

A Finer Definition of Neology in English: the life-cycle of a word

Abstract

It has been shown (Renouf, 1993a) that lexical and semantic neology can be identified in a text corpus at surface level by automatic means. According to hypotheses upheld, a lexical neologism is often a lexical item which occurs for the first time in a diachronic corpus of journalism; and found automatically by comparing each word in a stream of fresh corpus data with a baseline index, such that each previously unseen item is deemed to be a candidate neologism. Meanwhile, a semantic neologism has been shown to be identifiable through the change in the word's collocational environment (Renouf, 1993b).

Most neologisms, once recorded, never reappear in the corpus, while others remain. In this paper, we examine the changing status of neologisms across time, tracking the 'life-cycle' of a word (Renouf, 2007), from its first appearance in our text, through its fluctuations in frequency and popularity, to its possible assimilation into mainstream language, to its possible death and re-birth.

For convenience, we use the term 'word' as generic for all neologistic formations, though of course these may take the form of a simplex or complex word, or a multi-word unit of incipient or established fixedness. The study will be based on a corpus of 1.2 billion words of UK mainstream newspaper text covering the period 1989-2011.

Key words: lexical neology, semantic neology, neologism, collocation, Collocate Bank

1. Introduction

It has been shown (Renouf, 1993a) that lexical and semantic neology can be identified in a text corpus at surface level by automatic means. According to hypotheses empirically tested and upheld, a lexical neologism is usefully often a lexical item which occurs for the first time in a diachronic corpus of modern-day English, specifically one of mainstream journalistic text; and which can thus be found automatically by comparing each word in a chronological stream of fresh corpus data with a baseline index, with each previously unseen item emerging deemed to be a candidate neologism.

Meanwhile, a semantic neologism, or new sense of an existing word, has been shown to be realised in part by the change in the word's collocational environment. In Renouf, 1993b, we cited the word *charter*, which formally meant 'low-cost', in the context of *airlines, flights* and *holidays*; but in 1991 gained a newⁱ sense of 'governmental undertaking', in the context of *citizen's, patient's* and *parent's*.

However, the definition of a lexical neologism as a word at its first occurrence in a particular corpus is only part of the story, dictated by the nature of automated identification. Some words do indeed occur only once in a corpus; in our 1.2 billion words of UK mainstream newspaper text covering 1989-2011, 67% of the word types fall into this category; they include both 'nonce formations' – different kinds of one-off innovation, including context-dependent coinages such as *apple-juice seat* (Downing, 1977) and morphological irregularities, such as *heroidy* (*hero + oid + y*; Hohenhaus, 2005); hapax legomena (conventional, quite possibly future bona fide members of the lexicon, but so far recorded as occurring once only); and errorsⁱⁱ. But the other 33% of word types do recur. This distinction matters when potential users of the analysed data are taken into consideration. For some, such

as linguists and language teachers, any neologistic use is a matter of interest. Lexicographers require further that a neologism exhibit a degree of staying power in text over a period of time, and even show signs of establishing itself in the language. Meanwhile, for a translator or a terminologist for whom English is not the native language, the problem is exacerbated. On the one hand, they may need some objective endorsement of the degree to which a potential neologism is taking root in the language; as well as some benchmark to judge which of two competing neologisms is winning the race. On the other, they may need to know whether the English source language neologism is actually new and not already in technical or general text; that it is not a typographical error, or an aberrant invention by an English non-native-speaker writing a text (e.g. EU and other official documentation) in English for translation into other languages. Maria-Jose Caravina, Senior Terminologist at the EU Translation Centre in Luxembourg (personal communication), illustrates the dilemma with the term *serial maize*. The translator must determine whether this is a bona fide term meaning some form of super-productive, genetically-modified cereal crop, or a phonological confusion with *cereal maize* (corn flakes). A relevant control corpus, or even a large general corpus, including the web, can be of immediate help in such a case.

This paper is part of a larger study aimed at investigating aspects of the nature of neology which are represented in a dynamic corpus. In this case, it is a 1.2 billion word corpus of English mainstream daily news text from the *Guardian* and the *Independent* newspapers, stretching from 1989-2011 at the time of writing. In Renouf (2007), we examined lexical productivity and creativity; in Renouf (2012), we considered the needs of language professionals mentioned above; in Renouf (forthcoming), we shall look at the way neologistic types accumulate in text, to the point of constituting the start of a new register. In the current paper, we shall apply a combination of linguistic criteria and the lexical-statistical measures

created during the AVIATOR (Renouf, 1993a, b) and APRIL projects (Renouf, 2010) to identify the changing status of a neologism in a corpus across time, tracking the ‘life-cycle’ of a word from its first appearance, through the stages of its assimilation into mainstream language, to its possible death and revival.

1.1 Intuitive definition of a neologism

In the absence of any corpus evidence, it is difficult to know where to start in finding or analysing candidate neologisms. It poses a fundamental problem for lexicographers. The Merriam Webster dictionary websiteⁱⁱⁱ still talks of its editors devoting:

“an hour or two to reading a cross-section of published material ... this activity is called "reading and marking." The editors scour the texts for new words, new usages of existing words, variant spellings, and inflected forms– in short, anything to help in deciding if a word belongs in the dictionary, understanding what it means, and determining typical usage ... To be included in a Merriam-Webster dictionary, a word must be used in a substantial number of citations that come from a wide range of publications over a considerable period of time ... The number and range of citations needed to add a word to the dictionary varies. In rare cases, a word jumps onto the scene and is both instantly prevalent and likely to last, as in 1980s with AIDS. In such a situation, the editors determine that the word has become firmly established and should be entered in the dictionary...”.

This traditional approach to the identification of neologisms clearly begs questions about selectional criteria and the feasibility of their objective application. But fortunately, awareness of the need for corpus evidence is slowly increasing in dictionary publishing houses. Since May 2011, the Merriam Webster site adds that: “Citations are also available to

editors in a searchable text database (linguists call it a corpus) that includes more than 70 million words drawn from a great variety of sources.”

1.2 Neologisms in a static, synchronic corpus

The availability of corpus evidence such as this is an improvement. But where the corpus is a static entity, as is still the case for many publishing houses, its ability to provide information about neologistic behaviour is limited. A neologism can be visually spotted as being unfamiliar, and will often be signalled as new by certain punctuation or metalinguistic comment, or explicitly presented as such, in terms such as ‘first coined by X in Y’. The corpus could also be run automatically against a master word list, which would identify all words not already registered. What a static corpus cannot do, however, is pinpoint the first occurrence of the word, trace its evolution across time, or compare the course of different new words and their variants relative to each other in time.

1.3 Neologisms in a dynamic, diachronic corpus

In our evolving, diachronic corpus, the identification process is simpler: each word entering the database is matched automatically against a base-line corpus, so that each unseen word is pinpointed at its first occurrence, and deemed to be a candidate neologism. Whether it remains a nonce formation or becomes assimilated can then be estimated or ascertained by tracking its subsequent path automatically through time. Thus, our definition of neologism becomes: ‘the first *and subsequent* occurrences of a word’ – a perspective of more use to linguists and professional wordmongers alike.

The relevant software tools capable of filtering and analysing the evolution of our text reside in two automated systems developed in the Research and Development Unit for English Studies: AVIATOR (Analysis of Verbal Interaction & Automated Text Retrieval), developed

in 1990-93 (Renouf, 1993) and APRIL (Analysis and Prediction of Innovation in the Lexicon), in 1997-2000 (Renouf, 2010). As outlined in the Introduction earlier, AVIATOR identifies and classifies new words according to simple surface criteria. APRIL moves on from AVIATOR in carrying out a morphological analysis, using a chart parser which parses each new word at character level, then classifying it grammatically and according to word formation type. The output in this paper is extracted from the downloaded and processed news corpus via the WebCorp Linguist's Search Engine user-interface (<http://wse1.webcorp.org.uk/>).

1.4 Semantic neology

In our AVIATOR system, the syntagmatic axis is used as the gateway to the paradigmatic. The surface patterns of text – specifically the collocational patterns of a word - are seen to determine a word's sense. Thence, significant changes in a word's collocational profile over time are deemed to mirror a change in its meaning. By these criteria, semantic neology, or semantic change in existing words, can thus also be discovered.

1.5 Formal neology

Formal neology is another standard procedure, for which most of the word types in the English lexicon are eligible (for some exceptions, see Davies, 2004), whereby a neologism changes grammatical word class and thereby increases its influence on text. Our APRIL system parses new words at point of entry into our corpus, and proposes a grammatical word class; but we identify formal neologisms at a post-processing stage, since formal neology involves a later change, to existing words. Later, in Figure 10, we illustrate such a grammatical transition with the term *shock and awe*, as it converts from noun to verb.

2. The life-cycle of a word in diachronic text

The diachronic approach to the study of neologisms in text allows us to observe the existence of a measurable 'life-cycle' for each word. According to this metaphor, used by analogy with a human life-span, the life-cycle of a word is conceived as consisting of some or all of the following major stages: birth, or perhaps just first occurrence in text; possible increase in frequency of occurrence; productivity, creativity, settling down, assimilation and establishment in the language, obsolescence, possible death - and possible revival.

As itemised, this model more accurately resembles a 'series' or a 'trajectory' than a 'cycle', the latter entailing an element of repetition within a period of time, a return to the start. But in fact, our model needs to encompass just those extra features, since the one-way path from life to death for one word can be cyclical for another. At some point in its 'life', a word may acquire new or additional senses. Alternatively, it may appear to 'die', but subsequently be resuscitated by some topical vicissitude. In each case, the word in its new sense or connotation, or with its new topical reference, will re-enter the life-cycle at the beginning, and progress approximately as does a word for the first time.

In the rest of this paper, we shall describe and illustrate the stages of the life-cycle, together with its possible 'recursive' and other elaborations, with reference to a series of words which are or which have been neologisms in our data.

2.1 Stage in Life-cycle: Birth

A new word is formed by a number of means; rarely by totally original linguistic invention or coinage; typically by the application of formation rules, primarily by compounding

(*Cleggmania*) or derivation (*gruntologist* - *gruntology* + *ist*); blending (*Grexit* - *Greek* + *exit*)

or conversion (*ask* – VtoN). The vast majority of neologisms are constructed of existing language elements; words and affixes. In addition, neologisms can be foreign borrowings, or loans from technical domains.

The arrival of a word in our corpus, even after filtration through a vast base-line lexicon, does not mean that it is new in the language. It may have been in circulation in private conversation or in technical text, and suddenly become sufficiently topical or attractive to excite the imaginations of our newspaper journalists.

Proper nouns are a separate, and numerically significant, category. In our data, they may be newly coined to designate new inventions and discoveries; but they may also name previously existing but obscure phenomena - people and particularly places - which are propelled into the media glare due to a real-world event or popular preoccupation. Every news event ruffles the water of the lexicon, perhaps only briefly, and a flurry of topically relevant words, new and old, will suddenly emerge in response. Proper nouns are often used metonymically, particularly when naming locations, as we shall later show.

2.1.1 Birth: *Eyjafjallajökull*

The proper noun *Eyjafjallajökull* exemplifies the case of the toponym which becomes news. It is not a neologism, referring to an ice cap sitting on a volcano in Iceland which is reported to have shown activity as early as 550 AD. In its recent incarnation, however, the word newly refers metonymically to the particular eruption of its volcano on April 14, 2010. The name rises to prominence in our data in April and May 2010, when 57 of its total 159 occurrences in our corpus are devoted to reportage before the end of April; followed by 15 instances in May, and 8 instances in June and 6 in July, devoted to discussion of the consequences, before

it fades again. Its time-graph in Figure 1 displays a typical change pattern for a natural event like this: namely, a sudden rise and fall. But, as can also happen, the name *Eyjafjallajökull* reappears in late May 2011, caught up in the reportage and discussion surrounding Grímsvötn, the volcano involved in the subsequent further eruption in Iceland.

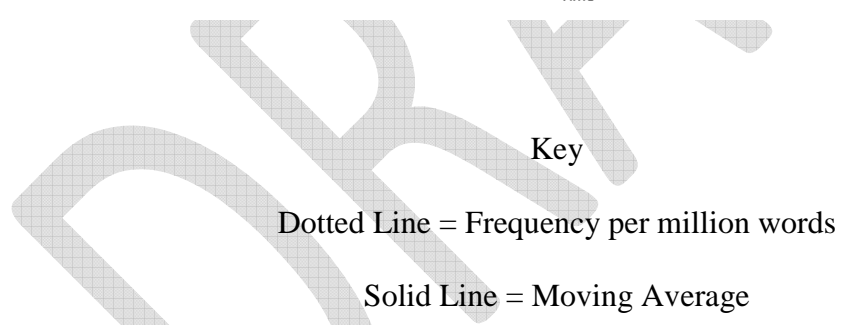
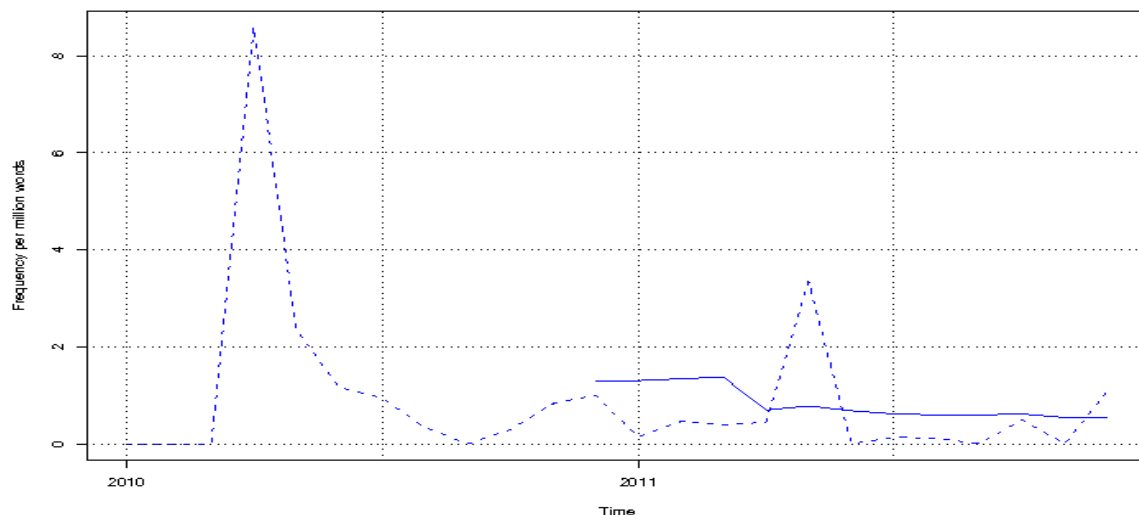


Figure 1: Time-graph for *Eyjafjallajökull* (159 occs.)

There was also, as predicted, a corresponding spike in established terms associated with this event and its consequences: *volcanic ash*, *silica dust*, *magma*, etc; as well as the emergence of new ancillary terms, including *volcanic insurance*, *volcanic tourism* and *Airborne Volcanic Object/Identifier and/ Imaging Detector (or Avoid)*, as contextualised in Figure 2.

15/04/10 the Guardian ran a piece last week on **volcano tourism**, which was timely.

22/04/10 "Travel free on Ryanair, when you buy our euro50 **volcano insurance**

04/06/10 The **Airborne Volcanic Object Identifier and Detector**, or Avoid, is based on technology tested on active volcanoes in Sicily

08/07/10 Any new eruption could boost demand for "**volcano insurance**"

08/07/10 The number of people searching on Google for "**volcano insurance**" increased

08/12/11 the **Airborne Volcanic Object Imaging Detector** (or Avoid) could be the weapon that tames the volcanoes

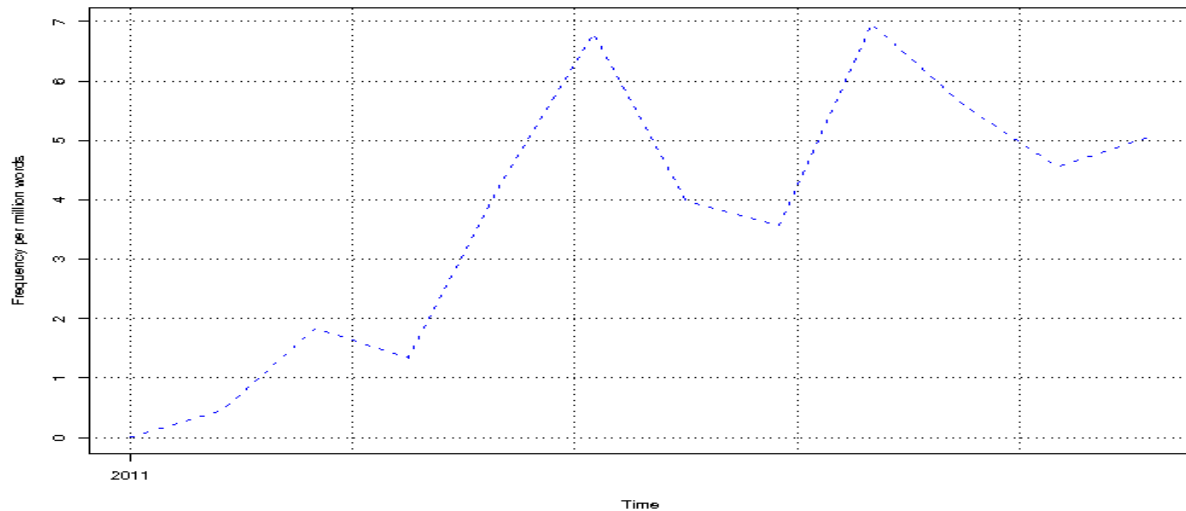
Figure 2: some ancillary coinages attendant upon the neologism *Eyjafjallajökull*

In this context, we might point to De Smedt's (2012) study of the sudden increase in the creation and use of compounds with *aske* (or *ash*)^a in the *Norwegian Newspaper Corpus* (Andersen, 2012), in the immediate aftermath of the same 2010 volcanic eruption.

2.1.2 Birth: Arab Spring

We saw the initial spike of frequency for the name *Eyjafjallajökull* as it mirrored topical events. In Figure 3, the time-graph for the proper noun phrase *Arab Spring* similarly lurches upwards in response to real-world events. This phrase names not a place but a concept; the idea that a new era of political and social enlightenment and liberation is about to begin in the Arab world. The metaphor equates this new era with the season of spring. It is a metaphor which had begun to circulate in the media in March 2005, when media commentators suggested that a spin-off benefit of the Iraq invasion would be the 'flowering' of Western-friendly Middle East democracies. In our data, however, it first appears only in February

2011, as a Tunisian street vendor dies after self-immolation, sparking anti-government protests in Tunisia and then other Arab nations. These protests became known collectively as ‘the Arab Spring’.



Key:

Dotted Line = Frequency per million words

Solid Line = Moving Average

Figure 3: Time-graph for *Arab Spring* (1574 occs.) case-insensitive

In Figure 3, the pattern of birth, or ‘rebirth’ here, is of a series of spikes across 2011, each representing newly heightened interest and discussion reflecting real-world developments. (A smoothing curve of moving average frequency is incalculable with such erratic results, and thus absent from the graph.)

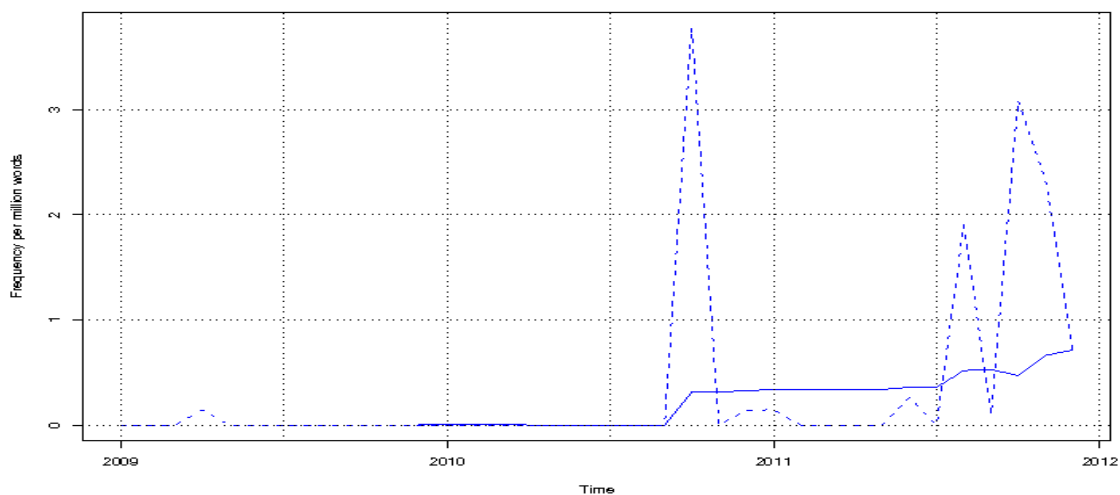
2.1.2.1 Analogous coining: *Arab Spring*

The term *Spring*, not new in itself but part of the bedrock of English metaphor, has been active in characterising events across this period. The term *Arab Spring* is almost certainly formed by analogy with *Polish Spring* (1950s), *Prague Spring* (1968), *Czech Spring* (1968), *Croatian Spring* (1968-74), *Peking Spring* (1978-9), *Moscow Spring* (1988), *Slovenian Spring* (1988), *Beijing Spring* (1989), *Slovene Spring* (1990), *Damascus Spring* (2000) and *Tehran Spring* (2000-4). Since *Arab Spring*, the search term [** Spring*] has continued its geographic meanderings, retrieving *Peshawar Spring* (2008) and *Palestinian Spring* (2011) by the time of writing.

2.1.3 Birth: *graphene*

Proper names are commonly coined for inventions, particularly scientific and technological. The first occurrence of the term *graphene* appeared in our news corpus in 2009, when it referred to a thin graphite layer capable of conducting electricity:

09/04/2009 **Graphene** could go faster. MIT researchers are using one-atom-thick pure carbon to make chips that could run at 500 GHz or more.



Key

Dotted Line = Frequency per million words

Solid Line = Moving Average

Figure 4: Time-graph for *graphene* (108 occs.), case insensitive

In Figure 4, the time-graph for *grapheme* allows for case insensitivity, with *Graphene* first appearing (19 occs.), swiftly shifting to *graphene* (89 occs.). In this time-graph, we see a rather protracted birth for the term. The probable reason is that, as a ‘neonym’ or technical neologism, *graphene* denotes an immensely valuable conductive material which has until recently has been researched in competitive secrecy. As the process of creating the material itself has evolved, so the term *graphene* has been coined more than once^{iv} (by Boehm, 1962, and Mouras et al, 1987), referring to earlier versions of the material and, metonymically, to the associated process which finally became officially recognised in the version developed by Geim and Novoselov at the University of Manchester in 2004.

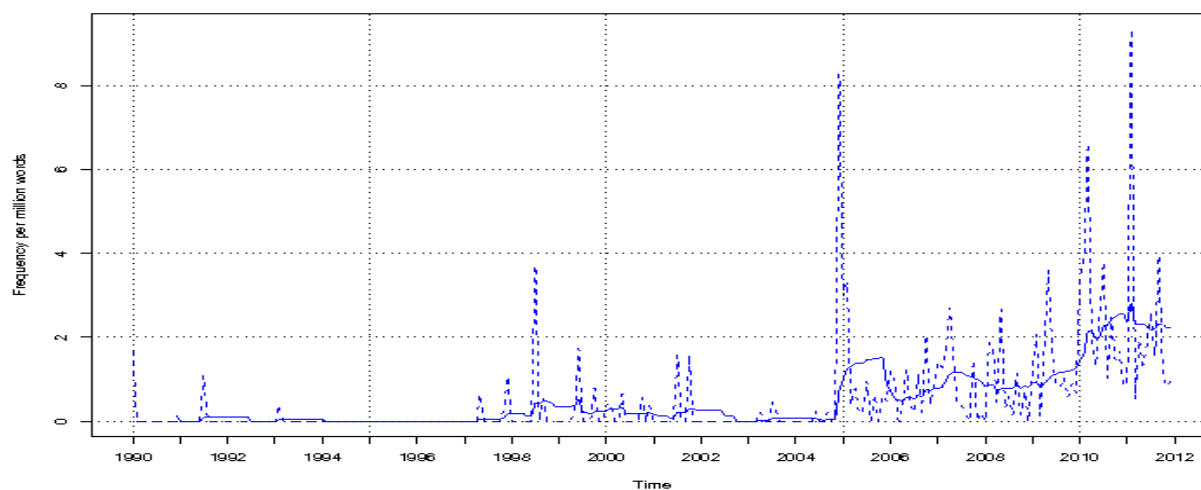
Our time-graph shows that ‘graphene’ suddenly finally newsworthy in autumn 2010. This was when Geim and Novoselov at the University of Manchester received the Nobel Prize in Physics. From the second half of 2011, the research team’s receipt of modest UK government investment maintains its mention^v. The term, and the invention, have yet to scale the heights of popularity, with only 108 mentions by the end of 2011, many of which emanate from relatively few dedicated news articles.

2.2 Stage in Life-cycle: Increase in frequency - gradual, or sudden if in vogue

There are several recognisable ‘change types’ for neologistic behaviour, variously representing slow or swift, slight or major, smooth or jerky change patterns, and these in combination. There are also a number of possible graphical representations of ‘time-series’ data, involving blocks and lines, among them the line graphs we have chosen to use in our time-graphs, to display the ‘birth’ and subsequent life-cycle stages in this paper. It should be borne in mind that the graphs organise themselves to best display the frequency changes across time, and that there are corresponding changes of scale which might make a trivial growth appear significant, or vice versa, unless one consults the scalar values on the vertical axes.

2.2.1 Increased frequency: *FOI*

As illustrated in Figure 5, the abbreviation *FOI*, standing for ‘Freedom of Information’, has put in rare appearances for some years, as discussions have come and gone to no avail about the need for such legislation. It does not occur in lower-case, since this form is dedicated to French *foi* (a loan applied to matters of faith and happily also misapplied to matters of the liver, as in *foi gras*), so case insensitivity need not be heeded in specifying the search string. Its frequencies increase significantly only from January 1, 2005, when the ‘Freedom of Information Act’ came into force.



Key

Dotted Line = Frequency per million words

Solid Line = Moving Average

Figure 5: Time-graph for *FOI/ FOIs* (836 occs.)

In Figure 6, a ‘heatmap’ graphically represents the collocational evolution of the abbreviation *FOI* over time, and shows frequency growth and peaking since 2005. The full contexts show that these peaks mainly reflect the large number of disparate requests being put into the FOI office by journalists and BBC programme teams seeking background for stories^{vi}.

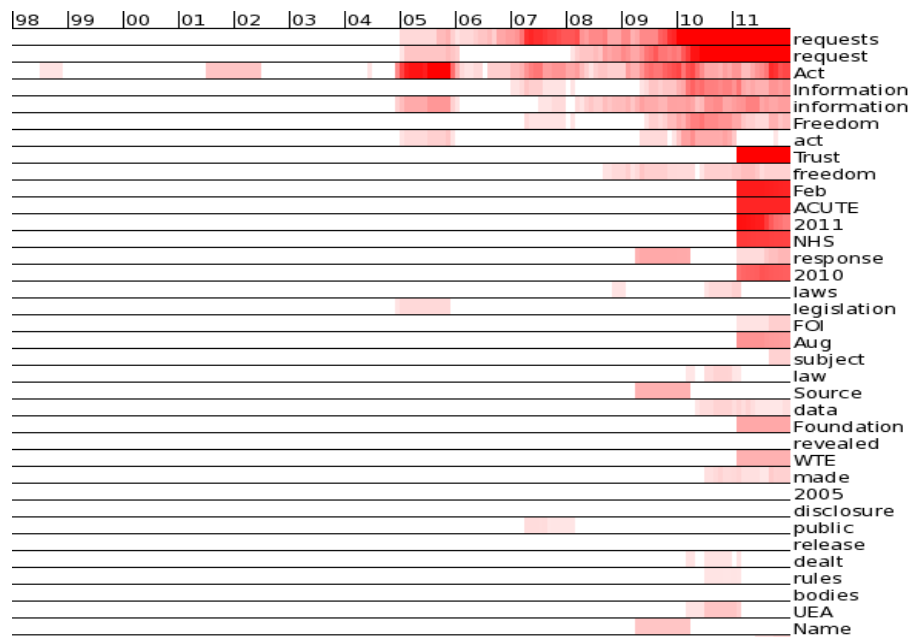


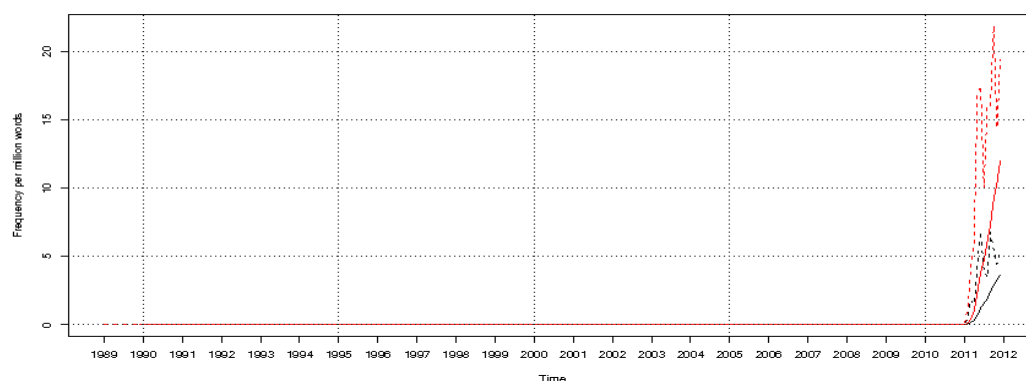
Figure 6: Heatmap for *FOI/FOIs* (839 occs.)

2.3 Stage in Life-cycle: Orthographic adjustment

As a word becomes more frequent in the corpus, it can change its manner of representation. A technical term, as with *Graphene/graphene* above, may convert to a lower-case initial word. A fully lexical term may be reduced to an abbreviation or an acronym. Abbreviations and acronyms may in turn shift from upper-case format, to a lower-case word with initial capital, or to a completely lower-case form. (We saw a counter-example with FOI, where there is a constraint set up by the existence of the unrelated but active lower-case form, *foi*.)

This orthographic progression is linguistically significant, since it is part of a process of lexicalisation and ‘de-acronymisation’ by which language users cease to perceive the word as an abbreviated phrase and see it as an ordinary lexeme, with all the cognitive ramifications of

that shift. Modern English journalistic text is increasingly populated with abbreviations and acronyms (notably in the public services, business and technology, where new terms are endlessly churned out), some acronyms being reverse-engineered to fit an existing, resonant word. They are probably fostered further by the commonplace use of these forms in social media). The orthographic variation associated with acronyms is also formally noteworthy, in terms of achieving maximum recall in searching for variants.



Key:

Arab spring: upper dotted line reaching 20+ wds. per mill. in late 2011

Arab Spring: lower dotted line reaching only 6+ wds. per mill. in 2011

dotted lines = frequency of term per million words

solid lines = smoothed lines of moving average frequency

Figure 7: Word History Plot for *Arab Spring/Arab spring*

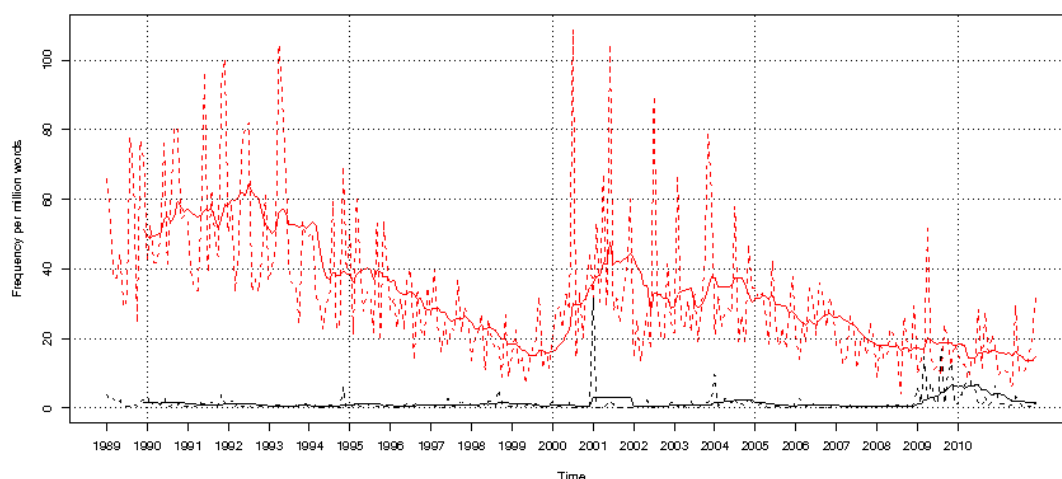
Figure 7 illustrates the frequency of the orthographic variants for the newly popular proper noun *Arab Spring*, from their common launching point of 2011. It shows that the

orthographically hybrid *Arab spring*, with 1194 occurrences, is surprisingly the favoured one. *Arab Spring* itself lags behind, at 379 occurrences. The original concordances show that *Arab spring* was the first variant and dominates throughout the corpus; until Dec 29th 2011, after which all instances of *Arab Spring* are printed in upper-case. This looks more like house style and less like a general trend in the language. The concordances also show that the search string [*Arab Spring**, *case insensitive*] also yields one instance of lower-case *arab spring*).

2.3.1 Orthographic change: AIDS

The technical acronym *AIDS* (‘Acquired Immune Deficiency Syndrome’) is neither a neologism, nor currently in vogue. However, its relatively high frequency of occurrence across time makes its word history plot easier to interpret than that for the newer *Arab Spring*. Figure 8 shows that the upper case form *AIDS* relinquishes the field to lower-case, word-initial capital *Aids* at once. Thereafter both orthographic forms are maintained, though *Aids* is hugely favoured.

Concordanced contextual information reveals that there are as yet no instances of *aids*. This variant is presumably blocked by the existing homograph *aids*, which is already doing dual service as the verb meaning ‘help’ and the plural noun meaning ‘sources of assistance’.



Key

upper lines = *Aids*; lower lines = *AIDS*

dotted lines = frequency per million words

solid lines = moving average

Figure 8: Word History Plot for *Aids* (37751 occs.) *AIDS* (1947 occs)

The OED intuitively judged this acronym to be a candidate for permanence in the language, and so far, they seem to be right. Like other acronyms referring to real-world events, it has an initial spike of popular concern. It then settles down, but at a respectable frequency level of average 30 words per million, so it remains ‘in the news’.

2.4 Stage in Life-cycle: Productivity

Productivity is an active, living quality in the language which is realised in the creation of newly derived and inflected variants of a word across time. There is uncertainty as to where

precisely the quality of productivity resides: whether an attribute of the affix, the morphological process itself, the underlying rules, or of the language system overall. Bauer discusses this in detail (e.g. 2001: p11ff). In our observations of language evolution in our diachronic corpus, we find it useful to think of the word itself as being productive, whenever the first occurrence of a word is swiftly followed by associated inflected and derived forms. But the affixes which typically accompany a neologism are often themselves productive: the commonest affixes (e.g. *-ness, un-*), and the more fashionable ones (e.g. *uber-/über-, -esque*).

Saussure's (1969 -1915: 228) view, cited by Bauer (2001, 14), at first appears to back us up. He says: "On pourrait classer les mots d'après leur capacité relative d'en engendrer d'autres" ('one could class words according to their relative capacity to engender others'), a view dismissed by Bauer as "idiosyncratic... or possibly just imprecise". But Saussure continues: 'selons qu'ils sont eux-mêmes plus ou moins décomposables' ('depending on the degree to which they are themselves analysable').

We depart from this view in that we find it useful to think of simplex words as also being productive. But it is more complicated than this, because, as Saussure himself implies in the same context, a new derivation may still not be formed by analogy with the complex word in question, but modelled on another.

There is not time in this paper to do the debate justice, since our focus here is on the fact that certain kinds of inflectional and derivational variants of a particular neologism often accompany or succeed it (or, indeed, precede it), rather than on which of several processes might have been at work.

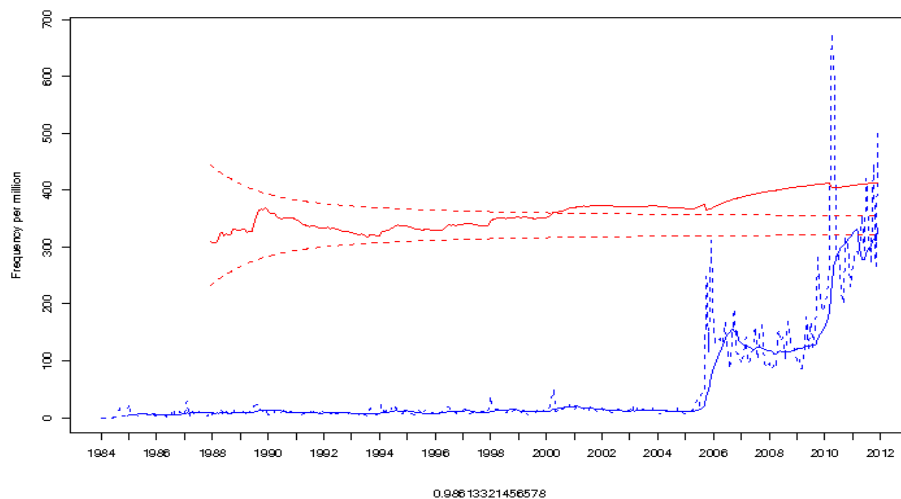
Some words (as well as abbreviations and acronyms) spawn more variants than others. The abbreviation *FOI*, standing for ‘Freedom of Information’, which was illustrated in Figure 5, appears to lack potential for productivity, so far yielding only:

FOIs, FOI-able, FOI-originated, FOI-exempt

This would be expected, since it is a term with restricted potential for application outside its narrow domain of reference. It is also, though very topical, not intrinsically amusing, an important factor in the fortunes and longevity of a word in journalistic text.

2.4.1 Productivity: *Cameron*

The proper noun *Cameron*, though in itself also not amusing, is prominent in the news, and does generate variants, as shown in Figure 9. The word is sufficiently frequent in our data to allow a ‘trend analysis’ to be done, and this shows as a ‘funnel’ containing a solid line. The direction this ‘trend’ line takes shows the significance of the change in frequency of the neologism, with any movement upwards out of the funnel indicating a definitely significant growth, such as we see in Figure 9 for *Cameron*.



Key

Dotted Line = Frequency per million words

Solid Line = Moving Average

Trend Analysis

Figure 9: Time-graph for [Cameron|Cameron's] (107176 occs.)

The low level of frequency shown in the years from 1984 to 2005 in Figure 9 is accounted for by the mention of all sorts of people named 'Cameron'. The first reference to David Cameron, Leader of the Conservative Party and later the Prime Minister from May 2010, appears in our data in 2005:

15/04/2005 "At a time when the Conservatives have been accused of looking backwards, **Cameron's** musical tastes prove him to be reasonably forward."

Thereafter, however, *Cameron* or *Cameron/s* in our data refers almost exclusively to him. From April 4 onwards, associated derived forms, compounds and hyphenated phrases begin to appear in our text, as listed in Table 1.

Cameronian	29	Cameronesque	4	Cameron-Hague	2
Cameron-Clegg	18	Cameronia	3	Cameronesquely	1
Cameron-led	16	Cameron-like	3	Cameron-speak	1
Cameron-style	12	Cameron-supporting	3	Cameron-Thing	1
Cameronism	7	Cameron-esque	3	Cameron-Tory	1
Cameron-Osborne	5	Cameron-cum-Clegg	2	Cameron-Brown	1
Cameronite	4	Cameronistas	2	Cameron-Blair	1

Cameronus	1	Cameron-Lite	1	Cameronomics	1
Cameronclegg	1	Cameron-inclined	1	Cameron-Cleggs	1
Cameronmoonface	1	Cameron-era	1	Cameron-ism	1
Cameronite-renegade		Cameron-Osborne-Clegg	1	Cameron-backed	1
-Blairite-BBCite	1	Cameron-sanctified	1	Cameron-endorsed	1
Cameron-Cable	1	Cameronised	1		
Cameron-connected	1	Cameron-in-			
Cameron-lite	1	sheep's-clothing	1		

Table 1: Early productivity of *Cameron* (*Guardian*, 09/10/2005 -28/07/2011)

Many of the variants in Table 1 have the look of neologisms created in the service of information and discourse structure. However, a brief look in Figure 10 at the contexts of some of the items in Table 1 indicates that David Cameron, as the Prime Minister, and his coalition party colleagues, do also become the object of humour, irony and derision in the press.

05/10/05 In other fashion news, the **Cameronistas** have instituted a ban on wax jackets.

17/12/05 Someone had to cry, "I am your man, trust me." But their leader didn't. Now David

Cameron-in-sheep's-clothing has.

30/05/06 This, to cynical British ears, might sound ever so slightly David **Cameronesque**

17/08/06 Davus **Cameronus**. Commonly known as the Disingenuously Crested Tit.

12/05/07 Open-necked shirts. Sorry, guys. Your summer look has been Blair/**Cameronised** for ever

15/04/10 moved from the era when Clegg is described as **Cameron-lite** to one in which Cameron is known as Clegg-lite

Figure 10: Morphological variants on the proper noun *Cameron*

2.5 Stage in Life-cycle: Creativity

There is a fuzzy interface, as we used to say, between productivity and creativity, as can be seen when we look at the range of new formations built upon the new referent *Cameron* above. By 'creativity', we mean an actual creative rule break; the manipulation of a neologistic word – and particularly a neologistic phrase - to create metaphor, word play or a pun, of the kind which pepper newspaper headlines, and are favoured by jaded journalists.

While many traditional proverbs and other sayings in English lend themselves readily to such treatment, it is rarer among new multi-word phrases, which are often technical terms, and not in themselves particularly humorous. It is perhaps not surprising that we find little play with the new phrases *Seasonal Affective Disorder* and *Carpel Tunnel Syndrome*, the entire repertoire so far being presented in 2.5.1 and 2.5.2:

2.5.1 Creativity: Seasonal Affective Disorder

08/12/09 After 25 years of **seasonal gender disorder**, Taylor has handed over the bloomers to Elkington, who has become Nottingham's new dame.

23/12/10 sufferers from acute **seasonal mobility disorder**

2.5.2 Creativity: Carpel tunnel syndrome

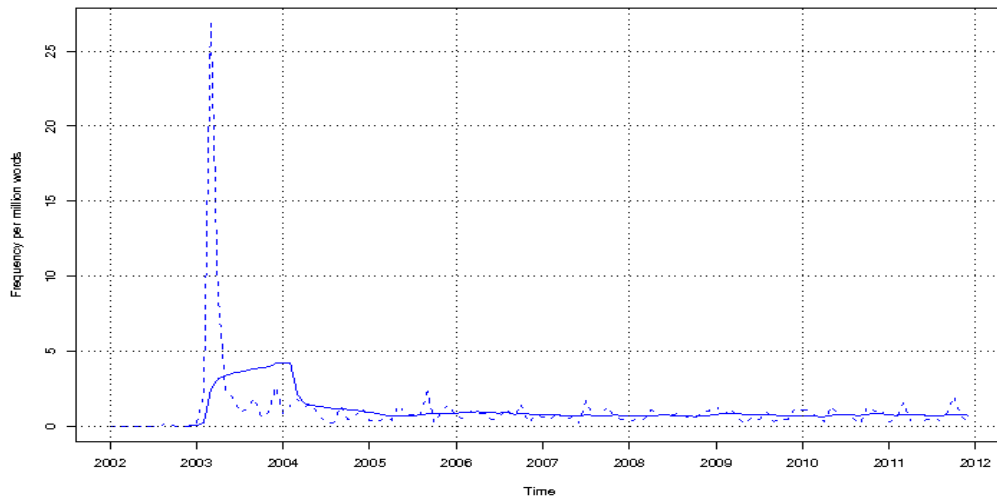
14/08/06 football folk are not alone in suffering from **furious tunnel syndrome**

The type of neologism which particularly lends itself to creative play is a popularised scientific term, such as *Mad Cow Disease*, or a colloquial and descriptive phrase, such as *weapons of mass destruction*. These two have been immensely, perhaps uniquely, creative over the years. They have the potential for humour or irony, and more prosaically, they contain sufficient lexical words (many terms being simply of the 'X of Y' pattern) to sustain some lexical substitutions while maintaining a recognisable framework of lexical reference for interpretation.

2.6 Stage in Life-cycle: Settling down

The next stage in our lexical 'life-cycle' is that which follows the first appearances and the initial growth and possible period of intense media interest. In Figure 3, we saw the term *Arab Spring* surge in popularity and frequency, then drop and flatten off, but not disappear. A similar pattern is next revealed in Figure 11 for the term *shock and awe*. This is the colloquial alternative for the technical term *rapid dominance*, both terms coined in 1996 to refer to a military doctrine instrumental in subsequent US war strategy. The military strategists^{vii} wrote: "rapid dominance will impose this overwhelming level of Shock and Awe against an adversary on an immediate or sufficiently timely basis to paralyze its will to carry on..." so that *shock and awe*, meaning 'a state of helplessness and lack of will', was technically the result of *rapid dominance*.

2.6.1 Change Type: *shock and awe*



Key

Dotted Line = Frequency per million words

Solid Line = Moving Average

Figure 11: Time-graph for *shock and awe* (667 occs.), case insensitive

Both *Shock and Awe*, and *shock and awe*, were allowed by a ‘case insensitivity’ search of our data, but the output was dominated by 538/667 lower case instances of the term. The term did not appear in our UK text until 2002, when US planning of the Iraq invasion was first reported on in UK newspapers:

06/09/02 One way is to march straight to Baghdad, blowing up everything in your way and then by **shock and awe** you cause the regime to collapse," Pike says.

The next mention was 5 months later, in similar vein, leading into a period of about 14 months of solid and frequent reportage. Since the term spawned so many like-named US

patents and products, it could have been expected to flourish for the duration of the war. Instead, there were social and political backlashes both in the US and elsewhere, questioning both the efficacy and ethics of the strategy, and the use of the term has thus subsided to a fairly stable if rare presence, at an average of 1 occurrence per million words, as seen the long tail in Figure 11. From early on, *shock and awe* came to be used metaphorically more than literally, to refer to a dramatic effect in almost any sphere of life, from architecture to advertising. By mid 2010, it is being used to criticise the EU governments' dealing with the financial crisis, as in Figure 12:

19/05/10 The bond markets have not been convinced by last week's giant "**shock and awe**" bailout of Greece.

25/10/11 "**shock and awe**" will not come cheap: the spread of the crisis to Italy and Spain means Europe needs at least euro2tn in the kitty.

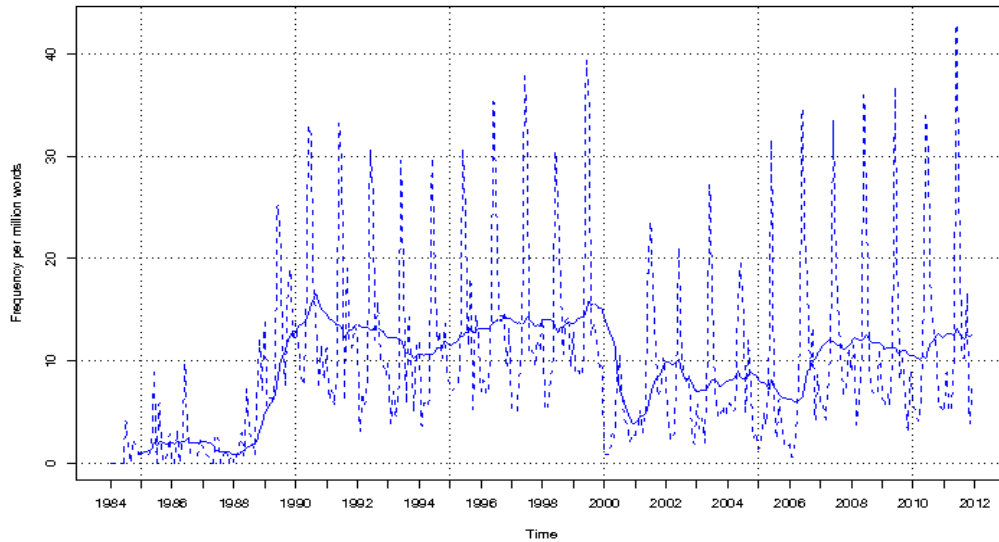
13/11/11 If Germany allowed the European Central Bank to print unlimited stocks of money, the ECB could **shock and awe** the bond markets into submission.

Figure 12: Contexts for *shock and awe* with reference to EU economic crisis

2.6.2 Change Type: *Ascot*

Ascot is an unassuming village in Berkshire, but, like the earlier-mentioned *Eyjafjallajökull*, its name has been appropriated metonymically to shift its reference from the immediate place, and extend it to encompass the local racecourse, founded in 1711, and thence the major horse races and associated events which are hosted there (see Halverson, 2012, for further discussion of metonymic extension and vagueness in place name use). In particular, Ascot

Racecourse stages Royal Ascot Week in June each year, a week of world-famous racing events attended by the Royal Family.



Key

Dotted Line = Frequency per million words

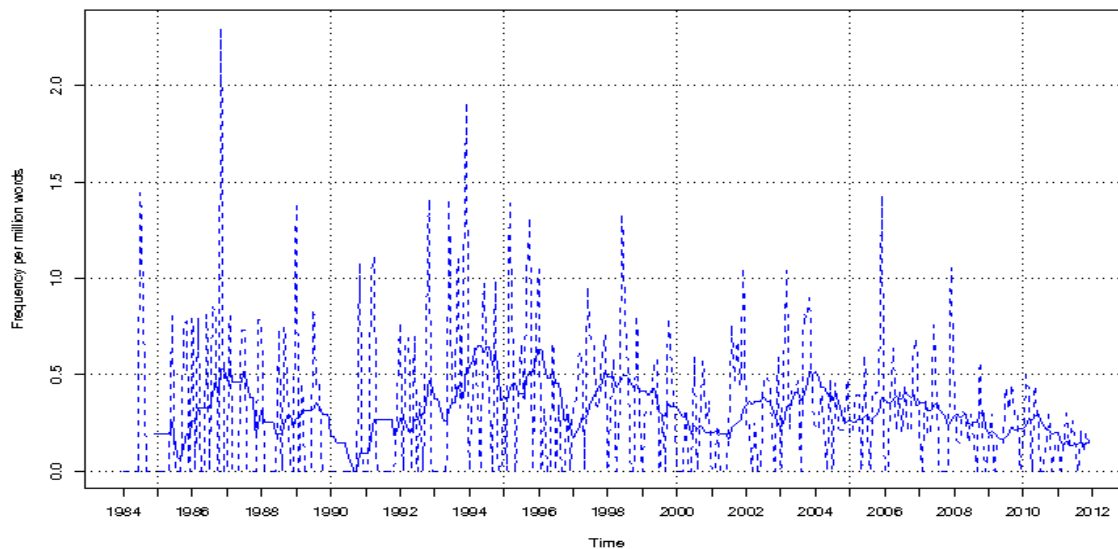
Solid Line = Moving Average

Figure 13: Time-graph for *Ascot* (12911 occs.)

In Figure 13, it can be seen that the occurrence of the proper noun *Ascot* peaks in a regular, annual pattern, suggesting that it is being used primarily to refer to the annual Royal Ascot racing event. The phrase *in Ascot* (as opposed to *at Ascot*) meanwhile, occurs a mere 149 times in our data, in which contexts *Ascot* refers to the actual village.

2.6.3 Change Type: *puce*

Another kind of change pattern across time is illustrated by the term *puce* in Figure 14. It displays a stable if lumpy profile – the spikes are actually of a very low frequency: only around 0.5 occurrences per million words.



Key

Dotted Line = Frequency per million words

Solid Line = Moving Average

Figure 14: Time-graph for *puce* (364 occs.)

This is an old-fashioned colour term, a shortening of the 18th French *couleur puce* or ‘flea colour’, and designating a colour in the purple range of the spectrum. As the term is patently no neologism, there is no initial spike in frequency. It continues largely in metaphorical use, in phrases such as *puce in the face*, *puce with rage*, though these sound increasingly dated today, and are probably generation-specific. There are a further 27 instances of upper-case initial *Puce*, of which most are irrelevant proper names, and a few refer humorously to the colour (including 5 references to a short film, ‘Puce Moment’):

23/02/93 Mr Branson revealed that the original plan had been to call the lounge the Mile High Club.

[Mr MacGregor turned] **Puce** to the hair roots, apparently

18/12/94 Hooray, it's a girl. But what to call it? Purple? Vermilion? **Puce**?

17/09/06 a single Black Wednesday or **Puce** Friday can open a wound that will not heal.

11/07/08 I ran out of colours to search for. Oh, I forgot periwinkle! Nope, not there. **Puce**?
Not there either.

2.7 Stage in Life-cycle: Obsolescence

By 'obsolescence', we mean 'the process of becoming obsolete or falling into disuse', and not 'the state of being obsolete'. This process can take place over centuries, decades or shorter periods.

2.7.1 Obsolescence: *video cassette*

Words which have been a staple for decades can slowly become out-dated. Others fade after a short period of jostling for position with competing claims to name a new concept. These are often technological terms, fading as technology advances. A few of the many examples are: *VHS tapes*, *VCR*, *dial-up Internet*, *CD*, *DVD*, *Walkman*. One such term is *video cassette*, its dwindling presence illustrated in Figure 15. The terms to which it has given way are *DVD*, and more recently *Blu-ray (disc)*.

The search string used to create the time-graph in Figure 15 was *video#cassette*, the wildcard symbol # allowing orthographic variants including *video-cassette* (hyphenated), *video*

cassette (separate word elements), and *videocassette* (solid compound) to be included. The variants are ranked as follows: *video cassette* (284 occs.), *videocassette* (43 occs.) and *video-cassette* (39 occs.); while the upper-case variants are a miscellany of rare items.

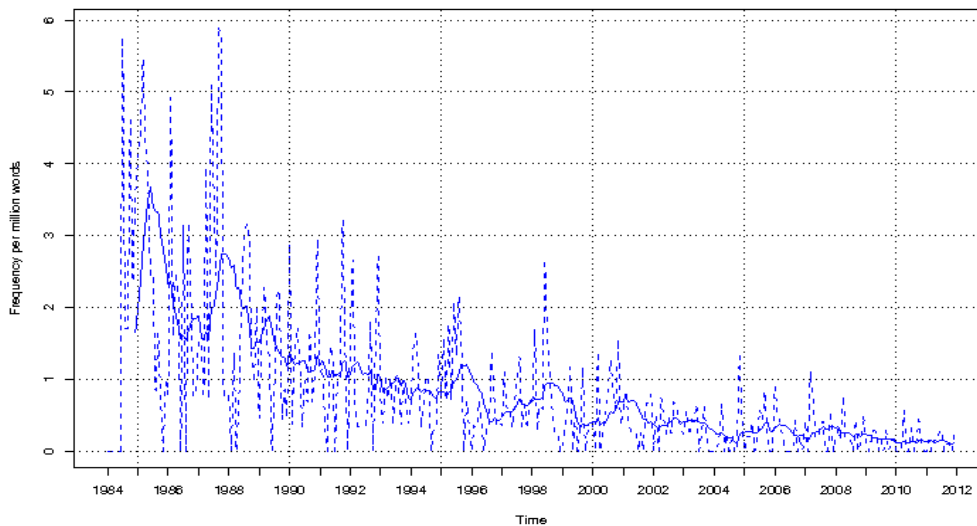


Figure 15: Time-graph for *video#cassette* (383 occs.), case insensitive

Some obsolescent terms have been found to return in new guise, such as *wireless*, which now plays a role in the new internet technology, in combination with new collocates *cable*, *networking*, *economy*, *technology*, *network* and *access* (Renouf, 2009). This is not the case, however, with *video cassette*.

2.7.2 Obsolescence: *dialling tone*

The term *dialling tone* (*dial tone* in the US), refers to a continuous sound which tells the hearer that a telephone is connected to the telephone system and is ready to be used. As indicated in Figure 16, this term is closer to disuse than *video cassette*. It is not showing a

downward curve but is simply stably rare – very rare, at around 0.1 occurrences per million words. This is probably because mobile phone users do not hear dialling tones. What keeps the term alive, as seen in the contexts of occurrence, is that there are occasional mentions of dialling tones in foreign systems, harking back to dialling tones in the past, and references to the absence of a dialling tone where one was expected.

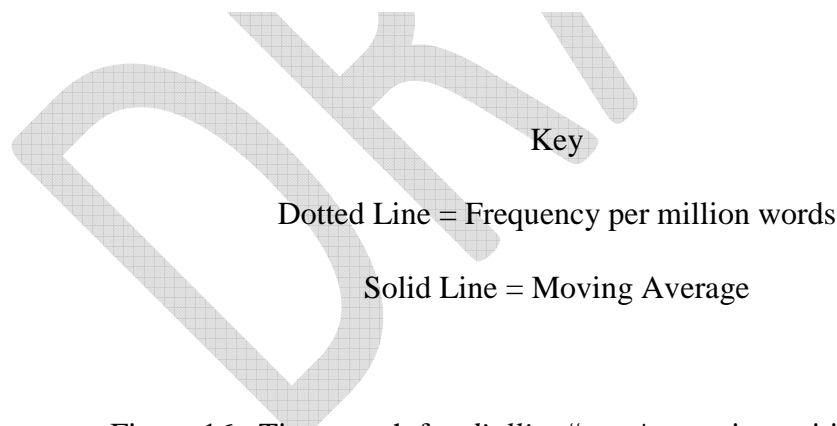
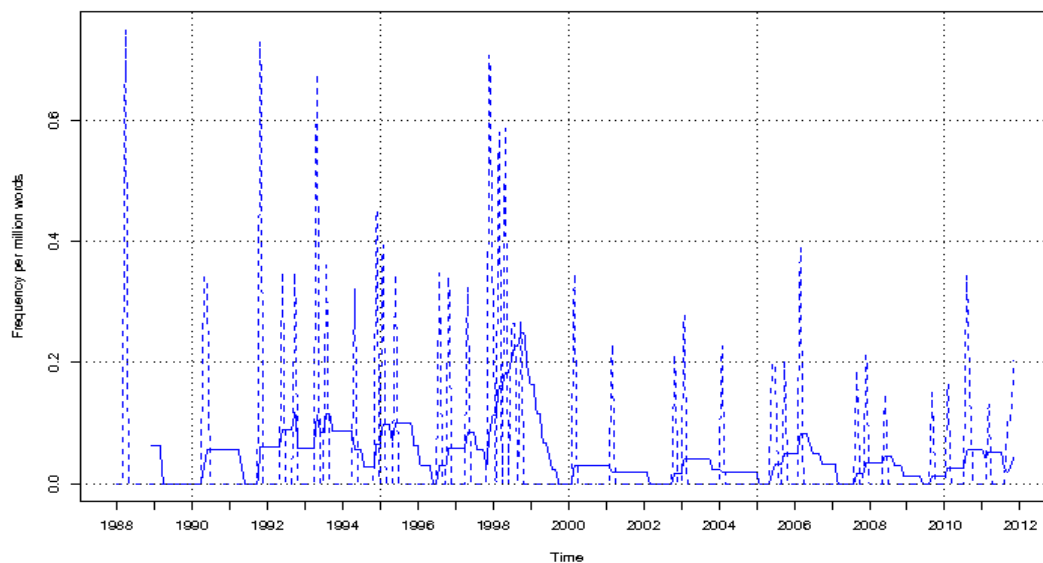


Figure 16: Time-graph for *dialling#tone**, case-insensitive (50 occs.)

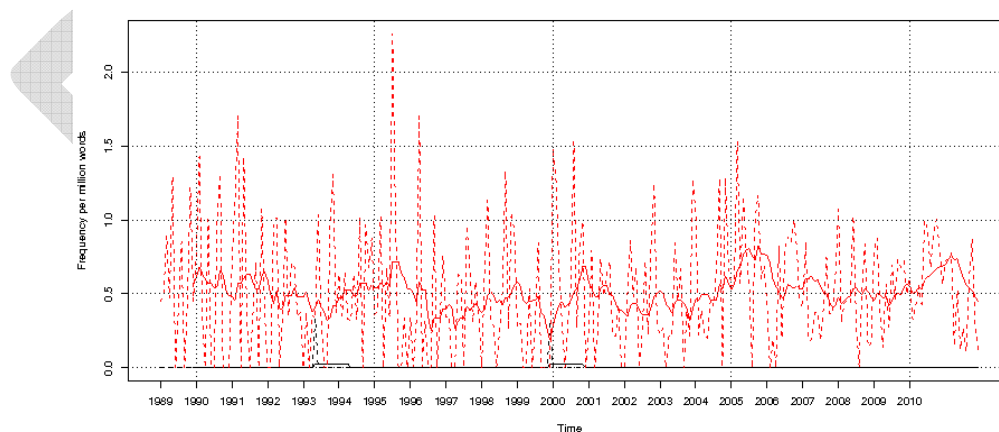
2.8 Stage of Life-cycle: Death

There is no set definition of ‘death’, though we have seen proposals that, to be dead, a word should not have occurred in a century, and so on. In our terms, it means that the occurrences

of a word are so few in our data that it is to all intents and purposes gone from the language. This is a stage on from ‘obsolescence’, where one sees a low frequency profile and a downward curve. It is where the pattern of occurrence is in fractions of an occurrence per million words, and where the word firmly hugs the bottom of the frequency graph across time.

2.8.1 Death: *obsoleteness*

The term *obsoleteness* is a perfectly good derived word, consisting of an adjective, *obsolete*, and a highly productive suffix *-ness*, which simply seems to have fallen out of use, and this certainly prior to 1984, when our monitoring began. Just two occurrences in 1.2 billion words of text are recorded for *obsoleteness*, each back in the 1990s, so that it is no more in daily language use, but replaced by the term *obsolescence*, which appears to have extended to mean both ‘moving out of use’ and ‘moved out of use’. The Word History Plot in Figure 17 compares the two trajectories.



Key:

Dotted Line = Frequency per million words

Solid Line = Moving Average

Figure 17: Word History Plot for *obsoleteness* (lower, almost flat line) and *obsolescence* (upper line) in *The Independent/Guardian*

The two instances of *obsoleteness* occur as follows, each in archaic contexts – the first concerning technology and the second the Church’s outdated model:

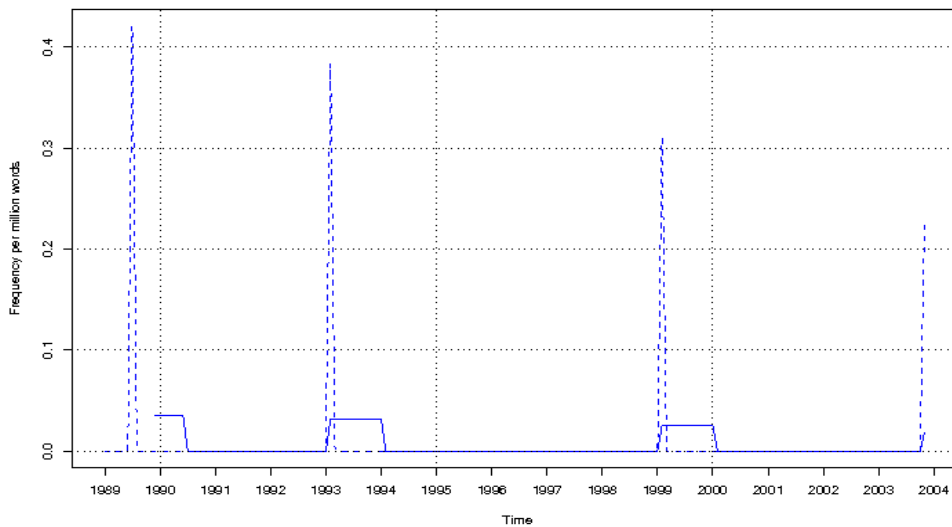
15/05/93 He associates opera with the vinyl disc, a fetish whose **obsoleteness** part of him recognises.

30/12/99 As for the Church of England, its form of establishment is inadequate to the point of crippling **obsoleteness**

2.8.2 Death: *donkey-brown*

Certain areas of the lexicon are more prone than others to obsolescence and death. We have demonstrated this with reference to technological terms, but it is also true, as we have mentioned *en passant* in previous studies, that colour terms, particularly fashion colour terms, fit this category.

In Figure 18, *donkey-brown* occurs only four times, and thus must be said to be all but dead.



Key:

Dotted Line = Frequency per million words

Solid Line = Moving Average

Figure 18: Time-graph for *donkey#brown*, case insensitive (4 occs.)

A complicating factor in our categorisation is that a word may be dead to all intents and purposes, but be revived by the media periodically for the purposes of irony or parody, and so on, as we see below.

1. 29/07/89 the sweep of the mountains, dappled khaki and **donkey brown**
2. 25/02/93 But I had my **donkey-brown** hipster loons, feather boas and Jimi Hendrix looks.
3. 27/02/99 ultra-simple porridge-coloured or **donkey-brown** notebooks encased in blanket thick felt

4. 29/11/03 the teak fancies upholstered in orange and **donkey-brown** stripes that decorated
our
homes

In context 1, the use of the colour term is licensed within the context of a poetic description of landscape; in 2 and 4, to bring humour in recalling past fashion; in 3, to bring a ‘retro’ feel to the description of merchandise.

Another colour term, *French navy*, contributes only two instances in 1.2 billion words. The use here is marked, to achieve irony on the topics of ‘gay speak’ and fashion jargon.

21/02/97 Gay men say 'taupe', 'crimson', '**French navy**', 'indigo', 'lime' and 'mauve

01/11/03 Take no notice when anyone tells you that scarlet (or **french navy** or winter white or mint green) is the new black.

3. Elaboration on the ‘life-cycle’ model: second comings

3.1 Stage in Life-cycle: Semantic neology

As we said in our introduction, our system exploits the surface patterns of text - specifically, a word and its collocates – such that any significant collocational change in a word’s profile during its life-cycle is deemed to mirror semantic change. By this method, semantic neology - semantic change in existing words – can be discovered. An evolving system of collocational change identification has been an integral part of our projects for the past 23 years. To discover semantic change in this paper, we have used the 1997-2000 APRIL project tools, as well as drawing on the 2004-07 WebCorplSE text processing and presentation tools.

The process involves building a ‘Collocate Bank’ containing a ‘Collocational/Collocate Profile’^{viii} of a specific span for each word, built from previous data. As the corpus augments, a new ‘collocational profile’ for each word in each new time-chunk is built, and this is compared with its existing profile in the Collocate Bank. The changes are stored until they reach a ‘significant change’ threshold (a minimum value calculated mathematically or statistically, which can be varied), at which point a sense change is deemed to have occurred. An example of a word which has accrued a new sense is *traction*.

3.1.1 Semantic change: *traction*

The word *traction* is an engineering term, in which domain it means ‘the maximum frictional force that can be produced between surfaces without slipping’. With our collocation calculation tool, we are able to identify a point in time when collocational change begins to occur with a specified significance measure, as shown in Table 2. The ‘tipping point’ (e.g. Gladwell, 2000) occurs at the start of 2000, and we are also able to scrutinise the emerging new collocates; in particular, *idea* and variants of the verb *gain*. The new collocational combination emerging, of ‘an idea gaining traction’, is clearly a metaphorical move into general language use by this previously technical term.

December 1999

control

suspension

engine

brakes

active

anti-lock

January 2000

steam

wheels

system

engines

electronic

electronic

handling	evidence
braking	electronic
drive	get
gain	idea
gaining	treatment
gained	pain
engine	getting
engines	steam
gains	political

Table 2: AVIATOR Collocate Bank output showing tipping point in significant collocational change, for the word *traction*

3.1.2 Semantic change: *viral*

The term *viral* is a medical term meaning ‘of, relating to, or caused by a virus’, where *virus* means ‘the causative agent of an infectious disease’. The heatmap tracing collocational change over time in Figure 19 corroborates this, showing *virus* in significant collocation with *infection, infectious and disease* from 1990 until the present day.

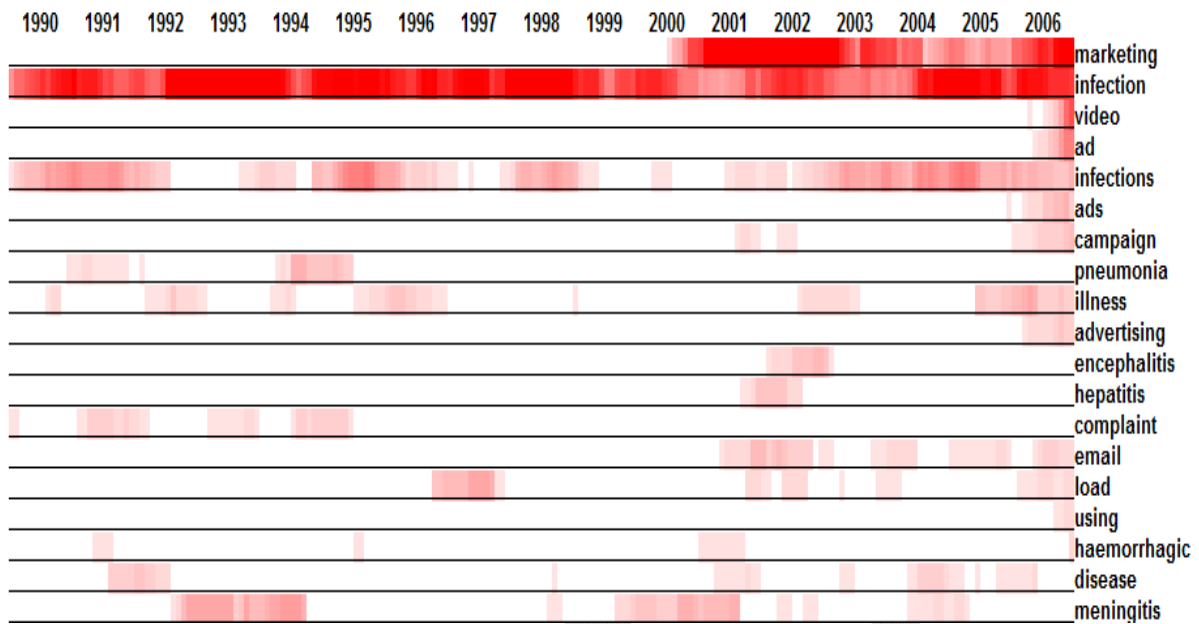


Figure 19: Heatmap for the term *viral*

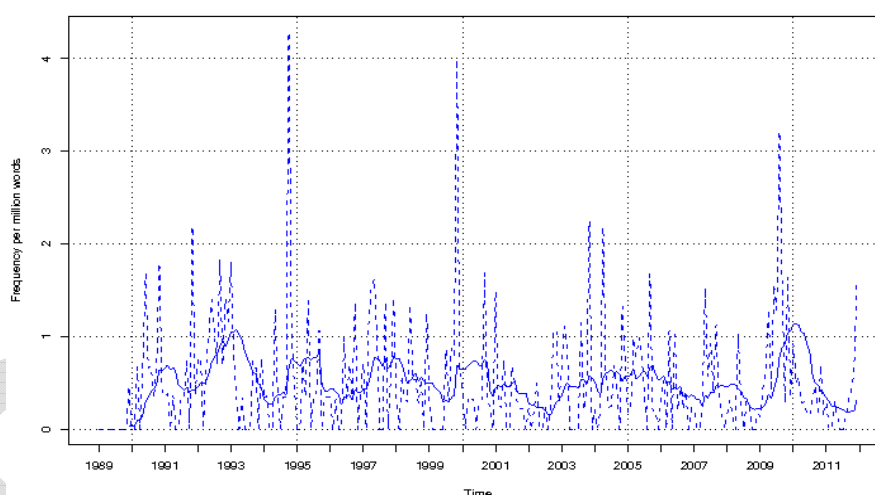
From 2000 on, however, the heatmap further reveals a new set of collocates emerging, including *ad*, *campaign*, *advertising*, *email*, and most significantly, *marketing*. Thus, we now have a series of emergent new multi-word units, *viral marketing*, followed by *viral ad/s* and *viral advertising*, *viral email*. What has happened is that the metaphorical sense of *viral* as ‘infectious’ has been modified to mean ‘swiftly passing from one (host) to another’. Thus it now can be construed to mean ‘using pre-existing social networks (such as Facebook, YouTube) to market product’.

3.2 Stage in life-cycle: Revival

We have said that there is a cyclical element in the life-cycle of a word; a word or phrase can seem to die or remain dormant but then return to active duty when real world events or media whim demand. To illustrate this, we take the phrase *velvet revolution*.

3.2.1 Revival: *Velvet Revolution*

Words which die and later return to active use do so over varying time-spans and gaps, many too long to be captured within our data covering only 1989-2011. Poll tax, for example, was levied in the 14th and 17th centuries in England, and introduced again in the 20th century; the full change pattern for the term *poll tax* can thus not be traced graphically using our system. We are conducting short-term diachrony, or ‘brachychrony’ (Mair, 1997). Such a long record would be more likely traceable through a text collection of the scope of the Helsinki Corpus (e.g. Rissanen, 2005), or across the BROWN family of corpora (Leech et al, 2009).



Key

Dotted Line = Frequency per million words

Solid Line = Moving Average

Figure 20: Time-graph for *velvet revolution** (573 occs.), case insensitive

The shorter-term change which we can monitor is exhibited in Figure 20 by the term *Velvet Revolution* (270 occs.; *velvet revolution*, 250 occs.). As with *Arab Spring*, the term *Velvet Revolution* refers to a series of similar events, but most dramatically to two sets of events: the earlier East European movements for democracy, and the corresponding later moves towards freedom in the Arab world. The graphic continuation of occurrence of the term obscures the time-gap between the real-world events, since journalistic reference to the earlier *Velvet Revolution* runs into and overlaps with the second. Nevertheless, though the two elements, *velvet* and *revolution*, of this neologistic compound have retained their literal meaning, we would argue that the referents of the earlier and more recent uses are sufficiently distinctive that the term could be said to have been quasi-coined twice.

During and after the Czechoslovakian revolution, the events were largely reported on as positive, particularly by Western journalists, with connotations of 'soft', 'gentle', 'peaceful', drawn from the physical quality of the fabric velvet, as shown in Figure 21.

01/08/09 “**velvet revolution**”. The phrase comes from the peaceful 1989 **velvet revolution** which overthrew decades of communism in Czechoslovakia.

01/08/09 Prosecutors read out an indictment outlining a plot by pro-reform parties to carry out a

“**velvet revolution**”, a popular, non-violent uprising to overthrow the Islamic Republic

02/08/09 a plot to carry out a “**velvet revolution**” to overthrow the Islamic Republic, similar to the

largely peaceful revolts that ended communist rule in eastern Europe.

25/08/09 The defendants were charged with fomenting riots and attempting to overthrow the ruling

system with a “**velvet revolution**” similar to uprisings in eastern Europe

Figure 20: Contexts for *velvet revolution* 2009: retaining the original 1989 sense

From 2009 on, the term re-emerges, used of Middle-Eastern government transition, but often in the context of what is seen as the evil in the East European reform. This is again a *velvet revolution*, but this time, *velvet* lends just one of its attributes, its ‘slipperiness’, and the term means ‘clandestine’, ‘stealthy’, and ‘foreign-orchestrated’ overthrow. This connotation is conferred by Iranian and other fundamentalist opponents, and picked up in Western reportage, as exemplified in Figure 22.

04/09 a conviction that western countries are attempting to overthrow the Islamic regime through a “**velvet revolution**” spearheaded by NGOs, charities and other international organisations

06/09 “**velvet revolution**” as the work of “Zionists” and “western media”

06/09 The supreme leader equates the demand for women’s equality with a western-backed “*velvet revolution*”

08/09 a “**velvet revolution**” a phrase from eastern Europe, used by the Islamic regime as shorthand for foreign-orchestrated subversion.

01/10 officials have accused “counter-revolutionary elements” with ties to the US and Britain of master-minding a “**velvet revolution**”

02/11 Khamenei depicts the Green Movement as puppets of a western-backed conspiracy bent on toppling the Islamic system in a “**velvet revolution**”

09/10 collaboration with foreign countries was the claim made by the Iranian authorities when they identified the womens' rights movement as being behind an attempted “**velvet revolution**”

Figure 22: Contexts for *velvet revolution*: new semantic value attributed

Thus we have two referents and, cutting across these, two connotations, depending on author. A different metaphorical spin comes to the meaning of *velvet*. Our corpus data thus lead us to question whether this is a reappearance of an existing phrase with a new connotation; or whether it is to some degree a new term with a new meaning.

4. Closing summary

Lexicographers, translators and terminologists need objective, measurable information about language use and language change which can allow them to identify and deal with neologisms systematically. We have proposed that the definition of neologism is more subtle and complex than it might seem when viewed intuitively, and that the tracking of a word in a diachronic corpus from its point of entry into written or spoken text, and subsequently, can reveal much about it which is otherwise inaccessible.

We have proposed that a word has a ‘life-cycle’, and have striven to illustrate the stages of that cycle with reference to a selection of neologisms. It is hoped that this exposition and analysis will help teachers and researchers, as well as lexicographers, translators and terminologists, to anticipate and make judgements about the status of a new coinage, and the likelihood of its being institutionalised in the language; as well as being alert to what it means and how that meaning might be changing.

5. References

- Algeo, John & Adele Algeo. 1988. Among the New Words. *American Speech* 63 (4): 235–236
- Andersen, Gisle. 2012. (ed.), *Exploring Newspaper Language - Using the web to create and investigate a large corpus of modern Norwegian*. Amsterdam: John Benjamins.
- Baayen, R. Harald & Antoinette Renouf. 1998. 'Aviating among the hapax legomena: morphological grammaticalisation in current British Newspaper English'. *Explorations of Corpus Linguistics: Proceedings of the 18th ICAME Conference, University of Liverpool, 21-25 May 1997* ed. Renouf, Antoinette. 181-189
- Bauer, Laurie. 2001. *Morphological Productivity*, Cambridge: Cambridge University Press.
- Davies, Ceri. 2004. A Corpus-Based Investigation Of Noun To Verb Conversion In English, unpubl. PhD Thesis, University of Liverpool.
<http://rdues.bcu.ac.uk/bibliog.shtml#PHD>
- De Smedt, Koenraad. 2012. Ash compound frenzy: A case study in the Norwegian Newspaper Corpus. Gisle Andersen. 2012. 241-56.
- Downing, Pamela. 1977. On the Creation and Use of English compound nouns. *Language* 53 (4): 810-842.
- Gladwell, Malcolm. 2000. *The Tipping Point*, Abacus, Little, Brown Book co., London.
- Halverson, Sandra. 2012. Metonymic extension and vagueness: Schengen and Kyoto in Norwegian newspaper language. Gisle Andersen 2012. 286-306.
- Hohenhaus, Peter. 2005. Lexicalization and Institutionalization, in *Štekauer & Lieber*. 363-365
- Leech, Geoffrey, Hundt, Marianne, Mair, Christian & Nicholas Smith. 2009. *Change in Contemporary English: A Grammatical Study*. UK: Cambridge University Press.

- Ljung, Magnus. (ed.) 1997. *Corpus-based Studies in English*, Amsterdam/Atlanta, GA: Rodopi.
- Mair, Christian. 1997. Parallel Corpora: A Real-Time Approach to the Study of Language Change in Progress. In Ljung (1997). 195-209.
- Renouf, Antoinette. 1993a. A Word in Time: first findings from dynamic corpus investigation. *English Language Corpora: Design, Analysis and Exploitation*, Aarts, Jan, de Haan, Pieter & Nelleke Oostdijk (eds.). Atlanta/Amsterdam:Rodopi. 279-288.
- Renouf, Antoinette. 1993b. Making Sense of Text: Automated Approaches to Meaning Extraction, in *Proceedings of 17th International Online Information Meeting, 7-9 December 1993*, 77-86.
- Renouf, Antoinette. 2007. Tracing lexical productivity and creativity in the British media: The Chavs and the Chav-nots?, in Judith Munat (ed.) *Lexical Creativity, Texts and Contexts*. Amsterdam: John Benjamins Publishing Company. 61-89
- Renouf, Antoinette. 2010. 'Identification automatique de la néologie lexicologique et sémantique : questions soulevées par notre méthode'. in Cabré, Maria-Teresa; Domènech, Ona.; Estopà, Rosa; Freixa, Judit; Lorente, Mercè (eds.). *Actes del I Congrés Internacional de Neologia de les Llengües Romàniques. Barcelona: Institut Universitari de Lingüística Aplicada; Documenta Universitaria, 2010*. 129-141. (1 vol. + CD-ROM ISBN 978-84-92707-08-9)
- Renouf, Antoinette. 2012. Defining neology to meet the needs of the translator: a corpus based perspective. *Neologica 6*, Humbley, John, and Jean-François Sablayrolles (eds.) Paris, Editions Classiques, Garnier.
- Renouf, Antoinette. forthcoming. A Finer Definition of Neology: from Word to Register? Proceedings of 33rd International ICAME Conference, Leuven. Publisher tbc.

Rissanen, Matti. 2005. *The Helsinki Corpus of English Texts: A window to the history of English*. Visnyk KNLU.

Saussure, Ferdinand de (1969) [1915], *Cours de Linguistique Générale*. Paris: Payot.

Štekauer, Pavol & Rochelle Lieber. 2005. (eds.), *Handbook of Word Formation*. The Netherlands: Springer

ⁱ actually a revived sense - see *Magna Carta*.

ⁱⁱ Church and Gale estimated a 10% error rate in news data. But this is of course a complicated issue: there are typographical errors of predictable and characteristic kinds, and our APRIL system weeds out the obvious candidates, e.g. according to basic rules of transposition and co-occurrence or absence; but there are manual spelling errors such as *accomodation* which are widely in use, and now recognised as bona fide variants by automated indexing systems. The most generous spin would be that today's error is tomorrow's new word, but that would be going too far, and errors certainly cause a great deal of post-editing where that matters. Fortunately, many errors are hapaxes, and so, though picked up by our system as candidate neologisms if they escape the error filter, they do not disturb our diachronic studies of later neologistic behaviour.

ⁱⁱⁱ http://www.merriam-webster.com/help/faq/words_in.htm

^{iv} Now was it a new concept. Back in 1859, the material yet to be called graphene was already studied by Brodie; in 1918, by Kohlschütter and Haenni; in 1947, by Wallace; and in 1966 by Hess et al.

^v as does reportage of the continuing threat to its development posed by stifling UK bureaucracy and research health and safety legislation; and copy-cat product development by global companies.

^{vi} These stories in turn reflect media concern about controversial plans to curb the UK debt crisis: such as proposed reforms of the BBC and the National Health Service (NHS), involving hefty cuts to WTEs (whole-time equivalent posts).

^{vii} Ullman, Harlan K. and James P. Wade. (1996). *Shock And Awe: Achieving Rapid Dominance*. National Defense University, 1996, XXIV.

^{viii} *Collocational /Collocate Profile* and *Collocate Bank* were terms which we coined for the main components of our methodology. It only became possible to create such things in late 1980s, the software being created then by Alex Collier. These terms may seem generic and obvious now, but they denoted unique and original concepts when coined.