

Processing and domain selection: Quantificational variability effects

Jesse A. Harris¹, Charles Clifton, Jr.², and Lyn Frazier¹

¹Department of Linguistics, University of Massachusetts Amherst, Amherst, MA, USA

²Department of Psychology, University of Massachusetts Amherst, Amherst, MA, USA

Three studies investigated how readers interpret sentences with variable quantificational domains, for example, *The army was mostly in the capital*, where *mostly* may quantify over individuals or parts (*Most of the army was in the capital*) or over times (*The army was in the capital most of the time*). It is proposed that a general conceptual economy principle, No Extra Times, discourages the postulation of potentially unnecessary times, and thus favours the interpretation quantifying over parts. Disambiguating an ambiguously quantified sentence to a quantification over times interpretation was rated as less natural than disambiguating it to a quantification over parts interpretation (Experiment 1). In an interpretation questionnaire, sentences with similar quantificational variability were constructed so that both interpretations of the sentence would require postulating multiple times; this resulted in the elimination of the preference for a quantification over parts interpretation, suggesting the parts preference observed in Experiment 1 is not reducible to a lexical bias of the adverb *mostly* (Experiment 2). An eye movement recording study showed that, in the absence of prior evidence for multiple times, readers exhibit greater difficulty when reading material that forces a quantification over times interpretation than when reading material that allows a quantification over parts interpretation (Experiment 3). These experiments contribute to understanding readers' default assumptions about the temporal properties of sentences, which is essential for understanding the selection of a domain for adverbial quantifiers and, more generally, for understanding how situational constraints influence sentence processing.

Correspondence should be addressed to Jesse A. Harris, Department of Linguistics and Cognitive Science, Pomona College, 185 East Sixth St., Claremont, CA 91711, USA. E-mail: jah24747@pomona.edu

This research was supported in part by Grant Number HD18708 from NICHD to the University of Massachusetts. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of NICHD or NIH. We thank Ashley Benatar, Maria Bonilla, Chelsea David, Brendan Davis, Margaret Grant, Jennifer Mack, Morgan Mendez, Desiree Svegliati, and Zachary Waegell for their invaluable assistance with data collection and analysis. We also thank Richard Bogartz, Mara Breen, Alexander Pollatsek, Florian Schwarz, Linnaea Stockall, and Adrian Staub, the attendants of the 24th Annual CUNY Conference on Human Sentence Processing in March 2011, where a version of this work was presented, for all of their insightful comments, suggestions, and criticisms, as well as two anonymous reviewers for their helpful feedback. The first author, now at Pomona College, delivered talks based on this research to audiences at Pomona College and the University of Tübingen, and wishes to thank the audiences there for their feedback.

Keywords: Quantificational variability; Semantic processing; Eye movement recording; Adverbs of quantification; Temporal structure.

Quantificational adverbs like *mostly* in (1) can quantify either over times or occasions as in the paraphrase in (1a), or over parts (or individuals, when the subject DP is plural or collective) as in the paraphrase in (1b) (Berman, 1987; Lewis, 1975; Hinterwimmer, 2008, among others):

- (1) Students mostly walk to school.
 a. On occasions when they go to school, most often they walk.
 b. Most students walk to school.

The phenomenon is interesting because it suggests that the domain over which a quantifier ranges can be determined contextually. This allows us to exploit the processing of sentences with quantificational variability to explore the mechanisms of language interpretation.

Quantified sentences generally contain an overt restrictor, a term that restricts the domain of elements over which the quantifier ranges. For example, the sentence in (2) contains a determiner quantifier *every*, an overt restrictor on its domain (the noun *student*), and the predicate that is asserted *received an A* (which is termed the “nuclear scope”; see Bach, Jelinek, Kratzer, & Partee, 1995):

- (2) Every student received an A.
 $\forall x$ (x is a student) (x received an A)

In addition to an overt restriction on quantificational domains, covert or implicit context can affect the domain of quantification in numerous other ways. For instance, it is common for quantifiers to be *implicitly* restricted by various contextual factors. Example (2) is not typically interpreted as a claim about students everywhere, but rather will probably be understood to have an implicit restriction on the domain of quantification, for example, the students in a particular class under discussion (von Stechow, 1994; Westerståhl, 1984).

What is striking in the case of quantificationally variable sentences is their permissiveness in finding a domain of quantification. In quantificationally variable sentences, implicit information may not only further restrict the *objects* in the domain, but also determine the *type of object* included in the domain. That is, in sentences with quantificational variability, the different interpretations of the sentence involve domains of distinct types, distinguished by the sorts of objects within them.

In addition to domains of parts and times (1), quantificational variability may involve domains composed of other kinds of objects, such as degrees. For example, Frazier, Clifton, and Stolterfoht (2008) examined quantificational variability for absolute scalar adjectives with a maximum (*clean*) or minimum (*dirty*) standard (Kennedy & McNally, 2005; Rotstein & Winter, 2004). As illustrated by the examples in (3), while *mostly* can quantify over either degrees (in this case, of cleanness) or over individuals (dishes), which reading is preferred depends on whether the degree standard is a maximum (100%) standard, as in *clean*, or a minimum (some nonzero degree) standard, as in *dirty*:

- (3) a. The dishes were mostly clean.
Maximum degree standard: quantification over degrees preferred
- b. The dishes were mostly dirty.
Minimum degree standard: quantification over parts preferred

In a written paraphrase selection study, examples like (3a) with a maximum standard adjective (*clean*) received degree interpretations 57% of the time, whereas minimum standard adjectives (*dirty*) received only 40% degree interpretations, so that a 60% majority of interpretations were quantity (*most of the dishes*) interpretations. Frazier et al. (2008) proposed that, in the absence of other factors, adverbial quantifiers preferentially take restrictors based on the content of “local” constituents, that is, constituents with fewer maximal projections separating the constituent from the adverb. Since an interpretation in which *mostly* quantifies over degrees is more local than one in which it quantifies over individuals, this principle accounts for the modest preference for *mostly* to modify degrees in (3a). Frazier et al. (2008) attributed the lack of preference for quantifying over degrees in sentences like (3b) to the fact that they contain minimum standard adjectives, with a vague (some nonzero) standard associated with them. In other words, a quantification over degree interpretation of (3b) would entail interpreting *mostly dirty* in this context as “most of the way to some uncertain nonzero amount of dirt”. As this interpretation is implausible, Frazier et al. (2008) proposed that the processor searched for an alternative domain of quantification, namely, individuals (dishes) rather than degrees of cleanliness/dirtiness.

The present paper further explores the processing of quantificationally variable sentences. It examines examples where *mostly* may quantify either over individuals (and, in Experiment 3, over parts of an entity or collective) or over occasions or times. We assume throughout that, like other quantificational elements, *mostly* takes two arguments in its semantic representation, a restrictor and a nuclear scope. But the restrictor need not be overt, as discussed above. The choice of implicit restrictor, or quantificational domain, can be viewed as a matter of potential semantic ambiguity resolution. Readers might deal with such semantic ambiguity in several different ways. The reader might compute all meanings, eventually accepting the one with the most supporting evidence. The reader could accept a semantically underspecified interpretation (Frisson, 2009), trusting that later material will further specify, if necessary, the quantificational domain. As a variant of underspecification, the reading process could break down, leaving the reader simply confused as to what was being quantified over, at least in a null or uninformative context. Alternatively, the comprehension system could make an unprincipled, “random”, choice of some possible domain, at little processing cost. Finally, it is possible that the processor follows some general principle to choose a quantificational domain, either by committing itself to a single analysis initially (as has been claimed in some cases of syntactic ambiguity resolution; Frazier & Clifton, 1996) or by integrating this principle with other contextual information in choosing an analysis.

In the present investigation, we advance and evaluate one such general processing principle, a cognitive economy preference principle proposed by Majewski (2006, in preparation). Majewski proposed that comprehenders do not in general postulate unnecessary times without evidence. She cast this as the No Extra Times (NET) principle in (4):

- (4) No Extra Times: A sentence describes a single occasion (unless there is evidence to the contrary).

She takes NET to be a general principle. A reader resists interpreting a sentence as referring to multiple events or situations that take place at different times, when the option of referring to a single event or situation is available. The motivation for NET is essentially conceptual economy. We assume that a single occasion can be represented as a single mental representation; multiple occasions require multiple mental representations, or at least a complex mental representation of multiple times, perhaps in terms of a mental model with multiple structures (Johnson-Laird, 1986) or as discrete event parameters within a situation (Barwise & Perry, 1983). Simply put, the reader will favour constructing a single representation when there is no need to create multiple or more complex ones.¹

Majewski originally developed NET to account for the interpretive preferences observed in reciprocals. For example, sentences with reciprocals like *each other* in (5) admit multiple readings in terms of the number of events involved: the “single event reading” in which both songwriters sued the other on a single occasion, and the “multiple event reading” in which each songwriter sued the other on a distinct occasion. To probe the idea that reciprocals preferentially describe one general event (Fiengo & Lasnik, 1973), Majewski presented sentences like (5) in an acceptability judgment-rating task in one of two contexts: ones supporting a single-event interpretation and ones supporting a multiple-event interpretation. She observed a general advantage in the acceptability ratings for sentences that appeared in the single-event context, suggesting that interpreting a sentence with reciprocals as describing multiple events is dispreferred:

- (5) Those songwriters sued each other.

The studies presented below are designed to test the further predictions of NET in a different domain: we report three experiments testing preferences in quantificationally variable sentences containing *mostly*. Experiment 1 obtained judgments of the discourse acceptability of continuations of sentences with *mostly* that presupposed quantification over parts versus quantification over times comparing them to the same continuations of sentences with unambiguous quantification like *Most of the NP . . .* vs. *Most of the time . . .* Experiment 2 addressed whether the preference observed in Experiment 1 could be attributed to a lexical bias associated with *mostly*, rather than a principle like NET. In this experiment, subjects selected their preferred interpretation of the domain—over parts, or over times—for sentences with an ambiguous domain of quantification for *mostly*. The sentences used in Experiment 2 differed from those used in Experiment 1 in that they required the reader to posit multiple times regardless of whether the quantificational domain ranged over parts or over times. If *mostly* were lexically biased so that a quantification over parts rendering were favoured, then subjects should prefer a quantification over parts interpretation across the board. In

¹As observed by an anonymous reviewer, complexity may manifest in multiple forms. In (1b), for example, the quantification over individuals rendering is complex in that the predicate *walk* is restricted to some, contextually determined majority subset of students—that is, those who walk to school. Of course, quantification over times is similarly complex; *mostly* limits the predicate to some contextually determined subset of times. Whether there is more underlying NET than conceptual economy, we are not in a position to say. One possibility is that there is a default that an episodic clause corresponds to or describes a single event. This default might arise not due to conceptual economy, but due to frequency or perhaps to a principle of language production. Of course, the idea that a single clause preferentially describes a single event or situation need not be incompatible with an account based on conceptual economy; in principle both could be true.

- e. Most of the time the students were here. Some of them weren't however.
(QTimes-QParts)
- f. Most of the time the students were here. Some of the time they weren't however.
(QTimes-QTimes)

A strong prediction of NET is that the first ambiguous sentences of the pairs in (6a–b) would initially be interpreted as quantifying over parts not over times. If this were the case, judgments of continuation naturalness for (6a, b) should be very similar to judgments of the unambiguous quantification over parts sentences (6c, d), and different from judgments of the unambiguous quantification over times sentences (6e, f), so that the incongruous pairs (6b, d) should receive lower naturalness ratings than the congruous ones (6a, c). The unambiguous versions (6c–f) also provided controls for the possibility that one type of continuation sentence would be preferred across the board, and allowed us to assess the possibility that no initial commitment to domain of quantification would be made in (6a, b) or that commitment would be random between quantification over times and over parts.

We introduced two additional sources of variance in the items. The first was that of predicate type. Half of the items were created with stative predicates (*were mostly drunk*) and half with locatives (*were mostly in the nursery*). Both kinds of predicates permit an interpretation in which the predicate applies to multiple individuals at one time, which is the interpretation we claim that NET favours over one that requires multiple events or multiple times. The second source of variation was in the quantifier used in sentence 2 continuations, for example, *some* in (6) above. The following four quantifiers were evenly distributed across the 24 items and predicate types: *some*, *a couple*, *a few*, and *several*.

The experimental items were combined with 56 two-sentences discourses from other experiments plus 44 fillers (22 of which contained discourse-unnatural continuations). Six counterbalanced list conditions were created by the program used to conduct the experiment, Linger (<http://tedlab.mit.edu/~dr/Linger/>), so that each subject saw four instances of each of the conditions illustrated in (6), and each of the 24 sentences was tested equally often in each of the six conditions. All experimental materials appear in Appendix A.

Subjects and procedure

Forty-eight undergraduate students from the University of Massachusetts Amherst participated in the questionnaire for course credit. Each was tested individually in a session lasting approximately half an hour, beginning with instructions presented on a computer monitor plus three practice items. Each two-sentence discourse was presented on the monitor as a whole. When a subject had read and understood the discourse, he or she pressed a key on the computer keyboard which brought up a 7-point naturalness scale, with 1 as “totally unnatural” and 7 as “completely natural”. The subject presented a number key to indicate his or her judgment of the continuation sentence. The experimental items and fillers were presented in an individually randomised order, using the program Linger, which recorded responses.

Results

The mean naturalness ratings and standard errors from each condition are presented in [Figure 1](#). It is apparent that contexts with implicit quantificational domains (6a, b)

Experiment 1: Mean rating by condition

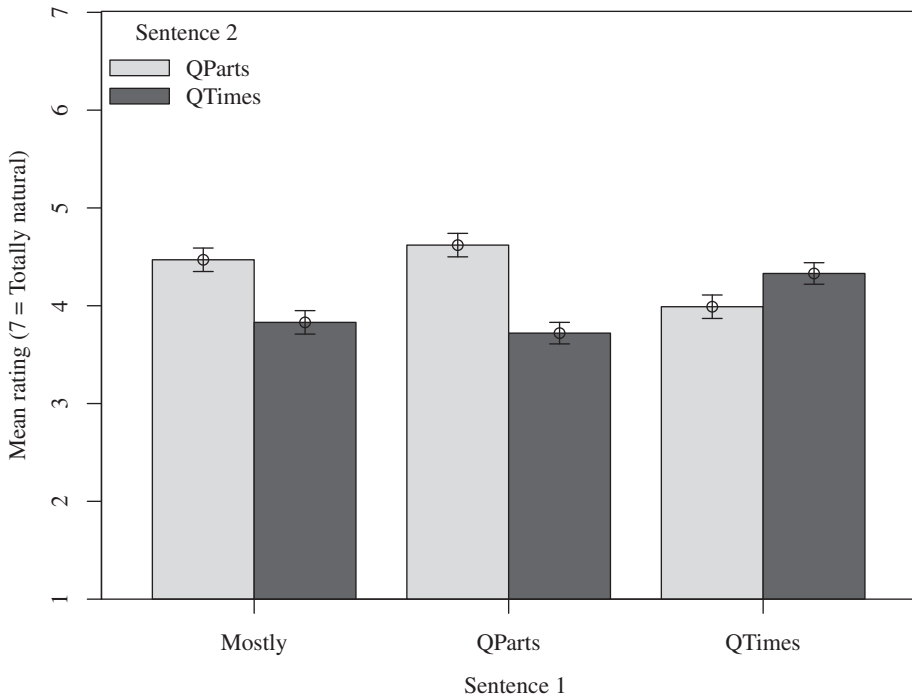


Figure 1. Experiment 1: Plot of mean naturalness ratings by condition, with standard errors indicated.

patterned with explicit quantification over parts (6c, d), as opposed to quantification over times (6e, f) interpretations, as was predicted by the NET hypothesis.

We analysed the data as a linear mixed effects model, using the statistical programming language R (R Development Core Team, 2008), crossing Sentence 1 (Mostly, QParts, QTimes) and Sentence 2 (QParts, QTimes) as fixed effect factors, with by-subjects and by-items random slopes for the fixed effects and their interaction.³ As we were interested in comparing the *mostly* condition to the two unambiguous conditions, we treated the ratings of the (6a) pair, with *mostly* in the sentence 1 context and a QParts continuation, as the intercept in the model. We used treatment-coding in the analysis, and report all significant effects below.

First, considering items with QParts continuations, that is, the (6a), (6c), and (6e) forms, items with explicit QTimes initial sentences (the (6e) forms; $M=3.99$, $SE=0.12$) were rated significantly lower than items with contexts containing *mostly* (the (6b) forms; $M=4.47$, $SE=0.12$), $t = -3.33$, $p < .001$, which did not differ significantly from items with QParts quantificational contexts (the (6c) forms; $M=4.62$, $SE=0.12$), $t = 0.80$. This was as predicted by NET: sentences containing *mostly* pattern with explicit QParts contexts rather than QTimes contexts. Second, for items containing *mostly* in the context, QParts continuations ($M=4.47$, $SE=0.12$) were preferred over QTimes continuations ($M=3.83$, $SE=0.12$), $t = -2.82$, $p < .05$. Again, this pattern was predicted by NET: if the domain for *mostly* is preferentially

³We thank an anonymous reviewer for suggesting we use random slopes in our error structure, although the effects did not differ appreciably from a linear mixed-effects model with the same fixed effects and simple by-subjects and by-items random intercepts.

interpreted as ranging over parts, then it should be penalised when paired with an incongruous continuation, for example, one supporting a quantification over times interpretation.

With respect to interactions, NET makes two predictions, which we explore in turn. NET predicts that contexts with ambiguous quantificational domains (6a, b) should be treated on par with explicit quantification over parts (6c, d), and thus the size of the penalty for QTimes continuations is not predicted to differ in this case. When the context explicitly quantified over parts, sentences that continued with a QParts interpretation were rated 0.90 better than sentences that continued with a QTimes interpretation, and when the context contained the ambiguous *mostly*, the difference between these continuations was nonsignificantly smaller at 0.64, ($t = -0.96$). That is, we found no statistically significant evidence that pairs with quantification over times continuations were easier to accommodate when the domain was implicit (6b) than when the domain was explicitly QParts (6d), strongly suggesting that *mostly* was given a QParts interpretation. In contrast, the interaction between the type of continuation and the contrast of ambiguous contexts and contexts disambiguated towards QTimes interpretations *was* significant. The 0.64 advantage effect of the QParts continuation following an ambiguous mostly context case was significantly larger than the 0.34 disadvantage following an explicit QTimes context, $t = 4.22$, $p < .001$. As a whole, these results are in keeping with the claim that the domain of quantificational adverbs like *mostly* are biased against postulating unnecessary times, as predicted by NET.

Finally, we conducted analyses to determine whether constructional differences between items had any effect. We found no effect of predicate type, that is, stative predicates did not differ from locative ones, and we found no effect of quantifier type in the sentence 2 continuation.

Discussion

Experiment 1 examined naturalness ratings for continuations in two-sentence discourses. It found that sentence contexts with an implicit quantificational domain behaved like sentence contexts with overt quantification over parts, in that—in both conditions—continuations supporting quantification over times were rated as less natural than continuations supporting quantification over parts. Our results straightforwardly support the NET principle by displaying a bias against unnecessary quantification over times interpretations. We found no evidence supporting the alternative hypotheses in which comprehenders (i) avoid commitment to an interpretation of the quantificational domain, or (ii) select a quantificational domain at random. Rather, we have initial evidence that comprehenders fix the quantificational domain early and do so in accordance with the domain-independent constraints imposed by NET.

EXPERIMENT 2

In our interpretation of Experiment 1, we assumed that the interpretation of the initial sentence context with an ambiguous quantificational domain affected the judgment of the naturalness of the continuation sentence by being biased towards quantification over parts. We interpreted the results in terms of NET: the bias for quantification over parts, in the absence of other, equally economical alternatives, reflects a general

preference against multiplying times without necessity in interpretation. However, identical results would have been expected if the adverb *mostly* is lexically biased towards the quantification over parts interpretation.

We consider such a lexical bias to be unlikely. For one thing, on their basic usage, adverbs modify verbal projections—predicates holding at particular times; if *mostly* is taken to modify the predicate, a QTimes interpretation would be likely. Nonetheless, Experiments 2a and 2b were designed to evaluate the possibility that *mostly* is lexically biased.

Experiment 2a

Experiment 2a tested the interpretation of sentences in which both quantification over times and quantification over parts interpretations encourage comprehenders to postulate multiple events at multiple times. In such cases, the NET principle does not predict a preference for avoiding quantification over times, as multiple times may be represented in any event. However, a preference for quantification over parts would still be expected if *mostly* were lexically biased to select that quantificational domain. Instead of using stative or locational predicates as in Experiment 1, Experiments 2a and 2b used predicates that are very likely taken to be habitual or to otherwise to describe events on multiple occasions, as illustrated in (7):

- (7) a. The students mostly walk to school.
- b. Students mostly walk to school.
- (8) What did that sentence mean?
 - 1. They walk to school most of the time.
 - 2. Most of them walk to school.

The presumably favoured habitual interpretation of the predicates used in Experiment 2 implies the existence of multiple times, even with a QParts interpretation of *mostly*; that is, sentences like *Most of the students habitually walk to school* require multiple situations involving walking, even if *mostly* is not taken to quantify over situations. If the predicate is not taken to be habitual, it is likely to be taken to describe multiple events that take place at multiple times. If interpretation of the predicate requires the representation of multiple times, NET would have no reason to favour a QParts over a QTimes interpretation (in contrast to the sentences used in Experiment 1, which could describe a single situation). Consequently, at least on the basis of NET, we do not predict a bias towards either interpretation in (7).

Lexical bias and NET are, to be sure, not the only factors that could affect the choice of quantificational domains. We tested one other possible factor by investigating sentences with definite description subjects (7a) and their bare plural counterparts (7b). A bare plural subject could encourage a habitual interpretation, and it is possible that the repeated times entailed by a habitual interpretation would encourage a reader to take *mostly* to quantify over times, rather than over parts. However, a bare plural subject could alternatively encourage a generic interpretation (in which the predicate generally holds true of members of the class denoted by the subject), which could favour taking *mostly* to quantify over parts, in the case of (7), over a set of contextually determined students. We had no strong expectations about which would be correct, but included the manipulation to explore potential factors other than NET that might influence quantificational domains are selected.

Methods

Materials. Twenty sentences like those in (7) were constructed (see Appendix B for a complete list). All sentences had present tense predicates, which are easily interpreted as habitual or generic. Two versions of each sentence were constructed, with (7a) and without (7b) the definite determiner, and a two-choice question with possible answers was written for each sentence, as illustrated in (8). One answer reflected the quantification over times interpretation of *mostly*, and the other reflected the quantification over parts interpretation. Presentation order of the answers was randomised individually for each subject. These items were combined into a questionnaire with 130 other sentences, 51 to be followed by two-choice questions, and the rest to be followed by acceptability or comprehension difficulty rating scales. Six of these 130 were “cheater” sentences with intuitively obvious answers or ratings, designed to identify subjects who did not follow instructions, and the rest were from unrelated experiments. The sentences were divided into two counterbalanced lists, each with 10 experimental sentences in each of the two conditions illustrated in (7). Each sentence was tested in one condition in one list and the other in the other list.

Subjects and procedures. A total of 55 University of Massachusetts undergraduates were tested in individual half-hour sessions. Five were eliminated for providing an incorrect answer to half or more of the “cheater” sentences, and an additional two were randomly eliminated so that 24 were assigned to each of the two counterbalanced lists.

The sentences were presented (following instructions and two practice sentences) on a computer terminal in individually randomised order, again using the program Linger. Each experimental sentence was presented on a computer monitor, as in Experiment 1, followed by presentation of a two-choice question with answers randomly assigned to the number 1 or 2 (or a rating scale, in the case of other experiments). Subjects responded by pressing the keyboard number key that corresponded to their preferred answer. The computer recorded the answers.

Results and discussion

The primary results are very simple: subjects chose the quantification over times answer (*They walk to school most of the time*) 56% ($SE = 2\%$) of the time when a determiner was present in the sentence (8a), and 49% ($SE = 2\%$) of the time when no determiner was present (8b). The principal result is that there is no bias in favour of a quantification over parts interpretation (*Most of them walk to school*), which was chosen 44% and 51% of the time for the two versions of sentences.

The 7% difference between interpretations of sentences with and without the determiner, while small, was significant. It was tested using a logistic linear mixed-effects model analysis, with by-subjects and by-items random slopes, analyzing the logit transformation of the observed proportions (see Baayen, Davidson, & Bates, 2008; Jaeger, 2008, for description and justification). The difference between the frequency of the two choices yielded $z = 2.36$, $p < .05$. Quantification over times was chosen significantly more often when the definite determiner was present than when it was not.

Experiment 2a thus shows that there is not a strong bias for *mostly* to quantify over parts when doing so requires an interpretation involving multiple occasions (as quantifying over times always does). The tendency for bare plural sentences to favour QTimes interpretations less than sentences with definite description subjects suggests

that the bare plural favours a (possibly nonhabitual) generic reading, encouraging quantification over parts. We conclude that whatever biases *mostly* itself might impose, it is not an overwhelming general bias in favour of quantifying over parts, and that multiple factors influence choice of quantificational domain.

Experiment 2b

Experiment 2b explored whether the present tense used in Experiment 2a was critical to the finding of no bias towards interpreting *mostly* in terms of quantification over parts.⁴ It is possible that the present tense encourages a habitual interpretation, and that this is responsible for the relatively high frequency of QTimes interpretations in Experiment 2a. Experiment 2b was conducted to test this possibility.

Methods

Materials. The 24 sentences of Experiment 2a were used, with two past tense forms added, as illustrated in (9):

- (9) a. The students mostly walk to school.
 b. Students mostly walk to school.
 c. The students mostly walked to school.
 d. Students mostly walked to school.

Subjects and procedures. The sentences were combined with 64 other sentences (all followed by two-choice questions with answers presented in random order) and assigned to four counterbalanced lists, so that each subject saw six sentences in each of the four forms, and each sentence was tested equally often in the four forms. Fifty-six undergraduates from the University of Massachusetts Amherst participated, using procedures identical to those described for Experiment 2a.

Results and discussion

Table 1 presents the percentages with which the QTimes interpretation was chosen. For analysis, we first fit the model with the planned effects of Determiner and Tense and their interactions as fixed effects and with by-subjects and by-items random slopes for those effects. No significant effects were observed. Following methods described in Crawley (2007), we reduced the regression model by successively removing the term that accounted for the least amount of observed variance. At each stage, the simplified

TABLE 1
 Percentages of choice of Quantification over Times interpretations (with standard errors in parentheses) for Experiment 2b

<i>Determiner</i>	<i>Tense</i>	
	<i>Past</i>	<i>Present</i>
Bare	46 (3)	43 (3)
Definite	50 (3)	53 (3)

⁴This possibility was suggested by a previous reviewer, who noted that the other experiments reported in this paper used past tense verbs.

model was compared against the previous, more complex model, via a chi-squared test, to verify that the simplification was warranted. The best fitting model, that is, the simplest model explaining the most amount of variance was one with a single predictor of Determiner, $z = 2.69$, $p < .01$. As observed in Experiment 2a, QTimes interpretations were more frequent for definite noun phrases than for bare plurals; this is expected if the bare plural noun phrase encouraged a generic interpretation. There was no effect of tense, and no significant interaction. For both past and present tense, there was little or no bias for QParts as opposed to QTimes.

As we concluded after Experiment 2a, *mostly* does not have a semantic bias to quantification over parts. Rather, the bias observed in Experiment 1 is best understood in terms of NET: when quantification over parts permits interpretation in terms of a single event and single time, as it did in Experiment 1, QParts is preferred over QTimes.

The results of Experiments 1 and 2 together suggest that, while there is no strong lexical bias associated with *mostly* that determines its quantificational domain, there is evidence for a general conceptual economy principle that penalises positing additional times for interpretation without evidence, which we have called the NET principle. In the third experiment, we used an on-line technique to determine the time course over which the NET-based preference for quantification over parts appears.

EXPERIMENT 3

Experiment 3 investigated the online interpretation of quantificationally variable sentences in an eye movement recording study examining the following three sentence types: (i) sentences intended to be neutral between quantifying over parts versus over times, (ii) sentences designed to be strongly biased towards the quantification over times (by combining a locative predicate with an individual, for example, an inspector, that could not plausibly be thought to be in different locations at one time), and (iii) sentences biased towards quantification over parts (by combining the same locative predicates with entities that can be divided into distinct parts, but cannot readily participate in events that occur at distinct times, either because they are events that happen at a single time or because they persist without change over time). It turned out to be very difficult to create completely neutral sentences, but we present normative data that indicate relative success in doing so. We predict that when both quantification over parts and quantification over times are possible, NET will create a bias towards quantification over parts. Making quantification over parts implausible will move the bias past neutrality towards a preference for quantification over times.

In sum, Experiment 3 explored quantificational variability involving quantification over parts versus times by manipulating the effect of a locative phrase, such as *in the capital*, on subject nouns with varying spatio-temporal properties in sentences with *mostly*. Consider (10), below. The Neutral sentences were like (10a) in that their subjects can be divided into natural parts of the same kind and also can participate in the predicated event at different times (e.g., *army*, *team*). Many of the nouns we used in the Neutral condition were ambiguous between a collective and a group entity. Sentence (10a), for instance, supports either a reading in which most of the army was in the capital, or a reading in which the army was in the capital most of the time. As mentioned earlier, if following NET quantification over times is dispreferred in the absence of other evidence, we expect that these “neutral” sentences will preferentially be interpreted with quantification over parts. The Nondivisible (10b) sentences had

subjects whose referents are not divisible in the relevant sense into natural parts (e.g., *inspector*, *mascot*), but could participate in the predicated events at different times; these were designed so that a locative phrase following *mostly* would favour the quantification over times. It is pragmatically odd to interpret sentence (10b) as asserting that some parts, but not others, of the inspector were in the capital on a single occasion. The Divisible sentences (10c) had subjects which referred to entities that can be divided into natural parts but that have a temporal unity intended to make it difficult to conceive of them participating in the predicated event at different times. We intended these sentences to be biased towards quantification over parts. For instance, consider example (10c): if an attack is spread out over separate times, it seems natural to refer to several attacks not a single attack.

We began with a norming study, investigating how ambiguous or biased the sentences we created were. We then conducted an on-line reading study to see if there are differences in processing difficulty between the different sentence types. In particular, if NET is correct, a quantification over parts interpretation should be assigned as long as it is compatible with the input. Consequently, long reading times should emerge in the locative region of the Nondivisible items, the point when the input is disambiguated or strongly biased against a quantification over parts interpretation.

Norming study

Methods

Materials. Twenty-seven sentence triples, as illustrated in (10), were constructed. In each triple, one sentence (a) was intended to be neutral between a quantification over parts and a quantification over times interpretation; one sentence (b) was intended to be unambiguously quantification over times; and one sentence (c) was intended to be biased towards quantification over parts. One pair of disambiguating interpretations was constructed for each sentence. The quantification over parts disambiguation had the form *Most of the army was in the capital* while the quantification over times interpretation had the form *The army was in the capital most of the time*.⁵

- | | | | |
|------|----|--|----------------|
| (10) | a. | The army was mostly in the capital. | (Neutral) |
| | b. | The inspector was mostly in the capital. | (NonDivisible) |
| | c. | The attack was mostly in the capital. | (Divisible) |

Three forms of a questionnaire were constructed. Each form contained nine sentences in each of the three forms illustrated in (10), and were followed by a quantification over parts disambiguation and a quantification over times disambiguation, in that order. Each questionnaire form contained the resulting 27 sentences in a different random order.

Subjects and procedure. Thirty-six University of Massachusetts undergraduates completed the questionnaire in individual sessions lasting 10 to 15 minutes. Twelve completed each form of the questionnaire. They were instructed to “choose the most

⁵As Patrick Sturt pointed out to us, the singular verb form for collective nouns like (10a) may be ungrammatical or at least dispreferred when compared to counterparts with the plural verb form in British English. However, in American English, collective nouns are singular in number. We follow American usage here.

accurate interpretation from the pairs listed beneath each sentence” by marking an X by the interpretation they found the most natural. They were told that they could check both interpretations if they found them both the same in naturalness and plausibility.

Results

The three sentences for which the manipulation of quantificational domain bias was least successful were eliminated. The frequency with which each version of each of the remaining sentences received a “quantification over parts”, a “quantification over times”, and a “both equally natural” interpretation was tabulated. These frequencies appear in Table 2. It is clear that the intended disambiguation towards quantification over times was successful. Further, the clear majority of quantification over parts items was interpreted in the intended way. However, a substantial majority of the intended-neutral items were also interpreted as quantification over parts. The mean percentages of parts interpretations differ among sentence forms by at least four SEs, as do the mean percentages of times interpretations. While disambiguation towards the parts interpretation is not perfect, and as we indicated earlier, the “neutral” items were not completely unbiased, the differences among the types of sentences are clearly substantial enough to permit a test of our claims.

In sum, there was a clear difference between items intended to be biased to quantification over parts (Divisible) and items intended to be biased to quantification over times (Nondivisible). There was a somewhat smaller difference between items biased to quantification over parts and items intended to be neutral between the two interpretations (Neutral). We considered the lack of complete ambiguity observed for the Neutral items to be a reflex of NET, rather than a confound in our items. However, we briefly address the issue of variability between items below.

Eye movement study

Method

Materials. The experimental materials consisted of 24 triplets like (11a–c):

- (11) a. The army ₁ was mostly ₂ in the capital, ₃ according to the newspapers. ₄
 b. The inspector ₁ was mostly ₂ in the capital, ₃ according to the newspapers. ₄
 c. The attack ₁ was mostly ₂ in the capital, ₃ according to the newspapers. ₄

The materials correspond to those that appeared in the norming study, except that an additional, sentence-final region (*according to the newspapers*) was added to separate processing effects associated with the regions of interest from well-known wrap-up

TABLE 2

Per cent quantification over parts, quantification over times, or both interpretations by condition (with standard error)

Sentence form	Domain of quantification		
	Parts	Times	Both
Neutral (army)	62.0 (3.4)	22.0 (2.9)	16.0 (2.6)
Divisible (attack)	85.0 (2.5)	9.5 (2.8)	5.5 (1.6)
Nondivisible (inspector)	0.5 (0.5)	99.0 (0.7)	0.5 (0.5)

effects (Just & Carpenter, 1980; Rayner, Kambe, & Duffy, 2000). As before, the items manipulated only the subject of the sentence. Approximately half of items had subjects containing two words (e.g., *mountain range*), rather than one, but were distributed evenly across conditions. Lexical level characteristics of the sentential subjects were obtained from the English Lexicon Project (Balota et al., 2007). If the subject contained two words in the region, only measures from the second were considered. Sentence subjects did not vary significantly by log HAL frequency, $F(2,23) = 2.09$, $p = .13$, with means of 9.62, 8.59, and 8.87 for Neutral, Nondivisible, and Divisible conditions, respectively. Additionally, the three conditions did not differ by length, $F(2,23) = 0.06$, $p = .95$, with means of 6.71, 6.83, and 7.00 for Neutral, Nondivisible, and Divisible conditions, respectively.

The experimental items were divided into three counterbalanced lists, each containing eight sentences in each of the three conditions illustrated in (11). Each list contained only one item from each of the 24 sentences, and each sentence appeared once in each list. The experimental items were interspersed with 100 filler sentences from four other experiments, of various type and complexity, following eight practice items. Forty of the sentences were followed by comprehension questions, including six experimental sentences (three of which were followed by questions that involved quantificational domain, and three by questions involving other aspects of the sentence).

Subjects and procedure. Thirty-four native speakers of American English participated in the eye movement study. All participants were members of the University of Massachusetts, Amherst community and were either given course credit or a small cash payment to participate. Prior to testing, participants received both written and verbal instructions. Participants were encouraged to read naturally for comprehension, and were not told about the experimental manipulations.

Participants were tested individually, their eye movements recorded using an EyeLink 1000 (SR Research, Toronto, Ontario, Canada) eyetracker. The eyetracker was interfaced with a PC computer, whose sampling rate was set at 1000 Hz. Participants were seated 55 cm from a CRT monitor displaying the stimuli. Although participants viewed the stimuli with both eyes, only data from the right eye were collected.

After participants received verbal instructions, the experimenter calibrated the eyetracker, recalibrating between trials when necessary. Trials were initiated when the participant fixated on a box at the left edge of the screen. Participants were instructed to immediately look to a coloured square affixed to the right edge of the monitor after each sentence. Participants signalled that they had completed the trial by pressing a trigger on a button box. After half of the trials, participants viewed a comprehension question. Responses to these questions were recorded with the button box. The experiment lasted approximately 45 minutes.

Results and discussion

Materials were divided into four regions for the purpose of analysis, as shown by the numbered | symbols in (11). Region 1 consisted of the definite determiner and the subject; region 2 of the past tense copula (*was*) followed by the adverb *mostly*; region 3 of a locative phrase; region 4 of an adjunct of varying semantic category. As mentioned, all regions but Region 1 were invariant within the triplet.

Prior to analysis, the eye movement data was cleaned of track losses, blinks, and long fixations, that is, those over 800 ms, that were not in Region 2. If these long fixations appeared in the Region 2, the entire trial was discarded. On this basis, 16 trials were removed from analysis, accounting for less than 0.02% of the total number

of trials. Following Rayner and Pollatsek (1989), we assume that short fixations, that is, those under 80 ms, do not contribute useful information to the reader. Thus, short fixations were incorporated into the nearest neighbouring fixation within three characters; otherwise, they were deleted.

We computed five reading time measures: *first pass time*, *go past time*, *second pass time*, *percentage regressions in*, and *percentage regressions out* (Rayner, 1998). *First pass time* reflects the sum of all fixations in the region before leaving the region to the left or right. *Go past time* reflects the sum of all fixations from when the reader first fixates in the region before leaving the region to the right, in addition to any time spent re-reading the region or prior regions before exiting to the right. *Second pass time* measures the time spent re-reading a region, once the region has been exited to the right, that is, the time after a regression into a region that the reader spends re-reading. *Percentage regressions in* is the percentage of all trials on which regressive saccadic movements were made into a region. *Percentage regressions out* is the percentage of trials in which a first-pass fixation made in a region was followed by a regressive saccade out of a region. Multiple eye-tracking measures were examined in an attempt to identify how quickly any influence of quantificational domain biases appeared. While some effects (including lexical and some syntactic and semantic effects) appear very early in the eye movement record (e.g., on first pass times), some semantic and pragmatic effects appear only after a regressive eye movement is made from a critical region (Clifton, Staub, & Rayner, 2007). Existing data support no clear expectations of just when any effects of quantificational domain biases will appear.

Data from each measure was cleaned of outliers. In those measures with a generally normal distribution—that is, first fixation duration and first pass time—scores above 3 standard deviations from the mean of each region were discarded. In go past and second pass measures, outliers were identified by visual inspection and removed. Cut-off points for these two measures were determined on a region-by-region basis. The resulting data loss never exceeded 2% of the data by region. All descriptive statistical measures and significance tests were computed over the cleaned data. Means and standard errors (in parentheses) for each condition and each region are presented in Table 3 for all analyses.

Using the R-language software for statistical computing (R Development Core Team, 2008), each measure and region of the cleaned data was then modelled as linear mixed effects regression models (Baayen et al., 2008) with by-subjects and by-items random slopes for condition (except for the analysis of proportion of regressions in and regressions out, where the random slopes model did not produce a better fit than the random intercepts model, and the analysis of second pass times, where conventional ANOVAs were used as described below). In all of our models, we treated the Neutral condition as the Intercept and computed whether the other two conditions showed signs of significant differences in reading time. Significance values were obtained from *t*-values computed from the *lmer* package (Bates & Maechler, 2009) in R, assuming a large *df*. (These significance values are anti-conservative for small *df*, but the bias is very slight for the large number of *df* in our design.) All significant effects are reported.

To foreshadow the general pattern of reading times, in Region 3, where the Nondivisible items force a quantification over times reading due to the nature of the noun and the predicate, the Nondivisible items exhibited disrupted reading, compared to the Neutral and the Divisible conditions. This disrupted reading appeared as disrupted reading in Region 3 for Nondivisible items; a longer go-past time for Region 3, longer second pass times for the preceding regions, and a higher

TABLE 3
 Eye-tracking measures, Experiment 3 (standard error in parentheses)

	<i>Region 1</i>	<i>Region 2</i>	<i>Region 3</i>	<i>Region 4</i>
	<i>Subject</i>	<i>was mostly</i>	<i>locative phrase</i>	<i>sentence final</i>
<i>First pass times</i>				
Neutral (army)	349 (14)	366 (10)	455 (14)	992 (23)
Divisible (attack)	374 (15)	333 (9)	435 (14)	1,010 (22)
Nondivisible (inspector)	342 (13)	329 (9)	444 (14)	984 (23)
<i>Go past times</i>				
Neutral (army)	358 (15)	412 (13)	509 (16)	1,044 (25)
Divisible (attack)	377 (15)	370 (11)	514 (18)	1,047 (24)
Nondivisible (inspector)	342 (13)	386 (13)	554 (20)	1,069 (25)
<i>Second pass times</i>				
Neutral (army)	38 (7)	32 (6)	12 (4)	NA (NA)
Divisible (attack)	21 (5)	30 (5)	10 (4)	NA (NA)
Nondivisible (inspector)	52 (8)	53 (8)	18 (54)	NA (NA)
<i>Percentage regressions out</i>				
Neutral (army)	NA (NA)	8.0 (1.7)	10.0 (1.9)	3.8 (1.2)
Divisible (attack)	NA (NA)	6.4 (1.5)	10.5 (1.9)	3.7 (1.2)
Nondivisible (inspector)	NA (NA)	9.2 (1.8)	17.2 (2.3)	8.0 (1.7)
<i>Percentage regressions in</i>				
Neutral (army)	13.3 (2.1)	11.0 (1.9)	1.3 (0.7)	NA (NA)
Divisible (attack)	8.2 (1.7)	11.2 (1.9)	1.76 (0.8)	NA (NA)
Nondivisible (inspector)	15.5 (2.2)	17.1 (2.3)	4.1 (1.3)	NA (NA)

frequency of regressions out of Region 3 for Nondivisible items than for Neutral items. An unexpected finding was that in Region 2 (*was mostly*), there was evidence of long reading times in the Neutral condition. We suggest below that this reflects the fact that many of our sentences used a singular term (e.g., *committee*) to pick out the subject referent, rather than a plural (*committee members*), which would have been more natural with the presumably preferred quantification over parts interpretation.

Reading after disambiguation. Go-past times for Region 3 (Figure 2) were longer in the Nondivisible condition ($M = 554$, $SE = 19$) than in the baseline Neutral condition ($M = 509$, $SE = 16$), $t = 2.11$, $p < .05$. In contrast, the Divisible condition ($M = 514$, $SE = 18$) did not exhibit any signs of an increased processing cost over Neutral items in this region, $t < 1.0$.

The Nondivisible condition elicited significantly more regressions out of Region 3 (17%; Figure 3) than the baseline Neutral condition (10%), $z = 2.48$, $p < .02$, which was treated on par with the Divisible condition (11%), $z < 1.0$. A similar pattern appeared in Region 4, where there were significantly more regressions out for the Nondivisible condition (8%) than for the Neutral condition (4%), $z = 2.12$, $p < .04$. As in the previous region, no differences between the Divisible condition and the Neutral condition (4%) were detected, $z < 1.0$, in Region 4.

After initiating a regression, readers were more likely to land either in Region 1 when viewing an item from the Nondivisible or Neutral conditions, or (at least numerically) in Region 2 if processing an item from the Nondivisible condition (Figure 4). Readers made significantly fewer regressions into Region 1 when viewing items from the Divisible condition (8%) as compared to the Neutral condition (13%),

Experiment 3: Go past times for regions 1 – 3

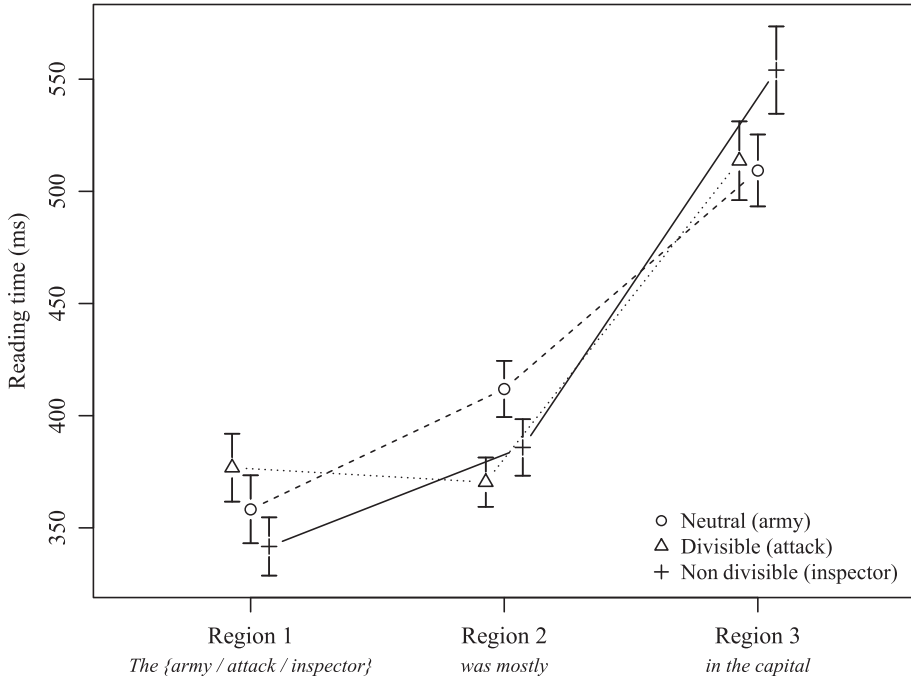


Figure 2. Go past times for Regions 1–3 for Experiment 3. Region 4 is omitted for clarity: no significant differences in go past times were observed in that region.

$z = 2.04$, $p < .04$. The Nondivisible condition did not differ significantly from the Neutral condition, $z < 1.0$, eliciting regressions into Region 1 on 16% of the observed trials. The Nondivisible condition prompted numerically more regressions into Region 2 than the other conditions, 17% of trials, compared to 11% for both Neutral and Divisible conditions, but the difference between the Nondivisible and the Neutral condition was not significant ($z = 1.59$, $p = .11$).

Because second-pass times are distributed in an extremely non-normal fashion on a trial-by-trial basis (the modal value is typically 0 ms), they were averaged over items and over subjects and analysed using F_1 and F_2 ANOVAs. There was a significant effect of condition in Region 1 second pass times, $F_1(2,66) = 4.29$, $F_2(2,46) = 4.73$, $p < .05$. The Neutral condition was compared to the Nondivisible and Divisible conditions using t -tests (with an error term taken from the ANOVA's MS error term). Increased regressions into Region 1 lead us to expect increases in second pass reading times in that region. However, items from the Divisible condition, which did not prompt as many regressions into Region 1 as the items from the Neutral condition, did not elicit significantly shorter second pass times ($M = 21$, $SE = 5$) than their Neutral counterparts ($M = 38$, $SE = 7$), $t_1(66) = 1.49$, $t_2(46) = 1.72$, $p > .05$. However, Region 1 second pass times were significantly longer for the Nondivisible ($M = 52$, $SE = 8$) than the Neutral condition, $t_1(66) = 2.09$, $t_2(46) = 2.41$, $p < .05$. A comparable difference was observed in Region 2. The main effect of condition was significant or marginally significant, $F_1(2,66) = 2.91$, $.05 < p < .10$; $F_2(2,46) = 3.90$, $p < .05$. Readers spent more time re-reading Region 2 in the Nondivisible condition ($M = 530$, $SE = 8$) than in the Neutral condition ($M = 32$, $SE = 6$), $t_1(66) = 1.83$, $0.05 < p < .10$; $t_2(46) = 2.07$, $p < .05$. As expected, items from the Divisible condition ($M = 30$, $SE = 5$) did not elicit increased re-reading times in this

Experiment 3: Percentage regressions out for regions 2 – 4

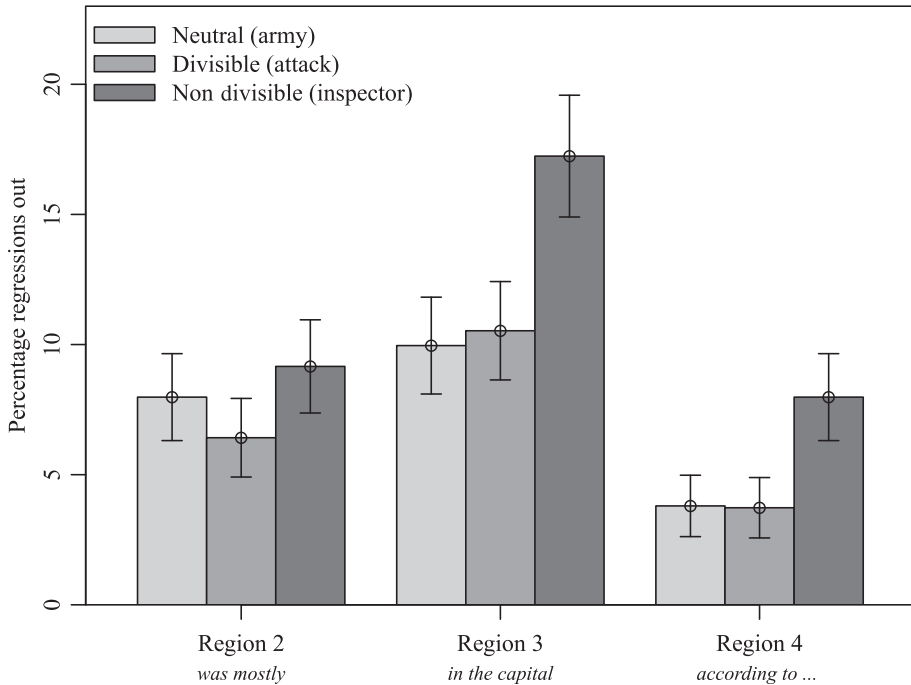


Figure 3. Percentage regressions out by region for Experiment 3. As the *regressions out* measure is not defined for item initial regions, region 1 is omitted from the plot.

region compared to the baseline Neutral condition ($t < 1.0$). No other significant effects were observed in later regions.

Reading before disambiguation. In contrast to this evidence for substantial re-reading prompted by a predicate favouring quantification over times (the Nondivisible items), no evidence for processing costs were observed in first pass reading time for Region 3. However, there was evidence for slower first pass reading of Region 2, (*was mostly*), in the Neutral condition than in the other conditions in the experiment. As shown in Table 3, readers spent significantly more time in Region 2 when processing items from the baseline Neutral condition ($M = 366$, $SE = 10$) than those from the Divisible ($M = 333$, $SE = 9$) or Nondivisible ($M = 329$, $SE = 9$) conditions, $t = 2.06$ and 2.53 respectively, p 's $< .05$. A similar trend also appeared in go-past times, although the differences did not reach significance under the model specified above: Neutral items elicited longer reading times than the other items in Region 2.⁶ As shown in Figure 2, items from the Neutral condition ($M = 412$, $SE = 13$) had numerically longer go-past times than items from both the Divisible condition ($M = 370$, $SE = 11$), $t = 1.77$, and the Nondivisible condition ($M = 386$, $SE = 13$), $t = 1.44$. In principle, this effect might be an ambiguity effect, though typically it is only lexical ambiguity that slows reading (see Clifton & Staub, 2008, for evidence that

⁶In addition, we found that a model of go-past times on Region 2 with by-subjects and by-items random intercepts better fit the data by the BIC criterion. In this model, the difference reached full significance between Neutral and Divisible conditions, $t = 2.70$, $p < .01$, and near significance between Neutral and Nondivisible conditions, $t = 1.71$, $0.05 < p < .10$.

Experiment 3: Percentage regressions in for regions 1 – 3

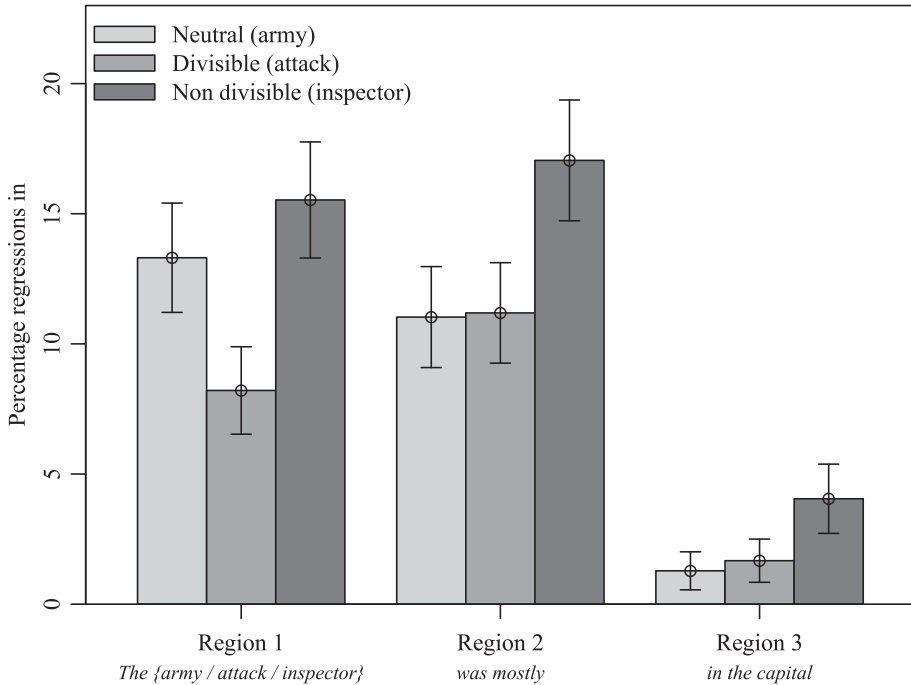


Figure 4. Percentage regressions in by region for Experiment 3. As the *regressions in* measure is not defined for item final regions, region 4 is omitted from the plot.

syntactic and semantic ambiguity actually speeds reading). We suggest that this effect may reflect a possible mismatch between the NET-favoured QParts interpretation and the use of a singular subject rather than a plural in Region 1. Suppose that, upon reading *mostly*, NET initially favours a QParts interpretation in all conditions. In some of our Neutral items (but not in the other items), there is a clear way to express that the referent of the subject noun is divisible into parts; for example, the collective noun *committee* is easily disambiguated to a parts interpretation if rendered as *committee members*, but *fishing trip* is not so easily disambiguated. We reasoned that the long first pass times observed in Region 2 could reflect the fact that a cooperative writer who wanted to quantify over the parts of a collective would have used a plural form if it was available; the use of the singular collective could produce some tension in the reader when *mostly* was taken to quantify over parts or individuals. To probe this idea, we categorised the nouns in the Neutral condition into three types: nouns that could be easily disambiguated to a parts rendering by use of the plural ($N=11$), those that could not ($N=10$), and those for which it was unclear ($N=3$).⁷ Consistent with our *post-hoc* hypothesis, first pass times on Region 2 for easily disambiguated nouns ($M=383.61$, $SE=16.89$) were numerically longer than either nouns that were not easily disambiguated ($M=352.61$, $SE=13.57$) or those for which it was unclear ($M=346.07$, $SE=22.54$). Though not definitive given the inadequate number of

⁷The noun phrases that we classified as easily disambiguated to a parts rendering were *army*, *committee*, *jury*, *platoon*, *police force*, *news crew*, *dance class*, *wrestling squad*, *soccer team*, *gang*, and *guided tour*; the noun phrases classified as not easily disambiguated to a parts interpretation were *mountain trek*, *business advertising*, *convention*, *conference*, *condo development*, *lab work*, *demonstration*, *hurricane*, *corruption*, and *fishing trip*; and noun phrases that we considered unclear were *herd*, *safari*, and *expedition*.

observations and lack of statistical power, the trend for longer times for the subjects that could be “pluralised” is highly suggestive.

GENERAL DISCUSSION

In Experiment 1 NET was supported: by interpreting *The students were mostly here* as quantifying over students, postulating a set of distinct times is unnecessary. Avoiding the postulation of unnecessary times can explain why neutral sentences were resolved to an interpretation quantifying over parts rather than times, as indicated by the naturalness ratings of context-continuation pairs of Experiment 1.

In Experiments 2a and 2b, we attempted to tease apart predictions based on the NET principle from the possibility that *mostly* itself had a lexical frequency bias for quantifying over individuals or parts. If so, a preponderance of quantification over parts interpretations should have been reported for *(The) students mostly walk to school*. This, however, was not observed, and suggests that *mostly* does not preferentially quantify over parts simply as a matter of its most frequent usage. Experiments 2a and 2b used sentences with habitual or multiple-event predicates that strongly encourage interpretations involving multiple times or events even when *mostly* is taken to quantify over parts. Thus, in contrast to the sentences used in Experiment 1 (which had stative or locative predicates, and were not likely to receive habitual or multiple-event interpretations), the NET principle does not favour their being interpreted as quantifying over parts, and is consistent with the observed lack of a preference for a QParts interpretation.

In Experiment 3, NET predicted that the Nondivisible condition would be difficult because the implausibility of the predicates in the experimental sentences quantifying over parts of a singular entity forces an interpretation in terms of multiple times. The Neutral and Divisible conditions should be easier to process since it is possible to assign the presumably preferred interpretation with quantification over parts. This is what was observed in Experiment 3 in eye-tracking measures that reflected reading after the region containing *mostly*. The processing profile of the Experiment 3 sentences does not appear to be regulated by principles specific to quantificationally variable items, but rather by the avoidance of postulating potentially unnecessary times. In short, the interpretation of the sentences was predicted by NET (Majewski, 2006, in preparation), a principle originally proposed in order to explain the processing of reciprocal sentences, not by principles specific to resolving quantificational variability.

Natural language is endlessly surprising. A priori one might have thought that it would be odd to make a quantificational assertion without indicating even the type of domain being quantified over. But the present findings suggest that readers are not particularly disturbed by quantificational variability. Rather they compute an analysis without extra times unless there is evidence for those times. We do not view the NET generalisation as a response to quantificational variability in particular. Rather, it is a general conceptual economy principle which militates against assuming entities without evidence. In this sense, the findings of the various studies discussed here suggest that theories of language interpretation need not contain any special principles that apply only to the resolution of quantificationally variable sentences. Rather, the results of these studies suggest that normal semantic biases come into play and determine the processing and preferred interpretation of such sentences, as in Frazier et al.’s (2008) earlier study of *The dishes are mostly clean/dirty*, where only general preferences needed to be invoked to explain the processing profile of the experimental sentences.

The present research indicates that the presence of quantificational variability per se does not impose huge processing costs, as long as normal interpretive defaults can be maintained. Indications of processing difficulty, including long reading times and increased regressions to previous portions of text, were observed when a default interpretation supporting the most economical processing decisions had to be abandoned. Further, we suspect that longer reading times also obtained when readers encountered a tension between the preferred or default reading and the choice of referring term used by the speaker, for instance, when the subject in the Neutral condition was phrased as a singular collective even though a salient alternative invoking parts was available. There is no reason to believe that the principles guiding the interpretation of the quantificational domain should be uniquely tied to quantification itself; general principles are enough.

That the NET principle should show powerful effects indicates, we think, how fundamental temporal structure is in determining the interpretation of the constituents of a sentence. How the constraints imposed by individual words and phrases can be satisfied depends in large part on whether what is being described is a single time/occasion or multiple times/occasions. In effect, the importance of the NET principle in characterising the processing of sentences containing *mostly* is just a reflex, we think, of the importance of this principle to constraining ongoing sentence interpretation in general.

Temporal processing has not been investigated extensively (though see Bestgen & Vonk, 1995; Bott, 2010; Coll-Florit & Gennari, 2011; Dickey, 2001). Compared to referential processing (Crain & Steedman, 1985 and many subsequent studies), attention to temporal reference has been limited. Yet temporal structure is essential for shaping what part of the world is taken to be under discussion, for example, whether it is one occasion or more. While it is generally assumed that discourse participants are not postulated without evidence, or multiplied without need, no comparable temporal assumption had been explicitly defended, or played a causal role in an analysis, until the work of Majewski (2006). Yet referential and temporal domains are tightly linked. Consider a multiply quantified sentence, as in *Every boy threw a stone*. If one occasion is involved, then necessarily multiple stones are involved. But if distinct occasions are involved, then it might be the same stone on each occasion. These kinds of examples illustrate the mutually constraining nature of constraints on the referential domain and constraints on the temporal domain. The typical discussion of preferred quantifier scope in scopally ambiguous sentences simply presupposes the operation of a principle like NET, and discusses the relative scope of an existential and a universal quantifier, on the assumption that one time is involved. We hope the current study will help this type of assumption to be identified explicitly, permitting a fuller more detailed picture to emerge of the role of temporal structure in natural language understanding.

Manuscript received 7 September 2011

Revised manuscript received 19 March 2012

First published online 28 August 2012

REFERENCES

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Bach, E., Jelinek, E., Kratzer, A., & Partee, B. (1995). *Quantification in natural languages*. Dordrecht: Kluwer Academic.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods*, 39, 445–459.

- Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge: MIT Press.
- Bates, D., & Maechler, M. (2009). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375–31. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Berman, S. (1987). Situation-based semantics for adverbs of quantification. In J. Blevins & A. Vainikka (Eds.), *University of Massachusetts Occasional Papers 12*, GLSA, University of Massachusetts, Amherst.
- Bestgen, Y., & Vonk, W. (1995). Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language*, 47, 74–87.
- Bott, O. (2010). *The processing of events*. Linguistics Today 162. Amsterdam: John Benjamins Publishing Company.
- Clifton, C., Jr., & Staub, A. (2008). Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass*, 2, 234–250.
- Clifton, C. Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. V. Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), *Eye movement research: Insights into mind and brain* (pp. 341–371). New York: Elsevier.
- Coll-Florit, M., & Gennari, S. P. (2011). Time in language: Event duration in sentence comprehension. *Cognitive Psychology*, 62, 41–79.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 320–358). Cambridge: Cambridge University Press.
- Crawley, M. J. (2007). *The R Book*. West Sussex: Wiley.
- Dickey, M. W. (2001). *The processing of Tense: Psycholinguistic Studies on the interpretation of tense and temporal relations*. Dordrecht: Kluwer Academic.
- Fiengo, R., & Lasnik, H. (1973). The logical structure of reciprocal sentences in English. *Foundations of Language*, 9, 447–468.
- von Fintel, K. (1994). *Restrictions on quantifier domains* (Doctoral dissertation), University of Massachusetts Amherst.
- Frazier, L., & Clifton, C. Jr. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., Clifton, C. Jr., & Stolterfoht, B. (2008). Scale structure: Processing minimum standard and maximum standard adjectives. *Cognition*, 106, 299–324.
- Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass*, 3, 111–127.
- Hinterwimmer, S. (2008). *Q-adverbs as selective binders: The quantificational variability of free relatives and definite DPs*. Berlin: Mouton de Gruyter.
- Jaeger, T. F. (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. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge: Cambridge University Press.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81, 345–381.
- Lewis, D. (1975). Adverbs of quantification. In E. Keenan (Ed.), *Formal semantics of natural language* (pp. 3–15). Cambridge: Cambridge University Press.
- Majewski, H. (2006, March). *Reciprocals, processing and event structure*. Poster presented at the Nineteenth Annual CUNY Conference on Human Sentence Processing. New York.
- Majewski, H. (In preparation). *Reciprocals and processing* (Doctoral dissertation), University of Massachusetts Amherst.
- R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: <http://www.R-project.org/>. ISBN 3-900051-07-0
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effects of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology*, 53A, 1061–1080.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher order modifiers. *Natural Language Semantics*, 12, 259–288.
- Westerståhl, D. (1984). Determiners and context sets. In J. F. A. K. van Bentham & A. ter Meulen (Eds.), *Generalized quantifiers in natural language* (pp. 45–71). Dordrecht: Foris.

Appendix A

Materials used in Experiment 1

Experiment 1: The full sextet is only displayed in the first two examples, while the rest show the a and b forms, from which the remaining conditions can be derived. Experiment 1 used the following seven point rating scale:

How natural was the second sentence? Totally unnatural..... Perfectly natural

1.
 - a. The students were mostly here. Some of them weren't however.
 - b. The students were mostly here. Some of the time they weren't however.
 - c. Most of the students were here. Some of them weren't however.
 - d. Most of the students were here. Some of the time they weren't however.
 - e. Most of the time the students were here. Some of them weren't however.
 - f. Most of the time the students were here. Some of the time they weren't however.
2.
 - a. The teachers were mostly in the cafeteria. A few of them weren't however.
 - b. The teachers were mostly in the cafeteria. A few times they weren't however.
 - c. Most of the teachers were in the cafeteria. A few of them weren't however.
 - d. Most of the teachers were in the cafeteria. A few times they weren't however.
 - e. Most of the time the teachers were in the cafeteria. A few of them weren't however.
 - f. Most of the time the teachers were in the cafeteria. A few times they weren't however.
3.
 - a. The actors were mostly in the theatre. Several of them boycotted rehearsal due to exhaustion.
 - b. The actors were mostly in the theatre. Several times they boycotted rehearsal due to exhaustion.
4.
 - a. The nurses were mostly at the front desk. Several of them were on break.
 - b. The nurses were mostly at the front desk. Several times they were on break.
5.
 - a. The childcare workers were mostly in the nursery. A couple of them were in the yard however.
 - b. The childcare workers were mostly in the nursery. A couple of times they were in the yard however.
6.
 - a. The repairmen were mostly out of the store. Some of them stayed in the workshop however.
 - b. The repairmen were mostly out of the store. Some of the time they stayed in the workshop however.
7.
 - a. The supervisors were mostly in the front office. A few of them entered the factory.
 - b. The supervisors were mostly in the front office. A few times they entered the factory.
8.
 - a. The pharmacists were mostly working. Several of them were chatting with the customers.
 - b. The pharmacists were mostly working. Several times they were chatting with the customers.
9.
 - a. The guests were mostly drunk. A few of them were sober though.
 - b. The guests were mostly drunk. A few times they were sober though.
10.
 - a. The accountants were mostly tired. A couple of them weren't.
 - b. The accountants were mostly tired. A couple of times they weren't.
11.
 - a. The waiters were mostly well-informed. Some of them were clueless.
 - b. The waiters were mostly well-informed. Some of the time they were clueless.
12.
 - a. The customers mostly paid. A couple of them defaulted.
 - b. The customers mostly paid. A couple of times they defaulted.
13.
 - a. The prisoners mostly behaved. Several of them didn't.
 - b. The prisoners mostly behaved. Several times they didn't.
14.
 - a. The comedians were mostly funny. Some of them weren't.
 - b. The comedians were mostly funny. Some of the time they weren't.
15.
 - a. The agents were mostly polite. A few of them weren't.
 - b. The agents were mostly polite. A few times they weren't.
16.
 - a. The mechanics were mostly in the garage. A couple of them were in the junkyard.
 - b. The mechanics were mostly in the garage. A couple of times they were in the junkyard.
17.
 - a. The farmers were mostly in the field. Several of them were in the barn milking cows.
 - b. The farmers were mostly in the field. Several times they were in the barn milking cows.
18.
 - a. The fishermen were mostly on the lake. Some of them were back at the camp tying lures.
 - b. The fishermen were mostly on the lake. Some of the time they were back at the camp tying lures.
19.
 - a. The monks were mostly in the abbey. A few of them were in the village collecting alms.
 - b. The monks were mostly in the abbey. A few times they were in the village collecting alms.
20.
 - a. The lions were mostly in the den. A couple of them went in the tall grass to hunt.
 - b. The lions were mostly in the den. A couple of times they went in the tall grass to hunt.
21.
 - a. The athletes were mostly training. A couple of them watched sports on TV.
 - b. The athletes were mostly training. A couple of times they watched sports on TV.
22.
 - a. The bartenders were mostly friendly. Several of them were a little curt.
 - b. The bartenders were mostly friendly. Several times they were a little curt.
23.
 - a. The students were mostly bored. Some of them were engaged in the topic.
 - b. The students were mostly bored. Some of the time they were engaged in the topic.

24. a. The librarians were mostly shelving books. A few of them met to discuss checkout policies.
- b. The librarians were mostly shelving books. A few times they met to discuss checkout policies.

Appendix B

Materials from Experiment 2a. Experiment 2b sentences were identical except two additional conditions were created as past tense variants.

1. The students | Students mostly walk to school.
2. The teachers | Teachers mostly take the extra training sessions.
3. The administrators | Administrators mostly apply for summer leave.
4. The employees | Employees mostly leave early on Friday.
5. The skaters | Skaters here mostly compete at the state level.
6. The drivers | Drivers mostly behave courteously around bicyclists.
7. The cooks | Cooks mostly enjoy cooking.
8. The architects | Architects mostly take classes in engineering.
9. The botanists | Botanists mostly study common plants.
10. The comedians | Comedians mostly perform in clubs.
11. The custodians | Custodians mostly stay indoors.
12. The local potters | Local potters mostly show in Leverett.
13. The amateur guitarists | Amateur guitarists mostly perform at the Black Sheep.
14. The foreigners | Foreigners mostly take the bus in Manhattan.
15. The Amherst policemen | Amherst policemen mostly investigate noise complaints.
16. The mechanics | Mechanics mostly work on combustion engines.
17. The pilots | Pilots mostly fly commercial planes.
18. The receptionists | Receptionists mostly speak English.
19. The contractors | Contractors mostly value their reputations.
20. The repairmen | Repairmen mostly know what they are doing.

Appendix C

Materials used in Experiment 3: Conditions are displayed in the following order: Neutral | Divisible | Nondivisible.

1. The army | attack | inspector | was mostly in the capital, according to the newspapers.
2. The committee | festival | CEO | was mostly in town, according to the press release.
3. The mountain trek | mountain range | mountain man | was mostly in Colorado, where the views are spectacular.
4. The jury | jail | judge | was mostly in the government compound, which was usually safe.
5. The business advertising | opening ceremony | business leader | was mostly in the city centre, where the transit lines meet.
6. The convention | campaign kickoff | governor | was mostly in Seattle, where it rains constantly.
7. The platoon | fire | platoon leader | was mostly in the barracks, where the new soldiers lived.
8. The conference | riot | organizer | was mostly in the campus centre, where there are many students.
9. The condo development | condo subdivision | condo developer | was mostly in Hadley, where there is a big shopping mall.
10. The lab work | lab presentation | lab instructor | was mostly in Morrill Hall, which is on North Pleasant.
11. The police force | police blockade | policeman | was mostly on campus, according to the radio report.
12. The news crew | news cast | news anchor | was mostly in the station, where there was a big commotion.
13. The dance class | dance audition | dance professor | was mostly in South College, according to the department website.
14. The demonstration | inauguration | city manager | was mostly in Town Hall, which didn't surprise anyone.
15. The wrestling squad | wrestling tournament | wrestler | was mostly in Boyden Gym, where the air smelled awful.
16. The soccer team | flood | team mascot | was mostly in Amherst, where Amherst College is located.
17. The herd | invasion herdsman | | was mostly in northern Greenland, where it is extremely cold.

18. The safari | coal mine | coal miner | was mostly in the jungle, which surprised most people.
19. The expedition | route | explorer | was mostly in East Africa, where the weather was very dry.
20. The gang | showdown | ex-con | was mostly in the south Bronx, according to the morning news.
21. The guided tour | sculpture garden | tour guide | was mostly in the Smithsonian, where many tourists gathered.
22. The hurricane | damage | rescue worker | was mostly in South Florida, where insurance rates are high.
23. The corruption | earthquake | mayor | was mostly in San Francisco, according to the city spokesman.
24. The fishing trip | fish farm | fisherman | was mostly in Lake Ontario, where the water is cold.