

Defining neology to meet the needs of the translator: a corpus-based perspective

Abstract

Since 1980, when text processing began in earnest, we have been studying patterns of neologistic behaviour across time. The knowledge gained has led to the development of automated systems which can record neologisms chronologically in news text since 1984. The systems identify new words and multi-word terms, semantic changes, and newly occurring synonyms for existing words.

The definition of neology varies for each user group. Part of the translator's task is to render the source text as closely as possible, in all its originality. This requires knowledge of, or at least access to, new features in the language of translation. It also entails the ability to use circumlocutions where a translation term does not exist, or is not known.

In this paper, we take a corpus linguistic approach to neology, both lexical and semantic, and try to demonstrate how this can be revealed to the translator through the automatic analysis of modern English diachronic text. The first part of the paper illustrates patterns of neology and change which can be detected; the second lays out the stages in the life-cycle of a neologism subsequent to its first occurrence.

Key words: lexical neology, semantic neology, formal neology, coining, neologism, language change, diachrony, life-cycle, productivity, creativity, translation, translator

Introduction

In this article, we use the term 'neology' to mean the study of new words, and 'neologism' to mean a newly coined word which may well become established in the language in time, and is not an obvious 'nonce formation'ⁱ. We generally avoid the term 'neonymy', because our source data are mainstream daily newspapers, not specialist texts, though they do report on popularised technical domains as well as lay and technical terms.

The study of neology in Britain has been a rather ad-hoc affair. As the reader will know, no official UK body monitors the evolution of English, or tries to standardise the incoming coinages and loans. The chief protagonists in documenting neonyms and neologisms have traditionally been the dictionary makers. Until recently, inventories of neologisms have largely been dictionary updates, (Cf. Algeo's full account of US neology, 1993). The Merriam Webster dictionaryⁱⁱ was, until May 2011, among those who on their sites still talked of reading and marking texts manually for neologisms.

But most major UK dictionary houses have by now followed the Collins-Cobuild lexicographic approach of the early 1980s, in accessing synchronically-held text corpora as a reference to word use. Collins WordBanks Online speakⁱⁱⁱ of "the unique resources of the Collins Word Web", in which "the majority of sources in the corpus date from between 2001 and 2005". Pearson ELT state that "all Longman dictionaries are compiled using the Longman Corpus Network – a huge database of 330 million words from a wide range of real-life sources such as books, newspapers and magazines".

Oxford Dictionaries Online^{iv} have a kind of ‘monitoring’ system of hybrid texts and notes across time, though they do not have the unbroken flow of text in which progress can be exhaustively and quantitatively recorded.

With the advent of the internet and new social communication opportunities, web sites have sprung up on which individual word enthusiasts and specialist groups gather new words as they come across them, though many of these, such as *Urban Dictionary*, tend to be US-originated. In the UK, there are a dozen or so translation associations, the main professional body being the Institute for Translators and Interpreters (ITI). This has several thousand members internationally, but offers no archival or other communal data infrastructure, the initiative falling to its informal regional member sub-groups, though it apparently does host small, specialist networks, one on terminology.

In this paper, we shall take a corpus linguistic approach to neology, both lexical and semantic, and try to demonstrate what can be revealed by the analysis of modern English diachronic text across time. In Part 1, we shall present the research background, reviewing our own work and the position of the translator. In Part 2, we shall consider neology in the context of its automatic identification in text, from the point of view of the translator. In Part 3, we shall develop a finer-grained diachronic definition of neology, again with the translator in mind.

Part 1. Research Background

1.1 Research and Development Unit for English Studies

Since 1990, when text processing was established but before the internet, we have been alone in studying patterns of neologistic behaviour, both lexical and semantic, systematically across time. This has been within the larger framework of our long-term research goal of exploring the relationship between regular surface patterns of word collocation, and word and text meaning; and of automating this study. Our research has been funded competitively by the UK Engineering and Physical Science Research Council (EPSRC) for its potential to improve understanding in language description, and performance in large-scale textual database access and management. Our flow of projects involving neology is as in Table 1:

Start Date	Project	research goal – the identification of English:
1990	AVIATOR	lexical/semantic neologisms
1994	ACRONYM ^v	nyms (synonyms, and other lexical sense relations)
1997	APRIL	lexical/morphological productivity
2000	WEBCORP	language use in live web text
2004	WEBCORPLSE	language use and change in web-derived corpora
2006	LEXICAL REPULSION	repulsion/close synonymy

Table 1: RDUES Projects dealing with neology, from 1990

The start dates are inevitably pre-dated by an initial design of a methodology. The projects have been conducted within the Research and Development Unit for English Studies^{vi}, and

are reported on in publications and demonstrated on the RDUES website at <http://rdues.bcu.ac.uk/>.

1.2 Diachronic corpus linguistics: the state of the art

A synchronic approach to English language study is often appropriate, and many synchronic corpora, both modern and historical, abound. But for the study of 'modern-day' English language use, the static, snap-shot corpus can only reveal current usage for so long^{vii}. There are mainstream conventions which hold, but at the margins, the language is always quietly and irrevocably moving on, and this subtly re-shapes the conventions over time. Neology is intrinsically diachronic. Neologisms will only appear in textual data which is up to date and up-dated; and their changing status can only be detected in chronologically-processed textual data. For these reasons, diachrony has underpinned our approach to language study since it became technically feasible in the late 1980s.

The awareness of the need for access to diachronic data for the study of neology is finally picking up steam elsewhere. The EnerG (English Neologisms Research Group) led by Schmid at LMU, Munich, is taking the neologistic updates from the OED and tracing their spread across the Internet. In Norway, the recent Norwegian NNS project in neologism and term detection (Andersen, this volume)^{viii} is well under way. In Barcelona, there is a new move by Teresa Cabre and her team into diachronic text monitoring at Pompeu Fabra^{ix}.

In France, the debate about the role of diachrony in the study of terminology has persisted across the decades. Humbley (forthcoming)^x points out that Guilbert (1973), a pre-corpus linguist, defined terminological innovation as both synchronic and diachronic, speaking of "une synchronie dynamique" (p.12), an idea echoed by Rondeau (1984: 121). Thirty-six years later, Dury and Picton (2009) eloquently enumerate the inhibitors which still hinder practical progress in the diachronic study of neology as: "les obstacles historiques, pragmatiques et psychologiques qui peuvent expliquer la frilosité des terminologies à l'égard de la diachronie", but they do go on to discuss possible ways forward in semi-automated diachronic text study.

The diachronic dimension is brought to our work in two ways: firstly, through the chronological capture, storage and processing of texts printed in the UK *Guardian* and *Independent* newspapers, providing an unbroken flow of data from 1984 to the present day. A twenty-six year period of textual activity is short (what Mair, 1997 calls 'brachychrony') in the context of long-term diachronic corpora held by such research centres as VARIENG^{xi} at Helsinki, which begin with the earliest manuscripts. It is also short in comparison with the century of text which is covered at thirty-year intervals by the 'Brown Family' of small, comparable corpora^{xii}. But it has the merit of allowing the unbroken tracking of language change across the last three decades of modern English. A corpus of news data is not a specialist corpus, but it does complement the specialist fields of terminology which are more widely published and read, particularly on the internet, including business, management and IT/ICT. It also illustrates the contexts and usage surrounding technical and semi-technical terms across a range of registers.

The diachronic dimension to our work is also achieved through the successive development of innovative processing software and lexical statistical methods^{xiii}, which have formed part of the above-mentioned projects.

The projects immediately relevant to this paper focus primarily on neology: AVIATOR (Analysis of Verbal Interaction and Automated Text Retrieval; 1990 – 93); and APRIL (Analysis and Prediction of Innovation in the Lexicon; 1997-00). The WebCorp (Web as Corpus, 2000-03) and WebCorpLSE (WebCorp Linguist's Search Engine, 2003-2007) projects also provide access to web-based neologisms which supplement those found within a newspaper corpus.

Project AVIATOR (Renouf, 1993) was dedicated to testing the hypothesis that new words can be tracked by matching new items against existing lexical inventories, and new senses by matching new collocational profiles against existing ones. Project APRIL (e.g. Baayen & Renouf, 1998) tested the hypothesis that the productivity of the language could be identified and classified at the level of hapax legomenon – that is, by treating all new words at their first point of entry – and extrapolating from the findings a set of words and sub-word morphemes which could be anticipated to recur, and thus be considered, for example, as useful candidates for an enlarged bank of search terms that could break the bottle-neck caused by the traditionally limited search term indexes. These AVIATOR and APRIL research hypotheses were ground-breaking and risky, both linguistically and technologically.

The WebCorp projects have also been ground-breaking, though simpler in concept. It was clear from 1997 that tailored diachronic corpora were too expensive to create and update, so the WebCorp project (2000-2003) (Renouf, 2003; Renouf, Kehoe, Banerjee, 2006) was conceived. This applied our previous experience of corpus processing to web-based text. WebCorp finds, examines and processes web text in real time, supplementing the view of language provided by our newspaper corpus both in immediacy and in text type. WebCorpLSE (2004-2007) (Kehoe & Gee, 2009; Renouf, 2009) followed pilot work done back in 2001-1 with a pioneering company, Searchengine.com^{xiv}, which had indicated that creating our own search engine and downloading a multi-billion word, growing newspaper corpus would allow us to monitor language change in a more statistically and chronologically reliable way than was possible with WebCorp's reliance on Google.

1.3 The needs of the translator

Our study of neology has largely been conducted from our own point of view; that of the corpus linguist, for whom any new word or use is of potential interest.^{xv} But we realise that the definition of neology varies for each user group. It is difficult to specify the needs of the professional fields involved. Practice varies between private, individual practitioners and those within the framework of government or commercial agencies. For the lexicographer, a new lexical item or new word sense formerly needed to show signs of assimilation before it warranted a dictionary entry; but with faster updating of dictionary editions, there is a greater opportunity and appetite for capturing new language use. For a terminologist, a new word similarly had to achieve an established role in the language, or be otherwise officially sanctioned, to qualify as a term; though now, terminologists often work more closely with translators, and the pressure is on to provide up-to-the-minute terminology to supplement the institutionally recognised canon of the lexicon.

As we understand it, the translator's perspective lies somewhere between those of the linguist, language teacher and terminologist. The task at hand is typically to render the

source text (SL) as closely as possible, in all its conventionality and originality. This requires knowledge of, or at least access to, both new and established features of the language of translation (TL). The primary SL is most often English. This will often be non-native speaking English (as found in EU documents). It will typically be a specialised SL, with evolving neologisms and other terms not yet found in a dictionary.

The translator needs to work on many fronts, including:

1. understanding the meaning of the SL neologism
2. confirming that it is not a typographical error^{xvi}
3. confirming that it is not an aberrant invention by a non native speaker
4. assessing its status in the language; i.e.
 - a. checking whether it is already in technical or general text
 - b. checking whether it is already in a dictionary
 - c. judging its likely staying power
5. identifying and weighing it against competing alternatives
6. finding an equivalent neologism in the TL
7. juggling with etymological and other complicating linguistic factors
8. rendering its meaning in a paraphrase or synonym

The translator is working at speed, making compromises on the hoof, needing to use circumlocutions and approximations where a translation term does not exist, or is not known.

In Part 1 of this paper, we have presented the background to our own work, the situation in diachronic corpus linguistics and also tried to assess the needs of the translator. In Part 2, we shall consider neology in the context of its automatic identification in text, from the point of view of the translator.

Part 2. An automated, corpus-linguistic definition of 'neologism'

In our automated, diachronic approach, a neologism is defined as the first occurrence of a word or multi-word unit found within a given corpus. It may actually be a nonce formation, with only one occurrence, or it may appear again later. It may be a typographical error, of which there are estimated to be at least 10% in text. But in any case, it is simply registered as new, a neologism, at its first point of entry to a series of analytical software filters, as shown in Figure 1.

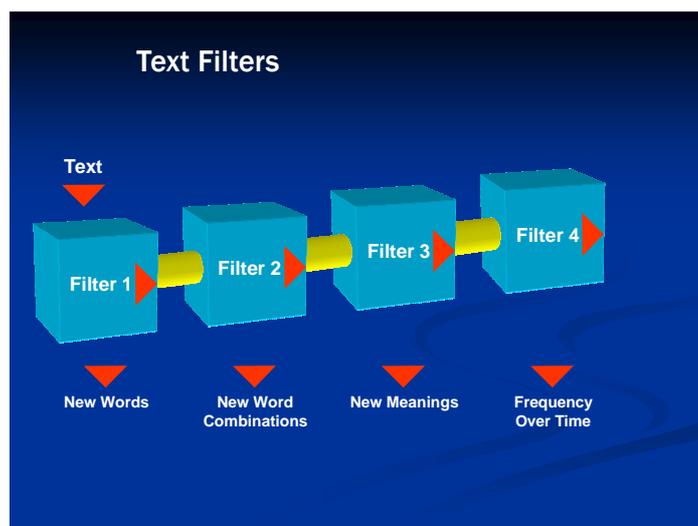


Figure 1: AVIATOR software filters to identify lexical and semantic neologisms

Filter 1 identifies new words in successive time-chunks of UK newspaper text from 1989 onwards. Each word in a chunk is compared with a master corpus of 5 years of *Guardian* newspapers from 1984 to 1988. Filter 2 narrows down candidate multi-word units among the neologisms. Filter 3 identifies semantic neology, and all new and existing words are recorded and stored with dates and contextual information by Filter 4 (Renouf, 2010).

2.1 Lexical neology

A sample of the neologisms extracted automatically from the *Guardian* in Dec 2010 by APRIL software is shown in Table 2:

- 1958's Scandinavian **depressionfest**
- this **cyber-putz** is hiding in England, but they can't find him
- the **pseudo-compilation** of proto-UK acid house
- Keith Richards has killed an orchid with cigarette smoke. News of the **botanicide** only surfaced this week
- Fox News **apocalypticist** Glenn Beck has predicted a new civil war
- endless **rescreenings** of *Gone With the Wind*
- the audience of **dissidents-in-exile** and Norway's great and good
- I want the body of a junkie , not a **jerkules**
- getting them to flower again involves **photoperiodism** and luck
- Jessie Cornish was a **pop-star-in-the-making**
- most car journeys are karaoke **holler-alongs**
- the **goose-whisperer** with legendary communication skills with birds
- books best suited as gifts for uncles, aunts and other **hard-to-buy-fors**
- Blocks like Grand Central are as banal as prefabricated **prole-stackers**

Table 2: Neologisms extracted from the *Guardian*, December 2010

In Table 2, we see a fairly typical range of word formation types. The majority of neologisms seem to be destined to remain 'nonce formations'. There are several derivations, both

standard, *-ist*, *re-*, *pseudo-*, and *vogueish*, *cyber-*. There is a compound, *depressionfest*, (though *vogueish* words like *-fest* are so flexible grammatically that they are arguably affixes; Baayen & Renouf, 1998). There are multi-word-hyphenated strings, e.g. *pop-star-in-the-making*, of various grammatical hues; compounds created by obvious analogy, as *holler-alongs* (from *singalong*), *goose-whisperer* (from *horse-whisperer*); a blended pun, *jerkules*. Just a few formations resemble technical terms, humorous, *botanicide*, *apocalypticist*, and scientific, *photoperiodism*.

Proper names are also a major component of neologistic output. In Table 3, we see some of the commoner varieties in the April 2010 output: the lexis of innovation – here technological (1-3); names of individuals brought to public notice - here actors and politicians (4-5); and names of real-world events (6-7).

1. Zangief	3 a fictional character in a Street Fighter video game series
2. Quickcut	7 a system to check files for compliance with a publication's criteria
3. Pagestore	8 a new company utilising Amazon's servers to crawl the web
4. Hotboxes	4 a guardian.co.uk 'hotbox', linked to relevant editorial content
5. Friis-Mikkelson	1 'actor <i>Friis-Mikkelson</i> was brought on board'
6. Cleggmania	49 popular reaction to Nick Clegg's pre-election speech
7. Eyjafjallajökull	61 Icelandic volcano
8. Bigotgate	16 Scandal resulting from a faux pas made by Gordon Brown

Table 3: New Proper Nouns extracted from the *Guardian*, April 2010

2.2 Multi-word lexical neology

Filter 2 refines Filter 1 output: if a new contiguous collocate or collocates appears, then a new word combination is deemed to be possibly forming. An example of this would be where *cow* begins to co-occur with *mad* and *disease*; or more precisely, where a regular permutation of those words begins to occur significantly over time.

2.3 Semantic neology

Filter 3 of the AVIATOR system identifies semantic neology. This knowledge is discoverable by comparing the collocates of each word entering the corpus with the collocational pattern recorded for it hitherto in an evolving 'Collocate Bank'. Existing words which exhibit continuing, significant change are deemed by us to have assumed a new or additional sense. The term *curate* exhibits this change. In fact, it exhibits multiple semantic change. A back-formation from the noun *curator*, it originally meant 'to undertake specialist work, primarily conservation and preservation, usually in relation to art or history collections, by a technical expert and scholar'.

- | | |
|----|--|
| 1. | 11/06/09 Holloway has curated a season of debates on politics |
| 2. | 16/08/10 the French pianist who curated the Aldeburgh festival. |
| 3. | 08/02/10 the gatekeepers that curated the content of the film |
| 4. | 04/07/11 an ongoing recital series curated by pianist András Schiff |
| 5. | 17/07/11 the festival which the band are curating in London this week |

-
6. 28/08/09 The company has collected and **curated** these data sets to use with Mathematica, a powerful computational programme
 7. 11/07/11 The app will deliver a daily edition of content, specifically **curated** for iPad
-
8. 16/06/11 a publishing house that has some interest in **curating** their authors' careers.
 9. 14/06/11 Whatever your reasons for **curating** your office or desk like it is.
 10. 03/06/11 He talks about "**curating**" shops and is to reinvigorate Waterstone's tired look.
 11. 26/06/11 a (web) journalist may add value in **curating** people and their information
 12. 10/07/11 an age of tweets and blogs, in which "**curating**" is the in-word (even DJs are now rechristened music curators),

Table 4: Extract for search term [*curate**{V*}/*curating*], illustrating some recent new senses

The tiny extract of data for the lemma *curate* in Table 4 illustrates how dynamic the language can be, and how tricky it is to keep track of words which, like *curate*, are re-discovered. This was almost a technical term, little used, which has been found to have a resonance and versatility which has suddenly propelled it out of obscurity and into multiple use.

- In 1-5, we see the verb *curate* in an extended sense, characterising the organisation of shows, music festivals and other events. It is true that festivals nowadays are more ambitious and more complex, technologically and organisationally, but the required skills still seem to be a step away from the scholarly nature of 'curating' as used to be, and the choice of this word seems in part to be to dignify the activity by implying that it requires recondite specialist knowledge.
- In 6-7, we see a further extension of *curate*, to mean 'deal with, manage, organise, arrange' in the technological context of '*data, data set, content, series, sets, app*'.
- In 8-12, we are alerted to a third vogueish use of *curate*: in web-based contexts, it is used to talk of 'managing things, enterprises and even people' – see collocates *desk, shops, careers, and people*.

I hesitate to see a 'classe d'objets' (Sablayrolles, 2009) emerging anywhere here, since there is no formal shift to accompany the semantic diffusion of *curate*, and the newly semantic similarities between *curated* and 'managed', 'organised' and so on is tenuous as soon as one tries to identify their shared collocates.

2.4 Semantic and formal neology

A semantic shift is, however, often accompanied by a grammatical reassignment. In Table 5, the contexts show that the noun *sunset* has clearly gained several metaphorical senses. It refers to the act of closure: from 1985, to an industry which is being closed down; from 1995, to a final review process before regulation takes effect; from 2004, to legislation which is ending; from 2008, to computing software services no longer available; and from 2009, to the redundant section of a newspaper.

In this metaphorical role, the lemma SUNSET now functions as noun modifier or verb. This 'formal neology' is a type of language change which deserves the attention of translators. Our system does not inventory neologisms by grammatical shift a priori, but any neologism identified by Filter 1 can be grammatically tagged as a search term prior to the retrieval of its

concordances from the corpus. It is also readily detectable in output through the accompanying collocational changes.

01/10/85	the dangers of expecting too much from sunrise as opposed to sunset industries
03/05/92	Floating voters mentioned the party's attachment to sunset industries
19/07/95	The report calls for ' sunsetting ', whereby all new regulation should be subject to an effectiveness review.
14/12/01	Part 14, relating to sunset clauses, has exercised all parts of the house
10/09/04	the bill that took automatic weapons off US streets is set to sunset
31/12/07	All want to make the Bush tax cuts, now scheduled to sunset in 2010, permanent
05/07/08	AOL has sent a memo that announces the " sunsetting " of AOL Pictures
19/10/09	As of this morning, we have decided to sunset the <i>Guardian America</i> front page.
03/09/09	They are " sunsetting ," or permitted to expire, unless Congress extends them.
20/12/10	the company proposed to " sunset " its social bookmarking service Delicious
21/12/10	Why I am sunsetting Yahoo.
06/06/11	Bills would bypass this process, but would always have a 12-month sunset clause
16/06/11	The rebels voted to introduce a " sunset clause" to the proposed referendum lock
17/06/11	They urged Salmond's government to add a " sunset clause" or review mechanism
30/06/11	the agricultural sector was seen as the " sunset industry" of Ireland
15/07/11	newspapers-in-print is a sunset industry

Table 5: *sunset/sunsetted/sunsetting*: instances of N-V conversion

In Part 2 of this paper, we have attempted to show how a corpus of mainstream UK journalistic text can provide insight into language which can help to meet a translator's needs to understand the meaning and use of SL neologisms; and to check that they actually exist and are neither typographical errors nor non-native usage. In Part 3, we shall develop a finer-grained diachronic definition of neology, again with the translator in mind.

Part 3. Establishing the changing status of a neologism: the life-cycle of a word in diachronic text

In the absence of ready textual evidence, there is uncertainty about the status of an SL neologism for translation purposes. Questions actually raised^{xvii} by translators include whether a loan adapted to the orthography or morphology of the established language is 'superior' in neological status to one which is not. Or, in the case of competing terms such as *e-learning* and *eco-learning* - both neologisms - whether the presence of *e-* in 'the dictionary' and the absence of *eco-* would indicate that *eco-* has a 'superior' neological status.

In this second part, we shall try to show how diachronic corpus analysis can reveal the status of a neologism at different points in time, beginning with whether it is a nonce formation or continues to appear in text, and then in what way it does so. Our corpus-linguistic definition of 'neologism' in this context is: 'the first **and subsequent** behaviour of a new word or multi-word unit found within a given diachronic corpus'.

We shall show some of these patterns across time, concentrating on the earlier stages - the birth of a word, and its increase in frequency, productivity and creativity – stages which are

of chief relevance to translators in indicating how a neologism can be judged to be gaining bona fide status in the language.

Words each have different patterns of occurrence, and their changes in frequency over time will be represented in this article by 'line graphs'. The method constructed by our lexical statistician, Professor Paul Davies, was a modification of standard ideas in control chart theory or change point theory (see, for example, Hand & Crowde, 1996; Ryan, 2011), which allowed for the fact that our neologistic data consisted, by definition, of low frequencies of occurrence.

Where each first occurrence is accepted as a bona fide word, the question of reliability arises. There are ways of identifying errors, in terms of impossible character combinations, but this is a rare phenomenon; more probable is, for example, the transposition of two characters. But in any case, in a diachronic corpus, it is the pattern of recurrence over time which counts. Errors generally fade away in time (or if they persist, the question is whether they will become bona-fide variants. For example, the erroneous *accomodation* occurs so frequently that it is treated as an alternative spelling in database lexicons.)

3.1 Birth, or the first occurrences in text of a neologism

There is a mismatch between the timing of a new real-world event or phenomenon, and the way it is referred to in news text. There can be delays. A mismatch can also exist between the perception of what is new, and what is really new: more than one person may well invent an idea or term, but it may not take hold the first time. Equally, a very similar event may occur at different points in history, but not be reported widely the first time, or be forgotten when the new event occurs and captures the public and media attention. For these reasons, many neologisms thrown up in news text are not actually new.

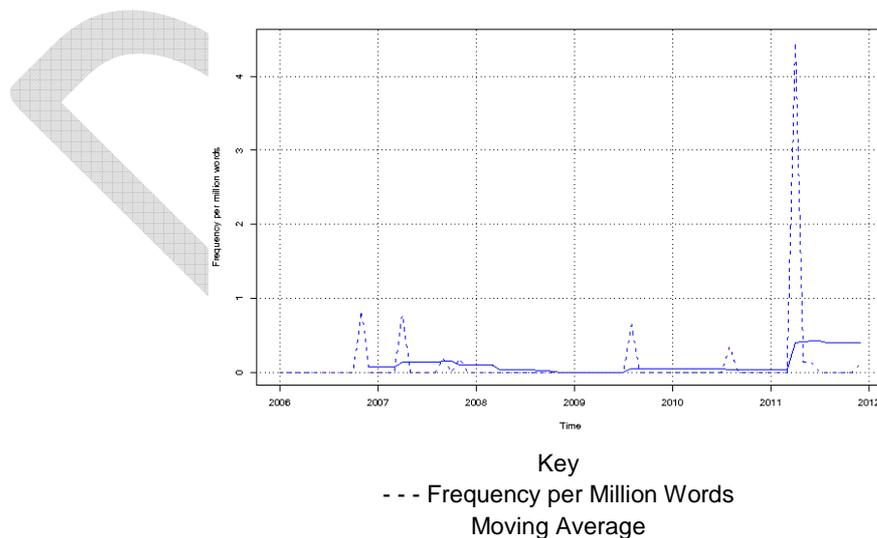


Figure 2: 49 instances of [burkini*|burqini*|burquini*] by end 2011

Figure 2 presents the pattern of frequency of occurrence of the term *burkini*/ burqini*/ burquini**, the wildcards allowing for plurality and derived forms. For reasons just explained, *burkini* is not actually new. There has been occasional mention from 1994, but only in 2011

does it briefly hit the news as a neologism, with reference to the holidaying apparel of a UK celebrity chef, Nigella Lawson. The graph is sketchy due to the few occurrences. Since the word is apparently perceived by the writers as new or strange to the reader, it is flagged as such by the conventional signalling punctuation and lexical and discourse markers (Renouf & Bauer, 2001), as illustrated in Table 6.

This coining^{xviii} is a blend of *bikini* and *burkha*, and, inconveniently for search purposes, its spelling has not yet settled down. In our system, though, one can allow for alternative spelling with a wildcard, as in *bur*ini*, which garners the same results.

meta-discourse signals (*aka; the term*)

2007 burkinis, **aka** beachgear for observant Muslim women

2011 **The term** burkini (or burqini) **was used by** an Australian designer **for** a swimming suit for Muslim women

inverted commas; glosses

2009 she was evicted from a public pool for wearing a “burkini”, **a veil, trouser and tunic, head-to-toe costume designed to allow Muslim women to swim piously.**

parenthesised gloss

2011 A celebrity donning a burkini (**or kaftan or any other voluminous outfit**)

definition:

2011 **A burqini (or burkini) swimsuit is a type of swimsuit for women designed by Lebanese Australian Aheda Zanetti**

Table 6: First occurrences of *burkini* with meta-signalling of novelty

Words exhibit one of a number of common change types across time. We can assume that the meteoric rise of *burkini*, given the ephemeral nature of women’s fashion, is likely to be followed by a similarly sudden fall.

3.2 Increase in frequency of occurrence

Subsequent to its birth, several events in the life of a word serve to indicate the likelihood of its assimilation as a bona fide part of the English lexicon. One is the degree to which its occurrence increases in frequency over time. This perspective is relevant for neologisms which form part of a specialist domain which nevertheless enjoys widespread communication, such as business and finance, particularly via the Internet, and of course IT/ICT.

In Figure 3, we represent the path of the neologism [*pop up/popup/pop-up* + N] across time, and see that it is picking up speed in 2011.

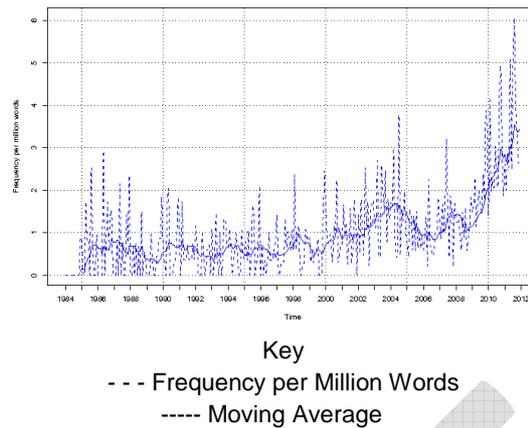


Figure 3: 1590 instances of the string *pop#up {N*}*'

It is clear what is causing the recent growth if one consults a 'heat map', a graphic representation of evolving data patterns which is generated by WebCorpLSE tools.

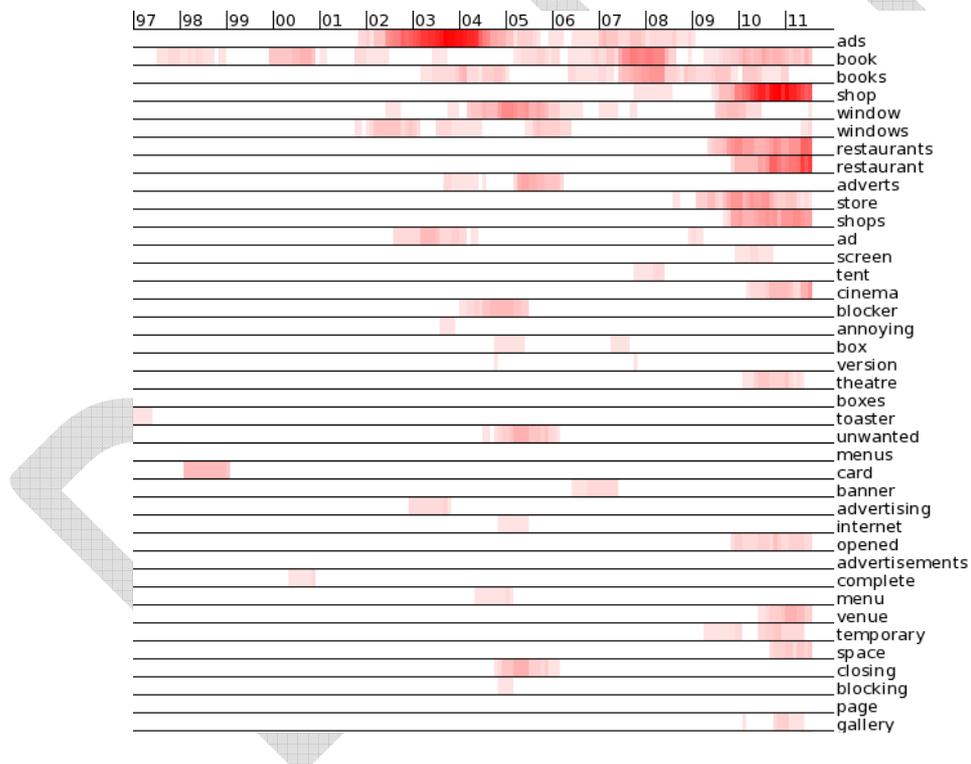


Figure 4: Heat map showing the changing collocational pattern for *pop-up*

The heat map in Figure 4 shows us how the collocational rankings for the word *pop-up* have been changing across time. This in turn reveals that the increased frequency is due not just to an increase in its established use as a verb-adjective conversion referring to a particular book format, or a kind of toaster. In around 2002, *pop-up* emerges strongly in the term *pop-up ad/s* to designate a form of online internet advertising intended to attract traffic or capture email addresses. In addition, in 2002, *pop-up* appears as a computing term, in combination with *window*, *windows*, *blocker*, *internet*. The heat map further reveals that in the past two years, *pop-up* has come to refer to the new real-world emergence of a flurry of temporary

buildings or venues set up to house small enterprises - *cinema, theatres, shops* and *restaurants* - which may also take advantage of a larger local event attracting custom.

3.3 Productivity

A sign of the increasing institutionalisation of a neologism in text is the appearance of associated inflected and derived forms. The definition of productivity is a matter of debate, the central view perhaps being that it is the affixes themselves which are productive (e.g. Bauer, 2001: 14ff). However, another perspective which we also find useful is of the word itself as being productive. This is in line with Saussure's (1969 -1915: 228) statement: "On pourrait classer les mots d'après leur capacité relative d'en engendrer d'autres", a view dismissed by Bauer as "idiosyncratic... or possibly just imprecise".

But Saussure continues: "selon qu'ils sont eux-mêmes plus ou moins décomposables" – i.e. 'complex words', whereas we find it useful to think even of simplex words as being productive. We could simply say that a word which is gaining in popularity begins to attract a range of productive affixes, and vice versa, though this does not account for the process of compounding.

In Renouf, 2007, we observed the fortunes of the simplex neologism *NEET*, one of the social stereo-typing acronyms which abound in British English and culture. *NEET* stands for 'Not in Education, Employment or Training'; and is used of disaffected youths of 18-24. Such a youth is 'a *neet*'. Coined at the end of 2004, the term gained in frequency, moving swiftly to lower-case orthography; but being a sober, governmental designation, it did not capture the media imagination, with a resulting absence of morphological variants except for pluralisation, and a formal shift to noun-modifier.

04/10/04 The health ministry estimates over 500,000 so-called neets
04/11/04 The neet group in Britain tends to be from the lower socioeconomic groups

However, this social phenomenon has continued to generate concern and news, and periodic reports and governmental initiatives have kept it in the news. By 2011, we see that it has accrued a modest range of derivations and compounds. Examples 3 and 4 in Table 7 take the opportunity to poke fun at members of the Royal Family, humour being an important driver for productivity.

03/06/07 these are Neets-in-waiting , too young to qualify as members of a 16-24 age group
15/09/09 Neet-free zone. One college's fight to keep young people on track.
27/09/09 Neetness is known to be common in families where there has been no experience of work for generations
27/09/09 William is already showing typical signs of Neet-related distress
13/07/10 one in six of those who experience two years of Neetism will be dead by 30

Table 7: Recent morphological variants on the neologism *Neet/s*

3.4 Analogical coining

Parallel-term proliferation is an observable reality in diachronic text, and a sign that the writer feels that the original form is sufficiently established for a neologistic variant to be recognisable by allusion. In Figure 4 earlier, the term *sunrise* was licenced by the new meaning of *sunset* in the context:

01/10/85 the dangers of expecting too much from **sunrise** as opposed to **sunset** industries

Pop-up

Another example is the hyphenated compound, *pop-up*. This is a conversion of the phrasal verb 'to pop up', functioning as noun modifier. Its constituents, *pop* and *up*, are common words in text, fulfilling many roles. *Pop* is a versatile verb: onomatopoeic, monosyllabic, simplex in morphology, with the sense of doing something suddenly, with speed and simplicity. The phonological effect of its combination with *up* is pleasing^{xix}.

In its computational sense, *pop-up* seems to be motivating the formation of associated coinings, of *pop* plus some of a set of available directional preposition: *pop-under*, *pop-over*, *pop-down* and *pop-out*. These are noun modifiers, with their plural inflections functioning as nominal abbreviations.

Pop-under is illustrated in Table 8. A '*pop-under advertisement*' is a software tool which hides a browser window under the active window and is not seen until the covering window is closed. A *pop-under ad* is also referred to as a *pop-under*.

30/08/01 you may be annoyed by X10's " pop-under " advertisements for a digital camera.
13/12/01 attempts to break through the surfer's disinterest: pop-ups, pop-unders
07/01/02 pop-ups and their even more annoying counterparts, pop-unders , were the staple tools of porn sites.
25/03/02 Banner ads, pop-up and pop-under ads
02/12/02 We are all irritated by pop-ups, pop-unders and spam, spam, spam, spam.
12/03/04 We ran a trial on pop-unders , and decided we didn't like them
12/03/04 you'll find a different species lurking under your browser window: the " pop-under "
01/04/04 "sponsored links" adverts are better than the banners, pop-ups, pop-unders , crawl-overs and other rubbish that infests the web.
25/01/07 A variation on the pop-up window is the pop-under advertisement.
15/09/08 Pop-ups and pop-unders are banned.

Table 8: 10 instances of *pop-under*

These hyphenated compounds occur frequently in text and can thus be expected to be polysemous. The nominalised *pop-over*, shown in Table 9, was previously the name of a type of US food, a light, hollow roll made from egg batter. But in our data, it now refers primarily to an advertisement that is part of an actual webpage, 'superimposed' in a transparent layer on top of the page, which cannot be blocked.

16/03/04 the Great Satan of web advertising, the flash pop-over ,
22/03/04 A reader compared pop-overs to the cat sitting on your newspaper as you read it.
16/03/04 Flash pop-overs : the Great Satan
16/05/10 as well as split screen views, popovers provide extra navigational aids.
19/07/10 it would be a short walk to a contextual pop-over for copy and paste functions.

Table 9: 5 instances of '[*popover**/*pop-over**]' in the computational sense

These pesky little coinings are fast-growing, destined to keep a translator on his/her toes. The concern among translators in France is evidenced in discussions on a French Moodle page, with references to what the Commissions Ministérielles de Terminologie suggest, at: <http://moodle.org/mod/forum/discuss.php?d=68365>

In Table 10, we see *pop-down* in its established computational sense of software, though in examples 2 and 4, it is extended to refer to a design feature of the hardware itself. More recent is example 5, where *pop-down* converts to a noun used to allude to and play on *pop-up* in its newest sense of temporary housing for events.

- | |
|---|
| <ol style="list-style-type: none"> 1. 01/09/85 Selecting a command causes a pop-down menu to appear 2. 16/07/93 the neat pop-down feet that hold the computer at a convenient angle 3. 13/05/97 Solo opens as a floating toolbar above your application, with pop-down menus 4. 25/06/97 This tiny little phone with its unique pop-down microphone 5. 14/05/11 you never know where a pop-up cinema is going to pop up, but this one is more of a pop-down, situated as it is in a subterranean venue. |
|---|

Table 10: all 5 instances of *pop-down*

In Table 11, we see *pop-out*, which had been fairly well established as a literal attribute of anything capable of that action, now being used of software applications. In 2011, it refers to a new digital re-creation of the pull tabs and spin wheels of the traditional hard-copy physical pop-up books of childhood.

- | |
|--|
| <p>08/10/01 Spam, unwanted pop-out ads, the crap people put into newsgroups.</p> <p>09/11/09 a tool that allows anyone to comment on anything on the web and have that comment displayed in a pop-out window alongside for all to see.</p> <p>13/05/11 Loud Crow Interactive's latest iPad app takes Beatrix Potter's Squirrel Nutkin book and gives it a pop-out tablet spin</p> |
|--|

Table 11: 3 instances of *pop-out* in the computational sense

Not all directional prepositions contribute to the pattern, however. This may be because *pop-offs*, *pop-ins*, *pop-by*, and so on, have already been claimed for other semantic purposes (they are all nonce formations as used in our data):

- | |
|--|
| <p>27/01/89 a nuance-free instrument which limits them to a pop-by numbers style.</p> <p>22/12/09 He would also get agitated at times, or have "pop-offs", as his father called them</p> <p>12/05/11 the festival opens in a welter of industry deals and celebrity pop-ins</p> |
|--|

or because the dimensions of a computer screen are primarily 'up-down', and 'over-under'; though *pop-out* must be understood in metaphorical terms, and the absence of *pop-above* has to be explained by the possibility that *above* is perceived as being positional rather than directional.

Knowledge of such serial coinage events in the source language can only help the translator to be more resourceful in rendering them in the translation language.

3.1.5 Lexical creativity

The examples above are evidence of words at play, within the normal range of morphological and lexical coining and re-combination. But half a step on from this is lexical creativity, by which we mean the creative manipulation of already existing formations. It is another clue to the possible staying power of a neologism. Journalists are alert to neologisms, whether their own or reported, with agreeable or amusing connotations, and potential for word-play. Once noticed, such a neologism will become more frequent, and if linguistically possible, creatively exploited. The most heavily creative neologistic phrase in our data has been *Weapons of Mass Destruction*, which, while not amusing in its real-world reference, offered huge potential for linguistic play, producing hundreds of alternative phrases, where *mass* and/or *destruction* were substituted with humorous alternatives (Renouf, 2007).

Another neologism taken up with relish by journalists was *Mad Cow Disease*, with high peaks of frequency in 1990-9, 1996-7 and lesser ones in 2001-2, as shown in Figure 5, echoing the scares.

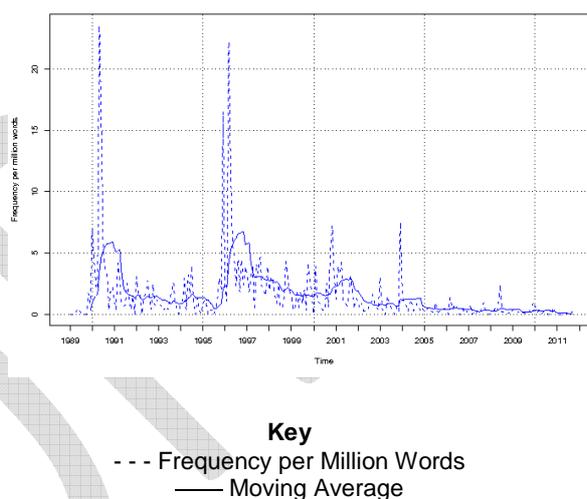


Figure 5: 1351 instances of *Mad Cow* Disease**, case insensitive

In order to investigate the creativity of a neologistic phrase, our system allows the suppression of one or more of its conventional elements, so that creative alternatives can emerge. To manipulate the phrase *M/mad C/cow D/disease*, for example, we suppress *COW*, *C/cow* and *C/cows*. The substitutions can be seen in Table 11, so that *mad+turkey+disease* is the creative phrase, and so on^{xx}.

turkey 5	Audit 3	Hod 2	car 2	mink 1	Kow-Tow 1
cat 4	calf 2	dog 2	Ostrich 1	vow 1	cowgirl 1
couch 4	Crowd 2	government2	potato 1	Euro 1	Artist's 1
Cowboy 3	Car 2	politician 2	cash 1	cow-like 1	male 1
sheep 3	mouth 2	cattle 2	Bishop 1	Kraut 1	chow 1

Treasury 1	judge 1	human 1	Comic 1	drivers 1	Person 1
curriculum 1	art 1	chair 1	antelope 1	Survival 1	crowd 1
driver's 1	bull 1	Cowes 1	moggie 1	reporter 1	Human 1
putting 1	fundamentalis1	beef 1	Jack 1	Jockey 1	cricket 1
ostrich 1	Grandmasters1	bear 1	chip 1	Cat 1	Bull
Cartoon 1	Couch 1	lamb 1	Elm 1	Dinosaur 1	
cowboy 1	motorist's 1	flu 1	investor 1	Cow's 1	
Moose 1	Philistines 1	Bercow 1	publicity 1	Bug 1	
Politician's 1	coup 1	media 1	icon 1	Hogg 1	

DRAFT

Table 12: 81 types of string *Mad * Disease*, case insensitive, with filters –Cow, -cow's, -Cow, -COW

The *mad turkey disease* revealed in Table 12 refers to a food scare in late 1996; the *mad cat disease* to a minor outbreak of the feline form of 'Bovine Spongiform Encephalopathy'. The remaining examples reflect two main phenomena. Firstly, the eagerness of the media to take linguistic advantage where possible, and the general pre-occupation with institutional overzealousness.

See, for example, the contexts for *Mad Audit Disease*, (where all three instances are from the same text about bureaucracy in Higher Education):

10/10/96 It is known as **Mad Audit Disease**.
 10/10/96 **Mad Audit Disease** sucks time away from more important duties.
 10/10/96 I blame Margaret Thatcher for **Mad Audit Disease**.

and with social malaise, as illustrated by the contexts for *Mad Couch Disease*

31/05/96 excessive viewing of sport on TV might lead to **mad couch disease**.
 09/06/96 televised sport could lead to a syndrome called '**mad couch disease**'.
 17/01/01 there is still work to be done in the iTV arena to stem '**mad couch disease**'.
 17/09/07 terms such as **Mad Couch Disease**

Secondly, the remaining internal collocates in Table 12 reveal the linguistic conventions of creativity. This is not a free-for-all, but a set of norms adopted to ensure that the reader comprehends the word-play in question (Renouf & Bauer, 2001). So the words above which substitute for *C/cow* tend to echo some aspect of the word's sound: in assonance, *couch*, *Crowd*, *Car*, *cow-like*, *mouth*, *car*, *vow*, *Kraut*, *Kow-Tow*, *cowgirl*, *chow*, *Cowboy*, *cowboy*, *Couch*, *coup*, *Cowes*, *Bercow*, *crowd*); in alliteration, *cat*, *curriculum*, *Cartoon*, *Comic*, *chip*, *Cat*, *cricket*); or allude to its semantics (*cattle*, *calf*, *antelope*, *sheep*, *Moose*, *mink*, *bull*, *dog*, *Person*, *Human*, *moggie*); or its reference. Some of these collocates conform to more than one norm, and the motivation for most is explicable within the detail of their contexts.

The rare but continuing occurrence of this term and these creative variants indicates that the collocational framework *Mad X Disease* is still available in the collective mind.

3.2 Competing synonyms

Synonymous terms frequently emerge and compete, and typically one gives way to the other at some point, or they become distinctively distributed^{xxi}. Competition can be judged in relation to frequency, productivity and creativity, trajectory and longevity. Quantifiable, diachronic corpus data such as we are using must be of great help to the translator in making these judgements.

mad cow disease vs BSE vs bovine spongiform encephalopathy

With relation to the multi-word unit, *M/mad C/cow D/disease*, there have been a number of well-known alternatives vying with it for position. *BSE*, the scientific co-referential

abbreviation, occurs 8,841 times in our data. (There are no instances of *bse*.) *M/mad C/cow D/disease* itself follows, with 1,351 occurrences, of which 1,243 are written in lower case, as *mad cow disease*, and only 40 with initial upper case, as *Mad Cow Disease*, (the rest a trail of odd orthographic combinations, puns and errors). Meanwhile, the full scientific term *bovine spongiform encephalopathy* lags behind, with 333 instances (there are no upper-case instances). So *BSE* is winning on the frequency stakes.

However, *BSE* shows no sign of generating puns or word-play (and likewise, *bovine spongiform encephalopathy* is not humorous and, like most classical scientific terms, leaves little room for word-play). So *mad cow disease* wins out in terms of productivity and creativity. In terms of longevity, all three terms have virtually faded from view by 2011, with downward trajectories since around 2005, as shown in the APRIL Word History graph in Figure 6 for the two synonyms which are sufficiently frequent to generate a clearly readable graph.

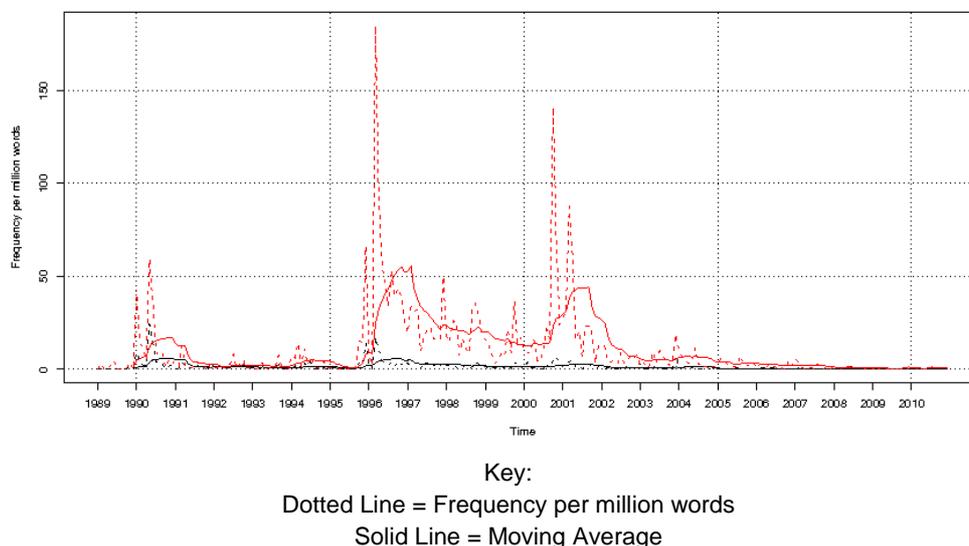


Figure 6: APRIL Word History Plot for *mad cow disease* (lower frequency line) and *BSE* (upper frequency line)

A notorious area of competing synonymy is that relating to the names for a portal data storage device. This nomenclature must be almost unique in not having settled down over the decades. Terms jockeying for position range from *flash memory* to *USB memory key*, *memory key*, *disk on key*, *USB key*, *data stick*, *finger stick*, *memory stick*, *info stick*, *USB stick*, *flash drive*, *keychain drive*, *thumb drive*, *pen drive*, *pocket drive*, *key drive* and *jump drive*, and this is without considering the commonly used trademarks of particular companies^{xxii}.

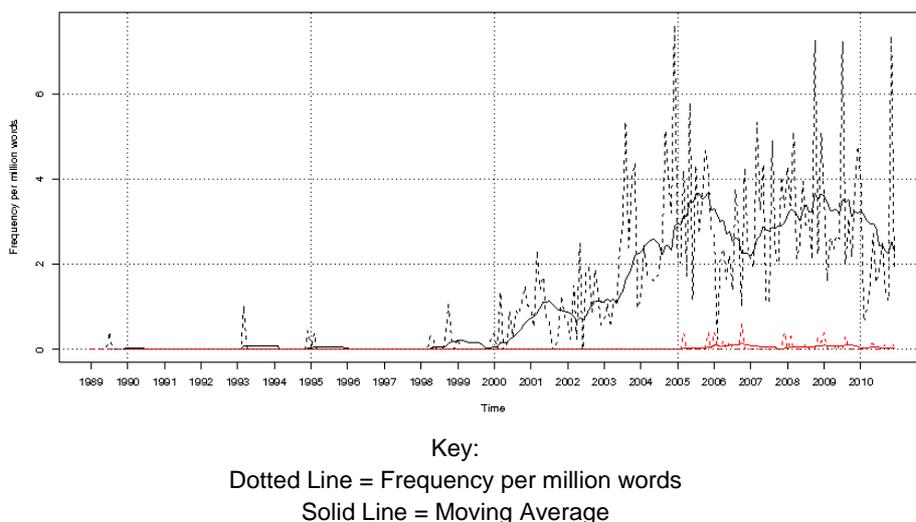


Figure 7: APRIL Word History Plot for *flash drive* (lower frequency line) and *USB* (upper frequency line)

This remains a topic of hot dispute on many web sites^{xxiii}. Several sites^{xxiv} host 2007 articles which report that: "Recently, 'USB flash drive' or simply 'UFD' has emerged as the de facto standard term for these devices. Many major manufacturers (SanDisk, Lexar, Kingston) and resellers use the term "UFD" to describe them. However, the myriad of (sic) different brand names and terminology used, in the past and currently, makes UFDs more difficult for manufacturers to market and for consumers to research".

In our data, we find only one instance of *UFD*, in 2004, and only 8 instances of *USB flash drive*, from 2005-2011. We find 67 instances of *F/flash D/drive*, from 2009-2011. In contrast, as shown in Figure 7, we find 1,807 instances of *USB*, from 1993 until 2011, with a continuing presence of 2 occs per million in 2011, so this appears to be the generic term of choice, at least in UK newspaper texts.

4. Conclusion

We have tried to show that the translator can benefit from access to modern English text corpora, and in particular, a large-scale flow of mainstream UK news text. A 'general' corpus like this can function as a reference corpus, to indicate whether an SL word is new, really exists and is showing signs of staying power. It also provides instances of the word in use which are valuable in guiding the translator towards natural phraseology. It needs to be chronologically stored and processed, so that trends and changes across time can be presented and quantified.

This does not, of course, obviate the need for access to specialised technical corpora. But, though a great volume of professional translation is highly specialised, there is also a huge volume of translation which has a direct link with what is going on in society in general. For example, the work on administrative texts of Rita Temmerman and Koen Kerremans (this volume) at the Erasmus Hogeschool in Brussels is done with the aim of harmonising the language content of forms so that no cultural or language group is disadvantaged in the EU

context. This work inevitably has as its starting point what ordinary people, rather than specialists, know. Here can be seen the relevance of newspaper corpora.

One field always has much to learn from another, and there are moves in translation studies which throw new light on lexicology and lexicography, e.g. on lexical metaphor (Kübler and Volanschi, 2011). The field of translation would benefit immensely from access to the knowledge and models of analysis which are established in neighbouring fields such as discourse analysis, information structure and pragmatics (see, for example, Pecman, 2012).

In this article, we have advocated corpus-linguistic methods of language observation - via concordances, and tabular and graphical representations of neology across time. There are corpora of many kinds available, not least the Web itself. Our own corpora and systems are available to researchers who register with bona fide projects at http://www.webcorp.org.uk/webcorp_linguistic_search_engine.html

5. References

- Algeo, John (1993), *Fifty Years among the New Words: A Dictionary of Neologisms 1941-1991*, Cambridge University Press.
- Andersen, Gisle 'Corpus approaches to neology and their relevance for dynamic domains', in *Neologica no. 6* eds. Humbley, John et Jean-François Sablayrolles, Paris; Editions Classiques Garnier. (this volume).
- Baayen, R Harald and Renouf, Antoinette (1998), 'Aviating among the hapax legomena: morphological grammaticalisation in current British Newspaper English', in *Explorations of Corpus Linguistics: Proceedings of the 18th ICAME Conference, University of Liverpool, 21-25 May 1997* ed. Renouf, A. 181-189
- Bauer, Laurie (2001), *Morphological Productivity*, Cambridge: Cambridge University Press.
- Dury, Pascaline et Aurélie Picton (2009), 'Terminologie et diachronie : vers une réconciliation théorique et méthodologique ?', in *Revue française de linguistique appliquée*. 31-41.
- Guilbert, Louis (1973), 'Théorie du néologisme', in *Cahiers de l'Association internationale des études françaises*, 25, 9-29.
- Humbley, John (forthcoming), *La néonymie : la néologie terminologique*, in Collection Jean Pruvost, Paris, Editions Honoré Champion.
- Kehoe, Andrew and Matt Gee (2009) 'Weaving Web data into a diachronic corpus patchwork' in Renouf & Kehoe (2009).
- Kübler, Natalie and Alexandra Volanschi (2011), 'The Impact of metaphorical framing on term creation in biology', in *Terminologie*, 17:2. Amsterdam: John Benjamins Publishing Company. 198-223.
- 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.
- Pecman, Mojca (2012), 'Tentativeness in term formation: a study of neology as a rhetorical device in scientific papers', in *Terminologie*, 18.1. Special issue: 'Neology in Specialised Communication'. Amsterdam: John Benjamins Publishing Company.

- Renouf, Antoinette (1993), 'A Word in Time: first findings from dynamic corpus investigation', in *English Language Corpora: Design, Analysis and Exploitation*, eds. Aarts, Jan, de Haan, Pieter, & Nelleke Oostdijk, Rodopi, Amsterdam. 279-288.
- Renouf, Antoinette (1996a), with Baayen, R. Harald, 'Chronicling the Times: Productive Lexical Innovations in an English Newspaper', *Language*, 72.1 ed. Sproat, Richard (1996) 69-96.
- Renouf, Antoinette (1996b), 'The ACRONYM Project: Discovering the Textual Thesaurus', in *Papers from English Language Research on Computerized Corpora (ICAME 16)*, eds. Lancashire, Ian, Meyer, Charles & Carol Percy, Rodopi, Amsterdam. 171-187.
- Renouf, Antoinette (1996c), 'Les Nyms : en quête du thesaurus des textes', in *Linguisticae Investigationes XX: 1*, John Benjamins, Amsterdam, 145-165.
- Renouf, Antoinette & Laurie Bauer (2001): 'Contextual Clues to Word-Meaning', in *International Journal of Corpus Linguistics*, Vol. 5(2), Amsterdam/Philadelphia: John Benjamin Publishing Co. 31-258.
- Renouf, Antoinette (2003), 'WebCorp: providing a renewable data source for corpus linguists', in Granger, Sylviane & Stephanie Petch-Tyson (eds.) *Extending the scope of corpus-based research: new applications, new challenges*, Amsterdam/Atlanta GA: Rodopi. 39-58.
- Renouf, Antoinette, Kehoe, Andrew & Jay Banerjee (2006), 'WebCorp: an integrated system for web text search', in Hundt, Marianne, Nadja Nesselhauf & Carolin Biewer (eds.), *Corpus Linguistics and the Web*, Amsterdam/Atlanta GA: Rodopi.
- 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 (2009), 'Corpus Linguistics beyond Google: the WebCorp Linguist's Search Engine' in 'New Paths for Computing Humanists', Siemens, Ray and Gary Shawver (eds.) in *Digital Studies / Le champ numérique Vol 1, No 1*, the Society for Digital Humanities / Société pour l'étude des médias interactifs (SDH/SEMI).
- Renouf, Antoinette & Andrew Kehoe (2009) (eds.) *Corpus Linguistics: Refinements and Reassessments*, Amsterdam/Atlanta GA: Rodopi.
- 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)
- Rondeau, Guy (1984), *Introduction à la terminologie*, Chicoutimi, Gaëtan Morin (ed.). 238.
- Sablayrolles Jean-François (2009), 'Néologie et Classes d'Objets', in *Neologica no. 3*, eds. Humbley, John et Jean-François Sablayrolles, Paris; Editions Classiques Garnier.
- Saussure, Ferdinand de (1969) [1915], *Cours de Linguistique Générale*. Paris: Payot.
- Temmerman, Rita and Koen Kerremans (2012), 'Secondary term formation in EU translations: a case of 'monitored laissez-faire'?', in *Neologica no. 6* eds. Humbley, John et Jean-François Sablayrolles, Paris; Editions Classiques Garnier (this volume).

6. Acknowledgement

I am most grateful to John Humbley for his generous advice on various translating issues.

ⁱ 'hapax legomenon' or one-time occurrence.

ⁱⁱ <http://www.merriam-webster.com/>

ⁱⁱⁱ <http://www.collinslanguage.com/content-solutions/wordbanks?gclid=CJ2h6fr5k64CFQlpfAodnT7QLg>

^{iv} <http://oxforddictionaries.com/words/how-do-you-decide-whether-a-new-word-should-be-included-in-an-oxford-dictionary>. "the Oxford English Corpus ... consists of entire documents sourced largely from the World Wide Web, while the Reading Programme is an electronic collection of sentences or short extracts drawn from a huge variety of writing, from song lyrics and popular fiction to scientific journals ... based on the contributions of readers ... We continually monitor (these) to track new words coming into the language: when we have evidence of a new term being used in a variety of different sources (not just by one writer) it becomes a candidate for inclusion in one of our dictionaries ... New terms can achieve enormous currency with a wide audience in a much shorter space of time, and people expect to find these new 'high-profile' words in their dictionaries. This presents an additional challenge to lexicographers trying to assess whether a term is ephemeral or whether it will become a permanent feature of the language". So Oxford Dictionaries has

^v See Renouf, Antoinette (1996b and 1996c).

^{vi} Established in 1988 at the University of Birmingham as the Research and Development Unit for English Language Studies, the Unit moved to the University of Liverpool in 1994 as the Research and Development Unit for English Studies; since 2004, it has been located at Birmingham City University.

^{vii} The widely-used and publicly available British National Corpus, a 100 million word synchronic collection of samples of written and spoken language, contains texts from 1991-1994.

^{viii} This NNS project is based on AVIATOR methodology (Andersen, 2012)

^{ix} This methodology was presented at CINEO 2008, and published in the proceedings (Renouf, 2010). It has since inspired diachronic corpus and software developments at the University of Pompeu Fabra.

^x Chapitre 3: Les Pionniers de la Néonymie

^{xi} <http://www.helsinki.fi/varieng/>

^{xii} This is a series of one-million word, comparable electronic corpora, of British and American written English, modelled on the first such corpus developed at Brown University in the US, by W Nelson Francis and Henry Kučera, itself ultimately derived from ideas gleaned from Professor Sir Randolph Quirk's pre-computational Survey of English Usage at UCL, London. The corpora currently span the period 1901 to 2011, with approximately thirty-year gaps between each, and are still being developed at a rate which makes an up-to-date inventory hard to find. See Zurich, Lancaster and Freiburg English Department websites.

^{xiii} See the RDUES website for a list of the major software developers across the decades: notably Alex Collier, Mike Pacey, Jay Banerjee, Andrew Kehoe and, more recently, Matt Gee. For shorter periods also Sue Blackwell, Eleanor Munby and Barry Morley. The lexical statistician has throughout been Professor Paul Davies, formerly of the University of Birmingham.

^{xiv} Managing Director, Ed Burt, 1995-2001.

^{xv} or rather, applied corpus linguist, since we identify and exploit surface patterns of lexis through the creation of, and in order to create, automated methodologies for information extraction.

^{xvi} Personal communication (2011) with Maria José Palos Caravina, Senior Terminologist, Language and Technology Section, Translation Centre for the Bodies of the European Union.

^{xvii} Personal communication (2010) with Xosé María Gómez Clemente – Professor at University of Vigo; President of 'Asociación de Tradutores Galegos', and author of dictionary of neologisms in Spanish.

^{xviii} Bauer's 2001 compromise term for formations lying between nonce formation and enduring neologism.

^{xix} There is potential for confusion with the base form *pop*, in that it is confusable with a homographic acronym, P.O.P. (plural P.O.Ps), which has several senses, all to do with electronic communication.

^{xx} Humbley (personal communication) sees similar 'défigement' in the French '*maladie de la table folle*' at http://french.feeder.ww7.be/spip.php?page=archive&rubrique=8&cal_date=2010-03-03&debut_syndic=100

^{xxi} Derivational affixes can also follow this pattern. See Renouf and Baayen, 1996a.

^{xxii} such as *disgo*.

^{xxiii} e.g. <https://bugs.launchpad.net/ubuntu/+source/usb-creator/+bug/275138>

^{xxiv} e.g. http://en.wikipedia.org/wiki/USB_flash_drive