Mathematics in language[1]

Richard Hudson, UCL

Abstract

Elementary mathematics is deeply rooted in ordinary language, which in some respects anticipates and supports the learning of mathematics, but which in other respects hinders this learning. This paper explores a number of areas of arithmetic and other elementary areas of mathematics, considering for each area whether it helps or hinders the young learner: counting and larger numbers, sets and brackets, algebra and variables, zero and negation, approximation, scales and relationships, and probability. The conclusion is that ordinary language anticipates the mathematics of counting, arithmetic, algebra, variables and brackets, zero and probability; but that negation, approximation and probability are particularly problematic because mathematics demands a different way of thinking, and different mental capacity, compared with ordinary language. School teachers should be aware of the mathematics already built into language so as to build on it; and they should also be able to offer special help in the conflict zones.

Keywords: numbers, grammar, semantics, negation, probability

1. Introduction

This paper is about the area of language that we might call 'mathematical language', and has a dual function, as a contribution to cognitive semantics and as an exercise in pedagogical linguistics. Unlike earlier discussions of mathematical language (Anon 1974; O'Halloran 2005; Shanahan, Shanahan & Misischia 2011) it focuses on elementary mathematics rather than the language of expert mathematicians.

Cognitive semantics, like cognitive linguistics in general, starts from the assumption that language is part of general cognition, without any special cognitive apparatus; in other words, "knowledge of language is knowledge" (Goldberg 1995: 5). The research aim is to produce an analysis without special formal apparatus. Clearly any analysis must make some formal assumptions, but the cognitive programme assumes that whatever formal assumptions are needed for language are also needed, independently of language, in general cognition. The empirical claim of cognitive linguistics is that the whole of language will turn out to be just like the rest of cognition, albeit cognition under the functional pressures of communication; this claim conflicts diametrically with the Chomskyan claim (now greatly reduced) that language constitutes a distinct 'module' of the mind (Chomsky 2011). My own preferred version of cognitive linguistics is Word Grammar (Hudson 1984; Hudson 1990; Hudson 2007; Hudson 2010; Gisborne 2010; Duran Eppler 2011; Traugott & Trousdale 2013) but in this paper I avoid using any of the special notions or notations of this theory.

The research strategy of cognitive linguistics emphasises the unity and interconnectedness of all knowledge, building on the massive evidence from cognitive psychology for a complex network structure in the mind which allows concepts to 'prime' other neighbouring, but disparate, concepts (as the word *doctor* primes neighbours such as *nurse, tractor*, *document* and *examine* – related to it

---

respectively through their meaning, morphology, phonology and syntax) (Hudson 2010). When applied to mathematics, the strategy seeks to find unity behind the apparent diversity of different available notations for representing mathematical notions such as those in (1) and (2):

(1) Two and three make five.
(2) 2 + 3 = 5

These very different notations are easily understood as implying very different knowledge structures belonging to different, and non-overlapping, domains of knowledge. As an exercise in cognitive semantics, this paper will seek the commonalities in semantic structure while recognising the differences of form.

The second purpose of the paper is to contribute to pedagogical linguistics, also known as educational linguistics (Hudson 2019). The link to education is obvious, given that mathematics is one of the main areas of knowledge that is normally learned through formal schooling, alongside the skills of reading and writing; together these constitute numeracy and literacy, the backbone of any educational system. The UK's national educational system performs poorly in both areas, and serves the least academic fifth of the population particularly badly, so there is ample room for improvement. The thesis of this paper is that elementary mathematics teaching would be more successful if it built more explicitly on the overlap between mathematics and ordinary language. In the words of a leader in maths education, "language analysis would support problem-solving, understanding, communication, reasoning, and analysis in mathematics" (Glaister 2019). This would require language analysis in the form of a careful study by teachers and students of certain details of the pupils' everyday language. It could of course also be argued that language analysis would help in the teaching of literacy; but although I am convinced this is also true, I won't pursue the idea here.

Taken together, these two goals mean that this paper will explore the cognitive structures of mathematics and relate them to those found elsewhere in the grammar of ordinary English; seen as pure research this will test the theoretical claims of cognitive semantics. However, the paper will also draw conclusions for school teaching about how mathematical structures might be explained better to school children who are already experts (as native or near-native speakers) in English.

2. Counting

We start with the basic building blocks of maths: numerals, i.e. individual words referring to numbers; we will consider number expressions with more complex structures in the next section. Numerals are clearly part of ordinary English, so we can contrast the numerals of English (*one, two, three, …*) with those of French (*un, deux, trois, …*), and the general grammar defines how numerals may be used and combined with other words – allowing, for example, *the three big books over there*, but not **three the big books over there*.

A numeral has a complex semantics, combining five different kinds of property:

1. distinctive, e.g. *three* is different from *two*
2. sequential, e.g. *three* follows *two*
3. quantitative, e.g. *three* is more than *two*.
4. set-referring, e.g. I'll have *three*, please.
5. purely numerical, e.g. -3 (the negative number minus 3).

(I use the word *set* here and below in the non-technical sense of a collection of objects; I say a little more about the differences in section 4, where I shall start to distinguish sets and collections.) When

small children learn numerals they probably learn their meanings in this order, except that #3 and #4 must evolve simultaneously.

The first challenge for a learner is to learn to distinguish and recognise the numerals, first as words and then later as written characters. In this respect they are of course just like other words, but unlike other words, a numeral may be used in such a way that its only function is to be distinctive. This is often the case in the adult world, where numerals and letters are used simply as arbitrary distinctive labels, as in telephone numbers or computer passwords. If your phone number ends, say, in 8143, there is no implication that it is some kind of meaningful neighbour of the phone whose number ends in 8144. Numbers are assigned randomly, and we treat them as random strings, naming each numeral separately: eight *one four three*, rather than *eight thousand one hundred and forty three*.

The next challenge for a small child is to invest numerals with sequential meaning: they simply learn the sequence *one – two – three – four - ….* Learning the numerals is just like learning the days of the week or the letters of the alphabet, or even the words of a nursery rhyme, where there is absolutely no link to measurement; so a child happily counts out the numerals while performing a repetitive operation such as hopping from stone to stone. This sequential meaning survives into areas of the adult world such as the numbering of houses in a street or the pages in a book; so if we know that a house is number 10 we guess that it stands between numbers 8 and 12, but we don't assume that it is larger than number 8. Alphanumeric sequencing plays a fundamental role in the way we organise information on the basis of the traditional order of letters in the alphabet plus the standard ordering of numerals.

Neither of these uses of numerals really qualifies as 'mathematical'. The main mathematically relevant function of numerals is their quantitative use, where they are used to compare quantities, though the sequential function comes into play when ordering a list. The quantitative use is clearly parasitic on their sequential function, so *three* not only follows *two*, but also indicates a larger quantity. This mental linkage between the two series in (3) and (4) is what defines the quantitative meaning, but it is combined with an extra unspoken assumption: that the difference between adjacent numerals is just 1.

(3) *one → two → three → four*
(4) *one < two < three < four*

This realisation brings the child to the point of counting, where they can correctly distinguish the meanings of expressions such as *two marbles* and *three marbles.* However, this move requires another conceptual leap: to the notion of sets or, more accurately, 'collections' (the linguistic counterpart of the mathematical set, discussed more fully in section 4). The point is that whereas an adjective such as *big* or *red* describes a property of an individual item (such as a marble), a numeral defines a property of a collection; so in order to understand *three big red marbles*, a child has to conceptualise not only the marbles but also the collection including them. In other words, the meaning of this phrase has the structure shown in Figure 1, where the italicised items are the words and the other items are elements in the conceptual structure that we call the phrase's meaning.
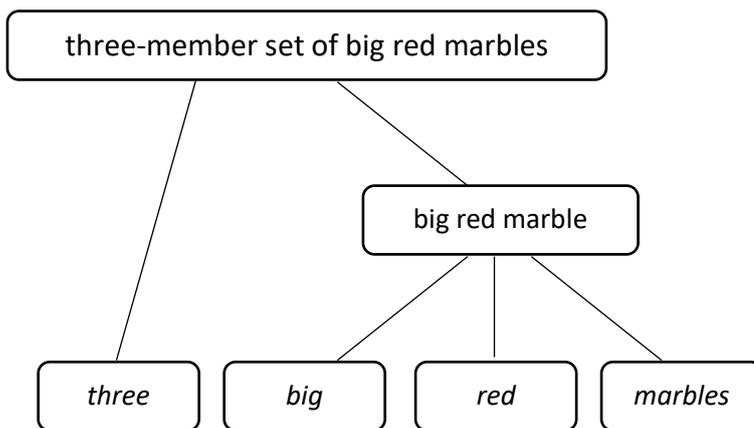
Figure 1: Three big red marbles

As Russell argued, recognising three marbles as a three-member collection requires a process of abstraction across all three-member collections – collections of three cars, people, apples or whatever. At this point we may conclude that the child has mastered the abstract concept 'three', but it may take time to make the same intellectual leap for all numerals. The child is now ready for mathematics.

This progression can be displayed in Figure 2, a diagram which includes the link between the quantitative meaning and the collections which makes concrete the fact that 3 is greater than 2: so a heap of three biscuits is greater than a heap of just two biscuits.
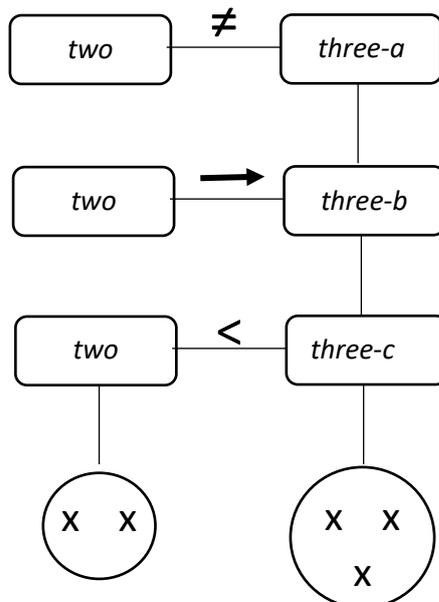

Figure 2: Three different meanings of <u>three</u>

The fourth use of *three*, to refer to a three-member collection, follows from a syntactic fact about English: that a numeral may be used on its own, without an immediately following noun, as in (5).

(5) I like the look of those apples; I'll have three please.

Of course, in this context *three* means 'three of those apples', but the fact is that the numeral can be used in this way, so a child hearing the word *three* has to process its meaning, even though it is merely an application of the quantitative function. On the other hand, it also reinforces the idea of a collection as something distinct from its members. In this respect it is comparable with collective nouns such as *crowd, herd* or *family*, each of which names a particular kind of collection; any child is therefore already familiar with the notion of collections before reaching mathematical sets.

Finally, we reach the purely numerical fifth step in the learning of numbers. A pure number combines all the associations of the earlier stages, but without being tied to the notion of a collection. The clearest evidence for this stage of development is arguably the ability to grasp the concept of negative numbers, since these cannot easily be identified as the size of a collection. Each number has now achieved the status of an ordinary concept, enriched by multiple properties of its own; for example, we can describe *three* as positive, odd, prime and so on. Having achieved independence from the collections that defined it initially, it can become the centre of a large and complex network of relations to other purely mathematical concepts.

The five-way ambiguity of numerals explains the subtle effects of agreement found in sentences such as (6) to(10):

(6) *Three* <u>is</u> a numeral.
(7) Three <u>comes</u> after two.
(8) Three <u>is</u> greater than two.
(9) Three <u>are</u> over there.
(10) Three <u>is</u> a prime number.

In the first three cases, *three* takes singular agreement (*is*) because it is the name of a single item: the word itself, an item in a sequence, or a quantity; but in example (9) it takes plural *are* because it refers to a collection of objects. The last example treats 'three' as a mathematical concept, independent of its word and of collections. These differences all have to be negotiated and internalised by a child on the way to mathematical insight.

The main pedagogical significance of this discussion is the complexity of the semantic structures around numerals, and the intellectual challenges that face a child on the way to understanding numerals as expressions of quantity. A primary-school teacher should be aware of these challenges, and should certainly not assume that knowledge of the numerals, as words, indicates an understanding of their role in counting and measurement. It is hardly worth stating that progress tends to be gradual and works by consolidating one level before moving on to the next. "Occasionally there are eureka moments when understanding happens consciously in discrete leaps or bounds. It is more typical, however, for understanding to take place unconsciously, by stealth, as the result of good study habits and perseverance." (Easdown 2006)

The further challenge for EAL children (who speak English as an additional language, in addition to their normal home language) means that the teacher needs, if possible, to be aware of other languages. The discussion so far, and indeed in the rest of this paper, focuses on English, but other languages can be very different (Hurford 2011). A few languages are reported to have no

numerals at all. At the other end of the spectrum, some languages combine numerals with 'classifiers', little words which add further information about typical members of the collection (e.g. that they are long and thin). Some languages have numerals with a base other than ten: 5, 12, and 20 are common, but other bases are also found. An EAL child's home language may be very different from English, so confusion is easy to imagine when confronted with elementary arithmetic in English. In other words, even if the "language of mathematics" is universal, the mathematics in language certainly is not.

3. Arithmetic, bases and larger numbers

The one-word quantity-numerals combine in ordinary English to define larger numbers – larger both syntactically (in terms of words) and semantically (in terms of the quantity defined). The combinations use the syntactic apparatus of ordinary words: dependency and coordination. For example, in *two hundred and six*, the word *and* signals a coordination while the juxtaposed numerals (*two hundred*) are held together by dependency. These two syntactic patterns carry different mathematical meanings: 'addition' for coordination, and 'multiplication' for dependency; in other words, the meaning of this expression is 2 * 100 + 6 (or 'two hundreds plus six').

To understand this system, the learner has to grasp the notion of a base. In modern English this is 10, but we still have traces of earlier bases in the words *dozen* (12) and *score* (20). Score is even more obtrusive in French, where 80 is *quatre-vingts* ('four twenties') and Danish, where both 60 and 80 are defined as a number of scores (*tres, firs*). The main point, of course, is that ten is by no means the only possible base, nor necessarily the most obvious one, so it is something that the child has to either be told explicitly or work out for themselves. Unfortunately, it is not at all easy to work out from the facts of English because the morphology of English numerals is quite opaque precisely at the point where it could be most helpful. This opacity is easy to see in Figure 3, where the semantic and syntactic units are in step for *two* and *forty two*, but out of step in between. (To avoid confusion, I use the Roman numeral X to indicate 'ten'.)
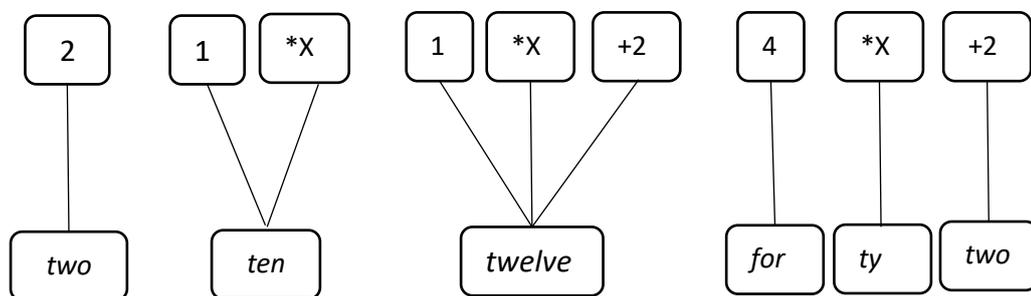

Figure 3: The opacity of low numbers

If the morphology had been transparent, 10 would not have been yet another monomorpheme; instead, it would have been *one ten*, or maybe *a ten* (like *a hundred*). Then 11 would have been *a ten and one* (like *a hundred and one*), and so on up to *two ten* (or *two ten*, like *two hundred*). This would have provided crystal-clear evidence for ten as the base. What we actually

have is a number system which remains morphologically opaque up to 12: … *ten, eleven, twelve*[2]. (Contrast the transparency of the equivalent French words: *dix, onze, douze*.) None of these words shows any similarity to any of the unit numerals; admittedly *twelve* starts like *two*, but the similarity in pronunciation is no greater than between *six* and *seven*. The higher teens are somewhat more transparent, but the elements are in the wrong order, and the *and* of addition is missing; so *thirteen* should be *ten and three*. As it is, *thirteen* sounds and looks remarkably similar to *thirty*, belying the fundamental difference between 10 + 3 and 3 * 10. And as for the numbers up to *ninety-nine*, they all continue to suppress *and*, which only appears transparently above a hundred (and even then is sometimes suppressed, especially in American English).

These comments on our numbers with complex structures point to two pedagogical conclusions, one negative and the other positive. The negative conclusion is that the intellectual challenge of learning how larger numbers work is considerable, because the morphology and syntax of English give very little help. Some children will crack the code easily, but others will struggle to achieve the deep understanding that arithmetic requires. This being so, the teacher needs to explain the system as clearly as possible, revisiting the explanation over and over again. The positive conclusion, however, is that anyone who can count in English already knows about addition and multiplication. The generalisation is simple: if a smaller number follows a larger one, the relation is addition; but in the reverse order it is multiplication. Thus in *twenty-three*, 3 is less than 20, so the meaning is 20 + 3; but in *thirty*, if we think of it as *three-ten*, 3 is less than 10 so the meaning is 3 * 10. This generalisation may be helpful with numbers below 100; above 100 it is still valid, but the presence or absence of *and* is an easier guide: *two hundred and thirty-two* can be analysed into 2 * 100 + 3 * 10 + 2.

4. Sets, collections and brackets

A set or collection is a conceptual unit which may have properties of its own and distinct from those of its members; the number of members (the mathematician's cardinality, which I shall call simply 'number') is one of these properties, but as we shall see below there are others. But as I noted earlier, although the sets found in natural languages such as English are similar to those of mathematics, there are also important differences of which any maths teacher should be aware. The differences include the following:

- A mathematical set may have any number of members, down to zero, but linguistic collections generally have at least two members.
- A mathematical set only allows 'collective' interpretations (as in *The crowd dispersed*, where dispersing has to apply to the entire crowd), whereas collections also allow 'distributive' interpretations (as in *The crowd are hungry*, where hunger applies to the individual members, not to the crowd as a whole).
- A mathematical set only recognises the 'and' relation between its members, whereas languages allow (and distinguish) a range of possible relations between members of a collection by means of conjunctions such as *or, but, then* and *nor.*

These differences justify the distinction made here between the 'collections' of language and the 'sets' of mathematics, but the similarities are also important, and collections presumably provide the conceptual foundations for sets, so we could call collections 'proto-sets'. However, for safety I shall

---

[2] Actually *eleven* and *twelve* can be broken down into meaningful parts if we ignore the ravages of phonological change. In Old English they were *endleofan* and *twelf,* where *leofan* or *lf* meant 'left over (after 10)', and *end* an *twe* of course meant 1 and 2.

keep to the terms *collection* and *set*, in the hope of encouraging mathematics teachers to be aware of the differences as potential snares along the conceptual road from collections to sets.

Returning to the property of number, it's easy to see the importance of number as the start of arithmetic, the point where numerals first acquire their quantitative significance. But more generally, collections provide the meaning for many parts of language, from collective nouns such as *crowd* or *herd* to the 'habitual' interpretation of verbs such as *got up* in (11), which may refer either to a single event or to a whole collection of them.

(11) He got up at eight.

Many languages distinguish systematically between one, two and many (i.e. more than two). English used to be such a language, and traces of the older system survive in sentence pairs like the following:

(12) He did it once/twice/three times. (NB *thrice* is no longer available.)
(13) They both/all came.
(14) Neither/None of them came.
(15) Did either/any of them come?

This rather simple two-way division of numbers is part of the mental apparatus that small children bring to school, and which they apply automatically when speaking, so we can be sure that every child is capable of distinguishing two from more than two.

As we have already seen, the number of members can be defined by numerals (which are grammatically distinct from adjectives). Confusingly, this property can also be defined by an adjective modifying the noun, so *numerous books* has a very different semantic structure from *big books*: a large collection of books for *numerous books*, in contrast with a collection of large books for *big books*. Figure 4 shows the mapping between syntax and semantics for *numerous big books*. Other adjectives that describe the collection rather than its individual members are *several, various* and *diverse.*
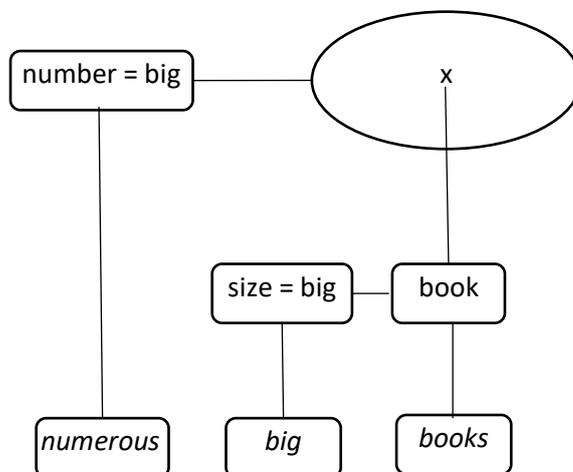

Figure 4: Numerous big books

Collections are also central to the meaning of two grammatical patterns: plural nouns and coordination. Plural nouns define a collection by intension (i.e. by providing a single definition of the

members), so *books* refers to a collection each of whose members is a book; whereas coordination defines by extension, i.e. by listing the members as in *John and Mary*, defining a collection with John and Mary as its two members. Beyond this difference, there are important similarities between the two kinds of collection.

Moreover, collections may 'project' their numbers in interesting and complex ways. If three boys make a cake each (i.e. if the reading is what grammarians call 'distributive'), then there will be three cake-making actions, leading to three cakes. But if they each make four cakes, then there will be $3 * 4 = 12$ cakes, since the interactions are multiplicative. This is the stuff of elementary arithmetic, but it is already built into our everyday language: language produces the initial formula $3 * 4$, and it's left to arithmetic to solve it. Of course, sometimes there are no specific quantities, as in *The boys each made some cakes*; but we all know that the total number of cakes reflects the number of boys and the number of cakes that each one made, even if we can't calculate the total. This projection of numbers is possible in almost any syntactic pattern, and not just in full clauses; for instance it also applies in noun phrases such as *the three boys' descriptions of four birds* which could refer to either four descriptions or up to twelve. Figure 5 shows how it can be generalised in terms of syntactic 'heads' and 'dependents', where the head of a clause is its verb (*made*) and that of a noun phrase is its noun (*descriptions*).
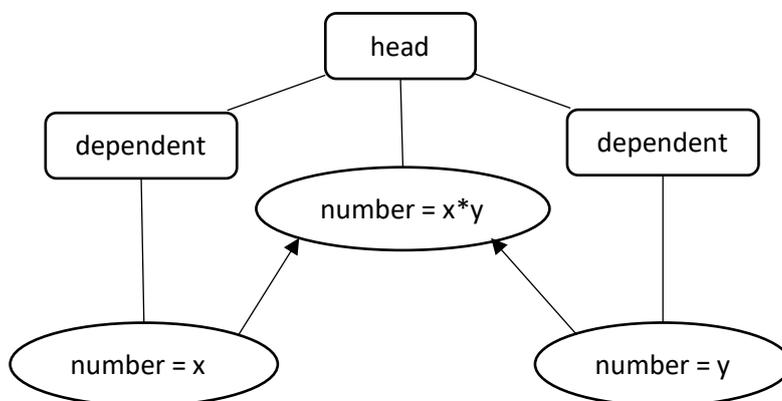

Figure 5: Generalising the projection of number

Putting number on one side, language also offers other ways of assigning properties to collections, including what are called 'participant roles'; so a collection may act as a participant in a situation. Indeed, some verbs such as *disperse* are specialised to take collections as their subject or object:

(16) The crowd/boys dispersed.
(17) The policemen dispersed the crowd.

In this example, 'dispersing' is one of the properties of the collection denoted by the noun, and the only kind of entity that can disperse or be dispersed is a collection, not an individual. Most verbs are not specialised in this way, so they accept either an individual or a collection, with corresponding differences of meaning. For example, *The boys made a cake* is ambiguous: either each boy made a cake (the distributive reading), or the boys collectively made one, meaning that 'making a cake' is

one of the collection's properties (the collective reading). It is the distributive reading that projects the collection's number up to the head as explained above. This ambiguity is a fact of ordinary spoken language, and no doubt it is as true of the language of small children as of that of adults. We can force a distributive reading by using *each* or *every* (*Each boy made a* cake, or *The boys each made a cake* or *The boys made a cake each*), and expressions like *between them* or *together* encourage a collective reading. But the main point for present purposes is that collections are already built into our grammatical thinking, as an essential part of our mental apparatus for interpreting plural nouns (or pronouns) and coordinations.

A further step towards arithmetic is the bracketing that is possible in syntax, and which is especially clear in syntactic coordination. For example,

(18) John and Bill and Mary played a trio.
(19) John, and Bill and Mary played a solo and a duet, respectively.
(20) John and Bill, and Mary played a duet and a solo, respectively.

These three sentences illustrate the three different ways of bracketing a series of three items: (1 2 3), (1 (2 3)) or ((1 2) 3). All are possible in ordinary language, and we recognise them all when pushed by the context. There are, of course, grammatical tricks for removing uncertainty about the bracketing, such as the removal of the first *and*, giving *John, Bill and Mary*; and punctuation can also help. But the point is that different semantic structures are available, and highly relevant because the differences are just the same as the distinctions made in mathematics by means of brackets, as between (a - (b − c)) versus ((a - b) − c).

Bracketing is pervasive in coordination, and not only when one coordination is part of another. Put most generally, bracketing allows the bracketed items to share their properties. In the case of (1 (2 3)), the items called 2 and 3 share their relation to 1; but the principle also allows coordinated items to share the same dependency relations to items outside the bracket. For instance, compare (21) and (22), where my bracketing is just an added notation to show the syntactic structure.

(21) He came in and he sat down.
(22) He (came in and sat down).

The second sentence is a reduced version of the first in which a single *he* is shared by the two 'conjuncts' (the elements of the coordination) rather than repeated; its syntactic structure is suggested in Figure 6, where the ellipse indicates a collection.
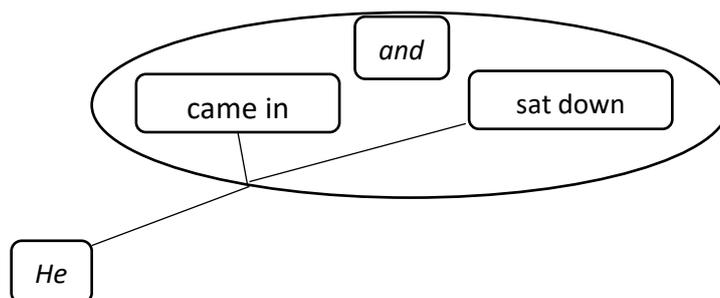

Figure 6: A collection combined with a word.

In this example, the shared item is a dependent, but it could also be the head (as in *John and Bill came in*) or both head and dependent (as in *John and Bill came in and sat down*); the latter possibility is displayed in Figure 7.
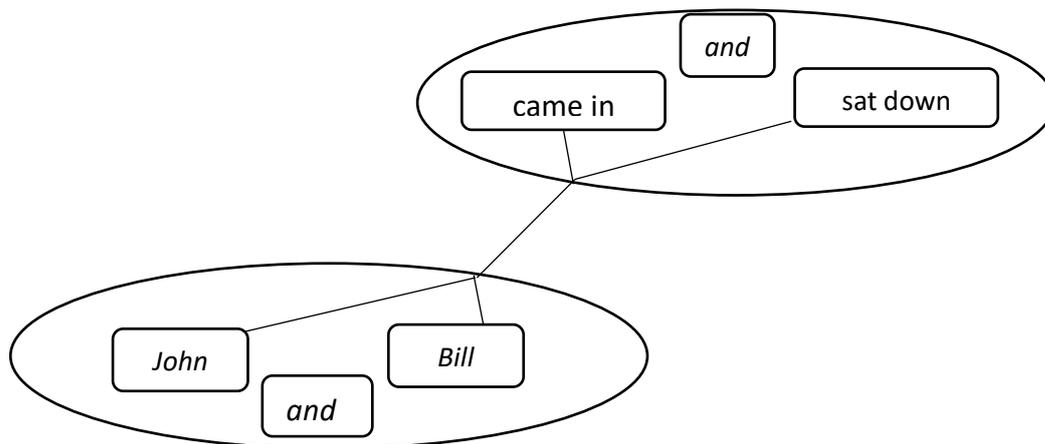

Figure 7: Two collections combined

Syntactic sharing examples such as these have precisely the same formal structure as a mathematical expression such as 3 (4 + 5), where (4 + 5) allows the two numbers to share their relation to 3. And as in maths, brackets are a useful way of removing ambiguities such as that in *old men and women*, which could be bracketed either as *(old men) and women* or as *old (men and women)*.

Once again the pedagogical conclusion is that teachers should try to build on the patterns of everyday language to illustrate and explain the idea of brackets in mathematical expressions. For instance, they could start by using brackets to distinguish the meanings of sentences like (18) to (20), before then moving on to mathematical equivalents.

5. Algebra and variables

Elementary algebra allows numbers to be replaced by letters to show either that the number is unknown, or that something is true for all possible numbers. The power of this approach comes from the assumption that each letter has the same value throughout the calculation. For example, in the equation $x + 3 = 5$, x has no specified numerical value, but the next step in the calculation, $x = 2$, gives it a value. By contrast, the identity $x + x = 2x$ is true whatever the numerical value of x may be. For simplicity (though not accuracy), we can call x in both cases a 'variable', meaning that its numerical value (i.e. its meaning) remains to be fixed by calculation, with 'true variable' for the case of $x + x = 2x$, where each instance of x is 'bound' by the other instances in the same equation.

Variables (in both senses) are commonplace in language, and especially so among pronouns. Personal pronouns (*I, you, she,* and so on) are variables whose value is fixed by the context of speaking, so the value of *I* varies according to who is speaking and that of *you* is tied to the person being addressed. A relative pronoun (e.g. *who*, as in *the student who wrote it*) has a value which varies with the noun to which it is added (its 'antecedent'), so in the example *… person who …*, it is the student. But of all the pronouns, the reflexive pronouns (*myself, yourself, herself, ….*) are arguably the most relevant because of their similarity to the true variables of algebra, which are bound within their mathematical expression. For instance, in (23) we don't know who *herself* is, but we can be certain that *herself* names the same person as *the woman* because English grammar

requires a reflexive pronoun to be bound to the subject of the clause containing it. (The rules are actually more complex than this, but the simple generalisation is a good first approximation.)

(23) The woman photographed herself.

Linguists tend to use algebraic subscripts to show this binding relationship, giving *The woman$_x$ … herself$_x$,* but a purely semantic analysis would replace *herself* by x, as in Figure 8.
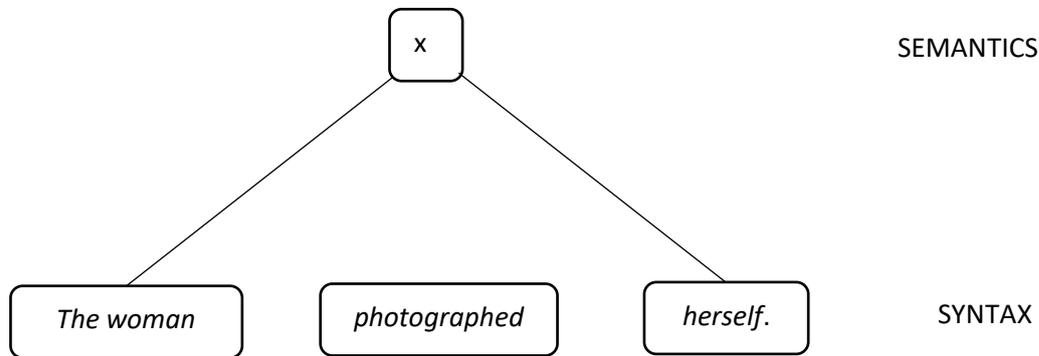

Figure 8: A variable as the meaning of a reflexive pronoun

This notation is particularly useful when there is no visible subject, as in (24).

(24) Photographing oneself is a sign of vanity.

In this case, a semantic representation of *photographing oneself* would be something like "x photograph x", where neither person is known but they are bound to one another.

These pronouns follow much the same principles as elementary algebra, and provide an excellent way in. Half an hour looking at examples such as these, and putting x on top of every word that names the same person as the reflexive pronoun, could make algebra seem very simple and obvious even to novices.

6. Zero and negation

One of the hardest areas of arithmetic, as of language, is the area occupied by zero and negative numbers. The discovery of zero was a late breakthrough in arithmetic compared with other numerals, and we still don't have a stable name for it – is it *zero* or *nought*? And negative numbers are counter-intuitive compared with their positive counterparts, as the process of subtraction (e.g. 5 - 3) turns into a kind of number which can be added (as in 5 + (-3), or even (-3) + 5). It's easy to understand 3 as the number of items in a three-member collection, but this route to understanding isn't available for negative numbers like -3.

The same is true In language, where negative meanings are much harder to process than their corresponding positives, as can easily be seen in the following little demonstration (Wason & Reich 1979). Consider sentence (25).

(25) No head wound is too trivial to ignore.

Does this seem reasonable? Now consider (26).

(26) No head wound is too trivial to treat.

Is this reasonable? Most people happily accept both these sentences as reasonable statements, but they must be wrong because the sentences contradict each other (ignoring a wound means not treating it). The little experiment shows that these two sentences are so hard to understand that we switch off our normal processing mechanisms, based on the words in front of us, and just guess. Why? Because each sentence contains too many hidden negatives. *No* is clearly negative, but so are *too* (if something is too big for purpose X, then X is not possible) and *trivial* (not important); and of course *ignore* contains another negative (not treat). In contrast, a positive version of (26) is much easier to process:

(27) Every head wound is sufficiently important to treat.

In this case it is obvious that *treat* is sensible, unlike *ignore*.

Moreover, language is the locus of a debate about 'double negation', as in the non-Standard *I didn't say nothing*. This has just the same double negation as the French equivalent, *Je n'ai rien dit* (literally 'I not have nothing said'), which of course is perfect Standard French; and the same used to be true in English until prescriptivists intervened in the 18th century. How can non-Standard *I didn't say nothing* carry the same meaning as Standard *I didn't say anything*? If it's just due to stupidity, why don't non-Standard speakers use a (much simpler) positive, *I said something*? And why doesn't an additional negation flip the meaning back to negative, as in *I didn't say nothing to no-one*? We seem to get confused very easily by negation, and it may be that the difficulty of negation in languages is related to the difficulty of zero and negative numbers in arithmetic.

The debate about double negation in grammar is interesting in relation to arithmetic because it's from arithmetic that the prescriptivist derived their inspiration. They argued that if two negatives make a positive, as they do in arithmetic so that (-2) * (-3) = 6, the same should be true in grammar, so *I didn't say nothing* must mean the same as *I said something*. One of the things that encouraged this linkage between negative numbers and negative sentences is the shared use of the word *negative* – an obvious example of language influencing our thought. So the prescription against double negatives can be blamed on the influences of both arithmetic and language (as well as on muddled thinking on the part of the prescribers).

The prescriptive argument relies on the linkage between negative sentences and negative numbers, but there is an alternative view: that the corresponding bit of arithmetic is actually zero, not negative numbers. Suppose the meaning of *not* is '0', and that the words to which it applies have a semantic structure like those described earlier, with a 'number' slot into which this 0 may fit. The number can now be taken as the number of instantiations reported. In that case, a positive sentence such as *It rained* could mean something like 'The number of examples of it raining = 1' – i.e. the raining happened. In contrast, *It didn't rain* would give the meaning 'The number of examples of it raining = 0' – i.e. the raining didn't happen.

Now the point of the change from negative numbers to zero as the mathematical model for linguistic negation is that the result of multiplying any number by 0 is always 0 – in contrast with negative numbers, where the result flips between positive and negative. In the zero-based approach to linguistic negation, *I didn't say nothing* is perfectly compatible with a negative interpretation, because both *didn't* and *nothing* offer 0 as their instantiation-number, so when these numbers project up (as described in Section 4) they both project 0, so the result is 0 – i.e. no instance of me saying something. The semantic structure is sketched in Figure 9.
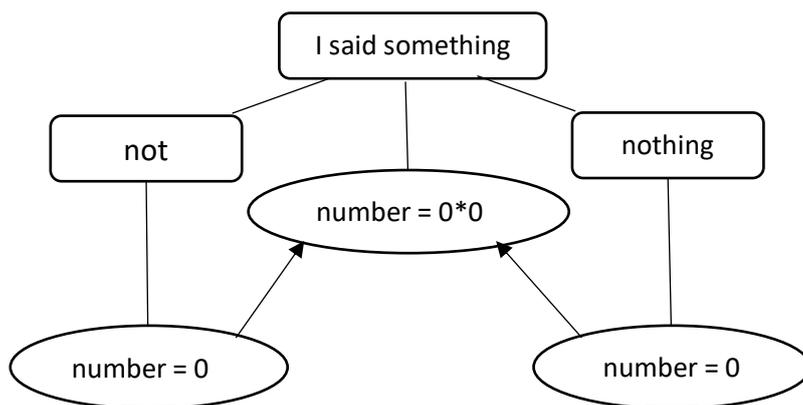
Figure 9: So-called double negation as zero multiplication

The Standard equivalent, *I didn't say anything*, is equally logical because the instantiation-number of *anything* (like every other *any-* pronoun) is unspecified, so again the result is 0. This analysis is shown in Figure 10.
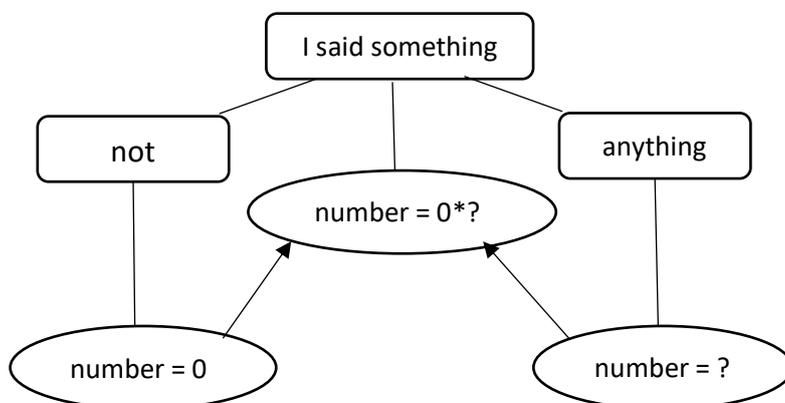

Figure 10: Single negation as unspecified number

And the difference between Standard and non-Standard grammar boils down to a simple matter of syntactic agreement: non-Standard applies a rule of 'negative concord' which is absent from Standard, whereby one negative element spreads its negativity across the whole clause (Labov 1972). Even Standard English applies a very limited version of this rule in sentences like (28), where the negative *wouldn't* triggers a spurious (and apparently contradictory) negative in the subordinate clause (here *didn't*).

(28) It wouldn't surprise me if it didn't rain.

Interestingly, this grammatical pattern has never attracted prescriptive comment, or even (as far as I know) any discussion by grammarians.

The message for pedagogy can once again be divided into two parts, one positive and one negative (in a very different and non-mathematical sense of *positive* and *negative*!). Starting with the negative part, negation and zero are a minefield of complexity in language. Given that zero is the absence of anything, we might expect it to be very simple, the baseline from which everything starts;

but for language, the baseline is positive: the grammar of negative sentences is always more complicated than that of positive sentences. (To see this, compare *It rained* with *It didn't* rain.) The default number in language is 1, not 0. This complexity is only increased by prescriptive condemnation of non-Standard negation (which, after all, is the grammar for the vast majority of school children). On the positive side, however, the grammar of ordinary English, both Standard and non-Standard, provides evidence for the effect of multiplication by zero. This observation may help some novices to understand this troublesome area of elementary mathematics.

7. Approximation

The next stop in this tour of points where language and mathematics meet is a fundamental difference between the two kinds of analysis. Mathematics is precise, and precision is at its root, whereas language is fundamentally 'quick and dirty', sacrificing precision for utility. Language is constrained by three considerations which don't apply to mathematics: speed, connectivity and broad coverage.

- Speed is vital, because information often needs to flow fast, so we need to be able to convey a lot of information in a few words.
- Connectivity is inevitable because language is connected intimately to anything relevant in our minds.
- Broad coverage is essential because we have to be able to cope with whatever life throws at us.

These constraints have many effects on language, but as far as mathematics is concerned, the most relevant effect is that language allows approximations which conflict with mathematical principles.

The need for speed is what motivates us to prefer short approximations to longer exactness; for example, if we are asked the time, we tend to round it to the brief *three o'clock* rather than the longer (and slower) *seven minutes to three* or (with a digital watch) *two fifty three*. For connectivity, one particularly relevant piece of evidence is the tendency for language to present quantities as points on a journey – a notion which seems to be deeply embedded in our general thinking. This emerges clearly in the famous contrast between *half empty* and *half full*, which allows us to think of a glass of beer either pessimistically (half empty) or optimistically (half full), according to whether we think of the beer as going out or coming in. Broad coverage is the motivation for metaphor and other kinds of extension of basic meaning, which allow us to apply familiar meanings to unfamiliar content (as when we extend *empty* to a glass which is actually full of air).

Returning to mathematical matters, take negation, the difference between positive and negative sentences such as *It rained* and *It didn't rain*. If we interpret this difference as argued in the previous section, negation distinguishes 1 (positive) and 0 (negative). But English allows approximations by treating this difference as a journey between the two extremes. The vehicles on this journey are the words *almost* and *hardly*, as in *It almost rained* and *It hardly rained*, both of which indicate something slightly less than a full, typical, 1, but approached from different directions. *Almost* approaches from 0, but stops just short of 1, whereas *hardly* approaches from 1, but presents the rain as moving towards 0; as I shall point out later, the effect is that almost raining is not raining, whereas hardly raining is raining. This journey is represented in Figure 11 by the arrow → between 0 and 1.
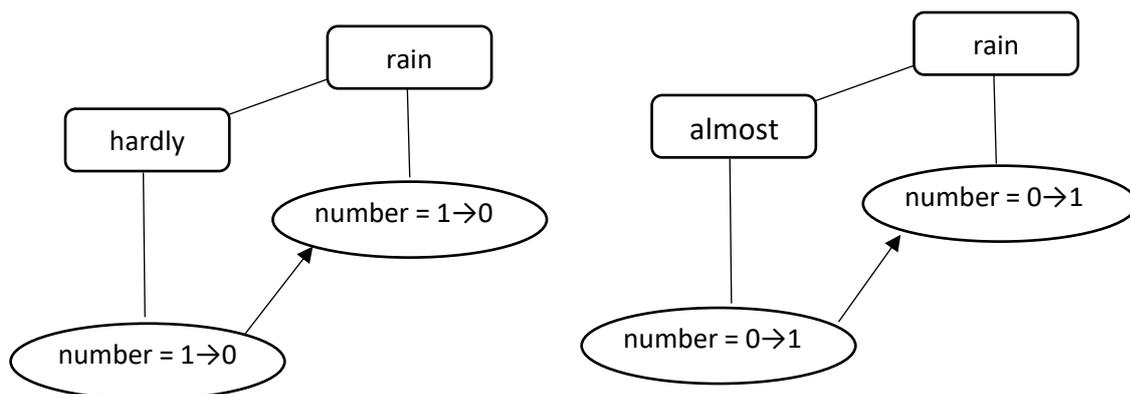
Figure 11: The meaning of hardly *and* almost

The effect of this journey-based approach is to introduce a mismatch between grammar and meaning because the words are classified grammatically according to the goal of the journey, rather than its source: we treat *almost* as grammatically positive, even though *It almost rained* means that it didn't rain, and we treat *hardly* as grammatically negative, even though *It hardly rained* means that it did rain. This claim about grammar is based on evidence such as the choice between *too* and *either*, which favours *also* in grammatically positive sentences and *either* in negatives:

(29) It rained too. (Not: It rained either.)
(30) It didn't rain either. (Probably not: It didn't rain too.)

Another test for grammatical negation is the 'tag question' added to a sentence, like *wasn't she* in *She was your friend, wasn't she?* This is a bit more complicated because you have to control the intonation on the question: if the intonation falls, then the 'polarity' of the question reverses that of the statement:

(31) She was your friend, wasn't she?
(32) She wasn't your friend, was she?

Now we return to our two vehicle-words, *hardly* and *almost* (or, incidentally, *scarcely* and *nearly*). This is how they behave in our two tests:

(33) It hardly rained either.
(34) It hardly rained, did it?
(35) It almost rained too.
(36) It almost rained, didn't it?

Conclusion: *hardly rained* is grammatically negative, although it did in fact rain, and *almost rained* is grammatically positive, although it didn't rain.

So what? The point of this discussion is to highlight the apparently complex ways in which language channels our thinking, all motivated by the functional pressures of speed, connectivity and broad coverage. The result is that when we speak, we not only give relatively low priority to precision but we also build our thinking around metaphorical journeys where the grammar is more concerned with the goal and direction of travel than with the current location. All these complexities seem a world apart from the clean world of mathematics.

8. Scales, relationships and comparison

Mathematics is arguably all about relationships, with the equality of '=' and the inequalities of '<' and '>' providing the foundations of a complex system based on comparison. Interestingly, the same is (arguably) true of language. Comparison plays a central role throughout the semantics of languages.

Take, for example, everyday gradient adjectives such as *big*, which identify a gradient scale – in this case size, but other scales indentified by adjectives are weight, intelligence, wealth and so on. A big dog is a dog which is relatively big, but 'relatively' means 'relative to the size of a typical dog'; so the meaning of *big dog* turns out to be the surprisingly complex 'a dog whose size is greater than that of the typical dog'. The typical dog is not the only possible standard of comparison, which can also be provided in the local context, such as the average size of a particular collection of dogs; but there is always, and necessarily, some standard of comparison; this is represented in Figure 12 simply as 'standard'. This being so, we expect the standard to vary from case to case, so a small elephant is much bigger than a big mouse.
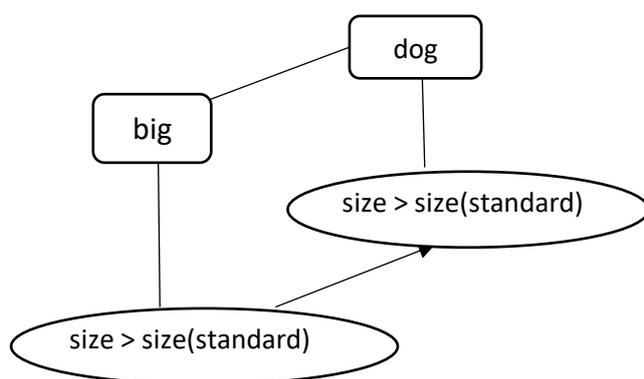

Figure 12: Size is relative

Another example illustrating the importance of comparison in language is the verb *to be*, as in (37).

(37) That building is the post office.

This is a special use of *be* which is exactly equivalent to the = of mathematics, meaning that anything which is true of the left-hand side (the grammatical subject) is equally true of the right-hand side (the grammatical complement). This use of *be* is different from the classificatory use in examples like (38).

(38) That building is a school.

In this case the relationship between the two sides is different: instead of equality we have class-membership (the building is an instance of the category 'school'); but it is still a relationship involving a comparison between the individual and the general category. This relationship is often called 'isa' in AI. The same verb finds yet another relational meaning in (39).

(39) That building is big.

Here the verb interacts with the adjective to define the relationship 'greater than' between the building's size and the size of the typical building.
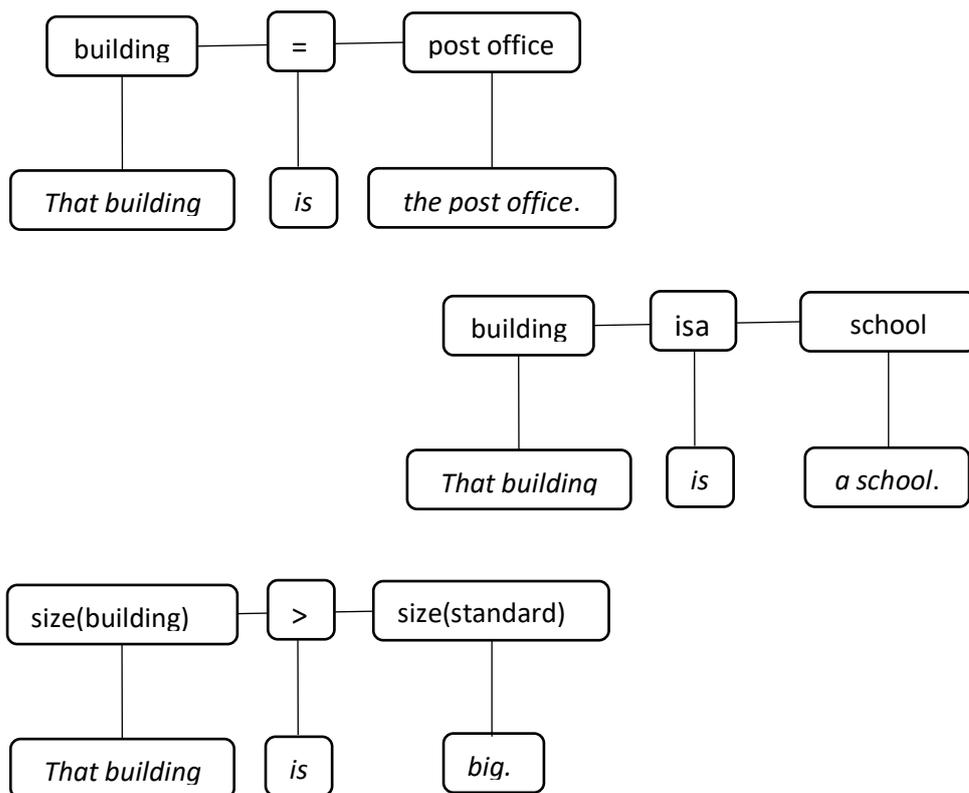
```
┌──────────┐   ┌───┐   ┌──────────────┐
│ building │───│ = │───│  post office │
└──────────┘   └───┘   └──────────────┘
      │          │            │
┌──────────────┐ ┌───┐ ┌──────────────────┐
│ That building │ │ is │ │ the post office. │
└──────────────┘ └───┘ └──────────────────┘


                  ┌──────────┐  ┌─────┐  ┌─────────┐
                  │ building │──│ isa │──│ school  │
                  └──────────┘  └─────┘  └─────────┘
                        │          │          │
                  ┌──────────────┐ ┌────┐ ┌───────────┐
                  │ That building │ │ is │ │ a school. │
                  └──────────────┘ └────┘ └───────────┘


┌──────────────┐   ┌───┐   ┌──────────────────┐
│ size(building) │──│ > │──│ size(standard)   │
└──────────────┘   └───┘   └──────────────────┘
      │              │              │
┌──────────────┐   ┌────┐       ┌──────┐
│ That building │   │ is │       │ big. │
└──────────────┘   └────┘       └──────┘
```

Figure 13: Three meanings of is.

These everyday words – adjectives such as *big* and the verb *be* – are typical rather than exceptional in having relationships at the heart of their meaning; and they clearly show how the basic relationships of mathematics are built into language even though (as in the case of the verb *to be*) there is potential ambiguity about which relationship is in play on a given occasion. Many of these relationships apply to some gradient scale such as the scale of size denoted by *big*. Some words are dedicated to defining relationships, with prepositions as the prime examples: *before, during, on, inside, because of,* and so on; and other words denote scales; these are mostly adjectives such as *big, heavy, clever, important* or nouns such as *size, weight, duration, time, loudness, quality, quantity,* and so on. It is these scales that permit quantitative comparisons, but (unlike mathematics) they usually have a directionality, with one end of the scale as its 'destination' – again, the journey metaphor applies, and once again it is the destination that determines the choice of words. With size, for instance, the destination is 'big' rather than 'small', so we query size with *how big* rather than *how small* (though, on occasion, when smallness is desirable, the latter is available).

One of the consequences of the directionality typical of linguistic scales is logical entailment. For instance, if a car will hold four people, this entails that it will also hold three. Given that scales are directional, entailment can work in either direction: downwards or upwards. The car example illustrates downward entailment because a claim about four automatically extends to all numbers below four. However, in some cases the entailment is upwards. Take, for instance, the word *enough*. If three people are enough to do a particular job, then four will be enough as well; and similarly for *too many*: if three people are too many to fit into a car, then four will also be too many. What these two examples share is a meaning linked to a particular threshold point on the score, and picking out

quantities that are higher than this threshold (as either enough or too many). The examples quoted here all involve the scale of number, but any scale offers the same possibilities of either downward or upward entailment. For example, if a problem is too difficult, then any other problem which is more difficult will also be too difficult (upward entailment); and if you can solve this difficult problem, then you can also solve any problem which is less difficult (downward entailment).

Turning to mathematics, the main scale is, of course, number, and directionality is implied by '<' and '>' (which defined my third stage of learning, where numerals are linked to the size of a collection). But there does not seem to be any implied journey from a source to a destination, though it is easier to imagine a journey starting at 1 than one starting at infinity.

Once again, the discussion suggests some important consequences for education. Maths teachers should be aware of the central role played in ordinary language by scales, comparison and thresholds (such as the size of a typical dog, or the threshold at which we declare 'too many'). General pedagogical principles favour starting with what learners already know and building on this in order to introduce new ways of thinking, so the obvious way to introduce the scales and relationships of mathematics is by looking first at those of ordinary language. Since children are manifestly expert in the latter, this should provide a firm and confident foundation on which to build the former. But in drawing the parallels with language, a mathematics teacher should also be aware of the differences, notable the journey metaphor that permeates language.

9. Probability

The last area of mathematics to be considered here is probability, the stuff of statistical inference. This is a particularly important area for this paper, because statistical thinking is probably the area of mathematics which most impinges on and challenges a typical adult. Public issues tend to have a statistical underpinning, whether they involve health (e.g. the risks of smoking), finances (the risks of losing your savings), education (the risks of getting or giving an unfair mark) or the weather (the risks of getting wet). The world in which our language evolved, and in which we live, is only partially predictable, and uncertainty is at the heart of language and, indeed, of everyday reasoning (with or without language).

One of the most important recent developments in cognitive psychology has been the widespread (but not universal) acceptance of the idea that concepts have the properties of a 'prototype' – a typical schematic member, rather than the neat classical definition in terms of necessary and sufficient conditions (Rosch 1973). This is easily demonstrated in everyday concepts such as 'fruit'. A prototype approach recognises typical cases, such as strawberries and apples, but also recognises borderline cases such as tomatoes, and untypical cases such as pineapples (which atypically grow directly in the ground). So if we wanted to explain to a child what the word *fruit* meant, we would give strawberries and apples as our examples, and hope the child wouldn't push the boundaries as far as tomatoes and pineapples. It's a matter of conjecture how we identify our prototypes, but it certainly reflects our personal experience (including our experience of explanations such as the above example), rather than a mathematically sound survey of data. Even a mathematically clear concept such as 'odd number' has a prototypical structure, where some odd numbers are considered better examples of the concept than others; e.g. *three* and *thirty-one* are better than *eleven* and *twenty-one* (Fodor 1998).
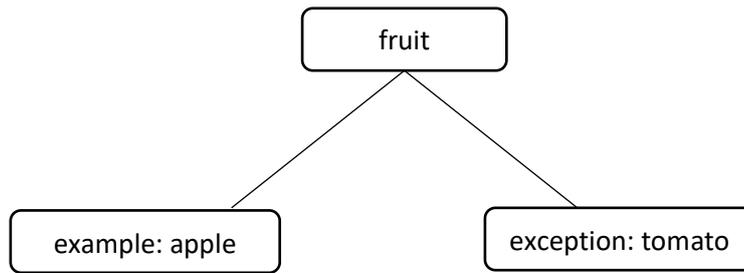
Figure 14: A prototype analysis of 'fruit'

If the prototype view of concepts is correct, then uncertainty and calculation of risk lie at the heart of all categorisation, because any given example of a concept may turn out to be untypical. Even an apparently innocent example of a potato, which you're happy to put in your mouth, could be untypically poisonous; but your risk-taking self takes a chance, and eats it.

English is full of tools for coping with uncertainty:

- modal auxiliary verbs: *may, will, might, can, …*
- adjectives: *likely, probable, impossible, …*
- adverbs: *maybe, perhaps, probably, ….*
- nouns: *risk, chance, probability, …*
- main verbs: *risk, gamble, doubt, estimate, think, …*

These important words allow a rich range of alternative ways of presenting possible or probable rain:

(40) It may rain.
(41) It's likely that it will rain.
(42) Maybe it will rain.
(43) There's a chance of rain.
(44) I think it will rain.

All these grammatical patterns are part of everyday language, so they represent well-oiled patterns of thought.

We also make subtle distinctions in our choice of tenses, as in (45) to (47).

(45) If I'm late, I will call you.
(46) If I was late, I would call you.
(47) If I had been late, I would have called you.

These tense changes (called in the TEFL literature 'the three conditionals') shift us conceptually into increasingly remote imaginary worlds. The simple future *will call* leaves us in the world of present reality, even though its details are only predicted. The past-tense *would call* moves us into an alternative reality which is still possible because it lies in the future; and the past perfect *would have called* puts us into an alternative world which was possible in the past, but which has now been overtaken by subsequent events. In Figure 15, the horizontal arrows link the real world, called 'world 1', to an alternative world 2, and the vertical arrows link earlier to later.
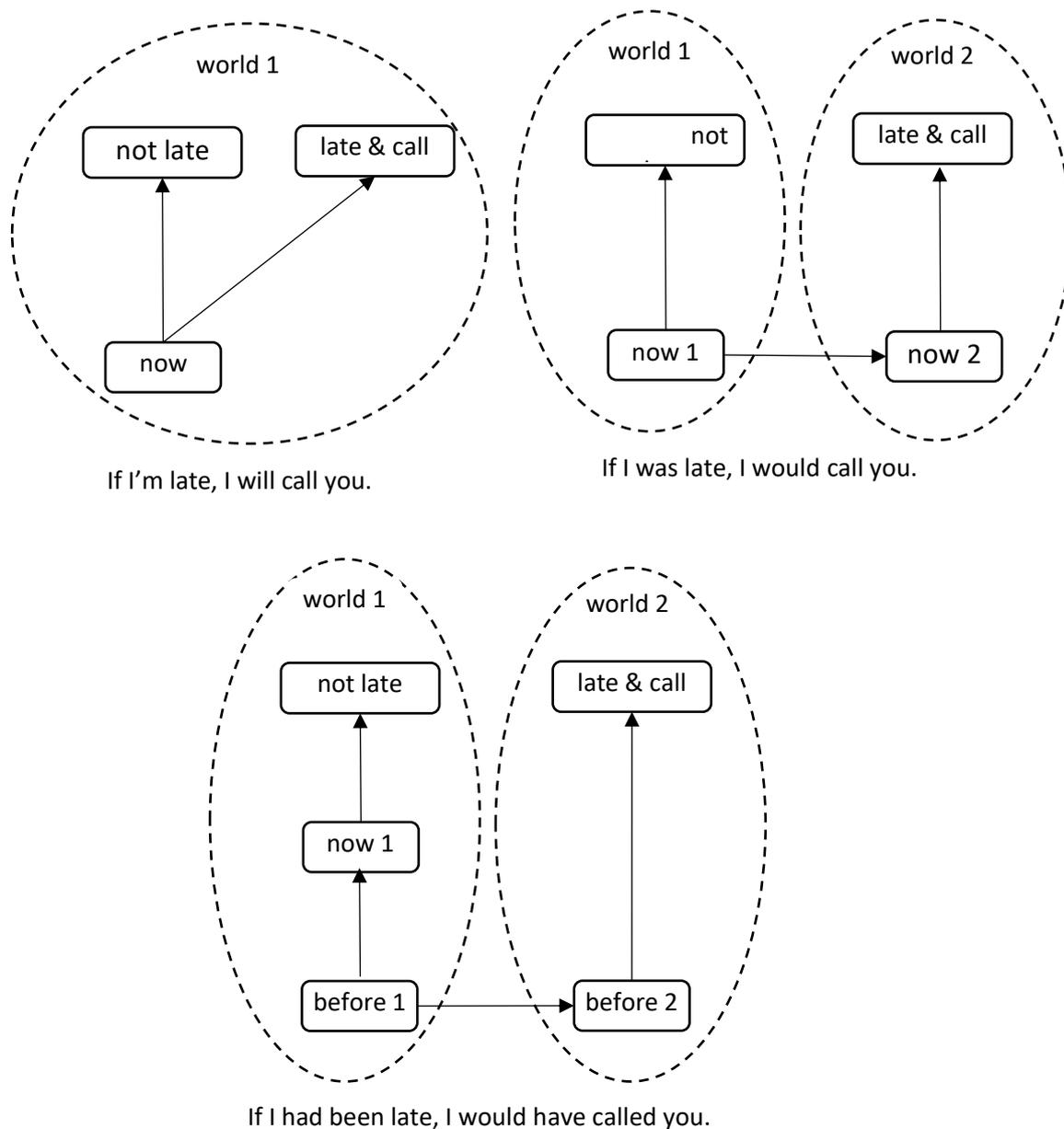
Figure 15: The three conditionals

In these sentences, the likelihood of my calling you dwindles from high to low to zero. The grammatical term for the low-probability options is *irrealis*, a semantic category which has many varieties and many different grammatical manifestations across languages of the world, enjoying names such as *subjunctive, conditional, optative* and *hortative* (see the Wikipedia page on 'irrealis').

Furthermore, we have sophisticated everyday vocabulary for discussing important areas of uncertainty, such as the verb *to risk* (Fillmore & Atkins 1992). This verb brings together a complex of elements in any risky decision:

- the person making the decision: *Jo risked everything.*
- the risky action: *Jo risked another drink*.
- the possible negative outcome: *Jo risked a fine*.
- the thing threatened by the negative outcome: *Jo risked her job*.
- the purpose of the risky action: *Jo risked everything for peace*.

Analysis of examples like these leads to a deeper understanding of risk-taking and the complex tension between possible gains and possible losses, both of which are evaluated quantitatively: how dire is the negative outcome and how good is the positive outcome? and how likely is each outcome?

And finally, there is massive evidence that we are astonishingly sensitive to statistical properties of the language that we experience and produce in speech. This evidence comes from the field of quantitative sociolinguistics, which focuses on variation which carries social meaning but not conceptual meaning – where there are different ways of saying the same thing, and each way is associated with some kind of social identity or value.

To take a simple (and now famous) example, some Americans pronounce the letter <r> in words like *fourth,* whereas others don't; but many (possibly most) alternate between the two pronunciations. William Labov launched quantitative sociolinguistics by a brilliant bit of research focused on one phrase, *fourth floor*, and carried out in three department stores in New York. In each store he located an item that was sold on the fourth floor, then asked dozens of shop assistants on the other floors where he might find this item; and after hearing the answer he got a repeated utterance by pretending not to have heard it. This gave him hundreds of data points in a single afternoon's research. Now the relevant fact about this <r> (which doesn't apply to syllable-initial <r> in words like *red*, which all Americans always pronounce) is that pronouncing it is considered to carry high social prestige, so it is strongly linked with upper-class and formal speech. What he found confirmed this precisely, with clear differences between the stores (which varied in the class of their target shoppers) and also between the first utterances and the more careful repetitions. The results are displayed in diagrams such as Figure 16 (Hudson 1996: 157), which give group percentages but also hint at systematic variation in the speech of individuals between the first and second pronunciation – and also between the words *fourth* and *floor*.
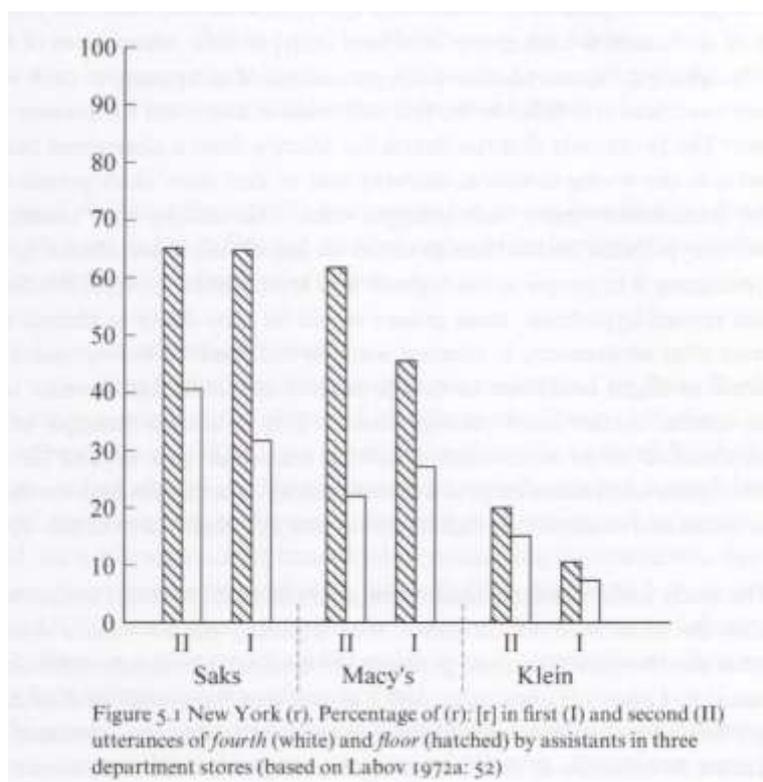

Figure 5.1 New York (r). Percentage of (r): [r] in first (I) and second (II) utterances of *fourth* (white) and *floor* (hatched) by assistants in three department stores (based on Labov 1972a: 52)

Figure 16: Pronunciation of syllable-final <r> in three New York stores

This work triggered an explosion of excellent research, which continues to the present. A particularly nice (and relevant) example is a study of the speech of eleven children in Buckie, a small town in the north of Scotland (Smith, Durham & Fortune 2007). This research studied large samples of speech between these children (aged between 2;6 and 4;0) and their carers, paying particular attention to two variable features:

- the pronunciation of words such as *house*, which in Buckie have two alternative pronunciations of the vowel: as a monophthong [u:] or as a diphthong [ʌʉ].
- the use of (standardly) singular verbs with plural subjects, as in example (48).

(48) Your leggies *are* cold. Your feeties *is* cold as well, *aren't* they?

Perhaps the most striking outcome of this research is the diagram in Figure 17, which shows for each child/carer pair the percentage of words such as *house* or *out* that were pronounced with a monophthong. The children are ordered by age, and it can be seen that the older children have percentages which are remarkably close to those of their carers (typically their mothers), which indicates very precise statistically-driven learning. The research also explored the use of a monophthong in individual lexical items, where there is also considerable and systematic variation, and found that even here the children tracked the adults closely.
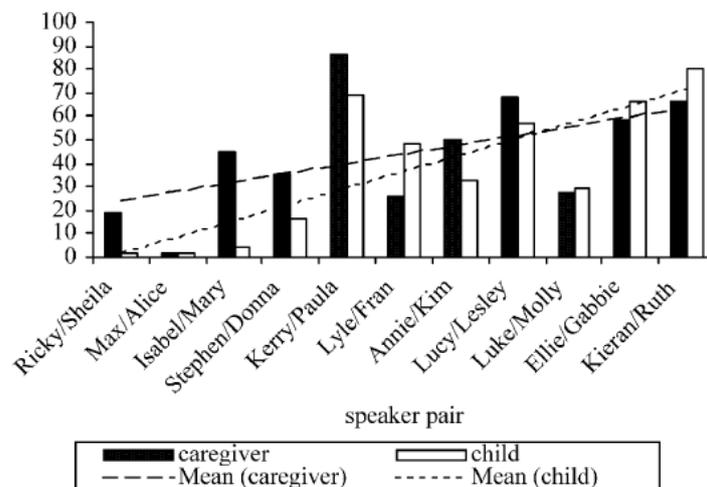

Figure 17: Child/carer uses of [u:] in Buckie

The other variable (singular verbs with plural subjects) showed a similar close match between the figures for the children and for their carers, but in other respects was different from the *house* variable.

The point made by this excursus into sociolinguistics is that our learning and our behaviour are driven by the statistics of experience: we tend to model our own speech on the speech of those around us, and this includes not only the fine details of grammar and vocabulary, but also the quantitative patterns that have been studied so extensively and intensively in sociolinguistics. This view is supported by equally massive evidence from cognitive psychology, which shows pervasive quantitative effects of experience, notably the effects of frequency and recency (Reisberg 2007: 145, 406).

What emerges from the discussion so far is that both linguistics and sociolinguistics show how well equipped we are with linguistic tools for thinking about uncertainty and hypothetical situations and for reacting in our behaviour to degrees of probability. And yet we seem to find it really hard to think about probabilities, as is easily demonstrated by the following example offered by Hans Gigerenzer (which, incidentally, introduces yet another meaning of the word *positive*):

(49) The probability that a woman of age 40 has breast cancer is about 1 percent. If she has breast cancer, the probability that she tests positive on a screening mammogram is 90 percent. If she does not have breast cancer, the probability that she nevertheless tests positive is 9 percent. What are the chances that a woman who tests positive actually has breast cancer? (Gigerenzer 2002: 5)

As Gigerenzer says, the answer probably seems 'foggy' to you. In contrast, the following reformulation is rather easy to understand:

(50) Think of 100 women. One has breast cancer, and she will probably test positive. Of the 99 who do not have breast cancer, 9 will also test positive. Thus, a total of 10 women will test positive. How many of those who test positive actually have breast cancer? (ibid: 6)

Rather obviously, only about 1/10 positive results actually indicates cancer – about 10%, not the 90% suggested by many people confronted with the original question. The radical difference between the two presentations emerges starkly when we diagram them, as in Figure. (The Venn diagram deliberately leaves part of the '1 cancer' ellipse outside the '10 positive' one in order to show that the woman with cancer may not test positive.)
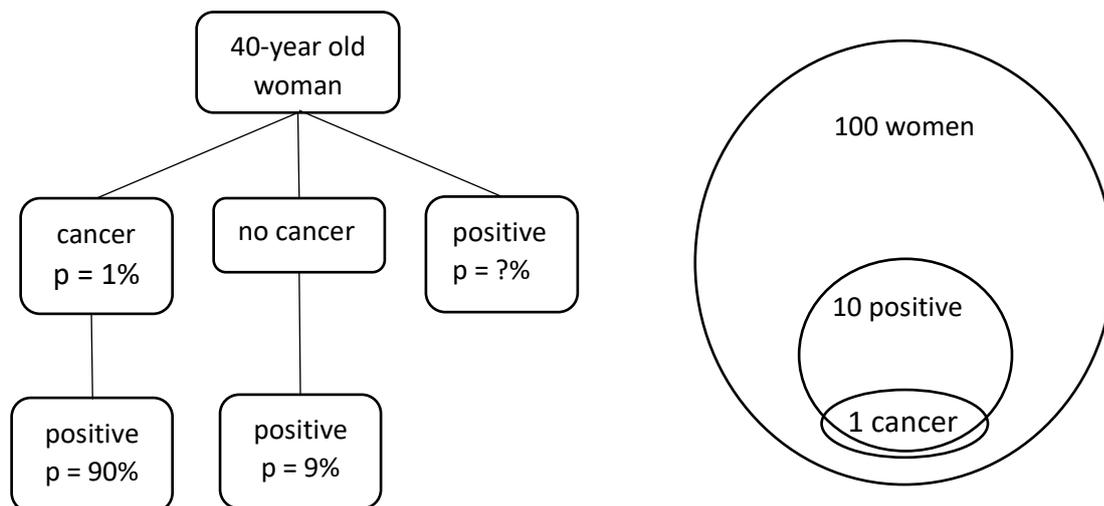

Figure: Two views of breast-cancer screening

This example suggests that the difficulty of probability lies in the language we use for discussing it. In short, the focus in teaching statistics should move from the numbers to the language and the underlying conceptual structures.

> Modern approaches to statistics emphasise the importance of formulating problems and questions, interpreting results, carrying out inference and communicating conclusions. These activities have to be carried out by people, not computers – and the medium in which they are carried out is language. (Sheldon 2019: 1)

I should like to make some suggestions along these lines for how we teach and use statistics, but first let's consider one possible source of difficulty: the intellectual difficulty of sentences with the syntactic form *The probability that … is … percent*.

How can a probability be a percentage? Those who are steeped in statistics may not see the issue because this terminology is so familiar, but it hides a serious problem linked to the notion of a gradient. I have already mentioned 'gradient adjectives' such as *big*. These are adjectives which identify a gradient parameter, a continuous scale on which objects can be located and compared. Since *big* is a gradient adjective, we can contrast different degrees of big-ness (i.e. size) using adverbs such as *very* or comparatives such as *bigger*: thus a very big book is bigger than a merely big book. The relevance of all this to probability lies in the fact that there are two different kinds of gradient adjectives differing according to whether their gradient is open-ended, like *big*, or closed, like *full*. If the gradient is closed, then it has an upper limit and allows adverbs such as *completely* or *utterly*, so a bottle can be completely full. But crucially it cannot be completely big.

Returning to probabilities, the problem is that the adjective *probable* defines an open-ended scale. Nothing can be completely probable; and since the noun *probability* takes its meaning from the adjective, it too shares this property. But if there is no upper limit to probability, how can it be measured as a percentage? To put the problem in another way, what would 100% probability be? The obvious answer is that it would be certainty, but that shifts the meaning from probability to a different gradient, certainty or confidence. Probability is a property of an event, whereas certainty applies to people: I can be certain of something, but I can't be probable of anything. Conversely, something may be probable and we may describe it as certain, but what we really mean by certainty is that we are certain about it. (So *certain* is in this respect like *comfortable* or *happy*, which can be applied to objects, such as a comfortable chair or a happy occasion, but which actually tell us that these objects make people comfortable or happy.) In short, if we link a claim to a percentage, the percentage must belong to the certainty of the claimant, not to the claim itself.

You could, of course, counter this objection by saying that I'm talking about everyday language, whereas probabilities expressed as percentages belong to the very special register of statistics. As a statement of the facts this is perfectly true, but it misses the point: a novice has to learn this register, and this particular item of learning is challenging, which means that statistics teachers should at least be aware of the challenge. It's even possible to avoid talking about probabilities altogether; for instance, the first sentence of (49) could have been rephrased like this, without even mentioning probability.

(51) About 1 percent of women of age 40 have breast cancer.

It's hard to see any loss of information from the original to this much simpler rewrite.

Given these concerns about the word *probability* and its meaning, it would be reasonable to conclude that statisticians would do well to avoid it when teaching novices or presenting their findings to the general public. The same conclusion arguably applies to words such as *average* and *significant*. For example, take the humps on a camel: either one or two. Most camels have one, but 6% (the bactrian camels) have two. The most helpful way to present these facts is to say that most camels have one hump, but exceptionally bactrians have two. Talking about averages is unhelpful, and leads to nonsense such as an average camel having 1.06 humps (Sheldon 2019). Similarly for *significant*: if a result is significant at the 5% level in a statistical test, it's unhelpful and misleading to attach the everyday meaning to the word 'significant' to the outcome. The outcome may be trivial and of no importance in the everyday sense. And even in the statistical sense there is still a 5%

probability that the outcome arose by chance. As with *probability,* the word *significant* has too much everyday baggage which may be easier for a statistician to ignore than for a novice.

What, then, is to be done with statistics? The way in which we teach and use statistics should build on the intellectual tasks that ordinary people are good at, and should where possible avoid those that are too hard for the typical human mind, or at least leave them till later in the curriculum. The human mind is brilliant at some things, but rather bad at others; one of our weak points is negation, which I discussed in section 6 in relation to the sentence *No head wound is too trivial to be treated*. We can now add the problems we noted above in connection with the notions of probability, average and significance. These topics do not belong in elementary statistics, which should instead start by building on the strengths that most students already have.

From the previous discussion, we know that most of us are relatively good at a number of things, with several skills standing out as particularly relevant to probability:

1. imagining alternative worlds (as in *If I was late, I would call you*).
2. linking numbers to collections.
3. thinking in terms of typical cases (i.e. prototypes).
4. thinking in terms of relative probability (as in *It may rain*).
5. thinking in terms of imaginary journeys (as in *It almost rained*).

The reformulation in (50) builds directly on strengths 1 and 2 in its first sentence: *Think of 100 women*, which invokes an imaginary 100-member collection of typical women.

This isn't, of course, the only possible way of introducing an imaginary collection; another would be to use a conditional sentence (with *if*):

(52) If 100 typical 40-year-old women were tested for breast cancer, 1 would actually have cancer and would probably test positive, but 9 of the others would probably also test positive, giving a total of 10 positive tests. How many of those who test positive actually have breast cancer?

Once again, the answer is obvious. Or we could reformulate in terms of a 'journey' (strength 5) through the women, using the word *almost* to signal an almost complete journey.

(53) Almost all 40-year-old women have no cancer and test negative when screened, but about 10% test positive, including almost 1% who actually have cancer. How many positive results indicate cancer?

Easy again. Or we could build on strengths 3 and 4, thinking in terms of typical cases and relative probability:

(54) A typical 40-year-old woman does not have breast cancer, and, when screened, tests negative; but about 10% of women test positive, though they typically don't have cancer. But exceptionally, 1% of all women test positive and actually have cancer. What are the chances that a woman who tests positive actually has breast cancer?

All of these reformulations are worth considering as more comprehensible alternatives to Gigerenzer's original version; and the key fact is that none of them mentions probability.

10.   Conclusions

This paper has focused on rather elementary areas of mathematics in order to compare them with the everyday language of children in a primary school or early secondary school. The purpose has been to ask how the semantics of ordinary language helps or hinders the teaching and learning of mathematics. I assume that expert mathematicians have evolved mental structures which allow them to think successfully about mathematical questions, so the question is how to help children who know no mathematics to move mentally towards the experts. The general principle must, as always, be to start where the children are, so the challenge for research is to work out exactly what this means. Where exactly are the children, in terms of linguistic knowledge and abilities? And what exactly are the special demands of mathematics? Since the mathematics is elementary, these questions demand more expertise in linguistics than in mathematics.

A linguistic analysis, paying attention to both meaning and grammar, leads to a number of relevant conclusions which may not be immediately obvious.

- Numerals, the building blocks of mathematics, have a complex semantics with five different types of meaning which have to be learned and consolidated one at a time. Moreover numbers of 10 or more are also suprisingly difficult to learn from the purely linguistic data. On the other hand, numbers with complex structures use both addition and multiplication. (Sections 2 and 3)
- Collections, brackets and variables are already part of the semantics of children's everyday language. (Sections 4 and 5)
- Negation arguably involves zero, but with a metaphorical journey that allows us to think in terms of 'hardly' and 'almost' as approximations to zero. Negation is inherently hard to think about. (Sections 6 and 7)
- Abstract scales and relationships are built into the semantics of language, including the meanings of =, > and <. (Section 8)
- Language is well adapted for recognising uncertainty and quantitative properties of collections, but fits poorly with probabilities. (Section 9)

In short, some important mathematical concepts such as 'base' and 'multiplication' are already part of a child's everyday language, but others are not, so the challenge for teaching is to build solid structures directly on the familiar concepts, and to keep building until the child is ready for the unfamiliar concepts. And on the way, of course, the teacher needs to beware of 'false friends' – words used in a mathematical way which also have other, and conflicting, uses in everyday speech.

What do these conclusions mean for teaching – for pedagogy, curriculum, school structures and the preparation of teachers? The most obvious conclusion is that maths teaching is intimately related to teaching of and about English. If a maths lesson needs to start (as I argue above) with the child's own language, then it starts as an English lesson; and if references need to be made along the way to links between language and maths, then it flips between English and maths. Fortunately, the way in which primary schools are organised guarantees that the same teacher is responsible for both subjects, so this principle doesn't present a problem; but it does mean that the teacher should be alert to cross-subject links.

Another conclusion is that primary teachers should be trained to recognise these links, which for most will take them into very unfamiliar territory. After all, a teacher whose own childhood training was divided cleanly into maths and English as separate subjects won't be used to

seeing them integrated in this way. And, of course, even without subject barriers the links themselves will not be obvious to most teachers, so they will need to be taught.

And finally, the arguments in this paper suggest the need for a radical change in the teaching of English, with far more emphasis on language analysis. For example, if the number system obscures the semantics of higher numbers, it would be helpful to start by looking, in class, at the morphology and syntax of English numerals in the everyday language of the children. But as things stand, this kind of analysis of everyday language is vanishingly rare in our classrooms, and very few primary teachers would face it with confidence.

References

Anon. 1974. Interactions between linguistics and mathematics. Final report of the symposium sponsored by UNESCO, CEDO and ICMI. Nairobi, Kenya, September 1-11, 1974. UNESCO. http://unesdoc.unesco.org/images/0001/000149/014932eb.pdf.

Chomsky, Noam. 2011. Language and Other Cognitive Systems. What Is Special About Language? Language Learning and Development 7. 263–278.

Duran Eppler, Eva. 2011. Emigranto. The syntax of German-English code-switching (Austrian Studies In English, Band 99. Herausgegeben von Manfred Markus, Herbert Schendl, Sabine Coelsch-Foisner). Vienna: Braumüller.

Easdown, E. 2006. Teaching mathematics: The gulf between semantics (meaning) and syntax (form). Proceedings of the 3rd International Conference on the Teaching of Mathematics at the Undergraduate Level. Istanbul: Turkish Mathematical Society. http://www.maths.usyd.edu.au/u/pubs/publist/preprints/2006/easdown-13.pdf.

Fillmore, Charles & Sue Atkins. 1992. Towards a frame-based lexicon: the semantics of RISK and its neighbours. In Adrienne Lehrer & Eva Kittay (eds.), Frames, Fields and Contrasts. New essays in semantic and lexical organisation., 75–102. Hillsdale, NJ: Erlbaum.

Fodor, Jerry. 1998. Concepts: Where cognitive science went wrong. New York: Oxford University Press.

Gigerenzer, Gerd. 2002. Reckoning with risk. Learning to live with uncertainty. London: Penguin.

Gisborne, Nikolas. 2010. The event structure of perception verbs. Oxford: Oxford University Press.

Glaister, Paul. 2019. The importance of language analysis to mathematical, quantitative and computational sciences. Languages, Society, Policy; also: Impact.

Goldberg, Adele. 1995. Constructions. A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press.

Hudson, Richard. 1984. Word Grammar. Oxford: Blackwell.

Hudson, Richard. 1990. English Word Grammar. Oxford: Blackwell.

Hudson, Richard. 1996. Sociolinguistics (Second edition). Cambridge: Cambridge University Press.

Hudson, Richard. 2007. Language networks: the new Word Grammar. Oxford: Oxford University Press.

Hudson, Richard. 2010. An Introduction to Word Grammar. Cambridge: Cambridge University Press.

Hudson, Richard. 2019. Towards a pedagogical linguistics. Pedagogical Linguistics 1. 8–33.

Hurford, James. 2011. The Linguistic Theory of Numerals (first published 1975). Cambridge: Cambridge University Press.

Labov, William. 1972. Negative contraction and negative concord in English grammar. Language 48. 773–818.

O'Halloran, K. 2005. Mathematical Discourse: Language, symbolism and visual images. London: Continuum.

Reisberg, Daniel. 2007. Cognition. Exploring the Science of the Mind. Third media edition. New York: Norton.

Rosch, Eleanor. 1973. Natural categories. Cognitive Psychology 4. 328–350.

Shanahan, Cynthia, Timothy Shanahan & Cynthia Misischia. 2011. Analysis of Expert Readers in Three Disciplines: History, Mathematics, and Chemistry. Journal of Literacy Research 43. 393– 429.

Sheldon, Neil. 2019. Statistics: insight not numbers. Languages, Society, Policy; also: Impact.

Smith, Jennifer, Mercedes Durham & Liane Fortune. 2007. "Mam, my trousers is fa'in doon!" Community, caregiver and child in the acquisition of variation in a Scottish dialect. Language Variation and Change 19. 63–99.

Traugott, Elizabeth & Graeme Trousdale. 2013. Constructionalization and constructional changes. Oxford: Oxford University Press.

Wason, Peter & Shuli Reich. 1979. A Verbal Illusion. Quarterly Journal of Experimental Psychology 31. 591–597.