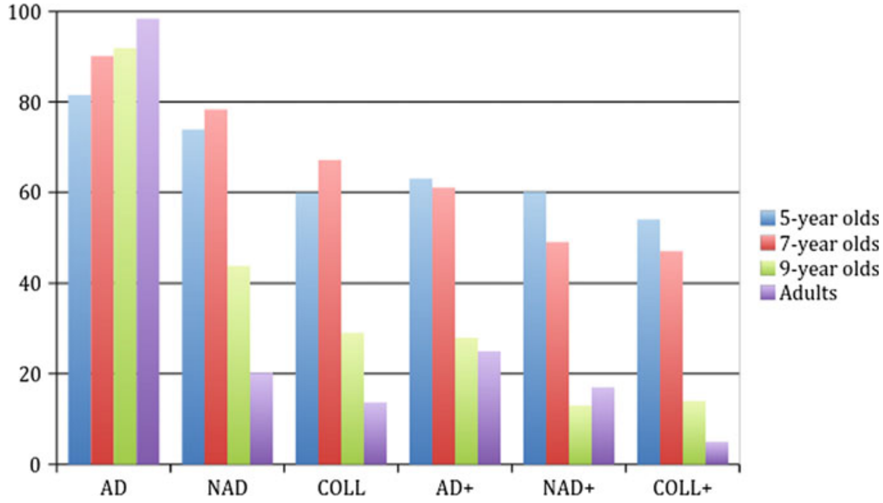


Fig. 1 Percentages of 'yes' responses across groups

7-year olds in general pattern like 5-year olds. No significant difference in percentages is found between the two groups for AD ($\chi^2(1) = 3.01$, $p = 0.08$), NAD ($\chi^2(1) = 0.39$, $p = 0.5$) or COLL ($\chi^2(1) = 1.04$, $p = 0.3$). On AD+, the percentages are almost identical: 63% (5-year olds) and 61% (7-year old). No significant difference is found between the two groups for NAD+ (60% for 5-year olds vs. 49% for 7-year olds) ($\chi^2(1) = 1.1$, $p = 0.2$), or for COLL+ (54% vs. 47%), ($\chi^2(1) = 1.05$, $p = 0.3$).

9-year olds differ importantly from 7-year olds on all conditions except on the target condition AD ($\chi^2(1) = 0.1$, $p = 0.7$), and are, overall closer to adults. No significant difference is found between 9-year olds and adults on the target condition AD ($\chi^2(1) = 4.57$, $p = 0.03$). The difference between adults and 9-year olds is found to be significant for NAD ($\chi^2(1) = 16.14$, $p < 0.001$), but not for COLL ($\chi^2(1) = 8.37$, $p = 0.004$) or COLL+ (χ^2 $(1) = 4.9$, $p = 0.03$).[13]

[13]For this type of study, the usual confidence level for statistical tests (such as Chi-Square) is 95%, and the result is considered as significant if $p < 0.05$. However, given that we have performed multiple comparisons, we applied Bonferroni correction. To do so, we divide the risk level (0.05) by the number of possible comparisons across treatments (or conditions) and age groups. In our experiment, this is equivalent to taking a risk level of 0.001. Therefore, results are considered significant when $p < 0.001$.

3.4 Discussion

3.4.1 Adults

We can say that overall adults results are in line with H in (21), that is, that sentences with *svaki* and *po* yield only atomic exhaustive distributive readings (98.4% of acceptance). We assume that the marginal acceptance of collective readings is due to the presence of *svaki*, given that the latter is compatible with collective construals (Sect. 2.3), and conversely that the marginal acceptance of non-atomic and non-exhaustive readings is due to the presence of *po,* given that the latter is compatible with non-atomic partitioning and non-exhaustive distribution (Sect. 2.4). Given the significant difference in acceptance under the atomic exhaustive context (AD) relative to all other contexts, we conclude that *svaki*NP-*po*NP enforces atomic exhaustive participant-distributive readings, as expected.

3.4.2 Children Versus Adults

Unlike adults, children overall allow non-atomic and non-exhaustive distributive, as well as collective, readings with *svaki*NP-*po*NP sentences. A slight preference for the target reading is attested, but the difference in acceptance rates between the target and non-target contexts becomes significant only at the age of 9. Children show non-adult like patterns on all conditions, with a significantly lower acceptance of the target condition and significantly higher scores on all other conditions, compared to adults.

Overall, 5-year olds and 7-year olds pattern alike. That is, children at the age of 7 are not sensitive to the strong distributive force of *po*, given that collective readings are not excluded. This finding is consistent with the results for distributive *po* in previous experiments (Knežević 2015), as well as with the data for Russian 5–6 year olds (Sekerina and Sauermann 2017, see Sect. 3.4.4). Nor are either 5-year olds or 7-year olds sensitive to the exhaustivity requirement triggered by *svaki (every)*, given that non-exhaustive readings are accepted.

The response pattern for 9-year olds falls halfway between that of 7-year olds and that of adults. We can see that children at age 9 start to dislike non-exhaustive readings, as compared to exhaustive readings. That is, 9-year olds and adults show similar percentages of 'yes' responses for the non-exhaustive conditions, and similar percentages of rejection for collective readings. They accept, however, non-atomic (NAD) reading at significantly higher rates than adults. In sum, 9-year-olds show patterns of responses similar to those of adults for all conditions, except the non-atomic distributive condition (NAD).

3.4.3 Possible Explanations. Acquiring Exhaustivity Prior to Atomicity. Acquiring *po* Prior to *svaki*

We have just seen that seen 7-year-olds do not reject collective readings with *po*, which suggest that they have not acquired the meaning of *po*. Nor do they reject non-exhaustive readings with *svaki*, which suggests that they have not acquired the meaning of *svaki* either.

Recall that we closed Sect. 3.1.3 with the question of whether high acceptance of non-atomic distributive readings with *po* in 7-year-olds could allow us to conclude that 7-year-olds know that distribution with *po* can be non-atomic. The answer is no. The results from the present experiment show us that 7-year-olds accept all non-target conditions with *po*.

We have further seen that 9-year-olds have the adult-like pattern on all conditions except the non-atomic distributive condition (NAD). We take this to show that 9-year-olds have acquired the meaning of *po* which, unlike *svaki*, enforces strong distributivity, since they reject collective readings, and that *po*, unlike *svaki*, allows both atomic and non-atomic distributive readings, since they accept both. Turning to the acquisition of *svaki*, they know that *svaki* enforces exhaustivity (since they reject non-exhaustive contexts), but they do not appear to know that *svaki* enforces atomic partitioning (since they accept non-atomic distributive construals). This in turn suggests that children acquire the exhaustivity requirement on universal quantifiers prior to the atomicity requirement. Putting all these findings together further suggests that children acquire *po* prior to *svaki*. We summarize these findings below.[14]

(22) a. 7-year-olds do not master the meanings of *svaki* and *po*
b. 9-year-olds know the truth conditions of *po* but not those of *svaki*.
That is, children acquire the distributive-share marker *po* prior to the universal quantifier *svaki* (*every*).
c. The exhaustivity requirement on universal quantification is acquired before the atomicity requirement.

In the closing section of the paper (Sect. 3.5), we address briefly the issue of how these findings bear on the analyses in the literature of distributive-share versus distributive- key markers. We first, however, comment on these findings from a cross-linguistic experimental perspective.

[14]A reviewer makes the very interesting suggestion that the acquisition path would be exhaustivity > distributivity > atomicity, while we have been less bold is our conclusions, claiming merely that exhaustivity > atomicity and that the acquisition of *po* precedes the acquisition of *svaki*. Although this conjecture is not incompatible with our findings, we do not have enough evidence however to conclude that distributivity is acquired either before or after exhaustivity since the difference between adults and 9 year-olds is not significant in collective contexts (see Sect. 3.3 and footnote 13 for statistical significance).

3.4.4 Cross-Linguistic Perspective: Sekerina and Sauermann (2017)

A lot of work has been done on the acquisition of distributive-key quantifiers such as *every, each* or *all*, across languages (Philip 1995; Brooks and Braine 1996; Brooks and Sekerina 2005/2006; Crain et al. 1996; Drozd 2001; Geurts 2003; Gualmani et al. 2003; Kuznetsova et al 2007, among many others). However, as far as we are aware of, only one acquisition study, Sekerina and Sauermann (2017), deals with the Slavic distributive-share marker *po* in sentences with a universal quantifier.[15] The experiment is an eye-tracking study, which focuses on quantifier spreading with Russian children. This term refers to different patterns of non-adult responses elicited with comprehension tasks of sentences with universal quantifiers. The nature and source of these errors are disputed (see, for example, Roeper and De Villiers 1993; Philip 1995; Rakhlin 2007), and we set this issue aside, since it is not relevant here. Sekerina and Sauermann focus on quantifier spreading, but also take into account the distributive force of *po* in Russian, and hypothesize that *po* will help children to spread less, since only a 'strict distributive' reading (in our terms, atomic-distributive) is available with sentences with the quantifier *kazhdyj (every)* and distributive *po,* henceforth *kazhdyj*NP-*po*NP sentences. The task was a sentence-picture verification. The sentences in (23) and (24) were matched with two experimental conditions, (i) containing extra alligators and (ii) containing extra bathtubs (as well as with distracter conditions). Their extra alligators contexts correspond roughly to our non-exhaustive atomic context (AD+) in (20).

(23) Kazhdyi alligator lezhit v vane.
every alligator lies in bathtub
'Every alligator lies in the bathtub.'

(24) V kazhdoj vane lezhit po alligatoru.
in every bathtub lies po alligator
'In every bathtub lies an alligator.'

There are, however, important differences between the two studies. First, Sekerina and Sauermann address different research questions from ours, because their primary focus is on quantifier spreading with the universal quantifier *kazhdyj* (*every*). Second, the design and materials differ in important respects. Their test sentences are unaccusative and come in two types: (i) sentences without *po* and with *kazhdyj (every)* in subject position, such as in (23), and (ii) sentences with *po*NP in subject position and *kazhdyj* (*every*) in a locative PP, such as in (24). Third, *po* in Russian always assigns case to its complement noun (behaving therefore as a preposition even in its distributive use) (see Harves 2003), while Serbian distributive *po* does not assign case, contrary to prepositional *po* (see Knežević 2015). Despite these

[15] See also Dotlačil (to appear) who investigates the processing of the universal quantifier *každý* and distributive *po* in Czech.

language-specific differences (which do not appear to bear on our discussion), both studies investigate children's understanding of sentences with the universal quantifier *kazhdyj*/*svaki* (*every*) in combination with distributive *po*.

Most importantly, Sekerina and Sauermann converge on some important conclusions that we share. Namely, 6-year-old children are not sensitive to exhaustivity, since they accept (23)/(24) in non-exhaustive contexts, with extra alligators or bathtub. Moreover, Russian 6-year-olds do not know the meaning of *po*, since they treat (23) and (24) alike, making quantifier spreading errors regardless of the presence or absence of *po,* contrary to predictions. The results from Russian and Serbian thus converge on the generalizations that the strong distributivity requirement imposed by *po*, just like the exhaustivity requirement imposed by universal quantifiers, are not mastered by 6, but later (by 9 for Serbian).

4 Conclusion. Implications for Theories of Distributive-Share Marking

We recapitulate our experimental findings below.

(25) a. 7-year-olds do not master the meanings of *svaki* and *po*
b. 9-year-olds know the truth conditions of *po* but not those of *svaki*
That is, children acquire the distributive-share marker *po* prior to the universal quantifier *svaki* (*every*).
c. The exhaustivity requirement on universal quantification is acquired before the atomicity requirement.

Recall (from Sect. 2.2) that the controversial issue in the theoretical literature is how to account for the variety of distributive readings that distributive-share markers yield. Is it that distributive-share markers allow distribution over both participants and spatiotemporal locations, as Choe (1987), Oh (2001) and Zimmermann (2002) argue? Or is it that they allow only distribution over spatiotemporal locations, with participant-based readings derived as a species of spatiotemporal distribution, as Balusu (2006), Cable (2014) and Knežević (2015) contend?

Now, although both Balusu (2006) and Knežević (2015) agree that distribution with distributive-share markers is over spatiotemporal units, there is an essential difference between the two approaches. Balusu argues that distributive-share markers involve universal quantification with an invisible distributive (universal) operator ranging over spatiotemporal units that are (exhaustively) distributed over. Knežević argues that distributive-share markers involve pluractionality, that is that the truth conditions of sentences with a *po n(umeral) NP* merely require a plurality of events, each of which involves *n*NP.[16] We can grossly paraphrase the meaning of (26a), according to the two approaches as in (26a) versus (26b), respectively[17,18]:

(26) a. Devojke farbaju po kutiju.
girl-NOM.F.PL paint-3.PL DIST box-ACC.F.SG
b. In every relevant (contextually determined) spatiotemporal unit, girls are painting a box. (Balusu)
c. There are (at least) two events of girls painting a box. (Knežević).

We close this paper by briefly commenting on the implications of our experimental findings for these two approaches to distributive-share marking.

On Balusu's (2006) analysis of Telugu, distributive-share markers involve an (implicit) universal quantifier, whose restriction is always the verb's event (spatiotemporal) argument. When there is no temporal distribution, spatial distribution arises. Participant-based distribution is reduced to spatial distribution because the sub-events are spatially partitioned in such a way that each sub-event involves one (atomic) participant, thus excluding situations where non-atomic groups are participants of different sub-events. This ensures that when distribution is not temporal, but only spatial, it has to involve atomic participants.

The issue with this proposal is that it does not fare well with our previous adult experimental results, since it predicts that non-atomic distributive readings are in fact excluded with the distributive-share marker *po* in Serbian, contrary to Knežević's claims, and to the experimental findings corroborating her claims (Sect. 3.1). As the comparison of the experimental results reported here (Sects. 3.3–3.4), with our previous experimental results (Sect. 3.1) clearly reveals, it is the universal quantifier *svaki,* not the distributive-share marker *po*, that is responsible for enforcing atomic distribution. This opens the question of the extent and the source of variation in the meanings of distributive-share markers across languages, here Telugu versus Serbian (or more generally, Slavic), issues that we leave open pending future cross-linguistic theoretical and experimental investigations.

Turning to Knežević's proposal (see (26c)), *po* is not a universal quantifier, but a pluractional marker, combining with a *n*(umeral)NP and yielding a plurality of events,

---

[16] Note that Beck and von Stechow (2005) develop an analysis of pluractionals for constructions such as 'dog after dog', which shares with ours certain advantages. The core idea is that the semantics of such constructions involves division of events into subevents and entities into subparts, obtained by a system of plural predication involving plural operators and a restriction on the relevant part-whole structures of events. Just like the semantics we give for *po*, the semantics they give for the adverbial itself is thus rather loose, both yielding pluralities of events (or division of events into subevents) and certain partitioning's of entities into subparts. It is not clear at this point, however, what the predictions of Beck and Stechow's analysis would be for the spatial-distributive readings, and as such for the atomic-distributive reading.

[17] As a reviewer points out, Balusu (2006) also has a plurality requirement. Note, however, that this is an additional requirement on the distributive-share ensuring that the latter referentially varies across (sub)events. For related discussion see Knežević (2015).

[18] See also Knežević et al. (2016) for an attempt to experimentally tease apart the two approaches.

each involving *n*NPs (i.e. one box, in (26a)[19]). The participant denoted by the *po*-less NP can be a plural set of individuals (i.e. girls in (26a)), partitioned atomically or not, and exhausted or not, or even a singular individual.[20] *Svaki*, on the other hand, is a universal quantifier, enforcing exhaustivity and atomicity. The prediction for children would then be that they can acquire distributive *po* and the universal quantifier *svaki* independently, at different ages. Since children at the age of 9 reject collective contexts in the presence of *po*, but accept both distributive atomic and non-atomic contexts in the presence of *svaki*, we could conjecture, on Knežević's proposal, that in languages that have both pluractional markers and universal quantifiers, such as Serbian, children acquire pluractionals before universal quantifiers. This conclusion is not altogether surprising in that the truth conditions of *po*-sentences are weaker than those of universally quantified sentences. There is no atomicity or exhaustivity requirement, merely event plurality, since (roughly) all that is required is that there be at least two events, each involving the participant(s) denoted by the *n*(umeral)NP that *po* combines with.

Our experimental findings are thus in line with Knežević's claim that distributive-share marking should not/cannot be uniformly reduced to universal quantification: distributive-share marking does not involve an (implicit) universal quantifier, but rather event plurality.

We hope that our experimental investigation of how Serbian adults and children comprehend the distributive-share marker *po* has contributed novel insights into the meaning of distributivity marking across languages and, as such, to theories of distributivity and universal quantification, as well as opening the door to more psycholinguistic investigation of distributive-share markers across languages.

Acknowledgements We are very thankful to all children and adults that voluntarily participated in the experiment, as well as the schoolteachers. We are also indebted to Boban Arsenijević, Ana Bosnić, Patricia Cabredo Hofherr, Ken Drozd, Katalin É. Kiss, Bart Hollebrandse, Angeliek van Hout, Brenda Laca, Orin Percus, Irina Sekerina, two anonymous reviewers as well as the all organizers and the participants of the *Workshop on Linguistic and Cognitive Aspects of Quantification* at the Research Institute for Linguistic in Budapest, for insightful comments and discussion.

[19]Recall from footnote 5, that *po kutiju* is analyzed as [*po* [ONE box]].

[20]As is the case in (i) below, where the *po*-less NP denotes a singular individual. Distribution in this case can only be over temporal locations.

(i) Devojka farba **po** kutiju.
girl-NOM.F.SG paint-3.SG DIST box-ACC.F.SG
'A/the girl is painting a box at different time intervals.'

Appendix

TEST SENTENCES	SCENARIOS
1. Svaka devojka pere po brod. (Every girl is washing DIST boat)	AD, NAD, COLL, AD+, NAD+, COLL+
2. Svaka devojka farba po kutiju. (Every girl is painting DIST box)	AD, NAD, COLL, AD+, NAD+, COLL+
3. Svaka devojka pakuje po knjigu. (Every girls is wrapping DIST book)	AD, NAD, COLL, AD+, NAD+, COLL+
4. Svaka devojka vuče po auto. (Every girls is pulling DIST car)	AD, NAD, COLL, AD+, NAD+, COLL+
5. Svaka devojka gura po stolicu. (Every girl is pushing DIST chair)	AD, NAD, COLL, AD+, NAD+, COLL+
6. Svaka devojka diže po kofer. (Every girl is lifting DIST suitcase)	AD, NAD, COLL, AD+, NAD+, COLL+

DISTRACTER SENTENCES	SITUATIONS
1. Devojke plešu. (The girls are dancing)	TRUE
2. Devojke kijaju. (The girls are sneezing)	TRUE
3. Devojke zvižde. (The girls are whistling)	TRUE
4. Devojke puze. (The girls are crawling)	TRUE
5. Devojke sedaju. (The girls are sitting down)	TRUE
6. Devojke se saginju. (The girls are bending)	TRUE
7. Devojke se grle. (The girls are hugging)	FALSE (The girls are clapping)
8. Devojke voze bicikl. (The girls are riding a bycicle)	FALSE (The girls are waving)
9. Devojke stoje na rukama. (The girls are doing a handstand)	FALSE (The girls are yawning)
10. Devojke pevaju. (The girls are sining)	FALSE (The girls are crying)
11. Devojke spavaju. (The girls are sleeping)	FALSE (The girls are crouching)
12. Devojke hodaju. (The girls are walking)	FALSE (The girls are bowing)

References

Balusu, R. 2006. Distributive Reduplication in Telugu. In *Proceedings of NELS*, vol. 36, 39–53.

Beck, S., and A. von Stechow. 2005. Dog After Dog Revisited. In *Proceedings of the Sinn und Bedeutung*, vol. 10, 43–54.

Bosnić, A., N. Knežević, J. Spenader, H. Demirdache, and B. Hollebrandse. 2016. Spatial Distribution and Exhaustivity Requirement in Serbian. Talk Presented at *Psycholinguistic Investigations into Number and Quantification in Natural Language*. Aug 29–30, Wroclaw.

Brooks, P.J., and M.D.S. Braine. 1996. What do Children Know About the Universal Quantifiers *All* and *Each*? *Cognition* 60: 235–268.

Brooks, P.J., and I.A. Sekerina. 2005/2006. Shortcuts to Quantifier Interpretation in Children and Adults. *Language Acquisition*. Special Issue: Theoretical Approaches to Children's Isomorphic Interpretations 13 (2): 177–206.

Cable, S. 2014. Distributive Numerals and Distance-distributivity in Tlingit (and beyond). *Language* 90 (3): 562–606.

Choe, J.-W. 1987. Anti-quantifiers and a Theory of Distributivity. PhD Dissertation, University of Massachusetts, Amherst.

Crain, S., R. Thornton, C. Boster, L. Conway, D. Lillo-Martin, and E. Woodams. 1996. Quantification Without Qualification. *Language Acquisition* 5 (2): 83–153.

Dotlačil, J. (to appear). Why is Distributivity so Hard? New Evidence from Distributive Markers and Licensors in Czech. In *Proceedings of FDSL 10.5.*

Drozd, K. 2001. Children's Weak Interpretations of Universally Quantified Questions. In *Language Acquisition and Conceptual Development*, ed. M. Bowerman and S.C. Levinson, 340–376. Cambridge, England: Cambridge University Press.

Gualmani, A., L. Meroni, and S. Crain. 2003. Children's Asymmetrical Responses. In *Proceedings of the Fourth Tokyo Conference on Psycholinguistics*, 135–158.

Guerts, B. 2003. Quantifying Kids. *Language Acquisition* 11 (4): 197–218.

Harves, S. 2003. Getting Impersonal: Case, Agreement, and Distributive Po-phrases in Russian. In *Formal Approaches to Slavic Linguistics*, vol. 11.

Knežević, N., and H. Demirdache. (to appear). On the Acquisition of the Distributive Numerals in Serbian. In *GALA 12 Proceedings*.

Knežević, N. 2015. Numerals and Distributivity in Serbian: At the Syntax-Semantics-Acquisition Interface. PhD Dissertation, University of Nantes.

Knežević, N., A. Bosnić, J. Spenader, H. Demirdache, and B. Hollebrandse. 2016. Testing Spatial Distribution and Exhaustivity in Serbian. Talk Presented at *(Co-)distributivity 2016*, UMR 7023, CNRS/Paris 8, Feb 11–12.

Kuznetsova, J., M. Babyonyshev, J. Reich, L. Hart, and E. Grigorenko. 2007. The Acquisition of Universal Quantifiers in Russian. In *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, ed. A. Belikova, L. Meroni, and M. Umeda, 224–232. Somerville, MA: Cascadilla Proceedings Project.

Oh, S.-R. 2001. Distributivity in an Event Semantics. In *Semantics and Linguistic Theory*, 326–345.

Philip, W. 1995. Event Quantification in the Acquisition of Universal Quantification. Doctoral Dissertation, University of Massachusetts, Amherst, GLSA Publications.

Rakhlin, N. 2007. A New Pragmatic Account of Quantifier-spreading. *Nanzan Linguistics* 8: 239–282.

Roeper, T., and J. De Villiers. 1993. The Emergence of Bound Variable Structures. In *Knowledge and Language, I, from Orwell's Problem to Plato's Problem*, ed. E. Reuland, and A. Werner, 105–139.

Sekerina, I.A., and A. Sauermann. 2017. Quantifier Spreading in Child Eye Movements: A Case of the Russian Quantifier *kazhdyj* 'Every'. *Glossa: A Journal of General Linguistics* 2 (1): 66. 1–18. https://doi.org/10.5334/gjgl.109.

Zimmermann, M. 2002. Boys Buying Two Sausages Each: On the Syntax and Semantics of Distance-distributivity. PhD Dissertation, University of Amsterdam.

The Distributive–Collective Ambiguity and Information Structure

Balázs Surányi and Levente Madarász

Abstract Sentences involving semantically plural NPs may be compatible with several interpretations. In one dominant interpretation, called the distributive reading, the predicate applies separately to each member of the plurality denoted by the NP. Another possibility, the collective reading, involves the predicate applying to the plurality as a whole. While previous theoretical and empirical work has identified a number of linguistic and non-linguistic factors that modulate the choice between these two interpretations in cases of semantic indeterminacy, the potential effect of a key aspect of linguistic meaning, namely Information Structure, has not been empirically investigated. In this pioneering study we show, based on a forced-choice sentence-picture matching experiment, that the information structural topic and focus roles both affect the resolution of the distributive–collective indeterminacy in sentences with semantically plural indefinite NPs. While the effect of focus status is uniform across the types of NPs examined, the effect of topic status is differential. The revealed pattern is argued to follow from the interpretive properties standardly associated with topic and focus roles.

Keywords Distributive · Collective · Information structure · Topic · Focus
Quantification · Hungarian

1 Introduction

A predicate can hold of a plural argument distributively, as in (1a), where the predicate 'died' applies to every member of the set of soldiers; or non-distributively. A common type of non-distributive interpretation is the collective one, illustrated in (1b), where

B. Surányi (✉)
Pázmány Péter Catholic University and Research Institute for Linguistics of the Hungarian Academy of Sciences, Benczúr utca 33. 1068, Budapest, Hungary
e-mail: suranyi@nytud.hu

L. Madarász
Central European University, Nádor utca 9. 1051, Budapest, Hungary
e-mail: Madarasz_Levente@phd.ceu.edu

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_7

'gathered around the house in a full circle' only applies to a set of soldiers together, but it does not hold of each soldier individually. According to semantic approaches to distributivity, while some predicates are lexically distributive, like (1a), or lexically collective, like (1b), other predicates, like (1c), are neither of the two (these are sometimes called 'mixed' predicates). Semantic treatments remain silent regarding the choice between the distributive and the collective interpretations in case the requirements of both are met, as in (1c) (e.g., Scha 1981; Link 1983; Hoeksema 1983; Landman 1989; Lasersohn 1993, 1995; Kamp and Reyle 1993; Landman 1989; cf. also Winter 2001).

(1) a. Two soldiers died. (distributive)
b. Ten soldiers gathered around the house in a full circle. (collective)
c. Two boys lifted the box. (distributive or collective)

While the potential impact of context and world knowledge in the resolution of the latter type of indeterminacy has been occasionally noted, an important question that has not been empirically investigated is whether and how Information Structure (IS) affects the availability of distributive and collective interpretations in semantically indeterminate sentences.

This paper presents an empirical study which, to the best of our knowledge, is the first experimental treatment of this issue. In particular, we report on a two-alternative forced choice picture-matching experiment that probed into interpretive preferences of different types of semantically plural quantified NPs (QNPs) in Hungarian sentences in which the QNP was either information structurally neutral, or it functioned as a focus, or as a topic. The results show that both the focus status and the topic status of QNPs significantly affect the rate of distributive and collective readings of sentences in which they occur. Furthermore, while the effect of focus status is uniform across the types of QNPs examined, the effect of topic status is differential.

We argue that the observed effect of the focus role is due to a difference between the distributive and collective readings with regard to the logical entailment relations that hold between some of the alternative propositions that constitute the basis of the interpretation of focus. The impact of topic status, on the other hand, ultimately derives from a preference for lexically presuppositional expressions to act as sentence topics. That the investigated QNP types do not react to the topic role in a uniform fashion is a result of their divergent presuppositionality properties.

The paper is structured as follows. In Sect. 2 we provide the relevant background concerning the nature of the distributive–collective indeterminacy, the interpretation of the topic and focus IS roles and the different types of QNPs that are to be empirically investigated in our study. Section 3 outlines the experimental methods employed and presents the results. The main findings are discussed in Sect. 4. Finally, Sect. 5 concludes with a summary and suggests avenues for futher research.

2 Background

2.1 The Distributive–Collective Indeterminacy

The term distributivity is generally used to refer to the application of a predicate to the parts of semantically plural entity. Distributive interpretations are commonly diagnosed by a distributive entailment, such as the one from (2a) to (2b), or to the paraphrase in (2c), which is an entailment down to singular individuals. They are also characterized by a cumulative entailment, from (2b) to (2a), which is the reverse of the distributive entailment (Landman 1989). Collective interpretations, on the other hand, arise when the predicate applies to the plural entity as a whole, in the absence of a cumulative entailment, or a distributive entailment down to singular individuals, as in (1b) above. A distributive reading may be obligatory, as in (1a) and (2c), excluded, as in (1b), or, in the case of 'mixed' predicates, the choice may be open between a distributive and a collective reading, as in (1c) and (2a). While the phenomenon is typically discussed in relation to the interpretation of subjects, the ambiguity extends to other dependent nominal phrases, including arguments and adjuncts alike (Dowty 1987; Lasersohn 1993, 1998).

(2) a. John and Bill lifted a box.
    b. John lifted a box, and Bill lifted a box.
    c. John and Bill each lifted a box.

Cases like (2a), in which the predicate phrase contains an indefinite, a further semantic indeterminacy is at play, orthogonal to distributivity. In particular, the object indefinite a *box* in (2a) may either refer to the same box or to two different boxes, even if only a distributive interpretation is considered. To be sure, in case two distinct boxes are involved, (2a) entails (2b) and (2c), hence it is interpreted distributively. Importantly, however, (2a) remains distributive if it reports two events that happen to involve the same box.[1] Indeed, the same type of interpretation is also available to both (2b) and (2c), which unambiguously correspond to the distributive interpretation. Thus, while the collective reading does not necessarily differ from the distributive one with regard to the number of boxes lifted in total, the two differ in terms of the number of lifting events: the latter reports a plurality of events, whereas the former only reports a single event.[2]

[1]This is also the case in (1c) above involving a definite object. The availability of the reading in which the same box partakes in two distinct lifting events crucially depends on whether the predicate can apply to the object more than once, i.e., whether the event is *repeatable*. Verbs that entail a change to their object that makes the event denoted by the VP unrepeatable are incompatible with a multiple-event reading (e.g., *eat a pizza*, *bake a cake*). Note that our discussion of the examples in (2) keeps the logico-semantic meanings, putting aside, for the sake of simplicity the novelty inference that comes with the indefinite article.

[2]The same box/different boxes indeterminacy is often described in terms of the existential scope of the indefinite object. The point is that the logical scope of the object is independent of whether the predicate is interpreted distributively or collectively with respect to the subject.

While the distributive–collective indeterminacy has occasionally been described as a case of underspecification or vagueness (Schwarzschild 1993; Kratzer 2007; see Winter 2000 for some critical discussion), the dominant approach, and the one to be adopted in this paper, postulates semantic ambiguity. According to a common implementation of this approach, distributive interpretations of ambiguous sentences differ from collective ones in that they involve a (normally) silent distributive operator that combines with the predicate phrase to yield a semantically distributive (or plural) predicate phrase (e.g., Link 1983; Dowty 1987; Roberts 1987; Landman 1989, 2000; Lasersohn 1995). Due to the distributivity operator, the distributive predicate phrase incorporates the distributive entailment, exemplified in (2b) and (2c) above, in its truth-conditional semantics. In other words, when a distributive predicate combines with a semantically plural individual (such as John and Bill in (2a)), the predicate is applied separately to each atomic individual contained in it.[3]

Psycholinguistic evidence for the ambiguity view is provided by Frazier et al.'s (1999) reading study. In this eye movement experiment it was found that in the absence of evidence for a distributive reading, the processor commits itself to a collective reading relatively early on, thus it has to perform costly revision in case a linguistic cue disambiguating in favour of a distributive interpretation comes up later. The authors interpret these results as supporting the ambiguity account. They capitalize on earlier literature indicating that readers—following a principle of Minimal Semantic Commitment—leave vagueness unresolved until meaning-specifying information is encountered. As opposed to vagueness, however, they do not postpone the resolution of ambiguities, but rather default to a favoured interpretation. Similar conclusions have emerged from Clifton and Frazier (2012), and from Boylan et al.'s (2011) visual world paradigm experiment.

The preference for the collective interpretation of ambiguous sentences extends to offline settings (Brooks and Braine 1996; Kaup 2002; Ussery 2008; Musolino 2009; Pagliarini et al. 2012; Knežević 2015). One possible explanation for this tendency may be the relative logico-semantic complexity of the distributive interpretation, which is hypothesized to include a covert distributive operator that the collective reading lacks (see Frazier et al. 1999). Alternatively, the collective bias may be due to an avoidance of the postulation of multiple events, which is taken to be cognitively costlier than the postulation of a single event (Harris et al. 2012).[4]

Beyond a general bias towards the collective reading, the choice between a collective and a distributive meaning in the case of mixed predicates is influenced by a variety of factors. The argument structural (or thematic) status of the dependent nominal, as well as its syntactic position within the sentence may both impact preferences in the resolution of the ambiguity (see Brooks and Braine 1996 for an experimental comparison of active subjects and passive *by*-phrases in English). Another promi-

[3]Following Link (1983), semantically plural individuals are treated on a par with singular individuals, ordered by the part-of relation. The former differ from the latter only in that they have other (and ultimately, singular, i.e., atomic) individuals as their parts.

[4]See also Pagliarini et al. (2012) for the potential role of a generalized conversational implicature in the collective bias.

nent factor affecting interpretive biases is the lexical semantic type of the dependent nominal. Essentially quantificational phrases, such as *every*-NPs, only give rise to distributive readings (Partee 1995, p. 564; Hackl 2000, p. 235; see also Kamp and Reyle 1993, p. 481; Endriss 2009, p. 231), while other QNPs do not impose such a restriction.[5] The morphosyntactic form of the dependent NP also contributes to preferences in non-trivial ways. For instance, overtly partitive QNPs are more easily interpreted collectively than their simple counterparts (*most of the boys* vs. *most boys*, Nakanishi and Romero 2004). Further, third person plural personal pronouns may favour collective readings more strongly than other definites (Kaup 2002).

In addition to lexical and grammatical properties of the nominal combining with the predicate, it is often noted that context and pragmatic information may exert palpable influence on the resolution of the distributive–collective ambiguity (for definite NPs, see Roberts 1990; Schwarzschild 1996). While the effects of extra-linguistic context and world knowledge are sometimes elusive, they are often easily discernible. For instance, it is extra-linguistic knowledge that makes the distributive reading readily accessible to *The circles surround the dot*, but unavailable to *The dots surround the circle*. The effect of current *linguistic* context may be more systematic, if less radical. For instance, plural anaphoric pronominals are less easily interpreted collectively if their linguistic antecedent appears in two distinct noun phrases than if it appears as a single (e.g., conjoined) noun phrase (Moxey et al. 2004, 2011).

A contextual linguistic factor whose possible effects have not been experimentally studied so far is information structure (IS). Our experiment addresses this paucity by looking at the influence of topic and focus status in Hungarian on the availability of distributive and collective interpretations of three different types of QNPs. In the remainder of this section, we introduce our assumptions regarding these three types of QNPs, preceded by a brief review of the particular notions of topic and focus to be drawn on.

2.2 Topic, Focus, and Quantified Indefinites

The terms topic and focus are understood here to correspond to specific notions of Information Structure. In particular, they are rooted in an approach to IS that takes it to be a manifestation of sentence-level information packaging that responds to the immediate communicative needs of interlocutors (Chafe 1976), or from a different perspective, the management of the Common Ground (Krifka 2008). Specifically, the notion of topic, unless otherwise noted, will be used to refer to the 'psychological subject' (von der Gabelentz 1896): that which the sentence is said to be 'about'

[5]Following Szabolcsi (1997, p. 133), those NPs are essentially quantificational whose determiner/quantifier is not purely intersective and which cannot denote (singular or plural) individuals. In this paper we reserve the terms 'quantificational' and 'quantifier' phrase for QNPs that are interpreted as quantificational in this sense. For the differential behaviour of different types of universally quantified NPs, see Gil (1995) and Beghelli and Stowell (1997).

(whence the term *aboutness topic*). This corresponds to the entity or set of entities under which the information expressed in the rest of the sentence should be stored in the Common Ground (Reinhart 1981; Portner and Yabushita 1998), and which serves as the pivot for truth-value assessment (Strawson 1964). Aboutness topic NPs do not necessarily have to be given or discourse-old (Reinhart 1981; Frascarelli and Hinterhölzl 2007), but sentences containing them carry a presupposition of existence: namely, that the set denoted by the NP (or its nominal restriction part) is non-empty (Strawson 1950, 1952; Cresti 1995). Accordingly, prototypical topics are referential definite NPs, which are lexically associated with the same presupposition of existence as that associated with aboutness topic status (in addition, they are also characterized by uniqueness/maximality). As for indefinites, when they function as an aboutness topic, they are interpreted as specific.[6] Thus, (3a), with 'a girl' being in a topic position, must be interpreted as involving a girl who is a member of a set of girls already introduced in the discourse. The same restriction is absent in (3b), in which 'a girl' is in an information structurally neutral, post-verbal position.

(3) a. Egy lányt megláttam az épület előtt.
a girl.ACC PRT.saw.1SG the building in.front.of
[DET+N]$_{TOP}$ PRT+V DET N P
'I saw a girl outside the building.'

b. Megláttam az épület előtt egy lányt.
PRT.saw.1SG the building in.front.of a girl.ACC
PRT+V DET N P [DET+N]$_{NEU}$
'I saw a girl outside the building.'

The other core IS status whose possible impact will be explored is focus. Focus presupposes the presence of alternatives that are relevant to the interpretation of the sentence (Rooth 1985, 1992). The alternatives may be of a wide range of different semantic types, including individuals, as in (4a) and (4b), as well as properties, locations, times, cardinalities, and so on, as long as their semantic type is identical to that of the focused element. The set of alternatives is restricted both by world knowledge and by context: in (4a) this set may be, for instance, just the neighbours. The focus (marked by subscript 'F' below) is contained in a constituent referred to as 'focus phrase' (marked by 'FOC') such that the set of focus-alternatives contains (the denotation of) alternative focus phrases that differ from each other in just the focused element (Krifka 2006). In (4b), for instance, the set of focus-alternatives

[6]For the specificity of ordinary (non-contrastive) aboutness topics, see Reinhart (1981), Cresti (1995), Erteschik-Shir (1997), Portner and Yabushita (1998, 2001), Portner (2002), Endriss (2009), Dalrymple and Nikolaeva (2011), Geist (2011). According to Farkas (2002), specificity of an NP corresponds to the restricted nature of the variation of values for the variable that the NP introduces. She distinguishes epistemic specificity (in terms of speaker's knowledge), scopal specificity (in terms of the ability to escape scope islands) and partitive specificity (in terms of being part of a set already present in discourse). According to É. Kiss (2005), it is partitive specificity that is relevant to the Hungarian topic position.

includes not just *any* alternatives to John's mother (such as Bill, or Bill's brother), but only individuals who are mothers of alternatives to John, such as Bill's mother. In (4c), focus-alternatives are different cardinalities of students.

(4) a. $[[\text{MARY}]_F]_{FOC}$ saw John's mother.
b. Mary saw $[\text{JOHN}_F$'s mother$]_{FOC}$.
c. Mary saw $[\text{THREE}_F$ students$]_{FOC}$.

The type of focus we will be concerned with is ordinary information focus, understood here to refer to that part of the sentence that constitutes new, non-presupposed information in relation to a current Question Under Discussion (Roberts 1998). A common way to diagnose information focus in a sentence is to set it in the context of an ordinary information question. In a congruent discourse, the *wh*-item of the question (e.g., in the case of (4b), *Whose sister did Mary see?*) corresponds to the information focus in the answer. Information focus is interpreted by default as providing an exhaustive answer, a phenomenon often modeled in terms of a silent exhaustivity operator (Groenendijk and Stokhof 1984), which has the same interpretation as the exhaustivity operator that gives rise to scalar implicatures (Chierchia et al. 2013).[7] On an exhaustive interpretation of a sentence S containing a focused element E, call it S(EF), certain counterparts of S(EF) are held to be false. All excluded counterparts differ from the broad focus version of S, call this S(E), only in replacing E with one of its alternatives E' in S(EF), call these broad focus counterparts S(E'). In fact, not all focus-alternatives are excluded in this way, but only those for which S(E') is not already entailed by S(E) (Schwarzschild 1994; Fox 2007; Chierchia et al. 2013).[8] To illustrate, if in (4c) the contextually relevant alternatives are natural numbers, then it can be inferred from (4c) that it is not the case that Mary met four students. This is so because four is a focus-alternative to three in (4c), and *Mary saw three students* does not entail *Mary saw four students*. On the other hand, since *Mary saw three students* entails *Mary saw two students*, the latter is not excluded by (4c).

The experiment to be presented below concentrates on the interpretation of three types of QNPs as a function of their IS status: bare numeral indefinites like *öt diák* 'five students', upward entailing comparative numeral phrases like *több mint három diák* 'more than three students', and *sok diák* 'many students'. We included these three types primarily because much of the existing literature concentrates on the distributivity–collectivityof definite NPs: typically, NPs introduced by a definite determiner, universal QNPs, or conjoined proper names. In other words, our experiment also sought to address the relative paucity of empirical data with regard to the distributive–collective interpretation of indefinites.

[7]See also Horvath (2007) for the view that such an exhaustivity operator is syntactically associated with the pre-verbal focus position in Hungarian.

[8]Wagner (2006, 2012) argues for a stronger restriction: the alternatives of an exhaustive focus must be mutually exclusive in the context of its sister constituent.

Table 1 Interpretations available to the investigated QNPs

	Cardinal NP	*many*-NP	Comparative numeral NP
Cardinal	yes	yes	yes
Quantificational	no	yes (=proportional)	yes

The inclusion of these particular QNPs also serves the purpose of studying any differences across plural indefinites since the three QNP types fall into two semantic subclasses. Bare numeral NPs are non-quantificational indefinites, whose plural numerals are cardinal modifiers predicating the number of atomic individuals in the denotation of the noun (Kamp and Reyle 1993; Winter 1997). The existential reading of cardinal indefinites comes about by Existential Closure (Heim 1982; Winter 1997, 2001). Comparative numeral NPs and *many*-NPs, on the other hand, are ambiguous between a genuinely quantificational and a non-quantificational interpretation. This is the standard view of *many*: its quantificational interpretation is the non-intersective, proportional reading ('Many students arrived' roughly meaning 'A great proportion of the students arrived'), and the non-quantificational indefinite reading is an intersective, cardinal reading ('A large number of students arrived') (Milsark 1979; Partee 1989).[9] A common approach to the comparative modified numeral QNP *more than n N* is that it is quantificational. The view that it is a simple generalized quantifier phrase has recently been both questioned (Hackl 2000; Takahashi 2006) and defended (Mayr and Spector 2012; Syrett 2018).[10] The choice between the alternative approaches to the quantificational reading of comparative numeral phrases will be immaterial for our present purposes. What is more important is that *more than n N* QNPs also have another, namely, an individual-denoting, cardinal indefinite reading (Constant 2012, 2014). This specific indefinite reading is exemplified by the sentence below, involving appositive post-modification:

(5) More than ten conference participants, who are incidentally all vegetarian, didn't come to the dinner.

Table 1 serves to summarize the interpretive options available to the three QNPs that our paper is concerned with.

With this much background in place, in the next section we present an experiment whose aim was to explore the potential impact of aboutness topic and information focus status on the resolution of the distributive–collective indeterminacy in sentences containing semantically plural quantified indefinites.

[9]For a different view of English *many*, see Solt (2009).

[10]Hackl (2000) suggests a decompositional approach to comparative QNPs. Takahashi (2006) argues for a different decomposition, according to which a comparative QNP like *more than three books* is decomposed into *two* generalized quantifiers: [*more than three*] and [n-many *books*].

3 Experimental Treatment

To investigate whether and how the Information Structural topic and focus roles affect the availability of the distributive and collective readings of quantified indefinites in semantically and pragmatically ambiguous sentences, a two-answer forced-choice (2AFC) sentence-picture matching experiment was performed. Each sentence contained as its grammatical subject a QNP belonging to one of the three types introduced in Sect. 2.2. The aim of varying the lexical type of the QNP was to explore whether the putative effect of Information Structure is uniform or differential across lexical QNP-types. The key question in this regard was whether the behaviour of QNPs that are interpreted unambiguously as cardinal indefinites would exhibit any differences from the behaviour of QNPs that have both a cardinal and a quantificational reading.

3.1 Pretests

In order to ensure that the target stimuli in the critical conditions are truly ambiguous, the target sentences were selected on the basis of a pretesting process. To warrant that no unwanted pragmatic effects bias the semantically ambiguous candidate sentences towards either of the readings, a series of reading availability tests had been conducted, in which participants judged the availability of distributive and collective readings.

3.1.1 Participants

One hundred ninety-nine linguistically naive participants were crowd-sourced through Facebook advertisements out of which forty-one were excluded due to the violation of at least one control item. In the ad-campaign, a random-lottery incentive system was utilized: by participating in the survey, subjects had a chance to win gift-cards. Participants were all native speakers of Hungarian. While the age and gender of the respondents were not logged, the advertistments calling for participation were only served to Hungarian residents above the age of 18.

3.1.2 Material

Aiming to control for variables that can potentially influence the resolution of the distributive–collective ambiguity (see Sect. 2.1), in all target sentences of both the pre-tests and the main test, QNPs were grammatical subjects and functioned as agents. In addition, verbs were transitive, telic and appeared in past tense, while the object nominal was introduced by an indefinite article. The ultimate goal of the reading availability pre-tests was to establish a set of ambiguous sentences in which both distributive and collective readings are available not only semantically but also pragmatically. Since the objective of the main experiment was to investigate the effect of topic and focus IS roles, QNPs occurred in both topic and focus positions. As a basis for comparisons in the main test, information structurally neutral, post-verbal QNPs were also included.

Owing to the fact that it has distinct sentential positions dedicated to topic and focus, Hungarian furnishes convenient testing ground to investigate the effect of these IS roles on the resolution of the collective-distributive ambiguity (for an overview, see É. Kiss 2002). Both topics and foci are fronted to the left of the finite verb, where they license different word orders. In neutral sentences without a narrow focus, the verb immediately follows its verbal particle (if it has one). In such sentences a pre-verbal nominal that precedes a non-focused pre-verbal sentence adverbial like *last week* or *in all likelihood* must be interpreted as a topic. By contrast, in sentences containing a narrow focus, it is the focused phrase (in the sense of Krifka 2006; see Sect. 2.2) that immediately precedes the verb, and the verbal particle must surface post-verbally.[11] When the noun is post-verbal, it is information structurally neutral: it is not construed as either a topic or a focus. Therefore in our sentences with the QNP in topic position, a high adverbial followed the QNP, which was in turn followed by a verbal particle and the verb (6). In sentences with the QNP in focus position, the verb was inverted to the left of the verbal particle, and the adverbial was post-verbal (7). Finally, information structurally neutral QNPs were positioned after the particle and the verb, while the adverbial occupied a pre-verbal position (8). Importantly, in all three word orders the subject QNPs linearly preceded the indefinite object.

(6) Öt turista tegnap felvert egy sátrat.
five tourist yesterday up.set a tent.ACC
$[Q+N]_{TOP}$ ADV PRT+V *an* N QNP in topic position
'Five tourists yesterday set up a tent.'

(7) Öt turista vert fel tegnap egy sátrat.
five tourist set up yesterday a tent.ACC
$[Q+N]_{FOC}$ V PRT ADV *an* N QNP in focus position
'It is five tourists that set up a tent yesterday.'

(8) Tegnap felvert öt turista egy sátrat.
yesterday up.set five tourist a tent.ACC
ADV PRT+V $[Q+N]_{NEU}$ *an* N QNP in neutral position
'Yesterday five tourists set up a tent.'

[11] Sentences with a pre-verbal focus are the canonical form of answers to ordinary *wh*-interrogatives. Taking them to provide a *semantically* exhaustive answer to the current question, É. Kiss (1998) refers to pre-verbal foci as 'identificational' foci. The pre-verbal focus position in Hungarian is pragmatically unmarked with respect to contrast: it is not inherently associated with contrastive interpretation (unlike, for instance, fronted focus in Italian; Rizzi 1997).

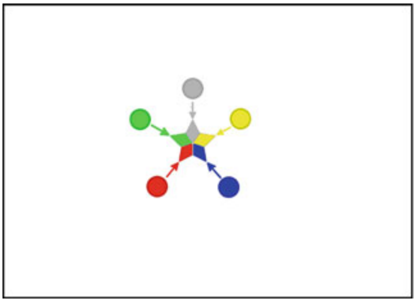
As indicated in Sect. 2.2, the quantified indefinite subject NPs fell into one of three categories: bare cardinal numeral NPs (e.g., *five kids*), upward entailing comparative modified numeral NPs (e.g., *more than three kids*) and *many*-NPs (e.g., *many kids*). The particular QNP types included in the investigation share three important properties. Morphologically, they are all singular in Hungarian and hence they trigger the same singular agreement on the verb when they function as subjects.[12] Also common to all three types is that they are not necessarily interpreted as essentially quantificational QNPs (which would limit their interpretation to distributive readings, see Sect. 2.1 above). Finally, in addition to an information structurally neutral post-verbal position, each of them is able to occupy both the topic position and the focus position, making the required comparisons possible.[13] A single quantifier expression was selected from each type: *öt* (i.e., five), *több mint három* (i.e., more than three) and *sok* (i.e., many). Importantly, these expressions can be visually depicted by sets of the same size (i.e., sets with a cardinality of five). This is desirable since the processing of the visual representations of divergent numerosities may introduce unwanted cognitive artifacts in judgment tasks.

3.1.3 Design and Procedure

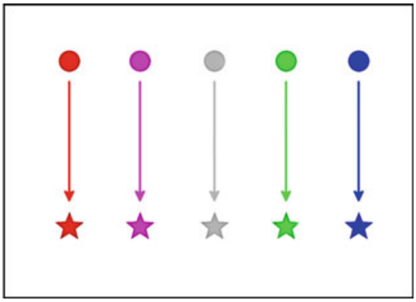
Since the pragmatic biases introduced by the different components of candidate sentences were beyond the intuitions of single individuals, a total of six sub-experiments were conducted, each with a different group of participants. The number of subjects who passed the controls varied across sub-experiments (median number of participants: 21.5 [18–50]). Each iteration consisted of 27 pairs of target sentences. The 27 target sentences were produced by taking three lexicalizations from each of the 3 × 3 (IS-role × QNP-type) factor-combinations. Candidate sentences were presented two times in a randomized order. At each presentation of the same sentence a static image was shown, displaying either a collective (Fig. 1a) or a distributive scenario (Fig. 1b). It was one of these two schematic visual representations that was presented together with each item. The pre-test also included 40 semantically ambiguous but pragmatically biased fillers. Further, it included 15 unambiguous sentences, having either only a collective reading or only a distributive reading. The latter served as controls, in order to be able to filter out inattentive or careless participants, or those who misunderstood some aspect of the task. In total, each participant judged 109 sentences. Filler and control sentences were identical in every version of the experiment and the presented visual scenarios were counterbalanced.

[12] In this regard Hungarian differs, for instance, from English, where only morphologically plural QNPs can combine with collective predicates like 'gather' or 'lift a piano together' (see Winter 2001, who analyzes these as set predicates).

[13] The latter does not hold of negative QNPs and other downward entailing QNPs like *kevés(ebb mint öt) diák* 'few(er than five) students' and *legfeljebb öt diák* 'at most five students', or the upward entailing synthetic comparative *háromnál több diák* 'more than three students', none of which can be topicalized, or the definite superlative *a legtöbb diák* 'most students', which cannot be focused (cf. Szabolcsi 1997).



(a) Visual representation corresponding to the collective interpretation of sentences.



(b) Visual representation corresponding to the distributive interpretation of sentences.

Fig. 1 Visual depiction of readings utilized in the reading availability judgment pretests as well as in the main experiment

Subjects were first introduced to the phenomenon via a practice phase. Here they were familiarized with the visual notational convention and the divergent reading types. In the test phase they were asked to judge whether a presented sentence has the meaning depicted by the image presented alongside experimental items (2AFC judgment task). They were told that there were no right or wrong answers.

3.1.4 Outcome

To establish the availability of readings, chi-squared test statistics were utilized. Candidate sentences, whose number of distributive and collective reading availability judgments did not differ significantly from each other ($p > .05$) were deemed ambiguous and were involved in a manual selection process. Out of the 162 candidate sentences 79 passed the insignificance criterion.

3.2 Main Test

Following the empirical validation of the availability of both readings in candidate sentences, the selected constructs were tested to see if the readiness of distributive and collective readings differs across QNPs in divergent informational structural roles.

3.2.1 Participants

In a different online ad campaign one hundred fifty-nine subjects were recruited to participate in the experiment. The campaign followed the same random-lottery incentive-system utilized in the pretests. Subjects were all adult native Hungarian speakers, their gender and age were mixed (120 female and 39 male participants, mean age: 43.55 [18–64]).

3.2.2 Design and Procedure

A two-answer forced choice sentence-picture matching task was designed and was implemented in IBEX (Drummond 2010). To minimize the chance of participants filling out the experiment in potentially distracting public areas, the usage of portable devices (e.g., smart phones, tablets) was disabled. The first screen explained and illustrated the distributive–collective ambiguity and participants were introduced to the visual notations that had to be paired with the textual stimuli later on (these were the same as the ones used in the pre-test; see Fig. 1). The presentation of stimuli began after this page. Each sentence was displayed on a separate slide printed above the schematic visual representations of the two readings. These visual representations were identical across all items. This was made possible by our choice of the three QNPs: each of them matched the same number of depicted elements (=five). The purpose of this was to avoid any confounding effects that differences in visual stimuli may have otherwise had on the results. The order of the two displayed scenarios was randomized and counter-balanced. For picking a reading, subjects were required to respond by pressing the F and J key on their computer keyboard (corresponding to the visual representations appearing on the left or right sides of their display). The presentation began with 6 practice items (2 obligatory collective, 2 obligatory distributive and 2 ambiguous sentences). This phase was followed by 18 targets, 39 semantically ambiguous fillers that were pragmatically biased to one of the readings, as well as a total of 17 semantically unambiguous distributive and collective control items shuffled together in an individually randomized fashion. Once responding to a sentence, a 1000 ms separator block followed, after which participants had to skip to the next slide by pressing a button. In total, participants judged 80 sentences, which required approximately 25 minutes. Besides registering judgments, reaction times were also measured for filtering out potential careless respondents (see Sect. 3.2.4).

3.2.3 Materials

Sentences followed the structure described in Sect. 3.1.2. Utilized constructs were selected from among those candidate sentences which passed the reading availability tests. Keeping in mind the 3×3 factorial design, in a manual selection process the sentences whose judgments differed the least were adopted as target items for the main experiment. In the manual selection process 3 lexicalizations were picked for each combination of factors (i.e., for each condition).

To compensate for the low number of lexicalizations, further items were generated by reintroducing QNP-type as a between-subject factor. QNP-type re-factorization was executed by rewriting each ambiguous target sentence with all three numerical determiners. By adopting this practice, three sub-experiments were created where each sub-experiment included only a single QNP-type. In this way, a total of 18 lexicalizations were obtained for each level of the original QNP-type factor. In the end, a total of 6 lexicalizations were constructed for each of the 3×3 conditions.

3.2.4 Analysis

An inherent issue of Internet-based experiments is the elicitation of high quality performance from an incentive driven crowd. Following the execution of the ad campaign, responses from each version of the experiment underwent a filtering process to exclude spammers and careless respondents. Filtering was done by a program which was developed for this purpose. Exclusion of participants was partially criteria dependent and was partially data-driven:

- No more than 4 'careless responses' were tolerated. Given that on an average humans read a single word in 320–370 ms (Trauzettel-Klosinski and Dietz 2012), we set up an RT-based criterion to spot careless responders. We discarded the responses of those participants who responded to at least 5 stimulus sentence faster than the number of words in the given sentence multiplied by 125 ms.
- No more than 3 'fast responses' were tolerated. Fast responses were button presses happening sooner than 400 ms post stimulus-onset. Participants committing fast responses were warned that they might be excluded from the incentive-lottery.
- No more than 4 'slow responses' were allowed. To make sure that subjects pay close attention to the task, the results of those who committed at least five such judgments, i.e., responded slower than the mean plus one standard deviation of all responses, were discarded. This value in our experiment was 42,000 ms.
- Participants who interrupted the survey at least once (i.e., responded slower than 120,000 ms) were also discarded.
- Participants committing to too many constitutive identical responses were also filtered. Since each participant saw a differently randomized version of the experiment, the response of consecutive responders were removed algorithmically, based on the median absolute deviation (Iglewicz and Hoaglin 1993) trends of their consecutive answer counts.
- No more than 3 control violations were tolerated (corresponding to roughly 80% accuracy). The responses of participants who provided erroneous collective/distributive judgments to at least 4 semantically and pragmatically unambiguous distributive/collective sentences were discarded.
- To reduce the probability that subjects pass the control conditions by chance, participants who were significantly faster (i.e., produced a t-score > 2) at responding to ambiguous target sentences than to unambiguous controls were also dropped.

Following the filtering process, the response of 114 participants were kept. The distribution of responses to sub-experiments centering around divergent QNPs was near-uniform (37 responses were collected to the sub-experiment containing *five* as a numeral, 38 to the *more than three* variant and 39 to the *many* variant).

To determine which factors or factor combinations usefully predict the results, a series of logistic mixed-effects models were fitted onto the merged sub-experiments.[14]

Candidate models were ranked following a complex metric which took both practical and statistical significance into account: to find the model which explains most of the observed variation, Pseudo-R^2 values were calculated (Nakagawa and Schielzeth 2013). Keeping in mind that selection based on these alone are subject to overfitting as more parameters in a model generally increase fitness, calculated fitness values were weighted by the Bayesian information criterion (BIC) score, which strongly penalizes additional parameters.[15] Candidate models are compared in Table 2.[16]

In addition, each sub-experiment was analyzed by comparing the ratio of collective responses per participant between the various levels of the IS-role factor, using the Wilcoxon signed rank test and Holm's multiple comparisons corrections. The same set of procedures was also employed to determine how judgments in different conditions deviate from chance-level. To contrast the behavior of QNP-types, Wilcoxon rank sum tests were utilized.

3.2.5 Results

Our primary finding is that both the IS status and the lexical type of QNPs modulate the resolution of the distributive–collective ambiguity. As shown in Table 2, candidate models lacking an IS parameter score low, and the top six candidates all contain both a QNP parameter and an IS parameter. ANOVA of Model 1 shows that both predictors produce a significant main effect: IS- ROLE($\chi^2(2) = 9.06, p = .01$), QNP-TYPE($\chi^2(2) = 6.87, p < .05$).

Collapsing the three QNP types, in sentences where the QNP occupies an IS-neutral position collective readings are preferred. The reading preference also deviates from chance level both in the topic position and in the focus position, this time in the direction of distributive readings (see Table 3a). It is not surprising then that when both factors are collapsed, the central tendency of collective-judgment ratios fail to show either reading to be more dominant ($V = 2399.5, p = .448$) (Fig. 2).

---

[14]The analysis was carried out by R 3.2.2. (R Core Team 2013). For fitting the candidate models, we relied on the function `glmer` from the package lme4 (Bates et al. 2014), while related fitness metrics were established by the function `sem.model.fits` from the piecewiseSEM package (Lefcheck 2015). Main effects were calculated by the function `Anova` of the car package (Fox et al. 2015).

[15]BIC weighted fitness scores were calculated based on the following formula: $BWF_i = \frac{1}{BIC_i} \times R^2_{GLMM(marginal)i} \times R^2_{GLMM(conditional)i}$. Reported scores are normalized: $\frac{BWF_i - min(BWF)}{max(BWF) - min(BWF)}$.

[16]While we assume that our filtering process should only bring about improved response quality by the reduction of spammer noise, we also acknowledge that any set of exclusion criteria may potentially affect model results. It must be noted, however, that—except for the more or less standard 80% control-passing criterion—filtering was executed on the basis of proxies of quality that are independent of the outcomes of the observed dependent variable (i.e., reaction time, reading time, response constituency).

Table 2 Comparison of fitted models. Row corresponds to fitted model, and × signs indicate included parameters. For fixed effects, colons indicate interaction between predictors. Plus signs in the case of random effects indicate the presence of random slopes corresponding to the fixed effects present in the model, colons indicate those configurations where an interaction is expected in the random slope structure. Models are ranked in order from best to worst based on their BIC weighted fitness score explained in Sect. 3.2.4

Model	Fixed effects			Random effects						BIC	Pseudo-R^2		BIC weighted fitness score
	IS	QNP	:	SUBJECT	+	:	ITEM	+	:		Marginal	Conditional	
1	×	×		×	×		×	×		2690.288	0.059	0.410	0.975
2	×	×		×			×	×		2596.966	0.057	0.395	0.943
3	×	×	×	×	×	×	×			2855.553	0.058	0.410	0.897
4	×	×	×	×			×	×	×	2853.684	0.060	0.395	0.894
5	×	×		×	×		×			2599.063	0.052	0.404	0.871
6	×	×	×	×	×	×	×	×	×	3174.154	0.059	0.414	0.827
7	×			×	×		×	×		2557.915	0.045	0.406	0.777
8	×			×			×	×		2530.277	0.044	0.400	0.753
9	×			×	×		×			2524.785	0.042	0.399	0.719
10	×	×					×	×		2648.797	0.055	0.304	0.681
11	×	×	×				×			2588.003	0.055	0.297	0.678
12	×	×	×				×	×	×	2905.424	0.058	0.306	0.659
13	×	×					×			2558.808	0.049	0.296	0.603
14	×						×	×		2591.845	0.044	0.301	0.537
15	×						×			2560.731	0.040	0.288	0.478
16	×	×	×	×	×	×				3126.527	0.049	0.144	0.225
17	×	×		×	×					2876.373	0.044	0.140	0.214
18	×	×	×	×						2807.818	0.047	0.115	0.190
19	×			×	×					2801.404	0.035	0.137	0.163
20	×	×		×						2786.421	0.042	0.109	0.158
21		×		×			×	×		2532.550	0.010	0.379	0.136
22		×		×	×		×	×		2570.222	0.010	0.380	0.131
23		×		×	×		×			2534.294	0.009	0.391	0.126
24	×			×						2778.705	0.033	0.109	0.119
25		×					×	×		2584.476	0.009	0.285	0.090
26		×					×			2548.563	0.009	0.296	0.086
27		×		×						2828.349	0.009	0.073	0.000
28		×		×	×					2866.200	0.009	0.073	0.000

Wilcoxon sign ranked contrasts of IS-roles collapsed over the QNP-type factor reveal that the Information Structural topic and focus status of QNPs both affect ambiguity resolution patterns significantly differently from the neutral status. Further, QNPs with topic and focus roles also behave differently from each other: topicalized QNPs are significantly more collective than their focalized counterparts (see

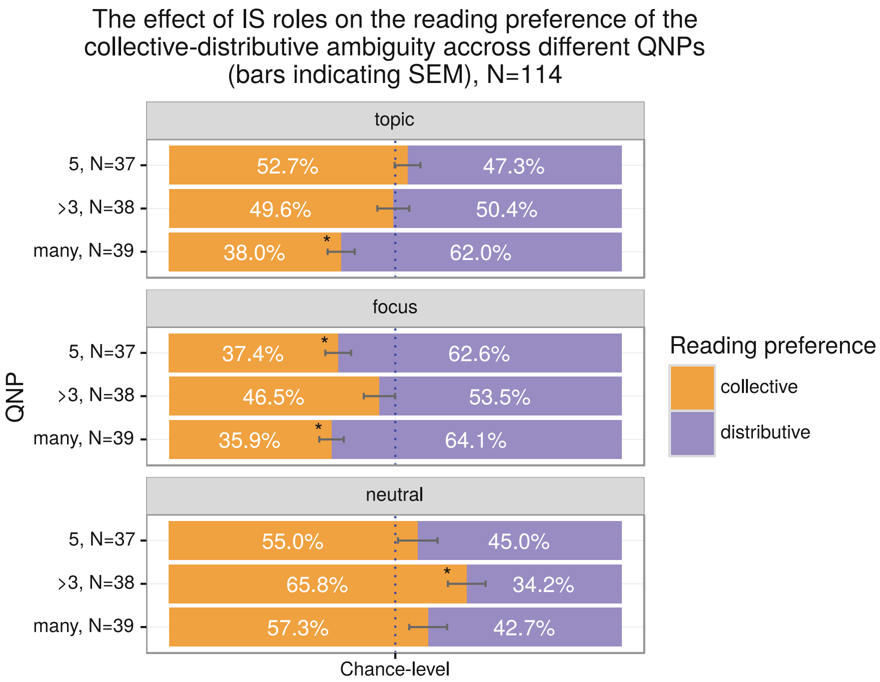


Fig. 2 Proportion of distributive and collective responses. Values on bars show the mean judgment ratios corresponding to collective and distributive readings, error bars report standard error based on the judgment ratio of individuals, asterisks next to error bars indicates significant deviation from the chance-level ($p < .05$)

Table 3b). This is mainly due to bare numeral QNPs: among the three QNP types in topic, only bare numeral QNPs differ from their focalized counterparts; in a topic position they show no deviation from their neutral variant. The other two QNP types, comparatives and *many*-NPs, display similar behavior in topic and in focus, with both variants diverging from their neutral counterparts in the distributive direction (see Table 5).

The effect of QNP-type is less pronounced when collapsed over the three IS statuses. Only *many*-NPs deviate from chance-level, namely, in the direction of distributive readings (see Table 4a). Wilcoxon rank sum tests reveal that bare numerals and comparative numeral NPs, as well as bare numerals and *many*-NPs display similar behavior; their differences fail to approach significance. Comparative numeral NPs and *many*-NPs cue reading judgments in a significantly different manner (see Table 4b). This difference can also be seen in the fact that in the neutral position, serving as the baseline, while *many*-NPs are not significantly different from chance level, comparative numeral NPs diverge from it in the collective direction (see Table 5a).

Table 3 Descriptive statistics (**a**) and corresponding IS-role contrasts (**b**) of experimental results collapsed over the QNP factor

(a) *Descriptive statistics of experimental results collapsed over the QNP factor, supplemented with Wilcoxon test results for deviation from randomness comparisons. Mean and SD values were calculated based on the collective-preference ratio of individual participants. Reported p-values are Holm-corrected.*

IS-role	mean	SD	V-value	p
Topic	0.466	0.202	828.5	<.005
Focus	0.399	0.190	299	<.001
Neutral	0.594	0.263	2927	.015

(b) *Summary of Wilcoxon contrasts of IS-roles collapsed over the QNP factor. P-values underwent Holm's multiple comparisons correction.*

Contrast	V-value	p
Topic–focus	2432.5	.008
Topic–neutral	975	<.001
Focus–neutral	454	<.001

Table 4 Descriptive statistics (**a**) and corresponding QNP-type contrasts (**b**) of experimental results collapsed over the IS factor

(a) *Descriptive statistics of experimental results collapsed over the IS factor, supplemented with Wilcoxon test results for deviation from randomness comparisons. Mean and SD values were calculated based on the collective-preference ratio of individual participants. Reported p-values are Holm-corrected.*

QNP-type	Mean	SD	V-value	p
5	0.483	0.163	251.5	.435
>3	0.539	0.171	434.5	.223
many	0.437	0.156	134.5	<.05

(b) *Summary of Wilcoxon contrasts of QNPs collapsed over the IS factor. P-values underwent Holm's multiple comparisons correction.*

Contrast	W-value	p
5–>3	550	.208
5–many	806.5	.376
>3–many	1001	.05

The same difference resurfaces both in the focus and in the topic position, where both are interpreted significantly more distributively (see Table 5b). In particular, in both of these IS positions only *many*-NPs diverge from chance level, namely, in the direction of distributive responses (see Table 5a).

Table 5 Descriptive statistics (**a**) and corresponding contrasts of sub-experiments (**b**)

(a) *Descriptive statistics of subexperiments, supplemented with Wilcoxon test results for deviation from randomness comparisons. Mean and SD values were calculated based on the collective-preference ratio of individual participants. Reported p-values are Holm-corrected.*

IS-role	QNP	N	Mean	SD	V-value	p
Topic	*5*	37	0.527	0.174	123	>.999
	>3	38	0.496	0.217	108	>.999
	many	39	0.380	0.187	59.5	<.001
Focus	*5*	37	0.374	0.173	10	<.001
	>3	38	0.465	0.213	113	.195
	many	39	0.359	0.169	6	<.001
Neutral	*5*	37	0.550	0.266	227	>.999
	>3	38	0.658	0.257	419.5	.05
	many	39	0.573	0.262	356	>.999

(b) *Summary of Wilcoxon contrasts of IS-roles per subexperiments. P-values per sub-experiment underwent Holm's multiple comparisons correction.*

QNP	Contrast	V-value	p
5	Topic–focus	990.5	<.005
	Topic–neutral	675.5	>.999
	Focus–neutral	433.5	.022
>3	Topic–focus	750.5	>.999
	Topic–neutral	452.5	.022
	Focus–neutral	411	.006
many	Topic–focus	778.5	>.999
	Topic–neutral	408	<.005
	Focus–neutral	374.5	<.001

4 Discussion

Before discussing our findings in detail, it is worth addressing an apparent contrast between the estimated salience of the collective readings in our experiment and that reported in several earlier empirical investigations (Brooks and Braine 1996; Frazier and Clifton 2001; Kaup 2002; Boylan et al. 2011; Bosnic 2016; see also Sect. 2.1). While earlier accounts emphasized the prevalence of collective readings, our treatment did not return a robust collective preference even for QNPs with a neutral IS role. Among such QNPs, a significant collective preference was only found with upward entailing comparatives, but even here the mean rate of collective responses only reflected a relatively weak bias.

Despite this difference, our results do not contradict earlier empirical generalizations, for at least two reasons. First, in prior empirical studies the collective preference was found to be strong in ambiguous sentences that contain non-quantified definite subject NPs. By contrast, the present experiment involved three types of indefinite

NPs, one of which only has a cardinal reading, while the other two have both a cardinal and a quantificational interpretation. Second, the objective of the pre-testing process resulting in the items to be investigated in the main experiment was precisely to find sentences for each QNP type that can reasonably be expected to have both a distributive and a collective reading. As will be recalled, the purpose of excluding examples inducing a strong distributive or collective preference was to weaken any general pragmatic biases with a potential masking effect, thereby allowing Information Structural manipulations in the main experiment to have a measurable impact. Accordingly, instead of the absolute rates of distributive and collective responses, what is directly relevant to the concerns of this paper is any significant effects exercised (within each QNP type) by the two IS roles, as compared to the neutral status.[17]

With this clarification in place, we can now proceed to a discussion of the actual outcomes. The overall results show that both the IS status and the lexical type of semantically plural quantified indefinite NPs affect the resolution of the distributive–collective ambiguity. (i) First, the focus status of QNPs affects ambiguity resolution patterns significantly differently from a neutral status in all three QP-types examined. In each case, focus status shifts the rate of responses in a distributive direction. (ii) Second, as compared to a neutral IS status, the topic role increases the rate of distributive interpretations in a similar way in the case of *many*-NPs and upward entailing comparative numeral NPs. (iii) Third, as opposed to *many*-NPs and comparative numeral NPs, the topic role had no significant effect on the ratio of distributive and collective readings assigned to bare cardinal numeral NPs, in comparison to their Information Structurally neutral status. In the remainder of this section we discuss each of these results.

The first finding, namely that focus shifts the rate of responses in a distributive direction in each of the three QNP types, can be understood, we argue, on the basis of the interpretation of focus. As noted in Sect. 2.2, focus presupposes the presence of a set of alternatives relevant in the given context to the interpretation of the sentence (Rooth 1985, 1992). The set of alternatives belong to the same semantic type as the focused phrase, and differ from each other only in the narrowly focused element within the focused phrase. Given the absence of context and the non-auditory mode of presentation, stimulus sentences in which the QNP was a focused phrase permit two focus-interpretations. On one of them, the focused element is the whole QNP, while on the other one what is narrowly focused is just the numeral/quantifier before the noun.[18] As pointed out in Sect. 2.2 (see Krifka 2006), focus alternatives are alternatives to the whole focused phrase, which in our case corresponds to the entire QNP. Thus, despite the fact that in the two possible readings we have just mentioned the alternatives vary along different dimensions, the two interpretations both presuppose a set of alternatives to (the denotation of) the whole QNP.

[17]For the same reason, no conclusions should be drawn from any direct comparison of responses across different QNP types within the same IS condition. Note that due to our experimental design, analyses of this latter type would necessarily involve between-group comparisons.

[18]The first reading would answer an information question with *who* as its *wh*-phrase, while the second one would answer questions with *how many*. Although a third reading, on which it is the noun that is narrowly focused, is logically possible, in the absence of context it is highly implausible.

Recall that one of the three QNP types involved in the experiment, namely *five*-phrases, has a cardinal indefinite reading only, while the other two, namely *more than three*-phrases and *many*-phrases, are ambiguous between a cardinal indefinite and a quantificational interpretation (see Table 1 above). Quantificational readings are generally distributive (see Sect. 2.1), and there is no reason to assume that focusing quantificational QNPs alters this.[19] What we can conclude from this is that in the case of *many*-phrases and *more than three*-phrases it is the cardinal interpretation that must have been affected by focus: it is on their cardinal reading that they are interpreted more distributively in focus than in their neutral position. This is corroborated by the fact that bare numeral QNPs, to which only a cardinal interpretation is available, also gave rise to more distributive responses in focus than they did in a neutral position.

The question, then, is why QNPs interpreted as existential cardinal indefinites are more easily understood distributively when in focus than when they are neutral. We argue that the reason for this preference lies in processing, in particular it is due to the fact that the set of their focus-alternatives can be kept smaller.

To see why this is so, recall from Sect. 2.2 that a sentence containing an exhaustively interpreted information focus excludes only those alternatives that are not entailed by the neutral, broad focus counterpart of the sentence. As illustrated in relation to (4c) above, alternatives to a narrowly focused bare cardinal numeral are other cardinalities. Thus, in (7) the focus-alternatives are sets of individuals of different cardinalities (limited to tourists, if narrow focus is on the numeral only). Crucially, if the predicate phrase ‘set up a tent’ is interpreted distributively, cardinalities smaller than the asserted cardinality are not excludable, due to the distributive entailment characterizing distributive interpretations (see Sect. 2.1). For instance, if there are five tourists that *each* set up a tent (which is the distributive entailment of (7)), then it follows for all smaller (positive) cardinalities n that there are n tourists that each set up a tent. Thus, focus-alternatives corresponding to ‘four tourists’, ‘three tourists’ and so on are not to be excluded on a distributive reading. It seems plausible to assume that those focus-alternatives that can, and do, get excluded are more relevant to the interpretation of information focus than those that are entailed and therefore cannot be excluded. Our proposal is that in the course of interpreting (7) distributively, entailed alternatives of smaller cardinalities are not included in the relevant set of focus-alternatives. On their cardinal interpretation, the same reasoning applies, mutatis mutandis, to the distributive reading of focused ‘many tourists’ and ‘more than three tourists’.

Importantly, the same option is unavailable to a collective reading of (7). This is because on the collective interpretation the fact that there are five tourists that set up a (single) tent together does not entail for any smaller (positive) cardinalities n that there are n tourists that set up a tent together. As before, the same carries over to the collective reading of the cardinal interpretation of focused *many*-phrases and *more than three*-phrases.

---

[19]It has been suggested that quantificational NPs are not focusable in the pre-verbal focus position of Hungarian (for relevant discussion, see É. Kiss 1998). The account we present immediately below is compatible with that view, as it pertains to the cardinal indefinite interpretation of the QNPs.

The question now reduces to the significance of the set of alternatives in processing. The psycholinguistic relevance of focus-alternatives has been confirmed in a range of recent experimental studies (e.g., Braun and Tagliapietra 2010; Fraundorf et al. 2013; Kim et al. 2015). There is also evidence that activating a set of alternatives incurs processing costs. Cowles (2003) found significant Late Anterior Negativity (LAN) in the processing of sentences in which focus-alternatives were relevant, as compared to sentences in which alternatives were not relevant. LAN is attributed to increased working memory load, found, for instance, in filler–gap dependencies. Assuming that a smaller set of alternatives incurs smaller processing costs, it is expected that the activation of the set of focus-alternatives to cardinal QNPs is less costly on a distributive reading than it is on a collective reading, since on the former reading, as we have argued, the set of focus-alternatives can be kept smaller. This explains why a distributive interpretation of cardinal indefinites may be relatively preferred by the processor to the collective reading. And it also explains why this relative advantage of the distributive reading of QNPs obtains in focus, in comparison to the neutral, post-verbal position. Namely, neutral QNPs are not associated with a set of alternatives. As a consequence, in the case of neutral QNPs the size of this set is irrelevant to the resolution of the distributive–collective ambiguity.[20]

Consider now findings (ii) and (iii), namely that, as compared to a neutral IS status, the topic role increases the rate of distributive interpretations in the case of *many*-NPs and *more then three*-NPs, but it fails to do so in the case of *five*-NPs. Recall from Sect. 2.2 that prototypical topics are referential definites, which have an existential presupposition and are also associated with uniqueness/maximality. Indeed, it is commonly assumed that ordinary aboutness topics must be entity-denoting (denoting an individual or a set of individuals), and that quantificational NPs are not suitable aboutness topics (Reinhart 1981; Portner and Yabushita 1998).

While many types of QNPs indeed cannot function as ordinary aboutness topics, it has been recognized that some apparently can, and accordingly, they can occupy an aboutness topic position too (Ebert and Endriss 2004; Endriss 2006, 2009). These QNPs include both *many*-phrases and upward entailing comparative numeral phrases (Repp 2009, p. 403; Krifka 2006, p. 268). It has been suggested that quantificational NPs can function as topics only if they are (re)interpretable in some fashion that makes them suitable topics. According to the implementation of Ebert and Endriss (2004) and Endriss (2006, 2009), when a quantificational NP is topicalized, it is the QNP's minimal witness set that is interpreted as the topic (following Szabolcsi 1997). On Constant's (2014) alternative account, such topic QNPs are turned into

[20] Szabolcsi (1997, p. 130), discusses the collective and distributive readings of examples similar to our sentences containing a focused QNP. She takes the immediately pre-verbal QNPs in those examples to be non-focused Predicate Operators that perform a counting operation. Concurrently, she claims that the collective interpretation involving such counting QNPs generally arises in a manner that is distinct from collective predication proper (a manner that she does not spell out explicitly). Importantly, her account takes both the collective and the distributive interpretation of immediately pre-verbal Predicate Operator QNPs to be unmarked in the general case. Elsewhere she notes, however, that when plural Predicate Operators combine with telic predicates, this gives rise to a marked 'It takes *n* Ns to PRED' reading; but it is not fleshed out why this is so.

an individual-denoting expression by a choice function. Without going into technical details, both of these approaches derive the asymmetry that exists between quantified NPs that can occur as topics and those which cannot from the nature of the (re)interpretation applying to topic QNPs in interaction with the nature of the quantified NPs themselves.[21]

Importantly, construing a quantified NP as a topic does not alter the interpretation of the sentence in other ways, including its distributivity/collectivity properties (see Endriss 2009, pp. 236–237 for a formal account). To take a simple example, 'every student' and 'most students' are genuinely quantificational and cannot combine with collective predicates, and this also characterizes their reinterpreted, topical counterparts (e.g., **Every student / *?Most students probably surrounded the teacher*).

Quantificational NPs have one key property that makes them highly suitable aboutness topics: they are presuppositional. The notion that quantificational (or more generally, strong) NPs presuppose the existence of the non-empty set that they quantify over (at least in contingent, i.e., empirical, contexts, Lappin 1998) has venerable history (McCawley 1972; de Jong and Verkuyl 1985; Diesing 1992; Jaeger 1995; Zucchi 1995). Advocating this general approach, Geurts (2007) derives their existential import from a stronger requirement. In particular, for strong NPs to be interpretable, suitable referents that they quantify over must be recoverable from the discourse context; thus a genuine quantifier signals that its domain is present in the discourse (Geurts 2007, p. 270). By contrast, non-quantificational cardinal indefinites (and weak NPs more generally) do not themselves carry the same presupposition.

It is this difference, we argue, that ultimately underlies the biasing effect that the aboutness topic function had on *many*-phrases and *more than three*-phrases, an effect that *five*-NPs failed to exhibit. Specifically, if a QNP can have either a quantificational or a cardinal indefinite reading, as in the case of *many*-phrases and upward entailing comparative numeral phrases, then, unless contextual cues dictate otherwise, the aboutness topic role favours the quantificational interpretation. This is because, as we have pointed out, on their quantificational reading presuppositionality is part of the lexical semantics of these QNPs.[22]

---

[21] See also Partee (1987, p. 132), who notes, in relation to her type shifting operators, that "there remain NPs for which none of our operations provide *e*-type [individual-denoting, BS] readings; these, not surprisingly, are the ones traditionally thought of as most clearly 'quantificational': *no man, no men, at most one man, few men, not every man, most men*." Constant (2014) demonstrates that topicalizable quantificational NPs all imply that their 'reference set' (the intersection of their restrictor and scope) is non-empty, and they can behave as individual-denoting in a range of other syntactic contexts. Though he discusses contrastive topics, his analysis also extends to cover quantificational NPs functioning as ordinary aboutness topics. It is implausible that topical *more than three*-NPs and *many*-NPs in our target sentences were interpreted as contrastive, as they lacked any context that would license such an interpretation. This is implausible also because a contrastive topic interpretation would require not only alternative topics in the context, but also the presence of a narrow focus following the contrastive topic in the sentence (Gyuris 2009).

[22] In the case of *many*-NPs this amounts to the claim that their proportional reading is more prominent in the topic position than in their neutral, post-verbal position. This accords with our own native speaker intuitions. On the assumption that the subject of intransitive individual-level predicates are aboutness topics (Milsark 1979; Ladusaw 1994; Erteschik-Shir 1997), this also agrees with

Recall from Sect. 2.2 that quantificational QNPs only admit of a distributive interpretation. As noted immediately above in relation to the aboutness topic interpretation of QNPs, this Information Structural role does not alter their distributivity properties. Therefore, when they function as aboutness topics, quantificational QNPs continue to be interpreted only distributively. In effect, by boosting the proportion of the quantificational interpretations of a QNP, the aboutness topic role concurrently elevates the rate of its distributive readings.

The net result is precisely what we find: *many*-NPs and *more than three*-NPs are more readily assigned a distributive interpretation when functioning as aboutness topics than in an Information Structurally neutral, post-verbal position. By contrast, since they only have a non-quantificational, cardinal indefinite reading, the aboutness topic role cannot affect the rate of the quantificational readings of *five*-NPs. Thus, the proportion of their distributive and collective readings is not significantly altered by the topic position, compared to their IS-neutral position.

5 Conclusion

The distributive–collective ambiguity that obtains with semantically plural NPs is known to be affected by a range of lexical and grammatical factors, as well as by world knowledge. This pioneering study has investigated whether and how sentence-level pragmatics, specifically, the IS roles of focus and topic affect the resolution of this semantic indeterminacy. According to the results of the forced choice sentence-picture matching experiment we presented, focus and topic roles both significantly affect the rate of distributive and collective readings. While the effect of focus status is uniform across the types of indefinite QNPs examined, the effect of topic status is differential. Focus incurs a higher rate of distributive interpretations in all the three QNP types, namely bare cardinal numeral NPs, *many*-NPs, and upward entailing comparative numeral NPs. Topic status, on the other hand, causes only the latter two types of QNP to shift in the distributive direction, while bare cardinal numeral NPs remain unaffected.

We argued that the observed effect of the focus role is due to the need to activate a smaller number of relevant focus alternatives in the course of the interpretation process in the case of the distributive reading, as compared to the collective reading. This difference is a consequence of distributive entailments down to atomic individuals inherent to the distributive interpretation, a property absent from the collective interpretation. The impact of topic status, on the other hand, ultimately derives from the presuppositional nature of aboutness topics. *Many*-NPs and upward entailing comparative numeral NPs are ambiguous between a quantificational and a cardinal

---

Diesing's (1992) observation that *many*-NPs can only be interpreted proportionally as the subject of individual-level predicates (i), as opposed to stage-level predicates (ii).
(i) Many firemen are altruistic. (proportional only)
(ii) Many firemen are available. (proportional or cardinal)

reading. As topics they are more likely to be interpreted on the basis of their quantificational meaning, as quantificational NPs lexically come with an existence presupposition, a property that also characterizes the topic function. Since quantificational NPs are lexically distributive, the result is an elevated rate of distributive readings. Because bare cardinal numeral NPs are unambiguously non-quantificational, their topic status has no effect on the proportion of distributive and collective interpretations.

It is worth drawing attention to two dimensions along which the generalizability of our conclusions may be tested by future research. First, the present study concentrated on three types of plural NPs that share the property of being indefinite. It remains an open empirical issue whether the interpretation of definite plurals, like conjoined names or NPs introduced by a definite determiner, is also affected in relevant ways by their information structural status. As definites are presuppositional independently of their topic or non-topic status, in their case we do not expect the topic role to have an impact on the resolution of the distributive-collective ambiguity. Second, although our experiment involved the two central information structural roles, namely topic and focus, we kept to a specific (albeit typical) instantiation of each. In particular, we examined the effect of aboutness topic and information focus status (see Sect. 2.2). Common approaches to information structure, however, admit of several other types of topic and focus, and only some of these share those interpretive properties of aboutness topic and information focus, respectively, that figured in our account (namely, presuppositionality, and the relevance of alternatives).[23] Whether and how such distinctions within the two key information structural roles may bear on the choice between a distributive and a collective interpretation is still to be explored.

From a broader perspective, our core finding that topic and focus status can affect the interpretation of sentences that are ambiguous between a distributive and a collective reading has a clear methodological repercussion: it points to a need to carefully control IS status in any experimental treatment of the interpretative processes that take part in the resolution of this ambiguity. This is pertinent not only to the study of the phenomenon in discourse configurational languages, like Hungarian, but also in languages, such as English, that do not routinely distinguish topic and focus functions in their syntax. As nothing in our account of the experimental outcomes directly implicated differences in syntactic positions across the different IS roles, we expect our results to carry over to such languages too. Whether this expectation is borne out remains to be verified by future work.

[23]Many approaches distinguish between aboutness topics and contrastive topics on the one hand (for references, see Büring 2016), and between alternatives based focus and newness based focus, on the other (for references, see É. Kiss 1998).

References

Bates, D., et al. 2014. *lme4: Linear Mixed-Effects Models Using Eigen and S4*. http://CRAN.R-project.org/package=lme4.

Beghelli, F., and T. Stowell. 1997. Distributivity and Negation: The Syntax of *Each* and *Every*. In *Ways of Scope Taking*, 71–107. Dordrecht: Kluwer.

Bosnic, A. 2016. Distributivity and Agreement Mismatches in Serbian. In *Studies in Language and Mind*, ed. S. Halupka-Rešetar and S. Martínez-Ferreiro, 77–105. *Selected Papers from the 3rd Workshop in Psycholinguistics, Neurolinguistics and Clinical Linguistic Research*. Novi Sad: University of Novi Sad.

Boylan, C., et al. 2011. A Matter of Ambiguity? Using Eye Movements to Examine Collective vs. Distributive Interpretations of Plural Sets. In *Abstract, 4th Biennial Conference of Experimental Pragmatics*, Barcelona.

Braun, B., and L. Tagliapietra. 2010. The Role of Contrastive Intonation Contours in the Retrieval of Contextual Alternatives. *Language and Cognitive Processes* 25: 1024–1043.

Brooks, P.J., and M.D.S. Braine. 1996. What Do Children Know About the Universal Quantifiers All and Each? *Cognition* 60 (3): 235–268.

Büring, D. 2016. (Contrastive) Topic. In *Oxford Handbook of Information Structure*, ed. C. Féry and S. Ishihara, 64–85. Oxford: Oxford University Press.

Chafe, W.L. 1976. Givenness, Contrastiveness, Definitness, Subjects, Topics and Point of View. In *Subject and Topic*, ed. C.N. Li, 25–55. New York: Academic Press.

Chierchia, G., D. Fox, and B. Spector. 2013. Scalar Implicature as a Grammatical Phenomenon. In *Semantics: An International Handbook of Natural Language Meaning*, ed. P. Portner, P. Maienborn, and P. von Heusinger, 2297–2332. Berlin: Mouton De Gruyter.

Clifton, C. Jr., and L. Frazier. 2012. Interpreting Conjoined Noun Phrases and Conjoined Clauses: Collective vs. Distributive Preferences. *Quarterly Journal of Experimental Psychology* 65 (9): 1760–1776.

Constant, N. 2012. Witnessable Quantifiers License Type-e Meaning: Evidence from Contrastive Topic. In *Proceedings of Semantics and Linguistic Theory*, vol. 22, 286–306.

Constant, N. 2014. *Contrastive Topic: Meanings and Realizations*. Doctoral Dissertation, University of Massachusets, Amherst.

Cowles, H.W. 2003. *Processing Information Structure: Evidence from Comprehension and Production*. Doctoral Dissertation, University of California, San Diego.

Cresti, D. 1995. Indefinite Topics. *Doctoral Dissertation*, Massachusetts Institute of Technology, Cambridge.

Dalrymple, M., and I. Nikolaeva. 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.

de Jong, F., and H. Verkuyl. 1985. Generalized Quantifiers: the Properness of Their Strength. In *Generalized Quantifiers in Natural Language*, ed. J. van Benthem, and A. ter Meulen, 21–43. Dorecht: Fortis.

Diesing, M. 1992. *Indefinites*. Cambridge: MIT Press.

Dowty, D.R. 1987. Collective Predicates, Distributive Predicates, and All. In *Proceedings of the Third Eastern States Conference in Linguistics*, 97–115, Columbus: Ohio State University.

Drummond, A. 2010. *Ibex (Internet Based Experiments)*. http://spellout.net/ibexfarm/.

É. Kiss, Katalin. 1998. Identificational Focus versus Information Focus. *Language* 74 (2): 245–273.

É. Kiss, Katalin. 2002. *The syntax of Hungarian*. Cambridge Syntax Guides. Cambridge: Cambridge University Press.

É. Kiss, Katalin. 2005. Event Types and Discourse Linking in Hungarian. *Linguistics* 43 (1): 131–154.

Ebert, C., and C. Endriss. 2004. Topic Interpretation and Wide Scope Indefinites. In *Proceedings of NELS 34*. Amherst: GLSA.

Endriss, C. 2006. *Quantificational Topics. A Scopal Treatment of Exceptional Wide Scope Phenomena*. Potsdam: University of Potsdam.

Endriss, C. 2009. Quantificational topics—A Scopal Treatment of Exceptional Wide Scope Phenomena. In *Studies in Linguistics and Philosophy*, ed. C. Ebert. Berlin: Springer.

Erteschik-Shir, N. 1997. *The Dynamics of Focus Structure*. Cambridge: Cambridge University Press.

Farkas, D.F. 2002. Specificity Distinctions. *Journal of Semantics* 19 (3): 213–243.

Fox, J., et al. 2015. *Car: Companion to Applied Regression*. http://CRAN.R-project.org/package=car.

Fox, D. 2007. Free Choice and the Theory of Scalar Implicatures. In *Presupposition and Implicature in Compositional Semantics*, ed. U. Sauerland, and P. Stateva, 71–120. Basingstoke: Palgrave.

Frascarelli, M., and R. Hinterhölzl. 2007. Types of Topics in German and Italian. In *On Informational Structure, Meaning and Form*, ed. S. Winkler, and K. Schwabe, 87–116. Amsterdam: John Benjamins.

Fraundorf, S.H., A.S. Benjamin, and D.G. Watson. 2013. What Happened (and What didn't): Discourse Constraints on Encoding of Plausible Alternatives. *Journal of Memory and Language* 69: 196–227.

Frazier, L., and C. Clifton. 2001. Interpreting Plural Subjects: Collective-Distributive Preferences. Manuscript. University of Massachusetts.

Frazier, L., H. Lasnik, and R. May. 1999. Taking on Semantic Commitments, II: Collective Versus Distributive Readings. *Cognition* 70 (1): 87–104.

Geist, L. 2011. *Indefiniteness and Specificity: Lexical Marking and Information-Structural Conditions*. Habilitationsschrift. Stuttgart: University of Stuttgart.

Geurts, B. 2007. Existential Import. In *Existence: Semantics and Syntax*, ed. I. Comorovski, and K. von Heusinger, 253–271. Berlin: Springer.

Gil, D. 1995. Universal Quantifiers and Distributivity. In *Quantification in Natural Languages*, ed. E. Bach, et al., 321–362. Dordrecht: Springer.

Groenendijk, J., and M. Stokhof. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Joint Doctoral Dissertation, University of Amsterdam, Amsterdam.

Gyuris, B. 2009. Quantificational Contrastive Topics with verum/Falsum Focus. *Lingua* 119: 625–649.

Hackl, M. 2000. *Comparative Quantifiers*. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge.

Harris, J. A., C. Jr. Clifton, and L. Frazier. 2012. *Processing and Domain Selection: Quantificational Variability Effects*. Manuscript.

Heim, I. 1982. The Semantics of Definite and Indefinite Noun Phrases. Doctoral Dissertation, University of Massachusetts, Amherst.

Hoeksema, J. 1983. Plurality and Conjunction. In *Studies in Modeltheoretic Semantics*, ed. A. ter Meulen, 63–83. Dordrecht: Foris.

Horvath, J. 2007. Separating "Focus Movement" from Focus. In *Phrasal and Clausal Architecture*, ed. S. Karimi, V. Samiian, and W. Wilkins, 108–145. Amsterdam: John Benjamins.

Iglewicz, B., and D.C. Hoaglin. 1993. ASQC Basic References in Quality Control. In *How to Detect and Handle Outliers*. ASQC Quality Press.

Jaeger, G. 1995. *Topics in Dynamic Semantics*. Doctoral Dissertation, Humboldt University, Berlin.

Kamp, H., and U. Reyle. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Kaup, B., S. Kelter, and C. Habel. 2002. Representing Referents of Plural Expressions and Resolving Plural Anaphors. *Language and Cognitive Processes* 17 (4): 405–450.

Kim, C.S., et al. 2015. Context-Driven Expectations About Focus Alternatives. *Cognition* 139: 28–49.

Knežević, N. 2015. *Numerals and Distributivity in Serbian: At the Syntax-Semantics-Acquisition Interface*. Doctoral Dissertation, Université de Nantes, Nantes.

Kratzer, A. 2007. On the plurality of verbs. In *Event structures in Linguistic Form and Interpretation*, ed. J. Dölling, T. Heyde-Zybatow, and M. Schäfer, 269–300. Language, Context, Cognition. New York: Mouton De Gruyter.

Krifka, M. 2006. Association with Focus Phrases. In *The Architecture of Focus*, ed. V. Molnar, and S. Winkler, 105–136. New York: Mouton de Gruyter.

Krifka, M. 2008. Basic Notions of Information Structure. *Acta Linguistica Hungarica* 55: 243–276.
Ladusaw, B. 1994. Thetic and Categorical, Stage and Individual, Weak and Strong. *Semantics and Linguistic Theory* 4: 220–229.
Landman, F. 1989. Groups, I. *Linguistics and Philosophy* 12 (5): 559–605.
Landman, F. 2000. *Events and Plurality*. Dordrecht: Kluwer.
Lappin, S., and T. Reinhart. 1998. Presuppositional Effects of Strong Quantifiers. *Linguistics* 26: 1021–1037.
Lasersohn, P. 1993. Lexical Distributivity and Implicit Arguments. In *Proceedings from Semantics and Linguistic Theory III*, ed. U. Lahiri and A. Wyner, 145–151. DMLL Publications.
Lasersohn, P. 1995. *Plurality, Conjunction and Events*. Berlin: Springer.
Lasersohn, P. 1998. Generalized Distributivity Operators. *Linguistics and Philosophy* 21 (1): 83–93.
Lefcheck, J. S. 2015. PiecewiseSEM: Piecewise Structural Equation Modelling in R for Ecology, Evolution, and Systematics. *Methods in Ecology and Evolution*, vol. 7. https://doi.org/10.1111/2041-210X.12512.
Link, G. 1983. The Logical Analysis of Plurals and Mass Terms: A Lattice Theoretical Approach. In *Meaning, Use, and the Iterpretation of Language*, ed. R. Bäuerle, C. Schwarze, and A. von Stechow, 303–323. Berlin: Walter de Gruyter.
Mayr, C., and B. Spector. 2012. Generalized Scope Economy—Not Too Strong. *Proceedings of Sinn und Bedeutung* 14: 305–321.
McCawley, J. 1972. Syntactic and Logical Arguments for Semnatic Structure. In *Three Dimensions of Linguistic Theory*, ed. O. Fujimura, et al. Tokyo: TEC.
Milsark, G.L. 1979. *Existential Sentences in English*. New York: Garland.
Moxey, L.M., et al. 2004. Constraints on the Formation of Plural Reference Objects: The Influence of Role, Conjunction, and Type of Description. *Journal of Memory and Language* 51: 346–364.
Moxey, L.M., et al. 2011. When do We Use “they” to Refer to Two Individuals? Scenario Mapping as a Basis for Equivalence. *Language and Cognitive Processes* 26 (1): 79–120.
Musolino, J. 2009. The Logical Syntax of Number Words: Theory, Acquisition and Processing. *Cognition* 111 (1): 24–45.
Nakagawa, S., and H. Schielzeth. 2013. A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models. *Methods in Ecology and Evolution* 4 (2): 133–142.
Nakanishi, K., and M. Romero. 2004. Two Constructions with most and their Semantic Properties. NELS 34.
Pagliarini, E., G. Fiorin, and J. Dotlacil. 2012. The Acquisition of Distributivity in Pluralities. In *36th Annual Boston University Conference on Language Development*, 387–399. Somerville: Cascadilla Press.
Partee, B. 1989. Many quantifiers. In *Proceedings of the Fifth Eastern State Conference on Linguistics*, ed. J. Powers and K. De Jong, 383–402. Columbus: Ohio State University Press.
Partee, B. 1995. Quantification and the Dynamics of Context-Dependence in Natural Language. In *Logica 1994: Proceedings of the 8th International Symposium*, ed. T. Childers and O. Majer, 141–162, Prague: Czech Academy of Sciences.
Partee, B. 1987. Noun Phrase Interpretation and Type-Shifting Principles. In *Studies in Discourse Representation Theory and Theory of Generalized Quantifiers*, ed. J. Groenendijk, D. de Jong, and M. Stokhof, 115–143. Dordrecht: Foris.
Portner, P. 2002. Topicality and (Non-)Specificity in Mandarin. *Journal of Semantics* 19 (3): 275–287.
Portner, P., and K. Yabushita. 1998. The Semantics and Pragmatics of Topic Phrases. *Linguistics and Philosophy* 21: 117–157.
Portner, P., and K. Yabushita. 2001. Specific Indefinites and the Information Structure Theory of Topics. *Journal of Semantics* 18 (3): 271–297.
R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna: Austria.
Reinhart, T. 1981. Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27: 53–94.

Repp, S. 2009. Topics and Corrections. In *Proceedings of Sinn und Bedeutung 13*, ed. A. Riester and T. Solstad, 399–414. Stuttgart: University of Stuttgart.
Rizzi, L. 1997. The Fine Structure of the Left Periphery. In *Elements of Grammar*, ed. L. Haegeman, 281–337. Dordrecht: Springer Netherlands.
Roberts, C. 1987. *Modal Subordination, Anaphora, and Distributivity*. Doctoral Dissertation, University of Massachusetts, Amherst
Roberts, C. 1990. Modal Subordination, Anaphora, and Distributivity. In *Outstanding Dissertations in Linguistics*, ed. J. Hankamer. New York: Garland.
Roberts, C. 1998. Focus, the Flow of Information, and Universal Grammar. In *The Limits of Syntax*, ed. P. Culicover, and L. McNally, 109–160. Cambridge: Academic Press.
Rooth, M. 1985. *Association with Focus*. Doctoral Dissertation, University of Massachusetts, Amherst.
Rooth, M. 1992. A Theory of Focus Interpretation. *Natural Language Semantics* 1 (1): 75–116.
Scha, R.J.H. 1981. Distributive, Collective and Cumulative Quantification. In *Truth, Interpretation and Information: Selected Papers from the Third Amsterdam Colloquium*, ed. J. Groenendijk, Theo M. V., and M. Janssen, 131–158. Published in 1984. New York: Walter de Gruyter.
Schwarzschild, R. 1994. *The Contrastiveness and Associated Foci*. Manuscript, New Brunswick.
Schwarzschild, R. 1993. Plurals, Presuppositions and the Sources of Distributivity. *Natural Language Semantics* 2 (3): 201–248.
Schwarzschild, R. 1996. *Pluralities*. Berlin: Springer.
Solt, S. 2009. *The Semantics of Adjectives of Quantity*. Doctoral Dissertation, The City University of New York, New York
Strawson, P. 1950. On referring. *Mind* 59: 320–344.
Strawson, P. 1952. *Introduction to Logical Theory*. London: Methuen.
Strawson, P.F. 1964. Identifying Reference and Truth-Values. *Theoria* 30 (2): 96–118.
Syrett, K., A. Brasoveanu, and A. Oho. An Experimental Investigation of the Scope of Object Comparative Quantifier Phrases. *Journal of Semantics*.
Szabolcsi, A. 1997. Scope Positions in Hungarian. In *Ways of Scope Taking*, ed. A. Szabolcsi, 118–122. Dordrecht: Kluwer.
Takahashi, S. 2006. More than Two Quantifiers. *Natural Language Semantics* 14 (1): 57–101.
Trauzettel-Klosinski, S., and K. Dietz. 2012. Standardized Assessment of Reading performance: The New International Reading Speed Texts IReSTstandardized Assessment of Reading Performance. *Investigative Ophthalmology and Visual Science* 53 (9): 5452–5461.
Ussery, C. 2008. Processing Plural DPs: Collective, Cumulative, and Distributive Interpretations. Manuscript.
von der Gabelentz, G. 1896. Ideen zu einer vergleichenden Syntax. Wort- und Satzstellung. *Zeitschrift für Volkerpsychologie und Sprachwissenschaft* 6: 374–384.
Wagner, M. 2006. Association by movement. Evidence from NPI-Licensing. *Natural Language Semantics* 14: 297–324.
Wagner, M. 2012. Focus and givenness: A Unified Approach. In *Contrasts and Positions in Information Structure*, ed. I. Kucerová, and A. Neeleman, 102–148. Cambridge: Cambridge University Press.
Winter, Y. 1997. Choice Functions and the Scopal Semantics of Indefinites. *Linguistics and Philosophy* 20 (4): 399–467.
Winter, Y. 2000. Distributivity and Dependency. *Natural Language Semantics* 8 (1): 27–69.
Winter, Y. 2001. *Flexibility Principles in Boolean Semantics: The Interpretation of Coordination, Plurality, and Scope in Natural Language*. Cambridge: MIT Press.
Zucchi, A. 1995. The Ingredients of Definiteness and the Definiteness Effect. *Natural Language Semantics* 3: 33–79.

Quantifier Spreading in School-Age Children: An Eye-Tracking Study

Irina A. Sekerina, Patricia J. Brooks, Luca Campanelli and Anna M. Schwartz

Abstract Children make quantifier-spreading errors in contexts involving sets in partial one-to-one correspondence; e.g., *Every bunny is in a box* is rejected as a description of three bunnies, each in a box, along with two extra boxes. To determine whether a signature pattern of visual attention is associated with the classic *q*-spreading error as it occurs in real time, eye-movements were recorded while children ($N = 41$; mean 8 y;9 m, range 5;8–12;1) performed a sentence-picture verification task, with *every* modifying either the figure or ground of locative scenes (*every bunny* vs. *every box*). On trials designed to elicit the classic error, children performed at chance (53.3% correct). Errors involved greater numbers of fixations to the extra objects/containers, time-locked to regions following the quantified noun phrase. Correct responses were associated with longer reaction times, indicating additional processing required for quantifier restriction; accuracy was uncorrelated with verbal or nonverbal intelligence and only weakly associated with age. The findings underscore the susceptibility of school-age children to make errors given a default expectation for distributive quantifiers like *every* to refer to sets in one-to-one correspondence and their inattention to sentence structure.

Keywords Quantifier-spreading (*q*-spreading) · Eye movements
Visual attention · Children · Universal quantifier *every* · Visual world paradigm

---

I. A. Sekerina (✉) · P. J. Brooks
Department of Psychology, College of Staten Island, City University of New York, New York, NY 10314, USA
e-mail: Irina.Sekerina@csi.cuny.edu

P. J. Brooks
e-mail: patricia.brooks@csi.cuny.edu

I. A. Sekerina · P. J. Brooks · L. Campanelli · A. M. Schwartz
The Graduate Center, City University of New York, New York, NY 10016, USA
e-mail: campanelli.l@gmail.com

A. M. Schwartz
e-mail: anna.m.e.schwartz@gmail.com

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_8

1 Introduction

Children's acquisition of quantifiers has been a topic of great interest due to the complexity of the mappings between semantic contrasts and lexical-syntactic structures, and children's apparent difficulties in learning these mappings. Inhelder and Piaget (1964) seminal work on class inclusion errors led to a proliferation of research on children's difficulties in restricting the domain of a universal quantifier to the appropriate noun phrase (e.g., Brooks and Braine 1996; Bucci 1978; Donaldson and Lloyd 1974; Freeman 1985; Philip 1995), with the term *quantifier-spreading* (*q*-spreading) coined as a description of these errors (Roeper and de Villiers 1993). The present study aims to shed light on the source of *q*-spreading errors in school-age children by examining patterns of visual attention in sentence processing in real time.

The terminology used to describe children's *q*-spreading errors is unfortunately very convoluted. Some authors have emphasized a distinction between *bunny*-spreading and classic *q*-spreading errors (Roeper et al. 2004), whereas others have emphasized under-exhaustive versus over-exhaustive search errors (Freeman 1985). Note, however, that the over-exhaustive error is identical to the classic *q*-spreading error. Our intention is not to prioritize one set of terms over the other, but to pay homage to both psychological and linguistic traditions in describing these errors. Figures 1 and 2 provide depictions of sets of objects in containers and corresponding sentences to illustrate each error type. *Bunny*-spreading errors occur when children extend the scope of a universal quantifier in sentences like *Every bunny is in a box* or *Every box has a bunny in it* to include extraneous objects that are neither bunnies nor boxes (e.g., cats or buckets, as shown in Fig. 1a, b).

Under-exhaustive and over-exhaustive (i.e., classic) errors occur in contexts involving sets in partial one-to-one correspondence—for example, three bunnies each in a box, along with one or more extra bunnies or boxes (Fig. 2a, b). Note that the specification of error types depends on the pairing of sentences with pictures. For the sentence *Every box has a bunny in it*, rejecting a picture with extra bunnies (Fig. 2a) is an over-exhaustive, classic error (i.e., the scope of the quantifier is extended beyond the boxes containing bunnies), and accepting the sentence as a

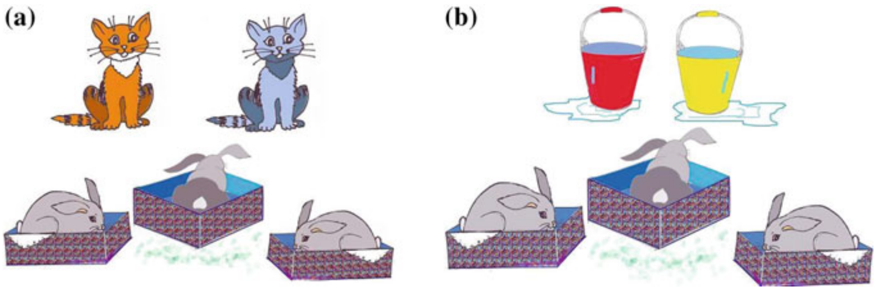


Fig. 1 Sample pictures with extra objects (**a**) or containers (**b**) designed to elicit *bunny*-spreading errors

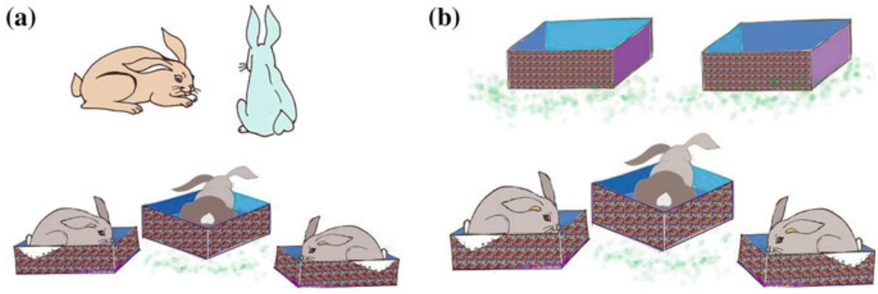


Fig. 2 Sample pictures of sets in partial one-to-one correspondence with extra objects (**a**) or containers (**b**) designed to elicit over-exhaustive (classic) and under-exhaustive errors

description of a picture with extra boxes (Fig. 2b) is an under-exhaustive error (i.e., some boxes fail to be included within the scope of the quantifier). For the sentence *Every bunny is in a box*, acceptance of Fig. 2a would constitute an under-exhaustive error, and rejection of Fig. 2b would constitute an over-exhaustive, classic error.

Whereas the under-exhaustive error has been documented only in young children and appears to be relatively rare (Freeman 1985; Roeper et al. 2004), the classic over-exhaustive, classic error has been reported in studies of school-age children (Bucci 1978), bilingual adults (Berent et al. 2009; DelliCarpini 2003; Sekerina and Sauermann 2015) and even monolingual adults (Brooks and Sekerina 2006; Minai et al. 2012).

Both *bunny*-spreading and under-exhaustive errors decline rapidly in early childhood and little attention has been paid to these errors in theoretical accounts; hence for brevity, we will use the term *classic q-spreading* to refer to the over-exhaustive error throughout the remainder of this chapter. Explanations for the classic error fall into two broad categories, linguistic and cognitive, but full treatment of the various explanations is beyond the scope of this chapter (see Rakhlin 2007). Within the Generative Grammar framework of language acquisition, researchers have attributed the error to children's immature, non-adult-like linguistic representations, which may lead to quantification over events rather than individuals (Philip 1995) or non-canonical mappings from syntax to semantics (Geurts 2003). Classic *q*-spreading has been attributed to weak quantification (Drozd 2001) and recovery from errors to syntactic restructuring (Roeper et al. 2004). In contrast, cognitive approaches attribute classic *q*-spreading to extra-linguistic factors that impact sentence processing, such as the pragmatics of the testing situation (Crain et al. 1996), weak cognitive control (Minai et al. 2012), or task demands (O'Grady et al. 2010).

In our prior work, we suggested that classic *q*-spreading might arise from shallow processing or lack of attention to sentence structure (Brooks and Sekerina 2005, 2006). Shallow sentence processing is thought to generate 'good enough' (underspecified) representations of sentence structures that under most circumstances are sufficient for comprehension (Clahsen and Felser 2006; Felser and Clahsen 2009; Ferreira et al. 2002; Sanford and Sturt 2002). When relying on shallow processing, individuals may associate sentence structures with canonical semantic representations by default, such as associating the first noun in the sentence with the role of actor, which would lead to an error in processing a passive sentence such as *The dog was chased by the cat* (Ferreira 2003). Brooks and Sekerina (2005, 2006) interpreted the occurrence of classic *q*-spreading in college students as evidence of shallow processing (as opposed to having a faulty or immature grammar) resulting in inaccurate mappings between syntactic and semantic representations.

In earlier work, Brooks and Braine (1996) suggested that children might use canonical collective and distributive representations as defaults when interpreting sentences with universal quantifiers: The canonical collective representation involves a group of individuals (or objects) performing an action together (*all of the men lifted a box,* with the interpretation that the men lifted a box together) or assembled in the same location (e.g., *all of the flowers are in a vase*, with the interpretation that the flowers are in the same vase). The canonical distributive interpretation assumes one-to-one correspondence, with individuals performing the same action but on their own (e.g., *each man lifted a box*, with the interpretation that each man lifted a different box) or in their own corresponding locations (e.g., *each flower is in a vase*, with the interpretation that there are as many vases as flowers, with one flower in each). For sentences with a distributive universal quantifier like *each* or e*very*, one-to-one correspondence is thought to be the default semantic alignment of the two sets of objects. When relying on shallow processing, the child may fail to consider the syntactic structure in determining which noun phrase is modified by the quantifier; consequently he or she will reject a distributive scene showing sets in partial one-to-one correspondence, resulting in a classic *q*-spreading error.

In documenting classic errors, researchers have tended to rely on two related methodologies: the picture-choice task and the sentence-picture verification task. The picture-choice task pits two or more pictures (e.g., one with extra bunnies, one with extra boxes) against each other, and asks participants to find the picture where, e.g., *Every box has a bunny in it.* Participants are required to choose one of the pictures, where typically only one picture is logically correct; hence the task is useful in determining error rates in using sentence structure to restrict the quantifier to the appropriate noun phrase. Brooks and Sekerina (2005, 2006, Experiment 3) tested adults using the picture-choice task with locative scenes (objects in containers) as illustrated in Fig. 2, with the position of the quantifier varying from trial to trial (e.g., modifying *bunny* or *box*). However, the visual display was more complex, with four pictures instead of two. One picture depicted the three pairs of objects in containers plus two extra objects, the second one had the three pairs plus two extra containers, and the remaining two were foils with two different objects or containers. The authors found that college students were only 75% correct (although they had no difficulties avoiding the foils that depicted extraneous objects). Street and Dąbrowska (2010) used the picture-choice task with two options—one with extra objects, one with extra containers, and tested adults who had low educational attainment. They attributed the group's poor performance on the task (78% correct for the sentence *Every X is in a Y* and 43% correct for *Every Y has an X in it*) to participants' lack of experience with the sentence structures.

The sentence-picture verification task (Clark and Chase 1972) presents pictures one at a time along with an accompanying sentence, and asks participants to decide whether or not the sentence is an accurate description of the picture. This task is advantageous for distinguishing different error types, as features of pictures and sentences may be manipulated independently. Studies using this task suggest that the number of extra objects influences error rates: In the most extreme case, with one extra object (e.g., three turtles each holding an umbrella, with one extra umbrella), Japanese adults achieved accuracy of only 59% correct when verifying sentences like *Dono-kame-mo kasa-o sashi-teruyo* 'Every turtle is holding an umbrella' (Minai et al. 2012). Other studies have reported high rates of classic errors amongst adult L2 learners of English with low proficiency (Berent et al. 2009; DelliCarpini 2003).

2 Current Study

In the current study, we examined eye-movements during sentence processing in school-age children to examine patterns of visual attention associated with classic *q*-spreading in the sentence-picture verification task. To date, only one prior study has examined eye-movements associated with susceptibility to classic errors in children. In a study with Japanese preschool-age children, Minai et al. (2012) varied the number of extra objects (e.g., umbrellas) in the sentence-picture verification task across blocks of trials; they reported very high rates of classic errors when there was just one extra umbrella as opposed to three, but only when the more difficult one-object condition was presented first. Perhaps due to the small number of children who responded correctly in the difficult condition—i.e., 25 of 29 children were categorized as *SR* (symmetric response) for consistently making the classic error—the researchers failed to find any evidence that patterns of eye-movements, recorded while the sentence was unfolding, distinguished children as a function of their accuracy in sentence-picture verification. However, they did report that children spent more time looking at the extra objects (one or three umbrellas) *prior* to the onset of the sentence, when compared to adults. Minai and colleagues interpreted their findings as suggesting that difficulties in the control of attention contribute to the occurrence of classic *q*-spreading in children.

Extending the study of classic errors to adult bilingual heritage speakers of Russian, Sekerina and Sauermann (2015) identified an attentional pattern of eye movements that distinguished incorrect (20%) from correct (80%) responses in heritage Russian-English bilingual adults performing the sentence-picture verification task in their weak language (i.e., Russian). This attentional signature was time-locked to the occurrence of the verb in the sentences (e.g., *Kazhdyj alligator lezhit v vanne* [Every alligator lies in bathtub]); thus, immediately after processing the quantified noun phrase (e.g., *kazhdyj alligator*), adults who were susceptible to the error showed increased looks to the extra objects in the picture.

The current study attempts to extend the attentional signature pattern associated with classic *q*-spreading to monolingual English-speaking children of ages

5–12 years. Children in this age range are still susceptible to classic errors and yet vary in performance (Brooks and Braine 1996; Brooks and Sekerina 2005, 2006, Experiments 1 and 2). Using the sentence-picture verification task, we examined visual attention as each sentence unfolded in real time. Our goal was to determine whether *q*-spreading would be associated with increased looks to the extra objects/containers and whether these looks would be time-locked to the region of interest immediately following the quantified noun phrase. We supplemented our analyses of eye movements with reaction time data to determine whether *q*-spreading errors were associated with greater or lesser processing time relative to correct responses. Finally, in addition to testing children across a broad age range, we administered assessments of non-verbal and verbal intelligence to determine whether either of these abilities would be associated with error rates after controlling for age.

Across trials we varied the structure of the sentences, with the quantifier *every* modifying either the object or container in locative events (e.g., *bunny* vs. *box* as illustrated in Figs. 1 and 2). In addition to trials designed to elicit classic errors, we also included two other types of trials: One type had the potential to elicit *bunny*-spreading errors and the other had the potential to elicit under-exhaustive errors. The latter trial type was treated as a control condition (e.g., *Every bunny is in a box* presented with a picture with extra bunnies), as we expected children to correctly detect the violation of one-to-one correspondence between bunnies and boxes with near perfect accuracy, and reject these sentences as descriptions of the pictures.

3 Method

3.1 Participants

Participants were 41 monolingual English-speaking children (23 girls and 18 boys, *M* = 8 y;9 m, *SD* = 1;11, age range = 5;8–12;1). Thirty children were recruited and tested in an afterschool program at a private Catholic school in Staten Island, NY; an additional 11 children were recruited from a child subject pool and tested in a laboratory at the College of Staten Island, CUNY. Informed consent was obtained from parents and assent from children. The children were from middle to upper middle class families; all had normal or corrected-to-normal vision. Children's receptive knowledge of vocabulary was estimated using the Peabody Picture Vocabulary Test, 4th Edition (PPVT-4, Form B; Dunn and Dunn 2007), *M* raw score = 149.6, *SD* = 23.9; *M* standardized score = 109.4, *SD* = 11.5; and their nonverbal intelligence was estimated with the Test of Nonverbal Intelligence, 3rd Edition (TONI-3; Brown et al. 1997), *M* raw score = 20.0, *SD* = 8.7; *M* standardized score = 108.9, *SD* = 15.3. Note that standardized scores on the TONI could not be computed for three children of ages 5;8–5;10 due to lack of age-referenced norms for children below 6;0. Due to lack of time, PPVT-4 and TONI-3 tests were not administered to four children. Due to equipment malfunctioning, eye-tracking data were lost for one child. Children received small gifts (e.g., stuffed animals) as rewards for their participation.

The 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. In accordance with the Declaration of Helsinki, informed consent was obtained from parents and assent from children.

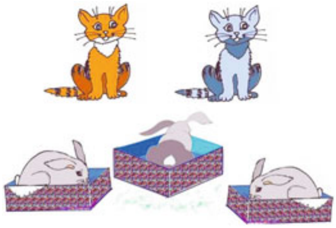
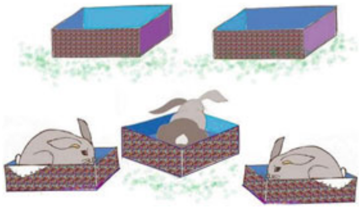
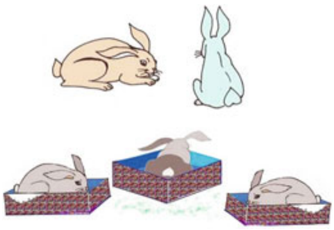
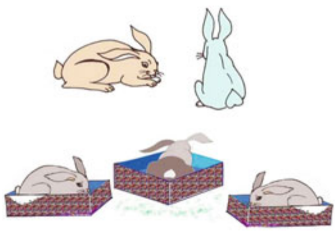
3.2 Design and Materials

Each trial of the sentence-picture verification task presented a picture paired with a spoken sentence that either matched or mismatched the picture (correct response = 'yes' or 'no'). Each sentence was recorded individually by a female native English-speaker using mono-mode sampling at 22,050 Hz. Sentences were spoken at a normal adult rate. The experiment presented 4 practice trials, 24 quantifier trials (i.e., sentences with the quantifier *every* modifying the figure or ground of a locative scene) interspersed with 8 active-voice and 8 passive-voice fillers (i.e., reversible sentences with two animate nouns), and 16 additional fillers, with the latter quantifier trials and fillers presented in a pseudo-randomized order.

Table 1 presents examples of trials for each condition using the set of bunnies in boxes to illustrate the quantifier trials. Note, however, that each trial of the experiment depicted a different set of objects in containers, with half of the sets depicting animate objects (e.g., alligators in bathtubs), and the other half inanimate objects (e.g., eggs in frying pans). Quantifier trials presented pictures of objects or animals in containers, with three object/container pairs in the foreground and two extraneous objects or containers (for *bunny*-spreading trials) or two extra objects or containers (for classic and control trials) in the background. The active/passive fillers depicted transitive actions with two animate nouns. The additional fillers used pictures that were visually similar to those used for quantifier trials except that they depicted five objects/animals in containers (e.g., five flowers in vases; five dogs on chairs), with the sentences referring to the number of objects, their color, or including a comparison (e.g., *There are more blue chairs than green ones*). These additional fillers were used to balance the number of trials where 'yes' versus 'no' was the correct response, while reducing the proportion of quantifier trials overall.

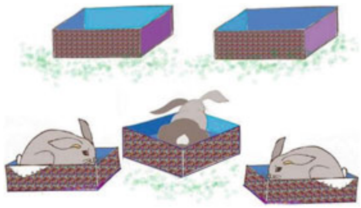
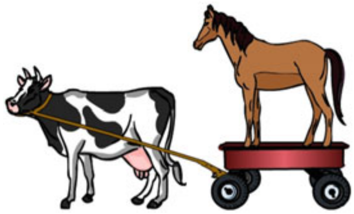
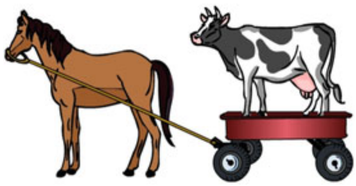
Six lists were created, using a Latin square to counterbalance sets across conditions, with children at each age randomly assigned to each list. The lists presented 8 trials in each of the three quantifier conditions: *bunny*-spreading, control (under-exhaustive), and classic (over-exhaustive), using object sentences (*Every X is in/on a Y*) in half of the trials and container sentences (*Every Y has an X in/on it*) in the other half. Note that the correct response for *bunny*-spreading and classic trials was always 'yes' whereas the correct response for control trials was always 'no.' Each list also included 8 active and 8 passive fillers, with half of the trials per condition

Table 1 Examples of sentences and pictures for each task condition

Condition	Sentence type	Example sentence	Example picture
Bunny-spreading	Object	Every bunny is in a box (correct = 'yes')	
Bunny-spreading	Container	Every box has a bunny in it (correct = 'yes')	
Classic	Object	Every bunny is in a box (correct = 'yes')	
Classic	Container	Every box has a bunny in it (correct = 'yes')	
Control	Object	Every bunny is in a box (correct = 'no')	

(continued)

Table 1 (continued)

Condition	Sentence type	Example sentence	Example picture
Control	Container	Every box has a bunny in it (correct = ‘no’)	
Filler	Active	The cow is pulling the horse (correct = ‘yes’) OR The horse is pulling the cow (correct = ‘no’)	
Filler	Passive	The cow is being pulled by the horse (correct = ‘yes’) OR The horse is being pulled by the cow (correct = ‘no’)	
Filler	Additional	There are more blue chairs than green ones (correct = ‘no’)	

associated with a correct response of ‘yes’ and the other half with a correct response of ‘no.’

3.3 Procedure

The sentence-picture verification task was programmed into a script run by DMDX, a free Windows-based software for language processing experiments (Forster and Forster 2003). Stimuli 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 simultaneously with the onset of the spoken sentence. Children were instructed to answer ‘yes’ if they thought that the sentence correctly described the picture, and ‘no’ otherwise. Accuracy was recorded using a gamepad attached to the 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.

Children’s eye movements were recorded using the ISCAN remote portable eye-tracking system (ETL-500). 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 computer and were recorded simultaneously with eye movements. Each child underwent a short calibration procedure prior to the experiment.

3.4 Data Treatment and Analyses

Eye movements were extracted from videotape using a SONY DSR-30 video tape-recorder with frame-by-frame control and synchronized audio and video. Nine trials (0.9%) were not recorded due to equipment malfunctioning and constituted missing data for the eye-movement analyses. For each trial, four categories were coded: looks to the three pairs of entities in the front of the picture, looks to the two ‘distractors’ (e.g., cats or buckets in the *bunny*-spreading condition), looks to the two ‘extra’ objects/containers (bunnies or boxes in the control and classic conditions), looks elsewhere in the picture, and track loss. Track loss and looks elsewhere constituted a small proportion of total looks (8.6%) and were removed from the eye-movement analyses; thus, fixations to the three object/container pairs in the foreground of each picture were in complimentary distribution with fixations to the distractors/extras in the back. We hypothesized that allocation of visual attention to irrelevant distractors/extras would co-occur with *q*-spreading errors; hence statistical analyses focus on proportions of looks to the distractors/extras as a function of response accuracy. Using fine-grain analyses, proportions of looks to the distractors/extras were analyzed in three separate time windows or regions of interests (ROIs) defined relative to the onset of each phrase (Table 2). Note that ROI 3 terminated when the child responded or one second after the offset of the stimulus sentence, whichever was earlier.

We conducted three sets of analyses with response accuracy, eye movements to distractors/extras, and reaction times as dependent variables. Mixed-effects logistic regression was used to analyze response accuracy and eye-movement data. The logistic part allows for modeling the nonlinear nature of the dependent variable, which is bounded by 0 and 1; this approach has been shown to be superior to an analysis of variance approach on transformed data (Jaeger 2008).

Response times were analyzed using linear mixed-effects regression with maximum likelihood as the estimation method. Although the distribution was slightly

Table 2 Regions of interest (ROIs) for each sentence type

Sentence type	ROI 1	ROI 2	ROI 3
Object	*Quantified NP*	*Verb–PP–NP-Loc*	*Silence*
	Every bunny	is in a box	
Container	*Quantified NP*	*Verb–NP–PP*	
	Every box	has a bunny in it	

skewed, response time data were kept in their original scale as analyses on transformed data produced the same pattern of results.

All models included crossed random intercepts for subjects and items (Baayen et al. 2008). Fixed effects and random slopes were examined during the model building process and retained only when they improved the model fit. We used a model comparison framework to contrast alternative models that were progressively more complex. This approach is preferable to significance tests of individual parameters in arriving at correct statistical inferences (Bliese and Ployhart 2002). The likelihood ratio test was used to compare the fit of competing models. Only age was retained as a covariate in all models independently of its statistical significance.

Outliers were trimmed in two steps: First, for each experimental condition, participants with average performance more than 3 standard deviations above or below the grand mean were excluded. Second, for each model, we examined its residuals and re-fitted it after removing observations with standardized residuals greater than 3 or smaller than −3. In none of the analyses were more than 3.5% of the data excluded. Data were analyzed with R version 3.1.0 (R Core Team 2014) using the *lmer* and *glmer* functions from the LME4 package, version 1.1–8 (Bates et al. 2015).

4 Results

4.1 Accuracy

To examine associations between accuracy in sentence comprehension and individual differences in verbal and nonverbal abilities, we computed partial correlations, between accuracy and PPVT and TONI raw scores for each sentence type and condition, controlling for age in months. Note that use of raw scores was necessitated by lack of age norms for 5-year-olds on the TONI. The partial correlation between TONI and PPVT raw scores approached statistical significance and showed a weak positive association ($r = 0.30$, $p = 0.076$).

As shown in Table 3, none of the correlations involving PPVT or TONI scores, except for one, reached statistical significance or showed a clear trend, thus pointing to the absence of any linear relationship between comprehension accuracy and verbal and nonverbal skills after controlling for age. The only significant association was

Table 3 Response accuracy ($N=41$) and partial correlations with PPVT and TONI raw scores (controlling for age in months) ($N=37$)

Condition	Sentence type	Means (SD)	Partial correlations	
			PPVT (raw)	TONI (raw)
Bunny-spreading	Object	80.3% (29.9)	0.26	0.21
	Container	82.3% (31.2)	0.25	0.13
Classic	Object	53.0% (35.4)	0.20	–0.07
	Container	53.5% (36.4)	0.02	–0.01
Control	Object	89.6% (21.6)	0.02	0.16
	Container	90.2% (23.6)	−0.08	0.15
Filler	Active	96.3% (7.0)	0.38*	0.12
	Passive	89.0% (13.7)	−0.22	0.19

Significance levels: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

between the filler active sentences and PPVT, which should be interpreted with caution because of the ceiling effect (comprehension accuracy = 96%). The high accuracy on both active and passive fillers indicates that children understood the instructions for the sentence-picture verification task and could process complex reversible passive sentences with a high degree of accuracy. We did not examine children's performance on the fillers sentences further, as it was unrelated to the main aims of the study.

In line with the previous findings in the literature, children's response accuracy in the classic condition was at chance ($M=53.3\%$; 95% CI = 43.5–62.8) whereas performance on the control condition approached ceiling ($M=89.9\%$; 95% CI = 87.3–96.7). Children averaged 81.3% correct in the *bunny*-spreading condition (95% CI = 69.1–94.0).

Figure 3 provides a histogram of children's scores (out of 8 trials) on the classic condition. Given that prior studies (Minai et al. 2012) split samples into subgroups of children who consistently made *q*-spreading errors versus logical responses, we examined our data for evidence of a bimodal distribution. To assess the likelihood that the children's scores on classic condition came from a normal distribution, we employed the Shapiro–Wilk test and the Anderson-Darling test of normality. Note that the null hypothesis is that the scores in the population are normally distributed. Results from Shapiro to Wilk test indicated that a normal distribution could not be assumed ($W=0.94$, $p=0.034$) whereas results from the Anderson-Darling test yielded a non-significant trend, suggesting a normal distribution could be assumed ($A=0.70$, $p=0.063$).

Given these ambiguous results, we did not attempt to split the sample to compare children who were consistently correct versus incorrect in the classic condition. Instead, we used logistic mixed-effects regression analyses to examine effects of age and sentence type (object or container) on response accuracy across conditions. These analyses included crossed random effects for subjects and items.

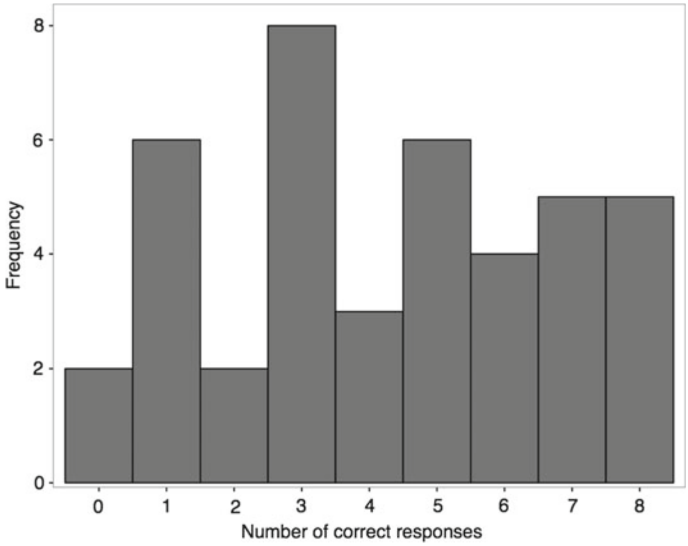


Fig. 3 Histogram of the number of correct responses (out of 8) for the classic condition ($N = 41$)

***Bunny*-Spreading Condition** Response accuracy increased significantly with age, $\beta = 0.07$, $SE = 0.04$, $z = 1.99$, $p = 0.046$, indicating that errors in this condition tended to be made only by younger children. The effect of sentence type was not statistically significant, $\chi^2(1) = 0.5$, $p = 0.818$.

Classic Condition The effect of age only approached significance, $\beta = 0.02$, $SE = 0.01$, $z = 1.72$, $p = 0.086$, indicating the occurrence of classic errors across the age range tested. There was no effect of sentence type, with children showing equivalently low accuracy on object and container sentences, $\chi^2(1) = 0.04$, $p = 0.847$.

Control Condition Similarly to the *bunny*-spreading condition, age was positively related to response accuracy, $\beta = 0.08$, $SE = 0.02$, $z = 3.70$, $p < 0.001$, indicating that the rare errors were made by younger children. As in the other conditions, there was no effect of sentence type on response accuracy, $\chi^2(1) = 0.05$, $p = 0.820$.

4.2 Eye Movements

Eye-movement analyses compared the proportions of looks to the distractors (*bunny*-spreading condition) or extras (classic condition) in the pictures as a function of comprehension accuracy. The control condition was not included due to ceiling effects on accuracy. Figure 4 and Table 4 present the results for each region of interest (ROI), as defined in Table 2. We used logistic mixed-effects regression models with crossed random effects for subjects and items and by-subject and by-item random slopes for sentence type. The analyses examined effects of age, accuracy, and sentence type on

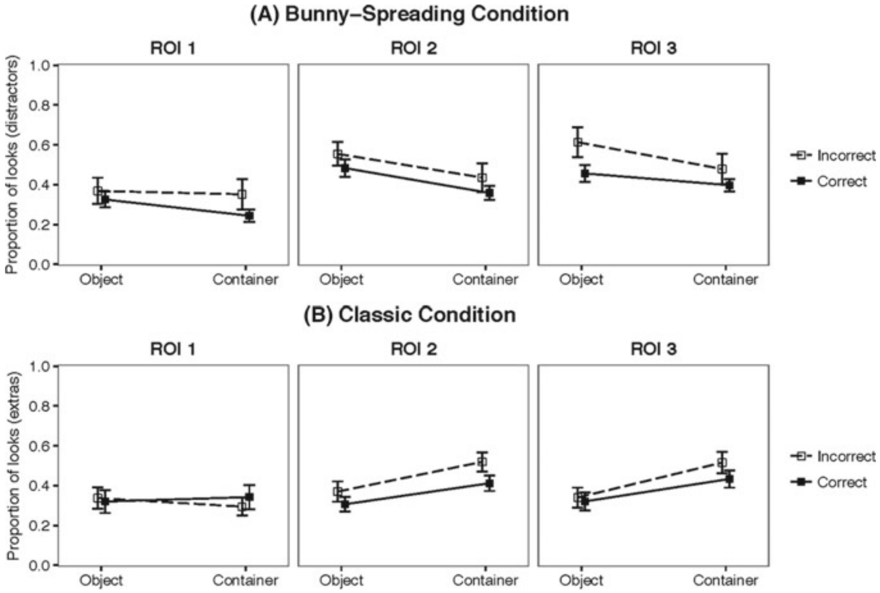


Fig. 4 Proportions of fixations to the distractors (*bunny*-spreading condition) or extras (classic condition) in each ROI. Solid line: correct trials, dashed line: incorrect trials

patterns of visual attention. Note that we retained the effect of age as a covariate in all models irrespective of its statistical significance. We also retained non-significant main effects of accuracy and sentence type if the interaction was significant.

***Bunny*-Spreading Condition** (Fig. 4a)
ROI 1 The first region of interest (ROI) consisted of the quantified noun phrase (e.g., *Every rabbit* or *Every box*). There was a significant effect of sentence type, qualified by a significant interaction of sentence type × accuracy. Children exhibited more looks to distractors when *every* modified the object rather than the container noun (i.e., they looked more at distractor cats than buckets). As shown in the left panel of Fig. 4a, sentence type yielded more of an effect on correct trials than on incorrect trials. Note also that while the left panel of Fig. 4a appears to suggest that children looked more often at distractors on incorrect trials, the main effect of accuracy failed to reach significance in ROI 1, $p = 0.064$.

ROI 2 The second ROI started at the verb (*is/has*) and continued to the end of the sentence (middle panel of Fig. 4a). Here looks to the distractors varied significantly as a function of accuracy, with children fixating more often on the distractors on incorrect trials. The effect of sentence type and the interaction of sentence type × accuracy were not statistically significant, $\chi^2(1) = 3.237$, $p = 0.072$ and $\chi^2(2) = 3.889$, $p = 0.143$, respectively.

ROI 3 The third region of interest consisted of the period of silence after the end of the sentence (right panel of Fig. 4a). Again, fixations to the distractors varied as a

Table 4 Summary of the mixed logistic analyses by ROI (fixed effects only) examining proportion of looks to the distractors (*bunny*-spreading condition) and extra objects/containers (classic condition)

	β	*SE*	*Z*	*p*-Value
Bunny-spreading condition	*ROI 1: quantified NP*			
(Intercept)	−1.33	0.53	−2.53	0.011
Age (months)	0.01	0.01	0.81	0.420
Accuracy (0 = incorrect)	−0.48	0.26	−1.85	0.064
Sentence Type (0 = container)	*1.43*	*0.53*	*2.69*	*0.007***
Accuracy × sentence type	*−0.75*	*0.32*	*−2.33*	*0.020**
	ROI 2: verb to end of sentence			
(Intercept)	−0.01	0.23	−0.03	0.976
Age (months)	−0.01	0.01	−1.26	0.210
Accuracy (0 = incorrect)	*−0.55*	*0.13*	*−4.14*	*<0.001****
	ROI 3: silence			
(Intercept)	0.23	0.21	1.09	.277
Age (months)	*0.01*	*0.01*	*2.03*	*.042**
Accuracy (0 = incorrect)	*−0.55*	*0.12*	*−4.48*	*<0.001****
Classic condition	*ROI 1: quantified NP*			
(Intercept)	−1.56	0.37	−4.22	<0.001
Age (months)	0.01	0.01	0.72	0.471
Accuracy (0 = incorrect)	0.16	0.17	0.95	0.341
Sentence type (0 = container)	−0.54	0.46	−1.18	0.238
Accuracy × sentence type	*0.63*	*0.24*	*2.66*	*0.007***
	ROI 2: verb to end of sentence			
(Intercept)	0.13	0.23	0.54	0.587
Age (months)	−0.004	0.01	−0.74	0.462
Accuracy (0 = incorrect)	*−0.46*	*0.09*	*−5.19*	*<0.001****
Sentence type (0 = container)	*−0.93*	*0.37*	*−2.49*	*0.013**
	ROI 3: silence			
(Intercept)	0.04	0.34	0.11	0.911
Age (months)	−0.01	0.01	−0.95	0.344
Accuracy (0 = incorrect)	*−0.28*	*0.11*	*−2.42*	*0.015**
Sentence type (0 = container)	*−1.06*	*0.39*	*−2.70*	*0.007***
Accuracy × sentence type	*0.52*	*0.17*	*3.06*	*0.002***

Significance levels: **p* < 0.05; ***p* < 0.01; ****p* < 0.001

Table 5 Mean reaction times for each condition as a function of response accuracy. Standard deviations are in parentheses

Condition	Sentence type	Correct trials	Incorrect trials
Bunny-spreading	Object	3256 (657)	3545 (893)
	Container	3311 (636)	3647 (1056)
Control	Object	3000 (599)	–
	Container	2822 (588)	–
Classic	Object	3433 (629)	3024 (695)
	Container	3767 (830)	3444 (764)

function of accuracy, with children looking more at the distractors on incorrect trials. Perhaps spuriously, age was also significant in the model, with older children tending to make more fixations to the distractors than younger children, although the older children made fewer incorrect responses. The effect of sentence type and the interaction of sentence type × accuracy were not statistically significant ($\chi^2(1) = 1.412$, $p = 0.235$ and $\chi^2(2) = 4.665$, $p = 0.097$).

Classic Condition (Fig. 4b)
ROI 1 In the first ROI, the only significant effect was a weak interaction of sentence type × accuracy (left panel of Fig. 4b).

ROI 2 In ROI 2, there were a significant effect of accuracy, with more looks to the extra objects/containers on incorrect trials. There was also a main effect of sentence type: Children looked more often at the extras in the pictures paired with container sentences than the extras in the pictures paired with object sentences (middle panel of Fig. 4b), perhaps because the extras were animate in half of the trials in the former condition. The interaction of sentence type × accuracy was not significant ($\chi^2(1) = 0.654$, $p = 0.419$).

ROI 3 In ROI 3, there were significant main effects of accuracy and sentence type, as well as an interaction of sentence type × accuracy. As shown in the right panel of Fig. 4b, the effect of accuracy on fixations to 'extras' was stronger for container sentences than for object sentences.

4.3 Reaction Time

Table 5 presents mean reaction times for each condition and sentence type, as a function of response accuracy. Note that in the control condition we did not examine response times for incorrect trials due to high accuracy (89.9%) yielding insufficient data for analysis.

Linear mixed-effects regression analyses were used to examine effects of age, sentence type (object vs. container) and accuracy (correct vs. incorrect trials) on children's response times.

***Bunny*-Spreading Condition** Reaction times for the *bunny*-spreading condition varied significantly with age, $\beta = -14.3$, $SE = 3.5$, $t = -4.10$, $p < 0.001$, with older children responding more quickly than younger children. Adding effects of accuracy, sentence type, and their interaction did not improve the model fit: $\chi^2(1) = 2.71$, $p = 0.099$, $\chi^2(1) = 0.13$, $p = 0.716$, and $\chi^2(3) = 4.01$, $p = 0.261$, respectively.

Classic Condition Likewise for the classic conditions, reaction times decreased significantly with age, $\beta = -13.3$, $SE = 2.9$, $t = -4.52$, $p < 0.001$. Reaction times also varied significantly as a function of accuracy, with slower responses on correct trials than on incorrect trials, $\beta = 368.5$, $SE = 113.5$, $t = 3.25$, $p = 0.001$, indicating that additional processing time was necessary for children to correctly restrict the universal quantifier. Reaction times also varied across sentence types, with faster responses when the quantifier *every* modified the object than when it modified the container ($\beta = -342.2$, $SE = 104.80$, $t = -3.27$, $p = 0.001$). Including the interaction of accuracy × sentence type did not improve the model fit, $\chi^2(1) = 0.02$, $p = 0.896$.

Classic Versus Control Conditions Next we compared the reaction times for children's correct responses in the classic and control conditions. In this analysis, the effect of age remained significant, $\beta = -13.6$, $SE = 2.6$, $t = -5.33$, $p < 0.001$, confirming faster responses in older children. The main effect of condition was also significant, $\beta = 867.8$, $SE = 110.1$, $t = 7.68$, $p < 0.001$, with slower correct responses to classic trials than control trials, which suggests that additional processing time was required for participants to avoid *q*-spreading errors. There was also significant effect of sentence type, $\beta = 227.5$, $SE = 91.4$, $t = 2.49$, $p = 0.013$, moderated by a significant interaction of condition and sentence type, $\beta = -568.6$, $SE = 156.4$, $t = -3.63$, $p < 0.001$). The reaction time difference that favored object over container sentences in the classic condition was absent (i.e., slightly reversed) in the control condition.

5 Discussion

The current study explored patterns of visual attention associated with *q*-spreading errors in school-age children (age range 5–12 years). At this age, performance of the group was expected to be at chance (~50% correct) on trials designed to elicit the classic error; thus, we sought to compare patterns of visual attention over roughly equal numbers of correct and incorrect trials with the goal of identifying a signature pattern of attention associated with committing the error. We administered the sentence-picture verification task using two distinct locative constructions: In object sentences, the universal quantifier *every* modified the designated objects (as in *Every bunny is in a box*), and in container sentences, it modified the designated containers (as in *Every box has a bunny in it*). Across trials, we varied the pictures to elicit different types of errors. For the *bunny*-spreading condition, we presented three bunny-box pairs along with two unrelated distractors (e.g., cats or buckets, as depicted in Fig. 1a, b). For classic and control conditions, we presented three bunny-box pairs along with two extra objects or containers (e.g., bunnies or boxes, as depicted in Fig. 2a, b). The classic condition was expected to elicit (over-exhaustive) *q*-spreading wherein the scope of the universal quantifier spreads beyond its subject. That is, we expected children to incorrectly reject the sentence *Every bunny is in a box* as a description of a scene with extra boxes. By swapping the pairings of sentences and pictures, the same stimuli were used for the control condition, wherein children were expected to correctly reject the sentence *Every box has a bunny in it* as a description of a scene with extra boxes.

The task was conceptually identical to the truth-value judgment task used in previous offline studies of classic *q*-spreading in children (see Rakhlin 2007, for an overview). We adapted procedures to allow for concurrent online recordings of eye-movements using the visual world paradigm (cf. Minai et al. 2012; Sekerina and Sauermann 2015). The paradigm allowed us to examine how children allocated their visual attention as each sentence unfolded in real time in order to determine whether increased looks to the distractors (*bunny*-spreading condition) or extras (classic condition) were time-locked to the region of interest immediately following the quantified NP. In addition to exploring how specific eye-movement patterns might be associated with *q*-spreading, we examined susceptibility to errors in relation to individual differences in non-verbal and verbal intelligence as well as age. We also measured reaction times as an additional variable to determine whether errors were associated with greater or lesser processing time relative to correct responses.

With regards to accuracy in performing the sentence-picture verification task, the children did quite well on the control (89.9% correct) and *bunny*-spreading (81.3%) conditions, as well as on the reversible active (96.3%) and passive (89.0%) sentences used as fillers. These findings indicate that children understood the task instructions and could succeed in sentence-picture verification with a variety of sentence structures. In contrast, the children exhibited the well-established classic error on trials that depicted extra objects and containers that were outside the scope of the universal quantifier, with the group performing at chance (53.3% correct).

For the *bunny*-spreading condition, accuracy increased with age, in line with prior work associating these errors with young children (Roeper et al. 2004). For the classic condition, the correlation between accuracy and age was not statistically significant ($p = 0.086$). The lack of a robust effect of age on classic *q*-spreading makes sense in light of findings from a group of college students performing the sentence-picture verification task with the same set of materials used in the current study (Brooks and Sekerina 2006), wherein one in five adults performed at chance in the classic condition. Taken together with other findings demonstrating classic errors in college students performing the picture-choice task (Brooks and Sekerina 2005, Experiment 3), the results suggest that classic *q*-spreading is less constrained by maturation than was previously thought and persists into adulthood. This set of findings is difficult to accommodate within frameworks that assume faulty grammar to be the source of children's errors (e.g., Philip 1995; Roeper et al. 2004), and are more consistent with shallow processing accounts that attribute errors to superficial processing of sentence structure (Brooks and Sekerina 2005, 2006).

In examining individual differences in relation to adult performance on the sentence-picture verification task, Brooks and Sekerina (2006) reported significant correlations between accuracy on classic trials and nonverbal intelligence (estimated using the Culture Fair Intelligence Test; Cattell and Cattell 1973) and need for cognition scores (Cacioppo et al. 1984). Similarly, in accounting for individual differences in adult performance on the picture-choice task, Street and Dąbrowska (2010) reported an association between classic errors and need for cognition, while also finding an association with self-reported time spent reading for pleasure. However, in the current study, we failed to find evidence in children for a relationship between comprehension accuracy and individual differences in either nonverbal or verbal intelligence (estimated using TONI-3 and PPVT-4 standardized assessment tests, respectively). Although the topic of individual differences in children's sentence processing clearly warrants additional research, the lack of any clear predictive relationships between classic *q*-spreading and nonverbal or verbal intelligence in children appears to underscore their broad susceptibility to the classic error.

Prior work with heritage speakers of Russian documented a signature pattern of visual attention associated with classic *q*-spreading in bilingual adults (Sekerina and Sauermann 2015): When committing the error, adults made a greater number of fixations to the extra objects/containers in the pictures, with the increased rate of fixations time-locked to the ROI immediately following the quantified noun phrase (ROI 2, defined as the verb). In line with well-established results from other studies using the Visual World Paradigm (Trueswell and Tanenhaus 2004), Sekerina and Sauermann (2015) interpreted the changing patterns of eye-movements to be a reflection of the participant's interpretation of the sentence as it unfolded in time. Thus, increased looks to the extras were a direct index of their spreading the domain of the universal quantifier beyond its subject, to encompass both the object and container nouns. Sekerina and Sauermann's findings contrast with those of Minai and colleagues (2012) who focused on executive control of attention in relation to classic errors in preschool-age children. In support of their hypothesis that a lack of control of attention in children increases their susceptibility to classic errors, the children showed a large increase in fixations to the extra object(s) *prior* to the onset of the sentence, when compared with adults. However, no evidence of aberrant patterns of visual attention *during* sentence processing was found.

Given these conflicting reports, the current study sought evidence for a signature pattern of visual attention associated with classic errors in a sample of school-age children. If classic errors involve a spread of visual attention as the scope of the universal quantifier is extended beyond its complement NP, we should see increased fixations to the extra objects/containers that are time-locked to when the error is made in sentence processing—presumably just as soon as the children have interpreted the quantified noun phrase (*every bunny* or *every box*). This is indeed what we found: *q*-spreading errors were associated with increased looks to the distractors (*bunny*-spreading trials) and the extras (classic trials) relative to correct responses. The increased fixations became significant in ROI 2 (extending from the verb to the end of the sentence), and remained significant throughout ROI 3 (silence). In other words, the eye-movement patterns associated with *q*-spreading gained strength as the spoken sentence unfolded, in line with the view that incremental sentence interpretation guides visual attention (for an overview, see Huettig et al. 2011), with the acknowledgement that effects can be bi-directional.

We analyzed reaction times as an additional dependent variable to determine whether classic errors were associated with greater or lesser processing time relative to correct responses. The analyses indicated a significant age-related decrease in overall reaction times; however this effect should be interpreted with caution as older children responded directly using the gamepad whereas younger children responded orally, with the experimenter registering their responses on the gamepad. In addition to the main effect, we found significantly slower responses on correct trials (3600 ms) than on incorrect trials (3234 ms) in the classic condition. This pattern is compatible with our hypothesis that additional processing time and effort are necessary for children to overcome shallow processing of sentence structure to correctly restrict the universal quantifier to the appropriate noun phrase. It argues against the hypothesis that errors are driven by distraction from the extra object/containers (i.e., a failure in executive control of attention), as distraction should lead to slower reaction times for incorrect trials. The faster response times associated with the classic error (and also with correct rejections of sentences in the control condition) support the view, initially proposed in Brooks and Braine (1996), that children adopt a default expectation that distributive quantifiers (e.g., *each* and *every*) refer to sets in one-to-one correspondence. Reliance on this default assumption leads to a quick rejection of pictures showing sets in partial one-to-one correspondence, which is overcome when children attend to linguistic cues that provide additional constraints on sentence interpretation.

Notably, in the classic condition, the children did not show higher accuracy for object sentences in comparison to container sentences, as had been reported in a prior study of classic errors in adults using the picture-choice task (Street and Dąbrowska 2010). Children did, however, show faster reaction times for object sentences than container sentences, which suggests that these sentences were somewhat easier to process. Eye-movement analyses also revealed an influence of sentence type on patterns of visual attention. In the *bunny*-spreading condition, object-sentences elicited greater numbers of looks to the distractors than container-sentences. That is, distractor cats in the context of the sentence *Every bunny is in a box* attracted greater attention than distractor buckets in the context of the sentence *Every box has a bunny in it.* In the classic condition, container-sentences tended to attract greater looks to the extra objects (e.g., bunnies) than object-sentences to extra containers (e.g., boxes). Thus, across both of these conditions, children's attention was drawn to distractors/extras that constituted the figure/object, as opposed to the ground/container, perhaps because the figure/object was animate in half of the trials. These findings complement research by Freeman (1985) demonstrating the impact of the visual context on children's errors in sentence comprehension.

In conclusion, the current study contributes to the vast literature on *q*-spreading in children by identifying a pattern of visual attention associated with committing the error in real time. We offer a unified account of school-age children's classic errors and the less-than-perfect performance of monolingual and bilingual adults that attributes errors to shallow sentence-processing strategies as opposed to immature or faulty grammar.

Acknowledgements Irina Sekerina and Patricia Brooks share first authorship and made equal contributions to the study. We extend our gratitude to the parents, teachers, and children at St. Thomas-St. Joseph School in Staten Island, NY, for their involvement in the study. We also thank Annemarie Donachie, Carolin Obermann Linxen, and Kasey Powers for their assistance in collecting and tabulating the data.

References

Baayen, R.H., D.J. Davidson, and D.M. Bates. 2008. Mixed-effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language* 59 (4): 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-effects Models using lme4. *Journal of Statistical Software* 67 (1). https://doi.org/10.18637/jss.v067.i01.

Berent, G.P., R.R. Kelly, and T. Schueler-Choukairi. 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: 251–290.

Bliese, P.D., and R.E. Ployhart. 2002. Growth Modeling using Random Coefficient Models: Model Building, Testing, and Illustrations. *Organizational Research Methods* 5 (4): 362–387. https://doi.org/10.1177/109442802237116.

Brooks, P.J., and M.D.S. Braine. 1996. What do Children Know About the Universal Quantifiers 'All' and 'Each'? *Cognition* 60: 235–268. https://doi.org/10.1016/0010-0277(96)00712-3.

Brooks, P.J., and I.A. Sekerina. 2005. Shortcuts to Quantifier Interpretation in Children and Adults. *Language Acquisition* 13: 177–206. https://doi.org/10.1207/s15327817la1303_2.

Brooks, P.J., and I.A. Sekerina. 2006. Shallow Processing of Universal Quantification: A Comparison of Monolingual and Bilingual Adults. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society,* ed. R. Sun, 2450. Mahwah, NJ: Erlbaum.

Brown, L., R.J. Sherbenou, and S.K. Johnson. 1997. *The Test of Nonverbal Intelligence: A Language Free Measure of Cognitive Ability*, 3rd ed. Austin, TX: Pro-Ed.

Bucci, W. 1978. The Interpretation of Universal Affirmative Propositions. *Cognition* 6 (1): 55–77.

Cacioppo, J.T., R.E. Petty, and C.F. Kao. 1984. The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment* 48: 306–307.

Cattell, R.B., and H.E.P. Cattell. 1973. *Measuring Intelligence with the Culture-fair Tests*. Champaign, IL: Institute for Personality and Ability Testing.

Clahsen, H., and C. Felser. 2006. Grammatical Processing in Language Learners. *Applied Psycholinguistics* 27: 3–42. doi:10.1017.S0142716406060024.

Clark, H.H., and W.G. Chase. 1972. On the Process of Comparing Sentences Against Pictures. *Cognitive Psychology* 3 (3): 472–517.

Crain, S., R. Thornton, C. Boster, L. Conway, D. Lillo-Martin, and E. Woodams. 1996. Quantification Without Qualification. *Language Acquisition* 5 (2): 83–153. https://doi.org/10.1207/s15327817la0502_2.

DelliCarpini, M. 2003. Developmental Stages in the Semantic Acquisition of Quantification by Adult L2 Learners of English: A Pilot Study. In *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002): L2 Links*, ed. J.M. Liceras, H. Zobl, and H. Goodluck, 55–63. Somerville, MA: Cascadilla.

Donaldson, M., and P. Lloyd. 1974. Sentences and Situations: Children's Judgments of Match and Mismatch. In *Problèmes actuels en psycholinguistique, Colloques internationaux du Centre*

National de Recherche Scientifique, ed. F. Bresson, 74–86. Paris: Centre National de Recherche Scientifique.

Drozd, K. 2001. Children's Weak Interpretations of Universally Quantified Questions. In *Language Acquisition and Conceptual Development*, ed. M. Bowerman and S.C. Levinson, 340–376. Cambridge, UK: Cambridge University Press.

Dunn, L.M., and D.M. Dunn. 2007. *Peabody Picture Vocabulary Test*, 4th ed. Pearson.

Felser, C., and H. Clahsen. 2009. Grammatical Processing of Spoken Language in Child and Adult Language Learners. *Journal of Psycholinguistic Research* 38 (3): 305–319.

Ferreira, F. 2003. The Misinterpretation of Noncanonical Sentences. *Cognitive Psychology* 47 (2): 164–203.

Ferreira, F., K.G. Bailey, and V. Ferraro. 2002. Good-enough Representations in Language Comprehension. *Current Directions in Psychological Science* 11 (1): 11–15.

Forster, K.L., and J.C. Forster. 2003. DMDX: A Windows Display Program with Millisecond Accuracy. *Behavior Research Methods, Instruments, and Computers* 35: 116–124.

Freeman, N.H. 1985. Reasonable Errors in Basic Reasoning. *Educational Psychology* 5 (3–4): 239–249.

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

Huettig, F., J. Rommers, and A.S. Meyer. 2011. Using the Visual World Paradigm to Study Language Processing: A Review and Critical Evaluation. *Acta Psychologica* 137: 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003.

Inhelder, B., and J. Piaget. 1964. *The Early Growth of Logic in the Child: Classification and Seriation*. Oxon, UK: Routledge.

Jaeger, T.F. 2008. Categorical Data Analysis: Away from ANOVAs (transformation or not) and Towards Logit Mixed Models. *Journal of Memory and Language* 59 (4): 434–446. https://doi.org/10.1016/j.jml.2007.11.007.

Minai, U., N. Jincho, N. Yamane, and R. Mazuka. 2012. What hinders Child Semantic Computation: Children's Universal Quantification and the Development of Cognitive Control. *Journal of Child Language* 39: 919–956. https://doi.org/10.1017/S0305000911000316.

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

Philip, W. 1995. Event Quantification in the Acquisition of Universal Quantification. Doctoral Dissertation, University of Massachusetts, Amherst, GLSA Publications.

R Core Team. 2014. R: A Language and Environment for Statistical Computing (version 3.1.0). Vienna, Austria: R Foundation for Statistical Computing.

Rakhlin, N. 2007. A New Pragmatic Account of Quantifier-spreading. *Nanzan Linguistics* 8: 239–282.

Roeper, T., and J. De Villiers. 1993. The Emergence of Bound Variable Structures. In *Knowledge and Language, Volume 1, from Orwell's Problem to Plato's Problem*, ed. E. Reuland and W. Abraham, 105–139. Kluwer Academic Publishers.

Roeper, T., U. Strauss, and B.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.

Sanford, A.J., and P. Sturt. 2002. Depth of Processing in Language Comprehension: Not Noticing the Evidence. *Trends in Cognitive Sciences* 6 (9): 382–386.

Sekerina, I.A., and A. Sauermann. 2015. Visual Attention and Quantifier-spreading in Heritage Russian Bilinguals. *Second Language Research* 31 (1): 75–104. https://doi.org/10.1177/0267658314537292.

Street, J.A., and E. Dąbrowska. 2010. More Individual Differences in Language Attainment: How Much do Adult Native Speakers of English Know About Passives and Quantifiers? *Lingua* 120: 2080–2094.

Trueswell, J.C., and M.K. Tanenhaus (eds.). 2004. *Approaches to Studying Word-situated Language Use*. Cambridge, MA: MIT Press.

Turning Adults into Children: Evidence for Resource-Based Accounts of Errors with Universal Quantification

Oliver Bott and Fabian Schlotterbeck

Abstract The present study shows that adults make errors of quantifier spreading similar to those commonly observed in preschool children when interpreting universally quantified sentences. In a resource demanding version of the truth value judgment task, adult participants often rejected scope disambiguated, universally quantified sentences (e.g. *Every kid is such that it was praised by exactly one teacher*) in situations where each kid was praised by exactly one teacher, but there was (A) an additional teacher praising no kids (= classic quantifier spreading) and/or (B) a teacher who praised more than one kid. While the classic spreading error has been studied extensively, spreading errors of the second type have not been attested in the acquisition literature. Neither type of error occurred in an ordinary picture verification task using the same materials. A third experiment ruled out the possibility that the errors observed in Experiment 1 are due to misrepresenting the situations in memory. Our results are most consistent with resource-based accounts of quantifier spreading (e.g. Geurts in Lang Acquis 11:197–218, 2003) but are unexpected under the discontinuity hypothesis (e.g. Philip in Event quantification in the acquisition of universal quantification. UMI, Ann Arbor, Michigan 1995) and accounts relying on plausible dissent (e.g. Crain et al. in Lang Acquis 5:83–153, 1996). We outline a novel explanation of quantifier spreading in terms of the computation and evaluation of default models that can account for the presented results as well as earlier findings reviewed in the introduction of the chapter.

Keywords Quantifier spreading · Spreading errors in adults · Extra-object error
Branching error · Resource-based account · Default interpretation
Incremental truth value judgment task

O. Bott (✉) · F. Schlotterbeck
University of Tübingen Project B1, Collaborative Research Centre SFB 833 & Project CiC, DFG Priority Program XPrag.de, Nauklerstr. 35, 72074 Tübingen, Germany
e-mail: oliver.bott@uni-tuebingen.de

F. Schlotterbeck
e-mail: fabian.schlotterbeck@uni-tuebingen.de

K. É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47,
https://doi.org/10.1007/978-3-319-91566-1_9

1 Introduction

Do children's and adults' logical concepts differ from each other? Consider, for instance, universal quantification in natural language expressed by the determiner *every*. It is commonly assumed that adult English speakers will interpret *every A is B* as *A*, the determiner's restrictor argument, is a subset of *B*, its scope argument (e.g. Barwise and Cooper 1981). However, since the pioneering work of Inhelder and Piaget (1959), it has been well established that children show certain nonadult responses when interpreting universally quantified sentences. A particularly striking finding is that children tend to judge a sentence like *every boy is riding an elephant* to be false in a context with each boy riding an elephant and an extra elephant without a rider. They, however, behave like adults when no extra elephant is present in the scene. This error, to which we henceforth refer to as extra-object (*eo*) error, is known as exhaustive pairing (Drozd and van Loosbroek 2006), Type-A error (Geurts 2003), classic spreading (Roeper et al. 2004; Sekerina et al. this volume), or overexhaustive spreading error (Sekerina and Sauermann 2015). Not only do children seem to restrict the quantificational domain (i.e. the set of individuals *every* depends on) to the set of boys but also falsely consider the whole set of elephants instead of only taking into account those elephants that are ridden by a boy. At first sight, at least, it thus seems as if children employ an essentially non-conservative concept of universal quantification—in sharp contrast with commonly held beliefs about universal constraints on natural language quantification (Hunter and Lidz 2013). Beyond this kind of spreading error, other errors have been reported in the literature, too.[1]

The present chapter presents an experimental study that provides evidence that not only children but also adults are prone to the extra-object error. It contrasts the classic extra-object error with a new type of interpretation error, which we henceforth refer to as branching error. Our study shows that both types of errors only occur in adults if they are in a highly resource demanding timed task. Taken together, these findings have important consequences for theoretical accounts of quantifier spreading and we outline a novel explanation of extra-object and branching errors.

1.1 Different Types of Quantifier Spreading

We would like to be more precise about what a spreading interpretation of *every* is. The truth conditions in (1-a) and (1-b) are a first approximation. A spreading interpretation of *every boy loves a girl* requires both (1-a) and (1-b), i.e. their conjunction, to be true. We refer to the first condition as the *regular truth condition* and the second

[1]See, e.g. Philip (2011) for an extensive review and discussion of the various errors with universal quantification.

as the *spreading condition*. Note that the spreading condition (1-b) results from (1-a) by interchanging the quantifiers' restrictor arguments with each other.[2]

(1) a. $\forall x(boy(x) \rightarrow \exists y(girl(y) \wedge love(x, y)))$
b. $\forall y(girl(y) \rightarrow \exists x(boy(x) \wedge love(x, y)))$

An extra girl without a lover clearly violates (1-b) leading to a spreading error. Interestingly, if we substitute *exactly one girl* for *a girl*, a spreading interpretation should give rise to another kind of error in addition to an extra-object error. The regular truth conditions are provided in (2-a) and the spreading interpretation results from their conjunction with the spreading condition (2-b).

(2) a. $\forall x(boy(x) \rightarrow \exists y(girl(y) \wedge love(x, y) \wedge \forall z(girl(z) \wedge love(x, z) \rightarrow z = y)))$
shorthand: $\forall x(boy(x) \rightarrow \exists^{=1} y(girl(y) \wedge love(x, y))$
b. $\forall y(girl(y) \rightarrow \exists x(boy(x) \wedge love(x, y) \wedge \forall z(boy(z) \wedge love(z, y) \rightarrow z = x)))$
shorthand: $\forall y(girl(y) \rightarrow \exists^{=1} x(boy(x) \wedge love(x, y))$

Consider a situation in which each boy loves exactly one girl but one girl is loved by two boys. Then (2-b) is false, although, under standard truth conditions (2-a), the sentence is true. Using *every* and *exactly one* thus offers the possibility to investigate whether beyond the extra-object error this second kind of spreading error occurs, too. This type of spreading error illustrates what we refer to as *branching errors* due to the feature *branching lines* ($+bl$); cf. the [$-eo$, $+bl$] conditions with branching lines illustrated in Fig. 1. Note that even though no extra object is present, the depicted situation still violates a one-to-one mapping between elements in the restrictor and the scope set. If both types of errors occur, we would like to know whether they are equally frequent and whether their proportions add up when an extra and a shared object are both present in the scene.

1.2 Theoretical Accounts on Spreading Errors

Three broad classes of theories have been proposed to account for quantifier spreading in children. Although originally invented to account for children's errors, these theories make crucially different predictions of how adults treat universal quantification and whether their representations differ from children's.

According to the discontinuity hypothesis, the linguistic representation of the universal quantifier in the adult system differs from the representation of *every* in child language. Along these lines, one view is that quantifier spreading is due to deficient syntactic representations (among others Roeper and Mattei 1974; Philip

[2] To keep things simple, in (1-a), (1-b), (2-a), and (2-b) first-order formulae are provided for the doubly quantified sentences investigated in this paper. Strictly speaking the terms restrictor and scope argument of course refer to the two arguments of generalized quantifiers of type $\langle 1, 1\rangle$.

1995; Roeper et al. 2004). For instance, Roeper and Mattei (1974) assume that children misinterpret *every* as a sentential adverb, i.e. they interpret *every boy is riding an elephant* as in all subevents involving a boy and/or an elephant, the boy is riding the elephant. Building on this idea, Roeper et al. (2004) offer a detailed account of how *every* starts as an adverbial intensifier and progressively develops into the adult determiner. One prediction that follows immediately from the discontinuity hypothesis is that quantifier spreading should not occur in adults. It is important to emphasize that the above cited studies on extra-object errors in L2 comprehension cannot be taken to disconfirm the discontinuity hypothesis. It is well possible that the universal quantifier in the L2 as well as in child language is qualitatively different from the universal quantifier in the adult system of the L1.

Several authors (Brooks and Braine 1996; Drozd 2001; Geurts 2003; Drozd and van Loosbroek 2006; Brooks and Sekerina 2006; Sekerina and Sauermann 2015; Sekerina et al. this volume) proposed that young children (and L2 comprehenders) employ an adult-like representation of *every* but face a performance problem when processing the universal quantifier. Geurts (2003), for instance, argues that spreading may be due to the particularly difficult mapping from syntax to semantics that is required to compute strong determiners like *every*. Weak determiners like *five* in *five boys are riding an elephant* are proposed to be easier, because they only depend on one set, namely the set of elephant riding boys. By contrast, to evaluate a strong determiner one has to consider the whole restrictor set, i.e. the non-riders, too. According to Geurts, spreaders incorporate all the information into the nuclear scope resulting in an empty restriction and, therefore, initially misconstrue *every* as a weak quantifier. However, since children employ an adult-like representation of *every* that requires restriction, they are in a subsequent step forced to restrict its domain on pragmatic grounds. This explains why the sentence *every boy is riding an elephant* can end up semantically as *every elephant is such that a boy is riding it* but does not have to necessarily. Salience is thought to lead the pragmatic choice of what information is used for purposes of restriction. Adding more extra elephants to the situation will decrease their salience and thus reduce spreading errors. This is exactly what Gouro et al. (2001) and Sugisaki and Isobe (2001) report. An interesting prediction of performance-based accounts is that adults should also make spreading errors in their L1 when in a task that severely limits their processing resources.

Finally, various types of pragmatic explanations have been proposed for the extra-object error. Their common core is that the testing situation under which children make extra-object errors are pragmatically not felicitous. What seems to be an error with universal quantification rather corresponds to the rejection or an implicit accomodation of this situation. Crain and colleagues argue that children's errors are an artifact of inappropriate testing procedures that do not satisfy the felicity conditions of yes-no questions which they call *plausible dissent* (Crain et al. 1996; Meroni et al. 2006). According to their view, uttering a yes-no question is only felicitous when at some point during the conversation both answer alternatives are under consideration. The question *is every boy riding an elephant* is thus only appropriate when at some time during an experimental trial the child can conceive of a situation different from the actual scene (i.e. where some boy is not riding an elephant). Crain et al. (1996, Experiments 3+4) showed that children making spreading errors in the classical task were almost flawless when the condition of plausible dissent was satisfied. Based on these results and theoretical considerations, they claim that even preschool children have full competence with universal quantification. If plausible dissent is not satisfied, even older children or adults may have problems with pragmatically odd material. In fact, an eyetracking study by Meroni (2002) employing the visual world paradigm (see, e.g. Tanenhaus et al. 1995) provided initial evidence for plausible dissent affecting adults' online comprehension. Participants spent longer looking at an extra object in a scene when the condition of plausible dissent was not fulfilled than when it was. However, the scenes varied with respect to the number of objects being present making it difficult to disentangle perceptual/attentional effects from effects of semantic interpretation.

Other pragmatic accounts of the extra-object error have been recently proposed by Philip (2011) and É. Kiss and Zétényi (2017). According to the relevance account by Philip (2011), children face a problem when having to decide on a universal quantifier's domain. They determine relevance by means of a perceptual mechanism that puts strong emphasis on a symmetrical pairing between the restrictor set and the scope set. He assumes that this cognitive mechanism is actually so strong that they tend to implicitly extend the quantificational domain and accomodate the testing situation by including an unseen object into their context representation in order to recover symmetry. What is important for the present chapter is that this explanation can account for the extra-object error but cannot be easily extended to the branching error reported on below. É. Kiss and Zétényi (2017) propose a different relevance theoretic explanation of the extra-object error. They hypothesize that in an experimental testing situation where the stimuli are not embedded in a context, all elements of the testing situation are inferred to be relevant due to ostensive effects (Csibra and Gergely 2009). Accordingly, the child is confronted with a pragmatically odd situation in which the extra object is presented as relevant but—at the same time—has to be ignored because it is truth conditionally irrelevant. This pragmatic violation leads them to reject the sentence, or rather the whole situation. Similar to Philip (2011), the explanation is tailored to the extra-object error and cannot be easily extended to branching errors reported below.

1.3 Spreading Errors in Adults?

Children's extra object errors with universal quantification are typically contrasted with that of adult control groups who are believed to be error free. This has recently been empirically challenged by a number of studies who reported extra-object errors in adults, too. Initial evidence for spreading in adults has been provided by Brooks and Sekerina (2006, Experiment 3). In a timed picture selection experiment, adult participants made approximately 20% errors in scenes containing an extra object. Related to this finding Street and Dąbrowska (2010) showed that less educated monolingual speakers of English made errors more than 20% of the time in a picture selection task of the same kind as often administered to children. They attributed their findings to different underlying grammatical rules reflecting qualitatively poorer linguistic experience in the less educated group.

There is growing evidence that adult second language (L2) learners are also prone to extra-object errors. DelliCarpini (2003) tested learners of various proficiency levels and observed 70% extra-object errors in a first group of learners with lower proficiency but only 35% errors in a second group with higher proficiency. The existence of extra-object errors in L2 comprehension has been corroborated more recently by a study by Berent et al. (2009), even though the authors did not find any differences between L2 comprehenders of different proficiency levels. Finally, Sekerina and Sauermann (2015) tested extra-object effects in a special group of bilingual adults, namely bilingual heritage speakers of Russian with English as their dominant language. They monitored heritage speakers' eye movements while they interpreted Russian sentences with universal quantifiers and provided a truth-value judgment for visually presented scenes with or without extra objects. The eyetracking record and the judgment data revealed that the heritage speakers made extra-object errors in their heritage language but not in their dominant language English. Sekerina and Sauermann (2015) proposed that extra-object errors with universal quantification are due to cognitive (over)load—a resource-based account of the extra-object error.

1.4 The Present Study

The present study investigates whether adults are prone to spreading errors in their L1 as well, and whether we find evidence for task demands to affect error rates. Moreover, we are interested in whether, besides the classic extra-object error, there is also evidence for a second kind of spreading errors, the above-outlined branching errors. Taken together, these two aspects are highly relevant for the theories reviewed above because they make rather different predictions. According to the discontinuity hypothesis, no spreading should occur in adults. Plausible dissent predicts essentially the same. In case the condition is satisfied there should be no quantifier spreading. By contrast, resource-based accounts predict spreading errors even in adults if the task is so demanding that they face cognitive overload while processing *every*. Last but not least, relevance-theoretic pragmatic accounts may be consistent with extra-object errors in adults but, since they are tailored to this particular kind of error, they do not predict other types of spreading errors such as branching errors.

We conducted three experiments to test these predictions. All of them used tasks that had been pretested in methodological work before. The first experiment employed an ordinary truth value judgment task without any time pressure to assess whether spreading errors occur in L1 adults when both sentence and picture are present throughout the trial. We drew upon the work by Bott and Radó (2007) who compared three different offline tasks to assess the meaning of quantified sentences. Their study revealed that a picture verification task like the one used here using abstract set diagrams is a highly reliable offline method to assess the interpretation of doubly quantified statements in adults. In a cross-methodological comparison the evaluation of abstract set diagrams turned out to be even less biased than the evaluation of concrete scenarios. The first experiment provided a baseline for the second experiment, which used a resource demanding version of the picture verification task testing the same materials. In the *Incremental Truth Value Judgment Task* participants first inspected a set diagram and had to uphold this picture in working memory while incrementally providing truth value judgments for universally quantified sentences presented in a self-paced fashion to them (see Bott and Schlotterbeck 2012, for methodological discussion of this task). This kind of dual-task setting imposed enhanced processing load on them and thus should give rise to spreading errors. Specifically, having to keep a picture in memory while incrementally comparing it to the incoming sentence should severly tax working memory. The last experiment was intended to rule out an alternative, non-linguistic explanation, namely that the observed errors are not due to semantic interpretation but are artefacts of the task employed in the second experiment.

2 Experiment 1

The first experiment employed an ordinary picture verification task in order to provide us with a baseline of how often the two discussed types of spreading errors, i.e. extra-object and branching-line errors, occur when participants have access to both a picture and a sentence throughout the trial, that is, when processing load is kept minimal. Moreover, in the first experiments there was no time pressure for providing an answer and the experiment was kept as short as possible to avoid errors due to fatigue.

2.1 Method

2.1.1 Participants

24 native German speakers with German as their L1 (mean age 24.7 years; range 20–35 years; 20 female) studying at Tübingen University took part in the experiment for course credit or 5€.

2.1.2 Materials

We constructed German doubly quantified sentences like (3) with the universal quantifier *jeder* (*every*) scoping over *genau ein* (*exactly one*). The quantifiers were separated by a clause boundary after *gilt* (*holds*) creating a scope island that blocks inverse scope (cf. Bott and Radó 2007, as well as Bott and Schlotterbeck 2012, for the same kind of constructions in the Incremental Truth Value Judgment Task). The universal quantifier was always picked up by the object pronoun in the embedded sentence.

(3) Für jeden Schüler gilt: ihn lobte genau ein Lehrer voller
for each pupil holds: him praised exactly one teacher full-of
Wohlwollen.
goodwill

'Each pupil is such that exactly one teacher praised him full of goodwill.'

The 16 experimental items were paired with four pictures each, like those in Fig. 1. The diagrams depicted sets and the relations between their elements. They were labeled as in Fig. 1 to provide participants with all the necessary information to interpret the diagrams without any ambiguity. The set diagrams always showed the set corresponding to the subject/agent on the left hand side and the object/patient on the right hand side. We manipulated the presence of extra (i.e. unconnected) objects in the set on the left as well as the presence of branching lines (i.e. an element in the set on the left that is connected to two elements on the right) according to a 2×2 design (EXTRA OBJECT ($\pm eo$) $\times$ BRANCHING LINE ($\pm bl$)). In all four conditions, each element in the set on the right was connected to exactly one element in the set on the left. Hence, the sentences are true in all four conditions.

Four lists were constructed according to a Latin square design. 34 filler sentences (20 false) involving various combinations of quantifiers were included in each list. All fillers were quantificational statements with a whole range of quantifiers (e.g. *no*, *most*, *less than n*...).

2.1.3 Procedure

In each trial participants had to decide whether a sentence matched a set diagram. Sentence-picture pairs were presented on the same screen and there was no time limit for providing an answer. Participants had to provide their answer by either clicking on a *yes, matches* or a *no, does not match* button. We logged both judgments and judgment times of correct responses.

The experiment started with a practice session of five trials which was followed by the 50 sentence-picture pairs which were presented in an individually randomized order. The experiment was conducted in a quiet computer pool at Tübingen University exclusively reserved for the experiment. The experiment was implemented as an internet questionnaire using WebExp2 (Mayo et al. 2006).

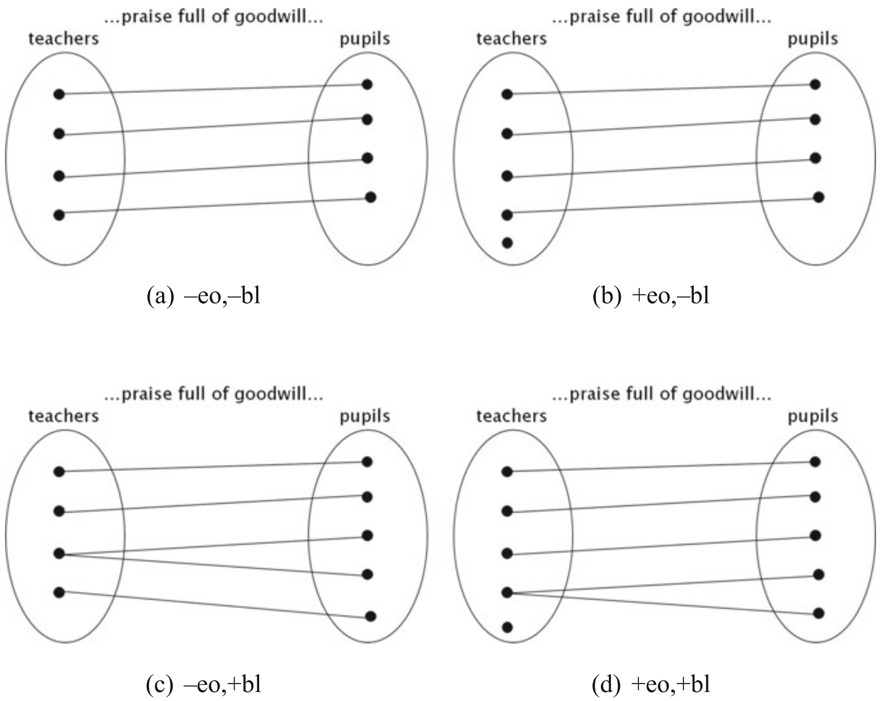


Fig. 1 Set diagrams used in Experiments 1 and 2 (labels translated from German). [+/–eo] indicates the presence/absence of an extra object; [+/–bl] the presence/absence of a branching line

2.2 Results

Inspecting the data for each individual participant showed that 22 out of the 24 participants were very consistent. Each of them had at most one error in any of the experimental trials. Only two made systematic errors. The first consistently made a branching error (rejection rates for $[-eo, -bl]$: 0%; $[+eo, -bl]$: 0%; $[-eo, +bl]$: 100%; $[+eo, +bl]$: 75%). The second had a rather high number of rejections for all four types of pictures but even more so in conditions with branching lines ($[-eo, -bl]$: 50%; $[+eo, -bl]$: 50%; $[-eo, +bl]$: 100%; $[+eo, +bl]$: 100%). We excluded the data of these two participants from further statistical analysis.[3]

The remaining 22 participants were included in the statistical analyses. The descriptive statistics are summarized in Table 1.

[3]We also computed a logit mixed effects model including all participants. This lead to a spurious main effect of *branching line* ($estimate = -1.59$; $z = -2.03$; $p < .05$). The main effect of *extra object* and the interaction were, however, far from significant (both z-values $< .5$).

Table 1 Mean proportions of errors and mean judgment times (+standard deviations in parentheses) in Experiment 1

	Errors in %	Judgment time in s
$[-eo, -bl]$	3.4	9.2 (4.4)
$[+eo, -bl]$	2.3	10.6 (5.0)
$[-eo, +bl]$	6.8	14.2 (6.8)
$[+eo, +bl]$	4.6	13.4 (5.4)

2.2.1 Accuracy

In a logit mixed effects model including the fixed effects EXTRA OBJECT, BRANCHING LINE and their interaction as well as random intercepts of PARTICIPANTS and ITEMS neither of the main effects was significant (EXTRA OBJECT: $estimate = 0.48$; $z = 0.47$; $p = 0.64$; BRANCHING LINE: $estimate = -0.81$; $z = -1.00$; $p = 0.32$) nor was their interaction ($estimate = -0.00$; $z = -0.00$; $p = 0.99$). Thus, acceptance rates did not differ between conditions.[4]

2.2.2 Response Latencies (RTs)

We corrected judgment times of "yes" judgments for outliers by removing all judgment times that were more than 2.5 standard deviations above a participant's mean (this affected 3.1% of the data). RTs were fastest in the $[-eo, -bl]$ condition. Participants took about 1 second longer to judge the $[+eo, -bl]$ condition. The two branching line conditions were judged slowest: the $[-eo; +bl]$ condition took on average five seconds longer to judge than $[-eo, -bl]$ and was even judged somewhat slower than the $[+eo, +bl]$ condition.

Judgment times were analyzed in a linear mixed effects model including EXTRA OBJECT, BRANCHING LINE, and their interaction as fixed effects as well as the maximal random effect structures of PARTICIPANTS and ITEMS. The analysis revealed a significant main effect of BRANCHING LINE ($estimate = 4903$; $t = 5.57$) reflecting longer judgment times for pictures with a branching line than for pictures without. In addition, the interaction between EXTRA OBJECT and BRANCHING LINE was marginally significant (ANOVA comparing the saturated model with a model without the interaction in the fixed effects: $\mathscr{X}^2(1) = 3.55$; $p = .06$). The interaction was due to the fact that $[+eo, -bl]$ pictures took longer to judge than $[-eo, -bl]$ pictures and that the pattern was reversed in the $[+eo, +bl]$ versus $[-eo, +bl]$ conditions. To find out whether the difference in RT between the $[-eo, -bl]$ and the $[+eo, -bl]$ condition was reliable, we computed a generalized linear mixed effects model on this

[4]Logit mixed effects models including the random slopes of EXTRA OBJECT and/or BRANCHING LINE failed to converge. No random slopes were therefore included into the statistical analysis of error rates. Since this was an issue for most logit mixed effects model analyses reported in this chapter, all models only included the random intercepts in order to guarantee comparability between experiments. All generalized mixed effects model analyses were computed with the statistical software environment R using the package *lme4*.

subset of the data. The pairwise comparison revealed a marginally significant fixed effect of EXTRA OBJECT (ANOVA comparing a model including the fixed effect of EXTRA OBJECT with a model without the fixed effect: $\mathscr{X}^2(1) = 2.76$; $p = .1$).

2.3 Discussion

In an ordinary offline version of the picture verification task adult L1 speakers of German did not make any spreading errors. This provides us with a baseline that any errors to be found in the next experiment testing the same materials must in fact be due to task demands. Interestingly, we found an increase in judgment time due to extra objects and an even bigger increase in judgment time when pictures contained a branching line. This already hints at greater task demands in conditions with extra objects and branching lines than in the $[-eo, -bl]$ condition.

3 Experiment 2

Next we investigated whether adults are prone to spreading errors when they are in a task that severly limits their processing resources. Adult participants provided truth value judgments for universally quantified sentences phrase by phrase while they had to keep a picture in mind. We modified the experimental task employed in the previous experiment. Memory load was enhanced by showing the picture separately from the sentence. In earlier research (Bott and Schlotterbeck 2012, Experiment 3) we have provided preliminary evidence that in this particular task, undergraduate college students make errors with universal quantification in their L1 reminiscent of those reported in the acquisition literature on quantifier spreading reviewed in the introduction.

3.1 Method

3.1.1 Participants

40 students from the University of Tübingen with German as their L1 (mean age 23.9 years, range 19–35 years, 31 female) took part in the study for a payment of 8€. None of them had participated in the previous experiment.

3.1.2 Materials

We took the materials from the previous experiment and constructed 16 additional items. (4) is the sample item repeated from (3) augmented with vertical lines to indicate segmentation in self-paced reading.

(4) Für jeden Schüler gilt: $|_{ROI2}$ ihn $|_{ROI3}$ lobte $|_{ROI4}$ genau ein Lehrer
for each pupil holds: | him | praised | exactly one teacher
$|_{ROI5}$ voller $|_{ROI6}$ Wohlwollen.
| full-of | goodwill
'Each pupil is such that exactly one teacher praised him full of goodwill.'

As in the previous experiment, the 32 experimental items were paired with four set diagrams each, like those in Fig. 1.[5] The picture materials of the first 16 items were taken from Expample 1. Again, we manipulated the presence of extra (i.e. unconnected) objects in the set on the left as well as the presence of branching lines (i.e. an element in the set on the left that is connected to two elements on the right) according to a 2 × 2 design (EXTRA OBJECT ($\pm eo$) × BRANCHING LINE ($\pm bl$)).

The sentence-picture pairs were distributed over eight lists according to a Latin square design. In addition, 82 filler sentences (41 false) involving various combinations of quantifiers were included in each list. All fillers were quantificational statements with a whole range of quantifiers. The 34 fillers from the previous experiment were part of this set.

3.1.3 Procedure

In each trial, participants first inspected a set diagram. After pressing a button, the diagram disappeared and the sentence was displayed self-paced using moving window presentation. At each segment, participants had to decide whether the unfolding sentence still matched the picture: they had to either choose "no, doesn't match" which aborted the trial or "yes, go on" to read the next segment.

The experiment began with a practice session of 10 sentences followed by three experimental blocks. Feedback was only provided in the practice session. Both the order between blocks and the order of sentences within blocks was randomized individually. An experimental session took about 30 min.

3.1.4 Statistical Analysis

Whenever a true sentence was aborted or a false sentence was accepted, the respective trial was counted as an error. Proportions of errors in the four experimental conditions were analyzed using a logit mixed effects model with fixed effects of EXTRA OBJECT, BRANCHING LINE, and their interaction as well as random effects of PARTICIPANTS and ITEMS (cf. Jäger 2008). In order to break down significant

[5]The experiment was part of Bott and Schlotterbeck (2012, Experiment 3) also investigating pictures that were only consistent with *exactly one* scoping over *every*. We thus had a total of eight conditions yielding four data points per participant in each condition. This is the same number of data points per condition and participant as in the previous experiment. Since this additional manipulation is not relevant here, we will restrict the discussion to the above-mentioned four conditions.

Table 2 Mean proportions of errors and mean *yes, go on* judgment RTs (+standard deviations in parentheses) of the second quantifier region in Experiment 2

	Errors in %	Judgment time in ms
$[-eo, -bl]$	1.3	744 (518)
$[+eo, -bl]$	11.3	900 (731)
$[-eo, +bl]$	45.0	1,706 (1,751)
$[+eo, +bl]$	44.4	1,724 (1,703)

interactions, we calculated models on subsets of the data as pairwise comparisons. The judgment RTs of *yes, go on* judgments for the second quantifier region (ROI 4) were analyzed in linear mixed effects model analyses including the fixed effects of EXTRA OBJECT and BRANCHING LINE and their interaction as well as the full random effect structure of participants and items (Barr et al. 2013); p-values were determined by computing model comparisons with ANOVAs. Judgment RTs were corrected for outliers by removing all judgments that were more than 2.5 standard deviations above a participant's mean judgment time (this affected 3.9% of the data).

3.1.5 Results

3.1.6 Accuracy

Table 2 presents the descriptive statistics of Experiment 2. Proportions of errors clearly differed between conditions. $[-eo, -bl]$ was rejected only 1.25% of the time, $[+eo, -bl]$ 11.25% of the time, $[-eo, +bl]$ 44.38% of the time, and $[+eo, +bl]$ 45.00% of the time. Fillers were overwhelmingly judged correctly indicating that participants were able to do the task (false fillers: 95.29% rejections; true fillers: 12.69% rejections).[6]

The logit mixed effects model analysis revealed significant main effects of BRANCHING LINE ($estimate = 5.02$; $z = 6.05$; $p < .01$), of EXTRA OBJECT ($estimate = 2.57$; $z = 3.04$; $p < .01$), and a significant interaction between the two factors ($estimate = -2.60$; $z = -2.95$; $p < .01$).

We computed two pairwise comparisons to further analyze the effect of BRANCHING LINE: The first model only included the $[-eo, -bl]$ and the $[-eo, +bl]$ conditions. It revealed a significant fixed effect of BRANCHING LINE ($estimate = 5.48$; $z = 5.53$; $p < .01$). A second model comparing the $[+eo, -bl]$ with the

[6]Because the reader may wonder about the relatively high percentage of false rejections for the true fillers a comment may be in order here. Some of the true filler trials involved doubly quantified sentences with combinations of quantifiers that turned out to be quite hard to process such as *fewer than three* and *no*, or *at most five* and *more than half* (see the experimental data and discussion in Bott et al. 2013, Experiment 2).

[$+eo$, $+bl$] condition also yielded a significant effect ($estimate = 2.49$; $z = 6.96$; $p < .01$). Thus, pictures with a branching line led to an increase in rejections irrespective of the presence or absence of an extra object.

A comparison of the [$-eo$, $-bl$] with the [$+eo$, $-bl$] condition revealed that the presence of extra objects also led to a significant increase in rejection rate ($estimate = 3.06$; $z = 3.00$; $p < .01$). Another model comparing the [$-eo$, $+bl$] with the [$+eo$, $+bl$] condition revealed no reliable difference between the two conditions (effect of EXTRA OBJECT: $estimate = -0.03$; $z = -.14$; $p = .89$). Thus, although there was a reliable EXTRA OBJECT effect when no branching line was present, this effect disappeared if the picture had a branching line.

3.1.7 Response Latencies (RTs)

The analysis of judgment RTs of *yes, go on* decisions on the second quantifier region revealed the same kind of interaction. The [$-eo$, $-bl$] condition had the shortest mean judgment RT. The RTs were approximately 150 ms higher in the [$+eo$, $-bl$] condition, and the two branching line conditions had the longest *yes, go on* RTs approximately twice as long as in the conditions without branching lines. The mixed effects model analysis revealed that the interaction was reliable (ANOVA comparing the saturated model with a model from which only the interaction was removed from the fixed effects: $\mathscr{X}^2(1) = 17.24$; $p < .01$). Two follow-up analyses, the first comparing the [$-eo$, $-bl$] with the [$+eo$, $-bl$] condition and the second comparing the [$-eo$, $+bl$] with the [$+eo$, $+bl$] condition, revealed a marginally significant EXTRA OBJECT effect in the comparison of the two [$-bl$] conditions (ANOVA comparing models with and without a fixed effect of EXTRA OBJECT: $\mathscr{X}^2(1) = 3.53$; $p = .06$) but no reliable difference in the comparison of [$+bl$] conditions ($p = .74$).

3.2 Discussion

In contrast to the previous experiment, the present experiment employing a highly resource demanding incremental version of the same picture verification task gave rise to a substantial both kinds of spreading errors in adults (cf. Table 1). Firstly, universally quantified sentences were rejected more often when they followed a picture with an extra object than after a symmetrical picture. Secondly, branching lines led to false rejections of the sentence. Compared to the classic extra-object error branching lines even led to about four times as many false rejections. The highly demanding task used here made adults favor a ‘symmetrical’ interpretation of *every* and reject universally quantified sentences when symmetry is violated similarly to children. The same pattern of effects was observed in the analysis of *yes, go on* judgment RTs of the second quantifier, i.e. the sentence region where the truth conditions of the universally quantified sentences could be determined. Surprisingly, the effects of extra objects and branching lines did not add up in the [$+eo$, $+bl$] condition. We will come back to this under-additive effect in the general discussion.

The obtained data are fully compatible with resource-based accounts of quantifier spreading. In the ordinary offline version of the picture verification task employed in Experiment 1 rejection rates did not differ reliably between the four conditions.[7] Any spreading errors in the present experiment must thus be related to the greater resource demands of the incremental version of the task. Coming back to the theoretical accounts reviewed in the introduction this suggests that children and adults may only differ in terms of resource limitations with no qualitative differences in their quantificational systems.

The findings are unexpected under any of the other types of theories discussed in the introduction. In the general discussion, we will relate the present findings to the reviewed spreading accounts from the introduction.

4 Experiment 3

We claimed that the errors in the resource demanding task of the previous experiment reflect non-standard, spreading interpretations of the quantificational statement. There is, however, a plausible alternative explanation to our findings: the errors could simply be due to encoding or retention failures of the pictures and, therefore, have nothing to do with interpretation problems.[8]

In storing images in visual short-term memory, people often remember features but fail to bind them correctly to objects (see, e.g. Wheeler and Treisman 2002). Binding problems may be especially likely to occur in a dual-task setting like the one employed in the Incremental Truth Value Judgment Task that requires to divide selective attention between keeping the picture in visual short-term memory and interpreting the incoming sentence. Note that the pictures illustrating quantifier spreading were more complex than the [-eo,-bl] controls; they contained additional visual elements which have to be properly bound to the other parts of the diagram. What kind of memory effects might be expected? First of all, participants may have confused the two sets leading to errors in all conditions except the [-eo,-bl] control. This would explain the extra-object errors. Moreover, extra objects may be easier to encode

[7] An anonymous reviewer suggested that the lack of effect in Experiment 1 may be due to the smaller number of participants. We, therefore, conducted an additional logit mixed effects model analysis directly comparing the proportions of errors in the two experiments including EXPERIMENT as an additional fixed effect besides those of EXTRA OBJECT and BRANCHING LINE, including all two-way interactions and the three-way interaction. The random effect structure only included the intercepts of PARTICIPANT and ITEM. The analysis revealed no significant three-way interaction but a marginally significant two-way interaction between the factors EXPERIMENT and EXTRA OBJECT (model comparison: $\mathscr{X}^2(1) = 3.68$; $p = .06$) and a significant two-way interaction between the factors EXPERIMENT and BRANCHING LINE (model comparison: $\mathscr{X}^2(1) = 14.27$; $p < .01$). Thus Experiment 2 in fact led to a marginally significant increase in the amount of extra-object errors and a significant increase of branching-line errors relative to Experiment 1.

[8] We are indebted to Jesse Snedeker (p.c.) for pointing out this explanation to us.

than branching lines because the latter are features relating two objects whereas the former only carry item-specific information (see, e.g. Einstein and Hunt 1980, and subsequent literature for differences in the memory of relational versus item-specific information). If this is correct, our findings would be fully expected, although there need not be any problem in adult's semantic interpretation of universally quantified sentences.

To rule out such a memory-based explanation, the present experiment employed exactly the same task as the second experiment and tested whether pictures with a branching line resulted in more errors than symmetrical pictures with a single line independent of universal quantification. In addition, we aimed at replicating the branching line effect for the universal quantifier of the second experiment. We therefore tested the $[-eo, -bl]$ and the $[+eo, +bl]$ conditions from the previous experiments and added conditions that did not involve a universal quantifier but contained numerical quantifiers instead.

4.1 Method

We created 24 new items and manipulated QUANTIFIER (*universal quantifier* versus *numerical quantifier*) and BRANCHING LINE (*symmetrical* versus *branching* diagram) according to a 2 × 2 within design. A sample item is provided in (5) and (6) with the corresponding set diagrams in Fig. 2. Note that the diagrams with a branching line always contained an extra object. The universally quantified conditions were included to replicate the difference of proportion of errors between the [–eo,–bl] and the [+eo,+bl] conditions of the first experiment.

The items were constructed in such a way that the universally quantified sentences always had an subject-before-object word order while the two numerical quantifier conditions had an object-before-subject word order. This was done for the following reasons: the sentence materials in Experiments 1 and 2 always had an object-before-subject word order, too. As in the previous experiments the set diagrams were constructed in such a way that the set denoting the subject NP was presented on the left hand side and the set denoting the object NP on the right hand side. The comparison with the numerical quantifiers thus involved the same (reverse) mapping between the two sets in the sentence and picture materials, whereas the universally quantified sentences involved a simpler, congruent mapping. The universally quantified sentences were changed in order to test whether the number of $[+eo, +bl]$ errors would be reduced in case of a congruent mapping. The critical region for RT analyses was again ROI 4.

(5) Für jeden Techniker gilt: $|_{\text{ROI2}}$ er $|_{\text{ROI3}}$ reparierte $|_{\text{ROI4}}$ genau einen
for each technician holds: | he | repaired | exactly one
Rechner.
computer
'Each technician is such that he repaired exactly one computer.'

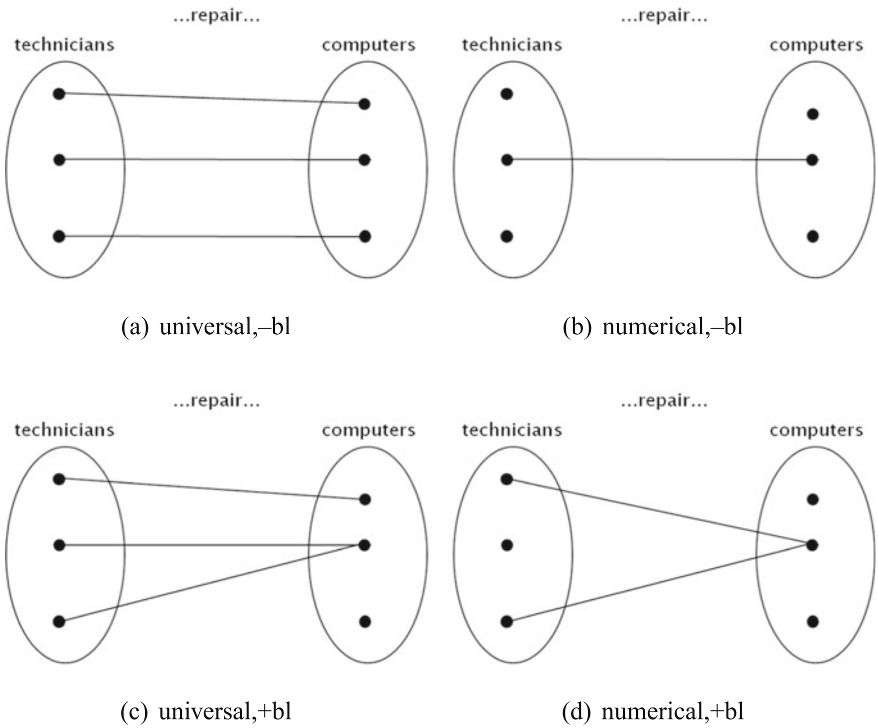


Fig. 2 Set diagrams used in Experiment 3 (labels translated from German). *Note universal/numerical* indicates the kind of quantifier; [+/–bl] indicates the presence/absence of a branching line

(6) Für einen Rechner gilt: |$_{\text{ROI2}}$ ihn |$_{\text{ROI3}}$ reparierte/n |$_{\text{ROI4}}$ einer/zwei
for one computer it holds: | it | repaired | one/two
der Techniker.
of the technicians
'One computer is such that it was repaired by one/two of the technicians.'

The 24 items, together with 102 fillers (35 false), were distributed over four lists in a Latin square design. 36 native German speakers with German as their L1 (mean age 25.0 years, range 20–40 years, 31 female) studying at the University of Tübingen participated for a payment of 8€. None of them had participated in the previous experiments. The procedure was identical to the second experiment.

Table 3 Mean proportions of errors and mean *yes, go on* judgment RTs (+standard deviations in parentheses) of the second quantifier region in Experiment 3

	Errors in %	Judgment time in ms
universal $[-eo, -bl]$	1.9 (13.5)	926 (857)
universal $[+eo, +bl]$	50.9 (50.1)	1,797 (1,348)
numerical $[-eo, -bl]$	2.8 (16.5)	1,005 (1,083)
numerical $[+eo, +bl]$	2.8 (16.5)	746 (540)

4.2 Results and Discussion

4.2.1 Accuracy

Table 3 summarizes the descriptive statistics of the present experiment. The universally quantified $[+eo, +bl]$ condition led to 50.9% errors, an even higher proportion of errors than in the first experiment, even though the present experiment employed subject-before-object sentences with a congruent mapping to the depicted sets.[9] Participants made virtually no errors in the other conditions. The universally quantified, symmetrical $[-eo, -bl]$ condition had 1.9% errors and the two conditions with the numerical quantifiers were indistinguishable from each other. They both had 2.8% errors. Branching thus had no effect on this type of sentences. Participants were equally good at remembering asymmetrical pictures with an additional line as they were when they had to keep symmetrical pictures in mind.

Error rates were analyzed in a logit mixed effects model analysis including the fixed effects of QUANTIFIER and BRANCHING LINE and their interaction and included the random intercepts of PARTICIPANTS and ITEMS. The analysis revealed a significant interaction of QUANTIFIER and BRANCHING LINE ($estimate = 4.68$; $z = 5.41$; $p < .01$), but the effects of QUANTIFIER ($estimate = -0.43$; $z = -0.60$; $p = .55$) and BRANCHING LINE ($estimate = 0.00$; $z = 0.00$; $p = 1.00$) were far from significant. Thus, branching only led to errors in the universally quantified sentences but not in sentences with numerical quantifiers.

[9] An anonymous reviewer pointed out that verbs like *repair* may bias the interpretation towards a one-to-one mapping because it is implausible that the same object is repaired twice. We therefore checked the verbs used in Experiment 3 for verbs that do not display such biases. Among these were seven psych verbs, *admire*, *amuse*, *bore*, *disappoint*, *fear*, *ignore*, and *hate*, which can simultaneously occur in different experiencers. We then compared this subset to the other seventeen items. The results were qualitatively highly similar across the two subsets of items with $\leq 2\%$ numerical difference in any of the experimental conditions. It is thus highly unlikely that the $+bl$ errors can be accounted for resorting to verb semantics.

4.2.2 Response Latencies (RTs)

The same pattern of effects was observed in the RT analysis of *yes, go on* judgments for the sentence final second quantifier region (region 4).[10] The universally quantified $[-eo, -bl]$ condition had a mean RT of 926 ms. However, the verification of the $[+eo, +bl]$ condition took much longer with a mean judgment RT of 1797 ms. Crucially, the two numerical quantifier conditions showed the opposite trend. The non-branching condition had a mean RT of 1005 ms and the branching condition was even judged faster with a mean RT of 746 ms. A linear mixed effects analysis including the fixed effects of QUANTIFIER and BRANCHING LINE as well as their interaction and the maximal random effects structures for participants and items revealed a highly significant interaction (ANOVA comparing the maximal model with a model with the interaction removed from the fixed effects: $\mathscr{X}^2(1) = 17.24$; $p < .01$). A planned comparison on a subset of the data including only the universally quantified conditions revealed that the more than 1000 ms difference between $[-eo, -bl]$ and $[+eo, +bl]$ was significant (ANOVA comparing a model with the fixed effect of BRANCHING LINE versus a model without fixed effects: $\mathscr{X}^2(1) = 11.14$; $p < .01$).

These findings rule out a memory-based explanation of the extra-object and branching line errors in the previous experiment. Instead, the difficulty observed for scenarios with branching lines must be attributed to processing difficulty related to the linguistic interpretation of universally quantified sentences.

5 General Discussion

What are the implications of our findings for the different types of theories outlined in the introduction? First of all, the experimental results are not compatible with the discontinuity hypothesis which states that language learners employ a completely different system of quantification (e.g. event quantification) than adults do in their L1. According to this hypothesis, adults should not make spreading errors irrespective of the task they perform.

The hypothesis of plausible dissent cannot explain the obtained results, either. All three experiments tested assertive sentences which are not subject to the condition of plausible dissent in the first place. A possible objection could be that participants may have construed an implicit yes-no question. However, this cannot fully explain our findings either, because even if plausible dissent should have caused a problem in the second and the third experiment, it remains unclear why it did not lead to spreading errors in the first experiment testing exactly the same materials. After all, the plausible dissent condition is not restricted by working memory limitations.

[10] As in the previous experiments judgment RTs were corrected for outliers by removing all judgment RTs that were more than 2.5 standard deviations above a participant's mean (this affected 5.1% of the data).

Other pragmatic accounts of the extra-object error cannot easily account for the observed findings, either. Neither Philip (2011)'s Relevance Account nor É. Kiss and Zétényi (2017)'s Ostensive Communication Hypothesis readily can account for the existence of branching errors.

At first glance, the findings nicely fit resource-based accounts of quantifier spreading. Like Brooks and Sekerina (2006, Experiment 3) and the work on L2 comprehension referred to in the introduction, we found spreading errors in adults. Most crucially, the contrast between the first two experiments shows that spreading errors occur when semantic interpretation takes place under severe resource limitations. Can the existing resource-based accounts explain our findings in all their particulars? Although they can account for the general difference between the online experiments, on the one hand, and the offline task, on the other, they do not offer a self-evident explanation why branching should lead to substantially higher proportion of errors than extra objects. Moreover, it remains completely unclear why in the [+*eo*, +*bl*]-condition extra-object and branching effects did not add up but gave rise to an under-additive interaction. We do not see how any existing theory can account for the obtained pattern of results and will therefore briefly outline a different explanation.

Like many others we assume that comprehenders favor symmetrical interpretations of *every*/*each* (see, e.g. Sekerina et al. this volume, and the references therein). We think that this preference results from comprehenders' automatic computation of a default model which is the cognitively simplest state of affairs making the sentence true. The intended notion of a default model is closely related to Johnson-Laird's notion of a mental model (see, e.g. Johnson-Laird 2010) but can also be related to the notion of a minimal model from logic programming and closed-world reasoning (Stenning and van Lambalgen 2008; Schulz 2014). The important point is that comprehenders are assumed to construct a maximally simple mental model for any sentence given. Let us consider the experimental conditions once more. When interpreting the sentence *every boy loves exactly one girl* following a picture with four boys, comprehenders will first restrict *every* to a set of four boys. It is important to emphasize that later on, when trying to accomodate the default model, no changes are allowed concerning the size of the restrictor set. This assumption is crucial because otherwise the standard (non-spreading) truth conditions of the universal quantifier would be completely flawed: it's meaning would become equivalent to that of *some*. Due to the distributivity of *every* the comprehension system will construct a minimal situation/event for each of the boys in this set. Essentially, comprehenders thus construct a one-to-one mapping corresponding to the [−*eo*, −*bl*]-condition which can be easily matched to an [−*eo*, −*bl*] picture. This explains the extremely low proportion of errors in this condition.

In the [+*eo*, −*bl*]-condition the default model cannot simply be matched to the diagram. However, the default model is a proper part of the [+*eo*, −*bl*] model. All comprehenders need to do is notice that the default model can be properly embedded in the model shown in the picture. The extra-object error occurs in case comprehenders stick to the minimality requirements of their default model and fail to acknowledge that their representation can be safely expanded without affecting truth. In a sense, this explanation is very similar to the one proposed by É. Kiss and Zétényi (2017) who take the minimal situation consistent with the quantified sentence to be the typical learning situation for universal quantifiers. Our explanation can also account for some of the findings reported in the eyetracking study by Sekerina et al. this volume. It is fully expected under our account that those trials in which an extra-object error occurred took on average less time than trials that received logically correct responses. We would like to suggest that the observed difference in judgment RT is due to the application of the proposed matching procedure required in the correct trials. What remains somewhat surprising, though, is why more attention during listening—as indicated by the proportion of looks—was allocated to the extra objects in the incorrectly answered trials than in the correct trials. Perhaps the observed difference can be related to the effort and time put into the first step, that is, simple matching of the initial semantic representation to the picture resulting in erroneous responses.

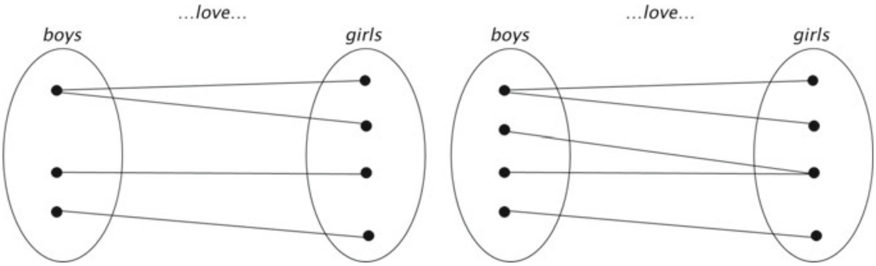
As for the branching conditions, 'model mapping' becomes impossible (as long as we prohibit adding additional elements to the restrictor set). Let us first consider the $[-eo, +bl]$ condition. Here comprehenders have to deal with a relation between the two sets that substantially differs from the one-to-one mapping of their default model. Unable to verify the sentence via a simple matching procedure, they have to check whether each of the boys is connected to exactly one girl. To put it differently, they have to verify the sentence on the basis of its standard (non-spreading) truth conditions instead of being able to rely on the representation they automatically came up with. Without further stipulation, this also explains why the effects of extra objects and branching lines did not add up in proportion of errors and RTs of the first experiment and the RTs of the second experiment: an extra object does not hurt the verification process if this process is based on truth conditions. Assuming that verification via truth conditions is more difficult than the outlined matching procedure, only the former should be affected by resource limitations. An explanation like this clearly accounts for the differences between the ITVJ experiments (Experiments 2 and 3) and the ordinary picture verification task (Experiment 1).

A possible concern might be that the just outlined approach is limited to the quantifiers *every* and *exactly one*. With these particular quantifiers it yields a model that satisfies the truth conditions of the spreading interpretation in (7-a) and (7-b) repeated for convenience from (2-a) and (2-b).

(7) a. $\forall x(boy(x) \rightarrow \exists y^{=1}(girl(y) \wedge love(x, y)))$
b. $\forall y(girl(y) \rightarrow \exists x^{=1}(boy(x) \wedge love(x, y)))$

What happens if *every* is combined with other quantifiers like *at least one* ($\exists^{\geq 1}$) as in (8-a)?

(8) a. Every boy loves at least one girl.
b. $\forall x(boy(x) \rightarrow \exists^{\geq 1} y(girl(y) \wedge love(x, y)))$
c. $\forall y(girl(y) \rightarrow \exists^{\geq 1} x(boy(x) \wedge love(x, y)))$



(a) default model (b) +bl model

Fig. 3 Two models for *every boy loves at least one girl*

In this case, our approach makes interesting predictions that differ from those of an account only relying on the combined truth conditions in (8-b) and (8-c). The quantitative relations in the default model of the linear scope interpretation will correspond to those illustrated in Fig. 3a. This is because *at least one* is only fully felicitous if there is at least one boy who loves more than one girl. Otherwise, *one* or even *exactly one* would have been more informative alternative expressions. Analogously to what we assumed for *exactly one*, the default model is constructed in a way that there is no girl that is loved by two boys. Consequently, the model in Fig. 3b should, according to our approach, lead to branching errors, and these errors should show up particularly under testing conditions when processing resources are severely limited. Considering the mere truth conditions in (8-b) and the spreading condition in (8-c), however, both models should be accepted under a spreading interpretation. We have the intuition that a picture like the one in Fig. 3b is in fact harder to verify than Fig. 3a, but the empirical investigation of this issue has to be left for further research.

Can this line of thinking also account for the effects of plausible dissent from Crain et al. (1996)? Answering a question which violates pragmatic felicity conditions can arguably impose a second task, too: children might have the feeling that they have to answer two questions at the same time, namely *is every boy riding an elephant* and *why did they ask this question in the first place*? So, in our view, the condition of plausible dissent can be fully captured by a resource-based account like the one just sketched. Moreover, our proposal applies equally well to L1 and L2 acquisition since children acquiring their L1 as well as adult L2 learners face resource limitations compared to adult monolinguals (see, e.g. Clahsen and Felser 2006, for an extensive review).

To conclude, we have presented evidence that the extra object fallacy is not restricted to L1 and L2 acquisition but can be elicited in adult L1 comprehension, too. We reported on a second kind of spreading error in adults, namely branching errors, that have not been investigated yet. Branching errors were even more frequent than the classic extra-object errors. Our findings support the view that, after all, universal quantification in adults is not that different than in children. Rather, our results are most compatible with the view that the apparent differences must be attributed to more general limitations of cognitive resources.

Acknowledgements This research was funded by the German Science Foundation (DFG) within project B1 of the Collaborative Research Centre 833 *The Construction of Meaning* (O. Bott and F. Schlotterbeck) and within project *Composition in Context* (CiC) of the Priority Program *XPrag.de* (O. Bott). The authors are indebted to Detlef Axmann, Janina Radó, Wolfgang Sternefeld, Jakub Szymanik, and Rolf Ulrich for valuable comments and suggestions. The authors would also like to thank Jesse Snedeker, Katalin É. Kiss, and five anonymous reviewers for their comments on earlier drafts of this paper. The usual disclaimer applies.

References

Barr, D., R. Levy, C. Scheepers, and H. Tily. 2013. Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal. *Journal of Memory and Language* 68: 255–278.

Barwise, J., and R. Cooper. 1981. Generalized Quantifiers and Natural Language. *Linguistics and Philosophy* 4: 159–219.

Berent, G., R. Kelly, and T. Schueler-Choukairi. 2009. Economy in the Acquisition of English Universal Quantifier Sentences: The Interpretation of Deaf and Hearing Students and Language Learners at the College Level. *Applied Psycholinguistics* 30: 251–290.

Bott, O., U. Klein, and F. Schlotterbeck 2013. Witness Sets, Polarity Reversal and the Processing of Quantified Sentences. In *Proceedings of the Amsterdam Colloquium 2013*, ed. M. Aloni, M. Franke, and F. Roelofsen, 59–66.

Bott, O., and J. Radó. 2007. Quantifying Quantifier Scope: A Cross-methodological Comparison. In *Roots: Linguistics in Search of its Evidential Base*, ed. S. Featherston and W. Sternefeld, 53–74. Berlin/New York: de Gruyter.

Bott, O., and F. Schlotterbeck. 2012. Incremental Truth Value Judgments. In *Empirical Approaches to Linguistic Theory*, ed. B. Stolterfoht and S. Featherston, 3–28. Berlin: Mouton de Gruyter.

Brooks, P., and M. Braine. 1996. What Do Children Know About the Universal Quantifiers 'all' and 'Each'? *Cognition* 60: 235–268.

Brooks, P., and I. Sekerina. 2006. Shortcuts to Quantifier Interpretation in Children and Adults. *Language Acquisition* 13: 177–206.

Clahsen, H., and C. Felser. 2006. How Native-like is Non-native Language Processing? *Trends in Cognitive Sciences* 10: 564–570.

Crain, S., R. Thornton, C. Boster, L. Cornway, D. Lillo-Martin, and E. Woodams. 1996. Quantification Without Qualification. *Language Acquisition* 5: 83–153.

Csibra, G., and G. Gergely. 2009. Natural Pedagogy. *Trends in Cognitive Sciences* 13: 148–153.

DelliCarpini, M. 2003. Developmental Stages in the Semantic Acquisition of Quantification by Adult L2 Speakers of English: A Pilot Study. In *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA2002): L2 Links*, ed. J. Liceras, H. Zoble, and H. Goodluck, 55–63. Sommerville, MA: Cascadilla.

Drozd, K. 2001. Children's Weak Interpretations of Universally Quantified Questions. In *Language Acquisition and Conceptual Development*, ed. M. Bowerman and S. Levinson, 340–376. Cambridge University Press.

Drozd, K., and E. van Loosbroek. 2006. The Effect of Context on Children's Interpretations of Universally Quantified Sentences. In *Semantics Meets Acquisition*, ed. V. van Geenhoven, 115–140. Dordrecht: Kluwer.

Einstein, G.O., and R.R. Hunt. 1980. Levels of Processing and Information: Additive Effects of Individual Item and Relational Processing. *Journal of Experimental Psychology: Human Learning and Memory* 6: 588–598.

É. Kiss, Katalin, and T. Zétényi. (2017). Quantifier Spreading: Children Misled by Ostensive Cues. *Glossa: A Journal of General Linguistics* 2 (1), 38: 1–20. http://doi.org/10.5334/gjgl.147.

Geurts, B. 2003. Quantifying Kids. *Language Acquisition* 11: 197–218.

Gouro, T., H. Norita, M. Nakajima, and K. Ariji 2001. Children's Interpretation of Universal Quantifier and Pragmatic Interference. In *Proceedings of TCP 2001*, ed. H. Shobo, 61–78.
Hunter, T., and J. Lidz. 2013. Conservativity and Learnability of Determiners. *Journal of Semantics* 30: 315–334.
Inhelder, B., and J. Piaget. 1959. *The Early Growth of Logic in the Child*. London: Routledge.
Jäger, T. 2008. Categorical Data Analysis: Away from Anovas (Transformation or Not) and Towards Logit Mixed Models. *Journal of Memory and Language* 59: 434–446.
Johnson-Laird, P. 2010. Mental Models and Human Reasoning. *PNAS* 107: 18243–18250.
Mayo, N., M. Corley, and F. Keller. 2006. Webexp2: Experimenter's Manual.
Meroni, L. 2002. Children's and Adults' Interpretation of the Universal Quantifier: Grammatical Non-adult Principles. Unpublished Manuscript: University of Maryland.
Meroni, L., A. Gualmini, and S. Crain. 2006. Everybody Knows. In *Semantics in Acquisition*, ed. V. van Geenhoven, 89–114. Dordrecht: Springer.
Philip, W. 1995. *Event Quantification in the Acquisition of Universal Quantification*. Ann Arbor, Michigan: UMI.
Philip, W. 2011. Acquiring Knowledge of Universal Quantification. In *Handbook of Generative Approaches to Language Acquisition*, ed. J. de Villiers and T. Roeper, 351–394. Dordrecht: Springer.
Roeper, T., and E. Mattei. 1974. On the Acquisition of *All* and *Some*. *Stanford Papers in Linguistics* 9: 63–74.
Roeper, T., U. Strauss, and B. Pearson. 2004. *The Acquisition Path of Quantifiers: Two Kinds of Spreading*. Amherst: Technical Report University of Massachussetts.
Schulz, K. 2014. Minimal Models Versus Logic Programming: The Case of Counterfactual Conditionals. *Journal of Applied Non-Classical Logics* 24: 153–168.
Sekerina, I., and A. Sauermann. 2015. Visual Attention and Quantifier-spreading in Heritage Russian Bilinguals. *Second Language Research* 31: 75–104.
Sekerina, I.A., P.J. Brooks, L. Campanelli, and A.M. Schwartz. (this volume). Quantifier Spreading in School-age Children: An Eye-tracking Study.
Stenning, K., and M. van Lambalgen. 2008. *Human Reasoning and Cognitive Science*. Cambridge, MA: MIT Press.
Street, J.A., and E. Dąbrowska. 2010. More Individual Differences in Language Attainment: How Much Do Adult Native Speakers of English Know About Passives and Quantifiers? *Lingua* 120: 2080–2094.
Sugisaki, K., and K. Isobe. 2001. Quantification Without Qualification Without Plausible Dissent. In *Proceedings of SULA*, ed. J.-Y. Kim and A. Werle , 97–100.
Tanenhaus, M., M. Spivey-Knowlton, K. Eberhard, and J. Sedivy. 1995. Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science* 268: 1632–1634.
Wheeler, M.E., and A.M. Treisman. 2002. Binding in Short-term Visual Memory. *Journal of Experimental Psychology: General* 131: 48–64.

Index

K.É. Kiss and T. Zétényi (eds.), *Linguistic and Cognitive Aspects of Quantification*, Studies in Theoretical Psycholinguistics 47, https://doi.org/10.1007/978-3-319-91566-1