

When numbers are dressed up to the nines: A short study of number-containing phrases in English

Dr Bertrand Richet

Abstract

Numbers are the product of our cultural and determinative interaction with reality. Within language, numbers are primarily found as grammaticalised forms of quantification and anchoring. But they are also regularly part of fixed phrases within which their purely numerical dimension sometimes disappear, loses its density or combines with a symbolical value which has historically been quite active. This study proposes to examine number-containing fixed phrases in English, determine the extension of the set and its recurrent characteristics. The question is indeed first that of outlines, as it depends on the definition one gives of fixed phrases, which are both detached from the rest of language and discourse and remarkably integrated within them. The focus will then turn on determining the preferred semantic areas covered by those idioms, which numbers are primarily found and how they interact with the rest of the phrase and surrounding discourse. The concluding remarks will be on the rhetorics of numbers. The study is based on a data set drawn from major English dictionaries as well as text and conversation corpora.

Résumé

Produits de notre interaction déterminative et culturelle avec le monde, les nombres se sont coulés dans la langue d'abord comme formes grammaticalisées de quantification et d'ancrage, puis au sein de locutions figées dans le cadre desquelles leur dimension proprement numérique a pu d'ailleurs disparaître, perdre de sa densité ou se combiner à une symbolique numérologique complémentaire très active historiquement. Dans le cadre de cette étude, nous nous proposons de dresser un état des lieux des expressions figées du nombre en anglais, d'en déterminer les contours et d'en mettre au jour les caractéristiques récurrentes. La question se pose d'abord des contours, précisément, de la définition à donner des locutions, tout à la fois

détachées du reste de la langue et du discours et intégrées à ceux-ci. Il s'agira ensuite de déterminer les lieux sémantiques privilégiés d'apparition des nombres, de leur nature préférentielle et de leur valence particulière dans les expressions et les discours qui les accueillent. Ceci nous amènera enfin à nous interroger sur la rhétorique du nombre. Notre étude se fondera sur les données fournies à la fois par les dictionnaires généraux et spécifiques, comme le *Brewer's Dictionary of Words and Fable*, et par des corpus de textes et de conversations.

Our new 3-door.
It looks like a million but costs \$995,005 less.
Advertisement for Hyundai Excel car, 1986

0. Introduction

The presence of numbers in language has always been a fascinating issue,¹ not least because numbers are themselves fascinating and have played a major role in the developing of trade-based human interaction (number words are thus said to be the comparatively fastest moving words since they have traveled the seas with merchants ever since antiquity, as shown by the indo-arabic origin of today's western figures²). Even though human beings are technically the same all over the surface of the earth, the extremely various ways in which each group has represented quantity over time tells us a lot about the cognitive and cultural complexity of that type of representation. It is true that there are many similarities between the different body-based systems, such as the universal existence of representations of very small quantities (*one, two, or three*) or the widespread use of base-ten quantification, but counting, just like language, is specific. It may even be more specific than language since even bilingual speakers tend to count preferably in one language only, the one they very intimately (in the strongest sense of the word) master, especially when it comes to silent counting.³

Number-containing phrases are in a way doubly-specific, since they combine the manner in which a given language has set its number system with the peculiarities of the corresponding culture to give birth to particular word combinations that have progressively moved away from “free

expressions” to become “formulas” (Jespersen [1924] 1992: 18). After a first study of the general presence of numbers in English language and discourse (Richet 2005) and another on number-containing words (Richet 2010), the idea is to explore this new territory and see whether what was observed at small-unit levels is also to be found at that more complex level. Beyond questioning the links between culture, cognition and language (basically, what kind of quantity can be or is represented?), one must explore the link between language and discourse. How do words become fixed phrases?

This study is based on a systematic gathering of relevant data in major British and American English-language dictionaries, the references of which are given at the end of this paper. First I propose a statistical examination of what is found in English, depending on number magnitude (from zero to the infinite) and number type (not just cardinal numbers, but also ordinal numbers and fractions). Then I move from figures to facts and a more semantic analysis of emerging patterns depending on number, focussing for the sake of brevity, on the first three numbers, which actually contain the most fundamental meanings of quantity. Finally I propose to come back to discourse and corpora and see how the idioms present in the list are represented in the Coca corpus.

1. Figures...

One’s intuition about the presence of numbers in fixed phrases is that only the most common numbers will be represented, *i.e.* numbers that have made their way into mainstream human experience rather than specialised mathematical forms such as decimals or square roots. Anyway, since more complex numbers are the result of a combination of simpler forms, it is possible to find them in the latter’s dictionary entries if they are indeed part of a fixed phrase.

It was decided to cover the whole spectrum, starting with cardinal numbers and going from zero to the infinite, including “odd” numbers such as *dozen* and *score*,⁴ continuing with ordinals and following the same path whilst distinguishing between ordinal and fractional usage. It is a well-known fact indeed that apart from *half* and *quarter* the form is the same (*third*, *fifth*, etc.) but the meaning is different though etymologically connected.

Typically only word combinations were taken into account and only those in which a figurative sense exists, though it may be difficult at times to draw a clear line between figurative and literal meanings and therefore between “fixed phrase” and “idiom” *per se*. The data was drawn from a variety of dictionaries, including the *Oxford English Dictionary*, the *Oxford Dictionary of Idioms*, *Brewer’s Dictionary of Phrase and Fable*, the *McGraw-Hill Dictionary of American Idioms*, the *Cambridge Dictionary of American Idioms*, etc. Register was taken into account but dialectal phrases and variations were left out, though they surely provide interesting additional data. The aim was to concentrate on mainstream language.

Data gathering proved rather difficult since the number word may not be the main entry to the phrase. For instance *to be three sheets to the wind* is potentially to be found at *three*, *sheet* and *wind* with redirections needed from one entry to the other but this is seldom the case. Additionally not all phrases are listed in entries, especially when they are numerous. Finding “*two* is used in many phrases such as [list], *etc.*” (my italics) is rather frustrating. It was decided, in order to expand the corpus, to use ready-made lists found on the internet’s many *idiom sites*⁵ and then move back to dictionaries for cross-reference confirmation.

Almost 550 different phrases have been found and included for analysis. *Table 1* (below) shows their distribution according to number type:

Table 1. Phrase distribution according to number type

Number type	Nb	%
<i>Cardinal</i>	397	72%
<i>Ordinal</i>	97	18%
<i>Fractional</i>	55	10%
Total	549	100%

Cardinal phrases make up almost three quarters of all fixed phrases, ordinal phrases just under a fifth and fractional phrases just a tenth. This is in line with one’s intuition about the connection between the cognitive complexity of non-cardinal numbers and their smaller statistical presence in language and discourse.⁶ Actually if one considers numbers in general in a corpus such as *Coca*, the difference between cardinal numbers and the rest is clear, as shown by *Table 2*, which includes data from the first ten cardinal, ordinal and fractional numbers. Ordinals and fractionals are put together since they share the same form, except for *second/half*.

Table 2. Coca occurrences of the first 10 numbers according to type

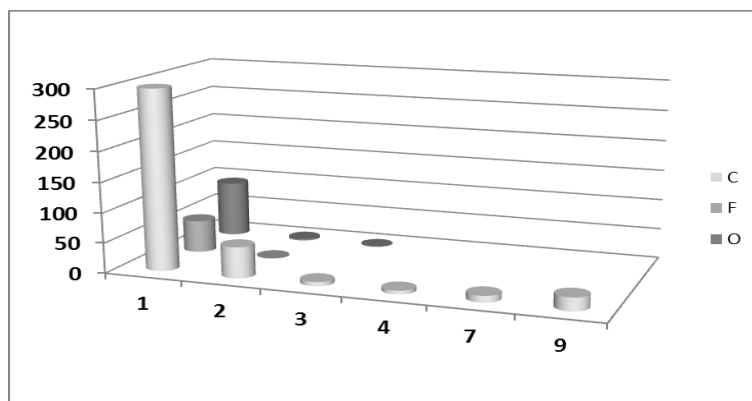
Cardinal	Ordinal / Fraction
2,145,755	852,903
72%	28%

Turning now to size, numbers have been grouped together according to their length. For practical reasons, a length number was attributed and it corresponds to the number of digits needed to express the quantity. For each category, an illustrative example of a phrase is given:

1. Units: From zero to nine (*a one-day wonder*)
2. Tens: From ten to ninety-nine (*a baker's dozen*)
3. Hundreds: From 100 to 999 (*The first hundred years are the hardest*)
4. Thousands: From 1,000 to 999,999 (*death by a thousand cuts*)
7. Millions: From 1,000,000 to the infinite (*a million-dollar wound*)
9. For phrases including words like "Number" (*There's safety in numbers*)

The results are shown in Chart 3:

Chart 3. Occurrences of phrases according to number length and type



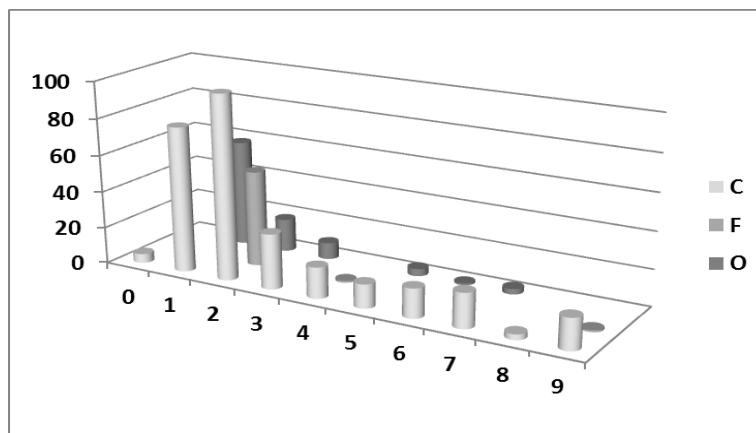
Units, *i.e.* phrases containing numbers from zero to nine, clearly make up most of the stock of fixed phrases in the English language. Counting typically ends when there are no more fingers available.

Tens are almost exclusively represented by cardinals. Of the 56 cardinal phrases, 14 (just a quarter) are based on *ten* and nine on the other traditional base *twelve*, which means that two fifths of a set potentially including ninety numbers (from *ten* to *ninety-nine*) make use of only the two most simple numbers. The appeal of round numbers is once more confirmed, though the actual definition of the roundness of a number is more complex than it seems, as shown by Sigurd (1988).

Actually, only units and, to a lesser extent, tens are represented in all number types (cardinal, ordinal, and fractional) and only cardinals are represented over the whole spectrum. Fractions quickly disappear because the quantity they represent becomes meaningless while ordinals keep on to the three-figure category⁷ but only marginally with the phrase mentioned above (*The first hundred years are the hardest*). Ordering elements is likewise quite meaningless after ten, except in very special circumstances (after a competition for example) that are easily filtered out by the phrase fixation process.

Such data lead me to redirect my attention solely on the 446 units of the corpus (81% of the total of 549 phrases). Results are shown in *Chart 4*.

Chart 4. Occurrences of phrases according to unit and number type



The situation is quite similar to that of the whole spectrum. Small quantities abound while larger numbers are much less represented. The

combination of the three basic numbers 1, 2, and 3 accounts for 78% of units (346 tokens out of a total of 446), which is on a par with the general pattern of numbers in discourse (80% of the nearly 3 million tokens).⁸

Still, phrases include all cardinal numbers while ordinals and especially fractions are more limited. As for fractions, only *half* is really represented, with 52 of the 54 fraction-based phrases. Its prominence can be explained by the fact that this is the simplest fraction available and that it does not necessarily refer to an exact division but to any kind of non-complete portion of the whole, except for extremes (near absence and near completion). It can also be applied to objects that are normally not easily quantifiable (*half-dead, too clever by half*, etc.). That explains why fractions are comparatively numerous for 2, more numerous for that matter than the ordinal *second*, which is semantically and statistically “dwarfed” by *first*. The latter is the only non-cardinal number which is nearly on an equal footing with cardinals as there are 58 phrases with *first* and 79 with *one*. Just as *half* emerges thanks to its extensive meaning, *first* comes out because of its qualitative and quantitative prominence and more generally because of time-induced unidimensionality and the derived importance of protocol. Conversely there is not much remaining after the “eternal” gold / silver / bronze trilogy.

Though it does not appear in *Chart 4*, another characteristic is worth mentioning. 108 of all phrases (20%) are full sentences, *i.e.* proverbs, which are equally present in cardinals, ordinals and fractions. Both features do not come as a surprise since proverbs can be seen as the ultimate form of phrase fixation, which is not dependent on the specific nature of the words assembled. Still, the number is comparatively high since there is a total of “only” about 1,100 proverbs in English (as found in the *Oxford Dictionary of Proverbs*), which means that 10% of all proverbs contain a number. This is compatible with the notion that proverbs provide simple rules of life based upon perceived patterns, in which numbers have a structural role. This also accounts for the nature of numbers found: 98 proverbs (91%) contain only numbers up to three, with an internal distribution of 44 proverbs with *one* (among which 18 with *first*), 34 with *two* (among which 10 with *half*) and 9 with *three*.

2. And Facts...

2.0. Introduction

The data presented above leads one to concentrate on the area where most phrases are to be found, *i.e.* the *one two three* trilogy. The question arises not so much of whether those three numbers are over-represented (since the number words themselves are also more present in discourse) than of *how* they are represented. In other words what is there in the meaning of *one*, *two*, and *three* that makes those numbers phrase-friendly beyond being human-friendly?

Their semantic specialisation will be first addressed and it will be shown that they efficiently enable a clear representation of the world and the speaker/hearer/world relationship. Then attention will turn to two different forms of combination, one implying vertical number variation around the same phrasal core, the other a linear association of two numbers which brings out into the open the connections between the two. Since there are so many phrases available, the full corpus cannot be presented here.

2.1. *One*: Unity and beyond

One has three fundamental values that are expectedly connected. Its first value lies within itself as it represents an entity, an object. Its second value is linked with that of *zero*: *one* is just a step away from zero, which can be considered positively as being *just above zero*, or negatively as being either *just above zero* in a downward movement or *just under completion* with only one element needed to complete a given set. Finally, one should also take into account its relation with plurality, starting with binary systems and evolving into more complex series of elements.

This boils down to saying that *one* is between *zero* and *two* but surely the implications go far beyond the perception of the mathematical sequence. Though the difference between *one*, *two* and *three* is the same as between *seventy-five*, *seventy-six*, and *seventy-seven*,⁹ it is clear that the latter are only sequentially related while the former are associated by a network of various symbolic connections which paradoxically gives them more individual value.

Let us first consider the *entity value*. *One* implies that we have something and that something is fundamentally self-explanatory: it is because it is. It allows to explain phrases like *one and only* or *one and the same*, which insist on the unicity of the referent. The connection with narcissism is easy, as

shown by *to take care about (or look after) number one*. This centralised vision also implies a form of hierarchy that both cardinal and ordinal phrases embody, as in *public enemy number one, first and foremost* or the peculiar *the firstest with the mostest*, which initiate potential series.

The *next-to-zero value* is connected to the minimal positive quantity that *one* symbolises. It is there and just a step away from not being there, the solitary figure thanks to which honour is saved or the much-resented fly in the ointment. From frequency (*once in a (blue) moon, a while, upon a time*) to precedence (*one step ahead of*) or minimally defined quality and quantity (*one way or another, every dog is allowed one bite*), *one* supports a reality that does not need much to disappear (*From the sublime to the ridiculous is only one step*) or that lacks something to be complete (*one sandwich short of a picnic*). It explains its regular use in negative statements and associated phrases (*not one iota, not to know the first thing about*) as well as in unexpectedly low-quantity phrases (*a one-day/-hit wonder, a one-horse race/town, a one-night stand, a one-trick pony*) and proverbs (*Don't put all your eggs in one basket, A bird never flew on one wing, In the country of the blind, the one-eyed man is king, There is more than one way to skin a cat*).

Finally, the *one-and-the-rest value* gives *one* a special status as a salient element of a series and what it represents is compared to those other elements. As can be expected proverbs are well-represented there (35 out of a total of 48 phrases) since they allow more complex linguistic forms than phrases and since such combinations imply comparison and therefore typical proverb matter such as the expression of popular thought. The saliency can be cardinal (*one of these/those days*) or ordinal (*at first glance/sight/hand/light, First impressions are the most lasting, First thoughts are best*) and has a one-contains-all flavour to it, as also exemplified by *year one, from day one* or *Once a fool, always a fool, If I have not told you once, I have told you a hundred times*, with a meaning of sequence-initiation (*One funeral/wedding makes many, One good turn deserves another, first of all*), continuation, either of open series (*one after another, one at a time, one by one, one step at a time*) or binary systems (*One hand washes the other, on the one hand, One nail drives out another, One thing leads to another, When one door shuts, another opens*) or finally sequence-ending (*one for the road*).

In a way, *one* can be used to represent any other quantity or quality since it *symbolises* quantity or quality. Just as in the French proverb *Qui vole un oeuf vole un boeuf* [He that steals an egg will steal an ox], quite a number of

British proverbs make use of equivalence with *one* associated to the locator against which another element is judged. Quality is expressed with *worth* (*One hour's sleep before midnight is worth two after, One volunteer is worth two pressed men*), *better* (*Better one house spoiled than two, Better to live one day as a tiger than a thousand years as a sheep*) or other forms (*One Englishman can beat three Frenchmen, One year's seeding makes seven years' weeding, Once a fool [and variants], always a fool*), with equivalence (*If I've told you once, I've told you a hundred times*), preference or sequence meanings, as in the card-dealing type *One for sorrow, two for mirth; three for a wedding, four for a birth*.

Actually such a binary pattern extends into the “over-One” domain (*Fall down seven times, get up eight, He wears a ten-dollar hat on a five-cent head, Six hours' sleep for a man, seven for a woman, and eight for a fool*). The rest is made up of more random combinations, such as *to be at sixes and sevens, Genius is ten per cent inspiration and ninety per cent perspiration or to talk [and variants] nineteen to the dozen*.

2.2. *Two*: The birth of a system

Just as was the case with *one*, *two* can be considered in three ways, *i.e.* for itself as in binary systems, in connection with *one*, which involves a gap, and compared to larger numbers. Binary systems are extremely common in our experience (bipolarity, causality, space association, *etc.*). In other words, *two* is often expected either because phrases are based on naturally binary reality (*to have eyes like two burnt holes in a blanket, to jump in with both feet, to stand on one's two feet*), couple-based systems (*It takes two to tango/be stupid/make a bargain/a quarrel, A game that two can play, That makes two of us*), some of which are quite culture-connected (*blues and twos, meat and two veg*), or proximal associations (*There are two sides to every question, to be two sides of the same coin, close as two coats of paint, like two peas in a pod, as cross as two sticks*). This extends to some fraction-based phrases, with an upper-level unity made up of two elements (*How the other half lives, To go halves*).

The connection with *one* depends on the status of *two* in phrases (as a cardinal, *two*, an ordinal, *second*, or a fraction, *half*) and whether *one* or *two* is prototypically expected.

As a cardinal it can signify that *one* is more appropriate, due to a radical impossibility of having *two* instead of *one* (as in *If you run after two hares, you will catch neither, No man can serve two masters, One cannot be in two places at once*), a sheer improbability (*Lighting never strikes the same place twice, Opportunity never knocks twice at any man's door*) or a lack of comfort connected to precedence, as in *A bird in the hand is worth two in the bush* or *If two ride on a horse, one must ride behind*, forced parallelism as in *to be of two minds, to be like a dog with two tails, to have two left feet*, or position (*between two fires, to fall between two stools*).¹⁰

Conversely there are many cases in which *two* is considered more appropriate even though *one* is primarily expected. This grants expansion of power (*to have two strings to one's bow, to kill two birds with one stone, to wear more than one hat*), which can be more ambiguous (*a two-edged sword*), reason (*to think twice, Two heads are better than one, Measure twice, cut once*) or possibilities (*a two-way street, to be two-fisted*). Quite interestingly the combination of two different activities, which is normally doomed to failure, may also be what is expected in some circumstances (*If you can't ride two horses at once, you shouldn't be in the circus*).

As an ordinal, position is essential and can be considered vertically or horizontally, *i.e.* with a hierarchy or a sequence in mind. Hierarchy entails a negative reading (*a second banana, a second-class citizen, to play second fiddle, to come off second best*) while sequence is more diverse, depending on the element it is associated with. Having *second thoughts* is a sign of caution and can therefore be viewed positively or negatively depending on how caution itself is regarded, in the same way as the meaning of the verb *second-guess*. *Second nature* also depends on what is indeed a second nature to someone. *Second childhood* is normally negative while *Second wind* is positive in a cycle-based representation of life.

When *two* is a fraction (*half*), there can be no definite univocal reading since 50%, which should normally be neutral, as in the case of sharing (*to go halves*), will typically be viewed positively or negatively depending on which side the speaker primarily considers, as exemplified by *seeing the glass half empty* or *half full*, irrespective of the nature of associated words. *Half* is clearly negative when considering a downward movement (*at half-mast, to be half the [dancer etc.] sb used to be, to be half-dead*) or a quantity lower than expected (*half an ear, an eye, half-baked, half-hearted, a half-wit, not to know the half of it*). Conversely, an upward movement and/or starting from zero is viewed positively (*A fault confessed is half redressed*,

[Doing sth] is half the fun/half the battle, Half a loaf is better than none/than no bread, Well begun is half done), to the point that in some cases *The half is better than the whole*, which reminds one of other phrases favouring small, manageable quantities to the detriment of larger groups, such as *Two boys are half a boy, and three boys are no boy at all, Three may keep a secret, if two of them are dead, or Two is company, but three is a crowd*.

The latter phrase brings us to a final development about *two*, when it is used with implicit or explicit reference to larger numbers. Whatever linguistically and culturally magnified differences may exist between *one* and *two*, the latter remains a small number which can basically correspond to *few* or *a few*. The *Few* interpretation is shown in *for two cents/pins one would [do sth], in two shakes (of a lamb's tail)/twos/ups or to put in (one's) two cents/penn'orth* while the *A few* interpretation is found in *One/Two combinations (to be short of a bob or two, to know a thing/a trick or two, to bring down a peg or two)*.

2.3. *Three*: From system to series

Unlike *two*, which is based on simple relation, *three* introduces the first complex system, which is completed by *four*, *seven*, *ten* and other intermediary bases. As shown in the data part, cardinals are by now the main source for phrases, since fractions are too difficult to easily conceptualise and therefore have restricted access to language sedimentation. Ordinals are present, but mainly with a horizontal, sequence meaning instead of a vertical, hierarchy use.

Three refers to a quantity above *two*. This sounds quite obvious but the implications are essential. It goes beyond the simple binary system that *two* represents. Even the very British *page three girl* owes her existence to the magic of three and the suspense of having to turn over the front page of tabloids and the reassuring fact of not having to twist one's head to the left in order to "read" page two. In other words, *Three* is more than *two* as well as a solid base for more complex representations, which accounts for an "Excess" reading as well as a "Pattern" reading of the associated fixed phrases.

Excess implies abundance (*a three-martini lunch, a three-time loser*, or conversely *say no more than three words with someone*), complexity (*a three-ring circus, a three-pipe problem, as queer as a three-dollar bill*), or intemperance (*a three-bottle*) with a possibility of negative connotations,

with quite a number of ordinal phrases there as ordinals memorise what comes before and can therefore suggest excess more easily (*a three-letter man, a third rail, a third tongue, a third wheel, to get the third degree*).

But *three* is also a practical means to stabilise reality¹¹, as in the notorious *Three trees* or *three-legged mare* (*to climb the three trees with a ladder*). *Three* suggests completeness and is extensively used for such purpose in “normal” discourse. Its stability makes it a good candidate for formulaic recipes (*the rule of three, the three Rs*) or forms of contentment (*three acres and a cow, three squares a day, three cheers*).

As it also goes beyond the simple one-to-one relation, it works well with proverbial cycles (*Bad things come in threes, From clogs to clogs is only three generations, It takes three generations to make a gentleman*). With *three* a series reaches an end or a new start, with both cardinals (*Three removals is as bad as a fire, Fish and guests smell after three days*) and ordinals (*third time lucky, third time's the charm, The third time pays for all*).

2.4. *Four* and beyond: Playing with bases

All of the numbers between 0 and 10 are represented in fixed phrases, as well as most of those between 10 and 20, to the exception of 14, 16 and 18. Still, of the first 100 numbers, only 30 are included in phrases, 25 of which are between 1 and 50. Each of the smaller numbers tends to have a specific value even though the number of phrases that each “produces” is variable, including outside the *one-two-three* trilogy. Still, none gives birth to more than 10 phrases. Three are rather productive numbers (*twelve, forty* and *one hundred*) and they are all related to two bases.

The first is dozenal, a base historically connected to planetary movement and time calculation and wildly used in daily trade, and phrases indeed contain references to quantity (*by the dozen, to be a dime a dozen*), abundance (*It is not spring until you can plant your foot upon twelve daisies, to put in the dozens*) or completeness, as can be expected when dealing with a base (*daily dozen, to strike twelve the first time, twelve good men and true*).

The second is decimal, with variations. *Forty* is originally found in the Scriptures, in association with time (days, as in the Deluge, or years, representing a generation). It consequently spread in western culture, with reference to a large quantity suggesting multitude and/or completeness (A

fool at forty is a fool indeed, like forty, to catch forty winks) and possibly a new start (*Life begins at forty*) since a first stage has been completed.

Some phrases include next-to-(cultural)-base numbers, either just under (*be on cloud nine, dress up to the nines, sweet seventeen*, with *eighteen* as the threshold to adulthood) or just over (*over the eight* with underlying reference to perfect *seven, the nineteenth hole, or a thousand and one*, all suggesting excess), while others seem to have appeared more randomly, including *twenty-three skidoo, seventy-three* (“goodbye” and variants in telegraph Morse code) or *to eighty-six something*.

2.5. Gradation

The apparently paradoxical question could be: how fixed is a fixed phrase? As far as number-containing phrases are concerned, the answer very much depends on the nature of the combination. As seen above, especially in the parts devoted to *one, two, and three* phrases, the role of the number is often quite specialised so that changing the number is not an option, either because the result would be meaningless (**It takes three to tango, *Five eyes see more than three*) or because it would imply a potentially extensive reorganisation of implications (*?The second duty of a soldier is obedience*). Still, quite a number of phrases might accept often jocular context-driven variations like *Second come, second served* or *A ten Martini lunch*.

There are basically two possibilities for non-problematic number variation: gradation and fuzziness.

Gradation implies that the number clearly expresses meaningful quantity (more meaningful than *Cloud nine*, though the number in the latter was not chosen at random). To exemplify what is meant four phrases have been selected for a more in-depth analysis. Two include *one* as the standard number (*a one-horse town* and *to be one brick short of a load*) and two include *three* as the standard number (*to be three sheets in the wind* and *to be a three-time loser*).

The choice of *one* is justified either by the fact that it corresponds to a necessary minimum (a historically typical Wild West town must have at least one horse) or by the symbolical value attached to *one* (*one* is enough to block completeness, as shown previously). The quite interesting thing is that though the standard number makes up a majority of tokens in a simple Google search (82% for *one-horse town* and 63% for *one brick short of a*

load), there are many variants. *Two-horse* and *Three-horse towns* make up most of what remains and retain both the small quantity meaning and a form of autonomy. Moving up the spectrum other more humorous uses are found, with an often explicit play on the original phrase:

- (1) *I'm bustin' out of this one horse town this weekend to go to Calgary (a six or seven horse town at least!) to see Oasis and Ryan Adams live!*¹²
- (2) *I know I live in a one horse town...er more like a one thousand horse town to be accurate, but that's not the saying now is it... drawback to small town is often having to drive to larger town for anything but the basics.*¹³

As for *bricks* there are very few tokens of variants over *three*, showing how powerful the symbolical value of a small number is. Still there are instances of large-scale numbers to refer to really dumb specimens:

- (3) *Well, talking to Murphy and David Foss is pointless, but I will bow to the will of the majority. Am I a humorless bastarrd as John Murphy asserts? Should I be banished to the outer darkness? Or is John Murphy evidently a couple hundred bricks short of a load?*¹⁴
- (4) *Does anyone know if it's actually "illegal" to be hiking in the field and forested area of Shannon Lake, in the trails opposite the Talus Ridge developments? I park up the new road and then hike my dogs from there, and the security guard (who is a few thousand bricks short of a load) keeps telling me that I'm not allowed back there... Luckily every day that he sees me he thinks it's the first time we've met, so I've only gotten warnings!*¹⁵

The possibility for exaggeration is also to be found in the *three*-based phrases. *Be three sheets in/to the wind* has two openings, one with atonement (*one sheet in the wind* if you are just tipsy), the other with intensification:

- (5) *Paul Revere had a drinking problem, in fact he was about ten sheets in the wind when he made that historic ride across the countryside. I wish I could have been there to see what it was really like. He must have been talking like that drunk Senator talking about healthcare.*¹⁶

The number may even go beyond a ship's possibilities:

- (6) *Now they say I got insane / A thousand sheets in the wind.*¹⁷

Be a three-time loser, which refers to the number of times somebody has been put in jail, allows likewise expansions though it is difficult to sort out the actual usage that is made of the phrase. One is typical of fixed phrases, with at least partially figurative meaning, while two others involve accurate counting even with larger numbers (sport commentary and court proceedings), which explains why *all* the numbers between 1 and 20 are to be found.

Fuzziness is typically connected to the expression of infinity. So it is less about “objective” intensification as is the case with gradation than about the various ways in which to represent a large quantity. Just as French allows the use of 36, 40, 50, 107, 1000, 1001, 36,000, 300,000, 1,000,000 and more context-driven or idiosyncratic variants to refer to a non-finite high number of items, English has many numbers to do so. A test on the Coca corpus with the phrase *x ways* starting from 24 shows tokens with 28, 37, 40, 41, 43, 44, 45, 50, 60, 85, 100, 163, 200, 500, 1,000, 2,000, 10,000 and 1,000,000 at least. Most appear only once, except 40, 50, 100, 1,000 and 1,000,000. The pattern is fixed but the surface quantifier is allowed some freedom within a given range, that is, little variation in meaning and a communicatively acceptable form, which therefore makes 347,962 an unlikely candidate.

3. Talking numbers

3.0. Methodological introduction

It is now time to turn to discourse and see whether all those 550 fixed phrases that have been compiled in dictionaries and other reference books deserve to be there, that is to say whether this is supported by presence in discourse. There are two difficulties though when it comes to moving from qualitative to quantitative analysis.

The first is a question of time. Dictionaries surely record usage but they are also about memory and some phrases can be included without being normally used today. Moreover not all freely-available corpora are adequate chronologically speaking.

The second is a question of usage. The original idea was to have a look at it on both sides of the Atlantic ocean, with the British BNC (100 million words

in 1980 to 1993 texts) and the American COCA (410+ million words in 1990 to 2010 texts) in order to account for dialectal variation. Not only do dates differ but the size and (partially) the constitution of the two corpora are not entirely balanced. Moreover, while fixed phrases tend to occur more frequently in spoken discourse, the corpora are mainly based on written data, all of this being sources of distortion.

Still, for practical reasons, only the more recent Coca data will be examined but one must bear in mind the limitations that have just been stated.

3.1. Statistical presence

The difficulty with phrases is that in many cases they resemble “normal speech”, which means the presence of form variation (especially for verbs) and a potential difficulty to draw the line between literal and figurative meanings. Searching in the Corpus of Contemporary American English has therefore meant a two-phase tailored procedure, first by deciding how open the search equation would be, then by sorting out the results. Coca offers the possibility of searching lemmas rather than words, *i.e.* all the forms of the same word. [be] therefore stands for “be”, “is”, “are”, etc. This is useful for semi-fixed phrases but not for proverbs for instance, which should remain monomorphic. In other cases, several lexical forms were tested for the same phrasal core (*at first glance*, *at first sight*). Results were then reviewed to eliminate non-figurative occurrences. That was when another difficulty cropped up: for some common combinations like *to do one, for two cents* or *take five* it was virtually impossible to distinguish without a really broad context between literal and figurative meanings and those phrases were finally excluded. Results are shown in *Table 5* just below. Excluded phrases, which make up just over one tenth of the total, are coded XXX:

Table 5. Number of phrases according to number of Coca occurrences

Tokens	Phrases	%
0	177	32%
1-4	98	17%
5-9	29	5%
10-19	28	5%
20-99	99	18%
100-999	48	8%
1,000-9,999	10	2%

10,000 and over	1	0%
XXX	59	11%
Total	549	100%

A third of all phrases do not appear at all in the corpus and a sixth appear less than 5 times, which means that half are either not or very little represented in the type of discourse to be found in the corpus. At the other end of the spectrum only 2% appear more than a thousand times.

Complete or almost complete absence is apparently surprising considering that fixed phrases are such because of recurrent usage and that a corpus is supposed to provide an accurate image of usage, with both spoken and written language. Two elements must be taken into account, though. One is date and the other is spontaneity. The corpus is contemporary, spanning over the last twenty years. Many phrases typically entered language long ago and may have lost some of their edge in the process. Actually nearly 40% of absent phrases are proverbs and 60% of all number-containing proverbs found in dictionaries are not recorded in the corpus. Many are presented as old-fashioned in the reference books (*Fools and bairns should never see half-done work* or *He that would the daughter win, must with the mother first begin*). Moreover their cumbersome form makes them difficult to integrate in discourse, which leads us to the second parameter, spontaneity. In a way the corpus is too modern and spontaneous to easily accommodate proverbs and other old-fashioned phrases (*six of the best, by the elevens*) and not enough so to accept other more slangish or specialised types of phrases (*a bunch of fives, to have eyes like two burnt holes in a blanket*). Dialectal specialisation is another matter, with many typically British phrases that are logically absent from the corpus (*A bird never flew on one wing, page-three girl, within the four seas*).

Let us now turn to the 22% of phrases with less than 10 tokens. The situation varies much between *one-* or *two-*phrases and *three-* or *four-*phrases, all of which represent two thirds of our corpus. A third of the former belong to the *Rare* category against half of the latter, *i.e.* not only are phrases containing larger numerals fewer in dictionaries but they tend to be less present in discourse, a case similar to the chicken and egg problem, though it is easier to find a way out: since it is more difficult to conceptualise larger numbers, they are less used in discourse, so they are less likely to be part of a phrase which itself is less likely to be used in discourse.

Of the 11 phrases appearing more than 1,000 times in the corpus none belongs to the *three* group or beyond. Apart from the phrasal verb *to zero in*

and *to think twice* (just around 1,000 tokens for both) all massively-present phrases belong to the *one* group. More specifically the top three group is made up of ordinal phrases, with *first of all* (12,142), *in the first place* (7,300) and *first and foremost* (1,599), showing how much hierarchy and precedence matter in our world.

3.2. Causes and consequences: Concluding remarks

The first conclusion to be drawn from the preceding results is that this is a low-number world displaying a wide variety of simple systems. Simple systems can indeed be varied because they help unearth basic relations between elements or sets of elements. That is the case for *two* phrases or binary-pattern phrases with *one* compared to something else. The simplest of all is of course unity, since it allows to embody presence (instead of absence), unicity (instead of plurality), precedence, sequentially or vertically (instead of equality). This low-number world gives rise to low-number phrases, many of them are used to accompany every aspect of that multiplicity of references. In turn the phrases are more often used than not, as they naturally correspond to speakers' experience of the world and their conceptualising ability. Conversely more complex forms of representation are less often found, fixed and used and they tend to aggregate around bases (10, 12, 20, etc.), *i.e.* larger forms of small numbers.

More generally speaking the question arises of the reason for the presence of numbers in fixed phrases and how relevant their presence is. One easy answer could be that they are found in phrases just because they are also found more generally in language and discourse. However that is only shifting the problem without really tackling it. There can be in fact several answers.

One is connected to what can be called the *magnitude effect*, which is especially felt with mainly culture-determined bases (7, 9, 12, 40, 64, etc.) and explains the presence of relatively large-number-containing phrases. The number is there to impress thanks to the combination of magnitude and harmony, a perfect representation of a complex world. The effect can be simpler, with gradable phrases in which the number plays the role of an intensifier.

Another effect is the *short-cut effect*. Numbers allow quick definition (what is within limits or outside them), identification (number are used as a

coded representation of reality) and stabilisation, that is to say the integration of an object or series of objects within a set. This is especially true of small-number-containing phrases showing reality as it is, or as it should be.¹⁸ The definitive dimension of numbers leaves one with no other option than accepting what is said as fact. In a way numbers are a confirmation, in the truest sense of the word, of what phrases achieve, the crystallisation of experience through existing patterns.

Notes

1. This does not mean that they have greatly retained the attention of grammarians. The singular *vs* plural distinction has but numerals have largely remained on the margins of linguists' interest, even long after Albrecht Goetze deplored "the sketchy way in which numerals are treated in the current grammars" (Goetze 1946: 186).
2. See Ifrah (2000) for a complete presentation of the journey of numbers.
3. Mackey (1968: 557) establishes a difference between the external functions of language, when a speaker is in direct contact with others, and its internal functions, such as when counting, praying, cursing or dreaming, in which case there is no social constraint. Moreover this is linked to the language in which the mental operations involved were learned at school, as shown by Grosjean (1978: 275). Repeated school exposure to fixed computing algorithms leaves a strong imprint on subjects.
4. The vigesimal system of counting (with 20 as a base) exists in many languages. Indo-european languages opted for a decimal system with a few vigesimal elements, such as *score* in English and *quatre-vingts* in French. For a discussion, see Nikl (1926).
5. Here are a few sites
<http://www.usingenglish.com/reference/idioms/>
<http://www.idiomconnection.com/>
<http://idioms.thefreedictionary.com/>
6. Though it is true that fractions are more complex than cardinals they are used in discourse, with competition from percentages for the "share" meaning and decimals for the "non-integer" meaning. Back in the trade world of the Middle Ages fractions were omnipresent and there were books specially designed for merchants to master them. See Benoit *et al.* (1992) for a fascinating history of fractions.
7. This is typically the case in sports with the final podiums. For a comprehensive study of podium ceremonies at the Olympics, see Barney (1998).
8. This is in keeping with results of a previous study. See Richet (2005: 144–146).

9. Interestingly enough, each of those three numbers has a specific value: three-quarters for *seventy-five*, a reference to a U.S. truck stop chain (*Union 76*) for the second or to the famous 1976 drought in Western Europe for the second and the upper Tintin-reading age limit for the third. Many other cultural references can certainly be found...
10. The spectrum of uses goes beyond Abrudeanu's presentation though the latter looks beautifully clear and convincing : "Le déterminant le plus fréquent dans les phraséologismes est *deux*, qui exprime l'idée d'oscillation ou d'indécision [The most frequent determiner in fixed phrases is *two*, which means oscillation or indecision]" (Abrudeanu 2001: 7).
11. Surely men *are* stable though only bipeds but this is achieved thanks to complex balancing procedures. Tripods do not need that extra brain processing.
12. From Craig Silliphant's blog on films and other cultural products.
(<http://www.newstalk650.com/blogs/craig-silliphant/all?page=15>).
13. From a review of Pet World, a pet shop located in Hollister, in the San Francisco area. The writer lives in the same town (<http://www.yelp.com/biz/pet-world-hollister>).
14. Rather inflamed discussion on a forum about baseball, namely the Baseball Think Factory's Hall of Merit [A Look at Baseball's All-Time Best]. John Murphy and David Foss are participants.
(http://www.baseballthinkfactory.org/files/hall_of_merit/discussion/biz_mackey/P200/)
15. Found on the Okanagan's discussion forum. This is the initial post of a thread called "Hiking up Shannon Lake".
(<http://forums.castanet.net/viewtopic.php?f=35&t=8357&start=0>).
16. Discussion on a forum thread initiated by a clearly Conservative webmaster posting a quotation by George Washington.
(<http://forums.augusta.com/viewtopic.php?p=205495&sid=a36523a1b70d14642096ee14a99c61ce>)
17. Punk band The Strap-Ons' "Replace Me" lyrics.
18. The difference explains some of the difficulties historians have in analysing centuries-old texts. One example is given by Moses 1996 and concerns the *Secret History of the Mongols*, the oldest surviving Mongolian-language literary work, written in the thirteenth century. In the text, "the expression of quantities ... has more to do with symbolism than with historical veracity" (Moses 1996: 75), that is, numbers are used for mnemonic reasons or to impress listeners in that oral tradition.

Dictionaries

Brewer's *Dictionary of Phrase and Fable*, 18th edition, Edinburgh: Chambers Harrap Publishers, (1870) 2009.

- Cambridge Dictionary of American Idioms* (edited by Paul Heacock). Cambridge: Cambridge University Press, 2003.
- Cambridge Idioms Dictionary* (edited by Elisabeth Walter), second edition, Cambridge: Cambridge University Press, (1998) 2006.
- McGraw-Hill Dictionary of American Idioms and Phrasal Verbs* (edited by Richard A. Spears), 2006, New York: The McGraw-Hill Companies.
- Oxford Dictionary of English Idioms* (edited by John Ayto), Third Edition, Oxford: Oxford University Press, (1999) 2009.
- Oxford Dictionary of Proverbs* (edited by John Simpson and Jennifer Speake), Fifth Edition, Oxford: Oxford University Press, (1982) 2008.
- Oxford English Dictionary* online edition (oed.com)

References

- Abrudeanu, Clara
 2001 La détermination dans la phraséologie – Etude comparée du français et du roumain [*Determination in fixed phrases*] in *Détermination et formalisation* [*Determination and formalisation*], Xavier Blanco, Pierre-André Buvet, and Zoé Gavriilidou (eds), 1–12. (*Linguisticae Investigationes Supplementa* 23.) Amsterdam: Benjamins.
- Barney, Robert K.
 1998 The Great Transformation: Olympic Victory Ceremonies and the Medal Podium. *Olympika* 7: 89–112.
- Benoît, Paul, Chemla Karine, and Jim Ritter
 1992 *Histoire de fractions, fractions d'histoire* [From the history of fractions to fractions of history]. Basel/Boston/Berlin: Birkhäuser (*Science network historical studies* 10).
- Goetze, Albrecht
 1946 Number Idioms in Old Babylonian. *Journal of Near Eastern Studies* 5 (3): 185–202.
- Grosjean, François
 2001 *Life with two Languages: An Introduction to Bilingualism*. 11th reprint. Harvard: Harvard University Press. Original edition: 1982.
- Ifrah, Georges
 2000 *The Universal History of Numbers: From Prehistory to the Invention of the Computer, Volume 1*. Hoboken (New Jersey): Wiley. Original edition: Paris: Robert Laffont, 1994.
- Jespersen, Otto
 1992 *The Philosophy of Grammar*. Chicago: University of Chicago Press. Original edition: 1924.
- Mackey, William
 1968 The Description of Bilingualism in *Readings in the Sociology of Language*, J.A. Fishman (ed.), 554–584. La Haye/Paris: Mouton.

- Moses, Larry
 1996 Legends by the Numbers: The Symbolism of Numbers in ‘*The Secret History of the Mongols*’. *Asian Folklore Studies* 55 (1): 73–97.
- Nykl, A.R.
 1926 The Quinary-Vigesimal System of Counting in Europe, Asia, and America. *Language* 2 (3): 165–173.
- Richet, Bertrand
 2005 Des chiffres et des lettres : expression(s) du nombre en anglais contemporain [*Numbers and words: expression(s) of number in contemporary English*] in *Six études de linguistique*, Claude Delmas and Mireille Quivy (eds), 141–164. (*Cercles*, The Occasional Papers Series, 2).
<http://cercles.com/occasional/ops2-2005/richet.pdf>
- 2010 Des nombres en noms : représentation du nombre en anglais et en français [*Numbers in words: the representation of number in English and French lexicon*] in *Res Per Nomen II: Langue, référence et anthropologie* [*Res Per Nomen II: Language, reference, and anthropology*], Pierre Frath, Laure Lansari, and Jean Pauchard (eds), 387–401. Reims: Épure.
- Sigurd, Bengt
 1988 Round Numbers. *Language in Society* 17 (2): 243–252.