Dimensions for automatic interpretation of approximate numerical expressions: An empirical study

Sébastien Lefort[1] sebastien.lefort@lip6.fr
Elisabetta Zibetti[2] ezibetti@univ-paris8.fr
Marie-Jeanne Lesot[1] marie-jeanne.lesot@lip6.fr
Marcin Detyniecki[1,3] marcin.detyniecki@lip6.fr
Charles Tijus[2] tijus@univ-paris8.fr

[1]Sorbonne Universités, UPMC Univ Paris 6, CNRS UMR 7606, LIP6, 75005, Paris, France
[2]Laboratoire CHArt-LUTIN, EA 4004, Université Paris 8, 2 rue de la liberté, 93526, Saint-Denis - Cedex 02, France
[3]Polish Academy of Science, IBS PAN, Warsaw, Poland

ABSTRACT
Imprecise numerical expressions, such as "*about 100 meters*", are pervasive in natural language. Mobile robotics, Geographic Information Systems, intelligent personal assistants as well as database querying applications are required to automatically and accurately interpret such expressions, called Approximate Numerical Expressions (ANE). The main challenge is to determine their numerical boundaries that sound plausible to users. The aim of this paper is to provide guidelines to interpret ANEs that are independent from the domain and the formal representations. We identified three arithmetical properties and examined their involvement in ANE interpretation as intervals of denoted values. The implicit assumption of symmetry of the intervals was also tested. To do so, 146 participants were asked to provide the intervals corresponding to 24 ANEs in a semantically neutral context. Results suggest that the properties of ANEs we identified are key factors in their interpretation while symmetry is not always maintained. This study contributes towards an understanding of how users process ANEs and its results can be used to improve intelligent interfaces that lead to better users' satisfaction and natural interaction between him/her and the system.

Author Keywords
Approximate numerical expressions; imprecision; database querying; about

ACM Classification Keywords
H.1.2. [Models and Principles]: User/Machine Systems---Human Factors; Human Information Processing

IUI 2017, March 13 - 16, 2017, Limassol, Cyprus

ACM 978-1-4503-4348-0/17/03□$15.00
DOI: http://dx.doi.org/10.1145/3025171.3025174

INTRODUCTION
Communication between human beings, in daily life, is rarely precise but rather vague [12]. Let's consider the example of a walker asking his/her way to someone. The latter may answer him/her using an Approximate Numerical Expression (ANE): "*Walk for* ***about 100 meters*** *and turn on your right*". Although the information conveyed by the ANE "*about 100 meters*" is imprecise, the walker will intuitively know which street is the correct one. Information systems do not work this way: they process precise information and, in the former example, the range of acceptable values to choose the street needs to be set.

While intelligent systems whose interaction mode relies on natural language become pervasive in daily life, interpreting such imprecise expressions remains a challenging issue. Application domains include robotics, database querying, such as Geographic Information Systems [2, 7] (e.g., looking for an area whose surface is *about 1000m²*), expert systems, such as the medical ones [24] (e.g., representing the information of a patient saying s/he has fever since *approximately one week*), or intelligent personal assistants, such as Apple Siri or Microsoft Cortana. These assistants may already embed ANE interpretation algorithms. However, either their underlying models are intuitively designed by engineers, or they are based on existing models [8, 13, 20, 22] that to the best of our knowledge have never been empirically tested with respect to the users' expectations.

The goal of this paper is thus to provide guidelines to design ANE interpretation models that are both anchored on psycholinguistics knowledge and proved to be consistent with the existing formal representations. By jointly taking into account the way the user represents ANEs as well as their mathematical formalism, our aim is to instill human empirical reasoning into intelligent systems so that their behavior will be smartly adapted to what is expected by the user with regards to imprecise queries. These guidelines are purposely meant to be domain independent in order to be potentially implementable into a wide spectrum of concrete applications requiring ANE interpretations.

Approximate Numerical Expressions (ANEs) are vague linguistic expressions of the general form "*about x*", where *x* is a number. Two elements can be distinguished in ANEs: (1) their semantic and pragmatic context and (2) their reference number. For instance in the expression "*about 100 sheep*", the reference number is 100, the semantic context is what is counted or evaluated, i.e., sheep. The pragmatic context here is unknown (possibly a farmer wanting to buy sheep). These two aspects of an ANE may have different effects on its interpretation [15, 19, 22, 23]. Without ignoring the impact of the context, in this paper we will exclusively focus on the arithmetical dimensions of ANE reference numbers, as they are factors involved in every ANE and are not specific to any context.

So far, three formal representations have been proposed to model the imprecision conveyed by ANEs: intervals [13, 15, 20, 22], fuzzy numbers [14, 16] and probability distributions [8]. The aim of this work is to provide empirically validated guidelines that are general and may be applied in each of these three representations. More specifically, our goal is to highlight the role of three arithmetical dimensions (namely: granularity, magnitude and the last significant digit) of the reference number embedded in ANEs for the determination of the range of values their imprecision denotes. The results of a deep examination and formalization of such dimensions as factors aim at later enable us to characterize not only the interval endpoints, but possibly also the membership functions of fuzzy numbers and the probability distributions corresponding to ANEs of the two other approaches. In this paper, we will use the interval approach because, to our knowledge, it is the only one that has been used in the literature to estimate the imprecision conveyed by ANEs in human communication [13, 15, 20, 22]. This approach therefore allows a precise understanding of the arithmetical dimensions of ANE reference numbers involved in their interpretation.

This paper is structured as follow: to start, we first detail the rationale of using the interval formal approach rather than fuzzy numbers or probability distributions. Two interpretation models from the literature of ANEs as intervals are examined and their strength points and limitations are discussed. In order to complete the rationale of our contribution, the cognitive bases of number processing in human beings is also discussed. Then a formal definition of ANE dimensions is proposed and tested by means of data collected through an empirical study. The aim, the hypotheses, the material and methods used in the study are detailed. Our results about the relevance of the ANE dimensions involved in its interpretation are then presented and discussed with regard to their congruence with the human way of processing numbers and quantities. We conclude by presenting the limits and future implications of our main outcomes.

RELATED WORKS

Among the three formal approaches proposed in the literature to model the imprecision conveyed by ANEs, we focus on the interval approach. Two ANE interpretation models are presented: a linguistic theoretical model [13, 20, 22] and an empirical model [8].

ANEs are a specific kind of vague expressions, related to numerical values, named "*imprecise expressions*" by [23]. In the more general field of vague expressions, [15] introduced the notion of *pragmatic halo*, to formalize the vagueness denoted by such expressions. A *pragmatic halo* is defined as the union of the entity explicitly denoted and the set of entities belonging to the same semantic group, that are implicitly denoted. For instance, in the proposition "*The distance between Berlin and Paris is about 900km*", the expression "*about 900km*" simultaneously denotes the exact distance, i.e., 900km, and a range of possible distances around the exact one, e.g., 850 to 950km.

From the point of view of logic, the truth value of a proposition including a vague expression is true if the actual entity, that may be unknown by the speaker, belongs to the pragmatic halo of the expression. For instance, the proposition "*The distance between Berlin and Paris is* ***about 900km***" is true if the actual distance is 892km and the pragmatic halo of "*about 900km*" is the interval [850, 950]. On the contrary, the proposition is false if the actual value is beyond the pragmatic halo (e.g., 825km).

Pragmatic halos of vague expressions can be modified using different approximators, such as "*about*" or "*roughly*" [14]. Indeed, the pragmatic halo of "*about 900km*" may include a narrower range of distances than the one of "*roughly 900km*". While "*about*" and "*roughly*" may convey symmetric pragmatic halos, i.e., centered around the reference number, other approximators may also affect their symmetry. For instance, the pragmatic halo of "*at least 900km*" may start at 900km and have no right endpoint.

On the basis of these former works, it can be posed that interpreting an ANE "*about x*" consists in estimating the pragmatic halo of the ANE, i.e., the range of values it denotes. The pragmatic halos of ANEs can then be represented as intervals and noted: $I(x) = [I^{-}(x), I^{+}(x)]$.

The main challenge when taking this point of view is then to find a way to identify, formalize and define relevant arithmetical dimensions as good predictors of the endpoints of intervals (i.e., the range of values denoted by "*about x*") used for the ANE interpretation while also considering their plausibility from a human cognitive perspective.

In the next two subsections we present two models from the literature that propose relevant properties of ANEs and provide either theoretical or empirical approaches to estimate the intervals of values corresponding to pragmatic halos of ANEs. In a third subsection we present the cognitive bases of number representation in human beings as it is exposed in the cognitive psychology literature.

Granularity and the scale-based approach

From a linguistic and pragmatic point of view, the main assumption of language production is that speakers tend to produce the simplest expression in order to be understood [10, 18, 25]. Therefore, in the numerical expression framework, speakers tend to prefer round number ANEs rather than exact expressions to convey numerical information because of their simplicity [13]. For instance, if someone buys a piece of furniture 96 euros, s/he may say "*it is worth **about 100** euros*". The approximator "*about*" is used by the speaker to widen the range of values denoted by "*100 euros*". In this case, the speaker prefers to give an approximation ("*about 100 euros*" rather than "*96 euros*") because it is a simple numerical expression that conveys the order of magnitude of the actual value, which is the information s/he wants to convey to the hearer.

The simplicity of numerical expressions and their saliency can be formalized through a scale system [13]. A scale system is defined by a set of scales, expressing different granularity levels of measurement. In the temporal framework, for instance, the scale system can be, from the coarsest to the finest granularity: days, hours, half hours, quarters of hours, minutes, etc. A numerical expression belonging to a coarser granularity level (e.g., "*one hour*") is then considered as simple while a numerical expression belonging to a finer granularity level is considered as less simple (e.g., "*one hour and a half*").

This pragmatic approach, proposed by [13], was further developed and formalized by [20] and [22] in the scale-based model of ANE interpretation. In their model each granularity level is represented as equidistant points on a line. The granularity is the distance between two points at the considered level. Formally, the scale system is defined as $S = \{s_1, \cdots, s_n\}$, where s_i is the ith granularity level, and $s_{i+1} > s_i$. A scale system is said optimal if $s_{i+1} = a_i \cdot s_i$, where $a_i \in N^+$ (e.g., the decimal system $S = \{1, 10, 100, \cdots\}$, where $\forall i, a_i = 10$). The set of granularity levels $Grans(x) = \{g_1, \cdots, g_m\}$ a number x belongs to, can be computed as:

$$Grans(x) = \bigcup_{\{\, s_i \in S \,|x \bmod s_i = 0\}} \{\, s_i \,\}$$

Consequently, a number can belong to several granularity levels. For instance, in the decimal system, the number 100 belongs to three granularity levels: 1, 10 and 100, while 112 belongs only to the 1 granularity level. The interval $I_S^i(x)$ of values denoted by a numerical expression x depends on the considered granularity level *i*. These denoted values are the ones closer to the reference number than to any other number on the considered granularity level $g_i(x)$, formally:

$$I_S^i(x) = \left[x - \frac{g_i(x)}{2}, x + \frac{g_i(x)}{2}\right]$$

The interpretation of a numerical expression can occur at any granularity it belongs to [22]. Let's consider the example of 200. The granularity levels it belongs to in the decimal system are $Grans(200) = \{1, 10, 100\}$. It is reasonable to think that its interpretation may occur either at the 10 granularity level, resulting in $I_S(200)$=[195, 205], or at the 100 level, resulting in $I_S(200)$=[150, 250].

Consequently, if as suggested by [15], approximators are used to modify the pragmatic halo of vague expressions, then in the particular case of numerical expressions, they are used by the speaker to explicitly convey the granularity level at which the interpretation should occur. In the case of ANEs, where the approximator is "*about*", the coarsest granularity level $Gran_C(x)$ is used [22]. Formally, $Gran_C(x)$ is computed as:

$$Gran_C(x) = \sup_{\{s_i \in S | x \bmod s_i = 0\}} s_i$$

The interval of values denoted by an ANE "*about x*" can therefore be formally defined as:

$$I_s(x) = \left[x - \frac{Gran_C(x)}{2}, x + \frac{Gran_C(x)}{2}\right]$$

For instance, $I_S(400)$=[350, 450] and $I_S(430)$=[425, 435]. As a consequence, according to this scale-based approach, the width of the interval corresponding to an ANE is equal to the coarser granularity level the ANE belongs to.

This approach has the advantage of taking into account the granularity of the ANE through a set of scales. However, it does not address the issue of the position of the expression in the granularity level: all ANEs at the same granularity level result in the same interval width. Yet, one may expect that the interval of "*about 100*", for instance, would be narrower than the interval of "*about 800*".

Moreover, the scale-based models are theoretical models that have, to the best of our knowledge, not been empirically validated. On the opposite, other works address the ANE interpretation issue by collecting real data from human beings.

Magnitude, granularity, *fiveness*: an empirical approach

To the best of our knowledge, [8] are the only authors who have proposed to collect empirical data to quantitatively define the imprecision conveyed by approximators. In their study, participants were asked to give the endpoint values of the intervals corresponding to different numerical expressions including approximators, such as: "*Greece enjoys **more than 250 days** of sunshine a year*" or "*Bats make up **about 20%** of all classified mammal species globally*".

The aim of the authors was to test the relevance of several ANE dimensions as predictors of the width of intervals through linear regression analyses. Beyond a granularity related dimension, called here roundness, the authors introduced two other arithmetical dimensions: the order of magnitude and the *fiveness*. The roundness $R(x)$ is defined as the decimal position of the last significant digit (e.g.,

$R(13) = 1$ and $R(130) = 2$) and is related to granularity: $R(x) = \log_{10}(Gran_C(x))$. The order of magnitude $z(x)$ is related to the actual value of the ANE reference number *x*: $z(x) = \log_{10}(x)$. The *fiveness* depends on the last significant digit. More precisely, it equals 1 if the last significant digit is 5 (e.g., 150), it equals 0 otherwise (e.g., 140). After defining these three arithmetical dimensions the authors found that their combination is a good predictor of the width of intervals corresponding to ANEs.

Beyond the identification of these three dimensions as factors involved in ANE interpretation, the authors have shown that they are involved in a logarithmic scale: the roundness and the order of magnitude are logarithms of granularity and value of the reference number. This result is congruent with the way the human cognitive system processes numbers and quantities. However, the relevance of using a logarithmic scale, compared to the linear one, remains to be tested.

Moreover, from a methodological point of view, although [8] controlled the type of unit related to ANEs (discrete, length, time, etc.), the questionnaire mixed several semantic contexts, which may result in different intervals for the same reference number [19]. Consequently, a study whose aim is to identify arithmetical dimensions involved in ANE interpretation should use uncontextualized ANEs, or at least a controlled context, as material for the participants.

From the examination of these two existing models it appears that the granularity (related to roundness in [8]), the magnitude and the value of the last significant digit of the ANEs are key factors in their interpretation. Nevertheless, the arithmetical dimensions these models are based on are not formalized in the same way, and their relevance with regards to the numerical cognition in human beings remains to be demonstrated.

Finally, it is noteworthy that both models implicitly consider the intervals corresponding to ANEs as symmetric, i.e., centered around the ANE reference number (e.g., *I*(100) = [95, 105]). To the best of our knowledge, this assumption has never been empirically validated and can therefore be questioned.

Cognitive representation of numbers and quantities

As ANEs involve numbers, it seems relevant to anchor the discussion of the interpretation factors on a cognitive ground and more specifically on the way the human cognitive system encodes and represents numbers and quantities.

Two different subsystems are involved in number cognition [4, 5]. The first one, called the Approximate Number System, relies on approximate, non-symbolic representations. The numbers are ordinaly represented on a logarithmically compressed mental line, where quantities are encoded according to the Weber-Fechner law [4]: two quantities x_1 and x_2 can be distinguished if their absolute difference is greater than a fraction c of the largest of both, formally:

$$\frac{|x_1 - x_2|}{\max(x_1, x_2)} > c$$

This indistinguishability between close quantities results in imprecision when estimating quantities. For instance, in estimating, without counting, the number of dots in an array, one cannot distinguish between the quantities 98 and 101. On the other hand, the Approximate Number System would be abstract, independent from the modality [1, 3]. Its role is to encode the magnitude, regardless of the considered dimension or the evaluated physical characteristic: number, space, time, brightness, pitch, etc.

The second cognitive subsystem is exact and symbolic [4, 9]. Based on language, and unlike the Approximate Number System, its representations do not suffer from imprecision. It is involved in the knowledge and the processing of arithmetical facts. In this system, some numbers are considered as more salient than others. Indeed, it has been shown than some numbers are more frequently expressed than others [6, 11]. For instance, analyses of corpuses of newspapers reveal that round numbers occur more frequently than non-round numbers. Similarly, numbers whose last significant digit is 5 or, to a lower extent, 2 are more frequent.

As a whole, these findings from the studies of numerical cognition lead us to consider a priori three arithmetical dimensions that should be taken into consideration to properly account for the way people interpret ANEs.

Firstly, as symbolic numerical expressions, ANEs involve the exact number system to interpret them, leading to consider the ANE interpretation issue as a formal problem as in the case of the scale-based approach [13, 20, 22]. Following this rationale, *granularity* and the *value of the last significant digit* emerge as two relevant dimensions to be considered. Secondly, because ANEs are real world estimations of quantities, they should rely on the Approximate Number System to represent them. From this perspective, interpreting an ANE consists in estimating the imprecision implied by its representation, which only depends on the magnitude of the reference number.

Hence, based on the ANE interpretation approaches described previously, and on the insights provided by the findings from the numerical cognition studies, our contribution pursues the following two objectives:

(1) to formally define the arithmetical dimensions on which the human processing of ANEs interpretation relies, that is to obtain the range of values denoted by "*about x*";

(2) to validate by the means of an empirical study (a) the involvement of these dimensions in the human interpretation of ANEs; (b) the relevance of simultaneously considering these dimensions in a logarithmic scale; (c) to

test the implicit assumption of symmetry of the intervals corresponding to ANEs.

Dimension	Formal definition	Example
Magnitude	x	7650
Granularity	$Gran(x) = 10^{i^*}$ where $i^* = \min\{\, i \mid a_i \neq 0\}$	10
Last significant digit	$LSD(x) = a_{i^*}$	5

Table 1. Dimensions of a positive integer $x = \sum_{i=0}^{q} a_i \cdot 10^i$, illustrated in the case of $x = 7650$ in the last column.

FORMAL DEFINITIONS OF ANE DIMENSIONS

The ANEs considered in this study are of the form "*about* x", where $x \in N^+$. The corresponding intervals, i.e., the range of values denoted by "*about* x", are noted $I(x) = [x - \Delta^-(x),\ x - \Delta^+(x)]$. In the decimal system, x can be written as $x = \sum_{i=0}^{q} a_i \cdot 10^i$, with $a_i \in [\![0,9]\!]$.

Now that we have laid the basic formal definitions of the ANE reference number and the corresponding interval in the decimal system, we can go further with our proposal of a formal definition of three arithmetical dimensions that according to us characterize an ANE (Table 1).

The aim of this section is to systematize the different dimensions proposed in the literature to reach a coherent and meaningful set of three dimensions likely to account for ANE interpretation.

Magnitude is defined as the actual value of x. It is meant to take into account the way quantities are encoded in the cognitive Approximate Number System. It is logarithmically related to the order of magnitude proposed by [8]: $x = 10^{z(x)}$.

Granularity $Gran(x)$ is defined as the order of magnitude at which the last significant digit of x occurs in the decimal system. This definition of granularity is related to the one proposed by [21] in the scale-based approach, $Gran_C(x)$. Indeed, $Gran(x) = Gran_C(x)$ when the selected scale-system is the decimal one. The granularity is also logarithmically related to the roundness proposed by [8]: $Gran(x) = 10^{R(x)}$.

The **value of last significant digit** $LSD(x)$ is more general than the *fiveness* proposed by [8] since the authors consider 5 as a special case of last significant digit, compared to the others, while we propose to consider all values of the last significant digit as distinct cases.

THE STUDY

According to our rationale and aims, three main hypotheses are formulated:

H1 – The magnitude, the granularity and the value of the last significant digit are the key factors in ANE interpretation. We should therefore observe the separate involvement of these three dimensions according to the three following hypothesized assessments:

H1.1 – Magnitude: [8] have shown that the magnitude of the ANE is a good predictor of its corresponding interval width. Thus, we predict that at constant granularity and last significant digit, different magnitudes result in different distances $\Delta(x)$ of the interval endpoints to the ANE reference number. More specifically, the higher the magnitude, the higher the distance, formally: if $x > y$, $Gran(x) = Gran(y)$ and $LSD(x) = LSD(y)$, then $\Delta(x) > \Delta(y)$. For instance, 8150 > 50; $Gran(8150) = Gran(50) = 10$; $LSD(8150) = LSD(50) = 5$, therefore $\Delta(8150) > \Delta(50)$.

H1.2 – Granularity: both models from the literature [8, 13, 20, 22] suggest that granularity is the most influential factor in ANE interpretation. More specifically, according to the scale-based model [13, 20, 22], the coarser the granularity, the wider the corresponding interval. We therefore predict to observe an effect of granularity on the intervals corresponding to ANEs: at constant value of the last significant digit, different granularities of ANE result in different distances $\Delta(x)$ of interval endpoints to the ANE reference number. More specifically, the higher the granularity, the higher the distance, formally: if $Gran(x) > Gran(y)$ and $LSD(x) = LSD(y)$ then $\Delta(x) > \Delta(y)$. For instance, $Gran(5000) = 1000$; $Gran(500) = 100$; $LSD(5000) = LSD(500) = 5$, therefore $\Delta(5000) > \Delta(500)$.

H1.3 – Last significant digit: despite the fact that this dimension is not taken into account in scale-based models, the way the human cognitive system represents numbers suggests that, at the same level of granularity, a different last significant digit should lead to different distances $\Delta(x)$ of interval endpoints to the ANE reference number. More specifically, the higher the last significant digit, the higher the distance, formally: if $LSD(x) > LSD(y)$ and $Gran(x) = Gran(y)$, then $\Delta(x) > \Delta(y)$. For instance, $LSD(80) = 8$; $LSD(50) = 5$; $Gran(80) = Gran(50) = 10$, thus $\Delta(80) > \Delta(50)$.

H2 – Relevance of a three-dimensional logarithmic scale account of ANE interpretation.

H2.1 – Relevance of a three-dimensional model: a model that takes into account the magnitude, the granularity and the last significant digit (MGLSD) should better account for the observed intervals than the two models based on the dimensions proposed in the literature.

H2.2 – Relevance of the logarithmic scale: as suggested by the work of [8], the relationship between the values of the intervals endpoints and the ANEs dimensions occurs in a logarithmic scale. Therefore, models in logarithmic scale should better account for the observed intervals than the ones in linear scale.

H3 – Interval Symmetry: according to the scale-based models [13, 20, 22], intervals are centered around the ANEs values. This leads us to hypothesize the symmetry of intervals, that is the left endpoint of an interval should be at the same distance from the ANE reference number than the right endpoint, formally: $\Delta^{-}(x) = \Delta^{+}(x)$. For instance, $I(100) = [100 - 10, 100 + 10] = [90,110]$.

Methods

Population

One hundred and forty six adults volunteered to take part in this study: 102 women and 44 men aged 20 to 70 (M = 38.6; σ = 14.2). All were recruited through an announcement, diffused on a mailing-list, and all were native French speakers.

Material and procedure

An online questionnaire was designed to collect intervals corresponding to the 24 ANEs listed in Table 2. Reference numbers of ANEs were selected to cover different combinations of dimensions to test the hypotheses: several magnitudes at constant granularity and last significant digit[1]; several granularities at constant last significant digit[2]; several last significant digits at constant granularity[3].

ANEs never were semantically contextualized, no cues were given to participants as to what was measured or counted. It may be that participants supplied their own context to the provided ANEs. However, it is reasonable to think that the context they supplied was the same across the experiment, such that one can detect the differences from one ANE to another. When the differences observed between ANEs are similar across participants, they should suggest a factor specific to the ANE reference number, independent from the context, such as the proposed arithmetical dimensions of ANEs.

The instructions, given in French to the participants, can be translated as "*In your opinion, what are the MINIMUM and MAXIMUM values associated with* ***about x****?*". The order of the ANEs was randomly set for each participant. Participants gave the left and right endpoints of an interval as one answer, by typing the values in the devoted space.

Once the participants finished providing the ANE endpoints, they were asked two questions meant to control inter-individual variability with respect to: (i) the use of mathematics and (ii) the subjective level of mental arithmetic. These two questions are meant to check whether a daily use of mathematics may affect ANE interpretation.

[1] 40/440; 100/1100; 500/1500; 30/4730; 50/150/8150

[2] 20/200/2000; 40/400; 50/500; 600/6000; 80/800/8000

[3] Tens: 20/30/40/50/80. Hundreds: 100/200/400/500/600/800. Thousands: 1000/2000/6000/8000

x	$Gran(x)$	$LSD(x)$	*Average endpoints*
20	10	2	[16.2, 24.5]
30	10	3	[24.5, 35.5]
40	10	4	[34.0, 44.8]
50	10	5	[40.9, 59.1]
80	10	8	[71.0, 87.6]
100	100	1	[86.7, 113.4]
110	10	1	[102.3, 123.0]
150	10	5	[131.5, 167.3]
200	100	2	[171.8, 232.9]
400	100	4	[357.8, 439.1]
440	10	4	[419.4, 457.4]
500	100	5	[458.6, 541.1]
560	10	6	[540.6, 577.2]
600	100	6	[552.7, 649.0]
800	100	8	[733.1, 858.5]
1000	1000	1	[878.0, 1131.4]
1100	100	1	[1038.4, 1168.0]
1500	100	5	[1351.9, 1643.7]
2000	1000	2	[1774.3, 2266.6]
4700	100	7	[4543.7, 4826.9]
4730	10	3	[4632.1, 4794.2]
6000	1000	6	[5574.8, 6589.8]
8000	1000	8	[7423.8, 8514.8]
8150	10	5	[8021.4, 8301.7]

Table 2. Reference number of ANEs used in the questionnaire and their dimensions: magnitude (x), granularity (Gran) and last significant digit (LSD). The last column presents the intervals formed by the average values of the left and right endpoints given by the participants.

Data cleaning

The answer, corresponding to the interval of ANE "*about x*", of participant p is noted $I_p(x) = [I_p^{-}(x), I_p^{+}(x)]$. These answers were transformed so as to get two absolute distances $\Delta_p^e(x)$ between the reference number x and the endpoint $e \in \{-,+\}$ of the interval: $\Delta_p^e(x) = |I_p^e(x) - x|$. Consequently, $I_p(x) = [\, x - \Delta_p^{-}(x), x + \Delta_p^{+}(x)]$ and $|I_p(x)| = \Delta_p^{-}(x) + \Delta_p^{+}(x)$. The analyses reported below are based on the absolute distances as the only dependent variable. This variable is more relevant than the interval

width, since it allows to compare them together, without losing the symmetry information. Indeed, two widths may be equal although the values of the endpoints are different (e.g., |[95, 110]| = |[90, 105]|).

In order to exclude outlier pairs from the set, data were processed according to the following three-step procedure:

Step 1: We considered an answer as an outlier if: (i) the right endpoint was below the reference number or the left endpoint was above it, formally: $I_p^+(x) < x$ or $I_p^-(x) > x$ (e.g., *I*(800) = [700, 750] or *I*(800) = [810, 850]); or (ii) one endpoint value was greater than ten times or less than one tenth the reference number of the ANE, formally $I_p^+(x) > 10 \cdot x$ or $I_p^-(x) < \frac{x}{10}$ (e.g., *I*(100) = [9, 1101]).

Step 2: Mean and standard deviation were computed for the remaining values of each endpoint of each ANE. Any endpoint beyond three standard deviations of the mean was considered as outliers.

Step 3: Participants with at least 70% missing values or outliers were considered as untrustworthy, and all their answers were excluded.

From 3504 intervals in the original corpus, 3177 (91%) were included in the analyses. 10 participants were excluded from the study.

Statistical analyses

Four types of analyses were performed, depending on the considered hypothesis.

H1 – Magnitude, granularity and last significant digit

To test the hypothesis of an effect of magnitude, granularity and the last significant digit, we compared the distribution of distances $\Delta(x)$ from the endpoint values to the reference number between across ANEs (e.g., "*about 20*" vs. "*about 30*"). Since all these distributions are not normal, comparisons between two ANEs were performed using the Wilcoxon signed rank test, designed for paired samples. When comparisons involved three or more ANEs, a Friedman test was used. In case of significant effect, post-hoc analyses, not reported here for space reasons, were done using Nemenyi test for pairwise multiple comparisons.

H2 – Three-dimensional model in logarithmic scale

To compare the two models from the literature (the scale-based approach [13, 20, 22] and the empirical approach [8]) to the three-dimensional account of ANE interpretation we propose (MGLSD), we used Bayesian analyses. As in [8], the models linearly combine dimensions of ANEs.

The scale-based approach is solely based on granularity $Gran(x)$, the model therefore takes the form:

$$\Delta(x) = \alpha_1 \cdot Gran(x) + \alpha_2$$

The empirical model [8] is based on a combination of magnitude, granularity and *fiveness* (x) , formally:

$$\Delta(x) = \beta_1 \cdot x + \beta_2 \cdot Gran(x) + \beta_3 \cdot f(x) + \beta_4$$

Finally, the three-dimensional approach we propose (MGLSD) considers the magnitude, the granularity and the last significant digit of the ANEs, formally:

$$\Delta(x) = \gamma_1 \cdot x + \gamma_2 \cdot Gran(x) + \gamma_3 \cdot LSD(x) + \gamma_4$$

For each model, the Bayes factor resulting of the comparison to the absence of relationship is reported.

To test hypothesis H2.2, predicting a better account of the observed intervals in a logarithmic scale, a second iteration of this process, using the logarithms of the variables, was performed. In this case the models respectively are:

- Scale-based approach:
$$\log_{10}(\Delta(x)) = \alpha_1 \cdot \log_{10}(Gran(x)) + \alpha_2$$

- Empirical approach:
$$\log_{10}(\Delta(x)) = \beta_1 \cdot \log_{10}(x) + \beta_2 \cdot \log_{10}(Gran(x)) + \beta_3 \cdot f(x) + \beta_4$$

- Our proposed MGLSD :
$$\log_{10}(\Delta(x)) = \gamma_1 \cdot \log_{10}(x) + \gamma_2 \cdot \log_{10}(Gran(x)) + \gamma_3 \cdot \log_{10}(LSD(x)) + \gamma_4$$

H3 – Interval symmetry

To test the symmetry hypothesis, we propose not to use standard Wilcoxon signed rank test. Indeed, this test appears significant when the mean difference of ranks between the distributions of endpoints is different from zero. Therefore, it is not adapted to detect mere differences between endpoints if there is no clear trend in these differences. To overcome this issue, we propose to evaluate the equality between two series of endpoints by counting the number of endpoints which are equal, with an allowed relative error of 10%, formalized as:

$$Eq(x_1, x_2, e) = \frac{1}{P(x_1, x_2)} \cdot \left| \left\{ p \in P(x_1, x_2) \mid \frac{|\Delta_p^e(x_1) - \Delta_p^e(x_2)|}{\min\left(\Delta_p^e(x_1), \Delta_p^e(x_2)\right)} < 0.1 \right\} \right|$$

where x_1 and x_2 are the reference numbers of the considered ANEs, i.e., the independent variable, *e* the endpoint to be compared, and $P(x_1, x_2)$ the set of participants whose endpoints are not outliers for x_1 nor for x_2.

Control: Arithmetical skills

The last items of the questionnaire are meant to assess whether participants regularly use mathematics and have good mental arithmetic skills, to ensure that these inter-individual variables do not affect the ANE interpretation. The effect of a daily use of mathematics was tested using Wilcoxon rank-sum tests. We compared the values given by the participants as endpoints of each ANE separately to investigate the difference between intervals given by mathematical users and non-mathematical users.

Similarly, we tested the effect of the subjective level in mental arithmetic on intervals using Wilcoxon rank-sum tests. To do so, in a first step, two groups were created, according to the reported level in mental arithmetic: participants who reported a score from 1 to 3 included were assigned to the low-level group; the high-level group consists in participants with a score of 4 or 5. The significance threshold was set for all analyses at p=.01.

RESULTS

In order to illustrate the obtained results, the right column of Table 2 presents the average left and right endpoints given by the participants. They are provided to give an indication of the interval widths. Please note that they are note representative of a typical answer.

Control of the mental arithmetic skills

From the 136 participants included in the analyses, 66 reported using regularly mathematics while 70 reported not using it. No significant difference was found for any endpoint of any ANE according to the two groups (W ranging from 1867.5 to 2414.5; p=N.S.).

78 participants were included in the low-level group of mental arithmetic and 58 in the high-level group. The analyses revealed no significant difference between the two mental arithmetic groups on any endpoint of any ANE (W ranging from 1767.5 to 2614; p=N.S.).

A daily use of mathematics or a subjective high level of mental arithmetic therefore seems to have no effect on the interpretation of semantically uncontextualized ANEs.

Magnitude, granularity and last significant digit (H1)

Effect of magnitude

To test the hypothesis of an effect of magnitude, we compared four pairs and a trio of ANEs with different magnitudes and constant granularity and last significant digit. Statistical analyses showed significant differences in the four comparisons of couples and in the trio (Table 3). The post-hoc analyses performed on the trio revealed significant differences in all combinations.

These results support the effect of magnitude: at constant granularity and last significant digit, ANEs with different magnitudes lead to different interval widths. More specifically, a larger magnitude results in a larger interval.

Effect of granularity

In order to test the effect of granularity, ANEs whose granularities are different, and with the same last significant digit, were compared (Table 4): four couples and two trios. Results show significant differences in all comparisons. Moreover, post-hoc analyses performed for the two trios revealed that, in both cases, all levels of granularity differ from each another. Therefore, these results tend to support the effect of granularity on ANEs intervals. More specifically, they show that higher granularity levels lead to wider intervals of denoted values.

Comparisons	Test (V: Wilcoxon; χ^2: Friedman)
40/440	$V = 115.5; p < .01$
100/1100	$V = 87; p < .01$
500/1500	$V = 237.5; p < .01$
30/4730	$V = 325; p < .01$
50/150/8150	$\chi^2 = 297.8; p < .01$

Table 3. Effect of magnitude: results of comparisons between ANEs of same granularity and last significant digit but different magnitudes.

Comparisons	Test (V: Wilcoxon; χ^2: Friedman)
40/400	$V = 92; p < .01$
50/500	$V = 252.5; p < .01$
100/1000	$V = 265.5; p < .01$
600/6000	$V = 62; p < .01$
20/200/2000	$\chi^2 = 388.5; p < .01$
80/800/8000	$\chi^2 = 407.7; p < .01$

Table 4. Effect of granularity: results of comparisons between ANEs with same last significant digit but different granularities.

Granularity level	Friedman test
20/30/40/50/80	$\chi^2 = 275.6; p < .01$
100/200/400/500/600/800	$\chi^2 = 542.5; p < .01$
1000/2000/6000/8000	$\chi^2 = 373.2; p < .01$

Table 5. Effect of the last significant digit: results of comparisons between ANEs of same granularity but different last significant digit values.

Effect of last significant digit

ANEs whose last significant digit differ and granularities are constant were compared to test the hypothesis of an effect of the last significant digit value on interval endpoints: tens, hundreds and thousands (Table 5).

The Friedman tests show a significant effect of the last significant digit at each granularity level (Table 5). However, post-hoc analyses revealed threshold effects, especially in hundreds and thousands. Indeed, one can notice that the interval widths corresponding to x = 500, x = 600 and x = 800 do not significantly differ from one another, nor the interval widths corresponding to 50 and 80. Similarly, the distances of the endpoints of 6000 and 8000 from the reference numbers are not significantly different.

These results seem ambiguous. On one hand, the Friedman tests revealed an effect of the last significant digit. On the other hand, while post-hoc analyses revealed significant

differences in most of the comparison, they also suggest a threshold effect: distances between endpoints and ANE reference numbers do not increase between 500 and 800 or between 6000 and 8000.

Model	Bayes Factor	
	Linear scale	*Logarithmic scale*
Scale-based	$1.63 \cdot 10^{238}$	$9.60 \cdot 10^{283}$
Regression	$1.16 \cdot 10^{319}$	$6.20 \cdot 10^{445}$
MGLSD	$\mathbf{1.40 \cdot 10^{329}}$	$\mathbf{4.52 \cdot 10^{446}}$

Table 6: Bayes factors obtained by Bayesian analyses of models, in linear and logarithmic scale.

Three-dimensional model in logarithmic scale (H2)

Table 6 presents the Bayes factors obtained when comparing the three models (scale-based, regression and our three-dimensional proposition, MGLSD), either in linear or logarithmic scale.

The results show that all three models better account for the collected intervals when they are in a logarithmic scale.

As revealed by the models comparison, the three-dimensional model MGLSD we propose better accounts for the collected intervals. This observation is valid both in linear and logarithmic scales, supporting the ANEs dimensions we propose. However, in logarithmic scale, MGLSD slightly better fits the data than the regression model (Bayes factor 7.28 times the one of the regression model). The scale-based model, based on granularity only, shows the weakest Bayes factor.

Taken together, these results support our second hypothesis. Magnitude, granularity and the last significant digit of ANEs are involved in their interpretation. A model that takes into account these three dimensions better fits the collected intervals corresponding to ANEs. Moreover, as revealed by the analyses, interpretation models should consider them in a logarithmic rather than in a linear scale.

Interval symmetry (H3)

The scale-based model implies that intervals are centered around ANE reference numbers. We tested using the equality criteria defined above whether this assumption holds on the collected data. On the whole dataset, 78.7% of the intervals are symmetric. This score ranges from 50.7% (4730) to 89.6% (1500). Low scores (< 70%) occur for ANEs with multiple significant digits and whose last is neither 1 nor 5: 440 (61.5%), 560 (67.7%), 4700 (65.7%) and 4730 (50.7%). A low score is also observed for 8150 (64.2%). Although its last significant digit is 5, the difference of order of magnitude between its granularity (10) and its magnitude may account for this low score.

These results suggest that the assumption of symmetry of the intervals does not hold for all ANEs. Indeed, multiple significant digits ANEs whose last significant digit is neither 1 nor 5 lead to less symmetric intervals.

DISCUSSION

Prior studies suggest that granularity and magnitude of ANEs are the two key factors of their interpretation [8, 13, 20, 22]. More specifically, the theoretical scale-based model of ANE interpretation [13, 20, 22] defines the granularity as the only dimension of ANE to be taken into account. Complementary, the empirical work of [8] suggests that the interval corresponding to an ANE is also affected by its magnitude and by its last significant digit, especially if it is 5. For instance, while the granularity of "*about 50*", "*about 150*" and "*about 8150*" is equal, the intervals corresponding to these ANEs should be different. The way the human cognitive system encodes quantities led us to posit that considering the last significant digit of an ANE may also be a factor influencing the intervals.

In order to determine the dimensions of the ANEs that affect their corresponding intervals, we first proceeded to a systematisation and formal description of the dimensions likely to account of the processing of ANEs and then we carried an empirical study to collect intervals corresponding to uncontextualized ANEs in order to test the relevance of these dimensions.

We performed analyses to separately examine the effect of the three ANEs arithmetical dimensions we propose. We found that the interval corresponding to an ANE is influenced by the magnitude, the granularity and the last significant digit of its reference number. While the granularity is the common factor proposed in the literature [8, 13, 20, 22], the magnitude has only been highlighted by the empirical study of [8]. We have shown that at constant granularity and last significant digit, the interval of denoted values is wider when the magnitude of the ANE is higher (e.g., $|I(8150)| > |I(150)| > |I(50)|$). The effect of the magnitude on the width of the intervals may be due to the representation of the quantities in the cognitive Approximate Number System. Indeed, to be distinguished by this system, the difference between two quantities must increase linearly with the higher quantity [4]. Applied to ANEs, this principle implies that the range of the indistinguishable values, i.e., the imprecision, should increase with the ANE reference number [13]. Our results are consistent with this interpretation.

Based on the same principle, we predicted that the last significant digit influences the width of the intervals. More specifically, at the same granularity level, the interval corresponding to ANEs should be larger when the last significant digit is higher (e.g., $|I(80)| > |I(50)| > |I(20)|$). Comparing several ANEs at the same granularity levels (i.e., tens, hundreds and thousands), we found significant differences between ANEs whose last significant digits are different. However, the results suggest that the relationship between this dimension and the width of the intervals is not linear. Indeed, we observed that the width does not necessarily increase at each incrementation of the last significant digit (e.g., between 30 and 40, or between 400

and 500). These results also do not support the conclusion of [8] regarding the relevance of the *fiveness* property as predictor of the interval width. According to the authors, ANEs whose last significant digit is 5 result in wider intervals than ANEs whose last significant digit is different. The analyses reveal no significant difference in the size of the intervals between 50 and 80, nor between 500 and 800. Moreover, results show that differences appear between different last significant digits at the same granularity level (e.g., $|I(600)| > |I(200)|$). Thus, it seems that the case of 5 as last significant digit is not more specific than any other digit.

The Bayesian analyses we performed to test the fitting of the models with regards to the collected data show that, compared to the models from the literature, the one that we propose (MGLSD), that simultaneously takes into account the magnitude, the granularity and the last significant digit of ANEs, better accounts for the intervals. Moreover, all three models better fit the data that we collected when they are in a logarithmic scale. The involvement of such scale in ANE interpretation may be related to the way the human cognitive system represents numbers. Indeed, in the Approximate Number System, the quantities are encoded on a logarithmically compressed mental line [4].

The last hypothesis concerns the symmetry of the approximator "*about*". Indeed, models from the literature [8, 13, 20, 22] posit that the imprecision conveyed by an ANE "*about x*" is equally distributed on the left and the right of its reference number, formally: $\Delta^{+}(x) = \Delta^{-}(x)$. This equality was tested on the collected data. Although the mean equality score of the whole dataset is high, five ANEs lead to lower scores: 440, 560, 4700, 4730 and 8150. These ANEs can be characterized by two properties: (i) their reference numbers have at least two significant digits; (ii) except for 8150, their last significant digit is neither 1 nor 5. From these two observations, we propose an explanation based on saliency: the saliency of a number can be defined as its ability to be activated in the human cognitive system, as revealed by the analyses of corpuses [6, 11]. These analyses show that round numbers are more salient than non-round numbers. Similarly, numbers whose last significant digit is 1, 2 or 5 are more salient. From this point of view, the numbers whose symmetry scores are low are close to salient numbers. For instance, 4730 is close to 4700 and 4750. As salient numbers tend to be more easily expressed, it may be that participants tend to be biased toward them. Moreover, since these salient endpoints are not necessarily at the same distance from the ANE reference number, the intervals are not symmetric. However, the experimental setup of this study does not allow us to determine if this bias is representational, on the number mental line, or at the language production level, because of the simplicity of their verbal expression [13].

Finally, the fact that participants tend to take into account the magnitude, the granularity and the value of the last significant digit may be interpreted as a compromise between the two numerical cognitive systems: the Approximate Number Systems deals with magnitudes while the exact, symbolic system may process the formal expression of the ANE reference number, and thus considers its granularity and last significant digit as the relevant properties in its interpretation.

CONCLUSION

The aim of this work was to provide guidelines to design interpretation models of Approximate Numerical Expressions to better fit user's expectations when s/he expresses imprecise queries in natural language. More specifically, the goal was to determine the relevant arithmetical dimensions of the numerical part of uncontextualized ANEs for their interpretation. We have shown, by the means of an empirical study, that the magnitude, the granularity and the last significant digit of ANEs are the key factors when they are considered in a logarithmic scale. Interpretation models should therefore involve these three properties in a logarithmic scale.

Two limitations of this study are that: (1) in daily life, people may not consciously set endpoints to the denoted range of values but may rather interpret *implicitly* ANEs, in terms of acceptable and unacceptable values; (2) ANEs are rarely uncontextualized in daily life. Future work should therefore address these issues by collecting data in an implicit way, for instance by asking participants if randomly generated values are denoted by an ANE and by studying the effects of different semantic and pragmatic contexts on ANE interpretation.

However, through this study we systemized and empirically validated the arithmetical dimensions of ANEs involved in their interpretation by human beings and we highlighted how to model them quantitatively by the means of intervals [17], fuzzy numbers [16] or as probability distributions.

The concrete outcome of our findings is their implementation, for instance, in database querying applications and expert systems aimed at automatically interpreting imprecise expressions and progressively adjusting to the user satisfaction. In the framework of the ReqFlex database flexible query system [21], our current work aims at designing an intelligent interface that models ANEs (i) without involving expert knowledge and (ii) to provide answers that fit the user's expectation such that s/he does not need to refine his/her query many times.

ACKNOWLEDGMENTS

This work was performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02.

REFERENCES

1. Barth, H., Kanwisher, N. and Spelke, E. The construction of large number representations in adults. *Cognition 86*, 3 (2003), 201-221.

2. Bennett, B. Spatial vagueness. In Jeansoulin, R., Papini, O., Prade, H. and Schockaert, S. (Eds) *Methods for Handling Imperfect Spatial Information. Vol. 256 of Studies in Fuzziness and Soft Computing*, Springer (2010), 15-47.
3. Bonn, C.D. and Cantlon, J.F. The origins and structure of quantitative concepts. *Cognitive neuropsychology 29*, 1-2 (2012), 149-173.
4. Dehaene, S. The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences 7*, 4 (2003), 145-147.
5. Dehaene, S. Origins of mathematical intuitions. *Annals of the New York Academy of Sciences 1156*, 1 (2009), 232-259.
6. Dehaene, S. and Mehler, J. Cross-linguistic regularities in the frequency of number words. *Cognition 43*, 1 (1992), 1-29.
7. Delboni, T.M., Borges, K.A., Laender, A.H. and Davis, C.A. Semantic expansion of geographic web queries based on natural language positioning expressions. *Transactions in GIS 11*, 3 (2007), 377-397.
8. Ferson, S., O'Rawe, J., Antonenko, A., Siegrist, J., Mickley, J., Luhmann, C.C., Sentz, K. and Finkel, A.M. Natural language of uncertainty: numeric hedge words. *International Journal of Approximate Reasoning 57* (2015), 19-39.
9. Gelman, R. and Butterworth, B. Number and language: how are they related? *Trends in Cognitive Sciences 9*, 1 (2005), 6-10.
10. Grice, H.P. Logic and conversation. In Cole, P. and Morgan, J.P. (Eds) *Syntax and semantics. Vol 3: Speech acts*. Seminar Press (1975), 41-58.
11. Jansen, C. and Pollmann, M. On round numbers: Pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics 8*, 3 (2001), 187-201.
12. Jucker, A.H., Smith, S.W. and Lüdge, T. Interactive aspects of vagueness in conversation. *Journal of pragmatics 35*, 12 (2003), 1737-1769.
13. Krifka, M. Approximate interpretations of number words: A case for strategic communication. In: Bouma, G., Krämer, I. and Joost, Z. (Eds) *Cognitive foundations of interpretation*. Koninklijke Nederlandse Akademie van Wetenschapen (2007), 111-126.
14. Lakoff, G. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of philosophical logic 2*, 4 (1973), 458-508.
15. Lasersohn, P. Pragmatic halos. *Language 75*, 3 (1999), 522-551.
16. Lefort, S., Lesot M.-J., Zibetti, E., Tijus, C. and Detyniecki, M. How much is "about"? Fuzzy interpretation of approximate numerical expressions. In *Proc. of IPMU'16 610*, Springer (2016), 226-237.
17. Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C. and Detyniecki, M. Interpretation of approximate numerical expressions: Computational model and empirical study. *International Journal of Approximate Reasoning* (under press).
18. Levinson, S.C. *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press (2000).
19. Sadock, J.M. Truth and approximations. In *Proc. of the 3rd Annual Meeting of the Berkeley Linguistics Society* (1977), 430-439.
20. Sauerland, U. and Stateva, P. Scalar vs. epistemic vagueness: Evidence from approximators. In *Proc. of SALT* (2007), 228-245.
21. Smits, G., Pivert, O. and Girault, T. ReqFlex: fuzzy queries for everyone. In *Proc of the VLDB Endowment 6*, 12 (2013), 1206-1209.
22. Solt, S. An alternative theory of imprecision. In *Semantics and Linguistic Theory. Vol. 24* (2014), 514-533.
23. Solt, S. Vagueness and imprecision: Empirical foundations. *Annual Review of Linguistics 1*, 1 (2015), 107-127.
24. Straszecka, E. Combining uncertainty and imprecision in models of medical diagnosis. *Information Sciences 176*, 20 (2006), 3026-3059.
25. Zipf, G.K. *Human behaviour and the principle of least-effort*. Addison-Wesley (1949).