Lexicalization of quantificational forces in adverbial and determiner domains

Johanna Alstott (jalstott@college.harvard.edu)
Harvard College, 28 DeWolfe St. (Mailbox 10)
Cambridge, MA 02138 USA

Masoud Jasbi (masoud_jasbi@fas.harvard.edu)
Department of Linguistics, Harvard University, Boylston Hall (3rd floor)
Cambridge, MA 02138 USA

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 sentence's 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. Operationalizing a quantifier's simplicity or complexity as its shortest logical denotation and its informativeness in terms of the probability of successfully communicating a given world model via that quantifier (p. 516), the author found that computationally generated sets of quantifiers that most resembled those of natural languages tended to have the highest degree of simplicity-in-informativeness optimization. Although Steinert-Threlkeld (2019)'s research focuses on the logical denotations of quantifiers rather than their quantificational forces and is thus not directly applicable to our research, similar issues of optimality can be explored in investigating the lexicalization of quantificational forces. For instance, in examining our two questions, we might find that natural-language quantifier systems converge on a few common quantificational forces as a way to balance the imperatives of simplicity (few, simple lexicalizations of forces) and informativeness or expressiveness (many fine-grained lexicalizations of forces). In the process, we could investigate whether what the optimal lexicalizations of force for determiners are also optimal for adverbials.

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 empirical testing (e.g. Lewis, 1975; de Swart, 1993; von Fintel, 1994). While this assumption, likely made based on intuition, has proven useful in constructing formal semantic theories of quantificational adverbs, it is a worthwhile endeavor to investigate the empirical grounds of this intuition.

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. Finally, we exclude definites (e.g. *the*, *this*, *John's*) and indefinites (*a*) from consideration in our hypotheses because of their unique semantic properties and controversial status as quantifiers (see, e.g., Heim, 1982); we additionally ignore cardinal numbers, as our focus is on quantificational systems rather than exact numerals.

Hypotheses for English

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]: 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 a Keenan (1996) list of English lexical determiners, 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* and *a 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 preliminary 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 trade-offs, 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. At this preliminary stage in our research, we do not rigorously determine which additional forces would be optimal; we choose to investigate clusters II., III., and V. simply because of their preexisting status in the literature. Once we have empirically specified what further clusters actually exist, one can begin to think more rigorously about their optimality. In motivating our hypothesis about adverbs (Hypothesis 1), we can suggest that similar considerations are operative in the adverbial scale.

To begin 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

---

[1] Lewis (1975) 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.

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

We sought 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. To that end, we looked at Keenan (1996)'s inventory of lexical determiners and Lewis (1975)'s list of adverbial quantifiers and excluded several lexical items judged to be potentially problematic with regards to empirical testing.[2] 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 listmates"; 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:

"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

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

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, moving the slider 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 because it would have lead to confusion: for example, seeing "Bill drove to work every day" and then "Bill drove to work no days."
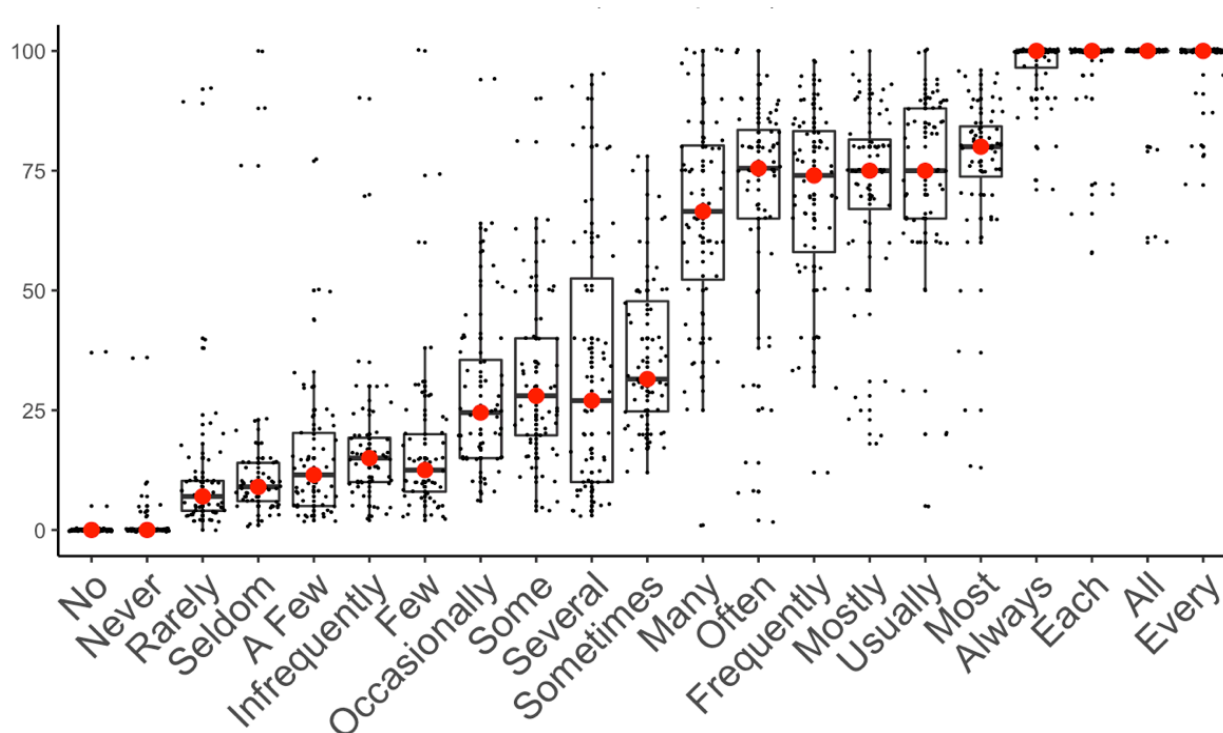

Figure 1: The results of the experiment on quantificational force. The y-axis represents the number of days (out of 100) and the x-axis shows the lexical items tested. The median of the responses is represented with a red circle.

Results

Figure 1 shows the results of our experiment after exclusions. From our original pool of 200 participants, we excluded those who self-reported a native language other than English (3 participants); those who failed an attention check involving movement of a slider to a specified number (3 participants); and those who interpreted a universally quantified sentence like (1a) with a response between 0 and 50 days or interpreted a negative existentially quantified sentence like (2b) with a response between 50 and 100 days (82 participants). We considered such responses to be indicative of insufficient attention to or lack of understanding of the task. The exclusion of such a high number of participants under this criterion was unexpected given the relative simplicity of the experiment. We hope to address this issue in replications via stronger task-independent attention checks and possibly a different participant pool. As there were 56 participants per condition post-exclusions and each participant was presented with one sentence per quantifier, each quantifier received 56 interpretations.

Figure 1 shows box-plots for all of the lexical items tested. The red circles represent the median response for each quantifier, while the black dots represent individual data points. Quantifiers on the x-axis are ordered based on participants' median responses. At the two sides of the axis, we have Tier I (*always*, *each*, *all*, *every*) universal and Tier VI (*no*, *never*) negative existential forces. In between these two extremes, we can detect roughly three clusters of forces: a majority force (*many*, *often*, *frequently*, *mostly*, *usually*, *most*) corresponding to Tiers II and III, a positive propor-

tional force (*occasionally*, *some*, *several*, *sometimes*) corresponding to Tier IV, and a negative proportional force (*rarely*, *seldom*, *a few*, *infrequently*, *few*) corresponding to Tier V. Overall, these results align well with the intuitions reported in the theoretical literature. In the next section, we apply a more formal analysis to recover the cluster of responses based on quantificational force.

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.
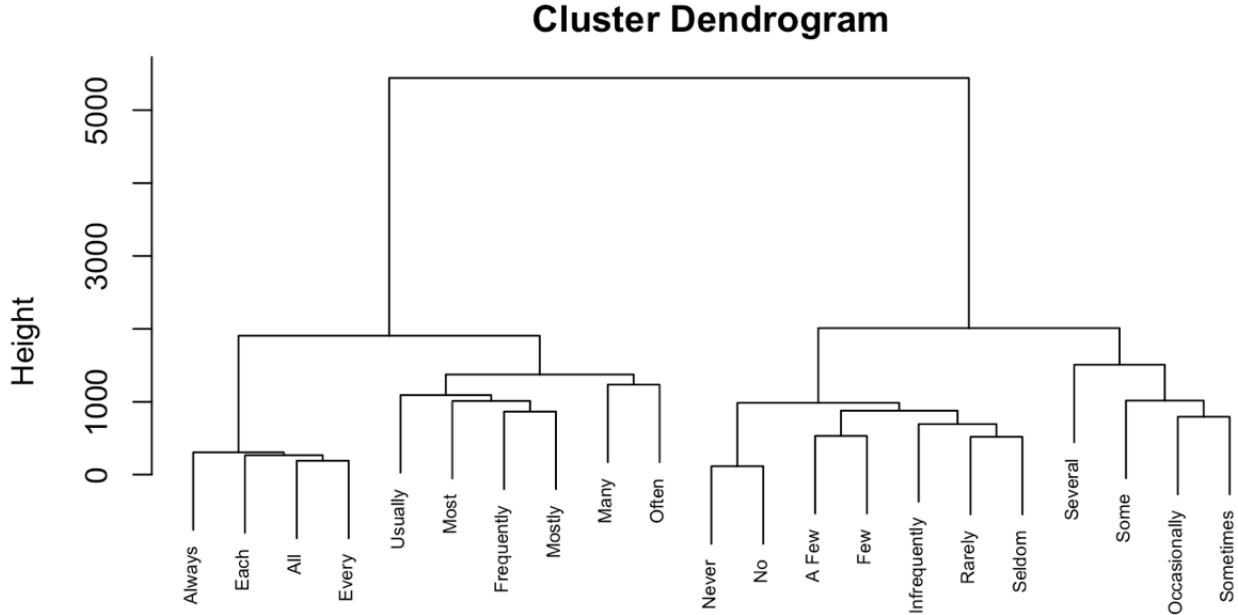
Cluster Dendrogram

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

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 expect that future replications of this work can recover these two clusters when we address the issue of participant inattention and reduce the amount of noise introduced this way.

Discussion

This study was concerned with two overarching questions: First, which quantificational forces do languages encode lexically? Second, are patterns of lexicalization similar across determiner and adverbial scales? In order to address these questions, we started with Lewis (1975)'s hypothesis

that English encodes 6 quantificational forces, shared across the determiner and adverbial domains. Using a web-based experiment and hierarchical cluster analysis of participants' responses, we managed to recover 4 of these quantificational forces systematically. In line with Lewis's hypothesis, these clusters contained both determiner and adverbial quantifiers, providing evidence that patterns of quantificational-force lexicalization may be similar in the determiner and adverbial domains in English.

We may not have clearly recovered all 6 hypothesized quantificational forces for two main reasons. First, it is possible that English simply encodes fewer than 6 quantificational forces. While this is a possibility, we do not think our results here are sufficient to support it. The second possibility, which we find more compelling at this point, is that due to excess noise in our data, our cluster analysis did not have enough precision to detect more fine-grained quantificational forces. As the data in Figure 1 shows, even for quantifiers such as *each, all,* and *every*, many responses lie between 50 to 80. This is after we excluded 82 participants for choosing values below 50 for such trials or above 50 for trials with *no* and *never*. Many responses for quantifiers like *rarely* and *few* are above 70 and for quantifiers like *most* and *often* below 30. These patterns suggest to us that we are dealing with a relatively high degree of noise even for a web-based experiment. This is despite the fact that we had included attention checks.

In future replications, we expect that with less noisy data, hierarchical clustering can at least recover the negative exis-

tential force (Tier VI) and separate it from the negative proportional ones (Tier V).[3] However, it is important to see if we can also recover Tier II and Tier III Future replications for English will also give us a better sense of which groupings in Figure 2 are stable and which are not; by doing so, we can obtain greater certainty about which groupings are legitimate quantificational forces and which are due to chance. In future experiments, we also hope to apply the approach presented here to quantificational nominals and adverbials in other well-studied languages such as German and French.

However, it is crucial to note that results obtained using this experimental paradigm do not offer anything conclusive with regards to the optimality considerations introduced above. Rather, discoveries made about what quantificational forces are lexicalized function as a kind of prelude to more rigorous investigation into issues of optimality and the lexicalization of quantificational forces. The discovery of a small number of lexicalized quantificational forces in multiple domains and languages would seemingly invite an optimality explanation, but such a discovery alone proves neither that the forces discovered are optimal nor that the optimal forces in the determiner realm are the same as those in the adverbial realm. As such, we leave more detailed conclusions about optimality to future research.

Finally, we would like to add that our research leaves untouched many further issues on the similarities and differences between the determiner and adverbial scales. 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. The evidence for this hypothesis can come from the observation that 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 important 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 short, despite the preliminary nature of this study, it lays the foundation for a systematic investigation of quantificational meaning and its lexicalization by world languages in the future.

Acknowledgements

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.

---

[3] These two types of quantifiers do branch off at a low level of height, but we can't draw conclusions from that bifurcation because of exclusions of data on the negative existentials.