

Exploring the Relationship between Language Change and Dictionary Compilation in the Age of the Collaborative Dictionary

Sharon Creese

Coventry University Priory Street, Coventry, CV1 5FB
creeses@uni.coventry.ac.uk

Abstract

The rise in collaborative ‘wiki’ dictionaries means that dictionary creation is no longer the purview solely of academics and publishing companies. Ordinary people can now create and share their own dictionary entries, whilst traditional publishing houses must compete against resources able to achieve levels of interactivity and immediacy that they simply cannot. These differences in the dictionary landscape may not be the only consequence of the rise of ‘wiki’ dictionaries, however; the very relationship between dictionary compilation and language change may be shifting, with the speed and ease of updating of ‘wiki’ dictionaries meaning that they not only reflect current use, but actually drive change.

This paper examines the possibility of this, through the findings of a pilot study featuring a new web-based corpus of youth neologisms, and media tracking of these new words. In it, I set out to determine the relationship between the *Wiktionary* definition and the grassroots use of particular words, as well as considering if and how this is changing as ‘wiki’ dictionaries become more and more firmly established.

Keywords: Wiktionary; wiki; collaborative dictionaries; language change; neologism; dictionary compilation, lexicography.

1. Introduction

Throughout history, the dictionary has always been a key tool in understanding how language should look and function. The rise of the Internet, however, and particularly the interactivity offered by Web 2.0, has fundamentally changed the dictionary landscape, with anyone now able to create and share their own ‘wiki’ contributions at the touch of a button (Meyer & Gurevych, 2012: 259; Leuf & Cunningham, 2005). The ease with which changes and additions can be made to these collaborative dictionaries means that they can be updated hundreds of times a day, offering a level of immediacy that cannot be achieved by mainstream electronic dictionaries. Though publishers may constantly add to and amend the entries in their dictionary wordlists, availability of this new information is governed by cost, meaning updates are often scheduled no more than four times a year.

The speed and ease of updating ‘wiki’ dictionaries opens up the opportunity for a more dynamic relationship between dictionary compilation and language change than has previously been the case, with the dictionary potentially not only reflecting language use, but actually driving change. Despite a growing body of literature on dictionary collaboration (see, for example, Meyer & Gurevych, 2012; Penta, 2011) this

possibility remains as yet unexplored. Evidence of such a shift in this relationship could prove valuable to dictionary publishers seeking ways to monetise and add value to their online offerings. Knowing that entry of a word into a 'wiki' dictionary leads to increased usage in the media could, for example, lead traditional publishers to consider working with the creators of 'wiki' dictionaries, in order to develop a stronger relationship with grassroots users of the language. This in turn might enable them to position themselves as more accessible than their competitors. Alternatively, a publisher that has featured new words which have gone on to be particularly active in the media might develop a marketing campaign around its success in recognising new words that stand the test of time.

This paper reports on a pilot study for a research project to examine the relationship between collaborative dictionary compilation and language change. It describes the design of a web-based corpus, WeBCoYN, to aid the identification of new words within teenage language. It subsequently discusses the pilot version of this corpus, and the process of media tracking potential neologisms in major newspapers and archives, to assess whether their use pre- or postdates appearance in the dictionary, and whether dictionary inclusion affects everyday patterns of use.

2. Wiktionary

Currently, the most influential 'wiki' dictionary is *Wiktionary*, launched in 2002. *Wiktionary* contributors come from all walks of life and educational backgrounds; they submit potential new entries to the dictionary by creating a new page featuring their word and its definition, which can be accepted as it is, edited and amended in the live file, or discussed in detail in the 'Tea Room' forum. These discussions can continue for weeks, and the entire conversation is available for others to review and join in, as is the 'revision history' showing changes made to the word's *Wiktionary* page. New discussions can be started at any time if a problem with an entry is identified, or a change in definition is proposed. A historical profile of the word's behaviour over time is therefore offered by the Tea Room combined with the 'revision history' page attached to each word. Revision histories comprise many lines of hyperlinks, every 'save' action having generated a new page in the history, accessed via a separate link. This can result in enormous amounts of loosely organised information, making it difficult to find evidence of a particular amendment.

Interestingly, although a long Tea Room discussion can provide some indication that a significant shift in meaning or usage has occurred, or that a new entry is controversial, some major changes seem to be accepted with little or no discussion, whereas minor issues can generate extensive threads. On the face of it, this often seems to depend on the individuals involved, some being more pedantic or prone to argument than others, some having a better command of English and some having more extensive knowledge of *Wiktionary* processes and content. It may also be that some people become overly concerned with the minutiae of an issue, or that

Wiktionary contributors' relative lack of linguistic knowledge and experience, as compared to that of professional lexicographers, deters them from entering into debate about complicated issues, and instead leads them to focus on less complex ones. This is an area which will be investigated in more detail during the main research project, for which this is the pilot study.

3. Materials and Methods

Online youth language was chosen for this project because young people conduct much of their lives in the electronic sphere, and may be responsible for taking neologisms coined to fill lexical gaps, for example in the technology marketplace (Janssen, 2013), and extending them into wider use. There is growing anecdotal evidence that young people play a major role in the spread and establishment of new words – see for example *Blorge*,¹ *The New York Times*,² and *Voxxi*.³ In time, WeBCoYN's corpus evidence may empirically demonstrate that this is the case.

The 16,000 word corpus used for this study is a pilot for WeBCoYN (the Web-Based Corpus of Youth Neologisms), a corpus of online 'youth' language. 'Youth' is defined in this case as those aged 12–25 (often also termed the 'teen' market, despite extending beyond the age of 19). WeBCoYN texts fall into two intersecting categories (Sinclair, 2004: 4):

- medium:
 - companion pages for 'teen' television programmes
 - online magazines/webzines
 - websites linked to trending franchise (e.g. the *Twilight* series)
 - independent 'teen' blogs
- type:
 - articles/features
 - biographies
 - personal comments (short entries referring to a previously mentioned topic)
 - blog posts (longer pieces on a new topic, possibly generating comments).

All texts are categorised according to the intersection between medium and type (see Figure 1). In the pilot study, each cell is approximately 1,000 words long. Contextual information was collected for each text, including date of collection, original

¹ Accessed at: <http://tech.blorge.com/Structure:%20/2009/08/31/teens-who-use-twitter-and-facebook-add-new-words-to-dictionary/>

² Accessed at: http://www.nytimes.com/2012/02/28/science/young-women-often-trendsetters-in-vocal-patterns.html?_r=2&

³ Accessed at: <http://www.voxxi.com/new-times-new-generations-new-words-genya-mujer/>.

publication date, and, where possible, the author's age, gender, location, and education level.

Texts for the pilot corpus were collected using Google searches and manual reading of websites to identify suitable sections (for the full study, a web 'crawler' programme will be used to automate this process [Fletcher, 2013: 5]). Texts were then POS (part of speech) tagged using Wmatrix software (Rayson, 2008), and manually tagged for potential neologisms, that is, words that looked 'new'.

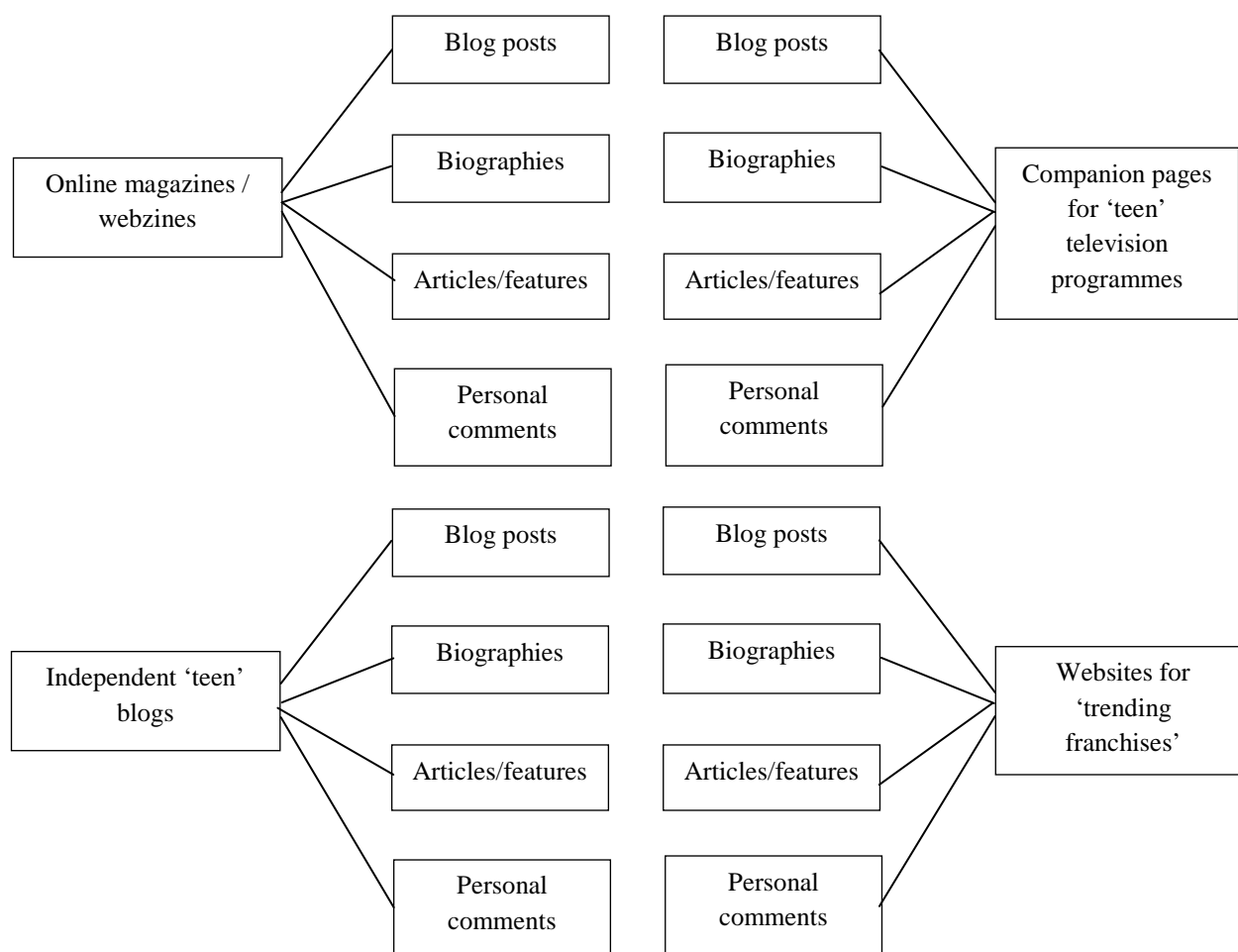


Figure 1. WeBCoYN cell structure (each cell contains an approximately equal number of words).

Once compiled and checked, the pilot corpus of 16,567 words was run through the Range programme,⁴ to exclude the most common 2,000 everyday words, and the top 1,000 academic words, as defined by the GSL (General Service List) and AWL (Academic Word List) (West, 1953; Coxhead, 2000). The resulting list of 2,452 words was then manually filtered, removing duplications, proper, place and trade names, obvious misspellings, and words which were clearly already well established. New

⁴ Accessed at: <http://www.victoria.ac.nz/lals/about/staff/paul-nation>.

word senses initially tagged as potential neologisms but which had been excluded during the Range filtering process on the basis of the original sense (for example ‘fetch’ and ‘genius’ [see Table 1]) were returned to the list, and the remaining 289 words were checked against eight dictionaries, to determine approximately when they entered the lexicon.⁵

Dictionaries from a number of different sectors were chosen here, in order to see whether new words appeared more quickly in standard reference works, in those aimed at second language learners, in non-British English dictionaries or in collaborative ‘wikis’:

- *Concise Oxford English Dictionary* (Soanes & Stevenson, 2006)
- *Macmillan English Dictionary for Advanced Learners of American English* (Rundell, 2002)
- *Macmillan English Dictionary for Advanced Learners* (Rundell, 2007)
- *Merriam-Webster* (2013) (online, accessed at: <http://www.merriam-webster.com/>)
- *Oxford Dictionaries* (2013) (online, accessed at: <http://oxforddictionaries.com/>)
- *Macmillan Dictionary* (British English/American English) (2013) (online, accessed at: <http://www.macmillandictionary.com/>)
- *Oxford English Dictionary* (OED) (2013) (online, accessed at: <http://www.oed.com/>)
- *Wiktionary* (2013) (accessed at: http://en.wiktionary.org/wiki/Wiktionary:Main_Page)

All words appearing in a dictionary before 2008 were deleted, since they can no longer be considered ‘new’, as were terms used only in an Internet context, unless they had already entered the dictionary and become established beyond their original sphere (e.g. ‘LOL, which entered *Wiktionary* in 2003, and appears in all of the dictionaries above). Google searches were then conducted to find evidence of use of the remaining 43 words. Evidence of significant usage, generating, for example, several pages of valid search results in multiple mass-media/social media/gaming contexts, or use in three or more ‘reputable’ sources (for example, websites produced by legitimate publishers), was deemed sufficient to consider the word ‘in use’. From the original 2,452 potential new words, 24 met these criteria, however only 14 had made it into a dictionary (in most cases, *Wiktionary*) and can therefore be considered established neologisms (see Table 1).

⁵ Unfortunately, online dictionaries produced by traditional publishers do not feature inclusion dates, and unlike *Wiktionary*, details of their lexicographical processes are not available to readers, meaning there is no way to know how, why or exactly when they were included.

Neologism (created since 2008)	Meaning	Part of Speech	Date entered Wiktionary	OED (online) 2013	OD (online) 2013**	MW (online) 2013**	MED (online) 2013**
fav	favourite	adjective	Feb-08	N	N	N	N
fetch	cool	adjective	Jan-13	N	N	N	N
genius	impressive	adjective	N	N	Y	N	N
girlchild	female child	noun	Feb-10	Y	N	N	N
gravatar	automatic avatar	noun	Mar-12	N	N	N	N
homeschooler	someone who is homeschooled	noun	Mar-09	Y	Y	Y	N
liveblogging	writing a real-time blog	verb	Dec-08	Y	Y	N	N
mischief- maker	creator of mischief	noun	N	Y	Y	N	Y
OMG	oh my god	exclamation	May-08	Y	Y	Y	Y
pre-visualize	imagine something before creating it	verb	N	Y	N	N	N
quick-release	single action release mechanism	adjective	N	Y	Y	N	N
sooo(o)*	emphatic version of 'so'	adverb	Aug-08	N	N	N	N
teared up	started to cry	verb	Jun-09	N	N	N	N
teenhood	period spent as a teenager	noun	Dec-09	Y	N	N	N

*This entry includes the variant 'soooo' found in the corpus, as indicated in brackets.

**OD = Oxford Dictionaries online; MW = Merriam-Webster online; MED = Macmillan Dictionary online

Table 1. Neologisms identified through analysis of WeBCoYN.

Twelve of these neologisms returned a frequency of one, with only 'sooo' (and the variant 'soooo') and 'OMG' appearing more than once (see Table 2). Given the size of the pilot corpus, it is perhaps unsurprising that the frequencies are so low. To get a wider view of the use of these words, they were also examined in Sketch Engine's SiBol/Port newspaper corpus.⁶

⁶ SiBol/Port draws data from three specific years: 1993, 2005 and 2010. Accessed at: https://the.sketchengine.co.uk/bonito/run.cgi/first_form?corpname=preloaded/sibolport_1;

Frequencies:	WeBCoYN	SiBol/Port
s000	3	65
s0000	2	63
s000(o) total	5	128
OMG	4	154

Table 2. Neologism frequency comparison – WeBCoYN and SiBol/Port.⁷

Five newspapers were chosen for media tracking, to cover the broad spectrum of target audiences (in terms of education level and socio-economic group) for this medium within the UK. In all cases, it was the online version of the newspaper that was consulted:

- *The Independent* (<http://www.independent.co.uk/>)
- *The Guardian* (<http://www.guardian.co.uk/>)
- *Daily Mail – Mail Online* (<http://www.dailymail.co.uk/home/index.html>)
- *The Sun* (<http://www.thesun.co.uk/sol/homepage/>)
- *Daily Express – Express* (<http://www.express.co.uk/>)

In addition to the main media tracking, digital newspaper archives were also interrogated, through the British Newspaper Archive.⁸

Newspapers were chosen to provide evidence of the use of new words/meanings in this pilot study because they are more able to keep pace with language change than books or magazines, since they are produced daily. They are also aimed at a wide cross section of the population – different ages, education levels, income brackets and social groups – meaning that new or amended words that appear in newspapers can be deemed to have moved beyond their original sphere, and become established within the language.

In all five newspapers, a search was conducted for the neologism, using the paper's online search engine.⁹ A number of problems were encountered, for example concerning the lack of consistency in how results are presented. *The Guardian* presents a list of the number of articles featuring the search word, broken down by

⁷ Size of corpora – WeBCoYN: 16,567 tokens, 3785 words; SiBol/Port: 387,585,716 tokens, 327,025,669 words.

⁸ See <http://www.bl.uk/>.

⁹ Since conducting the pilot study in April 2013, *The Independent* and *The Guardian* have changed their search functions. At the time of the original media tracking, the latter only searched mainstream articles; it now also includes data from interactive pages like blogs and comments. *The Independent* no longer limits initial search results to post-2010, and some articles included in the April 2013 results list are now excluded (presumably because the articles have been removed). *The Sun*, meanwhile, has rebranded its online presence to *Sun+* and now no longer allows for searching without creating a subscriber account. All of this means that conducting the same study in August 2013 could lead to different results than those reported here.

year, which the user can then click through to read.¹⁰ The other newspapers do not provide numerical results lists; *The Independent* simply says that the word has appeared ‘x times since 2010’, and provides links to the relevant articles, whilst *Mail Online*, *The Sun* and the *Express* merely give the total number of results, plus links, with no indication of the time frame. This problem was largely overcome by conducting manual year-on-year searches in *The Independent*, *Mail Online* and *The Sun* (using the ‘advanced search’ function), in order to obtain results comparable with those from *The Guardian*. The lack of an advanced search facility in the *Express*, however, meant it was not possible to do the same, and hence only flat figures were available, with no date context.

A further difficulty was that where the term under investigation is a new sense of an existing word (for example ‘fetch’ or ‘genius’ above), the number of search results is unmanageably high, since the search functions offered by these newspapers have no POS filter, and hence every instance of the word appears. Each article must then be individually examined, to determine if it contains the correct sense of the word. For the pilot study, this problem was resolved by selecting newly created words for media tracking, rather than new senses of existing words. This returned few enough results that each article could be individually checked, using corpus query software to generate concordance lines of all of the instances of the word, which could then be analysed for sense and meaning to ensure they were, indeed, the word under investigation.

For the main WeBCoYN study, a three-stage process will be employed to create a searchable corpus of newspaper articles containing the neologisms being media tracked. Firstly, files identified by the newspaper’s search engine as containing the relevant neologism will be automatically downloaded, to create a corpus of HTML files. A script will then be used to remove all HTML tags and output the files as pure text. Finally, these text files will be run through a POS tagger such as Wmatrix to add the structural mark-up required to enable identification of the correct use of the word, for example, the adjective form of ‘genius’ as opposed to the noun.

The two words chosen for media tracking in the pilot study (from the list in Table 1) were ‘gravatar’ (an automatic avatar) and ‘teenhood’ (the period of being a teenager).

Although both these words are in *Wiktionary*, neither of them has a discussion page in the Tea Room. This suggests that no-one has objected to the original definitions, and there has been no further development of the words. The use of these two words by teenagers, and their appearance in *Wiktionary* but their absence from most traditional dictionaries (online or print), makes them ideal candidates for

¹⁰ Following the changes to its search function, *The Guardian*’s results are now also presented differently, appearing as a chronological list, instead of year by year. This is less user-friendly than the previous format, and could hamper media tracking for the main WeBCoYN study.

examination of the impact of collaborative dictionaries on language change. They are new words which are still in the process of establishing themselves in the lexicon. By examining the frequency, date and context of their use in the media, we can consider the possible impact of entry into a ‘wiki’ dictionary on everyday use of a word.

4. Findings

Gravatar

‘Gravatar’ entered *Wiktionary* in March 2012. A blend of ‘globally recognised avatar’, it began life as a trade name, but is rapidly becoming a generic term. ‘Gravatar’ refers to an avatar linked to an email address via a central registration point; wherever that email address is used to post a comment on a website, the ‘gravatar’ is automatically imported.¹¹ Plugins are now available to allow ‘gravatars’ to be incorporated into independent sites.

‘Gravatar’ is a new word at the beginning of its lexical journey. So far, it has only entered *Wiktionary* (2013) (and *Wikipedia* [2013]); it does not yet appear in any of the other collaborative dictionaries, such as *The Free Dictionary* (2013) or the *Urban Dictionary* (2013) (although a film of the same name is included in the latter [2010]). ‘Gravatar’ is beginning to be used as an alternative to ‘avatar’, and it is possible that this may become more common as its use spreads from social networking and blogging sites, to more mainstream ones. Similarly, as users of the term grow older, they will likely carry the word with them, so we could reasonably expect to see ‘gravatars’, rather than ‘avatars’ on the comment pages of newspapers or other news outlets in the future.

Media tracking ‘gravatar’ in the five target newspapers returned no results, which is unsurprising given how new the word is and the fact that at present it remains firmly within the online sphere of use. (The British Newspaper Archive returned one result, but it was the name of a school, featured in an advertisement in 1883.) A Google search for ‘gravatar’, returned 170 million hits,¹² the first few pages being mostly blogging sites, Internet forums and compatible software.

1.	enabled sites such as this one. Using	gravatars	helps make our weblog a more friendly and personal
<i>Source: http://www.synchronoustechology.net/blog/how-to/set-up-your-gravatar/.</i>			
2.	activate the plugin, and it will add	gravatars	to your blog template and admin panel automatically
<i>Source: http://wordpress.org/extend/plugins/gravatar-favicon/.</i>			
3.	anyone know how I get my	gravatar	on my battlelog I've got it set up but just don't know
<i>Source: http://battlelog.medalofhonor.com/mohw/forum/threadview/2832654490161464530/.</i>			
4.	here Just wanted to say about your	gravatar.	We both have sketched birds out there. => Best
<i>Source: WeBCoYN pilot study, April 2013.</i>			

Table 3. Online concordances of ‘gravatar’.

¹¹ See <https://en.gravatar.com>.

¹² As at 18.4.13; by 18.8.13 this had risen to 211 million.

Table 3 shows concordance lines for ‘gravatar’ taken from these online sources, and the WeBCoYN pilot study. Lines 1 and 2 demonstrate the use of ‘gravatar’ in blogging contexts, giving guidance on how ‘gravatars’ can enhance websites. Both assume a level of understanding of what a ‘gravatar’ is and its purpose.

Concordance line 3 is taken from a gaming forum, and features a player who is seeking help with his ‘gravatar’. Many of the instances of ‘gravatar’ online are in forums or ‘help’ pages like this, offering advice on how to get the best out of a ‘gravatar’. Concordance line 4 (taken from the WeBCoYN pilot study) is slightly different, in that the reader of a blog is commenting on the ‘gravatar’ used by the blog’s author. Again, it is clear that the writer of the comment understands and is familiar with ‘gravatars’, and it is implied that the blog author is in a similar position. The fact that the comment is being made, suggests that the writer has seen and recognised the ‘gravatar’ from elsewhere on the web. This ‘transferability’ is, according to the company behind them, one of the key functions of ‘gravatars’.¹³

Teenhood

Although it also occurred only once in the pilot WeBCoYN study, ‘teenhood’ is a more established word than ‘gravatar’, having entered *Wiktionary* in December 2009 and already featuring in *OED* and the *Urban Dictionary*. It also appears in *UKWaC*, with concordances dating back to 2003.¹⁴ Despite this, media tracking of ‘teenhood’ returned only 20 results (from 2000–2012), the majority of which were confined to *The Guardian* (see Table 4), suggesting that it is still not particularly well established in the lexicon.

‘Teenhood’ also appears seven times in the British Newspaper Archive, with all instances carrying the same meaning. All of these come from the late 1800s, however (see Table 5). This suggests that while we may think of ‘teenhood’ as new, it is actually a word which enjoyed a brief period of use over a century ago, fell out of favour and was then reinstated, or was perhaps even created anew without awareness of its earlier existence. Unlike other reinstated words, such as ‘truthiness’, there is no indication in *Wiktionary* of this previous incarnation of ‘teenhood’.¹⁵

¹³ See <https://en.gravatar.com/>.

¹⁴ See <https://the.sketchengine.co.uk/bonito/run.cgi/first?iquery=teenhood&queryselector=iqueryrow&corpname=preloaded%2Fukwac2>.

¹⁵ See <http://en.wiktionary.org/wiki/truthiness>.

Year	<i>The Guardian</i>	Month	<i>The Independent</i>	Month	<i>Mail Online</i>	Month	<i>The Sun</i>	<i>Express</i>
2013	0		0		0		0	
2012	1	Dec	1	Nov	0		0	
2011	4	Mar, May*, Jun, Nov	0		0		0	
2010	1	Aug	0		1	Apr	0	
2009	1	Sep	2	Aug, Dec	0		0	
2008	2	Oct, Nov	0		1	Feb	0	
2007	2	Jun, Jul*	0		0		0	
2006	0		0		0		0	
2005	1	Aug	0				0	
2004	1	Jun*			0		0	
2003	0				0		0	
2002	1	Jul			0		0	
2001	0				0		0	
2000	1	Aug			0		0	
Total	15		3		2		0	0

*alternate spelling used: ‘teen-hood’

Table 4. Appearances of ‘teenhood’ in target media.¹⁶

Word	British Newspaper Archive	Date
teenhood	7	1870s
	1	1890s

Table 5. Appearances of ‘teenhood’ in digital archives.

Whilst ‘teenhood’ may not be strictly speaking ‘new’, the number of appearances found during media tracking was lower than expected, and was unexpectedly biased towards a single newspaper, *The Guardian*. Comparing the dates of these instances with the date of entry into *Wiktionary* – December 2009 – shows a marked increase in usage after inclusion. ‘Teenhood’ appeared in the five target newspapers 11 times between January 2000 and November 2009, and nine times from December 2009 to December 2012. Thus we see a doubling of the frequency of appearances, from an average of 1.2 per year pre-*Wiktionary*, to 3, post-*Wiktionary*. It is also interesting to note that, outside of *The Guardian*, only two appearances of ‘teenhood’ occurred before the word entered *Wiktionary*: one in *The Independent* just four months beforehand, and the other in *Mail Online* in February 2008. After its entry into

¹⁶ Some newspaper archives are more comprehensive than others, leading to gaps in the search results prior to 2005. Blank cells indicate that it was not possible to search that period. *The Express* search engine does not facilitate year-by-year searching.

Wiktionary, *The Independent* used ‘teenhood’ twice more, the *Mail Online*, only once.¹⁷

Examining sample concordance lines for these uses of ‘teenhood’ (see Table 6), we can see that there has been no change in the use or meaning of the word over this time. From 2000 through until 2010, ‘teenhood’ is used in the context of ‘adolescence’, with a sense of nostalgia for an earlier time in life. All of the concordances either refer to or imply powerful relationships, along with the sense of a journey, sometimes physical (concordance lines 1 and 2) sometimes emotional (3 and 4).

1.	about a car-crazy, rock ‘n’ roll midwestern	teenhood	in the late 50s and early 60s, Lucas made a movie
Source: <i>The Guardian</i> , August 2000			
2.	on a roadtrip to find their boyfriends from	teenhood?	It doesn’t mean love is dead: it merely means
Source: <i>The Guardian</i> , August 2005			
3.	early films captured the exquisite pains of	teenhood	growing in popularity to achieve cult status.
Source: <i>The Independent</i> , August 2009			
4.	and it’s not like you spend childhood and	teenhood	preparing for adulthood and then everything is
Source: <i>The Guardian</i> , December 2012			

Table 6. Concordance lines for ‘teenhood’ – media tracking April 2013.

The lack of any discussion over ‘teenhood’ in the Tea Room, and the mere five entries in its revision history (almost all of which occurred over a ten minute period) all indicate that the *Wiktionary* populace is happy with its definition of ‘teenhood’:

- ‘1. adolescence
- 2. state of being a teenager’ (2009).

The media’s corresponding use of the word, and its growing popularity, suggests that non-*Wiktionary* users are similarly satisfied with it.

5. Conclusion

The speed and ease of updating ‘wiki’ dictionaries opens up the opportunity for a more dynamic relationship between dictionary compilation and language change, with the dictionary potentially not only reflecting language use, but actually driving change. Whilst several authors are already working on the implications of ‘wiki’ dictionaries (see for example Meyer & Gurevych, 2012; Gurevych & Wolf, 2010; Penta, 2011), following on from earlier works on the wider field of electronic, and collaborative but non-interactive dictionaries (see Nesi, 2008 and de Schryver, 2003), this relationship has, as yet, gone unexplored.

¹⁷ Although one of *The Independent’s* post-*Wiktionary* uses was in a round-up obituary article featuring quotes from the earlier August 2009 piece. Both have since been removed from the site.

Of course not every word that enters *Wiktionary* will stand the test of time (Algeo, 1993). Whilst ‘gravatar’ is at too early a stage in its linguistic development to predict its future with any certainty, it appears that ‘teenhood’ is surviving and may, in fact, thrive. Media tracking in five newspapers identified that ‘teenhood’ was used only 11 times in the media prior to inclusion in *Wiktionary*, whilst it appeared nine times in the three years afterwards. It will be interesting to see over the coming months and years whether this increase is sustained and will lead to a successful reincarnation for ‘teenhood’, and whether this ultimately leads to recognition by traditional publishers of works other than *OED*, and incorporation into new editions of other mainstream dictionaries.

If it does, it may be that the project following this pilot study will reveal a similar pattern of entry into *Wiktionary*, followed by an increase in use and faster establishment of a place in the lexicon. This could suggest a new role for *Wiktionary* as an early predictor of successful neologisms. Determining this will require analysis of both successful and unsuccessful new additions to *Wiktionary* (defined by the longevity of the word).

Evidence of a new relationship between dictionary-making and language change would not only satisfy academic curiosity, but could prove useful to dictionary publishers seeking innovative ways to monetise their online offerings and set themselves apart from the competition. A clearer understanding of the behaviour of new words once they have entered the dictionary and begun to spread into wider spheres of use could enable these companies to better tailor their time and resources, whilst building a stronger relationship with grassroots language users. Collaboration with the producers of ‘wiki’ dictionaries could present traditional publishers with a unique selling point around which to promote their products.

6. Acknowledgements

My thanks to my PhD supervisor Professor Hilary Nesi (Coventry University) for her help and guidance on this paper and on the corpus pilot study. Thanks also to Sian Alsop (Coventry University) for her assistance in devising the automated media tracking process proposed for the next stage of this research.

7. References

- Algeo, J. (1993). Desuetude among New English Words. *International Journal of Lexicography* 6(4), pp. 281-293.
- Coxhead, A. (2000). An Academic Word List, *ELI Occasional Publications #18*, Wellington: School of Linguistics and Applied Language Studies, Victoria University of Wellington.
- Daily Express – Express* [online]. Accessed at: <http://www.express.co.uk/>.
- Daily Mail – Mail Online*. Accessed at: <http://www.dailymail.co.uk/home/index.html>.

- de Schryver, G-M. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2), 143-199.
- Fletcher, W.H. (2013). Corpus Analysis of the World Wide Web. In Chapelle, C.A. (ed.) *The Encyclopedia of Applied Linguistics* [online] Blackwell Publishing. Accessed at: <http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0254/full>.
- Gravatar [online]. Accessed at: <https://en.gravatar.com/>.
- Gurevych, I. & Wolf, E. (2010). Expert-Built and Collaboratively Constructed Lexical Semantic Resources. *Language and Linguistics Compass*, 4(11), pp. 1075-1090.
- Janssen, M. (2013). Lexical Gaps. In Chapelle, C.A. (ed.) *The Encyclopedia of Applied Linguistics*. [online] Blackwell Publishing. Accessed at: <http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0693/full>.
- Leuf, B. & Cunningham, W. (2005). *The Wiki Way: Quick Collaboration on the Web*, USA: Addison-Wesley.
- Macmillan Dictionary (British English / American English)* [online] Accessed at: <http://www.macmillandictionary.com/>.
- Merriam-Webster* [online]. Accessed at: <http://www.merriam-webster.com/>.
- Meyer, C. & Gurevych, I. (2012). Wiktionary: a New Rival for Expert-Built Lexicons? Exploring the Possibilities of Collaborative Lexicography. In Granger, S. & Paquot, M. (eds.) *Electronic Lexicography*. Oxford: Oxford University Press, pp. 259-292.
- Nesi, H. (2008). Dictionaries in Electronic Form. In Cowie, A.P. (ed.) *The Oxford History of English Lexicography*. Oxford University Press, pp. 458-478.
- Oxford Dictionaries* [online]. Accessed at: <http://oxforddictionaries.com/>.
- Oxford English Dictionary* [online]. Accessed at: <http://www.oed.com/>.
- Penta, D.J. (2011). *The Wiki-fication of the Dictionary: Defining Lexicography in the Digital Age*. Media in Transition 7 Conference, Massachusetts Institute of Technology, Cambridge, MA, USA, 13 May 2011.
- Rayson, P. (2008). From Key Words to Key Semantic Domains. *International Journal of Corpus Linguistics* 13(4), pp. 519-549.
- Rundell, M. (2002). *Macmillan English Dictionary for Advanced Learners of American English*. Oxford: Macmillan Education.
- Rundell, M. (2007). *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Education.
- Sinclair, J. (2004). Corpus and Text – Basic Principles. In Wynne, M. (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*. [online]. Oxford: Oxbow Books. Accessed at: www.ahds.ac.uk/linguistic-corpora/.
- Sketch Engine: SiBol/Port Corpus* [online]. Accessed at: https://the.sketchengine.co.uk/bonito/run.cgi/first_form?corpname=preloaded/sibolport_1;
- Sketch Engine: SiBol/Port Corpus – 'teenhood'* [online]. Accessed at: https://the.sketchengine.co.uk/bonito/run.cgi/first?corpname=preloaded%2Fsibolport_1&reload=&iquery=teenhood&queryselector=iqueryrow&lemma=&lpo

s=&phrase=&word=&wpos=&char=&cql=&default_attr=word&fc_lemword_wi
ndow_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all
&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&usesubcorp=
&sca_doc.author=&sca_doc.date=.

Sketch Engine: ukWaC (old WSG) – ‘teenhood’ [online]. Accessed at:
<https://the.sketchengine.co.uk/bonito/run.cgi/first?iquery=teenhood&queryselector=iqueryrow&corpname=preloaded%2Fukwac2>.

Soanes, C. & Stevenson, A. (2006). *Concise Oxford English Dictionary*. Oxford:
Oxford University Press.

The British Newspaper Archive [online]. Accessed at:
<http://www.britishnewspaperarchive.co.uk/>.

The Free Dictionary [online]. Accessed at: <http://www.thefreedictionary.com/>.

The Guardian [online]. Accessed at: <http://www.guardian.co.uk/>.

The Independent [online]. Accessed at: <http://www.independent.co.uk/>.

The Sun [online]. Accessed at: <http://www.thesun.co.uk/sol/homepage/>.

Urban Dictionary (2013). [online]. Accessed at: <http://www.urbandictionary.com/>.

Urban Dictionary – ‘Gravatar’ [online]. Accessed at:
<http://www.urbandictionary.com/define.php?term=gravatar>.

West, M. (1953). *A General Service List of English Words*. London: Longman, Green
& Co.

Wikipedia – ‘gravatar’. [online]. Accessed at: <http://en.wikipedia.org/wiki/Gravatar>.

Wiktionary (2013). [online]. Accessed at:
http://en.wiktionary.org/wiki/Wiktionary:Main_Page.

Wiktionary – ‘teenhood’ [online]. Accessed at:
<http://en.wiktionary.org/wiki/teenhood>.

Wiktionary – ‘teenhood’ revision history [online]. Accessed at:
<http://en.wiktionary.org/w/index.php?title=teenhood&action=history>.