A mathematical model of historical semantics and the grouping of word-meanings into concepts

Martin C. Cooper
IRIT
University of Toulouse III
118 route de Narbonne, 31062 Toulouse, France

cooper@irit.fr

Abstract

A statistical analysis of polysemy in sixteen English and French dictionaries has revealed that, in each dictionary, the number of senses per word has a near-exponential distribution. A probabilistic model of historical semantics is presented which explains this distribution.

This mathematical model also provides a means of estimating the average number of distinct concepts per word, which was found to be considerably less than the average number of senses listed per word. The grouping of word-senses into concepts is based on whether they could inspire the same new senses (by metaphor, metonymy, etc.), i.e. their potential future rather than their history.

keywords: polysemy, historical semantics, stochastic model, exponential distribution.

1 Introduction

Ambiguity is ubiquitous in natural language. It is most dramatic when it concerns the parsing of a sentence in examples such as

The High Court judges rape and murder suspects.

I heard a giant swallow after seeing a horse fly.

La petite brise la glace
('The girl breaks the mirror'/ 'The little breeze chills her') (from [44])

However, the most common form of ambiguity concerns the meanings of individual words, as in the following examples:

The *minister* decided to leave the *party*
('church minister'/'government minister', drinks party'/'political party')

He's a *curious* individual ('odd'/'nosey')

He fell down after taking a *punch* ('hit'/'drink')

Je *suis* un imbécile ('I'm following an idiot'/'I am an idiot')

The last two examples involve homographs (different words which happen to be spelt the same). However, it should be noted that only a small percentage of word-sense ambiguity is due to homography. (We obtained an estimate of approximately 2% by random sampling of English and French dictionaries [12, 21, 22, 24]). Many words have gained multiple senses by metonymy, figurative or metaphorical uses. The resulting senses are sufficiently different to be considered by lexicographers as distinct concepts (e.g. political party/drinks party). In information retrieval systems with natural language interfaces [51, 62] or in models of human language processing via networks of semantic links [43, 49, 66], a fundamental question is what should correspond to a basic semantic concept. Is it a word, a word-sense or a group of word-senses?

This paper presents a stochastic model of the evolution of language which allows us to answer this question. Applying the model to statistics obtained from a large number of monolingual and bilingual dictionaries [1]-[33] provides convincing evidence that neither words nor individual word-senses (as identified by lexicographers) correspond to concepts, but rather groups of word-senses. Our model will demonstrate that each word represents, on average, about 1.3 distinct concepts. This can be compared with the average 2.0 distinct senses per word listed in the dictionaries. This model also allows us to propose a novel and formal definition of the word "concept". There are clear applications in artificial intelligence [51, 62], cognitive science [42], lexicography and historical linguistics [35, 36, 61, 47].

2 The genesis of word-senses

Word origin and the evolution of spelling, pronunciation and meaning have long been studied by etymologists. Etymology tells us that many words in everyday use have a history that can be traced back thousands of years [55, 57]. This can be contrasted with the hundreds of new dictionary entries which lexicographers add to each new edition. These new entries are not only neologisms but also new senses for existing words. The history of the variations in spelling and pronunciation of particular words are not of direct concern here. We are interested in how (word,sense) pairs enter (or leave) a language. For each such semantic change we can try to identify the originator, the reason and the mechanism by which it occurs.

2.1 Origins of semantic change

Picoche [57] states that the majority of words in French have a scholarly origin and were introduced by clerics, jurists, intellectuals and scientists directly from Latin and Greek. However, it is clear that many words are of popular origin (e.g. *bike*, *trainers*, *OK* in English or *vélo*, *baskets*, *OK* in French) and have become accepted terms as the result of common use.

The principal reason why new (word, sense) pairs are introduced is to adapt language to new communicative requirements [61]. Discoveries and inventions can give rise to neologisms (e.g. *kangaroo*, *quark*, *internet*) or new senses for existing words (e.g. the 'armoured vehicle' sense of *tank* which coexists with the earlier 'large container' sense). Another driving force in historical semantics is the human tendancy towards efficiency of communication, by, for example, shortening words or expressions (e.g. clipping of *omnibus* to *bus*), ignoring unnecessary semantic distinctions and inventing new words to replace long expressions. Other reasons for semantic change have more to do with human psychology than with practical necessity. Taboo leads to the introduction of slang words or euphemisms, such as *to terminate* for 'to kill' or *senior* for 'old'. Litote is a special case in which a word is replaced by the negation of its opposite (e.g. *not bad* for 'good'). New words may be employed to make an old product sound more modern, exotic or appetizing (e.g. the old-fashioned British word *chips* is often replaced by *french fries* or *frites* on menus). The human tendancy to emphasize or exaggerate leads to the replacement of *severe* by *horrific* or *very* by *awfully* [61].

2.2 Mechanisms of semantic change

We can divide the mechanisms for neologism into three categories [35, 36, 39, 48, 61, 63]:

1. Word-creation from no previous etymon. This is rare but is the most likely explanation for echoic words such as *vroom*, *cuckoo*, *oh!* [38].

2. Borrowing from another language. This includes loan words (e.g. *strudel* from German, *pizza* from Italian) and loan translations (or calques) in which each element of a word is translated (e.g. *spring roll* from Chinese, *dreamtime* from the Australian aboriginal *alcheringa* [48] or *chien-chaud* which is the French Quebec version of *hotdog*).

3. Word-formation from existing etyma (words or word components). This includes

   (a) compounding (e.g. *bookcase*, *bushfire*)

   (b) blending (e.g. *brunch*, *motel*)

   (c) affixation (e.g. *<u>over</u>cook*, *<u>inter</u>national*, *like<u>ness</u>*, *privat<u>ize</u>*)

   (d) shortening (e.g. *petrol(eum)*, *radar*, *telly*, *AIDS*),

   (e) eponyms (e.g. *kleenex*, *sandwich*, *jersey*, *casanova*),

   (f) internal derivations [48] (e.g. *extend/extent* or *sing/song*),

   (g) reduplication [63] (e.g. *fifty-fifty*, *dum-dum*),

   (h) morphological reanalysis [61] (e.g. the non-existent verb *to edit* was formed from the noun *editor*; the word *cheeseburger* was derived from *hamburger* even though this word comes from the proper name *Hamburger*).

Mechanisms 1,3(f), 3(g), 3(h) are rare compared to 2, 3(a)-(e) [35]. Clearly, the above mechanisms are not exclusive. Borrowing and word-formation are obviously both at play in examples such as *blitz* which is a clipping of the German word *blitzkreig* and the French word *tennisman* which is a compound of two English words.

Word-creation, borrowing and word-formation generally produce a new word with a single sense, except when by coincidence the word already exists with a different sense. In the rest of the paper, we consider homographs produced by such coincidences as different words. In most dictionaries, homographs have distinct entries. For example, the term *bug*, meaning 'error in a computer program' was borrowed into French as *bogue* (by assimilation with the already existing word with the unrelated meaning 'husk'), but these two meanings of bogue are listed as two distinct words.

Nevertheless, we should mention three cases in which the neologism is often not recognised as a genuinely new word: ellipsis [36] (e.g. *daily (paper)*), zero derivation [53] (also known as conversion [48, 61]) (e.g. *to cheat* > *a cheat*) and borrowing of an already-existing word with a related sense (e.g. *to control* was borrowed into French as *contrôler* thus giving an extra sense to this French word meaning 'to verify').

The following is a list of mechanisms which can create a new sense for an already-existing word (adapted from [35]):

1. referential shift (e.g. *to print* now also refers to laser printers).

2. generalisation (e.g. *chap* used to mean 'a customer') or abstraction (e.g. *zest* denoted orange or lemon peel used for flavouring before being used in the abstract sense of 'gusto').

3. specialisation (e.g. in Old English *fowl* meant any kind of bird and *meat* any kind of food [55, 61]) or concretion.

4. metaphor (e.g. *kite* meaning 'bird of prey' applied to a toy).

5. metonymy (literally 'name change') i.e. naming something by any of its parts, accompaniments or indexes (e.g. *the crown* for the sovereign,

the City for the people who work there, *tin* for the container made of that metal, *cognac* for the drink originating from that region) [65].

6. clang association or folk etymology (e.g. *belfry* meant 'a movable tower used in attacking walled positions', but the first syllable was associated with *bell* and now the basic meaning is 'bell tower' [36]).

7. embellishment of language by using words which are more acceptable, attractive or flattering than existing terms (hyperbole, litotes, euphemisms, etc. as discussed above).

We use the general term *association* to cover all these cases.

3 The near-exponential rule

In order to study the relative importance of neologism, obsolescence and the creation of new meanings for existing words, we counted the number of senses listed for each word in several different dictionaries. We observed the following general empirical rule satisfied to within a fairly high degree of accuracy by all the monolingual dictionaries studied [9, 12, 19, 20, 21, 23, 24, 28, 31, 32]:

NEAR-EXPONENTIAL RULE: The number of senses per word in a monolingual dictionary has an approximately exponential distribution.

One way of testing this rule is by plotting $\log(N_s)$ against s, where N_s is the number of words in the dictionary with exactly s senses. If the Near-Exponential Rule is satisfied, then the resulting plot should be very close to a straight line with a negative slope. This is indeed the case for the dictionaries tested, with varying values of the slope depending on the dictionary. Figures 1 and 2 show the plot of N_s, on a logarithmic scale, against s for various English [19, 24, 28, 31] and French [9, 12, 20, 21, 23] dictionaries. Only those points (s, N_s) for which $N_s > 12$ are plotted in Figures 1 and 2.

For each dictionary, the values of N_s were obtained by sampling a random set of pages of the dictionary. Sampling was performed independently for each dictionary, meaning that the random sample of words was different in each case. We excluded entries corresponding to proper names, foreign words, spelling variants, derived words (such as past participles), regional

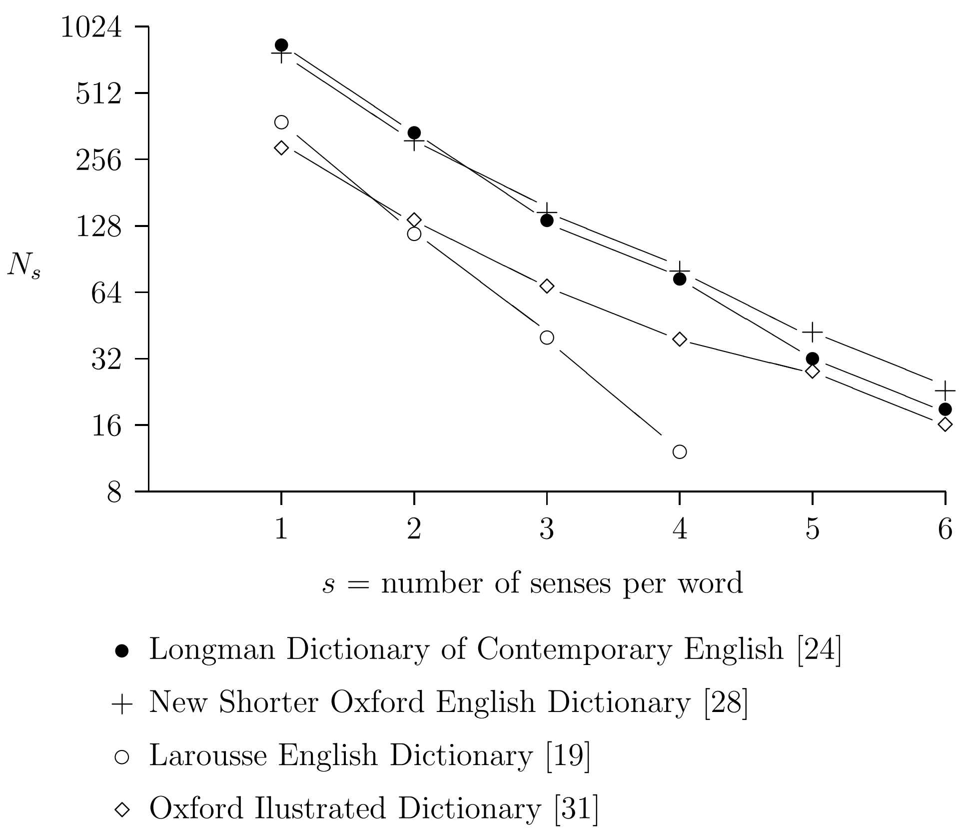

Figure 1: Plot of N_s (number of words with s senses) against s for various monolingual English dictionaries

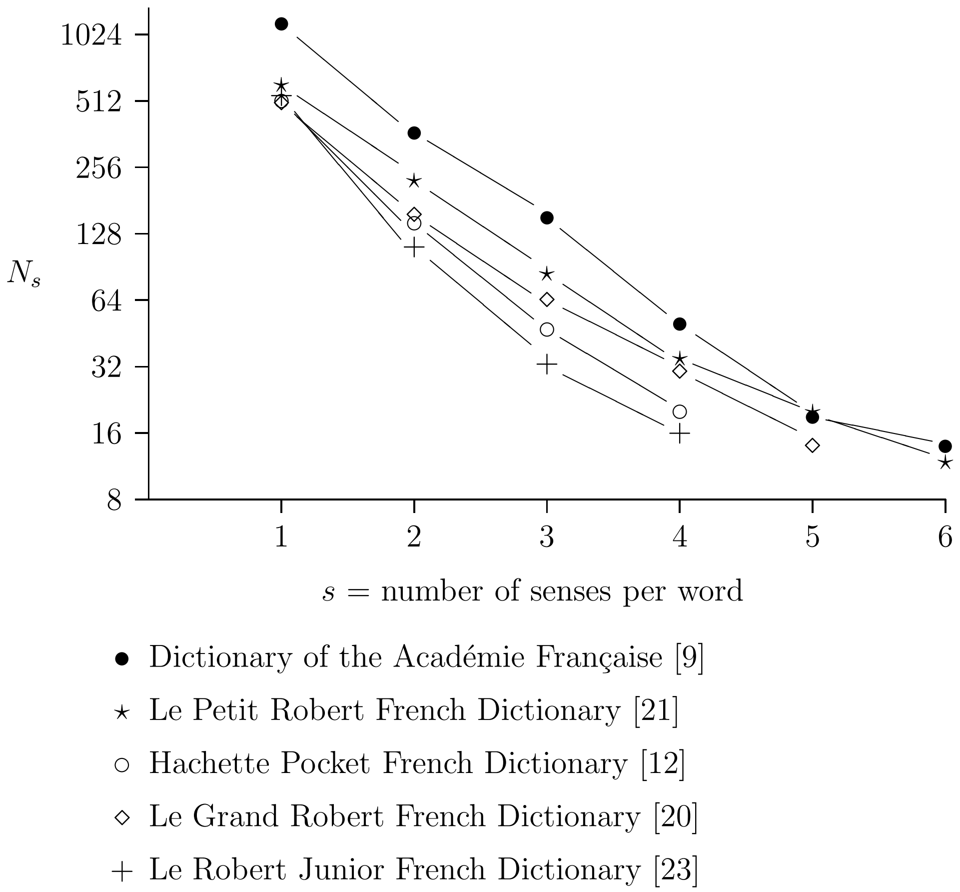

Figure 2: Plot of N_s (number of words with s senses) against s for various monolingual French dictionaries

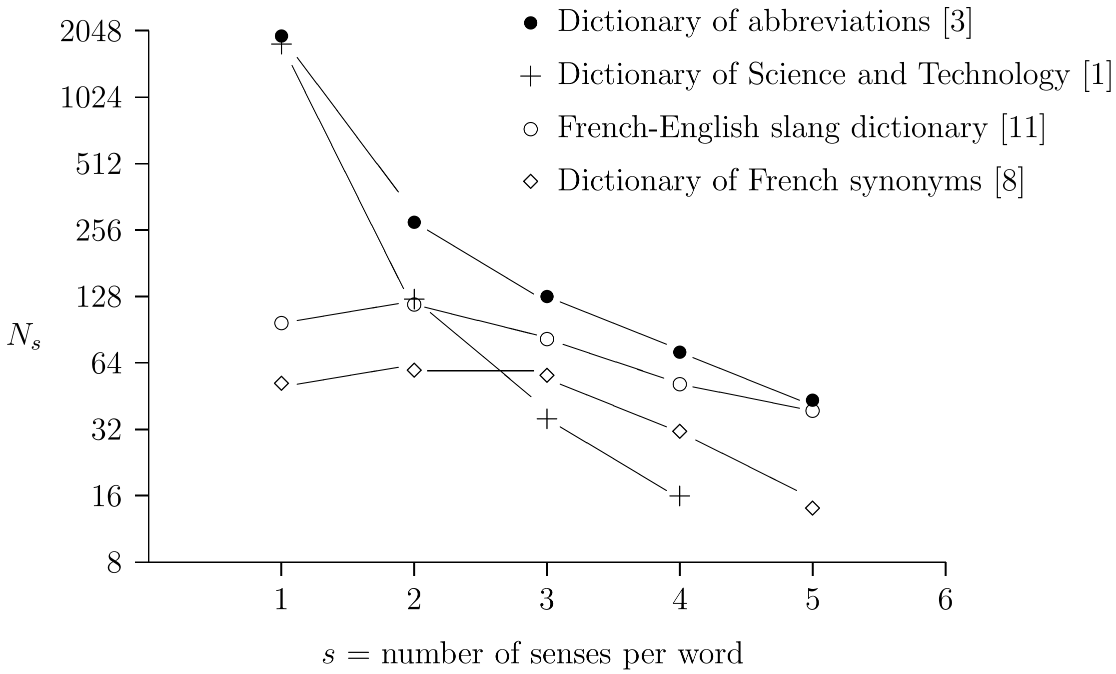

Figure 3: Plot of N_s (number of entries with s senses) for four different dictionaries, each showing a non-exponential distribution

words, abbreviations and expressions. We allowed hyphens within words but not spaces. Thus *cat-o'-nine-tails* counted as a word but *tower block* and phrasal verbs such as *to give up* did not. Only words forming part of British English or French spoken in metropolitan France were considered. All words were treated equally irrespective of their relative frequencies. Thus the words *to get* and *floccinaucinihilipilification* were given the same importance. The size of each dictionary is given in the reference section. The dictionaries vary in size from 20,000 to 80,000 words.

To ensure that the near-exponential rule was not simply an artifact of our choice of experimental procedure or of lexicographical practice, we performed the same analysis on a dictionary of abbreviations and acronyms [7], a dictionary of scientific terms [1], a bilingual dictionary of slang words [17] and a dictionary of French synonyms [10]. The resulting curves, shown in Figure 3, are far from straight lines.

We performed similar counts for bilingual dictionaries. Figures 4 and 5 show the number of words NT_t with t translations plotted against t for several pairs of languages. The NT_t scale is again logarithmic. Although

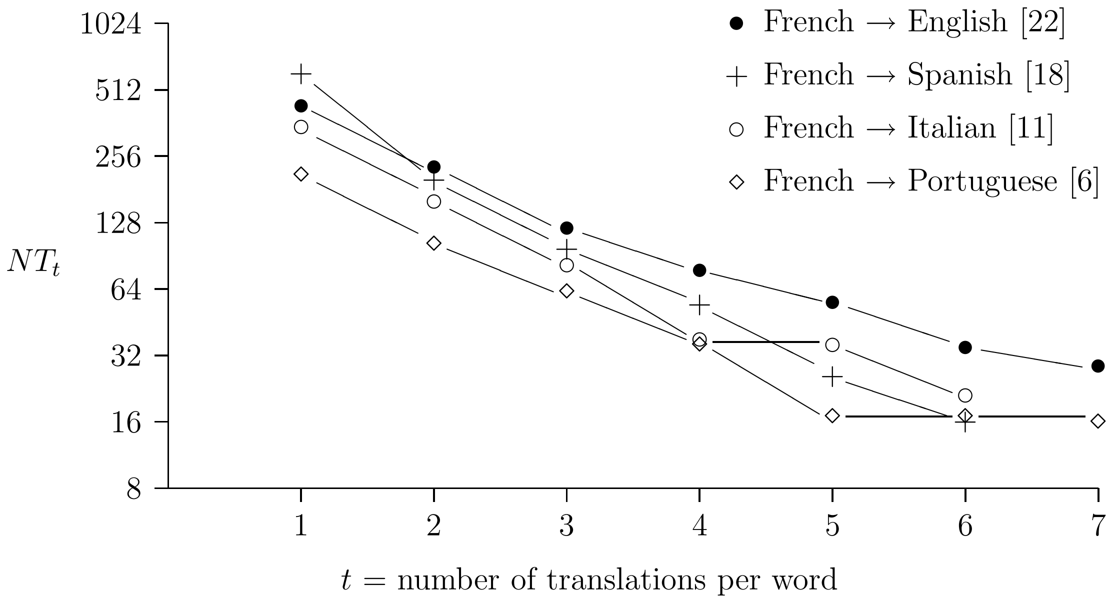

Figure 4: The number NT_t of words in French with t translations in English, Spanish, Italian and Portuguese

the near-exponential rule could also be said to hold for certain bilingual dictionaries, the curvature of the $\log NT_t$ against t curve varies considerably depending on the distance between the two languages. For pairs of languages with strong etymological connections (such as French-Spanish) the average curvature is positive (Figure 4), but for pairs of distant languages (such as Japanese-English) the average curvature is negative (Figure 5). A theoretical explanation of this phenomenon is outside the scope of the present paper, but is probably due to the greater differences in the segmentation of semantic space by distant languages (see [59] for some illustrative examples). It will be treated in detail in a follow-up paper.

4 Words, senses and concepts

In the following section we present a mathematical model which explains the near-exponential distribution observed in English and French dictionaries.

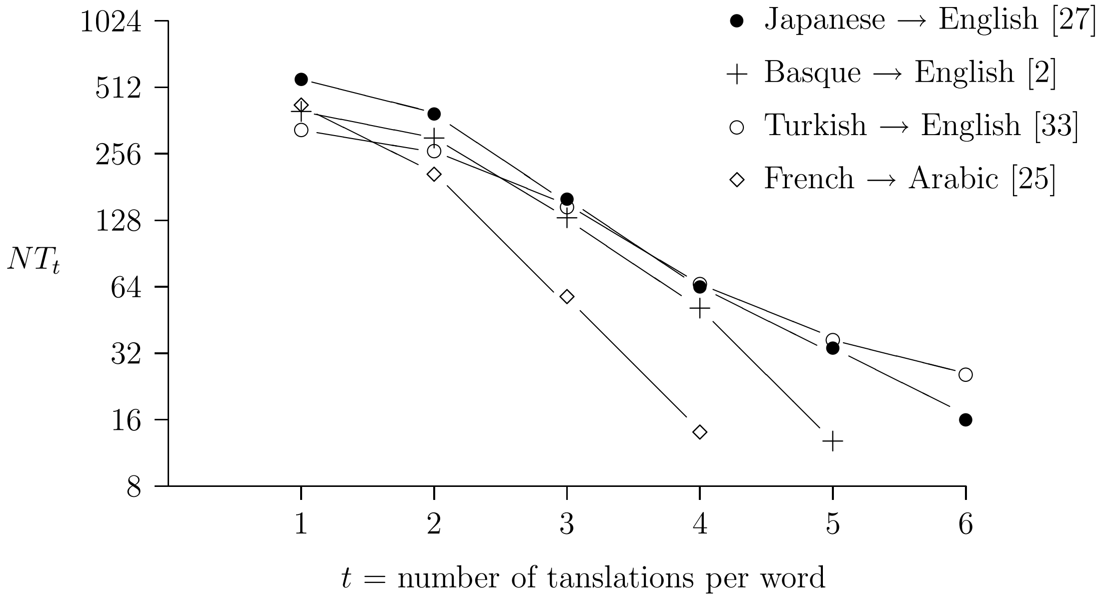

Figure 5: The number NT_t of words in langauge A with t translations in language B, for distant languages A and B

The curves of Figures 1 and 2 not only share the property of being close to straight lines (i.e. having curvature close to zero) but in each case the curvature that they do exhibit is always positive rather than negative. Although barely discernable for some of the curves, this positive curvature cannot be ignored. We fitted a straight line to the curves and then used a chi-squared test to judge the closeness of fit of this straight line to the data. For each curve the chi-squared test demonstrated a significant discrepancy between the model and the data. For example, the significance level was 15 standard deviations for the LCDE [24]. In order to find a satisfactory model to explain this slight but consistently positive curvature, we study in more detail the process by which words gain new senses.

The word *panel* provides a good example of a word whose number of meanings has grown since its introduction into English from Old French in the 13th century. Its original meaning was a piece of cloth placed under a saddle. Over the centuries it gained many meanings, by extension of this original sense, which can be grouped together in the following concept:

(C1) an often rectangular-shaped part of a surface (of a wall, fence, cloth, etc.) possibly decorated or with controls fastened to it.

The concept C1 covers four of the meanings of *panel* listed in the LDCE. However, during the 14th century *panel* also gained the following meaning: piece of parchment (attached to a writ) on which names of jurors were written (hence by metonymy) list of jurymen, jury [55]. Four of the meanings of *panel* listed in the LDCE can be considered to be covered by the following general concept:

(C2) a group of people (or the list of their names) brought together to answer questions, make judgements, etc.

If *panel* were to gain new meanings, such as

1. a side of a tower block

2. a school disciplinary committee

then these would be by association with the two concepts listed above (C1), (C2) respectively. Note that neither of these potential new meanings would correspond to a truly new concept, since they can be considered to be covered by the existing concepts (C1), (C2).

If, on the other hand, *panel* were to gain the following new meanings

3. a wall which divides a large room into smaller units but which does not reach the ceiling

4. a combined table and bench that could be used for example by a panel of experts

by association with the two concepts (C1), (C2) respectively, then these new meanings could be considered as corresponding to new concepts. These meanings are sufficiently different from the existing meanings listed in the LCDE that they themselves could give rise to further new meanings by metonymy, metaphor, etc. which would simply not be possible by direct association with the existing meanings. For example, the following meanings could theoretically be derived from the meanings 3, 4 above (but not directly from the concepts (C1), (C2)):

5. any division of something into smaller units

6. a combined desk and bench for a single person

We continue with another example, this time from French. The word *toilette* has eight meanings listed in Le Petit Robert [21], which we can translate and paraphrase as follows:

1. a small piece of cloth (from *toile*=piece of cloth) and, in particular, one that was used in the past to wrap up objects

2. a membrane used by butchers to wrap up certain pieces of meat

3. clothes, jewelry, comb, etc. (objects necessary to prepare one's appearance before going out and which used to be laid out on a small cloth)

4. the action of combing, making-up, dressing

5. a woman's style of dressing

6. the cleaning of one's body before dressing

7. a washroom, toilet

8. the cleaning, preparation of an object, text, etc.

We can group these meanings into three concepts:

(D1) a small piece of material (meanings 1,2)

(D2) the objects used for, the action of or the style of dressing, making-up, cleaning of a person or an object (meanings 3,4,5,6,8)

(D3) a washroom, toilet

We have grouped meanings together in this way because we consider that new meanings for *toilette* which could enter the French language by association with an existing meaning would be very similar for those meanings grouped into the same concept, but very different for those corresponding to different concepts.

This discussion leads us naturally to the following technical definition of concept.

Definition 4.1 *Two meanings of a given word correspond to the same* concept *if and only if they could inspire the same new meanings by association.*

We suggest grouping together different senses of a word, not only according to their parts of speech or to their etymology (i.e. the history) of word-senses, but also according to their potential future: whether or not they could inspire the same new meanings by association. This can be compared with the biological definition of species in terms of the ability to breed together to produce viable offspring rather than in terms of history or physical characteristics.

5 Mathematical model of word-sense genesis

This section describes a stochastic model of the creation of word-senses. This model not only explains the near-exponential rule but also provides a deeper insight into the process of naming. Let L_D be a language as defined by the set of (word,sense) pairs in a dictionary D. We consider the evolution of the language L_D over time. We must always bear in mind that L_D is, of course, only an approximate representation of the semantics of the corresponding natural language. For example, the compiler of a dictionary may choose to include archaic words as a historical record or exclude whole categories of words such as slang or technical terms.

Consider the evolution of L_D as a stochastic process in which each step is either

(a) the elimination of a word-sense (by obsolescence)

(b) the introduction of a new word (by creation, borrowing, word-formation, or any other mechanism)

(c) the addition of a new sense for an existing word (by association with an existing sense).

Let t be the probability of a step of type (a), u the probability of a step of type (b) and v the probability of a step of type (c). Note that $t+u+v=1$. The parameters of our model t, u, v are unknowns which will be estimated from the observed values of N_s (the number of words with s senses).

We make the following simplifying assumptions:

1. *New-word single-sense assumption:* when a neologism enters the language L_D, it has a single sense.

2. *Independence of obsolescence and number of senses:* the probability that a (word,sense) pair leaves the language L_D by obsolescence is independent of the number of senses this word has in L_D.

The new-word single-sense assumption is an essential part of our model. To test it we require two editions of the same dictionary. The 1994 edition of the Dictionnaire de l'Académie Française indicates which words are new compared to the 1935 edition. Less than 17% of these words are polysemic. Furthermore, this corresponds, according to our model and to within sampling error, to the proportion of originally monosemic words entering the language that can be expected to acquire new senses during the period between the publication of the two editions. Assumption 2 above is not as important since later we restrict ourselves to a no-obsolescence model.

As discussed in the previous section, the set of s senses of an ambiguous word may correspond to a number c of essentially distinct concepts, where c is some number between 1 and s. For example, the plumbing and anatomy senses of *joint* correspond to the same concept, since they could inspire the same new senses by association. The 'cigarette containing cannabis' sense of *joint* clearly corresponds to a different concept since it could inspire a very

	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9
SOED 1933	427	186	104	49	24	15	22	6	8
SOED 1993	403	176	86	44	32	16	14	7	1

Figure 6: The number N_s of words with s senses in samples from the 1933 and the 1993 editions of the Shorter Oxford English Dictionary.

different set of new senses by association. Associations inspired by distinct concepts are assumed to occur independently. We assume that a word with s senses in L_D represents on average $1 + \alpha(s - 1)$ concepts. We call α the *concept creation factor* (since, in a no-obsolescence model, α is simply the probability that a new sense for a word w can be considered as a new concept compared to the existing senses for w). We can now state a third assumption

3. *Associations are with concepts:* the probability that a concept gives rise to a new sense for a word w by association is proportional to the number of concepts represented by w in L_D, which is assumed to be on average $1 + \alpha(s - 1)$, where s is the number of senses of w and α is a constant.

The concept-creation factor α is another unknown which will be estimated from the values of N_s.

We make a fourth hypothesis in order to render the problem mathematically tractable:

4. *Stationary state hypothesis:* L_D considered as a stochastic process is in a stationary state, in the sense that the probability $P(s)$ that an arbitrary word of L_D has exactly s senses does not change as L_D evolves.

To test the validity of the stationary state hypothesis, we compared the 1933 and 1993 editions of the Shorter Oxford English Dictionary [32, 28]. In the space of 60 years the number of words in the SOED increased by 24%. Nevertheless the values of $P(s)$ $(s = 1, 2, \ldots, 9)$ remained almost constant. A chi-squared test revealed that the differences in the values of $P(s)$ $(s = 1, 2, \ldots)$ could be accounted for by sampling error. The corresponding values of N_s are given in the table of Figure 6.

The results of further experiments carried out to test the validity of the assumptions on which our model is based are given in a later section, so as not to clutter up the presentation of the model in this section.

Let m be the expected number of senses per word in L_D. Since

$$m \;=\; \sum_{s=1}^{\infty} s\,P(s) \tag{1}$$

and the values of $P(s)$ are constant by the stationary state hypothesis, m is also a constant.

The expected net increase in the number of word-senses in L_D during one step of the process is $-t + (1-t) \;=\; 1-2t$, since the probability that a word-sense is lost by obsolescence is t and the probability that a word-sense is gained is $1-t$. If r denotes the expected net increase in the number of words in L_D during one step of the process, then we must have $\frac{1-2t}{r} \;=\; m$ since m is a constant. Thus

$$r \;=\; (1-2t)/m \tag{2}$$

Note that the number of words in L_D would be constant if and only if $t = 0.5$.

Let $p_{out}(s)$ represent the probability that the next change in the language L_D is that a word with s senses loses one of its senses by obsolescence. Let $p_{in}(s)$ represent the probability that the next change in L_D is that a word with s senses gains a new sense. Note that $\sum_{s=1}^{\infty} p_{out}(s) \;=\; t$ and $\sum_{s=1}^{\infty} p_{in}(s) \;=\; v$, by the definitions of t and v.

By the stationary state hypothesis, the expected net increase in N_s (the number of words in L_D with exactly s senses) during one step must be proportional to $P(s)$. Denote the expected net increase in N_s by $\delta_s \;=\; dP(s)$, for some constant d. We then have $\sum_{s=1}^{\infty} \delta_s \;=\; d$, since $\sum_{s=1}^{\infty} P(s) \;=\; 1$. But $\sum_{s=1}^{\infty} \delta_s \;=\; r$, since the total expected increase in the number of words is r. Thus $\delta_s \;=\; rP(s) \;=\; (1-2t)P(s)/m$ (by Equation (2)).

We can also express δ_s, the expected net increase in N_s, in terms of the probabilities $p_{in}(s), p_{out}(s)$ which gives the following equation

$$(1-2t)P(s)/m \;=\; -p_{in}(s) \;-\; p_{out}(s) \;+\; p_{in}(s-1) \;+\; p_{out}(s+1) \tag{3}$$

since N_s is decremented when a word with s senses gains or loses a sense and N_s is incremented when a word with $s-1$ senses gains a sense or a word with $s+1$ senses loses a sense.

From the assumption of the independence of obsolescence and number of senses, it follows directly that $p_{out}(s)$ is proportional to $sP(s)$. Let $p_{out}(s) =$

$KsP(s)$, for some constant K. Then, since $\sum_{s=1}^{\infty} p_{out}(s) = t$, we have $t = \sum_{s=1}^{\infty} KsP(s) = Km$ by Equation (1). Thus $K = t/m$ and

$$p_{out}(s) = \frac{tsP(s)}{m}$$

Under the assumption that associations are with concepts, $p_{in}(s)$ is proportional to both $1+\alpha(s-1)$ and $P(s)$. Suppose that $p_{in}(s) = K' P(s)(1 + \alpha(s-1))$. Since $\sum_{s=1}^{\infty} p_{in}(s) = v$, $\sum_{s=1}^{\infty} P(s) = 1$ and $\sum_{s=1}^{\infty} sP(s) = m$, we have $v = \sum_{s=1}^{\infty} K'P(s)(1+\alpha(s-1)) = K'(1-\alpha) + K'\alpha m$. Thus $K' = v/(1-\alpha+\alpha m)$ and hence, for $s = 1, 2, \ldots$

$$p_{in}(s) = \frac{v(1+\alpha(s-1))P(s)}{1-\alpha+\alpha m}$$

Note that the creation of a new word with a single sense is a special case. By definition of u as the probability that the next step of the process is the creation of a new word,

$$p_{in}(0) = u$$

Summing Equation (3), for $s = 1, 2, \ldots$, gives

$$\frac{1-2t}{m} = -p_{out}(1) + p_{in}(0) = \frac{-tP(1)}{m} + u$$

Thus

$$u = \frac{1-2t+tP(1)}{m} \tag{4}$$

and, since by definition $v = 1-t-u$,

$$v = 1-t-\frac{1-2t+tP(1)}{m}$$

Plugging in the formulae for $p_{in}(s)$, $p_{out}(s)$ and v, our basic equation (3) becomes, after simplification, for $s > 1$:

$$\begin{aligned} & t(s+1)(1-\alpha+\alpha m)P(s+1) \\ & -\{(m-mt-1+2t-tP(1))(1-\alpha+\alpha s) \\ & \quad +(1-2t+ts)(1-\alpha+\alpha m)\}P(s) \\ & +(m-mt-1+2t-tP(1))(1-2\alpha+\alpha s)P(s-1) = 0 \end{aligned} \tag{5}$$

As observed in the previous section, empirical evidence indicates that $P(s)$ is a near-exponential function. In fact, if $P(s)$ were an exponential function then, since $\sum_{s=1}^{\infty} P(s) = 1$ and $\sum_{s=1}^{\infty} sP(s) = m$, we can easily deduce that $P(s)$ would be equal to $m^{-1}(1-m^{-1})^{s-1}$. The proof of the following result is simple but rather tedious and is hence is omitted.

Proposition 5.1 *The solution $P(s)$ to the set of equations (5) is the exponential function $P(s) = m^{-1}(1-m^{-1})^{s-1}$ if and only if*

$$\alpha = \frac{t}{m - 2tm + 2t}$$

Since the relationship between α, t and m given by the above Proposition did not seem to have any theoretical foundation, and since the observed values of $P(s)$ did not, in fact, follow a perfectly exponential distribution, we decided to estimate the values of the parameters m, α and t which would best explain the actual near-exponential distributions. We first set $m = \sum_{s=1}^{\infty} sP_{obs}(s)$, where $P_{obs}(s)$ are the observed values of $P(s)$ calculated from the values of N_s. Then we calculated the values of α and t which minimised the sum of the squares of the errors in the Equations (5). For six out of the ten dictionaries tested, the best-fit value occured when $t = 0$. The average value of the best-fit values of t was 0.04. These results led us to examine different editions of the same dictionaries in order to obtain an alternative estimate of t. We discovered that while hundreds or even thousands of words were added between different editions of the same dictionary [32, 28], very few words were removed due to obsolescence. For example, the number of words in the Dictionnaire de l'Académie Française [9] has increased by 28% in 59 years, whereas the total number of word-senses marked as obsolete in the latest edition is less than 1%. Our conclusion is that the English and French languages, as defined by dictionaries, are in a state of continual expansion with an almost negligible loss of word-senses by obsolescence.

We will therefore study in more detail the special case in which $t = 0$. The following result follows immediately from Equation (4) above, by setting $t = 0$.

Proposition 5.2 *When $t = 0$, the probability that the next addition to L_D is the creation of a new word is $u = 1/m$.*

Theorem 5.3 *When the assumptions 1,2,3,4 above hold and furthermore L_D is subject to no obsolescence, then the probability that an arbitrary word in L_D has s senses is given by*

$$\begin{aligned} P(1) &= \frac{1+\alpha m-\alpha}{m+\alpha m-\alpha} \\ P(s) &= \frac{1+\alpha m-\alpha}{m+\alpha m-\alpha}\prod_{i=2}^{s}\frac{(m-1)(1-2\alpha+\alpha i)}{m+m\alpha i-\alpha i} \qquad (s>1) \end{aligned} \tag{6}$$

Proof: When $t=0$, Equation (5) becomes, for $s>1$:

$$P(s)(m+m\alpha s-\alpha s) \;=\; P(s-1)(m-1)(1-2\alpha+\alpha s) \tag{7}$$

Summing Equation (7) for $s=2,3,\ldots$, gives

$$(1-P(1))m \;+\; (m-P(1))(m\alpha-\alpha) \;=\; (m-1)(1-\alpha) \;+\; m(m-1)\alpha$$

since $\sum_{s=1}^{\infty} P(s) \;=\; 1$, $\sum_{s=2}^{\infty} P(s) \;=\; 1-P(1)$, $\sum_{s=1}^{\infty} sP(s) \;=\; m$ and $\sum_{s=2}^{\infty} sP(s) \;=\; m-P(1)$. Solving for $P(1)$ gives

$$P(1) \;=\; \frac{1+\alpha m-\alpha}{m+\alpha m-\alpha}$$

The closed-form solution for $P(s)$ in the statement of the theorem then follows by an easy induction using Equation (7). ∎

6 Applying the model to experimental data

We make the no-obsolescence assumption throughout this section, i.e. that $t=0$. Proposition 5.2 allows us to estimate that, in French, approximately 60% of new word-senses correspond to the creation of a new word, and approximately 40% to the introduction of a new sense for an existing word. In English the split is approximately 50-50. There are, however, quite large variations (between 55% and 65% in French) depending on the dictionary. Variations are inevitable since different lexicographers have different interpretations of what constitutes distinct senses of a word. We conjecture that similar percentages exist for all natural languages, although there will be

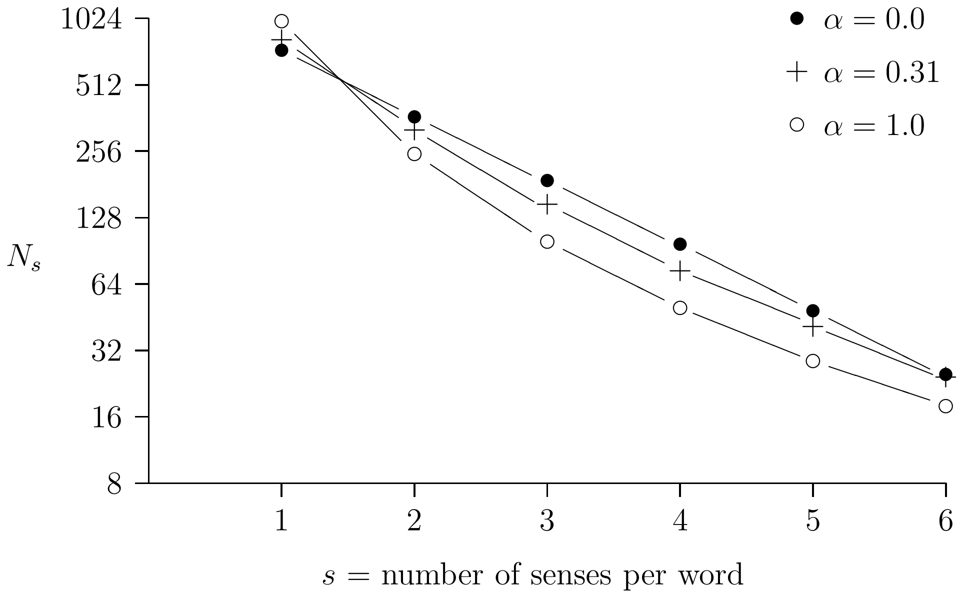

Figure 7: Plots of the predicted values of N_s for $\alpha = 0$,$\alpha = 0.31$ and $\alpha = 1.0$.

variations between languages depending, among other things, on the ease with which new words can be created.

The curves in Figure 1 are approximately straight lines but all have a slight positive curvature. This curvature can be explained by the fact that $\alpha > 0$. Note that, under the assumption $t = 0$, the concept-creation factor α is simply the probability that a new sense for an existing word is sufficiently different from previous senses for it to correspond to a new concept (capable of inspiring different associations to those that could be inspired by the existing senses). When $\alpha = t = 0$, Proposition 5.1 tells us that $P(s)$ is an exponential function. For $\alpha > 0$, however, the plot of $\log N_s$ against s does indeed have a positive curvature.

In order to evaluate visually the influence of the value of α on the predicted values of N_s, we generated the values of N_s using Equation 6 for various values of α. The results are plotted in Figure 7 (with the average number m of meanings per word set to be the same as the LDCE [24] in order to provide a concrete comparison). The observed values of N_s (for the LDCE)

DICTIONARY	m	α	c
Larousse (English) [19]	1.67	0.41	1.27
Shorter OED (English) [32]	2.26	0.22	1.28
New Shorter OED [28]	2.26	0.24	1.30
LDCE (English) [24]	2.04	0.31	1.32
Oxford Illustrated (English) [31]	2.46	0.22	1.32
Le Robert Junior (French) [23]	1.34	0.47	1.16
Académie Française (French) [9]	1.64	0.29	1.19
Hachette (French) [12]	1.53	0.38	1.20
Le Petit Robert (French) [21]	1.83	0.28	1.23
Le Grand Robert (French) [20]	1.79	0.46	1.36

Figure 8: The average number m of meanings listed per word, the concept-creation factor α and the average number c of concepts per word for various dictionaries.

coincide so closely with those predicted by our model with $\alpha = 0.31$ that the curves of observed and predicted values would be barely distinguishable if drawn in the same figure.

For each dictionary, we calculated the value of α which provided the best fit, in a least squares sense, between the observed values of N_s and those calculated from the values of $P(s)$ given by Equation (6). The best-fit values of α are given in the second column of the table in Figure 8 for each dictionary. The values of α vary between 22% and 41% for the English dictionaries and between 28% and 47% for the French dictionaries. Our conclusion is that, although nearly half of the words in a dictionary are ambiguous in the sense that they require more than one definition, only approximately one third of this ambiguity corresponds to ambiguity in the underlying concept (as defined by Definition 4.1).

The value of the concept creation factor α found for different dictionaries depends on the number of divisions into different senses the lexicographer chooses to list for each word. We can nevertheless calculate the average number of concepts per word in a dictionary. This number should be more independent of lexicographic choices. The table in Figure 8 also lists c the average number of concepts per word, which is given by $c = 1 + \alpha(m - 1)$, for each of the dictionaries studied. The average number of concepts per word is not the same, even for dictionaries of the same language. Variations are to be expected due to different lexicographical choices of which words and senses to include in the dictionary. We can note, in particular, that technical terms do not have the same distribution of number of senses per word as everyday words. Furthermore, many derived words do not have their own entries but are simply listed at the end of the entry for the root word. For example, in [24] *solidly* and *solidness* have no senses listed, and were hence ignored in our study, even though *solid* has 15 senses in the same dictionary.

Despite these inter-dictionary variations, we can nevertheless conclude that the average number of concepts per word (in the sense of Definition 4.1) is approximately 1.3 for English dictionaries and a little less for French dictionaries.

7 Further experiments to validate the model

As with any scientific theory, if our theory is correct we should be able to put it to the test by means of experiment. Playing the devil's advocate, we invented several experiments which, if unsuccessful, would demonstrate the invalidity of our mathematical model.

Firstly, we performed a chi-squared test to compare the observed values of N_s and the values of N_s predicted by our model (as calculated from Equation (6)). For nine out of the ten dictionaries tested, the χ^2 value was less than $\chi^2_{0.10}$ (the value which should be exceeded in only 10% of random trials). In the one other case, χ^2 was only marginally greater than $\chi^2_{0.10}$. These results are consistent with the hypothesis that the difference between the observed and predicted values of N_s are due to random sampling and that $E_s = N_s^{obs} - N_s^{pred}$ (for $s = 1, 2, \ldots$) are independent normally distributed random variables with mean 0 [50]. It is interesting to note that the difference between the observed values of N_s and those predicted by our model with $\alpha = 0$ (corresponding to the hypothesis that associations are with words) or $\alpha = 1$ (corresponding to the hypothesis that associations are with senses) are both statistically highly significant (at levels of 15 and 28 standard deviations, respectively, in the case of the LCDE [24]).

In order to test the validity of the stationary state hypothesis, we simulated the generation of a dictionary using the stochastic process model de-

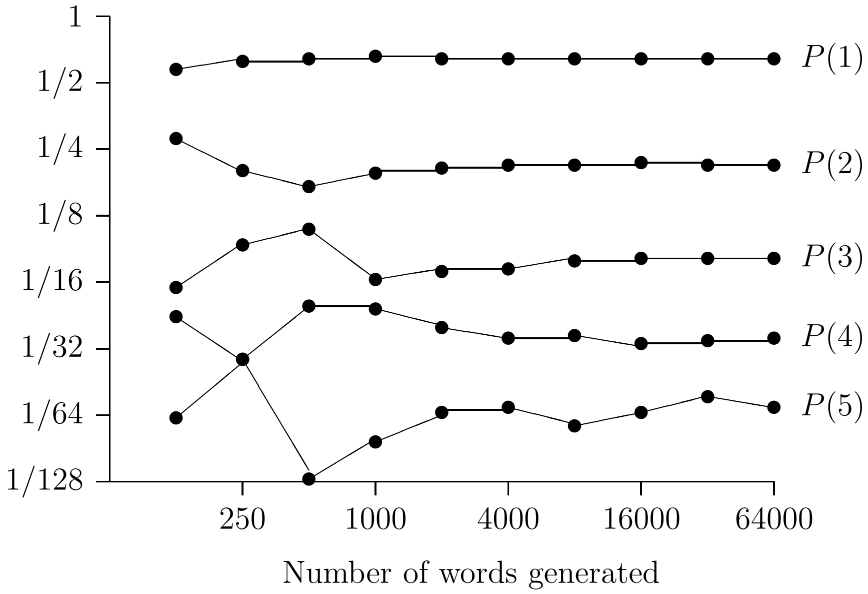

Figure 9: Values of $P(1), P(2), P(3), P(4), P(5)$ against the number of words generated in the simulation of the evolution of a dictionary

scribed in Section 5. We used a random number generator to decide whether the next step should be the creation of a new word or the creation of a new sense for an existing word. Figure 9 is a graphical summary of one such simulation, for the particular values $m = 0.6$, $t = 0$ and $a = 0.3$. The values of $P(1)$, $P(2)$, $P(3)$, $P(4)$ and $P(5)$ are plotted against the number of words generated. After the generation of only 1000 senses (which corresponds to less than 600 words), the values of $P(1)$, $P(2)$, $P(3)$, $P(4)$ and $P(5)$ are practically constant. We can deduce that a steady state has been attained long before the simulation generates a dictionary of size comparable to those studied (several tens of thousands of senses). We conclude that the stationary state hypothesis is, in fact, simply a mathematical consequence of our other assumptions, for dictionaries of any reasonable size.

To check the validity of our assumption that the average number of concepts corresponding to a word with s senses is $1 + \alpha(s-1)$, we tested a more general linear model $b + \alpha(s-1)$, for a constant b. The best-fit values of b for each dictionary were all found to be between 0.98 and 1.08, thus confirming our assumption $b = 1$.

Our conclusion that there is only a negligible loss of word-senses from dictionaries by obsolescence contrasts with the fact that 22% of the words in the Oxford English Dictionary [30] are marked as obsolete. Nevalainen [53] points out that many of these obsolete words were abortive attempts by pre-17th century writers to introduce new words which simply never caught on. Garner [45] attributes 1,700 neologisms to Shakespeare alone. Before the publication of the first monolingual English dictionaries in the early 17th century, both vocabulary and spelling were more a matter of personal taste than convention. Standardisation only occurred after the publication of Samuel Johnson's dictionary in the 18th century [8]. We should mention in passing that the very exhaustiveness of the OED makes it completely unsuitable (in the present context) as an accurate representation of the English language, since 90% of the senses listed are unknown to the majority of educated native English speakers. Thus our model cannot be expected to provide a faithful prediction of the evolution of the OED, since we assume that the set of word-senses in a dictionary is a an approximation to those available to people who create new senses for existing words. Instead of attempting to list all English words ever used, most dictionaries aim simply to list a set of words that an educated person might reasonably encounter during his/her lifetime, which is more in keeping with the assumptions of our model. Not surprisingly, therefore, fitting our model to values of N_s obtained from the OED gave incoherent values of the parameters ($\alpha = 1.51$ when, by assumption, we should have $0 \leq \alpha \leq 1$). We obtained a similar anomalous best-fit value $\alpha = 1.16$ for Webster's Third International Dictionary [34], no doubt because this dictionary is again so exhaustive.

It is worth going back to the counts of the number of senses per entry in specialised dictionaries [1, 7, 10, 17] plotted in Figure 3, to explain why these do not fit our model. The number of translations of a French word w in English slang [17] is related to the number of synonyms of w [10], since they both concern the onomasiological question of the different ways the same concept can be expressed in a language. This is the converse of the semasiological question of the development of different meanings of a given word, which is the problem our model addresses.

The number of meanings of abbreviations and acronyms [7] is closely related to the question of the distribution of homographs in a language, since abbreviations and acronyms almost invariably obtain new meanings by

DICTIONARY	m	α	c
Basque School[14]	1.35	0.48	1.17
Basque Learners[16]	1.36	0.50	1.18
Basque Modern[15]	1.38	0.55	1.21

Figure 10: The average number m of meanings per word, the concept-creation factor α and the average number c of concepts per word for three monolingual Basque dictionaries

coincidence rather than by association with existing meanings. For example, the 'temperature' and 'temporary' meanings of the abbreviation *temp* were clearly not derived by some direct semantic association between the notions of temperature and temporary (as would be required by our model).

The distribution of the number of meanings of scientific and technical terms [1] can, on the other hand, be partly explained by our model. The reason why the distribution is so far from satisfying the near-exponential rule is simply that 75% of the terms listed are composed of at least two words. When we only count single-word entries (as we did for all dictionaries in Figures 1 and 2), we obtain a distribution which can be explained by our model. We found that, although the average number of senses listed per word was much less than for English dictionaries of everyday language [19, 24, 28, 31, 32] (1.35 compared to 2.0), the number of concepts per word was approximately the same at 1.32.

In order to test the universality of the near-exponential rule, we also studied three monolingual Basque dictionaries [14, 16, 15]. Basque is a well-known language isolate. The curves of $\log N_s$ against s were again nearly straight lines with a slight positive curvature, and the values of N_s predicted by our model provided a very good fit to the observed values of N_s. The corresponding values of m, α and c are given in the table in Figure 10. The number of concepts per word was approximately 1.2 for all three dictionaries.

Our model assumes that there is no ambiguity in deciding what constitutes a word. However, such ambiguity is clearly present in fusional languages. In this paper, we have chosen the pragmatically simple definition that the words of a language can be approximated by those sequences of characters without spaces whose meanings are listed in a given dictionary. Applying this definition to a German monolingual dictionary [13], we ob-

DICTIONARY	m	α	c
Oxford Advanced Learner's [29]	1.56	0.41	1.23
Collins Concise [4]	2.22	0.24	1.29
Collins Learners [3]	1.74	0.39	1.29
Nelson [26]	1.72	0.43	1.31
Collins School [5]	1.64	0.56	1.36
Johnson [8]	1.55	1.00	1.55

Figure 11: The average number m of meanings per word, the concept-creation factor α and the average number c of concepts per word for six English dictionaries

served the usual slightly concave near-exponential distribution in N_s. The best-fit values of the parameters of our model were $m = 1.20$, $\alpha = 0.80$ and $c = 1.16$. The average number of meanings per word m and the average number of concepts per word c are low, no doubt due to the fact that many specialised terms which are expressed by a sequence of words in other langauges count, according to our definition, as a single word in German. Further research is required to test our model on other langauges with complex morphology.

Finally, we were surprised that the number of concepts per word was almost identical for the five English dictionaries tested (see table in Figure 8). However, we found that this was not always the case, since further trials on six other English dictionaries gave a larger range of values, shown in the table in Figure 11, varying from 1.23 to 1.55.

8 Relevance to computational linguistics

One application of our model is a simple method for testing whether an attempt to group word-senses into distinct concepts (according to Definition 4.1) has been successful. The number NW_i of words representing i concepts should demonstrate a distribution with a concept-creation factor close to 1, whereas the number NC_j of concepts covering j dictionary meanings should demonstrate a distribution with a concept-creation factor close to

0. Such a grouping of word-senses into concepts is clearly useful not only in computer models of natural languages but also in lexicography and historical linguistics. In lexicography, different rules have been proposed for identifying polysemy, based on etymology, statistical analysis of colocations in corpora, the existence of zeugma (such as *there is a pen on the table and one outside for the sheep), the existence of different synonyms (such as *present-now*, *present-gift*), antonyms (*right-wrong*, *right-left*) or paronyms (*race-racing*, *race-racist*), and the existence of ambiguous questions (such the canine/male ambiguity of the word *dog* brought out by the question 'Is it a dog?') [60, 37, 41]. In the context of computational linguistics, Milhacea & Moldovan relaxed these rules in order to find a more coarse grain representation in Wordnet, by grouping meanings based on similar synsets together with the existence of a common hypernym, antonym or pertainym [52]. The possible translations of a word w into several foreign languages is another useful practical tool for the grouping of the meanings of w into concepts [59].

We have introduced an equivalence relation between word-meanings: $S_1 \equiv S_2$ if the senses S_1, S_2 of a word w could give rise to the same new senses for w by metaphor, metonymy, etc. The grouping of meanings into the corresponding equivalence classes could be an essential part of an automatic system for the interpretation of non-standard uses of words. Words are often used with a meaning which is not explicitly listed in a dictionary. Metaphor and metonymy are obvious examples, but we can also mention meanings which are too specialised or too new to be listed in a general-purpose dictionary (such as the internet meanings of the words *provider*, *home*, *portal* for example).

Analysis of the plot of $\log N_s$ against s provides a method for identifying the criteria used in the compilation of a dictionary. A large positive curvature is characteristic of a dictionary whose aim is exhaustiveness. The OED [30] and Websters [34] are examples, and perhaps to a lesser extent Johnson [8]. A small positive curvature indicates a general-purpose dictionary whose aim is to list those words and meanings that an educated person can reasonably be expected to encounter during their lifetime. Machine-readable dictionaries play an important role in many natural language processing systems and the choice of dictionary is a critical one. In many applications, an exhaustive dictionary is inappropriate. Finding the best-fit value of the concept-creation factor α has allowed us to identify such dictionaries. Our model could also be used to estimate performance characteristics of systems which use machine-readable dictionaries, since we have given a formula for the expected number of words with s senses. For example, this may help us judge which is the best data-structure to use to store a dictionary.

An important aspect of the present work is the apparently universal nature of the near-exponential rule (with a slightly positive curvature when plotted on a logarithmic scale) for the number of words N_s with s dictionary meanings. This provides an insight into language in general rather than any one langauge in particular. Our mathematical model of historical semantics provides a very plausible explanation for this general rule. Various mathematical models of the evolution of networks have been proposed in recent years which explain other statistical phenomena in linguistics, such as the small worlds property of semantic nets [46]. It is worth pointing out that Price's [58] classical model for the number of nodes with s in-coming links in networks of citations of journal articles is mathematically identical to our predicted value of N_s if we set $\alpha = 1$. Since our model is a strict generalisation of Price's model, it may find applications, both within and beyond the frontiers of linguistcs, as a more general model of the prediction of network growth [54].

9 Conclusion

Empirical evidence indicates that the number of senses per word in a dictionary has an approximately exponential distribution. We have shown that a stochastic model of historical semantics can not only explain this near-exponential phenomenon but can also use the distance from an exponential distribution to estimate the average number of distinct concepts per word as well as the concept-creation factor (the percentage of new senses for existing words which can be considered as genuinely new concepts).

Further research is required to determine whether refinements to the mathematical model presented in this paper can produce a more accurate model by, for example, distinguishing between different parts of speech, distinguishing between everyday and technical terms and/or introducing the extra parameter word-frequency. The introduction to the OED [30] states that prepositions generally have more senses than verbs and adjectives, which in turn have more senses than nouns. Pagel [56] emphasises the fact that the evolution of language is not identical for all words. Fundamental vocabulary, including body parts, seasons and cosmological terms, are more stable than less basic words [64]. Zipf [67] was the first to notice a correlation between word-frequency and number of senses listed in a dictionary, in the form of a hyperbolic distribution. This can be summarised by saying that a word with frequency rank r has on average twice as many meanings as a word with frequency rank $10r$. A sophisticated stochastic model along the lines of the one presented in this paper but taking into account frequency would require information on the relative frequency of each different sense of each word. Unfortunately this is beyond the scope of this preliminary paper whose aim is simply to show that a simple stochastic model (based on grouping senses into distinct concepts liable to give rise to different new senses) can explain a universal property of monolingual dictionaries.

A more intriguing avenue of future research is the investigation of the possibility of using statistical analysis of dictionaries to model synonymy rather than (or as well as) polysemy. Indeed, it is an open question whether a similar approach to that followed in this work can be used to group together senses of *different* words which correspond to the same concept.

References

[1] *Academic Press Dictionary of Science and Technology*, ed. C. Morris, Academic Press, 1992 (*c.* 100,000 entries).

[2] *Basque-English Dictionary*, G. Aulestia, University of Nevada Press, Reno & Las Vegas, 1989 (*c.* 30,000 words).

[3] *Collins Cobuild Learner's Dictionary*, Harper Collins, London, 1996 (*c.* 24,000 words).

[4] *Collins Concise Dictionary of the English Language*, 2nd edition, Collins, London, 1988 (*c.* 37,000 words).

[5] *Collins School Dictionary*, Collins, Glasgow, 1989 (*c.* 17,000 words).

[6] *Dicionário de Francês Português*, Olívio da Costa Carvalho, Porto Editora, Porto, 1997 (*c.* 53,000 words).

[7] *Dictionary of Abbreviations and Acronyms*, 2nd edition, Tec & Doc - Lavoisier, 1992 (*c.* 50,000 entries).

[8] *A Dictionary of the English Language*, Samuel Johnson, facsimile edition, Times Books, London, 1979 (date of publication of original edition 1755) (*c.* 40,000 words).

[9] *Dictionnaire de l'Académie Française A-Enz*, Editions Julliard, 1994 (*c.* 56,000 words in the complete dictionary).

[10] *Dictionnaire des Synonymes de la Langue Française*, R. Bailly, Librairie Larousse, Paris, 1971 (*c.* 2,600 words).

[11] *Dictionnaire Général Français-Italien*, Larousse, 1994 (*c.* 34,000 words).

[12] *Dictionnaire Universel de Poche*, Hachette, 2000 (*c.* 32,000 words).

[13] *Duden Deutsches Universalwörterbuch*, Dudenverlag: Mannheim, 1996 (*c.* 122,000 words).

[14] *Europa Hiztegia - Eskola berrirakoa*, Adorez 6, 1993 (*c.* 24,000 words).

[15] *Euskal Hiztegi Modernoa*, Elhuyar Kultur Elkarten/Elkar SL, San Sebastian, 1994 (*c.* 38,000 words).

[16] *Euskara Ikaslearen Hiztegia*, Ibon Sarasola, Vox, Barcelona, 1999 (*c.* 26,000 words).

[17] *Harrap's English-French Slang Dictionary*, 1984 (*c.* 10,000 words).

[18] *Larousse Dictionnaire Français-Espagnol*, 1989 (*c.* 47,000 words).

[19] *Larousse English Dictionary*, Larousse-Bordas, 1997 (*c.* 29,000 words).

[20] *Le Grand Robert de la Langue Française*, 2nd edition, Le Robert, (9 volumes) 1992 (*c.* 80,000 words).

[21] *Le Petit Robert*, Dictionnaires Le Robert, 2000 (*c.* 46,000 words).

[22] *Le Robert & Collins English-French Dictionary*, 1978 (*c.* 31,000 words).

[23] *Le Robert Junior*, Dictionnaires Le Robert, 1999 (*c.* 20,000 words).

[24] *Longman Dictionary of Contemporary English*, 1978 (*c.* 38,000 words).

[25] *Mounged de Poche Français-Arabe*, 10th edition, Dar El-Machreq, Beirut, Lebanon, 1983 (*c.* 15,000 words).

[26] *Nelson Contemporary English Dictionary*, ed. W.T. Cunningham, Nelson, Walton-on-Thames, Surrey, 1977 (*c.* 20,000 words).

[27] *New Crown Japanese-English Dictionary*, Sanseido, 1968 (*c.* 55,000 words).

[28] *New Shorter Oxford English Dictionary*, Clarendon Press, 1993 (*c.* 78,000 words).

[29] *Oxford Advanced Learner's Encyclopedic Dictionary*, Oxford University Press, 1992 (*c.* 33,000 words).

[30] *Oxford English Dictionary*, 2nd edition, Clarendon Press, Oxford, 1989 (*c.* 290,000 words).

[31] *Oxford Ilustrated Dictionary*, 2nd edition, 1975 (*c.* 50,000 words).

[32] *Shorter Oxford English Dictionary*, Clarendon Press, Oxford, 1933 (*c.* 63,000 words).

[33] *Turkish-English Dictionary*, H.C. Hony, 2nd edition, Oxford University Press, 1957 (*c.* 16,000 words).

[34] *Webster's Third International Dictionary*, Encyclopaedia Britannica, Chicago, 1986 (*c.* 156,000 words).

[35] J. Algeo, "Vocabulary" in *The Cambridge History of the English Language*, Vol IV, ed. S. Romaine, Cambridge University Press, 1998, pp. 57-91.

[36] R. Antilla, *Historical and Comparative Linguistics*, Current Issues in Linguistic Theory 6, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2nd edition, 1989.

[37] J.R. Ayto, "On specifying meaning: semantic analysis and dictionary definitions", in R.R.K. Hartmann, ed., *Lexicography: Principles and Practice*, Academic Press, 1983, pp. 89–98.

[38] L. Bloomfield, *Language*, Henry Holt, New York, 1933.

[39] S. Chalker & E. Weiner, *The Oxford Dictionary of English Grammar*, BCA, London, 1994.

[40] M.C. Cooper, "Semantic distance measures", *Computational Intelligence* 16 (1), 2000, pp. 79–94.

[41] D.A. Cruse, *Lexical Semantics*, Cambridge University Press, 1986.

[42] D.A. Cruse, "Polysemy and related phenomena from a cognitive linguistic viewpoint", in *Computational Lexical Semantics*, ed. P. Saint-Dizier & E. Viegas, Cambridge University Press, 1995, pp. 33–49.

[43] C. Fellbaum, *WordNet, An Electronic Lexical Database*, MIT Press, 1998.

[44] C. Fuchs, *Les Ambiguïtés du Français*, Collection l'Essentiel Français, Ophrys, Paris, 1996.

[45] B.A. Garner, "Shakespeare's latinate neologisms", *Shakespeare Studies* 15, 1982, pp. 149–170.

[46] B. Gaume, K. Duvignau, O. Gasquet & M-D. Gineste, "Forms of Meaning, Meaning of Forms", *Journal of Experimental and Theoretical Artificial Intelligence*, 14(1), 2002, pp. 61–74.

[47] D. Geeraerts, *Diachronic Prototype Semantics: A Contribution to Historical Lexicology*, Clarendon Press, Oxford, 1997.

[48] S. Gramley, *The Vocabulary of World English*, Arnold, London, 2001.

[49] B. Hayes, "The Web of Words", *American Scientist*, March-April 1999, pp. 108–112.

[50] P.G. Hoel, *Introduction to Mathematical Statistics*, 5th edition, John Wiley and Sons, 1984.

[51] R. Mandala, T. Tokunaga & H. Tanaka, "Combining hand-made and automatically constructed thesauri for information retrieval", *International Joint Conference on Artificial Intelligence*, 1999, pp. 920–925.

[52] R. Milhacea & D.I. Moldovan, "EZ.WordNet: principles for automatic generation of a coarse grained WordNet", *FLAIRS Conference*, 2001, pp. 454–458.

[53] T. Nevalainen, "Early Modern English Lexis and Semantics", in *The Cambridge History of the English Language*, Vol. III, ed. R. Lass, Cambridge University Press, 1999, pp. 332–458.

[54] M.E.J. Newman, "The structure and function of complex networks", *SIAM Review* 45, 2003, pp. 169–256.

[55] C.T. Onions (ed.), *The Oxford Dictionary of English Etymology*, Oxford University Press, Oxford, 1966.

[56] M. Pagel, "The history, rate and pattern of world linguistic evolution", in *The Evolutionary Emergence of Language*, ed. C. Knight, M. Studdert-Kennedy & J.R. Hurfor, Cambridge University Press, 2000, pp. 391–416.

[57] J. Picoche, *Dictionnaire Etymologique du Français*, Dictionnaires Le Robert, 1992.

[58] D. J. de S. Price, "A general theory of bibliometric and other cumulative advantage processes", *J. Amer. Soc. Inform. Sci.* 27, 1976, pp. 292–306.

[59] P. Resnik & D. Yarowsky, "Distinguishing Systems and Distinguishing Senses: New Evaluation Methods for Word Sense Disambiguation", *Natural Language Engineering* 5 (3) 2000 pp 113–133.

[60] R.H. Robims, "Polysemy and the lexicographer", in R. Burchfield, ed., *Studies in Lexicography*, Oxford University Press, 1987, pp. 52–75.

[61] H. Schendl, *Historical Linguistics*, Oxford University Press, 2001.

[62] M. Stevenson & Y. Wilks, "Combining weak knowledge sources for sense disambiguation", *International Joint Conference on Artificial Intelligence*, 1999, pp. 884–889.

[63] R. Stockwell & D. Minkova, *English Words: History and Structure*, Cambridge University Press, 2001.

[64] M. Swadesh, "Lexico-statistic dating of prehistoric ethnic contacts", *Proceedings of the American Philosophical Society*, Vol. 96, 1952, pp. 452–463.

[65] E.C. Traugott & R.B. Dasher, *Regularity in Semantic Change*, Cambrige University Press, 2002.

[66] P. Vossen, "Condensed meaning in EuroWordNet", in *The Language of Word Meaning*, eds. P. Bouillon & F. Busa, Cambridge University Press, 2001, pp. 363–383.

[67] G.K. Zipf, *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*, Hafner Publishing Company, New York, 1972 (facsimile of 1949 edition, Addison-Wesley).