

**Tracing lexical productivity and creativity in the British Media:
'the Chavs and the Chav-Nots'
Antoinette Renouf**

1. Introduction

The aim of this chapter is to examine the processes of lexical productivity and creativity as they are revealed in the empirical observation of journalistic texts over time. This observation is assisted by software systems designed to identify and monitor lexical innovation and change in chronological, electronically-held texts, and to extract language resources in the form of analysed databases of lexical 'knowledge' which is not accessible or retrievable manually.

The chapter will look at some recent lexical coinages and revived uses which journalists have either produced themselves or which they cite, and the key concepts of productivity and creativity will be defined within the terms of modern diachronic corpus study.

With the diachronic perspective on lexis, the degrees of frequency, productivity and creativity indicate how active and important a word is in the language at a given point in time, and they provide the means whereby its 'life-cycle' may be charted across the years.

2. Data

In order to face the challenge of tracking and tracing lexical change over time (Renouf, 1994, 1997), we have accessed open-ended stretches of electronically held texts which are amenable to chronological analysis. The data underpinning this chapter thus comprise over 700 million running words, or tokens, of UK 'broadsheet' newspapers covering the period from 1989 to the close of 2005 in an unbroken flow.¹ Though lexical productivity and creativity vary according to the sub-domain of journalism in which they occur (sports reporting, for example, being surprisingly rich in word play and metaphor), we shall deal with these processes in 'undifferentiated' text across entire newspaper issues.

Data extracted from the World Wide Web by our 'WebCorp' tool will also be consulted.² Web-based text supplements the linguistic information found in even the most-up-to-date tailored text corpora by furnishing more recent coinages and usage (Kehoe & Renouf, 2002; Renouf 2003; Renouf et al, 2003, 2005 & forthcoming; Morley, 2006), although its lack of dating information inhibits the reliable diachronic study which a designed diachronic corpus permits (Kehoe, 2006).

3. Methodology and Tools

The research on which this chapter is based has followed a coherent intellectual path since 1990.³ Hypotheses about the relationship between meaning and surface textual patterning have proved to be sound, allowing us to cumulatively develop algorithms for automated systems capable of identifying a number of lexical and lexico-semantic phenomena in text across time. These phenomena include neologisms (Renouf, 1993b; Baayen & Renouf, 1996), new word senses⁴ (Renouf, 1993 c,d,e), sense relations and changing sense relations within text⁵ (Renouf, 1996; Pacey et al, 1998); and the kinds of productivity and creativity of new words entering news text⁶ (Pacey et al, forthcoming). The methodology for each of these procedures varies, but basically involves ‘feeding’ a specific time chunk of chronologically sequenced, fresh textual data through a set of software filters which detect novel words as well as new collocational environments of existing words.⁷ The illustrative material for this chapter has been produced via these various word filters, as well as by regular concordancing and word-processing software.

4. Productivity

Productivity is the term used to refer to the word formation processes wrought upon a lexeme. If a word is ‘productive’, it means that associated grammatical and derivational variants are being produced.⁸

Productive items may be words or sub-word morphemes. Our analytical systems reveal that the most common word formations to appear in chronological text include compounds, derivations, proper names, acronyms and words containing numerals; less common entrants include loans, blends and conversions (Pacey et al, forthcoming).

We have not yet ranked the individual roots or lemmata across time for productivity, though we will at a later stage. The analysis involved is only partially automatable, given morpheme-boundary ambiguity. But we do have a chronological record of the ranked inventories of morphological or, more precisely, affixal productivity for each year from 1989-2005.⁹ This is a rather pragmatic definition of affix, which more accurately equates to ‘combining element, hyphenated or unhyphenated’, and includes affixes, neo-classical combining forms and even words which seem to be grammaticalising – e.g. *cod(-)*, *faux(-)*, *(-)gate* (Renouf & Baayen, 1998). Within this framework, the annual ranking for each affix is interesting for the insight it gives into contemporary social and media preoccupations. It confirms that the level of productivity for each affix is remarkably stable, a stability sustained below the commonest, most established band of affixes - of the ‘*un(-)*’ and ‘*in(-)*’ variety - into the lower ranks. It also reveals some slightly more productive exceptions. Figures 1 and 3 isolate the few affixes showing an upward or downward trend in our data, their rankings

represented in terms of the frequency band into which they fall each year. There are ten bands (or ‘deciles’), with Band 1 ranking for affixes with the highest frequency score, and so on.

	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05
Cyber	8	9	8	9	8	5	3	2	2	2	3	2	3	4	4	5	5
Euro	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	4
Mid	2	3	3	3	4	4	3	3	3	4	4	4	5	5	5	5	5
Techno	6	7	6	6	5	5	3	3	4	4	5	4	5	5	7	6	6
Under	2	2	3	3	3	3	3	3	3	3	4	4	4	4	5	4	5
Dis	3	4	5	4	5	5	5	5	6	6	6	5	6	6	6	7	6
Inter	2	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	6
Mis	2	3	4	4	5	5	5	5	5	5	6	5	5	6	5	5	6
Mock	4	4	5	5	5	5	5	4	5	5	5	5	5	6	6	6	7
Faux	9	9	8	9	7	8	6	5	6	6	5	6	7	6	5	5	5
Trans	4	6	5	6	5	5	6	5	5	7	7	6	6	7	6	6	8
Poly	4	5	5	5	6	7	6	6	6	6	7	7	9	7	7	7	7
Uber	10	10	10	9	9	9	7	7	6	6	5	7	7	7	6	6	5
Fore	6	7	7	7	8	8	9	7	8	7	8	8	8	8	9	7	9
Vice	6	7	8	8	9	7	9	9	9	8	9	9	9	10	8	9	9

Figure 1. Prefixes with growth or decline in corpus frequency ranking across 16 years

In Figure 1, the ranking of prefixes reveals a noticeable rise for the vogue items *cyber(-)*, *faux(-)* and *uber(-)*. In addition, the prefix *techno(-)* shows definite but temporary peaking in the mid 90s.

cyber 2000

cybertainment
 cyber-squatting
 cyber-geeks
 cyber-potential
 cyberdream
 cyberpicketed

uber 2005

uberpundits
 uber-waif
 uber-hyped
 uber-talkshow
 ubermogul
 uber-media-saturated

faux 2005

faux-ruefully
 faux-handwritten
 faux-autism
 faux-clanger
 faux-cringes

techno 1996

techno-nerdish
 techno-glam
 techno-ambient
 technofreak
 techno-guru
 technoboffins

Figure 2: New prefixal formations in years of peak productivity for *cyber(-)*, *faux(-)*, *techno(-)* and *uber(-)*

In Figure 2, which itemizes the derived forms built upon these prefixes, we see that their rise is due to their function both as semantic components and as discourse signals of modernity and humour, and that they are ready to attach to most grammatical word-classes (an argument for classifying

modifiers like *faux(-)* not as adjectives but as grammaticalising affixes (Renouf & Baayen, op. cit.)).

Meanwhile, for the suffixes in Figure 3, we see an even more consistent profile across time, with a slight rise in rank only for vogue suffixes *(-)fest* and *(-)athon*.

	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05
Able	2	3	3	3	3	3	3	3	3	3	4	3	4	4	5	5	5
Ation	2	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5
Fest	7	7	7	7	7	6	6	5	5	4	5	4	5	5	5	5	5
Ment	4	5	6	5	7	6	6	7	7	7	7	5	7	7	7	7	7
Ville	3	4	4	4	5	4	5	4	5	5	5	5	6	5	5	5	6
Ful	4	5	5	5	5	6	5	6	5	7	7	6	7	7	7	6	7
Ship	4	6	5	6	6	6	7	5	5	6	7	6	8	7	7	7	7
Athon	9	8	9	8	9	7	8	7	7	8	8	7	8	7	7	7	7
Ize	5	6	7	7	7	7	6	7	5	6	7	7	6	6	7	7	8
Able	2	3	3	3	3	3	3	3	3	3	4	3	4	4	5	5	5

Figure 3: Suffixes with growth/decline in frequency across 16 years

Both *(-)fest* and *(-)athon* designate or allude to large-scale public events, and are thus ideal for news reporting. As seen in Figure 4, it is clear that at the end of 2005 there is a semantic set of words combining with *(-)fest*, of the *smarm*, *swank*, *schmalz*, *slush*, *gush* variety. The combining form *(-)athon* is meanwhile seen to attach to a semantic group connoting triviality: *froth*, *bitch*, *cringe*, *grope*, *yawn*. Both suffixes also attach to topical proper names, such as *Sven* (Sven-Göran Eriksson – English football team coach involved in scandals) and *Bard* (metonymic reference to a Shakespeare festival), demonstrating that they are used to add glamour to publicised events.

fest 2005	athon 2005
miaowfest	frothathon
backslapfest	Svenathon
smarmfest	Bardathon
Potterfest	bitchathon
kinderchucklefest	cringe-athon
swankfest	improvathon

Figure 4: Some new suffixal formations in year of peak productivity for *(-)fest* and *(-)athon*

Morphologically, once a neologism begins to take hold, it typically starts to spawn inflections and derivations. Such productivity may occur almost at once, especially if the word is in the public eye.

Alternatively, the productive process may take longer, or not occur at all. Reduced or delayed productivity may be due to one of a number of inhibiting factors. Possible ‘inhibitors’ are tentatively suggested below.

Inhibitor 1: a term may be introduced into general English which people find difficult to integrate into regular syntax; common examples are terms borrowed metaphorically from scientific domains, such as *symbiosis*. This first took on a non-biological sense in the 1600s, with *symbiosis* first recorded in 1622.¹⁰ The foreignness of this word and its variants is perhaps what has kept them stably rare in our text over the years. The first occurrence of *symbiosis* in our modern-day news corpus is in 1987, and of *symbiotic* in May 1986, where the latter is glossed as follows, possibly indicating that its comprehensibility is still not taken for granted:

86.05 Part of the reason lies in the **symbiotic** relationship between hardware and software - **one cannot work without the other**.

Figure 5 shows that there is little productivity over the years for *symbiosis*. Of the few formations, *symbiosis-travel* is part of a web address (www.symbiosis-travel.co.uk) rather than a genuine language item.

symbiosis 339	symbiotic 564
symbioses 5	non-symbiotic 1
<i>symbiosis-travel</i> 2	quasi-symbiotic 1
sapper-symbiosis 1	symbiotically 27

Figure 5: *symbiosis* and variants with frequencies in *Independent/Guardian* 1989-2005

Inhibitor 2: Other examples of words which may integrate with difficulty are foreign or classical terms, such as *synergy*.¹¹ Having its origin, like *symbiosis*, in the 17th century, this term featured in our modern text at a low level of frequency from the outset. However, its derived adjective *synergistic* first appeared in a non-scientific context in our news data only in March 1989, and then only once, in the following context:

89.03 Turner's broadcasting networks and MGM film archives are more 'synergistic'.

This is an instance of an established word form being presented in inverted commas, perhaps indicating that the writer feels it needs to be signalled as new because of its rarity and enduring strangeness.¹² The overall productivity of *synergy* in terms of the number of each variant form has remained fairly low over 16 years, as shown in Figure 6. The range of variants is also limited: there is no *synergistically* and there are few compounds.

non-synergistic 1	synergised 4	synesynergy 846
synergic 1	synergises 1	synesynergy's 1
synergical 1	synergising 2	synesynergy-driven 1
synergie 3	synergizing 1	synesynergy-less 1
synergies 906	career-synergy 1	synesynergy-unlocking 1
synergise 6	dis-synergy 1	

Figure 6: Productivity of lexeme *synergy* from 1989 – 2005

Inhibitor 3: A word which is not in widespread use may also be limited to one or two inflectional variants. The figures for *synergy* and *symbiosis* in Figures 5 and 6 indicate that the nominal form seems to be preferred for *synergy*, but both the nominal and adjectival forms appear for *symbiosis*. Some nouns may be retained in particular texts in their nominal form for rhetorical purposes, e.g. as thematic markers. Nouns which mark themes in news text include abstract nouns such as ‘*inequality*’ or ‘*unemployment*’ which have a sociological bias and are used in news reports on social issues. In these cases, the usual stylistic prohibition on exact repetition in written text may be obviated by inserting them into ‘nominal frameworks’ of the kind ‘*the NP1 of NP2*’.¹³

Indeed, while ‘*the NP1 of NP2*’ is commonly found within most textual domains and types, a web search shows that ‘*the NP1 of unemployment*’ is particularly frequent in news text, where its NP1 comprises nouns in various semantic areas, as follows:

- *problem, topic, issue, question*
- *effect, consequence*
- *level, rate*
- *risk, pressure*
- *spectre, curse, blight, tragedy, ghost, humiliation, misery, scourge*

The type of noun ranges from multi-purpose semi-lexical discourse organisers such as *question* and *issue*, to more specifically lexical, evaluative items like *spectre* and *scourge*. The nouns *problem* and *risk* occur most frequently in this lexico-grammatical pattern, especially when NP2 is actually a sequence of NPs. We find for example:

- 05.07 a summit dominated by **the problems of high inflation and growth.**
04.11 the social contract underwrites **the risk of unemployment, ill-health and old age**
97.06 a government to tackle **the problems of unemployment, low pay, and education and health services in decline**

Corpus frequency figures indicate that the nominal forms for *unemployment* and *inequality* predominate numerically - see below:

Unemployment	38,610	inequality	4,421
unemployed	18,542	inequalities	2,552
		unequal	1,983
		inequal	2

Inhibitor 4: A word may only be fleetingly topical at the outset, and subsequently only intermittently applicable to events or thinking, and thus not lend itself to productive word formation. The acronym *NIMBY*, standing for ‘Not In My Back Yard’, had a limited impact after its first appearance in 1989, and displays little variation over 17 years, as shown in Figure 7.¹⁴

NIMBY 7	Nimbyists 2	nimbyism 101
NIMBYS 1	Nimbyness 1	nimbyist 5
NIMBYism 2	anti-Nimby 1	nimbyists 2
NIMBYs 5	non-Nimby 1	nimbys 42
Nimby 124	Nimbys 62	nimbies 12
Nimby's 7	nimby 71	nimbys-come-lately 1
Nimby-ism 2	nimby-ish 1	not-in-my-back-yard 8
Nimbyish 5	nimby-ism 1	not-in-my-back-yarders 1
Nimbyism 83	nimby-ites 1	not-in-my-backyard 19
Nimbyist 1	nimbyish 2	

Figure 7: Productivity for lexeme *NIMBY/nimby*, 1989-2005

Although *NIMBY* very soon adopted a lower-case orthographic variant, *nimby*, this was not the sign of widespread currency that such an event usually represents (Renouf, 1993a). The sparse use of a limited repertoire of standard inflectional variants indicates that this acronym has not set the media world on fire over 17 years, and that, on the whole, what has been said about it has been factual rather than playful. Of course, it represents an entire phrase, but as a quasi-word it has not entered into any compound formations or, with only one exception, any creative use (see discussion of ‘Creativity’ in the next section).

In the context of productivity, it is pertinent to speculate as to whether anything in the nature of a generalisable principle can be deduced about the order in which these different inflections and variants are generated in language. As yet, we have not deduced any general pattern of evolution, and expect that it is likely either to be serendipitous as to which form occurs first, or determined by a particular text in which several references to the same concept arbitrarily follow one another, according to the writer’s style and the flow of the argument.

5. Creativity

Lexical creativity in journalism, as in literary prose or poetry, is employed for the purposes of achieving certain stylistic effects, such as humour or irony (for a full discussion see Munat, this volume). It is also a device used

to convey to the reader a sense of authorial learnedness, sophistication, distancing and so on. It manifests itself in punning and other word play, metaphorical extension, willful error and duplication or usurpation of the role of an existing formation. These devices have been classified more systematically and exhaustively by Kuiper (this volume), and other functions proposed by Hohenhaus (also this volume).

Creativity is typically thought of as the act or quality of an unpredictable departure from the rules of regular word formation, and in poetic texts it can take surprising and thrillingly unexpected turns. To the long-term observer of newspaper text, the repertoire of journalistic creative devices is rather more familiar and restricted, probably due to the pressures on journalists to produce large amounts of eye-catching copy to deadlines (Renouf, 2005a,b). However, one can still marvel at the ability of the writer to invent quite novel and allusive language, and of the reader to interpret it.

Different policies operate in the different newspaper publishers concerning the level of readers' real-world knowledge and regularity of readership that should be assumed by the reporting and editorial staff, and these doubtless also have a consequence for the degree and type of lexical creativity employed, which would be a subject for further study.

Whether the full creative potential of a word or phrase is actually realised is another matter. Each word has its own frequency profile across time. Like *NIMBY*, mentioned earlier, some remain stably rare. Other obsolescent or dormant words may suddenly find themselves revived and pressed into service (e.g. *poll-tax*, *cod-*; and *tsar*, which we shall come to later). Some new words emerge and display a surge in popularity on the crest of a new media topic or vogue, receding from view once media interest has shifted.

In journalism the time constraints may not only formalise linguistic experimentation but also induce the writer to prefer certain words and avoid others. At one extreme, with a consistently low-ranking term like the acronym *NIMBY*, we find just one creative treatment in our data, in July, 2001 - the phrase *nimbys-come-lately* - in:

01.07 Mitchell of the Alliance strongly denies representing bourgeois
nimbys-come-lately who don't want the hoi polloi lowering the tone¹⁵

referring to two Fulham football clubs, by analogy with the phrase 'Johnny-come-lately'. This is thin pickings for a lexical item so long in existence, considering that it did have some time in the sun.

It is clear that vogue terms, foregrounded by real-world events for whatever reason, will be in the front line for exploitation, as will words and phrases with inherent potential for allusion and punning. A highly topical or fashionable word or phrase is typically accompanied not just by a flurry of productivity but also of creativity.

6. Life-cycle

In order to demonstrate the evolution in the fortunes of a word or lexical unit, we take the noun phrase ‘*Weapons of Mass Destruction*’.¹⁶ This originally emerged in our data in 1989, in the context of the first Gulf War. However, it only really caught the media imagination in 2003, with the threat of a second Iraq war, as is shown by the ‘time series plot’ of its frequency pattern from 1989-2005, in Figure 8. The phrase is not specified for case, allowing upper and lower case counts to be conflated.

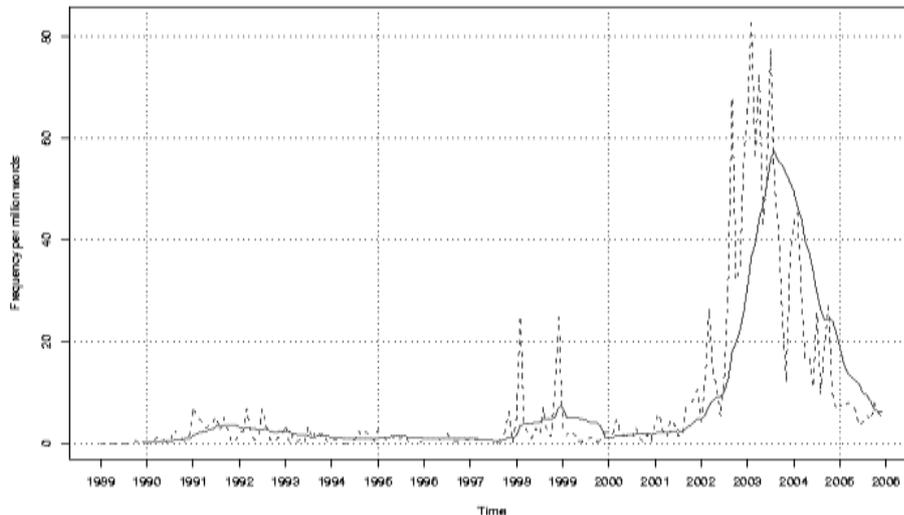


Figure 8: Time series plot: *weapons of mass destruction*, 1989-2005

Figure 8 indicates frequency bulges in the early and late 90s, (represented as a dotted line for simple ‘frequency per million words’, and as a solid line for the more interpretable ‘smoothed frequency’), as well as an enormous upsurge in the frequency line between 2002 and mid 2003, when the phrase was (re)introduced by the Bush administration. This pattern is followed by a slight downward trend, as the rate of occurrence slowed down again by the close of 2005.

The surge in frequency of the noun phrase ‘*Weapons of Mass Destruction*’ was accompanied by a surge in creative play with its lexical content. In order to discover more precisely what shape this new creativity took, we targeted the primary areas of creative potential within the phrase, namely the slots containing the words *mass* and *destruction*, and foregrounded these by the selection of 3 lexico-grammatical search frameworks, one [*weapons of * destruction*] another [*weapons of mass **], and the third [*weapons of * **]. A requirement for creativity was imposed by replacing the terms *mass* and *destruction* by wild-card open slots, in order to suppress the retrieval of the conventional phrase. The results obtained for each of these search patterns is presented in Figure 9.

<p><i>weapons of mass * (no destruction)</i> weapons of mass distraction weapons of mass obstruction weapons of mass disruption weapons of mass deception weapons of mass discomfort weapons of mass terror weapons of mass casualties weapons of mass entertainment</p>	<p><i>weapons of * destruction (no mass)</i> weapons of modest destruction weapons of scant destruction weapons of ass destruction weapons of indiscriminate destruction weapons of ministerial destruction weapons of market destruction weapons of most destruction</p> <p><i>weapons of * (no mass, no destruction)</i> weapons of bluff, brinkmanship & bloody-mindedness weapons of bumps and bruises weapons of nasty scratches</p>
---	---

Figure 9: Variants of *weapons of mass destruction* (Feb. 20, 2003)

As the results in Figure 9 show, on the same day that the fateful term *weapons of mass destruction* was newly uttered, Feb. 20, 2003, we already find quite a number of creative variants emerging. Some of these will have come from texts posted prior to the date, of course, but it is still a remarkable crop. Meanwhile, the noun phrase *Mass Destruction* is replaced by phrases which emphasise the seriousness of the issue through apparent ridicule and bathos.

However, these results are relatively few when compared with the tremendous surge of additional creativity which occurred the very next day, as shown in Figure 10.

<p><i>weapons of mass * (no destruction)</i> weapons of mass banalization weapons of mass construction weapons of mass consumption weapons of mass deconstruction weapons of mass distortion weapons of mass distractions weapons of mass effect (WME) weapons of mass hysteria weapons of mass instruction weapons of mass irritation weapons of mass media (WMM) weapons of mass partying weapons of mass protection weapons of mass repression</p>	<p><i>weapons of * destruction (no mass)</i> weapons of catastrophic destruction weapons of class destruction weapons of complete destruction weapons of further destruction weapons of global destruction weapons of godlike destruction weapons of great destruction weapons of less destruction weapons of local destruction weapons of mad destruction weapons of massive destruction weapons of media destruction weapons of mischievous destruction weapons of murderous destruction weapons of nuclear destruction weapons of pattern destruction weapons of terrible destruction weapons of wealth destruction weapons of moose destruction</p>
--	---

Figure 10: Creativity for *weapons of mass destruction* (Feb. 21, 2003)

Typically, in the life-cycle of a word or phrase in text, a rise in lexical creativity is followed by a fall. Accordingly, for *weapons of mass destruction*, in the time period following the initial linguistic and historic event, we see vastly reduced creative activity in web text (retrieved from the same number of web pages), as shown in Figure 11.

weapons of mass * (no destruction)

weapons of mass photography
weapons of mass media (WMM)

weapons of * destruction (no mass)

weapons of 'grass destruction'
weapons of singing destruction

weapons of * (no mass, no destruction)

weapons of math instruction

Figure 11: Decreased creativity of phrase *weapons of mass destruction* (Feb.28, 2002 - May 29, 2004)

7. Rules of Creativity

Creativity, as seen in Figures 8, 9 and 10, is shown not to be random but to follow a clear set of conventions, involving substitution on the basis of phonological, morphological, semantic and other types of similarity as well as allusion.

Destruction is replaced by words that have

- difference of one phoneme - *distraction*
- same initial letter – *distraction, disruption, deception, discomfort*
- same prefix - *deception*
- same suffix - *distraction, obstruction, disruption, deception*
- shared semantics - *disruption, discomfort, terror, casualties*
- other – (for effect) e.g. *entertainment*

Mass is replaced by words that have

- difference of one phoneme - *ass*
- same initial letter – *modest, ministerial, market, most*
- semantic similarity/contrast – *indiscriminate, most (vs. modest, scant)*
- other – (also collocates of *destruction*) – *indiscriminate*

The change in the meaning of a word, or the adoption of an additional meaning or metaphorical extension, is arguably a type of creativity. As we

outlined in our initial section on Methodology, a change or extension in meaning is suggested to us by our automated software which signals a change in the collocational profile of a word or phrase. Novelty is in part a linguistic response to a real-world event, and the new use will be exploited in the media in creative play. For example, we would expect to find creativity in the wake of events such as the shift of primary reference of the acronym *ASBO* from organisations such as *Association of School Business Officials*, to the newly-introduced *Anti-social Behaviour Order*.

8. Case Studies of some recently popular neologisms in the UK Media

Words which most clearly demonstrate such change patterns are those which are coined or revived in response to world events and topics that capture the public (and media) imagination for a period. One area of lexis of recurrent interest in British culture and reflected in UK journalism is the predilection for characterising sub-sets of UK society, through their clothing, behaviour or other attributes and associations. This preoccupation is peculiarly but not exclusively British; it is clear that this phenomenon occurs in other media and other societies, particularly in the US, whence many formations later make their way into UK text.

Some of these lexico-social attributions have been chosen from two currently topical and thus linguistically-active areas: namely

- words for young people of low social status and life-style, e.g. *chav*, *neet*, and *hoodie*
- words for highly-paid people appointed to trouble-shoot in industry, government and other institutions, e.g. *tsar*

These words are new and possibly just at the start of a longer life-cycle. We shall focus on the salient productive and creative activity that each has produced in its relatively short life.

8.1 Neologism 1 - CHAV

The lexeme *chav* first emerges in our text in a series of dialectal uses in the nineties: *chavo* (1995), *chaveys* (1997), *chava* (1999) – where, like these, it is a Romany (and possibly also Essex rural) term for a young child. However, it is only in 2004 that we find the term revived in our data, this time in reference to a young British person of low education, having insufficient means to live away from home though sufficient to indulge in the purchase and wearing of hitherto socially-prestigious items of clothing, such as Burberry caps and other fashion accessories. The name gained currency swiftly through its publicity and incorporation into the name of the

website, *chavscum.co.uk.*, a site pandering to, while ridiculing, the eponymous membership.

It is clear from our data that the base term *chav* itself emerged in 2004, at once rose precipitously to the top frequency ranks of voguish media vocabulary, and has remained there ever since. Figure 12 reveals that the **frequency** was accompanied by quite a healthy level of productivity.

0403 <i>chav</i> 170	0502 <i>chav-rap</i> 1	0506 <i>chav-nots</i> 1
0403 <i>chavs</i> 95	0502 <i>chavdom</i> 1	0506 <i>chavsta</i> 1
0403 <i>chavspotter</i> 1	0502 <i>chavalanche</i> 4	0506 <i>chavtowns.co.uk</i> 2
0403 <i>chavtastic</i> 4	0502 <i>chav's</i> 3	0507 <i>chavification</i> 1
0412 <i>chavvie</i> 4	0503 <i>chav-tastic</i> 2	0508 <i>chav-goth</i> 1
0410 <i>chav-spotting</i> 1	0503 <i>chav-in-the-street</i>	0508 <i>chav-infested</i> 1
0410 <i>chavscum.co.uk</i> 1	0503 <i>chavistocrats</i> 1	0509 <i>chavocrat</i> 1
0412 <i>chav-fave</i> 1	0503 <i>chavish</i> 4	0510 <i>chav-associated</i> 1
0501 <i>chavved-up</i> 1	0503 <i>chavvy</i> 4	0510 <i>chav-style</i> 1
0502 <i>chavocracy</i> 1	0505 <i>chav-central</i> 1	0511 <i>chavy</i> 1
0502 <i>chavier</i> 1	0505 <i>chavettes</i> 1	
0502 <i>chaviest</i> 1	0506 <i>chavette</i> 1	

Figure 12: Inflections of *chav* emerging during 2004-2005

The range of variation in productivity based on the word *chav* in Figure 12 is high, containing a considerable range of the canon of available suffixes and other post-positional combining elements. There is a striking predominance of *chav* compounding with colloquial or ‘youth’ terms like *spotter*, *spotting*, *fave* (‘favourite’), *rap* and *central*; we also find *chav* in apposition to *goth*; and *chav*-based derivations achieved through combination with less conventional affixes and affix-like elements.

The **creativity** associated with *chav* in its new sense is considerable, given its short life as compared with the 17-year-old acronym *NIMBY*. We find the following word play in *Independent* text from 2004:

- 04.02 **Chavs and chav-nots**: the eternal divide. Charvers, skangers, scallies - Britain's burgeoning peasant underclass may be revolting. But a threat to civilisation?
- 05.02 A bickering **chavocracy** that marries in register offices should suit advocates of a bicycling monarchy
- 05.03 the tulip-shaped skirt becomes de rigueur at Pangaea or wherever it is soap stars and **chavistocrats** hang out these days
- 05.05 you would think that our society is on the brink of collapse from a **chavalanche** of scary teenagers.
- 05.05 Vogue may be hoping to broaden their brand and attract a whole new readership but it seems pretty unlikely that the average **chav-in-the-street** is going to pick up a copy of Vogue at the local newsagent with their packet of fags
- 05.06 After the Selfish Revolution in the Eighties, better-off people just shut themselves away behind their garden walls and moaned about the **chav-nots**.
- 05.09 the polo-playing **chavocrat** who wears Nazi gear to a party and gets into a drunken brawl outside a nightclub
- 05.10 It's not simply a case of comedy for the **chavs and chav-nots**

Figure 13: Creative variants based on *chav* from 2004-5

It is plain from the word-play in Figure 13 that this fertility results from *chav* being used to refer not just to the British social sub-class of poor youth, but also to the Royal Family and other rich and aristocratic members of British society, with the clear implication that the lives of the privileged are similarly tasteless and undistinguished. The neologistic derivations *chavocracy* (2), *chavistocrat* (3) and *chavocrat* (7) above actually refer respectively to Charles and Camilla, 'the rich', and to Prince Harry.

One of these puns, namely *chavs and chav-nots*, appears to have caught on particularly, by analogy with 'the haves and the have-nots', meaning 'the rich and the poor'. Further instances of this phrase appeared in other broadsheet newspapers for the period, e.g:

Times 05.02 **Chavs and chav-nots**. Julie Burchill says there is nothing wrong with chav culture

Times 05.04 Sadly, in the battle between **the chavs and the chav-nots**, the ones with the bling-bling make the most noise.

8.2 Neologism 2 - HOODIE

We find a total of 159 occurrences of '*hoodie*' in our news data between 1989 and end 2005,¹⁷ and the evolution in its referential and connotative meaning is exemplified in Figure 14. Its creativity lies primarily in the use of metonymic extension and allusive reference. It first occurs in 1992 as a colloquial term for a 'hooded crow' (1). By Dec. 2001, however, *hoodie* was adopted to signify metonymically the hooded jacket or top made of soft material primarily associated with particular groups of society. Initially, a *hoodie* was just the preferred clothing for boys, young men and their peer groups (2), before it moved on to refer to an identification with fashion and youth icons in June 2002 (3).

1. 92.09 Here live only deer and eagles, the odd fox and **hoodie** crows.'
2. 01.12 He buys a black '**hoodie**' with 'Alien Ant Farm' on the back
3. 02.06 his progress is tracked by an army of **hoodie-wearing**, baggy-jeaned pre-teens. 'It's Matt Hoffman', they whisper reverently.
4. 02.10 you can't buy a decent **hoodie** from the Issey Misake or Jean Paul Gaultier sites.
5. 03.10 Small bags. Impossible in the winter to tote round a zip-up **hoodie**, extra vest...
6. 04.01 But Waynetta Slob [TV character] wore matching jogging pants and **hoodie**, if you catch my drift.
7. 04.11 How would anyone breaking a **hoodie Asbo** be caught? The hood turns any gentleman into a hoodlum.

Figure 14: Evolution in *hoodie* references in *Independent/Guardian* data 1989-2005.

Thereafter it was used to refer to adult fashion, as in (4) above (Oct. 2002); subsequently restricted to young female adult fashion but in the context of the need for warmth (see (5), Oct. 2003). It was not until Jan. 2004 (6) that the term *hoodie* was associated with the social underclass, even though it also continued to be used in reference to fashion items associated with sport, musical performers, ghetto stars, boy singers and music festivals. In this early period, only one significant collocate for *hoodie* was generated from its few occurrences by our analytical software, namely *crows*.

From June 2004, however, *hoodie* increasingly functioned as a negative evaluation of youth fashion sense, and in Nov. 2004, it adopted its first association with anti-social behaviour (7). From Nov. 2004, the negative sociological connotation of the term can be calculated by our analytical software. Its collocational pattern significantly increased and changed - no longer *crows*, but a list of collocates redolent of its new meaning, as shown in Figure 15. The accompanying figures represent respectively 'corpus frequency', 'plus/minus 1 span frequency' and 'collocate z-score', three statistical measures which combine to rank the collocates for significance.

wearing	4470	15+0	13.707
cap	1712	1+9	8.922
teenager	2109	5+2	5.930
ban	5974	1+5	4.829
baseball	774	0+5	3.981
wore	1298	5+0	3.968
black	13345	4+1	3.679
Bluewater	127	2+2	2.997
jeans	1078	1+3	2.978
debate	8248	0+4	2.832

Figure 15: Collocational pattern for *hoodie* in Nov. 2004

Similarly, until Feb, 2005, the plural form *hoodies* was too infrequent to have any statistically significant collocates. But thereafter, there is a stretch of increased frequency of occurrence as follows:

05.05	>	48
05.06	>	32
05.07	>	13
05.08	>	8
05.09	>	24
05.10	>	6
05.11	<	26

which is accompanied by an upsurge in significant collocates, revealing a clear change in connotation, as can be seen in Figure 16, with new items such as *ban*, *banned* and *yobs*.

wearing	3319	24+0	22.439
caps	799	1+8	7.945
ban	4090	8+1	7.724
baseball	586	2+6	6.964
boys	4367	7+1	6.735
wear	2700	6+1	5.853
banned	2370	4+2	4.887
young	19320	4+2	4.149
yobs	168	3+2	3.993
youths	606	4+1	3.975
teenagers	1773	5+0	3.927
shopping	2241	3+2	3.908
people	79115	6+0	2.262

Figure 16: Collocational profile for *hoodies* in Feb. 2005

The first outing for the doubly metonymic use of the term *hoodie* was in May 2005. The meaning of the term was transferred from the hood to the jacket and now to the wearer of the jacket. It was by then also imbued with the sense of the wearer being socially undesirable. In the example in question, we furthermore see the term used allusively, applied not to the usual hooded skateboarder or footballer, but to a tennis player.

05.05 As a result, Mancini is now officially as welcome in Genoa as a **hoodie** in the home counties

The neologisms *chav* and *hoodie* each evaluate and to some extent caricature members of the social underclass by reference to their choice of clothing, assigning respectively the vices of poor taste and fecklessness; and lack of social responsibility and petty criminality. They conjure up amusing images, have great metaphorical application, and allow one section of society to make jokes about and feel superior to another. They are thus ready candidates for use by the public, and not simply by the media, and probably for this reason have become so productive.

8.3 Neologism 3 - *NEET*

By contrast, the acronym *NEET*, (standing for ‘not in education, employment or training’) was coined not by the media but by the government Dept. of Health to classify still another social sub-set of British youth. Constant changes in the educational and social service provision in the UK lead to such a plethora of terms that acronyms are widely used to condense information. *NEET* is a bureaucratic term, neither particularly amusing nor readily exploitable by the public or even the media, so it is less productive than *chav* or *hoodie*. Until 2004, the word *neet* was simply a transliteration of the regional dialectal variant for ‘night’. We find in our news data six such instances, e.g (1) in Figure 17. Towards the end of 2004,

the new acronym *NEET* emerged (2-4). As in the earlier cited case of *NIMBY*, *NEET* immediately moved to lower-case orthographic status, *neet*, though again this was not the consequence of widespread currency (Renouf, 1993a). The only instances of productivity (5-6) show a minor syntactic shift, from noun to noun modifier.

1. 02.01 I've taken part in many a merry **neet** with hunting folk
2. 04.10 The health ministry estimates over 500,000 so-called **neets**
3. 04.11 Japanese researchers know their nation has **neets** too.
4. 04.11 The no. of young Japanese **neets** is an estimated 600,000
5. 04.11 The **neet** group in Britain tends to be from the lower socio-economic groups
6. 04.11 The **neet** phenomenon may be fundamentally different in both countries

Figure 17: Instances of *neet* since late 2004

With the meagre opportunity for creative use afforded by *NEET*, there has been just one single wordplay, as shown in Figure 18, a pun trading on the somewhat marginal phonological similarity between *neet* and the word 'nit', meaning 'louse'.

- 04.11 For most people, the suggestion that a sizeable number of British teenagers has a **neet** problem will prompt images of strong hair lotions and fine-tooth combs

Figure 18: Creativity of acronym *neet* since its inception in late 2004

8.4 Existing Term with New Meaning - *TSAR*

Many long-established words can lie fallow until a new use is found for them. The word *tsar* is an example.¹⁸ This used to have a fairly low-key but consistent life, bumping along in the bottom reaches of the lexicon, chiefly in relation to events in the Soviet Union.

It has become one of many similar terms - *guru*, *supremo*, *pundit*, *trouble-shooter* and so on - which have in recent years come back into fashion, used to characterise the New Labour government cronies, media personalities and other personages who oversee the functioning of a public institution or commercial body, or who are experts in some field. The terms often even appear in juxtaposition:

- 99.01 it is impossible not to trip over a **Government-appointed tsar, task force, regulator, or watchdog** surveying our lives
- 00.07 the arts council **supreme / guru / tsar** Gerry Robinson

Among these *tsar* and *pundit* are the most productive and creative, probably due to the linguistic accident that they are more easily manipulated and lend

themselves to affixation more readily than their synonyms. Since *tsar* is the more in vogue of the two, we shall select *tsar* as our object of investigation.

Our ‘time series’ graph of word frequency in Figure 19 shows that in recent history, specifically from 1989, this term has hovered in the lower reaches of the lexicon, initially restricted in reference to the former Soviet Union, to its imperial heritage and its past and present political leaders. As in Figure 8 for *Weapons of Mass Destruction*, the frequency pattern is represented as a dotted line for simple ‘frequency per million words’, and as a solid line for the more interpretable ‘smoothed frequency’.

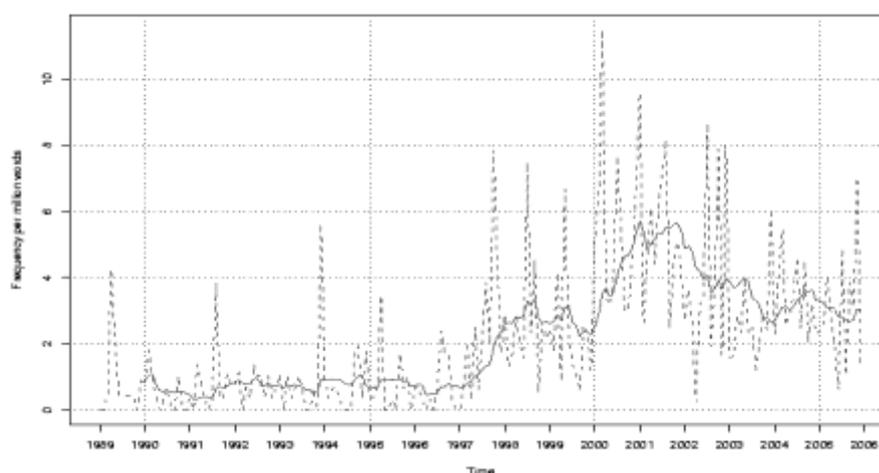


Figure 19: Time series frequency plot for *tsar* from 1989-2005

The long, steady life-cycle of the root form *tsar* has been shared with long-established inflections such as *tsarist*, *tsaritsa/s*, mostly transliterations of Russian titles used with literal reference. It can be seen in Figure 20 that the bulk of productivity has been rather conventional derivation, with a few more recent exceptions such as *e-tsar*, and a rash of creative wordplay, chiefly punning on *tsar/star*, which we shall consider in more detail later.

anti-tsarist 2	tsar's 94	tsarina's 3
anti-tsarists 1	tsar-gazing 1	tsarinas 10
co-tsar 1	tsar-hating 1	tsarism 17
e-tsar 1	tsar-in-waiting 1	tsarism's 2
ex-tsar 1	tsar-like 3	tsarist 234
mini-tsar 2	tsar-wars 1	tsarist-era 8
neo-tsarist 1	tsardom 9	tsarists 4
pro-tsarist 1	tsarevich 7	tsaritsa 1
super-tsar 1	tsarevna 1	tsaritsas 1
tsar 1702	tsarina 41	tsars 308

Figure 20: *tsar* and its inflections from 1989-2005

The frequency graph in Figure 19 shows that the stably low frequency of the word *tsar* suddenly rose steeply in our text from 1996, peaking in 2000 and 2001. The peaks corresponded to periods when it was the fashion for the Labour Party and other institutions to appoint public figures to oversee most aspects of national life. This historical context has had a particular effect on the productivity pattern for *tsar*.

Statistically significant collocates positioned immediately to the left of *tsar* in our data from 1996-2005 are ranked as in Figure 21:

drugs	job	sports	anti-corruption
drug	intelligence	economic	Drug
health	behaviour	car	disability
cancer	patient	anti-drugs	Cancer
homelessness	digital	ethnicity	Children's
Drugs	housing	equality	Crime
Russian	heart	economics	healthcare
care	maths	design	countryside
children's	IT	competition	Food
counter-terrorism	transport	policy	poverty
food	homelessness	Asbo	emergency
people's	architecture	film	arts
crime	rural	ex-drugs	TV
security	rail	anti-drug	education

Figure 21: Significant left-hand collocates for *tsar* from 1996-2005

What Figure 21 tells us is that, from about 2000, the productivity of *tsar* primarily takes the form of a multitude of compounds of the ‘*government's transport tsar*’ kind, with *tsar* as an all-purpose noun-head carrying a range of noun-modifiers. Together, these form designated roles and titles specifying a brand of advisor, usually in the public eye, highly-placed and highly paid, but exceptionally honorary, whose role is to monitor areas of practice and expenditure and solve problems to try and improve performance in different arenas, from government to sports. The term *tsar* seems to impart a certain glamour and power to these roles.

In contrast, in Figure 22, the collocational partners of members of the same semantic class - *guru*, *supremo* and *pundit* – whilst also designating assorted experts, bosses and top dogs, do not carry the same degree of national or governmental status or authority as do the *tsar* compounds.

guru	supreme	pundit
PR	PR	TV
management	policy	television
style	security	football
lifestyle	election	media
advertising	transport	political
design	marketing	racing
investment	media	soccer
fashion	press	radio
marketing	arts	Sky
Indian	communications	sports

Figure 22: significant left-hand collocates for *guru*, *supremo* and *pundit* from 1996-2005

Many compounds of *tsar* are longer multi-word units than those suggested by the collocational profile. Examples are shown in Figure 23 below:

arms control tsar
 EU foreign policy tsar
 the treasury's banking competition tsar
 the government's workplace equality tsar
 the European Commission's competition tsar
 hospital design tsar
 digital television tsar
 learning disability tsar
 The NHS information technology tsar
 the government's university admissions tsar
 the first criminal justice IT tsar

Figure 23: Multi-word compounds based on *tsar* from 2000-2005

Moreover, the de-contextualised listing of significant collocates in Figure 23 largely obscures the many humorous and parodic coinages. A sample of the effects achieved, both consciously and sub-consciously, is catalogued under the following headings:

- **compounds which humorously juxtapose colloquial words and basic concepts with the elevated *tsar***

- 97.10 South Boston has a '**trash tsar**' - in charge of street cleaning
- 97.10 **Bug tsar** appointed to stop millennium computer crash
- 97.10 **Rat tsar** needed to crack down on the booming population of rodents

- **compounds which label ludicrous official roles:**

(passim) 'NHS waiting lists tsar'; 'waiting list tsar'

- **compounds which apply ludicrous epithets to official roles**

- 01.07 Jean Glavany, the French agriculture minister and **tree tsar**
- 03.11 Every NHS trust is to appoint a **bug-buster tsar**

- **compounds which humorously paraphrase official terms**

homelessness tsar

- 98.07 the '**streets tsar**'
- 00.12 The government, **roofless tsar** and all
- 01.04 **Rough sleepers tsar** sweeps ahead with targets

older people's tsar

- 98.12 Call for '**grey tsar**' to champion the elderly
- 05.11 The **old-age tsar**, Professor Ian Philp

de-regulation tsar

- 00.12 The government's **red tape tsar**, Lord Haskins
- 03.03 Lord Haskins is the **multi-purpose regulation/farming/euro tsar**

food tsar

- 00.09 Loyd Grossman to be hospital '**foodie tsar**'

- **compounds which parody the official use of *tsar* in miscellaneous journalistic ways**

These generally involve making fun of public figures by characterizing them ironically as *tsars* - distinguished and powerful – while simultaneously highlighting their insignificance, triviality or other shortcomings, particularly arrogance. See for example:

- 99.03 Milton Hershey had become America's **caramel tsar**
- 00.08 a friend of lipless, **rightwing morals tsar** William Bennett
- 00.12 Labour's watchdog; a kind of **anti-William Hague tsar**
- 01.03 the **after-dinner tsar** Kenneth Clarke

- **compounds as part of everyday journalese**

- 99.01 The Foreign Office security consultant, or '**kidnap tsar**'
- 01.10 sources close to Radio 5 Live's **phone-in tsar**
- 01.10 In April, he appointed a gay activist as **Aids tsar**

Our selection of the term *tsar* from within its semantic group is not accidental; the term is fertile ground for the study of pun creation. Some words lend themselves more readily to wordplay than others: *guru*, for example, is less ambiguous, phonetically or otherwise, and so inherently less exploitable. For *tsar* we find that the phonetic similarity between *tsar* and *star* is heavily traded upon. Indeed, on some occasions it is almost impossible to avoid punning, or to know whether it is a deliberate device. See, for instance, the marginal case:

- 11.05 Learning disability tsar Rob Greig revealed that ministers were to appoint a learning disabled **co-tsar**

Figure 24 displays some of the deliberate creativity found in our corpus.

- 03.89 History has consigned previous US **tsar-wars** to oblivion
 03.91 exhibition ‘Twilight of the Tsars’ no chance for **tsar-gazing**
 05.92 he seems to have seen **tsar quality** in him
 03.98 Nowadays he'd be a **rock tsar**
 05.99 The **falling tsar**. Yeltsin brooks no rivals at court.
 05.99 In Cannes, **a tsar is born**: The Russians are coming to the film festival
 09.99 **A tsar rises in the east**: Helen Wallace on classical music
 10.99 Spoken like a **rising tsar**
 03.00 **Porn tsar** to save Utah from sex
 08.00 Managed by New Labour **pop tsar** Alan McGee
 10.00 Putin piles on **super-tsar** pretensions
 04.01 **When you wish upon a tsar**. Lithuanian magnate opens theme park.
 01.02 A **wandering tsar** blesses union between human and alien
 10.02 an army of unelected policy advisers - Blair's **galaxy of tsars**.
 02.03 New **Tsar is born** as artist gives facelift to regal past.
 10.03 A sprinkling of **tsar quality**. Sarah Eldridge meets the man
 06.04 **A tsar is born**: Prince Michael of Kent a cult hero in Russia.
 03.05 rise and fall of a **pop tsar** - pop's most colourful character.
 06.05 Bulgaria's **falling tsar** Simeon Saxe-Coburg faces bleak exit.
 08.05 **Tsar attractions**. From couture to high street, Moscow style everywhere
 11.05 **Tsar turn**: our rag trade king rocks Russia.
 11.05 Government has **tsars in its eyes**. Learning disabled tsar to be appointed

Figure 24: Puns on the vogue word *tsar* from 1996-2005

This shows that over the years, virtually every phrase and cliché involving the word *star* has been exploited, from *star-gazing*, to *star quality*, *rock star*, *rising star*, *porn star*, *a galaxy of stars*, *pop star*, *super-star*, *wandering star*, *star attraction*, *stars in one's eyes* and *star turn*. In addition, many of the puns have been wrung from song lyrics such as ‘(Catch) a Falling Star’ and ‘When you wish upon a star’, and film titles such as ‘Star Wars’ and ‘a Star is born’. Some allusions are more loosely made; it is not in fact ‘a star’ which *rises in the east*, but ‘the sun’. It is clear that these puns have been strategically created to add a humorous or evocative allusive quality to the writing, and are manipulated to fit the new contexts semantically. That the majority function as titles or sentence openers indicates their purpose as awakeners of reader interest.

13. Conclusion

In this chapter we have sought to make explicit some of the insights which can be revealed about the nature of lexical productivity and creativity in text

by employing a research methodology designed to trace lexical activity over time as it is used in UK journalistic text. We have focused on key concepts of ‘productivity’ and ‘creativity’, which we have defined within the terms of modern diachronic corpus study. We have conducted a modest case study of selected neologisms, and within the limitations of their relatively short life-spans and correspondingly limited occurrences, we have tried to illustrate their patterns of use.

Stepping back from this catalogue of minutiae about the individual words, one is able to make a few more general and hopefully thought-provoking observations, which may encourage the reader to revisit the data here and elsewhere in a fresh light. In tracing productivity and creativity chronologically over 16 years of journalistic text, one is struck by a number of findings which emerge from this methodological and philosophical approach.

Probably the central revelation is that words have a life-cycle consisting, in the most general terms, of birth or re-birth, followed by gentle or steeper upward trajectories in frequency of use and leading to brief or lengthier moments at the zenith of popularity, after which they take faster or slower downward paths, until they reach a stable level of use. During this life-cycle, words which make a sufficient impression on the public imagination will also spawn a number of productive and creative variants. Indulging in hyperbole for an instant, this is evidence of the marvellous self-generating and self-renewing power of language, words like stars soaring into the linguistic firmament in a blaze of popularity, spinning other words off hither and thither, and at length tumbling back down into the calm(er) reaches of servitude.

A related observation is that, for various reasons, not all words are equally destined for success. Some inhibitors have been suggested; another factor is that words do not all have universal franchise. Some stem from grass-roots movements in society, some from developments at more rarified social or scientific levels; and some are media coinages. The ability of a word to establish itself as an icon which furthers the media cause is the most crucial criterion for its securing a place in the lexicon of the press. The acronym *NIMBY* is a political term for a somewhat drab concept, and little used by the public or journalists. In contrast, a phrase like ‘*annus horribilis*’, coined by the Queen’s speech writer in 1992, whilst opaque to the majority of the British public, has fared better. Though foreign to many in meaning, etymology, provenance (from *annus mirabilis*¹⁹), and stylistic intention, its meaning is still guessable and so of service to journalists. Thus, following a brief if heady ascent and a fairly swift decline, it has settled at a stably low but resilient level of productivity, where it seems to be come in handy for journalists in their characterisation of key events at the close of each year. Our statistics on its occurrence indicate around double its average number of monthly occurrences between November and January, assigning an overall ‘seasonality’ score of ‘.04’, which means it is less seasonal than the terms

Wimbledon or *Christmas*, each with top scores of 0, but still quite significant.

A final observation is that our diachronic, empirical study of the data reveals and traces parallels between language and contemporary world events, and thus acts as a window on contemporary culture; in particular on recent world events and on aspects of youth culture as reflected in the media. These data and these findings raise issues which tip over into neighbouring disciplines: sociolinguistic questions, about the circumstances of language use and the fortunes of words; psycholinguistic questions, about the processes underlying the subtleties of mutual comprehension of allusion and wordplay; stylistic questions, about the preference for certain choices and not others, and so on. It is hoped that these concluding thoughts may inspire the reader to investigate such mysteries further.²⁰

ENDNOTES

¹ broadsheet newspapers comprising: *the Times* (timesonline.co.uk), *The Telegraph* (telegraph.co.uk), *the Guardian* (guardian.co.uk), *the Observer* (observer.co.uk), and/or *the Independent* (independent.co.uk)

² <http://www.webcorp.org.uk/>

³ In a series of major research projects funded by the UK Engineering and Physical Sciences Research Council, Department of Trade and Industry, commercial companies and successive universities; see <http://rdues.uce.ac.uk/>.

⁴ New word senses were researched in the AVIATOR (Analysis of Verbal Interaction and Text Retrieval) project <http://rdues.uce.ac.uk/aviator.shtml>

⁵ Sense relations in text were researched in the ACRONYM project (Automatic Collocation Retrieval of 'Nyms') <http://rdues.uce.ac.uk/acronym.shtml>

⁶ New words entering text were researched in the APRIL project (Automatic Prediction of Innovation in Language) project <http://rdues.uce.ac.uk/april.shtml>

⁷ *Neologisms* - As the new text flows through the filters, a word is deemed to be new if it has not appeared before in an accumulating master word list. It is then recorded as new and appended to the master list with accompanying records of the circumstances of its occurrence (and subsequent recurrence).

New Word Senses - The meaning of an existing word is deemed to be new if it begins to be surrounded by collocates which do not match those in the established 'collocational profile' or wordprint which is held for each word.

Sense Relations - Two (or more) words are deemed to be sense-related (e.g. synonymous) by establishing that they typically appear in text within very similar 'collocational profiles' or wordprints.

Productivity and Creativity - Words encountered for the first time are analysed and classified as candidate compounds, inflections, derivations and so on by software filters based on dictionaries and by character-level morphological analysers. This gives a picture of the pattern of productivity and creativity in text as a whole.

⁸ The corpus linguist who works diachronically tends to view productivity as an observed fact, and is not concerned with it as a probabilistic phenomenon in the way that theoretical linguists are (notably Aronoff, 1976). However, productivity is a probabilistic matter for corpus linguists working with sparse text such as early documents.

⁹ This is not 100% accurate for the reasons of morpheme-boundary ambiguity mentioned above, but we have used a unique and sophisticated backward-forward character-level chart parser to achieve the lowest-cost and thus most likely analysis in each case.

¹⁰ For the first instance of *symbiosis* in a transferred sense, OED online cites: Misselden, Edward (1622) in *Free trade; or the meanes to make trade flourish*: “To study and inuent things profitable for the publike Symbiosis”. OED does not cite *symbiotic* in a figurative sense before 1970: *Nature* 6 June 905/1: “Throughout its auspicious history the Botanical Society of Edinburgh has had a symbiotic relationship with the Royal Botanic Garden”.

¹¹ For the first instance of *synergy* in a transferred sense, OED online cites Heylin, Peter (1660), in *Historia quinqu-articularis; or a declaration of the judgement of the Western churches, particularly of the church of England, in the five controverted points reproached by the name of Arminianism* (1.9): “They speak only of such a **Synergie**, or cooperation, as makes men differ from a senseless stock, or liveless statua, in reference to the great work of his own conversion”. Meanwhile, the variant *synergistic* is cited in OED only in 1965 for Ansoff, Harry Igor, in *Corporate strategy: an analytic approach to business policy for growth and expansion* (v. 76): “This step certainly provides for some of the **synergistic** interactions”.

¹² As we know, inverted commas can be applied at any point, not just to signal unfamiliarity or assumed unfamiliarity on the part of the reader, but also authorial distance, etc. The enduring strangeness might also be due to the multiple technical terms deriving from this one etymology, introducing ambiguity which could further inhibit productivity.

¹³ See also ‘discontinuous collocational frameworks’, in Renouf A. and Sinclair J. (1991), *Collocational frameworks in English*. In: *English Corpus Linguistics: Studies in Honour of Jan Svartvik, Aijmer K. and Altenberg B.* (eds), Harlow: Longman, 128143.

¹⁴ *NIMBY*: a 1980s coinage brought to general attention in 1988 by a Conservative minister, Nicholas Ridley in reference to hypocritical colleagues supporting undesirable developments while opposing their implementation in their own constituencies.

¹⁵ *Nimbys -come-lately* - meaning approximately ‘someone new in town, whose credentials have not been proven’. From a song called “Johnny Come Lately”, from a 1943 film, from a 1839 publication (which read “She may be a Johnny-come-lately on the board, but she’s doing a fine job with publicity.”)

¹⁶ Some data for *Weapons of Mass Destruction* are also displayed on our web-site

¹⁷ Available in the US folk lexicon since the 1870s, semantically derivable from ‘hoodlum’ (etymology unknown), shortened form ‘hood’, specifically associated with organised crime, the term waited until a century later to appear in our UK data, and even then, it is first used in reference to the US (see *hoodies*). Some theories also associate it with *neighbourhood*.

¹⁸ What is said about the creativity of *tsar* does not hold for its spelling variant *czar*.

¹⁹ Dryden (1667) wrote a poem entitled ‘*Annus Mirabilis: the year of wonders 1666*’, in which he opines that the 1665 plague and 1666 fire of London could have been worse, in part no doubt because he himself survived both.

²⁰ Researchers interested in working with our data are welcome to access, occasionally via password, any of our systems at <http://rdues.uce.ac.uk/>

References

- Aronoff, M. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press).
- Baayen, R.H. and Renouf, A.J. 1996. ‘Chronicling the Times: Productive Lexical Innovations in an English Newspapers’, in *Language*, 72.1. 69-96.
- Kehoe, A. & Renouf, A.J. 2002. *WebCorp: Applying the Web to Linguistics and Linguistics to the Web*. World Wide Web 2002 Conference, Honolulu, Hawaii, 7-11 May 2002. (ISBN: 1-880672-20-0)

-
- Kehoe, A. 2006. 'Diachronic Linguistic Analysis on the Web using WebCorp' in Renouf, A.J. and Kehoe, A. (eds.), *The Changing Face of Corpus Linguistics*, 297-308. Amsterdam & New York: Rodopi.
- Morley, B. 2006. 'WebCorp: A Tool for Online Linguistic Information Retrieval and Analysis', in Renouf, A.J. and Kehoe, A. 283-296
- Morley, B. and Renouf, A.J. 2003. 'Linguistic Research with the XML/RDF aware WebCorp Tool' in *Proceedings of WWW2003, Budapest*. <http://www2003.org/cdrom/papers/poster/p005/p5-morley.html>
- Pacey, M., Collier, A. and Renouf, A.J. 1998. 'Refining the Automatic Identification of Conceptual Relations in Large-scale Corpora', in Charniak, E. (ed.), *Proceedings of Sixth Workshop on Very Large Corpora, Montreal, 15-16 August 1998*, COLING-ACL, 76-84.
- Pacey, M., Renouf, A.J., Davies, P. and Kehoe, A. (forthcoming): 'Monitoring Lexical Innovation in Journalistic Text Across Time'
- Renouf, A.J. 1993a. 'H.E.R.E. Today and 'here' tomorrow', unpublished paper delivered at 14th ICAME Conference, University of Zurich, Switzerland.
- Renouf, A.J. 1993b. 'A Word in Time: first findings from dynamic corpus investigation', in Aarts, J., de Haan, P., and Oostdijk, N. *English Language Corpora: Design, Analysis and Exploitation*, 279-288 Amsterdam and New York: Rodopi.
- Renouf, A.J. 1993c. 'Sticking to the text: a corpus linguist's view of language', in *ASLIB Proceedings, Volume 45/*, 131-136.
- Renouf, A.J. 1993d. 'What the Linguist has to say to the Information Scientist', in Gibb, F., (ed.), *The Journal of Document and Text Management*, vol. 1/2, 173-190.
- Renouf, A.J. 1993e. 'Making Sense of Text: Automated Approaches to Meaning Extraction', in *Proceedings of 17th International Online Information Meeting, 7-9 December 1993*, 77-86.
- Renouf, A.J. 1994. 'Corpora and Historical Dictionaries', in Lancashire, I. and Wooldridge, T. Russon (eds.), *Early Dictionary Databases*, University of Toronto, 1-8 Oct, 1993, 219-23
- Renouf, A.J. 1996. 'The ACRONYM Project: Discovering the Textual Thesaurus', in Lancashire, I., Meyer, C. and Percy, C. (eds.), *Papers from English Language Research on Computerized Corpora (ICAME 16)*. 171-187. Amsterdam & New York: Rodopi.
- Renouf, A.J. 1997. 'Tools for the Diachronic Study of Historical Corpora', in Nevalainen, T. (ed.), *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*, 'Mémoires de la Société Néophilologique de Helsinki', vol 52, 185-199.
- Renouf, A.J. and Baayen, R.H. 1998. 'Aviating among the hapax legomena: morphological grammaticalisation in current British

-
- Newspaper English', in Renouf, A. (ed.), *Explorations of Corpus Linguistics: Proceedings of the 18th ICAME Conference*, University of Liverpool, 21-25 May 1997, 181-189
- Renouf, A.J. 2003. 'WebCorp: providing a renewable data source for corpus linguists', in Granger, S. and Petch-Tyson, S. (eds.), *Extending the scope of corpus-based research: new applications, new challenges*. 39-58 Amsterdam & New York: Rodopi.
- Renouf, A.J., Kehoe, A. and Mezquiriz, D. 2003. 'The Accidental Corpus: issues involved in extracting linguistic information from the Web', in Aijmer, K. and Altenberg, B. (eds.), *Proceedings of 21st ICAME Conference, University of Gothenburg, May 22-26 2002*. Amsterdam & New York: Rodopi. 404-419
- Renouf, A.J. 2005. 'Phrasal creativity viewed from an IT perspective', in A. Hamm (ed.), *RANAM (Recherches Anglaises et Nord Américaines) n°38: Language chunks and linguistic units*. Université Marc Bloch, Strasbourg, 113-122.
- 2006: with Kehoe, A and Banerjee J. 2006. 'The WebCorp Search Engine: A holistic approach to web text search', in *Proceedings from the Corpus Linguistics Conference Series*, Vol. 1, no. 1, ISSN 1747-9398. www.corpus.bham.ac.uk/PCLC
- Renouf, A. & Kehoe, A. (eds.) 2006. *The Changing Face of Corpus Linguistics*, Amsterdam & New York: Rodopi.
- Renouf, A.J., Kehoe, A and Banerjee, J. 2007. 'WebCorp: an integrated system for web text search' in Nesselhauf, C. and Hundt, M. (eds.), *Corpus Linguistics and the Web*. 47-68. Amsterdam & New York: Rodopi.