Lexicalization of quantificational forces in adverbial and determiner domains

Anonymous CogSci submission

Abstract

Which quantificational forces do languages encode lexically? When a language features multiple quantificational scales (e.g. determiner and adverbial quantification), does the pattern of lexicalization of quantificational forces we discover for one scale correlate with those of other scales? We use English as a first test case for examining these questions, adapting the basic ideas of Lewis (1975) into the hypothesis that English lexical quantifiers unrelated to cardinal numbers or definite descriptions, determiner and adverbial alike, have one of six quantificational forces. To begin to test this claim empirically, we elicited speaker interpretations of a range of quantifiers in a web-based study. Dividing participants into an adverbial condition and a determiner condition, we gave a context specifying a 100-day period and asked participants to judge the quantificational force of quantified sentences denoting an individual’s daily activities during this period. We found evidence of cross-scale correspondences but fewer quantificational forces than expected. These results provide preliminary evidence for parts of our hypothesis but suggest a need for future research that covers more lexical items, languages, and quantificational scales.

**Keywords:** adverbs of quantification; quantifiers; Q-adverbs

# Introduction

Natural-language quantification—the use of linguistic expressions to make judgments of quantity—is one of the most well-studied phenomena in semantics and syntax. Since the development of Aristotelian logic, the literature on quantification has primarily focused on determiners that quantify over entities, such as “every,” “some,” and “no”; in modern generalized quantifier theory, such quantifiers denote relations between sets (see Barwise and Cooper, 1981). However, many languages, including English, additionally feature adverbial quantification; adverbs of quantification like “always,” “sometimes,” and “never,” also called Q-adverbs (Hinterwimmer, 2008), are variously taken to quantify over “cases”—ordered tuples of “admissible assignments of values” to a sentences’ free variables (Lewis, 1975; see also Kamp, 1981; Heim, 1982; Kratzer, 1989)—or over situations or events (de Swart, 1993; von Fintel, 1994; Hinterwimmer, 2008).

A central property of both determiner and adverbial quantifiers is quantificational force, which, in our use of the term, relates to how a quantifier resolves questions like “how many?” or “how much?” For instance, the quantifiers “every” and “always” in (1) have universal force: for each member of the set of days (1a) or for each “case” or situation (1b), Bill brushed his teeth before bed. Conversely, “no” and “never” in (2) have negative existential force—for no days (2a) or in no “cases” or situations (2b) did Bill arrive to work on time.

(1) a. Bill brushed his teeth before bed every day.

b. Bill always brushed his teeth before bed.

(2) a. Bill arrived to work on time no days.

b. Bill never arrived to work on time.

The universal and negative existential quantificational forces displayed in (1) and (2) represent the extremes of a spectrum of possible forces; for instance, quantifiers like “some” and “most” have quantificational forces that lie between those of the two extremes.

The recognition of this spectrum and of the existence of multiple quantificational scales (determiner, adverbial, etc.) leads to two interrelated questions: (1) Which quantificational forces do languages encode lexically? (2) When a language features multiple quantificational scales, does the pattern of lexicalization we discover for one scale correlate with those of other scales?

We consider it fruitful to examine these questions for two main reasons. First, these questions interact with recent research concerning optimality of lexically encoding quantificational meaning in language. Steinert-Threlkeld, (2019) argued that with respect to their quantifiers, natural-languages strike an optimal balance between simplicity and informativeness. Although Steinert-Threlkeld (2019) substantiated this claim with an analysis that relied on the logical denotations of quantifiers as relations between sets (p. 514), similar issues can be explored when we think about lexicalization of quantificational forces. For instance, in examining question (1), we might find that natural-language quantifier systems tend balance the imperatives of simplicity (few, simple lexicalizations of forces) and informativeness or expressiveness (many fine-grained lexicalizations of forces). Furthermore, an affirmative answer to question (2) could indicate the presence of an optimality constraint that prevents quantificational scales from differing dramatically in what forces they encode.

Second, the literature on the formal semantics of quantificational adverbs has assumed a system of truth-conditional correspondences between adverbs and determiners (e.g. between “often” and “many”) without subjecting that assumption to rigorous empirical testing (e.g. Lewis, 1975; de Swart, 1993; von Fintel, 1994). They appear to make this assumption merely based on intuition; empirical corroboration is needed to determine its soundness.

Before we turn to hypotheses, we make some notes about how we will approach our research questions in what follows. For the purposes of this paper, we will restrict ourselves to discussing these questions as they relate to English and its determiner and adverbial scales of quantification, though other languages and other types of quantification (such as the class of “frequency adjectives” discussed in, e.g., Gehrke and McNally, 2015) should come under scrutiny in future research on these questions. Additionally, our use of the term “quantificational force” restricts us to considering how speakers judge quantity in quantified sentences; no logical properties of quantifiers, such as domain restriction or polarity, will factor into our discussion.

We will treat Keenan (1996)’s inventory of lexical determiners and Lewis (1975)’s list of adverbial quantifiers as our master list of English lexical quantifiers; let us make some preliminary observations about these two lists and exclusions of certain words from consideration. First, Keenan (1996) includes many definite determiners in his list, such as “the,” “this,” and “John’s”; they have no adverbial counterparts in terms of force, but given that the status of definite determiners as quantifiers is controversial, we will exclude them from discussion altogether. We will also exclude the indefinite article because of its unique semantic properties (see, e.g., Heim 1982). Second, Keenan (1996) rightly includes cardinal numbers, but we consider it fair to posit *a priori* conclusions about them and the limited number of adverbial counterparts to cardinal numbers that exist; thus, we will only make passing mention of them in our hypotheses and discuss them in more detail in the Discussion section.

## Hypotheses

With these considerations and exclusions in mind, we turn to the literature and find two claims in Lewis (1975)’s paper on the formal semantics of adverbial quantifiers that can be adapted into working hypotheses for our two questions as they relate to English. First, Lewis (1975) posits a typology of adverbs of quantification that divides them into “six groups of near-synonyms” (p. 5). Though he does not elaborate on how the words are synonymous, we will assume a version of the typology that describes the tiers in terms of quantificational force. Here is our statement of the typology:

Hypothesis 1 (modification of Lewis 1975)[[1]](#footnote-2): In the adverbial scale of quantification, English lexically encodes six quantificational forces in addition to the three related to cardinal numbers (those embedded in “once,” “twice,” and “thrice”): I. the universal force of “always,” “invariably,” “universally”; II. the majority force of “usually,” “mostly,” “generally,” “ordinarily,” “normally”; III. the positive proportional force of “often,” “frequently,” and “commonly”; IV. the existential force of “sometimes” and “occasionally”; V. the negative proportional force of “seldom,” “infrequently,” and “rarely”; VI. the negative existential force of “never.”

Second, Lewis (1975), like other scholars of adverbial quantifiers after him (de Swart, 1993; von Fintel, 1994) posits truth-conditional correspondences between the lexical adverbs in the six tiers of his typology and lexical determiners; framing the correspondence in terms of “selective” and “unselective” quantifiers, he writes, “the unselective ∀ and ∃ can show up as the adverbs *always* and *sometimes*. Likewise *never*, *usually*, *often*, and *seldom* can serve as the unselective analogs of the selective quantifiers *for no x, for most x, for* *many x,* and *for few x.*” (p. 10) For our purposes, we treat such correspondences solely in terms of quantificational force and extend the correspondences to include the near-synonyms of “always,” “sometimes,” “never,” “usually,” “often,” and “seldom” listed in the typology.

This claim naturally leads to Hypothesis 2 below. Note the inclusion, per the Keenan (1996) list, of two additional universal quantifiers (“every" and “each”) as well as “several,” a determiner that we hypothesize occupies a tier with “some.”

Hypothesis 2: Excluding the quantificational forces embedded in cardinal numbers as well as those encoded in definite determiners, English lexically encodes six quantificational forces in the determiner scale of quantification, and they correspond to those of the adverbial scale: I. the universal force of “every,” “each,” “all,” and “both”; II. the majority force of “most”; III. the positive proportional force of “many”; IV. the existential force of “some" and “several”; V. the negative proportional force of “few”; VI. the negative existential force of “no” and “neither.”

Lewis (1975) posits his system of correspondences *a priori* and likely did so based on his own intuitions about these words, but we can give a more sophisticated motivation of our hypotheses 1 and 2 in terms of the balance between simplicity and informativeness. With regards to determiners (Hypothesis 2), it has been shown that the quantifiers in the Square of Opposition (some, no, every/each/all) have the least complex possible determiner meanings (van Benthem, 1986); thus, it makes sense that English would allocate forces to these three quantifier types (I., IV., and VI.), as they are both simple and informative. The inclusion of any additional forces would sacrifice simplicity for the sake of informativeness on two counts. First, the quantifiers in which these additional forces would be embedded (e.g. “most” in English) would be of a higher semantic complexity than those in the Square of Opposition. Second, the inclusion of more forces makes the quantifier system of a language more complex. However, these new quantifiers and forces also make a language more expressive and informative. Given these tradeoffs, it is plausible to posit a small number of additional lexicalized forces for determiners that cover wide swaths of the spectrum not covered by forces I., IV., and VI. Clusters II., III., and V. naturally suggest themselves. We can further motivate our hypothesis about adverbs (Hypothesis 1) by suggesting that the optimality constraints at work in the determiner scale are also operative in the adverbial scale.

To test Hypotheses 1 and 2, we elicited speaker judgments of the quantificational forces of a range of determiner and adverbial quantifiers in a web-based study; while this study did not test every English lexical quantifier, we designed it to provide a first test of the claim that the quantifiers under consideration form six distinct interpretive clusters.

# Methods

## Participants

In this study, participants (N = 200; 88 of these ultimately excluded—see “Results”) were recruited via the crowdsourcing platform Amazon Mechanical Turk (MTurk). After two introductory slides, participants were randomly assigned to an adverbial (N = 100 pre-exclusions/56 post-exclusions) or determiner condition (N = 100 pre-exclusions/56 post-exclusions). Regardless of the inclusion or exclusion of their data in analysis, participants received $1.50 for their time. Data were collected between September 21 and 22, 2019.

## Words tested

In order to provide a preliminary test of the system of correspondences laid out in Hypotheses 1 and 2 while keeping the survey short enough that participants would not lose focus, we excluded several lexical items judged to be potentially problematic with regards to empirical testing. First, we excluded “both" and “neither,” for these differ from the other words in terms of their presuppositions—they both presuppose that the number of items under consideration is two—and are thus not conducive to inclusion in an experimental framework geared towards those other quantifiers. Second, we excluded “invariably” and “universally,” for these are quite formal and might thus confuse participants who do not have these words in their working vocabularies. Third, we excluded “ordinarily” and “normally,” two words that Lewis (1975) flags as “[differing] semantically from their list-mates”; he exhorts readers to “omit them if [they] prefer” (p. 5). Finally, we excluded “commonly” and “generally" in the interest of keeping this preliminary study short. These exclusions leave us with the lexical items listed in Table 1, arranged according to their hypothesized tiers.

## Materials

The study, created on Qualtrics, consisted of a consent form, an introductory slide, several trials, and two “debrief” questions about age and native language. In the introductory portion of the study, participants were presented with the following context:

Table 1: Lexical items tested.

| Hypothesized Tier | Adverbs | Determiners |
| --- | --- | --- |
| Tier 1 | Always | Every, Each, All |
| Tier 2 | Usually,  Mostly | Most |
| Tier 3 | Often,  Frequently | Many |
| Tier 4 | Sometimes,  Occasionally | Some,  Several |
| Tier 5 | Seldom,  Rarely,  Infrequently | Few  A few |
| Tier 6 | Never | No |

“Bill is a college student who decided to stay home over summer vacation, which lasts for 100 days. Bill got a job at his local ice cream shop and has a single shift every day, meaning that Bill travels to and from work only once a day. Bill is on a diet and is limiting himself to three meals a day.”

After reading this context, participants proceeded to the bulk of the survey, in which they read quantified sentences denoting Bill’s daily activities and were asked to indicate the number of days during the 100-day period they believed the event had occurred. Depending on the condition to which they had been randomly assigned, participants either judged 10 sentences involving determiners, such as (2–3), or 11 sentences involving adverbs, such as (4–5).

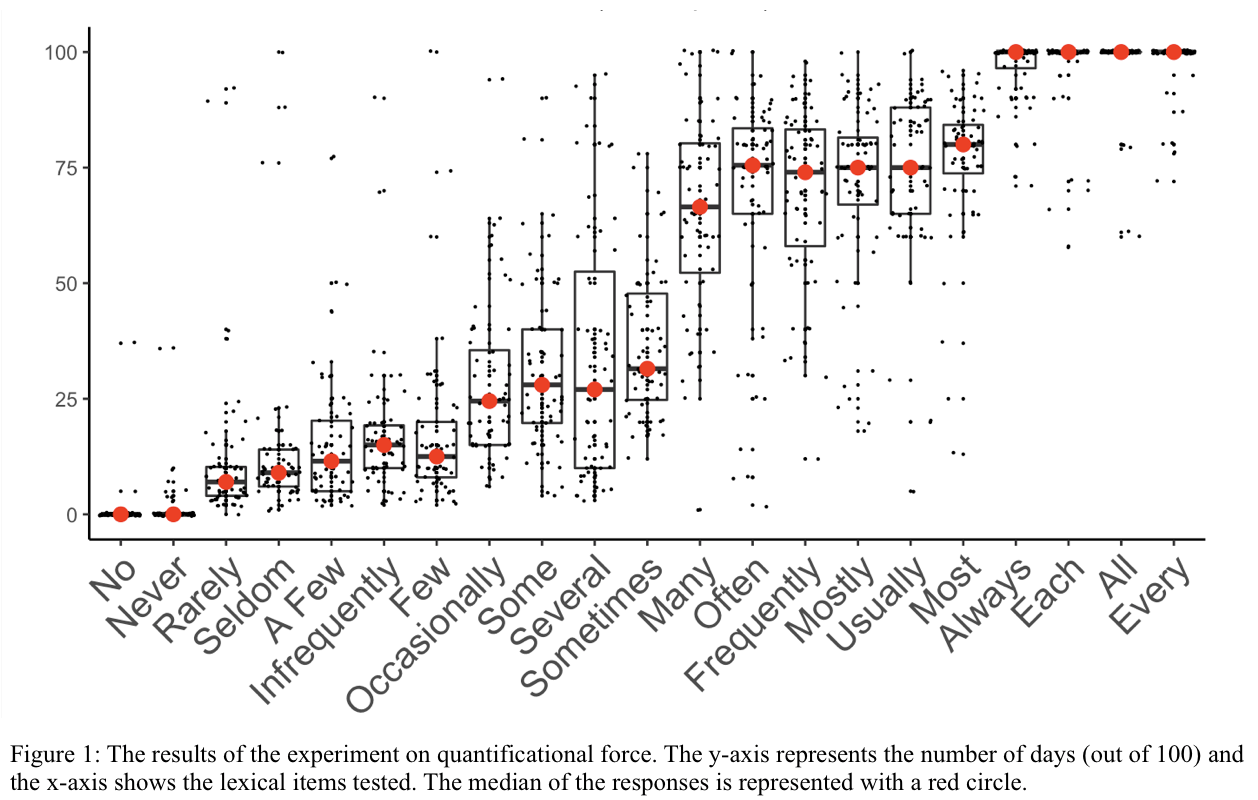
(2) Bill drove to work some days.

(3) Bill went to the grocery store after work every day.

(4) Bill mostly biked home from work.

(5) Bill seldom ate pizza for dinner.

In this way, we tested correspondences between certain adverbs and certain determiners with a between-subjects design and tested correspondences within each lexical class using a within-subjects design. Participants indicated their numerical interpretation of these sentences using a slider, which rested at the midpoint of the scale at the beginning of each question. Participants had to answer each question before proceeding.

In order to control for the effects of the sentential context in which participants made judgments of quantificational force, we implemented a system of randomization. We had a pool of eleven scenarios (e.g. Bill brushing his teeth before bed, arriving late to work, eating eggs for breakfast) and we coded questions with every possible combination of quantifier and scenario. In the study, participants were presented with one sentence for each member of the relevant class of quantifier (determiner or adverbial, depending on the condition). We randomized the order in which the quantifiers appeared; the scenarios that accompanied these quantifiers were also randomized, though we made sure that no scenario appeared twice in order to minimize participant confusion at seeing, say, “Bill drove to work every day” and “Bill drove to work no days.”

# Results

Figure 1 shows the results of our experiment after exclusions. As there were 56 participants per condition post-exclusions and each participant was presented with one sentence per quantifier, each quantifier received 56 interpretations.

Our exclusions, designed to restrict analyzed data to those collected from attentive, adult native speakers of English, were based on four criteria. First, two attention checks involving movement of a slider to a specified number were randomly placed throughout the survey; if participants failed either one, their results were discarded. Three participants’ data were excluded based on this criterion. Third, we included a question about native language immediately after the question about age and excluded all who reported a native language other than English. Three participants were excluded based on this criterion. Finally, we excluded all participants who interpreted a universally quantified sentence like (1a) with a response between 0 and 50 days out of 100 and those who interpreted a negative existentially quantified sentence like (2b) with a response between 50 and 100 days out of 100; we considered such responses to be indicative of insufficient attention or lack of understanding of the task. 82 participants were excluded based on this criterion.

# Analysis

We used hierarchical cluster analysis on participants’ responses to discover the lexical items that form tiers together with respect to quantificational force. In contrast to other common clustering techniques such as k-means clustering, hierarchical clustering has the advantage that the number of clusters (or tiers) is not predetermined. In addition, the clustering provides a tree-like structure called “dendrogram” which is easy to interpret. We used the *hclust* function from the “cluster” package in R with the “complete” linkage method to find maximally similar clusters of response patterns among the quantificational lexical items.

Figure 2 shows the dendrogram created after applying the hierarchical clustering algorithm. The most important aspect of the dendrogram is the height shown on the y-axis. The height of the dendrogram represents similarity among observations, with more similar observations fusing lower on the dendrogram and more dissimilar observations fusing higher up. To put this in the context of our experiment, lexical items that fuse at the bottom are quite similar to each other with respect to quantificational force and those that fuse at the top are quite different.

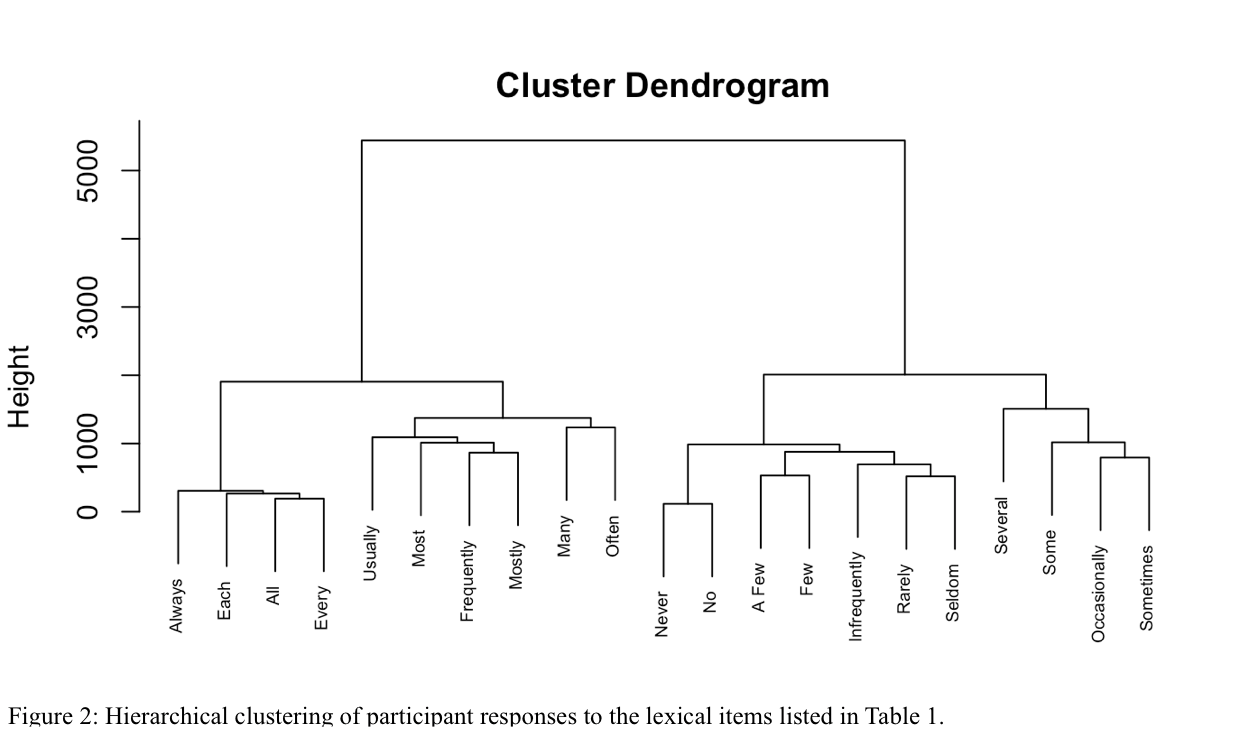
In order to identify clusters based on the dendrogram, we can make horizontal cuts at different levels of height. For example, cutting the dendrogram with a horizontal line at around the height of 3000 results in two overall clusters. The first cluster contains *always, each, all, every, mostly, most, frequently, usually, many, and often*, while the second cluster contains *never*, *no*, *a few, few,* *infrequently*, *rarely*, *seldom*, *several*, *some*, *occasionally*, and *sometimes*. This overall division corresponds to a “majority” vs. “minority” quantificational force. Cutting the dendrogram at the height of slightly below 2000 we get four clusters:

1. *always*, *all*, *each*, *every*
2. *mostly*, *most*, *frequently*, *usually, many, often*
3. *several*, *some*, *occasionally*, *sometimes*
4. *no*, *never*, *infrequently*, *rarely*, *seldom*, *few*, *a few*

Note that our use of universally and negative-existentially quantified sentences as attention checks limits our ability to make empirically supported conclusions about these quantifiers based on the cluster analysis. However, we consider the quantificational forces of these words to be *a priori* incontrovertible, which is why we chose them as attention checks.

# Discussion

These results preliminarily indicate the existence of correlations in quantificational force between the determiner and adverbial lexical quantifiers of English; each major cluster contained both adverbs and determiners. Because of this result, we tentatively find that English functions as a first piece of evidence for a positive answer to our second overarching question (“When a language features multiple quantificational scales, does the pattern of lexicalization we discover for one scale correlate with those of other scales?”)

However, the number of meaningfully distinct quantificational forces we discovered in English was fewer than hypothesized. Most tellingly, the forces embedded in “most,” “mostly,” and “usually” were not as distinct from the forces encoded in “many,” “often,” and “frequently” as expected. While “many” and “often” did branch off from the other four quantifiers at a low level of height, the force of “frequently” was inextricably tied to those of “most,” “mostly,” and “usually.” In our estimation, the low-level cluster involving “many” and “often” represents not a distinct quantificational force but a minor (if interesting) deviation. Indeed, our four main clusters appear to be much more stable than any lower-level clusters. As an additional hierarchical cluster analysis that included data from the participants who gave implausible interpretations of universally and negative existentially quantified sentences (see above) showed, this “many-often” cluster along with other lower-level clusters were contingent on our choice to exclude these participants, while our four overarching clusters were not. In this second analysis with the extreme outliers added in, we still found our four overarching clusters, but the branchings below that differed from those in figure 2. Though this is certainly not conclusive evidence against the claim that there may be distinct quantificational forces at lower levels of height, it does suggest the volatility of such “forces” in our experiment; for “many-often” to be considered a distinct force, we would certainly need further experimental corroboration.

The analysis also suggests an conflation between negative existential force and negative proportional force, but we suspect that this result is due to the fact that interpretations of quantifiers like “few” and “seldom” often approached zero, an extreme of our scale where negative existentials also naturally found themselves. Given this fact and our exclusions of certain implausible interpretations of negative existentials from analysis, we do not interpret the cluster analysis as supporting a conclusion about the relationship between the forces of negative proportionals and negative existentials.

As regards our first overarching question (“Which quantificational forces do languages encode lexically?”), our findings of fewer than six forces of course represent merely a first piece of evidence rather than a definitive answer. However, they constitute a starting point for future cross-linguistic inquiry; in such endeavors, the potential identity of positive proportional and majority force will warrant particular attention.

If reinforced by similar experiments, our results could hint at some form of optimality constraint of the sort described above, a constraint that causes the determiner and adverbial scales of quantification to parallel each other. Although we did not find the six quantificational forces predicted in our Hypotheses 1 and 2, the discovery of fewer forces actually strengthens the case that an optimality consideration is at play; the quantifier system of English lexicalizes few forces in the name of simplicity but extends beyond “all/always,” “some/sometimes,” and “no/never” in the name of informativeness. Despite this preliminary evidence of some sort of optimality constraint, though, future research and theoretical inquiry is needed to pinpoint the exact nature and scope of such a constraint.

Additionally, our results suggest that four or five quantificational forces in English not related to cardinal numbers or definite determiners are at a level of salience high enough that they warrant lexical encoding.

However, it is important to note that despite this evidence of correlations between certain determiner and adverbial lexical quantifiers of English, the system of correlations does not extend to all English lexical quantifiers. In particular, only four cardinal numbers have widely used adverbial counterparts. The four pairs are as follows: “zero”/“never,” “one”/“once”, “two”/“twice,” and “three”/“thrice.” “Thrice” has started to fall out of common usage, so the presence of the fourth correlation in the language of most speakers is dubious. Adverbial correlates for higher cardinal numbers, like “quadruply” or “quintuply,” are also marginal at best. In any case, there is no English lexical item that functions as an adverbial correlate of, say, 322. Note that the presence of an infinitely complex system of “forces” embedded in cardinal numbers does not nullify the claim that optimality considerations factor into a language’s system of cardinality-independent quantifiers.

Much further research on the similarities and differences between the determiner and adverbial scales remains. One outstanding question, raised by an anonymous reviewer, concerns the possibility that the adverbial scale of quantification has a more scalar representation than the determiner scale. Indeed, more gradable adverbs are found in the adverbial domain (“[very] seldom,” “[somewhat] infrequently,” “[quite] rarely,” “[extremely] often,” “[rather] frequently,” etc.) than in the determiner domain (seemingly only “[somewhat] many” and “[rather] few.”) It would be interesting to determine what effect this surface-level difference has on how speakers use the two types of quantifier, whether this trend holds cross-linguistically, and why it does if so.

In conclusion, given the scope of our guiding questions, our findings of cross-scale correspondences between English lexical quantifiers and fewer than six quantificational forces are partial. To more firmly establish the answers to our questions, we would need not only to rerun the experiment but also to run similar experiments with more quantifiers, more languages, and more quantificational scales. Despite the preliminary nature of this study, however, it lays the foundation for a far-reaching research agenda with the potential to make exciting findings.

# References

Barwise, J., & Cooper, R. (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, 4, 159–219.

van Benthem, J. (1986). Essays in Logical Semantics. Dordrecht, Netherlands: D. Reidel.

von Fintel, K. (1994). *Restrictions on Quantifier Domains.* PhD Thesis, University of Massachusetts, Amherst.

Gehrke, B., & McNally, L. Distributional Modification: The Case of Frequency Adjectives. *Language*, 91(4), 837-870.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases.* PhD Thesis, University of Massachusetts, Amherst.

Hinterwimmer, S. (2008). *Q-Adverbs as Selective Binders: The Quantificational Variability of Free Relatives and Definite DPs.* Berlin, Germany: Mouton de Gruyter.

Kamp, H. (1981). A Theory of Truth and Semantic Representation. In J. Goenendijk, T. Jannsen and M. Stokhof (eds.), *Formal Methods in the Study of Language* (pp. 277-322). Amsterdam: Mathematisch Centrum.

Keenan, E. (1996). The Semantics of Determiners. In E. Lappin (Ed.), *The Handbook of Contemporary Semantic Theory* (pp. 41-63). Oxford, England: Blackwell.

Kratzer, A. (1989). Stage-level and Individual-Level Predicates. ms. University of Massachusetts, Amherst.

Lewis, D. (1975). Adverbs of Quantification. In *Papers in Philosophical Logic* (pp. 5-20). Cambridge, England: Cambridge University Press.

Steinert-Threlkeld, S. (2019). Quantifiers in Natural Language Optimize the Simplicity-Informativeness Trade-off. Proceedings of the 22nd Amsterdam Colloquium, 513–22.

de Swart, H. (1993). *Adverbs of Quantification: A Generalized Quantifier Approach.* Doctoral dissertation, Linguistics Department, University of Groningen, 1991.

1. In addition to Lewis (1975)’s lack of explicit mention of quantificational force, he does not concern himself with our question of lexicalization, so his typology includes some non-lexical quantifiers; these have been removed. Furthermore, “once” has been extricated from tier 4, for it relates to cardinal numbers. [↑](#footnote-ref-2)