Numerical Concepts in Context

Paola Gega, Mingya Liu , and Lucas Bechberger

1 Introduction

Numerical concepts are an integral part of everyday conversation and communication. While mathematicians assign a precise interpretation to a natural number, e.g., 5 being exactly 5, the use and understanding of numerical expressions in natural language have a high variability. Broadly speaking, scientists use numbers more precisely when they discuss their research results (for example, 0.051 and 0.049 make a big difference in term of statistical significance) than street vendors at a flea market of Berlin (e.g., 51 or 49 cents for a broken antique glass are probably equally good results). In addition to broad context, narrower context such as questions under discussion (QUDs, Roberts 1996) or decision problems can influence the interpretation of numerical expressions as well: If a waiter asks “How many beers would you like to order?”, we mean exactly 10 when we say 10, no more no less. If a student is eligible for taking the exam with 2 assigned tasks, s/he is eligible with 2 assigned tasks—2 means at least 2. In contrast, if a student can pass the exam with 10 mistakes, 10 means at most 10. Furthermore, the interpretation of numerical expressions can also be subject to individual and developmental factors (e.g., Musolino 2004). In

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-69823-2_5) contains supplementary material, which is available to authorized users.

P. Gega (✉)

Institute of Philosophy, University of Bochum, Bochum, Germany

e-mail: paola.gega@rub.de

M. Liu (✉)

Department of English and American Studies, Humboldt University of Berlin, Berlin, Germany

e-mail: mingya.liu@hu-berlin.de

L. Bechberger

Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

e-mail: lucas.bechberger@uni-osnabrueck.de

© The Author(s) 2021

L. Bechberger et al. (eds.), *Concepts in Action*, Language, Cognition, and Mind 9,
https://doi.org/10.1007/978-3-030-69823-2_5

this paper, we will focus on the interpretive variability of numerical expressions in narrow linguistic contexts, namely, the nature of a number itself, and its co-occurring expressions.

Among others, the interpretation of numerical expressions depends on the perceived “roundness”: Round numbers (e.g., 50) can have both an imprecise or precise interpretation, whereas non-round numbers (e.g., 47) tend to have a precise interpretation. For example, Krifka (2002, 2007, 2009) proposes a “RNRI” (round numbers round interpretation) principle: “Round number words tend to have a round interpretation in measuring contexts”. Supporting evidence comes from the highly frequent use of round numbers in, among others, newspapers or street/distance signs, even though statistically speaking, it is very unlikely that the results of measurements are round more frequently than they are not (given sensitive instruments). In (1), taken from the Leipzig Wortschatz Corpus (Goldhahn et al. 2012), it is intuitive to assume that all the numerical expressions have an imprecise interpretation.

- (1) a. **Forty thousand people** in the state remained without water, and **26,000 people** were without electricity, she said, warning once again that people should stay inside.
- b. Gibraltar Airport - Located just **500 meters** from the city center, Gibraltar’s airport landing strip shares space with one of the island’s main roads.

Another piece of evidence is shown in the contrast between (2a) and (2b). Whereas (2a) is acceptable to characterize situations where John made 49 cupcakes, the use of (2b) is degraded in the same contexts. This shows that in contrast to round numbers, non-round numbers have a precise interpretation.

- (2) a. John made **50 cupcakes**.
- b. John made **48 cupcakes**.

A second factor contributing to the varying interpretation of numerical expressions is the type of approximator used in the expression. Precise approximators (e.g., *exactly*) impose a precise interpretation, whereas imprecise approximators (e.g., *roughly*, *approximately*, *about*) do the opposite, see (3a). However, due to the tendency of non-round numbers receiving a precise interpretation, it has been pointed out in Sauerland and Stateva (2011)¹ that it is odd to use them together with imprecise approximators, as can be seen in the contrast in (3b).

- (3) a. John made **exactly/roughly 50 cupcakes**.
- b. John made **exactly/?roughly 48 cupcakes**.

While the first and the second factors have received extensive treatment in the literature (a.o., Lakoff 1973; Rips et al. 2007; Krifka 2007, 2009; Sauerland and Stateva 2011; Kennedy 2013; Solt 2014), there is a third factor affecting the

¹What should also be noted with respect to approximators is Geurts’ (2006) sharp observation that precise approximators can only modify expressions that already have an exact meaning—while *exactly five sneezes* or *precisely half the cake* are perfectly acceptable expressions, *exactly tall* or *exactly some cookies* are not.

interpretation of numerical expressions which to our knowledge has largely been unexplored, namely, the unit of measurement. Consider (4): the combination of an imprecise approximator and a non-round number is not odd, which stands in contrast to “roughly 48 cupcakes” in (3b). The difference between the targeted expressions is that the unit “cupcake” in (3b) is discrete and the one in (4) “meter” is continuous.

(4) The tower is **exactly/roughly 48 meters** high.

The current paper examines these three factors in detail, as well as their ways of interaction. The paper is structured as follows. In Sect. 2, we provide a review of related works from theoretical linguistics. In Sect. 3, we report on a corpus-linguistic study with the following main findings: imprecise approximators occur more frequently with round numbers (e.g., *roughly* 50) than with non-round numbers (e.g., *roughly* 48). Furthermore, discrete units occur significantly less frequently than continuous units in the latter combination (e.g., *roughly 48 people* vs. *roughly 48 meters*), which indicates the imprecise nature of the continuous unit. In Sect. 4, we report a rating study testing the naturalness of imprecise approximators in combination with different kinds of numbers and different kinds of units. Our results show both effects by Number and Unit but no interaction between them. Section 5 provides a general discussion and concludes the paper.

Generally speaking, this chapter provides insights into the representation and application of numerical concepts. We focus our research on the usage and interpretation of these concepts in natural language texts, using the results of both a corpus study and a rating study. In our literature review, we summarize different formal models for representing the meaning of numerical expressions, which can be seen as (partial) representations of numerical concepts. In our two studies, we then seek to confirm the qualitative predictions made by these models about the practical usage of such numerical expressions. Our work can be related to the contribution by Gust and Umbach (Chap. 4) who also consider the granularity of interpretation for natural language phrases. While their work targets similarity expressions of varying kinds, we put our focus on expressions that involve concrete numbers. Our experimental rating study can be related to the procedure by Scerrati et al. (Chap. 6) who record binary responses on individual words, while we make use of Likert scale ratings on complete sentences. Finally, the focus on the interpretation of natural language phrases is also investigated by Vernillo (Chap. 8), who uses a theoretical analysis of individual verbs based on image schemata, while we perform a corpus study and a rating study on more complex phrases.

2 Theoretical Background

In this section, we provide a detailed discussion of the three linguistic factors influencing the overall interpretation of numerical expressions, based on the literature. As our concern is on their semantics and pragmatics, we assume a simplified “NumP”

(i.e., number phrase) structure for them consisting of a NumP-modifier (e.g., *exactly*), a Num head (e.g., *fifty*), and an NP complement (e.g., *people*), but are open to alternative syntactic structures.

2.1 Number: Round Versus Non-round

The discussion of round in contrast to non-round numbers is heavily intertwined with the topic of the granularity of scales in which we think. Thinking on a coarse-grained level can be seen as thinking in gross bins. A fine-grained, possibly continuous (i.e., maximally fine-grained) scale is simplified by turning it into a discrete scale with fewer values, therefore coarse-grained thinking means simplified thinking. While these few values are salient and meaningful in the way that we can quickly process and interpret them in a given context, using coarse-grained scales potentially results in less precise reports in measuring contexts.

If we look at scales of different granularity levels such as (5), we will find that round numbers appear both on fine-grained and on coarse-grained scales. This is not the case for non-round numbers—the more coarse-grained a scale becomes, the fewer non-round numbers it contains.

- (5) Scales progressing in steps of 10, 5 and 1 respectively
 - a. 0.....10.....20
 - b. 0.....5.....10.....15.....20
 - c. 0..1..2..3..4..5..6..7..8..9..10..11..12..13..14..15..16..17..18..19..20

Only values on a coarse-grained scale however can represent a whole range of other values; thus, since the values appearing on coarse-grained scales usually are round numbers, round numbers logically allow for an imprecise interpretation. In contrast, non-round numbers do not appear on coarse-grained scales and therefore only lend themselves to a precise interpretation. Thus, one would rather interpret expressions imprecisely that make available an imprecise interpretation than expressions that do not allow such an interpretation. This is why we tend to interpret round numbers imprecisely and non-round numbers precisely.

But what does ‘round’ really mean? The concept of roundness depends on the context. Solt (2014) speaks of a gradient nature of roundness, meaning that there is a ‘more’ and a ‘less’ to roundness: the hierarchical ordering of scales with respect to granularity yields this gradient. For example, 5 can be considered round since it also appears on the more coarse-grained scale (5b), but less round than 10, which appears on an even more coarse-grained scale (5a). In some cases, a number might be considered round if it only has—or is rounded to—two decimal places. In other cases, non-round numbers can take on the same function as round numbers, such as 12 or 24 h in a coarse-grained time scale (see more examples in Krifka 2007). In other words, the availability of an imprecise interpretation of a number does not necessarily depend on it being round; it rather depends on its coarse-grainedness within a system of representation. As our numerical reasoning most commonly makes

use of the decimal system however, which is a base-ten numeral system, round numbers like 10, 100, etc. and simple fractions of them most frequently coincide with coarse-grainedness and are thus more likely to be interpreted imprecisely.

Krifka (2007, 2009) assumes two general pragmatic principles from which he derives (and which shall explain) the RNRI (“Round Numbers, Round Interpretations”) phenomenon: (I) weak preference for simple expressions, (II) strict preference for truthful interpretations. The first principle explains why round numbers are used more imprecisely than non-round numbers. The second principle explains why round numbers are interpreted more imprecisely than precisely.

In more detail, Krifka assumes a conditional preference for simple expressions, which explains the approximate usage of round numbers in contexts that do not require high precision. If a speaker has the choice between uttering forty-eight or fifty, he will most likely choose the simpler expression, for reasons of communication efficiency. The preference is conditional in the sense that it can only come into effect if the difference between the two numbers is not relevant in the context (e.g., with specific QUDs or decision problems). Under a precise interpretation, however, the preference cannot come into effect; the speaker does not have the choice between one expression or the other. Krifka models the virtual equivalence between two measure expressions in low-precision contexts in the following way: Under an approximate interpretation, numbers represent ranges which can be characterized by a mean, i.e., the number which the interval is centered around, and a standard deviation, defining the borders of the interval, which also indicates the level of imprecision.² Naturally, ranges of two numbers can overlap if the values are close to each other. Two numbers are said to be indistinguishable from each other under an approximate interpretation if the ranges they represent overlap in such a way that their means are within their standard deviations. Under an approximate interpretation, forty-eight could for instance represent the range [46, 47, 48, 49, 50] (having the mean 48 and the standard deviation 2), whereas fifty would represent [48, 49, 50, 51, 52] in that case. Their means are within their standard deviations, so they are considered indistinguishable under this approximate interpretation. However, fifty has the advantage over forty-eight in that it has a simpler form (and is also otherwise more cognitively salient). The speaker thus chooses to utter fifty instead of forty-eight in a context where approximate interpretations are licensed. This also explains why non-round numbers are not interpreted in an approximate way: Once there are several indistinguishable alternatives one could make use of when reporting a measurement, the alternative with the simplest form is chosen, which excludes non-round numbers from the race.

Under a precise interpretation, numbers denote only themselves: forty-eight denotes 48 and fifty 50. The possibility of choosing between alternatives does not arise because their denotations are clearly different.

- (6) a. John made **50 cupcakes**.
- b. John made **48 cupcakes**.

²More specifically, Krifka models an imprecise number as a normal distribution which is centered around the number. To simplify things, he confines his discussion to a representation in terms of intervals.

Assuming a context which licenses an approximate interpretation, Krifka's model explains the acceptability of (6a) since fifty represents the range [48, 49, 50, 51, 52] which includes 48 and 51. If the context requires a precise interpretation, fifty represents only 50; the usage of this numeral thus would make (6a) false in situations where John made 48 or 51 cupcakes. Similarly, forty-eight in (6b) could represent the range [46, 47, 48, 49, 50] under an approximate interpretation. However, the speaker would have uttered fifty in such a situation, since under an approximate interpretation fifty is indistinguishable from forty-eight, and it is simpler. Thus, forty-eight cannot be interpreted imprecisely here—instead, it must denote solely its own value.

The second principle ought to explain an assumption specific to Krifka's theory. By way of principle (II), the preference for truthful interpretations, Krifka explains why an approximate interpretation of an encountered round number is more sensible than a precise one. Krifka holds the assumption that we prefer an imprecise interpretation of round numbers and therefore usually interpret round numbers imprecisely (an assumption challenged by Ferson et al. 2015). He argues that an imprecise interpretation maximizes the probability of truth of the statement: It is more likely that the value of a reported measurement is in the range of the interval around the reported number (which amounts to an approximate interpretation) than it is likely that the value is the number itself (which amounts to a precise interpretation). And since Krifka also assumes that we follow principle (II), he concludes that the approximate interpretation is the preferred one. On the other hand, an addressee can conclude from an utterance containing the more complex expression that a precise interpretation must have been intended since this is the only context where complex expressions are used—whenever possible, i.e., under an approximate interpretation, the simpler expression (which coincides with round numbers in this case) is chosen over the more complex alternative.

So far, Krifka's argumentation had little to do with a theory of granularity. One might ask however why it is generally the case that round numbers are simpler than non-round numbers. It turns out that the superficial simplicity argument can be reformulated in terms of the scale granularity framework. Krifka points out that it is not just the simplicity of the form of some expression that contributes to whether it is interpreted precisely or imprecisely. Instead, what matters even more is the expression's simplicity in terms of representation. This is where scale granularity becomes important. The simplicity of representation is marked by whether a value is cognitively salient on the scale of reference.

A numerical representation might be perceived as simple (more easily graspable) if it appears on coarse-grained scales of the unit. It becomes clear that the term *simple* here refers to how easily we can process the conveyed bit of information, as in the aforementioned example of time scales {0, 12, 24, 36, 48, ...}. Notice that twenty-four is neither simpler than twenty-three in terms of form nor round. It is because of the expression's simplicity of representation and persistence throughout scales of different granularity levels that a speaker might choose twenty-four over twenty-three under an approximate interpretation.

We can conclude that a simple representation promotes an imprecise interpretation because it allows one to reason on a coarse-grained level of scales. Krifka additionally

argues that in many cases, simplicity of expression and simplicity of representation coincide—not coincidentally, but because the frequency of use dictates such a development. Simplification of expressions is a result of an increase of frequency due to their additional approximate use: “salient representations tend to be shorter, and tend to be shortened in language change” (Krifka 2007).

Generally speaking, a characteristic of a round number is that it is simple: self-contained (no infinite decimal places) and conceptually graspable and decodable; it is a number that exists in a simple system of representation (for instance a system of multiples of tens)—the system depends on the context of use. In this paper, we will restrict our empirical analyses to a limited set of (conventionalized) round numbers (e.g., 10-roundness and 5-roundness, which do not need contextual support) in contrast to their non-round close numbers.

2.2 *Approximator: Approximate Versus Exact*

While we have discussed that (im)precise interpretations of numerals can arise from implicit assumptions about the numbers themselves, there is also an overt means for marking the intended level of precision. Approximators like *exactly*, *precisely*, *around*, and *approximately* are classified as hedges (Lakoff 1973): Expressions which modify the certainty, force, or precision implied by statements. Also belonging to this class are expressions like *maybe* or *I assume* (called *shields*), which can modify whole sentences. Approximators are a means of explicitly marking the degree of precision with which a measure expression is to be interpreted, but on a different level, the use of approximators also reveals something about the certainty with which a speaker utters something. The latter is evident if we consider uses of the approximators as speech-act adverbs, e.g., *Roughly speaking, I have 50 students in my class*. We leave it for future studies what differences such sentences have compared to *I have roughly 50 students in my class*.

When a speaker intends to indicate a high certainty about the accurateness of the uttered numeral, they likely use precise approximators. When doing so, the speaker simultaneously decreases the risk of conveying false information, which is higher with an unmodified alternative. In other words, using approximators increases the probability of the truthfulness. Thus, using imprecise approximators can also signal the speaker’s uncertainty in addition to imprecision in measuring, which is emphasized in Ferson et al.’s (2015) work.

While Krifka’s (2007) work is not concerned with the effect of approximators on numerical expressions, Solt (2014) extends the granularity-based framework to provide an account of these modifying expressions. She also introduces a new formalism for determining truth or falsity of sentences with numerical expressions that includes a contextually determined granularity level. In her analysis, the overt use of approximators in combination with numerals is modeled as a mapping from point-denoting expressions (the bare numerals) to intervals around these expressions.

Explicitly modified numerals thus denote a scalar segment. Solt formally defines the semantics of approximators as in (7):

$$(7) \quad [[\text{APPROXIMATOR } n]]^g = (n - \text{gran}'/2, n + \text{gran}'/2)$$

For imprecise approximators, gran' is the coarsest possible unit for a granularity level one could choose given the context. For precise approximators, gran' is the finest possible choice of a granularity level given the context. Thus, $[[\text{about } 50]]^{g[\text{gran}'=10]}$ would denote the interval (45, 55) in the appropriate context. It becomes clear that the denotation of a modified measure expression differs from the original numeral in that it (roughly) denotes the range of values halfway between the neighboring values on the coarse-grained level. In formal semantic terms, this complex expression however still is of type “degree” despite not denoting a point.

$$(8) \quad [[\text{exactly fifty}]]^{g[\text{gran}'=0.01]} = (50 - 0.01/2, 50 + 0.01/2)$$

Notably, Solt’s analysis of approximators, as shown in (8), yields as a result that precise approximators can make an expression more imprecise after being combined with the approximator. Although the granularity level is very fine-grained (with gran' being 0.01), the resulting complex expression denotes a more coarse-grained degree than the bare, unmodified numeral, namely, (49.995, 50.005) instead of 50. On the one hand, the analysis of the complex expression is not counterintuitive since in some contexts the usage of a precise approximator does not signal maximal but only increased precision. However, what seems unintuitive is that the bare numeral in contrast can never denote anything more imprecise than the maximally precise point it always denotes. The denotation of the numeral modified by a precise approximator is more imprecise than the denotation of the unmodified numeral. This conflicts with the empirical findings of Ferson et al.’s (2015) study that precise approximators (*exactly* and *precisely*) rather reduce a previously assumed range of imprecision associated with a numeral instead of making numerals more imprecise.

Since Solt’s theory does not assume numerals to denote ranges in the first place, there is no way she can model how an approximator can reduce the interval of imprecision that might be associated with a numeral. Thus, this analysis cannot explicitly model situations in which the context favors a default imprecise reading of a numeral while the approximator is used to override this reading. This is only possible within theories that overtly model the imprecision of a numeral such as Krifka who lets unmodified numerals denote ranges under an imprecise interpretation. These representational issues Solt’s theory faces due to the assumption of a monosemous exact denotation of numerals might not pose problems in terms of truth-conditional analyses. However, they show that Solt’s model is also not entirely optimal as it seems odd to assume that *exactly fifty* denotes a coarse-grained degree while *fifty* does not.

An alternative relates to Lasersohn’s theory (1999) of pragmatic halos in which he also proposes an analysis of approximators. Lasersohn takes precise approximators to be narrowing the so-called “pragmatic halos” of an expression: “Suppose, for illustration, that there are two points in time close enough to *i* that the difference

between them and i is ignored in context, so that the halo of *three o'clock* is the set $\{i, j, k\}$, ordered according the relation of closeness to i The real effect of *exactly* is on pragmatic halos: we want the pragmatic halo of *exactly three o'clock* to include those elements of the halo of *three o'clock* which are closest to i (that is, to the actual time of *three o'clock*), eliminating outlying elements.” (Lasersohn 1999: p. 528). In this analysis, precise approximators have no effect on the semantic level, however, they reduce the pragmatic slack with which one may speak and thus have an effect on whether an utterance can be used felicitously or not. This is not the case for imprecise approximators: They are analyzed to have the effect of expanding the denotation of the expression (they are combined with) into its halo. Thus, they have a clear truth-conditional effect in that the resulting denotation is ‘enriched’ by similar denotations, constituting a set.

Combining Sects. 2.1 and 2.2, a natural question arises as to how numbers interact with approximators. We will not be able to work out a formal analysis here, but focus on the distributional constraints due to the different levels of precision encoded in them.

2.3 Unit: Discrete Versus Continuous

Seeing numbers as part of a mathematical system, we find that at the most basic level, number systems permit the description of quantities by means of expressions consisting of a numeral and a unit, where the unit specifies the scale of measurement. Units can, for instance, be ‘people’, ‘buildings’, ‘chairs’ for discrete quantities, but also ‘days’, ‘acres’, ‘metres’ for continuous quantities.

Accordingly, a numeral can be an integer or real-valued; it furthermore can be expressed in words or numerical digits. Since units measure either discrete or continuous quantities, they can influence the numerals they appear with. Those units measuring discrete quantities restrict the numeral they combine with to the domain of integers. When measuring quantities physically, the numerical expressions used for description are almost always used imprecisely, especially in the case of measuring continuous quantities. Ferson et al. (2015) thus suggest a distinction between the mathematical and the ‘real world’ interpretation of a numerical quantity. Following this distinction means assuming that in non-mathematical contexts an unmodified scalar number already elicits an interpretation with an interval of imprecision; the expression might refer to any value within this interval. In contrast to this suggestion, however, Ferson et al.’s (2015) empirical study found that participants (who were asked to specify an interval the numbers can stand for) interpreted bare, unmodified numbers precisely in 94% of the time, despite the fact that the expressions were embedded in a natural language context.

What are the effects of units on the distribution and interpretation of number words and expressions? We will provide partial answers to this understudied question in the rest of the paper.

2.4 Summary

In summary, the use and understanding of numerical expressions are subject to influences from both broad discourse contexts and narrow linguistic contexts. In the paper, we will not provide formal analyses for numerical expressions; instead, we focus on the empirical testing of the observations from the literature and the current work. In the following, we will discuss numerical expressions with two goals: First, we will provide empirical (i.e., corpus- and psycholinguistic) evidence for the generalizations related to the distinction between round and non-round numbers. Second, we will provide empirical evidence for the effect of unit in the interpretations of numerical expressions.

3 Corpus Study

3.1 Hypotheses

The aim of the corpus study is, first of all, to support the initial observation made, namely that round numbers seem to appear more frequently in natural language contexts than expected if they only had a precise usage. If confirmed, this more frequent appearance is taken as support for the claim that round numbers, in addition to denoting their own values, are used imprecisely due to context (e.g., when imprecision prevails over precision, or when the speaker is uncertain about the actual precise values). Their additional use for this purpose would explain the prevalence of round numbers throughout natural language data. Furthermore, the analysis has been conducted to shed light on the distribution of approximator (null/precise/imprecise), numeral (round/non-round) and unit (discrete/continuous), as well as possible patterns in their conjoint appearance.

Based on the theoretical considerations in Sect. 2, we started with the following hypotheses where π_i denotes the probability of the number i occurring in natural language communication:

- (9) H0: $\pi_1 = \pi_2 = \dots = \pi_{500}$
 H1: $\pi_1 \neq \pi_2 \neq \dots \neq \pi_{500}$

In the null hypothesis H0, each numeral is assumed to appear with an equal probability in the corpus. The corpus study restricts numerical analysis to numerals in the range between 1 and 500, hence the notation above. Say the probability of appearance of each numeral is $1/500$, then we expect round numbers (i.e., numbers ending with a 0 or 5) to appear 20 percent of the time (100 out of the 500 numbers are round) whereas non-round numbers should appear 80 percent of the time (the remaining 400 out of 500 numbers are non-round).

Our first hypothesis is captured in the H1: We expect that the probability of appearance is not equal for every numeral. More specifically, related to H1, we assume that round numbers appear more often than expected (i.e., >20%).

Secondly, we assume that the default interpretation of numerals in general is precise, following Ferson et al. (2015) and the findings in their study (and contrary to Krifka 2007). As a consequence, a precise interpretation often does not have to be signaled explicitly whereas imprecise approximators are needed to signal an intended imprecise interpretation. Thus, our second hypothesis is that precise approximators appear less frequently than imprecise approximators.

Thirdly, in terms of combinations of approximators and numerals, let us recall example (3b) or (10a) from Sauerland and Stateva (2011), which they take to be odd. Since imprecise approximators usually signal a coarse granularity level, the appearance with a non-round number (which only appears on more fine-grained scales) strikes the reader as peculiar. We will therefore expect that imprecise approximators tend to appear with round numbers.

- (10) a. # What John cooked were approximately 49 tapas.
b. The rope is approximately 49 metres long.

Furthermore, theoretical accounts so far mainly focused on the interaction between approximators and numerals. Ferson et al. (2015) examined a potential influence of the unit on the interpreted imprecision of a numeral, a hypothesis that was not supported by the results of their study. To our knowledge, little attention has been paid to the potential interaction between unit, approximator, and numeral, see for instance, (10b). Whereas (10a) is odd to the reader, this oddity disappears in (10b), which is completely natural. This can be attributed to the fact that the continuous unit implies that 49 m can already be used imprecisely (49 is round compared to 48.7) whereas this is not the case for discrete numbers (49 is the most precise possible in this case and has no imprecise reading). The results of the corpus study will also be inspected with respect to this effect.

3.2 *Methods*

The study was based on the Leipzig Wortschatz corpus (Goldhahn et al. 2012), containing 1 million English sentences sourced from online news reports and general web crawling results. The corpus was searched for numerical expressions in the Approximator-Number-Unit fashion. The code was written in python and is publicly available online (<https://github.com/lbechberger/CorpusStudyNumerals>). The matches were analyzed with respect to the following variables:

- (11) Variables
a. Approximator: precise, imprecise, null
b. Number: round, non-round
c. Unit: discrete, continuous

Counts kept track of the different combinations. Numbers were counted as round if they ended with a zero or five; we only used integer numbers (excluding decimal numbers) in the analysis. We only included number words up to five hundred in the counts. The categories for the approximator matched for the following words:

- (12) Categories of approximators (Approx.)
 - a. Precise Approx.: ['exactly', 'precisely']
 - b. Imprecise Approx.: ['about', 'approximately', 'roughly', 'around', 'round about', 'roughly around', 'some'³]
 - c. Asymmetrical Approx.: ['more than', 'nearly', 'over', 'almost', 'approaching', 'below', 'above', 'fewer than', 'less than', 'at most', 'at least', 'close to', 'near to', 'up to', 'as high as', 'as low as', 'not quite']
 - d. Null Approx.: every expression preceding a numeral that does not match the words above

Asymmetrical approximators (based on Ferson et al.'s (2015) list of approximators used in his study) were not included in the statistical analysis. Yet, they were also matched to obtain an estimate of the frequency of their usage and have a more accurate account of the unmodified versus modified numerals ratio. Their appearance with either round or non-round numbers was neither recorded nor analyzed (although asymmetrical approximators, a.k.a. comparatives, are also a subject of debate in current accounts of imprecision (Solt 2014)). The unit was first matched as any word following the numeral and subsequently evaluated using WordNet (Princeton University 2010) for whether it belonged to one of the following categories:

- (13) Categories of units
 - a. continuous: ['time period', 'time unit', 'linear unit', 'magnitude relation', 'monetary unit', 'unit of measurement']
 - b. discrete: ['organism', 'human activity', 'group', 'location', 'transport', 'material']

All numerals occurring with matches that did not belong to any of the categories have been excluded; the remaining matches were used for the analysis. The data consequently had the nature of frequency counts of the aforementioned (*Approximator*)-*Number-Unit* sequences and of the respective counts of approximators, numerals, and units separately. The analysis consisted of testing the match counts against their expected frequency: The main hypothesis, that the frequency of round numbers is different from their expected frequency, was tested for significance using the Binomial Test. The effects in the Number (roundness) * Approximator and Unit * Approximator contingency tables were tested using the χ^2 Test.

³Such as *Some 50 students joined the protest.*

Fig. 1 Numeral counts from 0 to 100

Fig. 2 Numeral counts from 0 to 50

3.3 Results and Interpretation

As can be seen from Figs. 1 and 2, there are “spikes” of counts for round numbers (also visible in the range between 0 and 100) already indicating a marked appearance of round numerals in the corpus. The general distribution (a few numbers with very high counts and a tail to the right) suggests that numeral occurrences seem to follow a power law distribution, specifically one related to Benford’s law (Benford 1939). The extraordinarily high count for the numeral 1 can be explained by the frequent usage of the number word in many contexts (e.g., ‘He had one goal.’, ‘A government has the energy for only so many fights at one time.’, etc.).

Table 1 Frequency counts of matches in the corpus: Approximator * Number (roundness) * Unit

n = 182,895	Discrete		Continuous		
	Round	Non-round	Round	Non-round	
Precise	4	7	15	36	(62)
Imprecise	2,975	423	3,217	2,081	(8,696)
Null	21,215	64,990	30,535	57,397	(174,137)
	(24,194)	(65,420)	(33,767)	(59,514)	

Table 2 Frequency counts of matches in the corpus: Number (roundness) * Approximator

n = 182,895	Precise	Imprecise	Null	
Round	19	6,192	51,750	(57,961)
Non-round	43	2,504	122,387	(124,934)
	(62)	(8,696)	(174,137)	

182,895 of the matched numerals were used for the analysis (another 369,384 in that range were discarded due to unit constraints). The null hypothesis thus expects 36,579 of these numerals to be round and 146,316 numerals to be non-round.

Generally, as in Table 1, we observe the following tendencies: First, non-round numbers appear, in absolute terms, more often than round numbers. Second, unmodified numerals appear most frequently with a count of 174,137, followed by numerals modified by an imprecise approximator (8,696 counts) and lastly, numerals modified by precise approximators (62 matches). Third, numerals with discrete units (89,614 counts) appear almost as often as numerals with continuous units (93,281 counts), with a ratio of approximately 0.49/0.51.

More specifically, our findings are stated as follows, see Table 2. First, round numbers appear more frequently than expected. As we can read from the tables, a total of 57,961 (as opposed to the expected 36,579) round numbers and a total of 124,934 (as opposed to the expected 146,316) non-round numbers were counted. Instead of an expected 0.2/0.8 ratio, we found a ratio of approximately 0.32/0.68. This effect is particularly pronounced if the numerals appear with a continuous unit—the ratio between round and non-round numbers is roughly 0.36/0.64 there. Binomial testing reveals that this is a significant departure from the expected frequency ($p < 0.01$, one-sided).

Second, imprecise approximators appear more frequently than precise approximators. Table 3 shows a total count of 8,696 imprecisely modified numerals as opposed to the few 62 occurrences of precisely modified numerals in the given range. This undoubtedly supports our assumption that the default interpretation of numerals is precise which makes imprecise approximators an important tool to signal that the imprecise interpretation is intended, whereas precise approximators are unnecessary most of the time.

Table 3 Frequency counts of matches in the corpus: Unit * Approximator

n = 182,895	Precise	Imprecise	Null	
Discrete	11	3,398	86,205	(89,614)
Continuous	51	5,298	87,932	(93,281)
	(62)	(8,696)	(174,137)	

Table 4 Breakdown of Table 1 with respect to Unit

	Discrete		Continuous		
Imprecise	Round	Non-round	Round	Non-round	
	2975	423	3217	2081	(8696)

Third, imprecise approximators tend to appear with round numbers, especially if the unit is discrete. This is one of the most impressive results from the study: Even though generally and in absolute terms, non-round numbers occur more often than round numbers, we can read from Table 2 that if numerals occur with an imprecise approximator, the proportions are almost swapped. Even in absolute terms, imprecisely modified round numerals occur more often than imprecisely modified non-round numerals. This represents strong evidence for our hypothesis that imprecise approximators predominantly appear with round numbers. The deviations from the expected frequencies in Table 2 were significant using the χ^2 Test, i.e., χ^2 ($df = 2$, $n = 18,2895$) = 6585.259, $p < 0.01$, $\varphi_c = 0.19$. Conversely, this finding can be framed in terms of the infrequent appearance of imprecise approximators with non-round numbers (see Sauerland and Stateva's (2011) oddity example (10a) mentioned). Arguably, 2,504 occurrences of non-round numerals appearing with imprecise approximators is still a substantial count. Resolution however comes from looking at Table 4 where a further breakdown of the data with respect to the unit category is presented:

We see that this effect is particularly strong if we are looking at the discrete domain: There were 2,975 occurrences of the imprecise approximator-round numeral combination, whereas only 423 non-round numerals appeared with an imprecise approximator there (roughly an impressive 0.88/0.12 ratio). This is in line with Sauerland and Stateva's (2011) observation about imprecise approximators occurring with non-round numbers. In contrast, in the continuous domain, this effect vanishes for the most part (compare (10b)). This is also reflected in our counts: Although it is still the case that imprecise approximators occur more often with round numbers in this condition, the count for imprecise approximators appearing with non-round numbers is almost equally high and in absolute terms not negligible. This indicates that the oddity of imprecise approximators appearing with non-round numbers is drastically reduced if these numbers are continuous. We have thus encountered evidence for the claim that the unit has an effect on the co-occurrence behavior of approximators and numerals.

Last but not least, precise approximators tend to appear with continuous units. We see that if precise approximators appear at all, they tend to co-occur with continuous units (51 occurrences with continuous quantities vs. 11 occurrences with discrete quantities, see boldfaced numbers in Table 3). This makes sense to the extent that for continuous quantities, the precise interpretation is not trivial. These observations, however, should be taken with a grain of salt as we did not have many occurrences of precise approximators overall.

- (14) a. Trump announced his candidacy for the Republican nomination **exactly three months** ago.
 b. Belgium's federal prosecutor's office says authorities have so far made **(?exactly) three arrests** linked to the deadly attacks in Paris.

While *exactly* adds nothing to the already precise interpretation of (14b), in (14a), it makes a contribution to the interpretation of the numeral. Since the used numeral in (14a) can never be entirely accurately describing the actual time span between Trump's announcement and the report, the degree of accurateness needs to be marked explicitly to indicate "how precisely" the expression is meant. In (14a), one can assume that the speaker intended an interpretation accurate to the day (i.e., the report was made on the same date of the third subsequent month). Unless the numeral is of special interest, *exactly* in (14b) in contrast, appears redundant.

4 Psycholinguistic Experiment

To investigate the effect of the unit on the acceptability of numerical expressions, we tested English numeral expressions using a 2×2 factorial design, with the factors *Number* (round vs. non-round) and *Unit* (discrete vs. continuous).

4.1 Materials and Predictions

We used 24 different matrix sentence items, each in four conditions, see the Appendix for the entire list of the items. The experimental items were constructed under the following objective: The setout was to choose sentences containing the sequence *imprecise approximator—round number*, which has been motivated to evoke no perception of oddity. The sentences were picked from the Leipzig Wortschatz corpus. Before selecting the sentences, we determined the round numbers that they ought to contain. For this, 12 round numbers were randomly chosen in the range from 10 to 1000. This yielded the numbers 10, 60, 70, 100, 350, 400, 700, 750, 800, 900, 950, 1000. We then scanned the corpus for sentences containing imprecise approximators (*about*, *around*, *approximately* and *roughly*, six occurrences each) and the randomly chosen round numbers that would appear with either discrete or continuous

units, resulting in equally many sentences for both the ‘discrete’ and the ‘continuous’ condition.

Based on the experimental items for the *round* conditions, we created their non-round counterparts by changing the round number of each sentence into a close-by non-round one. This way we ensured that the non-round number would appear in a plausible context and linguistic environment. The oddity could thus only arise from the pairing of a non-round number with an imprecise approximator. The four conditions are exemplified in (15).

- (15) a. **r-disc**: As of then, about 60 Cubans had arrived in the Yucatan coast in 2015.
 b. **r-cont**: Brigham City is about 60 miles north of Salt Lake City.
 c. **nr-disc**: As of then, about 61 Cubans had arrived in the Yucatan coast in 2015.
 d. **nr-cont**: Brigham City is about 61 miles north of Salt Lake City.

Additionally, we used 48 filler items as distractors, which were news report sentences of comparable length that we also sourced from the Leipzig Wortschatz corpus. We did not revise these, as the pragmatic difference at focus is subtle and thus fillers containing ungrammatical or odd phrases would be inappropriate.

- (16) *The drug investigation began in August 2013 at Edwards Air Force Base in California.*

Based on Sauerland and Stateva’s (2011) observation that non-round numbers are odd with imprecise approximators and our corpus-linguistic finding that this effect is stronger with discrete units than continuous units, we had the following predictions: First, there will be a main effect of Number. More specifically, the condition “nr-disc” will be rated worse than “r-disc”, and the condition “nr-cont” will be rated worse than “r-cont”. These predictions are in accordance with the oddity suggested by Sauerland and Stateva. Due to the observations made in the corpus study, we included a second prediction, namely, there would be a main effect of Unit and possibly an interaction between Number and Unit due to a stronger worsening effect with discrete than with continuous units.

We used a Latin Square design, that is, each participant read one set of 72 sentences in total. As seen above, the participants’ attention was directed towards the phrases of interest by marking the relevant phrase visually, both in experimental and in filler items.⁴ For the filler items, the marked phrases were mostly DP’s or PP’s (i.e., determiner phrases or prepositional phrases).

⁴We highlighted the critical phrases, as in the pretest without doing so, the subjects did not distinguish the conditions, raising the question whether it showed no evidence for the effect of unit, or whether it was due to methodological issues.

4.2 Procedure and Participants

The experiment was set up with Ibx Farm (spellout.net/ibexfarm/), a website that provides free hosting for online psycholinguistic experiments. Experimental data was gathered using Amazon MTurk, a crowdsourcing platform where human intelligence tasks (HITs) can be carried out by participants who receive compensation for each HIT completed. Requesters were provided the link to the experiment and compensated with \$4. Native English-speaking workers on Amazon MTurk (N = 72) signed informed consent and participated in the study.

Before entering the experimental phase, participants first completed a practice session where 12 practice items were to be rated. During the experimental phase, they first read an entire sentence and then were asked to rate the naturalness of the underlined phases (which were shown again separately) on a 7-point Likert scale (1 = unnatural, 7 = natural).

4.3 Data Analysis and Results

The descriptive statistics is provided in Table 5 and visualized in Fig. 3. As can be seen in the table, descriptively, the “r-cont” condition received the highest mean rating, whereas the “nr-disc” condition received the lowest mean rating. The standard deviation was also highest for the “nr-disc” condition, indicating an overall lower consistency in ratings for this condition.

We analyzed the data using R. All analyses were performed using mixed effects linear regression models; the models were constructed using the *lme4* package in R

Table 5 Mean naturalness ratings (1 = unnatural, 7 = natural) and standard deviations (SDs) and standard errors (SEs)

Conditions	Mean	SD	SE
r-disc	5.82	1.47	0.07
r-cont	5.90	1.39	0.07
nr-disc	5.11	1.77	0.09
nr-cont	5.33	1.70	0.09

Fig. 3 Naturalness ratings of the experiment

(Baayen et al. 2008; Bates et al. 2012). All contrasts of interest were sum coded and included as fixed effects in the model. The reported model is the maximal model that converged. The model included *Number* and *Unit* (with interaction term) as fixed effects. Furthermore, we included random intercepts for subjects, items, and stimuli order, as well as random by-subject slopes for the effects of *Number* and *Unit* (and their interaction).

We found a significant main effect of *Number* ($t = 4.15$, $p < 0.0001$). Tukey's HSD for multiple comparisons of means indicates that round numbers were rated significantly more natural than non-round numbers with both continuous ($t = 3.96$, $p < 0.005$) and discrete ($t = 3.84$, $p < 0.005$) units. Furthermore, we found a significant effect of *Unit* ($t = 2.11$, $p < 0.05$) in that continuous conditions were rated better than discrete conditions. However, there is no interaction between the two factors, which suggests that the effect of neither factor is influenced by the presence or absence of the other.

In this study, we were able to confirm our first predictions about the effect of *Number* and *Unit*. We will leave the reason for the lack of an interaction for future studies.

5 Discussion and Conclusion

In this paper, we tried to gain insight into our understanding and interpretation of numerical expressions with regard to questions such as whether numbers are imprecise at the semantic level.

5.1 *Numbers and Number Concepts*

We must keep in mind that the development of the number system as we know it now has been a process of cultural construction and added knowledge over generations and centuries of historical time. When analyzing how we interpret numerical expressions in natural language contexts, insight might be provided by looking at the innate numerical concepts humans (and non-human animals) are equipped with for reasoning quantitatively.

Our understanding of number proceeds from concepts that do not conform to the structure and characteristics of the natural numbers (Rips et al. 2007). Two main mechanisms for quantitative reasoning have been identified for numerical ability in infants and non-human animals: On the one hand, a system works with internal analog magnitudes—perhaps some type of continuous strength or activation—which is a linear function of the input. On the other hand, infants' skills for quantitative reasoning may also draw on discrete and distinct representations of objects that are kept in short-term memory—however only less than four items can be represented this way.

Explained shortly, a mental (i.e., internal analog) magnitude is an internal representation of a quantity—this can be the cardinality of a set, but also duration, length, or volume of whatever is registered by the organism. What is special about this representation is that it is assumed to represent an objective magnitude in a direct linear relationship—in that it constitutes a continuous quantity (e.g., activation strength) represented mentally that adjusts to achieve a measure of a quantity. It is thus suggested that mental magnitudes share the formal properties of real numbers (Gallistel and Gelman 2005). However, analog magnitude representations are noisy, and the noise linearly increases the bigger the quantities become. This means, the bigger the measured values, the more imprecise the representation.⁵ Analog magnitude representations of large sets are thus only approximate; they are a coarse representation, contrasting with the precision associated with natural numbers.

The other mechanism makes for an infant's ability to predict the total number of objects in small sets (less than 4) and might be considered conceptually closer to the elaborate concept we have of integers.⁶ It depends on attentional or short-term memory mechanisms that represent individual objects as distinct entities. For each object, there is a distinct representation within the four-object capacity limit. A set exceeding three items cannot be held in the infants' short-term memory (Carey 2004).

Many psychologists believe that full-fledged mathematical thinking mainly originates from these two innate concepts that are also shown to be existent in non-human animals. Although other researchers argue that these abilities do not seem to be adequate prerequisites for forming the mathematical concept we have of numbers within a number system (see Rips et al. 2007 for a discussion of this issue), they are still shown to have relevance in quantitative and even arithmetical reasoning (Gallistel and Gelman 2005). In specific, analog magnitudes are shown to play a role in arithmetical computations: comparison of two values, and also addition and subtraction (Carey 2004). Indeed, if analog magnitude representations are made use of in mathematical contexts, which would most of all require high precision representations, it is likely that they are also employed when encountering numerical expressions in a natural language context.

How, however, do these mechanisms play into the interpretation of numerical expressions in natural language, if they do so at all? Krifka (2007) argues that the existence of these two distinct systems of representation provides plausibility for both an exact and an approximate interpretation of numerals since they work in parallel and are not hierarchically ordered in any way. Which one of the two is the "original" meaning of a numeral is not settled by this argumentation, it might even be that there is none and that both interpretations are equally prevalent. All the findings in developmental research however do not comprise or imply an inherent distinction

⁵In this, mental magnitudes follow Weber's law, according to which the discriminability of two values is a function of their ratio: The bigger the physical magnitudes (and consequently the analog magnitudes), the harder discrimination between pairs of values that are separated by the same absolute difference becomes.

⁶Short-term object representations do have the discreteness of natural numbers; however, they do not form a set representation of the tracked objects and, consequently, cannot represent cardinality. This in turn is represented by mental magnitudes.

of round vs. non-round numbers with respect to impreciseness. Thus, at least the imprecise interpretation of round numbers (not the general imprecise representation of quantities) seems to be a phenomenon “on top of” the basic interpretation of numerals, which likely only started to develop after the formation of more elaborate mathematical systems.

5.2 Contributions and Outlooks of the Current Study

In the current study, we provide a critical discussion of numerical expressions based on the recent formal (compositional) semantic literature, focusing on the imprecise and precise interpretation of numerical expressions. While the interpretation of numerical expressions depends on both broad discourse context and narrow linguistic context, we only dealt with the latter. Our corpus and experimental studies show that the interpretation of numerical expressions is subject to the kind of numbers, the kind of units, as well as whether and what approximators co-occur with them.

It is to note that the results we obtained in our study are certainly contingent on, for example, the specific corpus study or experimental design, the specific numerals (i.e., 0–500) we used, and the specific contexts they occurred (in our case, naturally occurring contexts instead of made-up contexts as in usual experimental works), thus, whether and to what extent they apply to numerical expressions in general need to be investigated in further studies. Furthermore, approximators might differ among themselves. For example, even within the imprecise category, *roughly* and *some* as in *some 50 people* might have syntactic, semantic, or pragmatic differences, which we were not able to handle here. The same holds for Unit which might differ in terms of aspects other than discreteness. Another question for future studies is how the interpretation of numerical expressions is manipulated by broad context (such as QUD, decision problems, developmental, or individual differences, purely information exchanging vs. strategic communication, counting vs. measuring contexts, to just name a few parameters). Despite of this, we believe that the method and the findings of the paper have made further steps to understanding numerical concepts and related concepts that they modify.

Acknowledgements For this research, Mingya Liu received financial support from the Forschungspool of Osnabrück University in the program of “Halten & Holen forschungsstarker Wissenschaftler*innen”. We thank Ulf Krumnack, Nikola Kompa and Chris Kennedy for discussions related to parts of the paper, and Juliane Schwab for her help with data acquisition. All mistakes are our own, of course.

Appendix: Test Items of the Experiment (I./C. For Item/Condition)

I./C. Sentence

- 1/1 People from about 10 municipalities around the wrecked plant still cannot go home.
- 1/2 About 10 hours after the terror attack, someone approached police officers near a synagogue and started shooting.
- 1/3 People from about 9 municipalities around the wrecked plant still cannot go home.
- 1/4 About 9 hours after the terror attack, someone approached police officers near a synagogue and started shooting.
- 2/1 As of then, approximately 60 Cubans had arrived in the Yucatan coast in 2015.
- 2/2 Brigham City is approximately 60 miles north of Salt Lake City.
- 2/3 As of then, approximately 61 Cubans had arrived in the Yucatan coast in 2015.
- 2/4 Brigham City is approximately 61 miles north of Salt Lake City.
- 3/1 Reportedly, around 70 members of the security forces have been killed in attacks blamed on the PKK.
- 3/2 BBC Travel says it is currently taking around 70 minutes to get through the traffic.
- 3/3 Reportedly, around 69 members of the security forces have been killed in attacks blamed on the PKK.
- 3/4 BBC Travel says it is currently taking around 69 minutes to get through the traffic.
- 4/1 Experts say present-day cars have about 100 computers on board, and that will double in a few years.
- 4/2 The helicopter was about 100 feet off the ground when it went out of control and smashed into a parking lot.
- 4/3 Experts say present-day cars have about 101 computers on board, and that will double in a few years.
- 4/4 The helicopter was about 101 feet off the ground when it went out of control and smashed into a parking lot.
- 5/1 The tornado hit on Monday morning, leaving around 350 homes destroyed.
- 5/2 A hard-charging fire spread across around 350 acres in a matter of hours, the county fire department said.
- 5/3 The tornado hit on Monday morning, leaving around 349 homes destroyed.
- 5/4 A hard-charging fire spread across around 349 acres in a matter of hours, the county fire department said.
- 6/1 In total, about 400 competitors are entered in this year's Dakar Rally, in cars, trucks, motorbikes and quad bikes.
- 6/2 The Dutch designed Guyana's drainage system about 400 years ago.

- 6/3 In total, about 401 competitors are entered in this year's Dakar Rally, in cars, trucks, motorbikes and quad bikes.
- 6/4 The Dutch designed Guyana's drainage system about 401 years ago.
- 7/1 Nigerian armed forces attacked Boko Haram's headquarters and killed about 700 people, including its leader.
- 7/2 It was confirmed that there was about 700 tons of the deadly chemical stored at the warehouse that blew up.
- 7/3 Nigerian armed forces attacked Boko Haram's headquarters and killed about 699 people, including its leader.
- 7/4 It was confirmed that there was about 699 tons of the deadly chemical stored at the warehouse that blew up.
- 8/1 The run is taking place in approximately 750 communities across Canada.
- 8/2 The storm was currently situated approximately 750 miles east of the Leeward Islands, the National Hurricane Center said.
- 8/3 The run is taking place in approximately 751 communities across Canada.
- 8/4 The storm was currently situated approximately 751 miles east of the Leeward Islands, the National Hurricane Center said.
- 9/1 Russian military and security forces have an inventory of approximately 800 drones, all believed unarmed.
- 9/2 South Indian Lake is approximately 800 kilometres north of Winnipeg.
- 9/3 Russian military and security forces have an inventory of approximately 799 drones, all believed unarmed.
- 9/4 South Indian Lake is approximately 799 kilometres north of Winnipeg.
- 10/1 Last year in Manitoba approximately 900 people were diagnosed with colorectal cancer.
- 10/2 The fire, which is being driven by the wind, is approximately 900 acres in size.
- 10/3 Last year in Manitoba approximately 901 people were diagnosed with colorectal cancer.
- 10/4 The fire, which is being driven by the wind, is approximately 901 acres in size.
- 11/1 The study included an analysis of roughly 950 papers and reports on fracking and water supplies.
- 11/2 The bus plunged near Joinville, roughly 950 kilometres southwest of Rio de Janeiro.
- 11/3 The study included an analysis of roughly 949 papers and reports on fracking and water supplies.
- 11/4 The bus plunged near Joinville, roughly 950 kilometres southwest of Rio de Janeiro.
- 12/1 Roughly 1,000 residents in the area of a resort town remain under evacuation orders.
- 12/2 The plane in which he was flying crashed in the Atlantic Ocean, roughly 1,000 miles off course.
- 12/3 Roughly 1,001 residents in the area of a resort town remain under evacuation orders.

- 12/4 The plane in which he was flying crashed in the Atlantic Ocean, roughly 1,001 miles off course.
- 13/1 Authorities then tracked down about 10 men who'd previously been questioned and asked for DNA samples.
- 13/2 The two car bombs in Baghdad went off about 10 minutes apart late Saturday in the Karrada district.
- 13/3 Authorities then tracked down about 11 men who'd previously been questioned and asked for DNA samples.
- 13/4 The two car bombs in Baghdad went off about 11 minutes apart late Saturday in the Karrada district.
- 14/1 Around 60 firefighters rushed to Adair Tower in North Kensington after a blaze in a two-room apartment.
- 14/2 According to the defense official, Vietnam has reclaimed around 60 acres of land since 2009.
- 14/3 Around 59 firefighters rushed to Adair Tower in North Kensington after a blaze in a two-room apartment.
- 14/4 According to the defense official, Vietnam has reclaimed around 59 acres of land since 2009.
- 15/1 Approximately 70 people have left Norway to become foreign fighters in Syria or Iraq.
- 15/2 Approximately 70 miles southeast of the camping resort, a tornado raked Coal City and damaged several subdivisions.
- 15/3 Approximately 71 people have left Norway to become foreign fighters in Syria or Iraq.
- 15/4 Approximately 71 miles southeast of the camping resort, a tornado raked Coal City and damaged several subdivisions.
- 16/1 Around 100 buildings in the city have been flooded, prompting calls for aid from residents.
- 16/2 Temperatures hovered around 100 degrees, raising concerns of wildfire.
- 16/3 Around 99 buildings in the city have been flooded, prompting calls for aid from residents.
- 16/4 Temperatures hovered around 99 degrees, raising concerns of wildfire.
- 17/1 Officials asked that the roughly 350 people who live in the area known as Morgan's Point remain inside their homes.
- 17/2 The mine is located roughly 350 kilometres from Mali's main northern city of Gao.
- 17/3 Officials asked that the roughly 351 people who live in the area known as Morgan's Point remain inside their homes.
- 17/4 The mine is located roughly 351 kilometres from Mali's main northern city of Gao.
- 18/1 This year's conference attracted roughly 400 activists from countries including China, Thailand, and Nepal.
- 18/2 The area spanning roughly 400 yards had been closed since early Wednesday morning.

- 18/3 This year's conference attracted roughly 399 activists from countries including China, Thailand, and Nepal.
- 18/4 The area spanning roughly 399 yards had been closed since early Wednesday morning.
- 19/1 The mayor said about 700 families have voluntarily left the area and more were applying for relocation help.
- 19/2 The last recorded radar return showed the plane at an altitude of about 700 feet not far away from the crash site.
- 19/3 The mayor said about 701 families have voluntarily left the area and more were applying for relocation help.
- 19/4 The last recorded radar return showed the plane at an altitude of about 701 feet not far away from the crash site.
- 20/1 Roughly 750 people drowned when their trawler sank between Libya and southern Italy.
- 20/2 The redesigned Ford F-150 pickup is roughly 750 pounds lighter than the older truck.
- 20/3 Roughly 749 people drowned when their trawler sank between Libya and southern Italy.
- 20/4 The redesigned Ford F-150 pickup is roughly 749 pounds lighter than the older truck.
- 21/1 The prison complex holds approximately 800 inmates in separate male and female lockups.
- 21/2 The Ryan Aeronautical struck a tree and crashed on a golf course approximately 800 feet from the runway.
- 21/3 The prison complex holds approximately 801 inmates in separate male and female lockups.
- 21/4 The Ryan Aeronautical struck a tree and crashed on a golf course approximately 801 feet from the runway.
- 22/1 The United States alone has flown roughly 900 combat missions over Iraq since August.
- 22/2 The National Hurricane Center said that Storm Ida was centered roughly 900 miles west of the Cape Verde Islands.
- 22/3 The United States alone has flown roughly 899 combat missions over Iraq since August.
- 22/4 The National Hurricane Center said that Storm Ida was centered roughly 899 miles west of the Cape Verde Islands.
- 23/1 An extended Bundeswehr mission in the Mediterranean could in future involve around 950 soldiers.
- 23/2 Hpakant, around 950 kilometers northeast of Myanmar's biggest city, Yangon, is the industry's epicenter.
- 23/3 An extended Bundeswehr mission in the Mediterranean could in future involve around 951 soldiers.
- 23/4 Hpakant, around 951 kilometers northeast of Myanmar's biggest city, Yangon, is the industry's epicenter.
- 24/1 The warning was for around 1,000 customers in the Hamlin area.

- 24/2 A handful of prior oil spills have taken place on the line, the largest around 1,000 gallons.
- 24/3 The warning was for around 999 customers in the Hamlin area.
- 24/4 A handful of prior oil spills have taken place on the line, the largest around 999 gallons.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Bates, D. M., Maechler, M., & Bolker, B. (2012). *lme4*: Linear mixed-effects models using Eigen and Eigen++ (R Package Version 0.999999-0). <https://cran.r-project.org/web/packages/lme4/index.html>.
- Benford, F. (1939). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78, 551–572.
- Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, 133(1), 59–68.
- Ferson, S., O’Rawe, J., Antonenko, A., Siegrist, J., Mickley, J., Luhmann, C. C., et al. (2015). Natural language of uncertainty: Numeric hedge words. *International Journal of Approximate Reasoning*, 57(1), 19–39.
- Gallistel, C. R., & Gelman, R. (2005). Mathematical cognition. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 559–588). Cambridge University Press.
- Geurts, B. (2006). Take five. In S. Vogeeler & L. Tasmowski (Eds.), *Non-definiteness and plurality* (pp. 311–329). Amsterdam: Benjamins.
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora collection: From 100 to 200 languages. In *Proceedings of the 8th International Language Resources and Evaluation (LREC’12)*.
- Kennedy, C. (2013). A scalar semantics for scalar readings of number words. In I. Caponigro & C. Cecchetto (Eds.), *From grammar to meaning: The spontaneous logic of language* (pp. 172–200). Cambridge University Press.
- Krifka, M. (2002). Be brief and vague! And how bidirectional optimality theory allows for verbosity and precision. In D. Restle & D. Zaefferer (Eds.), *Sounds and systems. Studies in structure and change. A festschrift for Theo Vennemann* (pp. 439–458). Mouton de Gruyter, Berlin.
- Krifka, M. (2007). Approximate Interpretations of NumberWords: A Case for Strategic Communication. In G. Bouma, I. Krämer & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 111–126). Royal Netherlands Academy of Arts and Sciences.
- Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In E. Hinrichs & J. Nerbonne (Eds.), *Theory and evidence in semantics* (pp. 109–132). Stanford: CSLI Publications.
- Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2(4), 458–508.
- Lasnik, P. (1999). Pragmatic halos. *Language*, 75(3), 522–551.
- Musolino, J. (2004). The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition*, 93, 1–41.
- Princeton University. (2010). *About WordNet*.
- Rips, L. J., Bloomfield, A., & Asmuth, J. (2007). From numerical concepts to concepts of number. *The Behavioral and Brain Sciences*, 31(01), 623–642.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics*, 49, 91–136.

- Sauerland, U., & Stateva, P. (2011). Two types of vagueness. In P. Egge & N. Klinedinst (Eds.), *Vagueness and language use* (pp. 121–145). Palgrave Studies in Pragmatics, Language and Cognition.
- Solt, S. (2014). An alternative theory of imprecision. In T. Snider, S. D’Antonio & M. Wiegand (Eds.), *Semantics and linguistic theory (SALT)* (Vol. 24, pp. 514–533). Washington DC: LSA and CLC Publications.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.