Learnable Classes of Natural Language Quantifiers: Two Perspectives

Denis Paperno

December 13, 2011

1 Introduction

This paper discusses one aspect of what Chomsky dubbed as "the logical problem of language acquisition", applying theoretical results from the computational learning theory. The specific issue addressed here is learning quantifier meanings, more precisely, determiner meanings. I present a new characterization of the natural class of simple quantifier meanings of natural language, and show that this class is learnable (identifiable in the limit). I then demonstrate that several previously identified Gold learnable classes of quantifier meanings are not PAC learnable [19], i.e. can not be identified with a given degree of precision after a reasonable (polynomial) amount of input data and computing time.

The paper discusses two computational approaches to learning. Identification in the limit, or Gold learning, provides the simplest model of learning success [9]. Under this model, a learner is successful if it eventually identifies the target. In contrast to Gold learning, PAC learning model does not require exact identification of the target concept at any given step. Learning can be imprecise, but the precision improves with increasing probability after reasonable amount of data (polynomial in the measures of probablity and precision). PAC learning can be seen as a more realistic model of human learning success than identification in the limit. First, PAC learning allows for some degree of imprecision; and indeed, human cognition is not error-free, and has been argued to allow for some conceptual vagueness. Second, the definition of PAC learning success places a bound on the amount of necessary data, while Gold learnable classes include those where identification of the target hypothesis is theoretically possible but not computationally feasible.

I start by discussing previous results on the Gold learnable classes of determiners such as first order definable and left upward monotone determiners. I then identify a novel learnable class of determiners that covers all underived logical quantifiers in the language sample of [13]. I then turn to PAC learnability, and prove several negative results for classes of quantifier meanings discussed earlier.

2 Formal preliminaries

Generalized quantifiers are functions from sets and relations in a given model to truth values. A generalized quantifier has type $(k_1, ..., k_n)$ iff for an arbitrary model it is a map $\mathcal{P}E^{k_1} \times ... \times \mathcal{P}E^{k_n} \longrightarrow \{0, 1\}$, where E stands for the domain of a model. Denotations of natural language noun phrases (DPs) like *every student*, *John*, *some but not all teachers* have type (1). They map subsets of the model to truth values, e.g. 'John' maps a set P to 1 iff John is a member of the set P. Denotations of determiners in natural language (*every*, *three*, *his*, *the*, *three quarters of*, *between six and seven*) have type (1,1), mapping pairs of sets to truth values.

The following properties are typical for denotations of determiners in natural language:

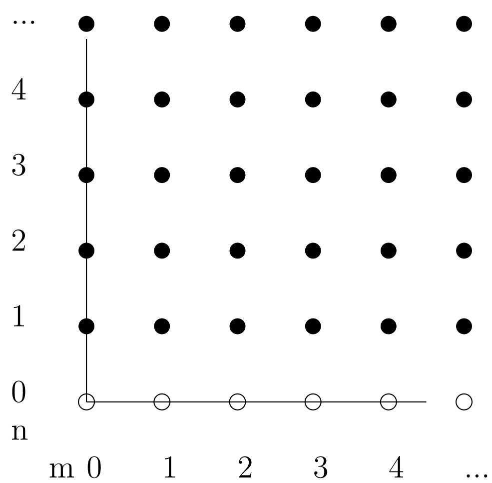
- Conservativity [14]. $Q_E(A, B) = Q_E(A \cap B)$
- Extension. For $A, B \subseteq E \subseteq E'$, $Q_E(A, B) = Q_{E'}(A, B)$
- Quantity (isomorphism invariance). For $f : E \cong E'$, $Q_E(A, B) = Q_{E'}(f(A), f(B))$
- Variation: for each E, there are E', A, B, C, such that $E \subseteq E', A, B, C \subseteq E'$, such that $Q_{E'}(A, B)$ and $\neg Q_{E'}(A, C)$

([16] propose only Conservativity and Extension as essential properties of natural language determiners and dedicate a few sections of the book to the study of determiners that fail Quantity)

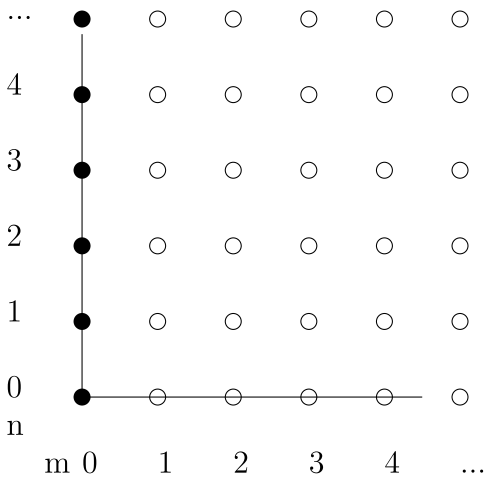
Properties of conservativity, extension, and quantity guarantee that $Q(A, B)$ depends just on the sizes of $A \cap B$ and $A - B$ (or, equivalently, just on the sizes of $A \cap B$ and A if A is finite). Quantifiers that satisfy these three properties are called *logical*. If we further assume that all semantic differences between natural language quantifiers can be observed on finite models, we can take the sizes of any subsets of the model to be natural numbers (This excludes from consideration mathematical quantifiers like *infinitely many* and *a continuum of*. It is not clear to me if these expressions in the mathematical meaning really belong to English as a natural language or to an artificial mathematical jargon). So we can characterize type (1,1) quantifiers that satisfy these requirements as functions from pairs of natural numbers to truth values. If n=$\mid A \cap B \mid$ and m=$\mid A - B \mid$, a logical quantifier Q maps (n,m) to Q(A,B). One can also think of pairs of natural numbers as two-dimensional vectors; then determiner denotations can be represented by sets of vectors.

Examples. Here are some examples of quantifiers interpreted as predicates of vectors. 'both' is true of just one pair (2,0): *both As are Bs* is true iff A contains exactly two elements and is a subset of B. 'all but five' is true of a vector (n,m) iff m=5. Indeed, *all but five A are B* is true just in case m=$\mid A - B \mid$ =5; n=$\mid A \cap B \mid$ is irrelevant. 'an even number of' is true of (n,m) iff n is even. 'exactly two thirds of' is true of (n,m) iff n=2m ($\mid A \cap B \mid / \mid A \mid = 2/3$).

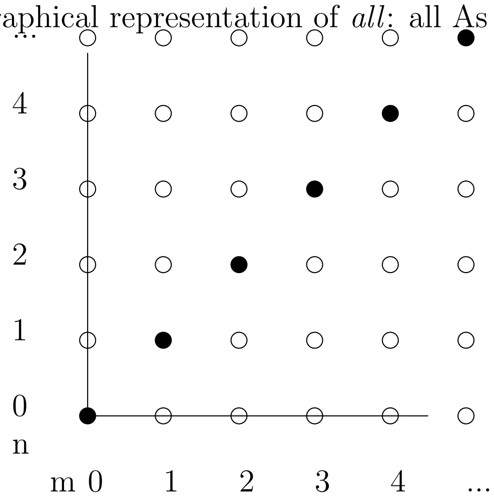
The sets of pairs of numbers that quantifiers correspond to can be represented visually on the Cartesian plane, or an infinite matrix where rows an columns are numbered.



Graphical representation of 'at least one' ($\approx$*some*): at least one A is B is true iff $n = \mid A \cap B \mid > 0$



Graphical representation of *all*: all As are Bs is true iff $m = \mid A - B \mid = 0$



Graphical representation of *exactly half*: exactly half As are Bs is true iff $m = \mid A - B \mid = n = \mid A \cap B \mid$.

If a relevant class of quantifiers is isomorphic to properties of vectors of natural numbers, we can encode them numerically; some such properties can be expressed by linear formulae.

Def. (m-dimensional) Linear basis is a sequence of m-dimensional vectors $v_i \in \mathbb{N}^m$, e.g. $\langle (1,4),(2,5) \rangle$. A linear set with basis $\langle v_0, v_1, ..., v_k \rangle$ is the set of vectors $\{v_0 + n_1 v_1 + ... + n_k v_k \mid n_i \in \mathbb{N}\}$.

Example. The set of number pairs that represents the meaning 'all but five' is linear. In this set, every member is a pair (n, m) where $m = 5$ (m stands for the number of As that aren't Bs, as relevant for the truth of *all but five As are Bs*). So the denotation of 'all but five' includes pairs (4,5) (4 As are Bs, 5 aren't), (100,5) (100 As are Bs, 5 aren't), etc. This set of vectors (pairs) is represented as $\{(0,5)+n_1(1,0) \mid n_1 \in \mathbb{N}\}$. The notation states that (0,5) is included, and so is any vector obtained from (0,5) by increasing the first member by any natural number n_1.

Def. A set is *semilinear* iff it is a finite union of linear sets.

3 Limitations of logical Generalized Quantifiers as a model for natural language determiners

Logical quantifiers as defined above can serve only as a very crude approximation of the meaning on natural language determiners. There are various quantificational phenomena that can not be adequately represented in terms of natural number pairs.

1. Natural language quantifies objects mostly on the integer scale, so *kissing three people* has a natural interpretation but *kissing three and a quarter people* makes no sense. But sometimes non-integer quantification is also allowed. One can *squeeze half of a lemon*, *eat two and three-quarters of an apple*, *paint three and a half chairs.* Lemons, apples, and chairs are count terms which overall tend to be quantified in integers; but with mass terms non-integer quantification is completely natural. Examples could be *drinking half of the water*, *buying 3.6 liters of juice*, *inhaling 0.932 cubic feet of air*, and *burning half of a ton of wood.*

2. The semantic values of interrogative determiners like *which* and *how many* are clearly not generalized quantifiers but a different type of denotation, formalized in various ways (Hamblin sets [10, 11]; Rexach's higher-type functions [17]; Karttunen's functions that include a quantifier and additional intensional semantic structure [12]).

3. Generalized quantifier perspective fails to draw linguistically important and cross-linguistically valid semantic distinctions, e.g. the collective vs. distributive distinction for quantifier words like *each*, *every*, and *all* [8].

4. Some natural language expressions that linguists analyze as determiners can be treated as generalized quantifiers but are not logical because they fail Quantity. Some classes of such determiners are analyzed in [16]. Possessive determiners like *your*, *Peter's*, *the president of America's* do not satisfy quantity as their value depends on identity of specific individuals (*you*, *Peter*, *America*), so it won't be preserved in isomorphic models. Values of other quantifiers depend on a set rather than an individual, e.g. *some student's* which depends on the set of students in a given model. The same applies to exceptive expressions like *every ... except John*, *no ... except some students*.

5. *Only* as in *Only students are in the room* is a famous example of a non-conservative quantifier [7]. Indeed, to assess the truth of this sentence it is not sufficient to know just the set of students and the set of students in the room; it crucially depends on whether some people in the room are *not* students. One might argue that *only* is not a determiner in English but rather a focus particle found in diverse syntactic environments. But other languages have equivalents of *only* that have obvious determiner properties. For instance, in Russian *odin* 'only' is used with NPs and not other kinds of phrases, and agrees with the noun in gender, case, and number. English *alone* 'only' as in *We trust John alone* is similar in the sense of just occurring in noun phrases, note though that it occupies an NP-final position unusual for English determiners.

6. Certain quantity expressions, like many natural language expressions generally, have vague meanings. It is impossible to specify exactly how many items would constitute 'roughly one hundred' or 'up to ten'. While in these examples fuzzy notions might be helpful, with 'roughly one hundred' being less true of 90 than of 91, other notions are not just fuzzy but context-dependent. How many items count as 'many' or 'few' depends on how many of them should be expected in a particular context; a carpenter can have many hammers and few nails even if he has more nails than hammers. The meaning of *many* is not isomorphism invariant, i.e. does not satisfy Quantity. Another example that demonstrates this comes from [14]. Indeed, imagine a medical congress which attracts fewer participants than usual this year, but they all happened to be lawyers in addition to being doctors. *Many doctors*

attended the meeting this year vs. *Many lawyers attended the meeting this year* can have different truth values even if doctors and lawyers are sets of equal size. This assumption guarantees the existence of an automorphism on the model that maps all doctors to lawyers and thereby changes the truth of the sentence *Many doctors came to the congress this year*. Shalom Lappin [15] analyzes such 'value judgment' quantifiers as intensional, with a proposal on how to introduce modality into their semantics.

Given all these limitations, any analysis of natural language determiners in terms of logical quantifiers involves major simplifications. Inasmuch as learnability is concerned, it is natural to assume that negative results have more weight than positive ones; if some class can't be learned in a simplified representation, so can't a broader, more complicated class.

4 Results on learnability

Let me now turn to some results on Gold learning (identification in the limit) of determiner denotations by Hans-Joerg Tiede [18].

4.1 Identification in the limit

Identification in the limit is a simple mathematical model of (unsupervised) learning from positive data. To learn a concept, in this perspective, is to identify it among a set of possible concepts. Applying the idea to quantifier meanings, learning a quantifier concept (e.g. the meaning of the word *two*) involves choosing the correct concept ('two') among its alternatives ('three', 'all', 'most' etc.).

Throughout this paper concepts are modelled simply as sets. Learning in the Gold perspective proceeds on the basis of examples (e.g. examples of situations where most As are Bs are the source of generalization of the meaning of *most*). The criterion of success is identifying the correct concept (set) after a finite number of examples (elements of the set).

After these informal preliminaries, let me give definitions of identification in the limit.

Def. A *learner* is a function from finite sequences of examples, $x_0, x_1, ...x_k$, for $x_i \in S$, into $\wp S$.

Def. A learner ϕ *identifies* an infinite sequence $X = x_0, x_1, ...$ iff there is $n \in \mathbb{N}$ such that for all $k \geq n$, $\phi(x_0, ...x_k) = \{x_i \mid i \in \mathbb{N}\}$.

Def. Learner ϕ identifies set C iff ϕ identifies any enumeration of C.

Identifying a particular concept is trivial. Indeed, if the concept to learn is fixed from the beginning, there's nothing to learn from data; data are uninformative. A much more interesting problem is identifying one of a set of possible concepts, based on positive data. This task is a rough approximation of human learning. That is, for each lexical item, a language learner identifies a concept from the set of possible lexical meanings. For instance, a child learns the meaning of the word *cat* while being shown examples of cats. Thanks to arbitrariness of linguistic sign, lexical meanings are not predetermined, so a child has no way to know in advance that *cat* is a word for an animal. So for each word there is a huge space of hypotheses on its meaning that can't be excluded *a priori*. In this paper, I restrict attention (somewhat artificially) to quantifier words and expressions. The motivation for this restriction is that quantifier concepts, unlike other lexical meanings, are mathematically well-understood.

If it is possible to identify a lexical meaning from data, the class of possible meanings is learnable. Here's a precise definition of a learnable class:

Def. A class of sets G is identifiable in the limit (Gold-learnable) iff there's a learner ϕ that identifies any set $C \in G$.

The property of finite thickness guarantees identifiability:

Def. A class of sets G has the distinguishing subset property iff for any $C \in G$ there is a finite distinguishing subset $D \subseteq C$ such that there is no $C' \in G$ such that $D \subseteq C' \subset C$.

Theorem. A class of sets G has is identifiable in the limit iff for it has the distinguishing subset property.

Corollary. If a class of sets $G \subseteq \wp S$ contains every finite $S' \subseteq S$ and at least one infinite set S'', G is not identifiable.

Proof. Take any finite $D \subset S''$. Since D is finite, $D \in G$, so $D \subseteq D \subset S''$, and D is not a distinguishing subset of S''. S'' has no distinguishing subset and G doesn't have the distinguishing subset property.

Def. A class of sets G has finite thickness iff for any s there are finitely many $C \in G$ such that $s \in C$.

Theorem [2]. If class G has finite thickness, G is learnable.

4.2 First Order Definability

Theorem 4.1. [18] The set of all first-order definable quantifiers is not identifiable.

Proof. The proofs relies on the fact that classes containing every finite set and one infinite set is not identifiable. Every finite set (of number pairs) is first order definable and there is an infinite set that is first order definable.

Example 4.1. One can easily construct a non-learnable subclass of FO: $\{$*exactly one*, *one or two*, *between one and three*, *between one and four*, ..., *between 1 and 100*, ...$\} \cup \{$*one or more*$\}$. It can be easily demonstrated that each quantifier in this class is first order definable. This class, however, is not identifiable in the limit per Angluin's theorem: *one or more* does not have a finite distinguishing subset that no member of the class is a superset of and a subset of *one or more*.

[6] discusses learnability of first order definable quantifiers in a setting with a *minimally adequate teacher* (MAT) [2]. A minimally adequate teacher evaluates the learner's hypothesis and provides it with counterexamples. Compared to already idealized model of identification in the limit, MAT learning setting simplifies the learning problem to almost trivial, but it is not even remotely realistic as a model of language acquisition, because humans acquiring language don't have access to explicit representations that a teacher could evaluate; their input is primary linguistic data, and not metalinguistic judgements.

4.3 Monotonicity

Monotonicity properties of quantifiers have received some scrutiny in the research on generalized quantifiers. Tiede discusses the relation of monotonicity to learnability.

Def. Left upward monotonicity (↑MON): quantifier Q is left upward monotone iff for all models $Q_E AB$ and $A \subseteq A' \implies Q_E A'B$. In vector notation this is equivalent to $Q(p, q) \implies Q(p + n, q + m)$, for all $n, m \in \mathbb{N}$.

Theorem 4.2. [18] The set of left upward monotone quantifiers is identifiable in the limit.

Proof. Define $succ(a, b) = \{(n, m) \mid n \geq a, m \geq b\}$. Take ↑MON and find a distinguishing subset for it $D = \{least(a, b) \mid succ(a, b) \subseteq Q\}$, i.e. $D = \{(a, b) \mid succ(a, b) \subseteq Q$ and for all $a', b', succ(a', b') \supset succ(a, b) \Rightarrow succ(a', b') \subsetneq Q\}$. By König's lemma, D is finite. Now if for a ↑MON quantifier $Q', Q' \supseteq D$, then $Q' \supseteq Q$, Q' must contain everything in Q. Indeed, take arbitrary $(a, b) \in Q$. Then $\succ (a, b) \subseteq Q$. If $(a, b) \in D, (a, b) \in Q'$. If $(a, b) \notin D$, then for some $(a', b') \in D, succ(a, b) \subset succ(a', b') \subseteq Q$ by definition of D (a full proof involves induction on a, b).

Fact. Other three monotonicity classes (left downward, right upward, right downward) are not identifiable in the limit ([18], proof just for MON↓).

Tiede cites Barwise and Cooper's tentative universal about monotonicity of natural language quantifiers in relation to these learnability results. However, Barwise and Cooper made no claims about left upward monotonicity, the only monotonicity property that guarantees learnability. In fact, they discuss monotone properties of NPs rather than determiners. Their conjecture is quoted below:

> **U6.** Monotonicity constraint. The simple NP's of any natural language express monotone quantifiers or conjunctions of monotone quantifiers.
> [3, 187]

Note that upward or downward monotonicity of the whole type (1) quantifier denoted by an NP translates into the corresponding *right* monotonicity of the determiner that the NP is built with, hence this proposed universal makes no positive identifiability predictions. In fact, the class in our example 4.1 contains only conjunctions (intersections) of right monotone quantifiers, but as we saw it is not identifiable in the limit.

4.4 (Semi)linearity

After first order definability and monotonicity properties of quantifiers, let us now turn to quantifiers that correspond to semilinear sets. The main learnability result on these is again negative.

Theorem 4.4. The set of semilinear sets is not identifiable.

Proof. It suffices to say that all finite and some infinite sets of vectors are semilinear. Our example 4.1 of a non-identifiable class also contains only semilinear sets.

Def. A set of vectors is in Abe normal form iff it is represented as a finite union of linear sets generated by at most three vectors.

Theorem. [1] Every 2-dimensional semilinear set has Abe normal form.

Fact. [18] One can find an Abe normal form for any semilinear set with exactly three vectors in the definition of each linear subset.

Def. $\mathcal{B} = \{\{v_0 + m_1v_1 + m_2v_2 \mid m_1, m_2 > 0\} \mid v_0, v_1, v_2 \in \mathbb{N}^2\}$

Theorem. [18] Any semilinear set can be represented as a union of sets in $\mathcal{B}$.

We will call such a representation the Tiede normal form of a semilinear set.

Theorem 4.7. $\mathcal{B}$ has finite thickness, therefore finite elasticity, therefore identifiable.

Def. Define a hierarchy of classes as follows:

$\mathcal{B}_0 = \mathcal{B}$;

$\mathcal{B}_{n+1} = \{b \cup b' \mid b \in \mathcal{B}_n, b' \in \mathcal{B}\}$.

Thus $b \in \mathcal{B}_{n-1}$ if it is a union of n sets in $\mathcal{B}$.

Theorem. [18] For any n, $\mathcal{B}_n$ is identifiable.

5 Monomorphemic quantifiers

We can now try to apply the mathematical results on quantifier learnability to natural language. Studies of development revealed a rich (probably innate) notion of quantity in human babies, but the research on quantifier word meaning identification primarily concentrated on the acquisition of numerals, which is of course not a representative class [4, ch. 9]. It is clear that in language acquisition not all quantifier denotations need to be learned from positive examples, as the Gold paradigm assumes. Some quantifiers are learned through explicit definitions, e.g. *prime number of*, *cubic number of* and other notions of arithmetic (it seems however that *odd number* and *even number*, though mathematical notions, might be used sufficiently in everyday contexts to be learned from examples before children are taught mathematics). Many quantifiers such as *between two and seven*, *three quarters*, *not many*, and *three hundred fifty seven* are structurally complex, with meanings predictable from the meanings of parts. To understand them it is not necessary to observe positive or negative examples; instead it is sufficient to learn the meanings of parts and the rules of semantic composition. This is very different from simple learning from positive data. Of course identifying a pattern of semantic composition requires prior identification of some basic notions through positive data, but the overall process is more complicated so let us set these cases aside. The only quantifiers that need to be learned through observation of examples are those with unpredictable meanings – either monomorphemic ones or idiomatic ones. Here are some data on monomorphemic determiners in several natural languages, taken from articles in the volume *Handbook of Quantifiers in Natural Language* [13]. These lists are certainly incomplete for some languages, but they give us an idea of the variety of monomorphemic quantifiers:

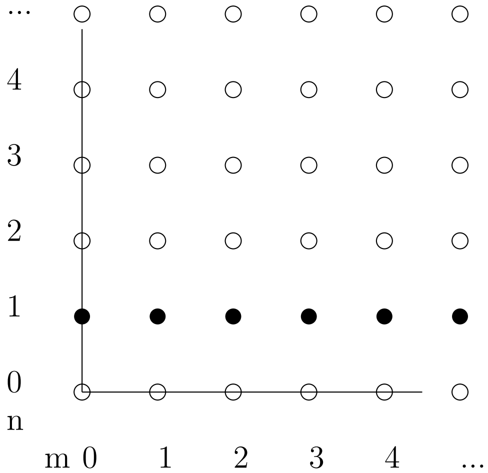
- **Adyghe**: 'one' and other numerals; 'some'; 'several'; 'a number of'; 'only' (!); 'many'; 'few'; 'little'; 'every', 'all'.
- **Basque**: 'always'; 'many times'; 'a lot'; 'often'; 'few times'; 'all'; 'many'; 'few'; ('half' - only occurs in combinations with other quantifiers); 'one'.
- **Garifuna**: 'all', 'one', 'many'.
- **German**: numerals 1-12; 'all'.
- **Hebrew**: 'always'; 'once'; 'often', 'very often'; 'majority of'; 'all'; 'one'; 'many'; 'few'; 'no'.
- **Hungarian**: 'every'; numerals 1, 2, 3, 10...; 'many'; 'few'.
- **Italian**: 'never'; 'always'; 'often'; 'all'; 'one'; 'many'; 'few, little'.
- **Malagasy**: 'all'; 'few'; 'many'; 1, 2 etc.; 'always'; 'usually'.
- **Mandarin Chinese**: 'no'; 'all'; 'often'.
- **Pima**: 1-9, 100, 1000; 'half' not monomorphemic but idiomatic;
- **Russian**: 'who'; 'what'; 'all'; 'every'; 'any'; numerals 0-10, 40, 100, 1000; 'many, a lot'; 'half'; 'one and half'; 'both'.
- **Telugu**: 'a'; 'several'; 'many/ much'; 'little/ few'; 'all'; 'always'; 'many' is used for 'most'.
- **Western Armenian**: 'all'; 'one'; 'many'; 'half'; 'often'; 'always'; 'few'; 'more'; 'less'; 'how many'; 'which'; 'what'; 'where'; 'when'; 'who'; numerals 0-10, 20, 100, 1000, million, billion.

We can summarize this list saying that typical monomorphemic quantifiers include 'all' and 'some', numerals, and value judgment cardinals 'many', 'few', 'several'; proportional 'half' and 'most' are rare but attested. These represent all three major semantics classes of determiners: cardinal (generalized existential), co-cardinal (generalized universal), and proportional. This list also includes some non-logical determiners: interrogative words, 'only', and value judgment cardinals. I don't have much to say about these since they are apparently more semantically complex than logical generalized quantifiers. Interrogatives can be analyzed as existential quantifiers coupled with a question operator [12]. Value judgment quantifiers like *many* can be seen as ordinary cardinal quantifiers which are vague. This would mean they take different specific values depending on the situation; in one context *many* is synonymous with *more than 5*, in another context with *more than 97*. So interrogative can be treated as logical, but restricted to special intensional contexts, and vague quantifiers might be seen as logical, modulo the context variable. 'Only' is semantically and syntactically distinct from all other quantifiers in many ways: it is conservative on the second argument and not the first, it is not a determiner, it combines with various syntactic categories, so it may not be relevant here.

Restricting our attention to logical quantifiers we can observe that all these determiner denotations are linear with at most two vectors in the basis. This means they are all Abe normal form linear sets, a class identifiable in the limit (subset of Tiede's class $\mathcal{B}_4$).

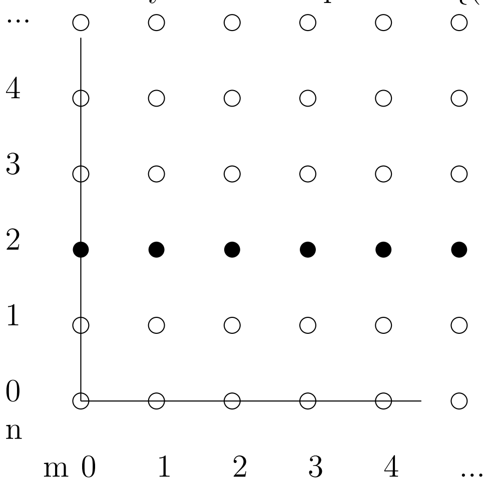
Denotations of cardinal numerals, whether we take the basic meaning to be 'exactly k' or 'at least k', are linear. For instance,

'exactly one' denotes the set n, m where $n = 1$. This can be represented as the vector set $\{(1, 0) + n(0, 1)\}$, or graphically as

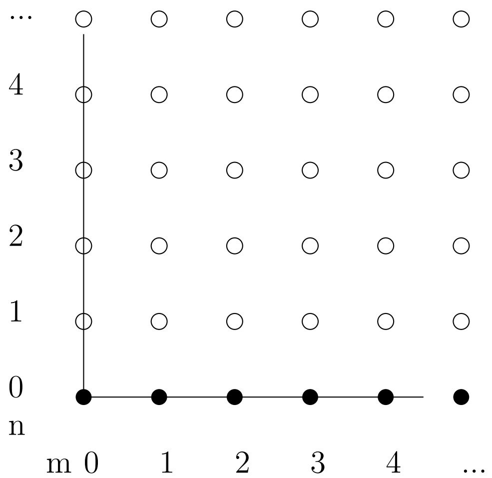


Similarly, 'at least one' denotes the set of pairs (n, m) where $n \geq 1$, represented as $\{(1,0) + n_1(0,1) + n_2(1,0)\}$; compare the graphical representation above.

'Exactly two' corresponds to $\{(2,0) + n(0,1)\}$, :



'at least two' corresponds to $\{(2,0)+n_1(0,1)+n_2(1,0)\}$, etc. The denotation of an arbitrary numeral k can be expressed either as $\{(k,0) + n(0,1)\}$ (the 'exactly k' reading) or as $\{(2,0)+n_1(0,1)+n_2(1,0)\}$ (the 'at least k' reading). This also covers 'no' (= 'exactly zero'). :



The denotation of ‘all’ is also arguable, but all the options are still linear. The most controversial issue is whether *All As are Bs* is true if there are no As. If we assume that *All As are Bs* can be vacuously true, the set of vectors that ‘all’ denotes is $\{(0,0)+n(1,0)\}$. Otherwise, if *All As are Bs* is untrue in the absence of As, this simply excludes one pair (0,0) from the denotation, so ‘all’ $=\{(1,0)+n(1,0)\}$.

Attested proportional quantifiers are also linear. ‘Half’ is the clearest example of a monomorphemic proportional quantifier, but there are other historically complex examples where the etymology is not quite transparent synchronically, like the English *quarter*. Generally, a statement of the form ‘Exactly $\frac{j}{k}$ As are Bs’ is true iff $\mid A \cap B \mid / \mid A \mid = \frac{j}{k}$, or, equivalently, $k \mid A \cap B \mid = j \mid A \mid = j \mid A \cap B \mid + j \mid A - B \mid$. So $(k-j) \mid A \cap B \mid = j \mid A - B \mid$, and ‘exactly $\frac{j}{k}$’ denotes the set of (n,m) such that $(k-j)n = jm$. Assuming the gcd of k and j is 1, the Abe normal form for such a set is $\{(0,0)+n(j,k-j)\}$. For instance, ‘exactly half’ = ‘exactly $\frac{1}{2}$’ translates into $\{(0,0)+n(1,1)\}$.

‘At least half’ is $\{(0,0)+n_1(1,1)+n_2(1,0)\}$, and ‘most’ = ‘more than half’ $=\{(1,0)+n_1(1,1)+n_2(1,0)\}$.

Conjecture. Any natural language monomorphemic quantifier is linear in Abe normal form.

This tentative universal property covers logical quantifiers listed above (*two*, *some*, *half*, *no* etc.), but also does not exclude meanings like *an even number of* or *an odd number out of the even number of*.

There are some quantifier denotations that this conjecture does exclude from being monomorphemic, e.g. *beween two and five*, *fewer than ten.* Most ↑MON quantifiers are not linear in Abe normal form, e.g. ‘at least five out of at least 10 or at least two out of at least 20' (recall that ↑MON quantifiers are identifiable). A more interesting example is *a prime number of*, which is not semilinear and I would argue is not a natural language quantifier. *A prime number* contrasts with *even* and *odd number* in this respect: the latter are linear, the former isn't. While both *prime* and *even* are apparently technical arithmetic terms, the notions of *even* and *odd* are used far beyond mathematics, e.g. in Chinese culture even number are considered lucky; in Russia, an odd number of flowers should be given to a living person, and even numbers of flowers are put on a grave. I don't know comparable uses of *prime number*.

6 Probably Approximately Correct learning (PAC)

We have applied to learnability of natural language determiners the fact that the class of linear sets in Tiede normal form is Gold-learnable. However, there is no guarantee that identification of the correct linear set will take a reasonable number of examples. To the contrary, it may take arbitrarily long. Turning to a different notion of learnability, we can show that linear sets in Tiede normal form are not PAC-learnable.

The notion of PAC learnability employs a different criterion of success. The learner is not required to identify the concept exactly (as in Gold-learning) but only approximately, allowing a small probability of misclassification. But the approximation is required to become more and more precise as more data come in. In other words, the probability to reach any given level of precision; the precise definition is given below:

Def. Assume a probability distribution P on the set S. Elements are drawn from S according to the probability distribution P, labelled for membership in the concept H (that is, a sample of positive and negative data is drawn). A ϕ learns H *probably approximately correctly* iff for any $\epsilon, \delta > 0$ there is a sample size n such that with probability at least $1 - \epsilon$, $P[H \ominus \phi(x_0, ..., x_{n-1})] \leq \delta$ for any probability distribution P, and n is polynomial in $\frac{1}{\epsilon}$ and $\frac{1}{\delta}$. Class $C \subseteq \wp S$ is *PAC-learnable* iff there is ϕ that learns any $H \in C$ probably approximately correctly.

Def. Class C shatters a set A iff for any $B \in \mathcal{P}A$, there is $B' \in C$ such that $B = A \cap B'$.

Def. A class C is said to have infinite VC-dimension (or *algebraic dimension*) iff for any $n \in \mathbb{N}$ there is a set with n elements that is shattered by C; otherwise C is said to have finite VC-dimension.

Theorem. [5] A class of sets is PAC-learnable iff it has finite VC-dimension.

Theorem. Semilinear sets have infinite VC-dimension and therefore are not PAC-learnable.

Proof. Any finite set of vectors is semilinear. Take an arbitrarily large finite set of vectors; all its subsets are semilinear, so semilinear sets shatter an arbitrarily large finite set, and their VC-dimension is infinite.

Theorem. First order definable determiners are not PAC-learnable.

Proof. Analogous to the previous theorem. The class of all first order definable determiners has infinite VC-dimension because it contains all finite sets of vectors.

Theorem. Linear sets in Tiede normal form have infinite VC-dimension and therefore are not PAC-learnable.

Proof. Take an arbitrary positive $n \in \mathbb{N}$ and show that linear sets in Tiede normal form have VC-dimension of at least n. Designate by p_k the k^{th} prime number, e.g. $p_1 = 2, p_3 = 5$. Let a_k be the product of all prime numbers up to the n^{th} except the k^{th}, $a_k = p_1 p_2 ... p_{k-2} p_{k-1} p_{k+1} p_{k+2} ... p_n$. Then consider the set of natural numbers $A = \{a_1, a_2 ... a_{n-1}, a_n\}$=$\{p_1 p_2 ... p_{k-2} p_{k-1} p_{k+1} p_{k+2} ... p_n \mid 0 < k \leq n\}$. This set has n distinct elements and is shattered by linear sets in Tiede normal form. Indeed, any subset $\{a_{i_1}, a_{i_2}, ..., a_{i_m}\} \subseteq A$ is the set of all positive integers in A that divide simultaneously by $p_{i_1}, p_{i_2}, ..., p_{i_m}$. The property "is positive and divides by x" is linear, expressed by $\{0 + xm \mid m > 0\}$, so an arbitrary subset of A $\{a_{i_1}, a_{i_2}, ..., a_{i_m}\}$=$A \cap \{0 + p_{i_1} p_{i_2} ... p_{i_m} k \mid k > 0\}$. The proof is done for the 1-dimensional case, but its extension to higher dimensions is trivial. Indeed, one can take all dimensions but one to be constant and proceed as in the one dimensional case. For instance, in case of 2-dimensional vectors one can take a_k to be $(0, p_1 p_2 ... p_{k-2} p_{k-1} p_{k+1} p_{k+2} ... p_n)$.

The other quantifier property that guarantees identifiability in the limit is left upward monotonicity. The class of quantifiers that satisfies it is again not PAC-learnable.

Theorem. Left upward monotone determiners ($\uparrow$MON) have infinite VC-dimension and are not PAC-learnable.

Proof. Take an arbitrarily large $n \in \mathbb{N}$ and show that VC-dimension of $\uparrow$MON is greater than n. Consider the set of $n + 1$ vectors $A = \{(i, n - i) \mid 0 \leq i \leq n\}$. It is shattered by left upward monotone functions. Let $B = \{(i_1, n - i_1), (i_2, n - i_2), ...(i_k, -i_k)\}$ be a subset of A. First observe that no vector $(i, n-i)$ is in $succ(j, n-j)$ if i is distinct from j. Then $\{(i_1, n-i_1), (i_2, n-i_2), ...(i_k, n-i_k)\}$=$A \cap B'$ where $B' = (succ(i_1, n-i_1) \cup succ(i_2, n-i_2) \cup ... \cup succ(i_k, n-i_k))$; B', as any union of successor sets, is left upward monotone.

This series of negative results suggests that all the identifiable classes of determiners discussed above are too broad. If we want to impose more realistic restrictions on learning, we should further narrow down the target class of meanings.

7 Extension to other types of quantifiers

In addition to types (1) and (1,1), study of quantification in natural language quantifiers also recognizes quantifiers of other types. For example, the quantifier denoted by *more ... than ...* in sentences like *More students than teachers sing* can be treated as having the type $(1, 1, 1)$.

MORE$(A, B, C) = true$ iff $\mid A \cap C \mid > \mid B \cap C \mid$. To represent quantifiers of this type, we might need to reserve a 4-dimensional space where coordinates are $\mid A \cap C \mid, \mid A - C \mid, \mid B \cap C \mid, \mid B - C \mid$. This particular quantifier is represented by the linear formula $(1, 0, 1, 0) + n_1(1, 0, 0, 0) + n_2(0, 1, 0, 0) + n_3(1, 0, 1, 0) + n_4(0, 0, 0, 1)$ and is a four-dimensional half-space. The expression *as many ... as ...* in *As many students as teachers came to the party* denotes a three-dimensional plane in the same four-dimensional space. Unfortunately, too few higher type quantifiers have been studied, so it is not clear what natural class of vector sets can serve as their model.

8 Conclusion

This paper discusses natural classes of determiner meanings that are identifiable in the limit (Gold learnable). Previously known learnable classes of determiners did not relate to the set of natural language determiners: first order and ↑MON determiners exclude common quantifiers like *most* and *half*, and Tiede's hierarchy of learnable semilinear classes is not linguistically motivated. The main positive result of the paper is that it identifies a Gold learnable class of determiners that includes all monomorphemic logical determiners attested in a representative sample of natural languages. This positive result is counterbalanced by a series of negative results on probably approximately correct learning: none of the Gold learnable classes of determiners turns out to be PAC learnable. It suggests that for a realistic semantic learning model, the initial set of hypotheses on quantifier meanings must be more restricted.

To what extent models of learning from positive data are an adequate formalization of human learning remains an open issue. One could ask, in particular, whether large numerals like *million* are learned and whether they could be learned by observing positive data. Are large numbers perceptually distinct at all? While learning big numerals from positive examples might be feasible, research on language acquisition [4] suggests that English-speaking children acquire meanings of numerals greater than 3 by mastering counting, while smaller numerals are indeed learned through observation of usage. But this does not prove that large numerals are not learnable, only that counting is a more efficient strategy for the specific task of learning numerals than generalizing from examples. Humans clearly can identify numbers of objects greater than three without counting, although this skill is of limited use.

References

[1] Naoki Abe. Polynomial learnability of semilinear sets. Technical report, University of Pennsylvania, 1989.

[2] Dana Angluin. Learning regular sets from queries and counterexamples. *Information and Computation*, 75(2):87–106, November 1987.

[3] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2):159–219, 1981.

[4] Paul Bloom. *How children learn the meanings of words.* MIT Press, 2002.

[5] Anselm Blumer, A. Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the vapnik-chervonenkis dimension. *J. ACM*, 36:929–965, October 1989.

[6] Robin Clark. Learning first order quantfier denotations: An essay in semantic learnability. Technical report, Institute of Research in Cognitive Science, University of Pennsylvania, 1996.

[7] SJAAK De MEY. only as a determiner and as a generalized quantifier. *Journal of Semantics*, 8(1-2):91–106, 1991.

[8] David Gil. Universal quantifiers and distributivity. In *Quantification in Natural Languages*, pages 321–362. Kluwer Academic Publishers, 1995.

[9] E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[10] Alan Hagstrom. *Decomposing Questions*. PhD thesis, Massachusents Institute of Technology, 1998.

[11] C. L. Hamblin. Questions in montague english. *Foundations of Language*, 10(1):pp. 41–53, 1973.

[12] Lauri Karttunen. Syntax and semantics of questions. In H. Hiz, editor, *Questions*, pages 165–210. Reidel, 1978.

[13] Edward L. Keenan and Denis Paperno, editors. *Handbook of Quantifiers in Natural Language*. Springer, to appear.

[14] Edward L. Keenan and Jonathan Stavi. A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9(3):253–326, 1986.

[15] Shalom Lappin. An intensional parametric semantics for vague quantifiers. *Linguistics and Philosophy*, 23(6):pp. 599–620, 2000.

[16] Stanley Peters and Dag Westerståhl. *Quantifiers in Language and Logic*. Clarendon Press, Oxford, 2006.

[17] J. Guitierrez Rexach. Questions and generalized quantifiers. In Anna Szabolcsi, editor, *Ways of Scope Taking*, volume 65 of *Studies in Linguistics and Philosophy*, pages 409–452. Kluwer Academic Publishers, 1997.

[18] Hans-Joerg Tiede. Identifiability in the limit of context-free generalized quantifiers. *J. of Language and Computation*, 1(1):93–102, 1999.

[19] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, November 1984.