Quantificational Force in Temporal and Nominal Domains

Anonymous CogSci submission

Abstract

Which quantificational forces do languages tend to 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 all English lexical quantifiers not related 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 an online study. Dividing participants into an adverbial condition and a determiner condition, we gave a context specifying a 100-day period and then provided various quantified sentences that denoted daily activities performed by an individual during this period; for each sentence, the participants judged the number of days they believed that the event denoted had taken place. [RESULTS] These results have psycholinguistic implications and suggests 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”. 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)—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 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), John 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 John arrive to work on time.

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

b. John always brushed his teeth before bed.

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

b. John 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, existential quantifiers like “some” have a quantificational force that lies 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? 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.

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, though we understand that lists of lexical items of this sort almost always suffer from lack of completeness; before making hypotheses related to our two questions above, 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 descriptions as quantifiers is controversial, we will exclude them from discussion altogether. [“A”?] 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. [“A FEW?”]

With these considerations and exclusions in mind, we turn to the literature and find two claims in Lewis (1975)’s seminal paper on adverbial quantifiers that can be adapted into working hypotheses for our two questions as they relate to English. First, Lewis 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, listed as Hypothesis 1:

(2) 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”): (1) the universal force of “always,” “invariably,” “universally”; (2) the majority force of “usually,” “mostly,” “generally,” “ordinarily,” “normally”; (3) the positive proportional force of “often,” “frequently,” and “commonly”; (4) the existential force of “sometimes” and “occasionally”; (5) the negative proportional force of “seldom,” “infrequently,” and “rarely”; (6) the negative existential force of “never.”

Second, Lewis (1975) posits truth-conditional correspondences between the lexical items in the six tiers of his typology and lexical determiners[[2]](#footnote-3); 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.”

(3) 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: (1) the universal force of “every,” “each,” “all,” and “both”; (2) the majority force of “most”; (3) the positive proportional force of “many”; (4) the existential force of “some" and “several”; (5) the negative proportional force of “Few”; (6) the negative existential force of “no” and “neither.”

While Lewis (1975) posits his complex system of correspondences *a priori*, these two hypotheses stand in need of empirical verification. To jumpstart this process of verification, we elicited speaker judgments of the quantificational forces of a range of determiner and adverbial quantifiers in an online 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; 6 of these ultimately excluded) were recruited via the crowdsourcing platform Amazon Mechanical Turk (MTurk). Data were collected between September 21 and 22, 2019. After two introductory slides, participants were randomly assigned to an adverbial (N = 97) or determiner condition (N = 97). In order to restrict analyzed data to those collected from attentive, adult native speakers of English, we developed three criteria for inclusion in analysis. 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. Second, we included two questions about age and native language, positioned at the end of the survey so as to minimize the incentive to lie. Six participants combined did not meet one of these criteria, so their data were discarded. However, they still received compensation ($1.50).

## 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 included 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 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 how do not have these words in their working vocabularies. Third, we excluded “ordinarily” and “normally,” two words that Lewis flags as “[differing] semantically from their list-mates”; he exhorts readers to “omit them if [they] prefer” (p. 5). [“COMMONLY” AND “GENERALLY”?]

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:

| 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 |
| Tier 6 | Never | No |

Table 1: Lexical items tested.

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 (5–6), or 11 sentences involving adverbs, such as (7–8).

(5) Bill drove to work some days.

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

(7) Bill mostly biked home from work.

(8) 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 complex 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, after participants were sorted into either the determiner or the adverbial condition, we presented them with one sentence for each member of the relevant class of quantifier, randomizing the order in which the quantifiers appeared; the scenarios that accompanied these quantifiers were also randomized, though we made sure that not 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

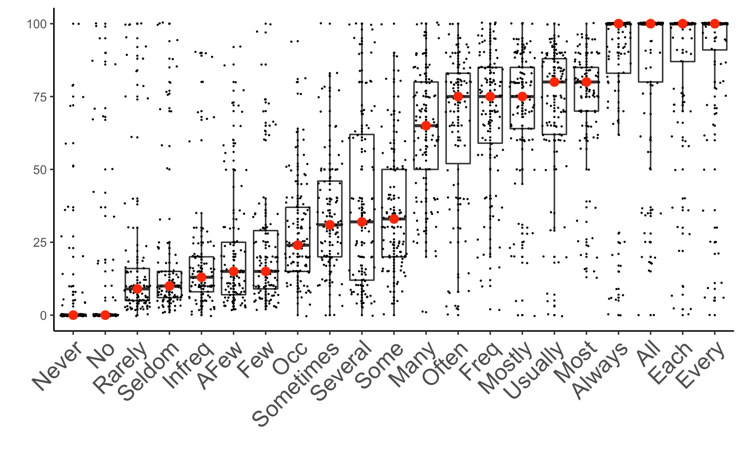


Figure 1: The results of the experiment on quantificational force. The y-axis represents the number of days (out of 100) and the x-axis shows the lexical items tested. The median of the responses is represented with a red circle.

Figure 1 shows the results of our experiment for all the tested quantificational items.

# Analysis

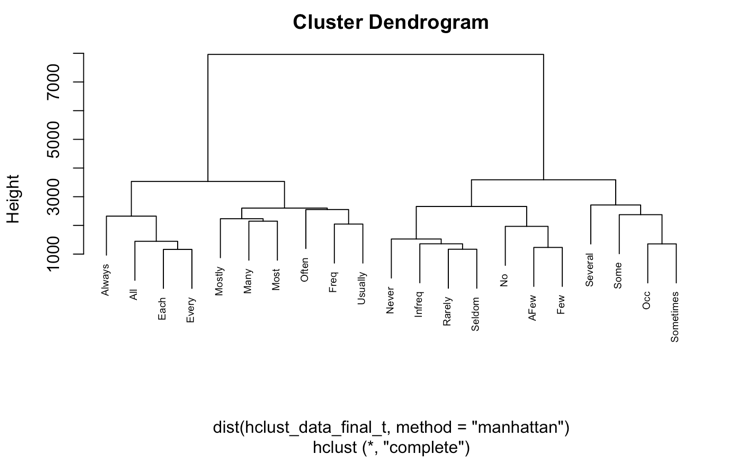


Figure 2: Hierarchical clustering of participant responses to the lexical items listed in Table 1

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 5000 results in two overall clusters. The first cluster contains *usually*, *frequently*, *often*, *most*, *many*, *mostly*, *every*, *each*, *all*, *always*, while the second cluster contains *no*, *never*, *infrequently*, *rarely*, *seldom*, *few*, *a few*, *several*, *some*, *occasionally*, and *sometimes*. This overall division corresponds to a “majority” vs. “minority” quantificational force. Cutting the dendrogram at the height of 3000 we get four clusters:

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

It is worth noting that the inclusion of negative quantifiers *no* and *never* in the fourth cluster along with quantifiers of low frequency such as *rarely* and *seldom* are most likely the artefact of noisy data provided by participants in the online study. We expect that reducing such noise will result in *no* and *never* cluster separately than lexical items such as *rarely* and *seldom*.

# Discussion

First level headings should be in 12 point, initial caps, bold and centered. Leave one line space above the heading and 1/4 line space below the heading.

# Acknowledgments

In the **initial submission**, please **do not include acknowledgements**, to preserve anonymity. In the **final submission**, **place acknowledgments** (including funding information) in a section **at the end of the paper**.

# References

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510-516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior.In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill.

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year old. *Cognition and Brain Theory*, *6*, 287-317.

Matlock, T. (2001). *How real is fictive motion?* Doctoral dissertation, Psychology Department, University of California, Santa Cruz.

Newell, A., & Simon, H. A. (1972)*. Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Ohlsson, S., & Langley, P. (1985). *Identifying solution paths in cognitive diagnosis* (Tech. Rep. CMU-RI-TR-85-2). Pittsburgh, PA: Carnegie Mellon University, The Robotics Institute.

Shrager, J., & Langley, P. (Eds.) (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.

1. In addition to Lewis’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)
2. This is not peculiar to Lewis (1975)—de Swart (1993), despite her dramatically differing account of how best to capture the meanings of adverbially quantified sentences, makes an identical assumption. [↑](#footnote-ref-3)