University of Groningen

Strategies for scope taking

Brasoveanu, Adrian; Dotlacil, Jakub

Published in:
Natural Language Semantics

DOI:
[10.1007/s11050-014-9109-1](https://doi.org/10.1007/s11050-014-9109-1)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Brasoveanu, A., & Dotlacil, J. (2015). Strategies for scope taking. *Natural Language Semantics*, 23(1), 1-19. <https://doi.org/10.1007/s11050-014-9109-1>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Strategies for scope taking

Adrian Brasoveanu · Jakub Dotlačil

Published online: 30 October 2014
© Springer Science+Business Media Dordrecht 2014

Abstract This squib reports the results of two experimental studies, a binary choice and a self-paced reading study, that provide strong support for the hypothesis in Tunstall (PhD thesis, 1998) that the distinct scopal properties of *each* and *every* are at least to some extent the consequence of an event-differentiation requirement contributed by *each* (Tunstall herself conducted an experiment that did not reveal the predicted effect). However, we also show that the emerging picture is more complex than Tunstall suggests: English speakers seem to fall into at least three groups regarding the scopal properties of *each* and *every*.

Keywords Event differentiation · Quantifier scope · Experimental semantics · Gaussian mixture models

1 Introduction

One of the important early contributions in the formal semantics of quantification is that scopal behavior is not a uniform phenomenon. It is nowadays commonly accepted that the analysis of scope must distinguish several classes of expressions, including indefinites, distributive quantifiers, negative quantifiers, etc. Ioup (1975) was one of the first to argue that quantifiers are hierarchically ordered with respect to their wide-scope taking abilities. This is exemplified in (1) below, where quantifiers to the left, i.e., higher on the scale, take wide scope more readily than quantifiers to the right. For example, *every* is a better wide-scope taker than *all*, and they are both worse wide-

A. Brasoveanu (✉)
UC Santa Cruz, Santa Cruz, CA, USA
e-mail: abrsvn@gmail.com

J. Dotlačil
University of Groningen, Groningen, The Netherlands

scope takers than *each*, as shown in (2). We use $\gg$ to indicate both higher wide-scope preference (in (1)) and actual scope-taking behavior (in (2)).

- (1) *each* $\gg$ *every* $\gg$ *all* $\gg$ *most*
- (2) a. She knows a solution to all problems. $??all \gg a, a \gg all$
 b. She knows a solution to every problem. $every \gg a, a \gg every$
 c. She knows a solution to each problem. $each \gg a, ?a \gg each$

This squib studies the difference in scopal behavior between the first two elements in Ioup's hierarchy, namely *each* and *every*. This difference has been previously analyzed from a syntactic and lexical-semantic perspective in Kroch (1974), VanLehn (1978), Beghelli and Stowell (1997), and Tunstall (1998), among others. We focus here on the hypothesis put forth in Tunstall (1998) that the distinct scopal properties of *each* and *every* are due to an event-differentiation requirement contributed by *each*. To the best of our knowledge, this hypothesis was tested only once (see Tunstall 1998, Experiment 3, Chapter 5), and the experimental study failed to reveal the predicted effect.

We partially replicate Tunstall's Experiment 3 (a binary choice task; we omit the depictive predicates and focus only on resultatives), and also report the results of a new self-paced reading experiment. These experiments provide strong support for the event-differentiation account of the scopal behavior of universal distributive quantifiers. However, we also show that the emerging picture is more complex than the one painted in Tunstall (1998): English speakers seem to fall into at least three groups regarding the scopal properties of *each* and *every*.

The squib is structured as follows. Section 2 discusses Tunstall's account of the two distributive quantifiers. Section 3 presents the first experiment, which is the partial replication of Tunstall's Experiment 3, and Sect. 4 presents the self-paced reading experiment. Section 5 concludes.

2 The meaning and scopal behavior of *each* and *every*

Suppose I have a basket full of apples and say (3a) or (3b) below. Even though both sentences use distributive quantifiers that are commonly treated alike in formal semantics, there is a subtle difference in meaning between them. As Vendler (1962) observes, if I utter (3a), I do not care how you take the apples as long as you take them all. In contrast, (3b) suggests that I expect you to take the apples one by one. This is a somewhat odd expectation, hence the question marks in front of the sentence. But the oddness of (3b) disappears if we justify this expectation, e.g., by adding ...*and examine them in turn*, or if we use a different predicate that warrants the individual treatment of the apples, e.g., *weigh* in (4).

- (3) a. Take every one of these apples.
 b. ?? Take each of these apples.
- (4) Weigh each of these apples.

Based on these examples, Vendler (1962) concludes that the distributive quantifiers *every* and *each* differ in their emphasis: *every* stresses exhaustiveness, while *each*

“directs one’s attention to the individuals as they appear, in some succession or other, one by one” (Vendler 1962, p. 150). This contrast is further discussed in Beghelli and Stowell (1997) and Tunstall (1998). For example, the latter takes the contrast between (5a) and (5b) below to show that *each*, but not *every*, requires a separate subevent for each element in its restrictor—this is the reason for the incompatibility of *each* with the qualification *not separately* in (5b).¹

- (5) a. Jake photographed every student in the class, but not separately.
- b. # Jake photographed each student in the class, but not separately.

Tunstall (1998) conjectures that the Differentiation Condition in (6) is the reason for the infelicity of *each* in (5), since individual students do not appear in separate picture-taking subevents. This condition also requires the addressee of (3b) to not take all the apples in the basket at once.²

- (6) *The Differentiation Condition* (Tunstall 1998):

A sentence containing a quantified phrase headed by *each* can only be true of event structures which are totally distributive. Each individual object in the restrictor set of the quantified phrase must be associated with its own subevent, in which the predicate applies to that object, and which can be differentiated in some way from the other subevents. (p. 100)

How can events be differentiated from each other? The subevents in the previous examples differ with respect to their temporal trace, but differences in other dimensions can be utilized for this purpose too. For example, (7) below (also from Tunstall 1998) is most likely taken to mean that Carol mashed potatoes one at a time, but it is also possible to imagine that the differentiation of the mashing subevents is achieved by varying other event characteristics, e.g., their location.

- (7) Carol mashed each potato.

Another possibility is to vary event participants, and this is relevant for the resolution of quantifier scope. Consider the contrast between the two examples in (8) below. Sentence (8a) does not signal that different dyeing events occurred: this sentence could be felicitously used if all the shirts were dipped together in a vat of dye. In (8b), however, *each* requires separate subevents of dyeing to have taken place, and this information can be filled in by the hearer in various ways. For example, the hearer might conjecture that each shirt was dyed in a different location or at a different time.

- (8) a. A helper dyed every shirt.
- b. A helper dyed each shirt.

¹ Apart from the works mentioned above, the difference in meaning between *each* and *every* has received relatively little attention in formal semantics, but it is commonly discussed in English ‘usage and style’ guides. For instance, Allen (2008) notes that “*each* regards the people or things concerned separately, whereas *every* regards them collectively” (p. 239).

² Beth Levin (p.c.) remarks that event differentiation seems to be related to the singulative number markers found in Welsh, Arabic, Nilo-Saharan languages, etc. We leave an investigation of this intriguing connection for a future occasion.

Probably the easiest way to accommodate the event differentiation requirement is to let the distributive quantifier *each shirt* take wide scope and distribute over the indefinite subject. The resulting reading—that each shirt is dyed by a (possibly) different helper—satisfies event differentiation by ensuring that different dyeing subevents have different agents (since different participants signal different events; cf. Carlson 1984; Landman 2000 among others). That is, the inverse scope reading of (8b) satisfies the differentiation requirement contributed by *each*, and therefore is more likely to be considered by English speakers. In contrast, there is no comparable pressure to consider the inverse scope reading with *every* in (8a).

As Tunstall (1998) notes, the situation changes if a resultative adjective is added, e.g., the adjective *blue* in (9). Resultatives introduce a new state into the meaning representation of the sentence and specify that that state is caused by the matrix event (see Hovav and Levin 1998; Kratzer 2005 among others for more discussion). In the case at hand, *blue* introduces a new state, that of being blue, which is the result of dyeing. Since the result state of *being blue* is an inherently distributive property of individuals, every shirt is now associated with a separate state. Therefore, the resultative satisfies the differentiation condition.³

- (9) a. A helper dyed every shirt blue.
- b. A helper dyed each shirt blue.

But shouldn't telic predicates like *dyed every/each shirt* in (8) automatically include result states and therefore satisfy the differentiation condition on their own?⁴ One possibility is that telic predicates do not actually contribute such states on their own, as is the case in mereological approaches to telicity like the one proposed in Krifka (1989). Alternatively, we can take the result state to be implicitly present (Dowty 1979; Pustejovsky 1995), but not sufficient to satisfy differentiation precisely because of its implicit nature. That is, the hearer still needs to do inferential work to conclude that the event of dyeing has an inherently distributive result state as one of its components that can satisfy the differentiation requirement contributed by *each*. This is different in (9b), which specifically mentions an inherently distributive result state, so no inferential work is required from the interpreter.

Because of the difference between (8) and (9), *each* is licensed in (9b) without any extra work: there is no need to look for a dimension with respect to which the dyeing subevents can be differentiated. In particular, the quantifier *each shirt* in (9b) does not need to take inverse scope over the indefinite. Tunstall (1998) therefore predicts that *each* in (8) should exhibit a higher preference for inverse (i.e., wide) scope than *every* to satisfy the differentiation condition, while the addition of the resultative in (9) should neutralize this higher preference for inverse scope exhibited by *each*.

As we have seen, the differentiation condition can shed light on one (albeit small) part of Ioup's hierarchy in (1) above. Unfortunately, the predictions made by the differentiation condition are hard to assess in traditional ways (via introspective judgments)

³ We assume that for the purpose of event differentiation, the distinction between states and events does not play any role.

⁴ We are indebted to an anonymous reviewer for emphasizing this point and for related discussion.

since the inverse-scope interpretation of *each* is never required, only *more or less likely*, and such distinctions are very subtle and fluid.⁵ But this is precisely the type of question that can be experimentally investigated. The differentiation-based account predicts that the universal quantifier in (8b) is more likely to be interpreted as taking inverse scope compared to the minimally different sentence in (9b). So in an experimental task, we expect participants to select the inverse scope reading more frequently for (8b) than for (8a). In addition, we expect this increase in frequency to disappear when a secondary distributive predication is added, i.e., *each* in (9b) should display the same (lack of) preference for inverse scope as *every* in (9a).

Tunstall (1998) designed an experiment to test precisely this prediction (Experiment 3 in Chapter 5). Unfortunately, Tunstall did not find a statistically significant interaction of quantifier type and presence or absence of secondary predication. That is, she found no significant difference between (8a) and (8b) on the one hand, and (9a) and (9b) on the other hand. Tunstall was able to detect an effect only post hoc, after considering only the first half of her study, and she conjectured that fatigue during the experimental task obscured all differences between experimental conditions in the second half of the experiment. But since these findings are based on data selected post hoc, the results cannot be taken to provide sufficiently strong evidence for the differentiation-based account of the scopal behavior of *each* and *every*. The main goal of the present squib is to supply this missing evidence.

3 Experiment 1: Binary choice between paraphrases

Our Experiment 1 was a partial replication of Tunstall's study. The replication is partial because Tunstall used secondary predicates which were either resultatives (*A helper dyed each/every shirt blue*) or depictives (*A helper dyed each/every shirt damp*), while we chose to focus exclusively on resultatives. Tunstall mentions that speakers disliked the depictive predicates, preferring prenominal adjectives instead, and this might have influenced the overall results: participants might not have been sensitive to the fine-grained distinction between surface and inverse scope if they found the use of secondary predicates marginal or ungrammatical in many cases. For this reason, we considered only resultative secondary predicates.

3.1 Experimental method

3.1.1 Materials and procedure

The four conditions that each experimental item was passed through are exemplified in (10) below. The experiment had a 2 (*each* vs. *every*) $\times$ 2 (presence vs. absence of resultative predicate) design:

⁵ Quantifier scope judgments are subtle to begin with. But introspectively comparing preferences for one scope or another when secondary predicates are present or absent is basically impossible to do in a reliable manner.

- (10) a. NO RESULTATIVE PREDICATE:
- i. A helper dyed every shirt. (EVERY)
 - ii. A helper dyed each shirt. (EACH)
- b. RESULTATIVE PREDICATE:
- i. A helper dyed every shirt blue. (EVERY)
 - ii. A helper dyed each shirt blue. (EACH)

After reading a stimulus sentence, i.e., an item in one of the four experimental conditions, the participants had to choose one of two paraphrases that disambiguated towards the surface or inverse scope of the universal quantifier. To illustrate, the paraphrases for item (10b-i) above are provided in (11) below: (11a) disambiguates towards the inverse scope reading while (11b) disambiguates towards the surface scope reading. The parenthesized text appeared only when the resultative predicate was present in the stimulus sentence and the round brackets themselves were omitted. The order in which the surface and inverse scope paraphrases appeared was randomized for every item and every participant.

- (11) A helper dyed every shirt blue.
- a. Each shirt was dyed by a possibly different helper (until it was the color blue).
 - b. All the shirts were dyed by the same helper (until they were the color blue).

There were 18 experimental items (similar to the experimental items of Tunstall 1998, with depictive predicates excluded) and 30 fillers, 6 of which were control fillers used to assess whether participants were paying attention to the experimental task rather than selecting answers arbitrarily. No filler included the quantifiers *every* or *each* or resultative predicates. The 6 control fillers consisted of very simple, unambiguous sentences. An example is provided in (12) below; the correct paraphrase is (12a).

- (12) Only one boy enjoyed the show on the beach.
- a. Nobody but one boy enjoyed the show on the beach.
 - b. Nobody enjoyed the show on the beach.

Four lists were created for the 18 items, and each list included a different condition for each experimental item (the items were rotated through the 4 conditions). Every experimental item appeared exactly once in each list. The participants were cycled through these four lists. The order of the 48 stimuli (18 experimental items + 30 fillers) was randomized for each participant.

3.1.2 Participants

All 43 participants were UC Santa Cruz undergraduate students who participated in the experiment for course credit or extra-credit. All participants were native speakers

Fig. 1 Exp. 1 (binary choice): mean probabilities of inverse scope and corresponding SEs

of English. The experiment was conducted using an installation of the Ibex platform⁶ locally hosted on the UC Santa Cruz servers. Only 2 participants had more than one mistake in the control fillers, and only one participant failed more than 3 out of the 6 control fillers. That participant was excluded from all subsequent analyses.

3.2 Results and discussion

3.2.1 Main results

Mean probabilities of inverse scope and (normal approximations of) the corresponding standard errors for the 4 experimental conditions are plotted in Fig. 1.

Given that the experimental data is binary categorical, we analyzed it with mixed-effects logistic regression models (using the `lme4` R package⁷), with SCOPE (SURFACE vs. INVERSE) as the dependent variable ('success' level: INVERSE SCOPE). We considered the full fixed-effects structure, i.e.: (i) QUANTIFIER (EACH vs. EVERY, the latter being the reference level), (ii) RESULTATIVE (PRESENCE vs. ABSENCE of resultative predicates; the latter being the reference level), and (iii) the interaction of QUANTIFIER and RESULTATIVE.

We also included the maximal random-effect structure that converged (following Barr et al. 2013) for both subjects and items, i.e., random intercepts and QUANTIFIER

⁶ See <http://code.google.com/p/webspr/>.

⁷ See R Core Team (2014) and Bates et al. (2014).

random slopes for subjects and random intercepts only for items. But the model with the maximal (converging) random-effect structure did not have a significantly lower deviance (as assessed by a Likelihood Ratio Test: $\chi^2 = 1.28$, $df = 2$, $p = 0.53$) relative to the model that had only random intercepts for subjects and items. We will therefore report the random-intercepts-only model. In addition, including only random intercepts rather than the full random-effect structure that converged will make it much more straightforward to enhance the model so that we can cluster participants according to their scope preferences (see Sect. 3.2.2 below).

We found that EACH increases the probability of inverse scope: $\beta = 0.7$, $p = 0.01$ ($SE = 0.28$). The 0.7 estimate for the main effect of EACH is on the logit scale. An approximate way of converting this logit-scale estimate to the more intuitive probability scale is to divide by 4 (see, for example, Gelman and Hill 2007). Thus, the probability of inverse scope for EACH is approximately 17 % higher than for EVERY. This can also be seen in Fig. 1: the probability of inverse scope for EACH (in dark-gray; or blue online) is higher than the probability of inverse scope for EVERY (in light-gray; or red online) whether a resultative predicate was present or not. No other effect was significant, i.e., neither the main effect for RESULTATIVE predicates nor the interaction of QUANTIFIER and RESULTATIVE.

To further study the interaction of quantifiers and the presence/absence of resultatives, we considered two data subsets: one including only the stimuli with EVERY, and one including only the stimuli with EACH. The former subset did not show any significant result even though numerically, the PRESENCE of resultative predicates increased the probability of surface scope for EVERY (as we can already see in Fig. 1). But the PRESENCE of resultatives had a significant effect in the *each* subset. This is suggestive evidence that resultative predicates decrease inverse scope but only in case of *each*, as Tunstall's differentiation hypothesis would predict.

However, the full dataset fails to provide evidence for this since the interaction of QUANTIFIER and RESULTATIVE was not significant. We conjectured that forced binary choice—an off-line task—is not a sensitive enough experimental method to detect this effect, so we ran a follow-up self-paced reading experiment to investigate the same issue. Self-paced reading (Just et al. 1982) is a real-time methodology that can detect early, incremental effects of scope, and which has been successfully used to investigate the processing of inverse scope (see Tunstall 1998 and Anderson 2004, among others).

3.2.2 Strategies for scope taking

Before discussing the self-paced reading study, we want to examine more closely the behavior of the participants in the binary choice task.

As the normal Q–Q plot in Fig. 2 shows, the subject random effects are clearly not normally distributed, which violates one of the assumptions of the mixed-effect logistic regression models implemented in `lme4` (Bates et al. 2014). The non-normality of the distribution is confirmed by the Shapiro-Wilk test: $W = 0.94$, $p = 0.04$, so we reject the null hypothesis that the subject random effects are normally distributed.

The plot in Fig. 2 actually gives us a lot more information than the Shapiro-Wilk test: the subjects seem to fall into three groups, which we have explicitly indicated in the plot. The central group really does seem to be normally distributed: it follows

Fig. 2 Exp. 1 (binary choice): Q–Q plot of subject random effects

the dotted Q–Q line closely. But in addition, we have two other groups at the extreme ends: the more numerous one at the negative end, which exhibits a strong bias for surface scope; and the less numerous one at the positive end, which exhibits a strong bias for inverse scope.

We can test if this impressionistic generalization holds under scrutiny by modeling the subject random effects as a mixture of three Gaussian distributions and checking whether different subjects are reliably classified as coming from one of the three Gaussian components. There is no off-the-shelf implementation that allows us to estimate such models, so at this point we switch to a Bayesian framework and estimate the exact same mixed-effects logistic regression model as above (full fixed-effects structure, intercept-only random effects for subjects and items) except that we take the distribution for the subject random effects to be a mixture of three Gaussians instead of a single Gaussian. All three Gaussian components have the same prior weight (the weights have a *Dirichlet*(1, 1, 1) prior distribution), i.e., before seeing any data, we assume that a subject could come from any of the three groups with equal probability.⁸

⁸ The prior for the mean of the first Gaussian component was a normal distribution with $\mu = 0$ and λ (precision) = 0.001 (where precision is the reciprocal of variance). This is a very low information prior on the standard logit scale. The means of the second and third groups were shifted from the first mean by

The mean posterior weights for the three groups, i.e., the estimated proportions of the three groups in the overall subject population after seeing the data, were as follows:⁹

- 21.2 % for the group with a surface scope bias; 95 % credible interval (CRI): 4.1 %, 42.6 %;
- 69.4 % for the central group with no clear scope bias; 95 % CRI: 41.4 %, 90.2 %;
- 9.4 % for the group with an inverse scope bias; 95 % CRI: 0.9 %, 30.2 %.¹⁰

Further examination of the data showed that the subjects in groups 1 and 3 (the groups with a clear bias towards surface or inverse scope) consistently made the same choice, i.e., they universally selected either the surface or the inverse scope reading, depending on their bias. Only the central group of subjects showed varying behavior. These results support informal observations that English speakers exhibit dialectal variation with respect to scope interpretation, either always assigning surface scope, always assigning inverse scope, or showing a varying behavior (Carden 1970; Gil 1982; Reinhart 2006). We also see that, unsurprisingly, the surface-scope dialect is more frequent than the inverse-scope dialect and the mixed dialect is the most common.

The mixture model did not change the previous results regarding the fixed effects: only the main effect of QUANTIFIER was significant (with basically the same magnitude), while the main effect of RESULTATIVE and the QUANTIFIER $\times$ RESULTATIVE interaction remained non-significant. The self-paced reading follow-up mentioned above, and to which we now turn, was meant to address this problem.

4 Experiment 2: Self-paced reading

It is often noted that forcing an inverse scope reading leads to processing difficulties (Kurtzman and MacDonald 1993; Tunstall 1998; Anderson 2004; Pykkänen and McElree 2006; Hackl et al. 2012, among others). Given this observation, we expect to see more processing difficulties associated with scope in (13a) than in (13b) because *each* is more likely to take inverse scope in the former case (due to the event-differentiation requirement, Sect. 2). This prediction is made regardless of the source of the processing difficulties associated with inverse scope. That is, whether we have a syntactic explanation for the difficulty of inverse scope (Tunstall 1998; Anderson 2004; Pykkänen and McElree 2006; Reinhart 2006, among others) or a semantic/pragmatic explanation (Fodor 1982; see also Crain and Steedman 1985; Altmann and Steedman 1988), we still make the prediction that (13a) should be more difficult than (13b).

Footnote 8 continued

a δ_1 and a δ_2 difference, respectively. The priors for the δ_1 and δ_2 differences were two folded (positive) normal distributions with $\mu = 0$ and λ (precision) = 0.1. The priors for the standard deviations of the three Gaussian components were (continuous) uniform distributions: *Unif*(0, 0.75) for the extreme groups 1 and 3 and *Unif*(0, 1.5) for the central group 2.

⁹ We used JAGS (Plummer 2013) to estimate the posterior distributions of the model parameters. Details of the MCMC estimation: 3 chains, 500,000 iterations per chain, 100,000 burn-in, and 200 thinning.

¹⁰ Most subjects were consistently classified in one of the three groups, indicating that a Gaussian mixture with three components (rather than two or four, for example) as the subject random effect distribution was the correct assumption.

- (13) a. A helper dyed each shirt without thinking about it.
 b. A helper dyed each shirt **blue** without thinking about it.

Of course, the processing difficulty associated with inverse scope cannot be directly observed by simply comparing (13a) and (13b) because the two sentences also differ with respect to the presence or absence of the resultative. In particular, although the presence of the resultative predicate in (13b) might alleviate the processing difficulties associated with the inverse scope of *each*, it might contribute processing difficulties of its own that could cancel or even reverse the alleviation effect. It is therefore crucial to add two control conditions and compare (13) with (14) below, where *every* is substituted for *each*.

- (14) a. A helper dyed **every** shirt without thinking about it.
 b. A helper dyed **every** shirt blue without thinking about it.

The universal quantifier *every* in (14) is not subject to the differentiation condition and consequently, its inverse scope reading should not be influenced by the resultative predicate (as per the argument in Sect. 2 above). Thus, the difference (if any) in processing difficulty between (14a) and (14b) must arise independently of the inverse scope processing cost.

Thus, the differentiation condition that is contributed by *each* but not *every*, coupled with the assumption that inverse scope is associated with processing costs, predicts that we should see a *relative* decrease in processing cost in (13b) compared to (13a). This decrease is relative to the difference in processing cost (if any) between (14b) and (14a).

We tested this prediction using a self-paced reading task (Just et al. 1982) in which the behavioral correlate of processing cost is reading time. We therefore expect the predicted decrease in processing cost to manifest itself as faster reading times.

4.1 Experimental method

4.1.1 Materials and procedure

The experimental items were passed through four conditions, exemplified in (15) below:

- (15) a. NO RESULTATIVE PREDICATE:
 i. A helper dyed every shirt without thinking about it. (EVERY)
 ii. A helper dyed each shirt without thinking about it. (EACH)
 b. RESULTATIVE PREDICATE:
 i. A helper dyed every shirt blue without thinking about it. (EVERY)
 ii. A helper dyed each shirt blue without thinking about it. (EACH)

The experimental items were identical to the items used in Experiment 1 except that a spillover region consisting of at least three words was added after the universal

quantifier/secondary predicate. Adding this region is necessary if we want to detect the processing cost of inverse scope: this cost is often observed in self-paced reading studies (as well as in other types of real-time studies, e.g., eye-tracking) only after the quantifier phrase is read (see Tunstall 1998 and Hackl et al. 2012, among others). In addition, directly measuring reading times on quantifier phrases when the resultatives are absent, and on the resultative predicates themselves when they are present, is bound to be misleading: universally quantified phrases and resultative predicates differ in many respects (frequency, length, syntactic category, etc.) and these orthogonal differences are possible confounds for our experimental manipulation.

There were 4 practice items, 18 experimental items, and 86 fillers. Three fillers included the quantifier *each* and one filler included *every*, but unlike in the experimental items, the quantifiers in the fillers appeared in subject position. Three fillers included resultative predicates but lacked any quantifiers. In each trial, the participants first saw a screen in which all the words were hidden (dashes were displayed where the words would appear) and the participants had to press the SPACE bar to reveal the next word (the preceding word was automatically hidden). In all, 32 stimuli were followed by a yes/no comprehension question, displayed on a new screen.

4.1.2 Participants

In this experiment, 44 UC Santa Cruz undergraduate students participated for course credit or extra-credit. All participants were native speakers of English. The experiment was conducted using an installation of the Ibex platform locally hosted on the UC Santa Cruz servers.

On average, participants failed to answer 8 % of the comprehension questions correctly. Three outlier participants were excluded from the analysis because they failed to answer more than 25 % of the comprehension questions correctly, and two participants were excluded because the means of their log-transformed reading times (henceforth log RTs) were more than two standard deviations away from the grand mean.¹¹ The final number of participants included in the analysis was 39.

4.2 Results

Following Trueswell et al. (1994), we factored out the influence of word length and position by running a linear mixed-effects regression with log RTs as the dependent variable. The regression had intercept-only random effects for subjects and two fixed effects: word length (in characters) and word position in the sentence. The resulting residualized log RTs were used for all subsequent analyses.

The analysis focused on the four words following the resultative (if the resultative was present; otherwise, the same words would follow the quantifier phrase). These words, underlined in the example below, were identical across all four conditions.

- (16) A helper dyed every/each shirt (blue) without thinking about it.

¹¹ We log-transformed the data to mitigate the right-skewness characteristic of reaction time data.

The processing cost of inverse scope and the alleviation contributed by the resultative predicate in the case of *each* should be visible in this underlined region. The mean residualized log RTs and corresponding standard errors for the four words are plotted in Fig. 3.

Each word was analyzed using a mixed-effects linear regression model with the residualized log RTs as the dependent variable, the full fixed-effects structure, and the full random-effects structure for subjects and items that converged. That is, the fixed effects were the same as in Experiment 1: QUANTIFIER (EACH vs. EVERY, the latter being the reference level), RESULTATIVE (PRESENCE vs. ABSENCE of resultative predicates; the latter being the reference level), and their interaction.

Only the third word (the word *about* in Fig. 3) showed a significant effect. On that word, EACH led to a significant slowdown in reading times: $\beta = 0.07$, $p < 0.05$ (SE = 0.035, $t = 2.04$, Wald 95 % Confidence Interval: 0.003, 0.139).¹²

More importantly, the same word shows a significant interaction of QUANTIFIER and RESULTATIVE: $\beta = -0.12$, $p < 0.01$ (SE = 0.04, $t = -2.99$, Wald 95 % Confidence Interval: -0.19, -0.04). The interaction is negative: that is, the presence of the resultative predicate leads to lower reading times, but *only* when the quantifier EACH was used. In the case of EVERY, the presence of the resultative predicate actually leads to higher reading times.

This is exactly what Tunstall's differentiation account predicts: the presence of the resultative predicate should alleviate the processing difficulties associated with inverse scope, but only for *each*. Since the resultative predicate satisfies the differentiation requirement contributed by *each*, we expect that when the resultative is present, *each* is less likely to undergo QR. This decreases processing cost, reflected in faster reading times. Since the scope of *every* is not driven by the differentiation requirement, resultatives have no comparable facilitating effect on that distributive quantifier.¹³

Finally, we note that unlike Experiment 1, Experiment 2 revealed no split of subjects into groups. In other words, the subject random effects did not violate the null assumption of being normally distributed (Shapiro-Wilk test for the third word: $W = 0.96$, $p = 0.14$).

Recall that in Experiment 1, the split of subjects into groups signaled that there were at least three subpopulations: one with a strong surface scope bias, one with a strong inverse scope bias, and finally, the largest one that had no clear bias and varied in their choice of surface vs. inverse scope. The fact that we do not see similar groupings among subjects in Experiment 2 might signal that the subpopulation effect regarding the interpretation of surface/inverse scope appears only when people explicitly introspect their judgments.

¹² A numerically negative interaction was present in the first word after the quantifier/resultative (the word *without* in Fig. 3) as well, but it failed to reach significance.

¹³ The observed interaction between quantifiers and resultatives might have an additional source. When the resultative is absent in the EACH condition, inverse scope is more likely, but not obligatory. The processor might leave scope intact and rather infer that the subevents are distinct from each other along some temporally or spatially specified dimension. In that case, the processor has to attribute particular properties to the event contributed by the main verb, and this extra inferential work should incur an additional cost. It is conceivable that this cost would also increase reading times, even though we are not aware of a study showing that self-paced reading can detect such effects.

Fig. 3 Exp. 2 (self-paced reading): mean residualized log RTs and corresponding SEs

But other interpretations are also possible. Most importantly, in Experiment 2 we measured reading times, not the subjects' choice of interpretation. Therefore, any split into three groups could only emerge if the subjects would cluster around three mean reading times. But this kind of clustering would be hard to observe even if certain subjects consistently select only the surface scope reading and other subjects consistently select only the inverse scope reading. These groups of subjects would be hard to distinguish from each other and from the larger subpopulation with no clear scope bias because reading times vary fairly widely due to other factors (level of text comprehension, reading habits, etc.; see Rayner 1998, among others). This inter-subject variation for orthogonal reasons is likely to conceal any splits into subpopulations according to the three scope-taking strategies.¹⁴

5 Conclusion

Tunstall (1998) hypothesized that *each* is subject to a lexical differentiation condition that influences, among other things, its preference for widest scope. However, no unambiguous experimental evidence was provided to support this hypothesis. We presented two experimental studies, a binary-choice task and a self-paced reading task, that provide strong support for the differentiation hypothesis. The first (binary choice) experiment supports the differentiation condition when we compare the *every* and *each* data subsets. The second (self-paced reading) experiment provides strong, direct support for the hypothesis: the interaction of quantifier and resultative predicate is highly significant, and in the predicted direction. We also provided experimental evidence for informal observations made in the previous literature that the population of English native speakers can be grouped into at least three subpopulations depending on which particular scope-taking strategy they favor: a strong bias for surface scope, a strong bias for inverse scope, or no clear bias one way or the other.

While we argued that the differentiation condition plays an important role in accounting for the scopal behavior of distributive quantifiers, we do not want to commit ourselves to the conclusion that event differentiation is the *only* cause for the distinct scopal properties exhibited by *each* and *every*. That other factors might play a role is suggested by the results of Experiment 1, which revealed only a main effect for *each*. That is, regardless of the presence or absence of resultatives, *every* remained a worse inverse scope taker than *each*. The main effect for *each* might be a consequence of the experimental setup, e.g., the *each* condition without resultatives might prime inverse scope for *each* even when this is not necessary, but it might also be showing that the distinct scopal behavior of the two quantifiers has other causes in addition to event differentiation.

Acknowledgments We are indebted to a *Natural Language Semantics* anonymous reviewer for comments that greatly improved the initial version of this paper.

¹⁴ One might wonder whether the lack of subject clustering is due to the fact that we used residualized log RTs in our analysis, since most of the inter-subject variation is factored out at that stage of the analysis. We re-ran all analyses with non-residualized log RTs as the dependent variable and did not find any clear groupings of subjects for the 'raw' log RTs either.

A Appendix: Experimental items

A.1 Experiment 1 (binary choice)

- (1) A maid polished every/each mirror (spotless).
 - a. Each mirror was polished by a possibly different maid (until it was spotless).
 - b. All the mirrors were polished by the same maid until they were spotless.
- (2) A helper dyed every/each shirt (blue).
 - a. Each shirt was dyed by a possibly different helper (until it was the color blue).
 - b. All the shirts were dyed by the same helper (until they were the color blue).
- (3) A janitor dusted every/each bookcase (spotless).
 - a. Each bookcase was dusted by a possibly different janitor (until it was spotless).
 - b. All the bookcases were dusted by the same janitor (until they were spotless).
- (4) A gardener pruned every/each bush (short).
 - a. Each bush was pruned by a possibly different gardener (until it was short).
 - b. All the bushes were pruned by the same gardener (until they were short).
- (5) An aide tied every/each package (closed).
 - a. Each package was tied by a possibly different aide (so that it was closed).
 - b. All the packages were tied by the same aide (so that they were closed).
- (6) A servant ironed every/each sheet (smooth).
 - a. Each sheet was ironed by a possibly different servant (until it was smooth).
 - b. All the sheets were ironed by the same servant (until they were smooth).
- (7) A teenager painted every/each chair (yellow).
 - a. Each chair was painted by a possibly different teenager (so that it was yellow).
 - b. All the chairs were painted by the same teenager (so that they were yellow).
- (8) A man cut every/each board (short).
 - a. Each board was cut by a possibly different man (so that it was short).
 - b. All the boards were cut by the same man (so that they were short).
- (9) A child froze every/each banana (solid).
 - a. Each banana was frozen by a possibly different child (until it was solid).
 - b. All the bananas were frozen by the same child (until they were solid).
- (10) A boy sliced every/each carrot (thin).
 - a. Each carrot was sliced by a possibly different boy (into thin pieces).
 - b. All the carrots were sliced by the same boy (into thin pieces).

- (11) A woman stained every/each table (brown).
 - a. Each table was stained by a possibly different woman (until it was brown).
 - b. All the tables were stained by the same woman (until they were brown).
- (12) A clerk crushed every/each box (flat).
 - a. Each box was crushed by a possibly different clerk (so that it was flat).
 - b. All the boxes were crushed by the same clerk (so that they were flat).
- (13) A boy scout filed every/each knife (sharp).
 - a. Each knife was filed by a possibly different boy scout (until it was sharp).
 - b. All the knives were filed by the same boy scout (until they were sharp).
- (14) A cook mixed every/each batter (smooth).
 - a. Each batter was mixed by a possibly different cook (until it was smooth).
 - b. All the batters were mixed by the same cook (until they were smooth).
- (15) A girl squeezed every/each washcloth (dry).
 - a. Each washcloth was squeezed by a possibly different girl (until it was dry).
 - b. All the washcloths were squeezed by the same girl (until they were dry).
- (16) A volunteer folded every/each poster (flat).
 - a. Each poster was folded by a possibly different volunteer (so it was flat).
 - b. All the posters were folded by the same volunteer (so they were flat).
- (17) A girl toasted every/each muffin (crispy).
 - a. Each muffin was toasted by a possibly different girl (until it was crispy).
 - b. All the muffins were toasted by the same girl (until they were crispy).
- (18) A worker trimmed every/each wire (short).
 - a. Each wire was trimmed by a possibly different worker (so that it was short).
 - b. All the wires were trimmed by the same worker (so that they were short).

A.2 Experiment 2 (self-paced reading)

- (1) A maid polished every/each mirror (spotless) before attending to other tasks.
- (2) A helper dyed every/each shirt (blue) without thinking about it.
- (3) A janitor dusted every/each bookcase (spotless) on the last day of the month.
- (4) A gardener pruned every/each bush (short) in the popular park.
- (5) An aide tied every/each package (closed) while talking to the supervisor.
- (6) A servant ironed every/each sheet (smooth) before retiring to bed.
- (7) A teenager painted every/each chair (yellow) during spring break.
- (8) A man cut every/each board (short) during the workshop for carpenters.
- (9) A child froze every/each banana (solid) after the instructor left the room.
- (10) A boy sliced every/each carrot (thin) as required by the recipe.

- (11) A woman stained every/each table (brown) without making any mistakes.
- (12) A clerk crushed every/each box (flat) as the regulations specified.
- (13) A boy scout filed every/each knife (sharp) after the trip in the mountains.
- (14) A cook mixed every/each batter (smooth) before switching the oven on.
- (15) A girl squeezed every/each washcloth (dry) before starting the washer.
- (16) A volunteer folded every/each poster (flat) at the end of the final session.
- (17) A girl toasted every/each muffin (crispy) during the very long lunch break.
- (18) A worker trimmed every/each wire (short) to reduce the chance of fire.

References

- Allen, R. 2008. *Pocket Fowler's modern English usage*. Oxford: Oxford University Press.
- Altmann, G., and M. Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30: 191–238.
- Anderson, C. 2004. The structure and real-time comprehension of quantifier scope ambiguity. PhD thesis, Northwestern University.
- Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–278.
- Bates, D., M. Maechler, B. Bolker, and S. Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–6.
- Beghelli, F., and T. Stowell. 1997. Distributivity and negation. In *Ways of scope taking*, ed. A. Szabolcsi, 71–109. Dordrecht: Kluwer.
- Carden, G. 1970. A note on conflicting idiolects. *Linguistic Inquiry* 1(3): 281–290.
- Carlson, G. 1984. On the role of thematic roles in linguistic theory. *Linguistics* 22: 259–279.
- Crain, S., and M. Steedman. 1985. On not being led up the garden path: The use of context by the psychological syntax processor. In *Natural language parsing: Psychological, computational and theoretical perspectives*, ed. L. K. David Dowty and A. Zwicky, 320–358. Cambridge: Cambridge University Press.
- Dowty, D. R. 1979. *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*. Dordrecht: Reidel.
- Fodor, J. D. 1982. The mental representation of quantifiers. In *Processes, beliefs and questions*, ed. S. Peters and E. Saarinen, 129–164. Dordrecht: Reidel.
- Gelman, A., and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models (Analytical methods for social research)*. Cambridge: Cambridge University Press.
- Gil, D. 1982. Quantifier scope, linguistic variation, and natural language semantics. *Linguistics and Philosophy* 5(4): 421–472.
- Hackl, M., J. Koster-Hale, and J. Varvoutis. 2012. Quantification and ACD: Evidence from real-time sentence processing. *Journal of Semantics* 29(2): 1–62.
- Hovav, M. R., and B. Levin. 1998. Building verb meanings. In *The projection of arguments: Lexical and compositional factors*, ed. M. Butt and W. Geuder, 97–134. Stanford: CSLI.
- Ioup, G. 1975. Some universals for quantifier scope. In *Syntax and semantics*, ed. J. Kimball, vol. 4, 37–58. New York: Academic Press.
- Just, M. A., P. A. Carpenter, and J. D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111(2): 228–238.
- Kratzer, A. 2005. Building resultatives. In *Event arguments: Foundations and applications*, ed. C. Maienborn and A. Wöllstein-Leisten, 177–212. Tübingen: Max Niemeyer Verlag.
- Krifka, M. 1989. Nominal reference, temporal constitution, and quantification in event semantics. In *Semantics and contextual expressions*, ed. R. Bartsch, J. van Bentham, and P. van Emde Boas. Dordrecht: Foris.
- Kroch, A. 1974. The semantics of scope in English. PhD thesis, MIT.

- Kurtzman, H. S., and M. C. MacDonald. 1993. Resolution of quantifier scope ambiguities. *Cognition* 48: 243–279.
- Landman, F. 2000. *Events and plurality: The Jerusalem lectures*. Dordrecht: Kluwer.
- Plummer, M. 2013. rjags: Bayesian graphical models using MCMC. R package version 3–10.
- Pustejovsky, J. 1995. *The generative lexicon*. Cambridge, MA: MIT Press.
- Pylkkänen, L., and B. McElree. 2006. The syntax–semantic interface: On-line composition of sentence meaning. In *Handbook of psycholinguistics*, ed. M. Traxler and M. A. Gernsbacher, 537–577. New York: Elsevier.
- R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124: 372–422.
- Reinhart, T. 2006. *Interface strategies: Optimal and costly computations*. Cambridge, MA: MIT Press.
- Trueswell, J., M. Tanenhaus, and S. Garnsey. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33: 285–318.
- Tunstall, S. 1998. The interpretation of quantifiers: Semantics and processing. PhD thesis, University of Massachusetts, Amherst.
- VanLehn, K. A. 1978. Determining the scope of English quantifiers. Technical Report. Cambridge, MA: MIT.
- Vendler, Z. 1962. Each and every, any and all. *Mind* 71(282): 145–160.