University of Pennsylvania Working Papers in Linguistics

Volume 24 | Issue 1 Article 21

4-2-2018

The Semantics and Pragmatics of "some 27 arrests"

Jon Stevens
The Ohio State University

Stephanie Solt
Leibniz-Zentrum Allegemeine Sprachwissenschaft

The Semantics and Pragmatics of "some 27 arrests"

Abstract
In this paper we examine the "some + n" construction, where the apparent quantifier "some" is used with a numerical expression, as in "some 27 students were arrested." Contrary to previous claims in the literature, we show that while many speakers prefer an approximative interpretation for some + n, it is untenable to analyze "some" as an approximator akin to "about" or "roughly." We survey some constraints on the distribution of some + n, and propose a semantic analysis based on recent theories of indefinite determiners (e.g. Alonso-Ovalle and Menendez-Benito 2010) which is able to explain these constraints. On our account, "some" introduces a manipulation of the domain of quantification, either restricting it to contextually relevant pluralities or widening it to include pluralities whose cardinality is *approximately* that of the associated numeral. In contexts where approximation is disfavored and there is no obvious restriction on the domain, the meaning contribution of "some" is essentially vacuous. We claim that this vacuousness, in conjunction with Horn's (1984) division of pragmatic labor, explains why some + n is most felicitous in emphatic contexts, e.g., "some 17 Republicans ran in the primary!" (where 17 is higher than expected) as opposed to "?Some 5 Democrats ran in the primary" (where 4 is average).

The Semantics and Pragmatics of "some 27 arrests"

Jon Stevens and Stephanie Solt*

1 Introduction

We examine the use of the quantifier *some* with a numeral, a construction which we call *some* + *n*, exemplified by the newspaper headline in (1).

(1) Some 27 arrests in one morning in anti-Mafia blitz
(The Italian Insider, 12 January 2016)

Previous authors have attributed an approximative meaning to the *some* + *n* construction (Sauerland and Stateva 2007, Anderson 2014), such that (2a) is taken to have roughly the same meaning as (2b).

(2) a. There were some twenty people at the party.
b. There were about twenty people at the party.

This has been captured via analyses that assign *some* a lexical entry comparable to that of an approximator, such as *about*. Sauerland and Stateva (2007) implement this using the mechanism of scale granularity, analyzing *some* + *n* as denoting *n* at the coarsest level of granularity (3). Somewhat similarly, Anderson (2014) draws on Lasersohn's (1999) notion of pragmatic halos, proposing that *some* + *n* denotes a value in the set formed as the union of *n* and its halo (4):

(3) $[\![some\ twenty]\!]^{gran} = coarsest(gran)([\![twenty]\!])$
(4) $[\![some\ twenty]\!]^{c} = f([\![twenty]\!] \cup halo_c([\![twenty]\!]))$

We first argue that, while an approximative reading of *some* + *n* is possible for many speakers, it is a mistake to treat *some* + *n* as an approximator, strictly speaking, and it is untenable to treat it as identical to *about* + *n*. We then note a particular emphatic effect that arises from the non-approximative use of *some* + *n*. We give a semantic analysis that is in line with more general work on quantification and indefinites (Kratzer and Shimoyama 2002, Alonso-Ovalle and Menéndez-Benito 2010, 2011), whereby *some* + *n* introduces existential quantification over a restricted domain of pluralities with numerosity (close to) *n*. Finally, we sketch out a pragmatic model of how our proposed denotation of *some* + *n* leads to the emphatic effects associated with the construction.

1.1 *Some* + *n* Is Not (Always) an Approximator

If one reads the article for which (1) is the headline, it is clear that no approximative interpretation is intended—*exactly* 27 relevant arrests in total were made in two separate raids that morning. This is not unusual. Numerous examples of *some* + *n* in contexts that suggest precision can be found in the Corpus of Contemporary American English (Davies 2008-), including the following.

(5) Some 1,841 retirees pulled down more than $100,000 a year in pension checks.
(6) Of some 206 students who responded to the survey, 52% were female.

Nothing in these examples, or in the contexts in which they appear, suggests a meaning similar to *about* + *n*.

That said, not all speakers agree on this. Anecdotally, intuitions among native speakers conflict about the felicity of the above examples, with some speakers showing a strong preference for *about*-like uses as in (2). Experimentally, Solt et al. (2017) use a numerical range task on examples like (1) and (2) to show that indeed there is a mixture of: (i) speakers who prefer approximative

*We thank the audiences of the 91st LSA Annual Meeting, PLC41, and XPRAG2017 for helpful discussion, and also Brandon Waldon for assistance with the corpus and experimental research. This research was supported in part by the American Council of Learned Societies (ACLS), and by the German Science Foundation (DFG) via grant SO1157/1-2 to the second author.

interpretations of *some* + *n* across the board, (ii) speakers who prefer approximative interpretations of *some* + *n* only when it is used with round numbers (ten, twenty, etc.), and (iii) speakers who tend to always interpret *some* + *n* as being precise.

We aim to give a semantic analysis of this construction that partially explains its variation in meaning. We do so by positing a representation that contains an underspecified domain restriction function (Alonso-Ovalle and Menéndez-Benito 2010) which can easily be, but need not be, enriched with an approximating function. Before spelling that out, let's examine other, more fundamental ways in which *some* + *n* differs from *about* + *n*, before turning to the pragmatic effects of using *some* + *n* in non-approximating contexts.

1.2 *Some* + *n* Has a More Constrained Distribution Than *About* + *n*

Unlike *about* + *n* and other approximators, like *roughly* and *approximately*, *some* + *n* can only occur in contexts where the associated numeric phrase can be interpreted as a plural individual; that is, as the concatenation or sum of some number of atomic individuals.

(7) The meeting lasted (about/some) three hours.

(8) The tree is (about/some) twenty feet from the house.

(9) The meeting took place at (about/*some) three o'clock.

(10) The tree was planted in (about/*some) 1989.

In (7), the phrase *three hours* can be interpreted as the sum of three 60-minute intervals, and *some* + *n* is perfectly acceptable. Similarly, *twenty feet* in (8) can be understood as a spatial extent formed as the concatenation of twenty 1-foot extents, again allowing *some*. In (9), on the other hand, the phrase *three o'clock* cannot be so interpreted—it makes no sense to speak of three individual "o'clocks" being summed together—and in that case, *some* + *n* is illegal. A similar point can be made about (10). This contrasts with *about* + *n*, which is licensed in all of these contexts.

An analysis of 500 *some* + *n* tokens extracted from COCA found that cardinality-based measures such as *some twenty students* were the most common (61% of tokens). After these, temporal and spatial extents (corresponding to (7) and (8), respectively) were the next most frequent. Also occurring were expressions corresponding to other additive measures, including land area, volume and proportion. Punctual temporal expressions as in (9) and (10) were unattested, however, in either this or a larger sample.

These data show that, while *about* + *n* instantiates pure approximation of a numeral or other measure expression, *some* + *n* is constrained by the semantic type of the denotation of the measure phrase in which the numeral occurs.

Moreover, *some* + *n* lacks a true degree interpretation. In other words, *some* + *n* cannot by itself, with no associated noun phrase, refer to a number or numerical range. This is demonstrated by its infelicity in mathematical statements and as the answer to a *how many* question. Again, this is contra the behavior of *about* + *n* and other similar constructions.

(11) Seven times fourteen is (about/roughly/approximately/*some) one hundred.

(12) Q: How many students passed the test?
A: (About/roughly/approximately/??some) fifty

But note that the degraded status of *some* + *n* in (12) is ameliorated if a pronoun is present (e.g., *some fifty of them*).

Taken together, examples (7)–(12) suggest that *some* + *n* is the result of combining *some* with a numeral phrase that can be interpreted as a plurality of entities of some sort. This, along with the propensity of many speakers to assign an approximative interpretation, especially with round numbers, will be the basis for our semantic analysis of this construction. We end our introduction by discussing an interesting pragmatic effect of non-approximative *some* + *n*.

1.3 *Some* + *n* Gives Rise to Emphasis

In cases where *some* + *n* does not give rise to an approximative reading, what is its function? It is clear that *some* + *n* is not identical to a bare numeral in such cases. Imagine the following sentences, each uttered in an out-of-the-blue context.

(13) a. I raised two children.
b. ?? I raised some two children.

It is pragmatically odd to utter (13b) (which is surely not meant to approximate) when (13a) would have sufficed to convey the exact same meaning. But we know from examples like (1), (5), and (6) that non-approximative *some* + *n* can be perfectly natural in some contexts. Such contexts, we argue, are ones in which the numeral phrase can be interpreted emphatically. That is, the speaker wishes to draw attention to the numeral phrase, inasmuch as she finds the numerosity to be noteworthy or surprising in some way. The following minimal pair illustrates the difference in felicity between *some* + *n* in ordinary and emphatic contexts.

(14) a. ? Some five Democrats ran in the 2016 Presidential primary.
b. Some *seventeen* Republicans ran in the 2016 Presidential primary!

The emphatic (14b) is a much more natural utterance, all things being equal, than (14a). The effect seems to be that in (14b), *some* + *n* is being used to highlight the number of Republican candidates, which is unusually large by modern historical standards. In (14a), on the other hand, the number of Democratic candidates is much more in line with what one would expect, and thus use of *some* + *n* seems unjustified.

In Section 3 we will show how the semantic analysis we give in the following section, which is used to derive the facts in the previous sections, can give rise to emphasis by making vacuous contributions that take on additional emphatic meaning via Horn’s (1984) division of pragmatic labor.

2 Semantic Analysis

The data discussed in Section 1.2 suggest strongly that *some* + *n* does not have a denotation in the domain of degrees, but that, instead, *some* operates on individuals of some sort. Taking a broader perspective, this should not come as any surprise: in its more common instantiation, *some* is an indefinite determiner, which is most commonly analyzed as introducing existential quantification over individuals. The core idea that we pursue here is that the *some* of *some* + *n* is in fact no different from ‘ordinary’ quantificational *some*. That is, the claim is that all of the examples in (15) are based on the same lexical entry for *some*:[1]

(15) a. Some student was arrested.
b. Some students were arrested.
c. Some twenty students were arrested.

More specifically, we draw on proposals from the literature on indefinites, according to which certain indefinite determiners are existential quantifiers that also introduce an operation or restriction on the domain of quantification. It is well known that quantification in natural language is made over contextually restricted domains. To say for example that *everyone had a good time* is not to make a claim about everyone in the entire universe, but rather to make a claim about some contextually specified set of individuals; say, those who attended a particular party. It has been argued that certain quantificational determiners encode the manipulation of the domain as part of their lexical semantics. Perhaps most famously, Kadmon and Landman (1993) propose that *any* is a domain widener, and Kratzer and Shimoyama (2002) make a somewhat similar claim for German *irgendein* ‘some’. Conversely, other indefinite determiners been proposed to restrict the domain of quantification (similarly to the case of *everyone* discussed above) via a subset function that maps the entire

[1] Farkas (2002) makes a similar claim, but attributes to *some* + *n* an inherently approximative meaning.

domain to some contextually relevant subset thereof. Such is the analysis given by Alonso-Ovalle and Menéndez-Benito (2010) to Spanish *un* 'a' and (with some further content) *algún* 'some'.

Building in particular on this latter work, we propose the following as the basic lexical entry for *some*, where *f* is a function from sets of individuals to sets of individuals:

(16) $[\![some]\!] = \lambda P \lambda Q.\exists x[f(P)(x) \wedge Q(x)]$

Below we will consider in more depth the nature of the function *f*, which we will argue is the locus of the interspeaker variation in the availability of the approximative interpretation, among other effects. First, however, let us see how this entry allows a compositional analysis of *some* + *n*, and how it accounts for the restrictions on its distribution.

To start, we follow authors including Krifka (1999) and Landman (2004) in assuming a non-quantificational semantics for numerically quantified noun phrases. Cardinal numerals themselves denote cardinality predicates; that is, predicates over plural individuals that specify their number of atomic parts, per (17a). These may compose intersectively with a plural noun phrase, forming a complex predicate, as in (17b).

(17) a. $[\![twenty]\!] = \lambda x_e.|x| = 20$
b. $[\![twenty\ students]\!] = [\![twenty]\!] \cap [\![students]\!] = \lambda x_e.students(x) \wedge |x| = 20$

In the absence of an overt determiner, a nominal expression of the form in (17b) composes via set intersection with the sentential predicate, with quantificational force arising via existential closure, yielding the following for a simple example:

(18) Twenty students were arrested.
$\exists x[students(x) \wedge |x| = 20 \wedge arrested(x)]$

Returning to *some* + *n*, the set expression in (17b) is also of the right semantic type to saturate the first argument slot of *some*. We thus derive (19) for *some twenty students*, and (20) for a full sentential example:

(19) $[\![some\ twenty\ students]\!] = [\![some]\!]([\![twenty\ students]\!])$
$= \lambda Q.\exists x[f([\![twenty\ students]\!])(x) \wedge Q(x)]$
$= \lambda Q.\exists x[f(\lambda y.students(y) \wedge |y| = 20)(x) \wedge Q(x)]$

(20) Some twenty students were arrested
$\exists x[f(\lambda y.students(y) \wedge |y| = 20)(x) \wedge arrested(x)]$

In words, (20) says that in the set obtained by applying the function *f* to the set of twenty-member pluralities of students, there is a plurality of individuals who were arrested.

Recall from Section 1.2 that *some* + *n* is infelicitous in contexts in which a numerical expression is interpreted as referring to a numerical value or range; this includes mathematical statements as well as answers to *how many* questions (see (11)-(12)). This restriction falls out from the present analysis. *Some* + *n* does not denote a degree; in fact, it is not a constituent at all. Thus it cannot occur in positions that involve reference to degrees. This contrasts with *about* + *n*, which plausibly has a purely degree-based denotation along the lines of (3) or (4) (see Sauerland and Stateva 2007 and Solt 2014 for specific proposals regarding the semantics of approximators such as *about*).

The present account also provides a route to explaining the constraints on the sort of measure phrases that may occur in *some* + *n*. On the present analysis, *some* takes sets of entities as arguments, and more specifically introduces a function from sets to sets, which we have characterized as a manipulation of its domain of quantification. In the cardinality-based example worked out above, the sets in question were sets of ordinary plural individuals; that is, of type $\langle e,t \rangle$. Certain measure expressions, such as those exemplified in (7) and (8) above, can also be analyzed as denoting sets, albeit of a different sort of plural entity. Durations, in particular, can be conceptualized as temporal extents; that is, as convex intervals on the timeline (Krifka 1989). An expression such as *three hours* can then be analyzed as denoting the set of such intervals that have a length of three hours; that is, that are equivalent to the concatenation of three 1-hour intervals. A corresponding analysis can be given to spatial extents, which in Vector Space Semantics (Zwarts 1997, Zwarts and Winter 2000) are analyzed as vectors; that is, as directed line segments between points in space. Thus *twenty feet* denotes the set of vectors of length twenty feet, while *twenty feet from the house* denotes the subset of that set whose origin is the house.

(21) $[\![\textit{three hours}]\!] = \{t : convex(t) \wedge hours(t) = 3\}$
(22) $[\![\textit{twenty feet}]\!] = \{v : |v| = 20'\}$

Expressions such as these are of the right general sort to serve as arguments of *some*. In semi-formal terms, our earlier example (8) receives the analysis in (23), which like the previous cardinality-based example features a domain-manipulation function *f* and an existential quantifier.[2]

(23) The tree is some twenty feet from the house.
$\exists v[v \in f(\{v : |v| = 20'\}) \wedge v \textit{ originates at the house} \wedge v \textit{ terminates at the tree}]$

By contrast, punctual temporal measures such as *three o'clock* or *1989* cannot be construed as individuals that can be members of sets. Rather, expressions such as *three o'clock* and *1989* are better understood as rigidly designating specific points or segments of the timeline. In that these expressions do not have set-based interpretations, they do not provide an appropriate argument for *some*. This explains the ungrammaticality of examples such as **some three o'clock* or **some 1989*.

Let us now return to a crucial component of our proposed semantics for *some*, namely the domain-manipulation function *f*. We claim that the nature of this function is crucial to the pragmatic effects that obtain with *some* + *n*, as well as the interspeaker variation in the interpretation that this construction allows. In Alonso-Ovalle and Menéndez-Benito's analysis of Spanish *un* and *algún* that served as the starting point for our analysis of *some*, the domain of quantification introduced by the indefinite determiner is a subset of the total domain. For example, *un estudiante* 'a student' introduces existential quantification over some subset of the set of all students in the universe. This captures the intuition that indefinite determiners semantically encode domain restriction of the sort involved covertly in universally quantified examples such as *everyone had a good time*, discussed above. Adopting Alonso-Ovalle and Menéndez-Benito's analysis, we propose that in the default case, the domain-manipulation function *f* lexicalized by *some* is likewise a subset function, which maps a set to a subset of it.

In the case of *some* + *n*, however, we propose that speakers differ in the value that they assign to this function. Specifically, for those speakers who attribute an approximative meaning to *some* + *n*, the function *f* does not *narrow* the domain of quantification, but instead *widens* it to include pluralities close in cardinality to *n*. The two possible interpretations of *f* are thus the following:

(24) a. **Default *some*:** $f([\![\textit{twenty students}]\!]) \subseteq [\![\textit{twenty students}]\!]$
b. **Approximating *some*:** $f([\![\textit{twenty students}]\!]) = \lambda x.students(x) \wedge |x| \approx 20$

That *some* should sometimes induce a sort of widening of the domain of quantification is not unexpected in the context of the cross-linguistic landscape of indefinite determiners, some of which are known to have a domain-widening function (e.g. German *irgendein*, discussed above). We moreover believe there to be solid pragmatic reasons for the interpretation in (24a) to be reanalyzed along the lines of (24b). Assuming that *f* is interpreted simply as a subset function, a sentence such as *some twenty students were arrested* is from the perspective of the hearer truth conditionally indistinguishable from its simpler counterpart *twenty students were arrested*: since the hearer does not know which subset function the speaker has in mind, existential quantification over $f([\![\textit{twenty students}]\!])$ is not distinguishable from existential quantification over the full domain $[\![\textit{twenty students}]\!]$. This plausibly has lead to a pressure to attribute additional semantic content to the *some* variant.

[2]The reader might note that the lexical entry for *some* in (16) is not precisely what we need to derive the interpretation in (23) compositionally. This might be dealt with by incorporating type flexibility into the semantics of *some*, or alternately by pursuing a decompositional approach according to which the lexical content of *some* is only the domain-manipulation function *f*, with existential quantification arising via existential closure. Selecting between these and other options would require a more in-depth investigation of the semantics of measurement than is possible in the present paper; however, we believe that such an extension is feasible.

Recall that the experimental investigation of Solt et al. (2017) found that some speakers attribute an approximative interpretation to *some* + *n* only in combination with round numbers (e.g. *some twenty*) while others interpret the construction approximatively across the board. Krifka (2007) points out that bare round numerals themselves allow an approximate interpretation, while bare non-round numbers must be interpreted precisely. We hypothesize that for the first group of speakers (round only approximate), *some* + *n* is interpreted as maximally domain widening with respect to the numerical value, such that *n* is interpreted in the broadest possible way compatible with its semantics; this yields an approximate reading for round numbers, but a precise interpretation for non-round ones. For the second group of speakers (across the board approximate), we propose that the approximating function has been further conventionalized in the semantics of *some*, which may then add an approximative meaning to the interpretation of the numeral.

To summarize, an analysis of *some* + *n* couched in the tradition of recent theories of indefinite determiners provides a means to account for the construction's similarity in meaning to unmodified numerical expressions, as well as the availability of an approximative interpretation for some speakers.

In concluding this section, we briefly consider other semantic effects found with *some*, and how they relate to the analysis developed here. Alonso-Ovalle and Menéndez-Benito (2010, 2011) observe that *algún*, the Spanish counterpart of *some*, produces an ignorance effect when it composes with a singular noun phrase, but not when it composes with a plural nominal. They account for this by proposing that *algún* requires that its (restricted) domain of quantification be a non-singleton set. In the singular case the ignorance effect can then be derived as an implicature relative to singleton-set alternatives that the speaker could have used; this implicature is blocked in the plural case. English *some* shows a similar pattern of ignorance effects with singular and plural noun phrases. For example, *some student was arrested* suggests that the speaker does not know (or care) which student it was, while this effect is absent in the plural *some students were arrested*. On this basis, we might similarly propose that the output of the domain-manipulation function lexicalized by *some* likewise must be a non-singleton set. However, an account along the lines of Alonso-Ovalle and Menéndez-Benito's would also predict an ignorance effect for *some* + *n*. This prediction is not borne out: *some twenty students were arrested* does not convey speaker ignorance as to which twenty students were arrested, though it may suggest that their number is more important than their identity.

These data suggest that a somewhat different account of variable ignorance effects with *some* is needed. We hypothesize that the solution to this puzzle may lie in redefining what it means to be a non-singleton set; exploring this in depth is, however, beyond the scope of this paper, and must be left for future work.

3 Deriving Emphasis

The semantic analysis above predicts that in certain contexts in which *some* + *n* is not interpretable as denoting approximation, the truth-conditional contribution of *some* + *n* beyond that of a bare numeral is completely vacuous. Let's take (13b) as a starting point, reprinted below as (25).

(25) ?? I raised some two children.

Let's assume two points: (i) the speaker in (25) knows how many children he or she raised, and (ii) there is no contextually obvious restriction on the set of two-child pluralities in this case. These assumptions entail that there is no obvious candidate for the output of the *f* function in the denotation of *some* + *n*, apart from the entire domain of two-child pluralities. That is to say, the hearer is forced to interpret *f* as if it were the identity function, and as a result, we obtain the same denotation for (25) as we would have using the simpler bare numeral phrase, *two children*.

$$
\begin{aligned}
[\![\text{I raised some two children}]\!] &= \exists x[f(\lambda y.[\mathit{children}(y) \wedge |y| = 2])(x) \wedge \mathit{raised}(\mathit{speaker})(x)] \\
&= \exists x[\lambda y.[\mathit{children}(y) \wedge |y| = 2](x) \wedge \mathit{raised}(\mathit{speaker})(x)] \\
& \text{ (if } f \text{ is interpreted as if it were the identity function)} \\
&= \exists x[\mathit{children}(x) \wedge |x| = 2 \wedge \mathit{raised}(\mathit{speaker})(x)] \\
&= [\![\text{I raised two children}]\!]
\end{aligned}
$$

The contribution of *some* in this context is completely vacuous. It is a short step from vacuousness to emphasis, as the following examples illustrate.

(26) a. No fewer than seventeen Republicans ran in the 2016 Presidential primary.
b. Not ten, not fifteen, but seventeen Republicans ran in the 2016 Presidential primary.
c. Fully seventeen Republicans ran in the 2016 Presidential primary.

In all of these cases, the numeric phrase is embedded within a construction (*no fewer than n*, *not n but n′* (where $n < n'$) or *fully n*) that is either entailed by or strongly implicated by the denotation of the corresponding bare numeral phrase. There is effectively no semantic contribution made by these constructions, and yet they can be perfectly natural, and not at all redundant, precisely in those contexts that favor (14b), reprinted below as (27).

(27) Some *seventeen* Republicans ran in the 2016 Presidential primary!

The constructions in (26) all have two things in common with *some* + *n*: (i) their truth-conditional contributions in context are slim to none, and (ii) they are felicitous as expressions of emphasis. Another construction that is even more similar to *some* + *n* in that it not only has an emphatic effect but can be used for approximation, is the discourse particle *like* (Beltrama and Hanink to appear).

(28) a. There were like, seventeen Republicans in the 2016 Presidential primary.
b. There were like fifteen Republicans in the 2016 Presidential primary.

When followed by a prosodic pause as in (28a), *like* carries a meaning of surprise, whereas in (28b) it is most naturally interpreted as an approximator. The generality of this phenomenon suggests a very general explanation. We propose the facts can be explained, at least in part, by applying Horn's (1984) *division of pragmatic labor*, paraphrased below.

(29) **Division of pragmatic labor:** Marked forms signal marked meanings.

This was proposed as a principle to explain implicatures arising from what Grice called Manner violations, where lengthy or otherwise convoluted expressions give rise to alternative meanings.

(30) a. Lee stopped the car.
b. Lee made the car stop.

In (30a), the unaccusative *stop* being used with an agent *Lee* gives rise to what one might construe as the default interpretation for $cause(Lee)(stop(car))$, namely that Lee pressed the brake pedal, causing the car to come to a stop. But in (30b), the explicit causative construction, *made stop*, which is a marked way of conveying $\lambda x.\ cause(x)(stop(car))$, gives rise to the unusual interpretation that Lee caused the car to stop via some other means, perhaps by standing in the path of the moving car until the driver stopped, or by crashing the car into a road sign.

Van Rooy (2004), expanding on the work of Parikh (2001), provides a mathematical analysis of why the division of pragmatic labor should hold generally within and across languages. The details of that account, which involves evolutionary game theory, cannot be explicated here, but in the remainder of this section we present a simple algebraic model that captures the spirit of that analysis and applies it to the use of *some* + *n* (or other forms like in (26)).

The key notion to modeling the division of pragmatic labor is *utility*. One can think of the utility of an utterance as its payoff—by definition, interlocutors strive to maximize the utility of their conversational contributions. In game theory and related frameworks like decision theory, utility is usually a function from (sets of) choices—such as which forms to use to convey which meanings—to real numbers. However, we do not need any specific numbers in our model. We can instead represent utility values purely algebraically. But first let's define some additional terms.

Utility-based models often encode Grice's Maxim of Quantity by giving higher utility to utterances that carry more information (see e.g. Frank and Goodman 2012, Franke and Jäger 2016). To this end, we take the utility of an utterance to be equal to the *surprisal* of the utterance, minus the *cost* of the utterance. Surprisal of utterance u is equal to $-\log \mathrm{Prob}([\![u]\!])$, where $\mathrm{Prob}([\![u]\!])$ is the prior probability of u being true. Thus, the more surprising u is (i.e. the less likely one is to believe

$[\![u]\!]$ before hearing u), the higher is the utility assigned to u. Again, we do not need specific values for the quantities in this model, and thus other measures of information content may also work.

In being emphatic, the speaker is signaling that the surprisal of u is higher than the hearer might have expected a priori. For simplicity we will only consider two possible surprisal values: a baseline surprisal value, which we will label σ, and an higher surprisal value associated with an emphatic interpretation, which we will label $\sigma+\Delta$, where Δ is the change in surprisal from a normal to an emphatic context.

The cost of an utterance is taken to be an encoding of how complex it is, either semantically or phonetically. We compare a single simple form (a bare numeral phrase) against a single complex form (*some* + *n*). For our purposes, it does not matter which measure of complexity we choose, because *some* + *n*, when compared to a bare numeral, is phonetically longer and also, if our hypothesis is correct, has a more complex semantic representation. For simplicity we can assume that the simple form is costless, and assign the complex form a fixed positive cost, which we will label c.

The speaker's goal in this model is to signal that the information being conveyed by her utterance has higher-than-baseline surprisal. We refer to contexts in which surprisal is higher than baseline as *emphatic* contexts (EMPH for short), and the other contexts as *baseline* contexts (BASE for short). The hearer does not know a priori which kind of context she is in. If she correctly determines that she is in an EMPH context, the utility for the speaker is $\sigma+\Delta$, whereas if she determines that she is in a BASE context, the utility for the speaker is only σ.

We now want to compare three different logically possible systems for marking emphasis:

1. **Complex marker (CM):** The complex form marks EMPH, and the simple form marks BASE (complies with Horn's division of pragmatic labor).
2. **Simple marker (SM):** The simple form marks EMPH, and the complex form marks BASE (the "anti-Horn" strategy).
3. **No marker (NM):** Emphasis is not marked.

Our chosen measure of how efficient a system is for maximizing utility in the aggregate is the *expected utility* of using that system. The expected utility (EU) of one of the above systems is the average utility of successfully employing that system, weighted by how often emphatic contexts arise, which we will label P_{EMPH}, and how often baseline contexts arise, P_{BASE}, which is $1-P_{\text{EMPH}}$. For the CM system, expected utility is the product of P_{EMPH} and $\sigma+\Delta-c$ (the utility of using a complex form to signal emphasis), plus the product of P_{BASE} and σ (the utility of using a simple form to signal a baseline context). For the SM system, expected utility is the product of P_{EMPH} and $\sigma+\Delta$ (the utility of using a simple form to signal emphasis), plus the product of P_{BASE} and $\sigma-c$ (the utility of using a complex form to signal a baseline context). Expected utility for not signaling emphasis is simply σ.

$$
\begin{aligned}
EU(\text{CM}) &= P_{\text{EMPH}}(\sigma+\Delta-c)+P_{\text{BASE}}\sigma \\
EU(\text{SM}) &= P_{\text{EMPH}}(\sigma+\Delta)+P_{\text{BASE}}(\sigma-c) \\
EU(\text{NM}) &= \sigma
\end{aligned}
$$

Armed with high school algebra, we can calculate the conditions under which the observed CM system (the system that employs Horn's division of pragmatic labor) is the best of the three. We do this by solving two inequalities: $EU(\text{CM}) > EU(\text{SM})$ and $EU(\text{CM}) > EU(\text{NM})$. The solutions are given below.

$$
\begin{aligned}
EU(\text{CM}) &> EU(\text{SM}) \\
P_{\text{EMPH}}(\sigma+\Delta-c)+P_{\text{BASE}}\sigma &> P_{\text{EMPH}}(\sigma+\Delta)+P_{\text{BASE}}(\sigma-c)
\end{aligned}
$$

$$\Downarrow$$

$$P_{\text{EMPH}} < \tfrac{1}{2}$$

$$
\begin{aligned}
EU(\text{CM}) &> EU(\text{NM}) \\
P_{\text{EMPH}}(\sigma+\Delta-c)+P_{\text{BASE}}\sigma &> \sigma \\
P_{\text{EMPH}}(\Delta-c) &> 0
\end{aligned}
$$

$$\Downarrow$$

$$P_{\text{EMPH}} > 0 \text{ and } \Delta > c$$

The conditions that arise from these inequalities can be summarized as follows: (i) emphatic contexts must occur less frequently than baseline contexts, (ii) emphatic contexts must nonetheless occur, and (iii) the boost in utility associated with emphasis must outweigh the cost of using a complex form. So, if emphasis is not too common,[3] and if talk remains cheap, it is best to employ CM.

4 Conclusion

We have showed that in many instances, contra previous literature, *some* + *n* is not an approximating construction. In doing so, we have noted some constraints on the semantic type of the measure expression contained within the *some* + *n* phrase, which are consistent with an account where *some* + *n* is the simple composition of *some* with a numeral phrase. We assume a semantics for *some* that involves existential quantification over a restricted domain derived via an underspecified function from sets to sets, f. Under this proposal an approximative interpretation can arise via filling the f function variable with a function that approximates. In cases where no approximative interpretation is possible, and in which the restriction on the domain is not obvious to the hearer, we argue that use of *some* + *n* is semantically indistinguishable (at least for the hearer) from the equivalent sentence with a bare numeral phrase. This explains why the use of *some* + *n* in these contexts can have the pragmatic effect of emphasizing the cardinality of the plurality associated with the numeral phrase, insofar as Horn's division of pragmatic labor predicts such uses. Finally, we developed a simple algebraic model to show that the division of pragmatic labor is communicatively more effective than both an "anti-Horn" system and a system that does not mark emphasis at all, namely, because it allows the speaker to convey when an utterance is more surprising or interesting than the hearer might have otherwise thought, and to do so with the least amount of effort on average.

Future work will further connect the semantics of *some* + *n* with the semantics of indefinites and quantifiers more generally. One avenue of exploration is the lack of ignorance effects with *some* + *n*, discussed in Section 2 above. Finally, we hope to further explore the interspeaker variation with respect to the approximative effect of *some* + *n*, and to look for any relevant diachronic trends. We note that an analysis of tokens from the Corpus of Historical American English (Davies 2010-) shows the rate of occurrence of *some* + *n* with overt markers of approximation has steadily decreased since the 19th century. This suggests, perhaps, that approximative use of *some* + *n* (which would preclude the necessity of a co-occurring approximator) is an innovation, perhaps a conventionalization of variation in how the domain of quantification is determined.

References

Ahern, Christopher, and Robin Clark. 2017. Conflict, cheap talk and Jespersen's cycle. *Semantics and Pragmatics* 10:1–41.

Alonso-Ovalle, Luis, and Paula Menéndez-Benito. 2010. Modal indefinites. *Natural Language Semantics* 18:1–31.

Alonso-Ovalle, Luis, and Paula Menéndez-Benito. 2011. Plural epistemic indefinites. In *Proceedings of NELS 40*, ed. S. Kan, C. Moore-Cantwell, and R. Staubs, 17–30. UMass.

Anderson, Curt. 2014. Approximation of complex cardinals using *some*. In *Proceedings of WECOL 2013*, 131–143.

Beltrama, Andrea, and Emily Hanink. To appear. This is, like, a mirative construction! "Like" between uncertainty and surprise. Ms., University of Konstanz.

Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present.* URL `http://corpus.byu.edu/coca/`.

Davies, Mark. 2010-. *The Corpus of Historical American English (COHA): 400 million words, 1810-2009.* URL `http://corpus.byu.edu/coha/`.

Farkas, Donka. 2002. Varieties of indefinites. In *Proceedings of Semantics and Linguistic Theory 12 (SALT XII)*, ed. B. Jackson, 59–83.

[3] Overuse of emphatic forms is the exception that proves the rule, as it is often accompanied by phonetic reduction and semantic simplification over time. For example, the famous Jespersen's cycle involves the reduction and bleaching of emphatic negation forms into baseline negation forms. See Ahern and Clark (2017) for a game-theoretic analysis of Jespersen's cycle.

Frank, Michael C., and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336:998.

Franke, Michael, and Gerhard Jäger. 2016. Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft* 35:3–44.

Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In *Meaning, Form and Use in Context: Linguistic Applications*, ed. D. Shiffrin, 11–42. Washington, D.C.: Georgetown University Press.

Kadmon, Nirit, and Fred Landman. 1993. Any. *Linguistics and Philosophy* 16:353–422.

Kratzer, Angelika, and Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. In *Proceedings of the 3rd Tokyo Conference on Psycholinguistics*, ed. Y. Otsu, 1–25. Tokyo: Hituzi Syobo.

Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. In *Semantics and Contextual Expressions*, ed. R. Bartsch, J. van Benthem, and P. van Emde Boas, 75–115. Dordrecht: Foris.

Krifka, Manfred. 1999. At least some determiners aren't determiners. In *The Semantics/Pragmatics Interface from Different Points of View*, ed. K. Turner, 257–291. Oxford: Elsevier.

Krifka, Manfred. 2007. Approximate interpretations of number words: A case for strategic communication. In *Cognitive Foundations of Interpretation*, ed. G. Bouma, I. Krämer, and J. Zwarts, 111–126. Koninklijke Nederlandse Akademie van Wetenschappen.

Landman, Fred. 2004. *Indefinites and the Type of Sets*. Oxford: Blackwell.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75:522–551.

Parikh, Prashant. 2001. *The Use of Language*. Stanford, CA: CSLI Publications.

Sauerland, Uli, and Penka Stateva. 2007. Scalar vs. epistemic vagueness: evidence from approximators. In *Proceedings of Semantics and Linguistic Theory 17 (SALT 17)*, ed. M. Gibson and T. Friedman, 228–245.

Solt, Stephanie. 2014. An alternative theory of imprecision. In *Proceedings of Semantics and Linguistic Theory 24 (SALT 24)*, ed. T. Snider, S. D'Antonio, and M. Weigand, 514–533.

Solt, Stephanie, Jon Stevens, and Brandon Waldon. 2017. "Some" approximations: An experimental investigation. Presented at XPRAG 2017, Cologne.

Van Rooy, Robert. 2004. Signalling games select Horn strategies. *Linguistics and Philosophy* 27:493–527.

Zwarts, Joost. 1997. Vectors as relative positions: A compositional semantics of modified PPs. *Journal of Semantics* 14:57–86.

Zwarts, Joost, and Yoad Winter. 2000. Vector space semantics: A model-theoretic analysis of locative prepositions. *Journal of Logic, Language and Information* 9:169–211.

Jon Stevens
Department of Linguistics
The Ohio State University
Columbus, OH 43085
stevens.400@osu.edu

Stephanie Solt
Leibniz-Zentrum Allgemeine Sprachwissenschaft
Berlin, Germany, D-10117
solt@leibniz-zas.de