

User-generated content (UGC) in online English dictionaries

Robert Lew, rlew@amu.edu.pl

1. Introduction

User-generated content rests on two pillars: the philosophy of the role of the collective of ordinary people, and the technical capability to realize this philosophy in practice.

The philosophy behind UGC is actually anything but new. It is based on the belief in the *wisdom of crowds* (Surowiecki 2005), which in essence puts the collective opinion of a group of people above that of a single expert. A practical and important social application of this principle is *trial by jury*, which functions in numerous societies and has many advocates.

The technological underpinnings of the concept of user-generated content have to be sought in the rise of *Web 2.0*. This playful name (which draws on the convention of referring incrementally to an improved version of a software application) stands for the advanced interactive stage in internet experience, and it is the modern affordances that come with it which make user-generated content happen. Users of Web 2.0 are no longer passive recipients of packaged content. Increasingly, they actively contribute to the creation and provision of self-made content. This double capacity of newly empowered users can be aptly captured in the neologism *prosumer*, which is a blend of *producer* and *consumer*. In this way, web users' social roles become blurred. In the context of online lexicography, the new model puts dictionary users in the shoes of lexicographers.

Another concept relevant to UGC is *new media*. In opposition to traditional media such as newspapers or motion pictures, new media involve a two-way exchange of information: it is *interactive* or *conversational*. Another of its characteristics is the immediacy with which new media can be accessed, being available on demand, essentially at any time, and on a range of digital devices (PC's, netbooks, tablets, mobile phones, ...). It is generated in real time and largely unregulated. Unlike in traditional media, the creation, publishing, distribution, and consumption of new media involve a level of democratization.

2. What is User-Generated Content?

It is instructive to consider what UGC actually signifies. A workable characterization of the concept (albeit under a slightly different name of *User-Created Content*) is provided in the relevant OECD report (Vickery/Wunsch-Vincent 2007); this publication cites three broad characteristics that have to obtain in conjunction (although the authors are open to a degree of fuzziness in the third criterion):

Publication requirement: A principle [sic!] characteristic is that the work is *published* in some context, for example on a publicly accessible website or on a page on a social networking site only accessible to a select group of people (e.g. fellow university students), even though UCC could be made by a user and never published online or elsewhere. This characteristic excludes e-mail, two-way instant messages and the like.

Creative effort: A certain amount of creative effort has to be put into creating the work or adapting existing works to construct a new one; i.e. users must add their own value to the work. UCC could include user uploads of original photographs, thoughts expressed in

a blog or a new music video. The creative effort behind UCC may also be collaborative, for example on websites that users edit collaboratively. Merely copying a portion of a television show and posting it on an online video website (a frequent activity on UCC sites) would not be considered UCC. Nevertheless the minimum amount of creative effort is hard to define and depends on the context.

Creation outside of professional routines and practices: User-created content is usually created *outside of professional routines and practices*. It often does not have an institutional or commercial market context and UCC may be produced by non-professionals without expectation of remuneration or profit. Motivating factors include: connecting with peers, achieving fame, notoriety or prestige, and expressing oneself. (Vickery/Wunsch-Vincent 2007: 18)

The *creative effort* stipulation becomes especially problematic in the context of lexicographic activity. This is so because dictionary-making has traditionally been an activity of an accretive, and often duplicative nature. For centuries, numerous dictionaries were written by copying and collating content from dictionaries already in existence, and a considerable degree of acceptance of this fact exists in the general public to this day, even among people directly involved in lexicography (Landau 2001: 403).

3. Motivation behind UGC

It is interesting to examine the motivation behind user-generated content. The gains of site hosts are rather obvious: being able to obtain potentially relevant content that is essentially cost-free, keeping the site alive with activity, and attracting users. But what exactly is it that drives web users to donate their time and (sometimes) expertise in order to create online content? It seems that at least three types of factors are at play as far as motivation for creating UGC is concerned: psychological, social, and economic.

3.1. Psychological motivation

Many individuals find contributing content *psychologically satisfying*. There is the do-gooder feeling of benefiting others which comes under the general rubric of *altruism*. Another psychological aspect important for some is that contributing content can fulfil the individual's need for self-expression. Finally, being able to provide advice on language, as often happens in lexicographic resources, makes people feel like teachers, and there are individuals who derive enjoyment out of mentoring others.

3.2. Social motivation

The second category of reasons which drive users to create UGC has to do with social aspects. One reason is the urge to be part of an online community, connecting and interacting with others: it is much the same motivation that is responsible for the current success of the phenomenon of social networking. Of course, in order for this aspect to be effective, the service needs to have a reasonable-sized active user base. In other words, this type of social incentive requires *scale* in order to be effective. This is also the prerequisite for another reason in the social category. Quite a few individuals engage in content creation in hopes of becoming famous (or notorious, as the case may be). Top contributors of UGC often succeed in securing certain prestige in the community. Another reason for engaging in user-supported lexicographic projects is that many individuals are honestly and publicly enthusiastic about

words, their meanings and uses, and matters of language usage. This *word enthusiast* factor provides an important social incentive for lexicographic UGC.

3.3. Economic motivation

Apart from psychological and social motivation, there may also be material rewards for supplying UGC. This may take the form of prizes, coupons for physical or virtual goods, access to premium features, etc.

4. UGC on the Web

Unlike in its early years, a significant section of today's World Wide Web relies on user-generated content. Web 2.0 is an interactive experience, with distinctions between consumers and producers of content becoming increasingly blurred.

With respect to specific websites, UGC can be exploited to varying degrees. This, of course, is a continuum, but for practical reasons it may be convenient to distinguish between at least three configurations with regard to the proportion of content originating from users:

1. content is entirely or almost entirely user-generated;
2. UGC is combined with professional (institutional) content; this option is sometimes termed half-collaborative;
3. professional/institutional content dominates, but is supplemented or enriched with user input; the latter may take the form of reviews, comments, etc.

As we shall see below, these three models are all represented in online dictionaries. Before we examine such dictionaries, let us take a quick look at a selection of pre-lexicographic examples, so as to get some view of the models underlying UGC in online lexicography.

Web 2.0 has brought in an opportunity for people to share their (amateur) artistic output, and rate and comment on the work of other users. [Quizilla](#) (Figure 1) is a relatively early (2002) website with user-generated content, which started as a platform for exchanging quizzes, but was soon extended to allow for sharing amateur prose and poetry, thus providing an example of how a community-based site is shaped according to genuine needs of its members. One of the key features that is now so much part and parcel of Web 2.0 is the user-rating of content.

An urge to produce and share creative content is one that is present in many humans and it is websites like [Quizilla](#) that allow such content to be shared and discussed with netizens of like interests. One of the most popular services of this type, but dealing with visual rather than verbal art, is the photograph-sharing service [Flickr](#) (Figure 2).



Imagine it Create it Share it

SEARCH GO

QUIZZES STORIES POLLS POEMS LYRICS GAMES MYSTUFF HELP ?

Poems Hub

TOP RATED MOST POPULAR NEWEST

Freedom (From My Ready Rope)

[READ A RANDOM POEM!](#)

Created by Angeltruth233 on 06/23/2012

[READ MORE!](#)

Title	Author	Published	Viewed	Rating
He Got Away.	thisheartbelongs	06/29/2012	763	☆☆☆☆☆
{ Miniscule Glints }	AutumnScorch	07/05/2012	120	☆☆☆☆☆
"Disaster Diasies"	SevenTongues	07/05/2012	162	☆☆☆☆☆
Nothing, Something.	RayaYarbrough	07/05/2012	234	☆☆☆☆☆
"I Miss You, Rain."	XtacoZxX	07/05/2012	131	☆☆☆☆☆
Goodbye.	RayaYarbrough	07/05/2012	341	☆☆☆☆☆
A Poem For The World	forever6red	07/05/2012	258	☆☆☆☆☆
Haunted	sexyrose111	07/06/2012	290	☆☆☆☆☆
[Collar Scrabble]	GardenUnique	07/08/2012	90	☆☆☆☆☆

LOG IN

username [Forgot Password?](#)

password or [Register](#)

Remember Me

GOT AN IDEA? GET STARTED!

Poem

title

NEW TO QUIZILLA?

Feel like taking a personality quiz or testing your

Figure 1: Quizilla, an early UGC website for the sharing of prose and poems (originally quizzes)



Home The Tour Sign Up Explore Upload

Explore / Most recent photos and videos



From Circles of ...



From kevjin



From p11118



From El Bingle

» Popular

» Creative



From ISAurora
Titans...



From Eidenmark



From Pattys-photos



From edlaw01



From kimilife



From Actua



From sleffert



From Justin Porterfield



From gwazulu



From SVNews



From tiredbees



From randylim27

Figure 2: The photo-sharing website Flickr

Of course, the use of UGC-based or UGC-enhanced websites does not stop at aesthetic appreciation. Just as many [Flickr](#) users value the art appreciation side of the service, for many others it is mainly about the practical utility of being able to conveniently share family photos from holidays, trips, or family events. One very practical application of the UGC-model is the air travel website [SeatGuru](#). The idea of this service is to assist potential travellers with flight seat selection. Its database accumulates data on plane seating quality for most of the world's airlines, listing specific models of aircraft in their fleets, including flight number information. Much of the actual detailed data come from the users themselves. Via feedback forms (Figure 3), travellers supply their comments on the quality of the seats they have occupied on their recent flights. Thanks to all this data, prospective passengers are in a position to make an informed choice of seating, particularly when checking in online. They can be warned of potential inconveniences, such as legroom restrictions due to media boxes, backrests failing to recline in front of bulkheads, or proximity to a lavatory. What this service has in common with lexicographic tools is that it serves specific consultation needs, in this particular case a type of needs which has sometimes been called *cognitive* (Tarp 2008; Ptaszyński 2009).

The screenshot shows the SeatGuru.com interface. On the left is a sidebar with a list of aircraft models under 'General Information', including Airbus A319-100 NEK, Airbus A320-200 NEK, Airbus A321 NEK, Airbus A330-300 (333), Airbus A340-300 Vers. 1 (34V), Airbus A340-300 Vers. 2 (34P), Airbus A340-300 Vers. 3 (343), Airbus A340-600 (346), Airbus A380-800 (388), Avro RJ85 Avroliner (AR8), Boeing 737-330, Boeing 737-530 NEK, Boeing 747-400 (744) Version 1, Boeing 747-400 (744) Version 2, Boeing 747-400 (744) Version 3, Boeing 747-8i Intercontinental, Bombardier Q-400, Canadair CRJ-700 (CR7), Canadair CRJ-900 (CR9), Embraer ERJ-195 (E95), Malaysia Airlines, MDLR, Mesa Airlines, Monarch Airlines, Oman Air, Openskies, Philippine Airlines, Porter Airlines, Airlines: O-R-S-T, and Airlines: U-V-W. The main content area features a 'Feedback for Lufthansa Airbus A319-100 NEK' form. It includes a 'My flight number was:' field, a 'My comments are for seat:' field, and a 'Your Comments:' text area. There are radio buttons for 'I agree', 'I disagree', and 'I have something to add'. A 'Submit' button is located below the form. To the right of the form is a 'Seatmap Key' section with various icons and labels: Good Seat (green square), Some Drawbacks (yellow square), Poor Seat (red square), Standard Seat (grey square), Blocked Seat (black square), Crew Seat (blue square), Power Port (black dot), Emergency Exit (red double lines), Overhead TV (black V), Galley (yellow circle), Lavatory (blue circle), and Closet (orange circle). Below the form is a 'Please Enter your Email Address:' field with a 'Submit' button. At the bottom, there is a 'Find Cheap Flights!' section with a search form for flights from Warsaw, Poland to Frederik. The search form includes fields for 'From:', 'To:', 'Depart:', 'Return:', and 'Travelers:'. The 'Depart:' field is set to 'Aug 4' and the 'Return:' field is set to 'Aug 11'. There is a 'Go' button and a 'NEW!' badge.

Figure 3: Seat feedback form on SeatGuru.com

One important area of online content where internet users provide meaningful contributions is citizen journalism, in which non-professionals get involved in reporting news on websites such as Korea's [OhmyNews](#) (Figure 4). But many mainstream news services also try to involve their regular readers in collaborative journalism. For instance, [CNN](#) has *iReporters*: registered individuals who contribute reports, including video recordings. Many other professional services now include similar kinds of user involvement, and it is in fact quite rare today to have news services which would have no provision for reader comments at least.



NoCut News: The lowest of Korea's tabloids

17
07 2012

by Curator-in-chief posted in [Citizen Journalism and Mainstream Media](#)
[No Comments](#)

[+1](#) 0
 [Tweet](#) 0
 [Share](#)
 [Like](#) 0

Refusing to be outdone in the xenophobe/racism department by TV station MBC, NoCut News has published a series of stories titled, "The Reality and Twisted Values of Some White Men." This is on the tail of a series of stories about a foreign graduate student who allegedly (totally relying on one anecdotal source in typical tabloid style) filmed himself having sex with Korean women.

English language blogs [The Marmot's Hole](#) and [Gusts of Popular Feeling](#) break down these articles and point out the bad journalism and inaccuracies.

[+1](#) 0
 [Tweet](#) 0
 [Share](#)
 [Like](#) 0

☆☆☆☆☆☆☆☆☆
 Rating: 0.0/10 (0 votes cast)
 Rating: 0 (from 0 votes)

become a curator

Recent Comments

Keith: Now there are over 8,000 members one week later!

peter byron: MBC has shown a lack of professionalism before by instigating the mad cow debacle with its lies and...

Luc: "asking users to only post content they created," Hahahaha You can't upload pictures on...

Curator-in-chief: When I lived in Germany the custom was that if a pedestrian was anywhere near a crosswalk that a...

CPY: Well, I think both is at fault. One thing for sure, I believe Singaporean must change their mindsets on the...

Figure 4: [OhmyNews](#), an example of citizen journalism

Film lovers have for some years enjoyed the extensive collaborative database [IMDb](#) (short for Internet Movie Database), where, subject to subsequent IMDb staff approval, registered users can contribute most of the pivotal data (film titles, appearances by actors, etc.), as well as provide supplementary content such as summaries and reviews, user ratings, or detailed content advisory for concerned parents.

5. UGC in lexicography

5.1. [Wiktionary](#) as an example of a collaborative dictionary project

As we have seen above, user-generated content has spread to many domains of the Web, thus it should not be surprising that it has entered online lexicography as well. What may be thought of as an immediate precursor to crowdsourced dictionaries is a reference work with an encyclopaedic rather than lexical focus: the collaborative encyclopaedia [Wikipedia](#) (Figure 5). This extensive resource is a large-scale implementation of the idea of the *wisdom of crowds* (Surowiecki 2005). It relies on the assumption that in the online world, there will be experts on a specific topic willing to share their time and expertise pro bono with the world at large. There is little doubt that this model has turned out to be highly successful, at least in some areas. A peer-review based comparative assessment of science articles in [Wikipedia](#) and the highly respectable *Encyclopaedia Britannica* (Giles 2005) found an average of four errors in [Wikipedia](#) as compared to three in the latter, suggesting that the factual accuracy of the two resources was at a comparable level. Presumably, [Wikipedia](#) got better rather than worse since that time.



Figure 5: An entry for *lexicography* in [Wikipedia](#), the collaborative encyclopaedia

Initiated as a sister project of [Wikipedia](#), [Wiktionary](#) came to life in 2002, initially as an English-language edition only, to be joined less than two years later by separate language editions of the then 143 active languages (Meyer/Gurevych 2012: 262). At the time of this writing (12 January 2013), the English edition of [Wiktionary](#) reports nearly three million entries, and three further editions stand at more than a million articles each. However, a non-trivial proportion of items are automatically generated inflectional forms; another reason behind the high numbers is the wholesale incorporation of entries from older out-of-copyright dictionaries. Even more importantly, only a fraction of the entries in the English edition are actual English words: a given language edition A uses language A as the primary language of explanation for its entries, but the headwords themselves are not restricted to language A. And so, for example, the English [Wiktionary](#) covers the English lexicon (which it explains with definitions in English), but also words from numerous other languages (which it explains with English translations and/or English definitions). Meyer and Gurevych (2012) stress that this aspect is poorly understood, even by some experts. As shown in Figure 6, the English Wiktionary entry for HANDLE covers the English word (at the top), but also the Danish, German, and Norwegian words so spelled in those languages.

[Wiktionary](#) is compiled through a collaborative process by a large community of Web users, in this context called *Wiktionarians*. Wiktionarians pattern into a pyramid-like structure with the top represented by (democratically elected) 98 *administrators* (Meyer/Gurevych 2012: 271), who hold the most extensive permission rights. *Registered users* are users holding personal accounts, and it is they who do the greatest proportion of the work. However, as in many collaborative enterprises, the distribution of the workload tends to be Zipfian, with only a very small minority of users taking care of most of the work. For example, the English [Wiktionary](#) has about 400,000 registered users, but only some 4,000 (a mere one per cent)

have at least ten edits to their names (Meyer/Gurevych 2012: 272). Unregistered users can also edit entries, but this happens rarely, with only about five per cent of all article edits being by unregistered users (Meyer/Gurevych 2012: 272).

The image shows the top portion of the Wiktionary entry for the word "handle". At the top left, it identifies itself as a multilingual tree encyclopedia. The Wiktionary logo and tagline are present. Navigation tabs for "Entry", "Discussion", and "Citations" are visible, along with "Read", "Edit", and "History" buttons and a search box. The main content area features a "Contents" table of contents with sections for English, Danish, German, and Norwegian, each with sub-sections for pronunciation and etymology. A sidebar on the left contains various utility links like "Main Page", "Community portal", and "Toolbox". At the bottom, there is a "Pronunciation" section with IPA and audio links, and a "Wikipedia has articles on:" box with a globe icon.

Figure 6: Top of the English [Wiktionary](#) entry for HANDLE

The collaborative aspect is supported in [Wiktionary](#) with the usual wiki-type functionality. Each article has a *talk page* for the exchange of comments. Differences of opinion are resolved through discussion, and, where consensus has not been reached, through voting. There are also pages in [Wiktionary](#) discussing more general matters of policy, such as criteria for inclusion of terms and senses in the dictionary.

As far as I am aware, a systematic evaluation of the lexicographic quality of [Wiktionary](#) articles has not yet been attempted. Those assessments that are available tend to address aspects analyzable by computer, such as coverage of terms and senses, or sense ordering (Krizhanovsky 2011; Meyer/Gurevych 2012). Unfortunately, such automated evaluation, though fast and convenient, cannot at present even begin to address the essence of the quality of lexicographic description. Isolated comments about the latter have been made, however, and overall have been critical of [Wiktionary](#)'s lexicographic quality (Fuertes-Olivera 2009; Hanks 2012). Hanks (2012: 77-78) observes that “[i]n the English Wiktionary, the etymologies are taken from or based on those in older dictionaries, as are the definitions, which are extremely old-fashioned and derivative, taking no account of recent research in either cognitive linguistics or corpus linguistics”. This observation is certainly largely true, and a close comparison of a Wiktionary entry with one from a quality professional dictionary for the same headword can be a sobering experience to enthusiasts of crowd-sourced lexicographic description.

handle (*third-person singular simple present handles, present participle handling, simple past and past participle handled*)

- (*intransitive*) To use the hands.
 - They have hands, but they **handle** not - *Psalm 115:7*
- To touch; to feel with the hand.
 - Handle** me, and see; for a spirit hath not flesh - *Luke 24:39*
- To use or hold with the hand.
 - About his altar, **handling** holy things - *John Milton*
- To manage in using, as a spade or a musket; to wield; often, to manage skillfully.
 - That fellow **handles** his bow like a crowkeeper - *Shakespeare, King Lear, IV-vi*
- To accustom to the hand; to work upon, or take care of, with the hands.
 - The hardness of the winters forces the breeders to house and **handle** their colts six months every year - *Sir W. Temple*
- To receive and transfer; to have pass through one's hands; hence, to buy and sell; as, a merchant **handles** a variety of goods, or a large stock
- To deal with; to make a business of.
 - They that **handle** the law knew me not - *Jeremiah, 2:8*
 - 2011** December 16, Denis Campbell, "Hospital staff 'lack skills to cope with dementia patients'", *Guardian*:
The findings emerged from questionnaires filled in by 2,211 staff in 145 wards of 55 hospitals in England and Wales and 105 observations of care of dementia patients. Two-thirds of staff said they had not had enough training to provide proper care, 50% said they had not been trained how to communicate properly with such patients and 54% had not been told how to **handle** challenging or aggressive behaviour.
- To treat; to use, well or ill.
 - How wert thou **handled** being prisoner - *Shakespeare, Henry VI, Part I, I-iv*
- To manage; to control; to practice skill upon.
 - You shall see how I'll **handle** her - *Shakespeare, Measure for Measure, V-i*
- To use or manage in writing or speaking; to treat, as a theme, an argument, or an objection.
 - We will **handle** what persons are apt to envy others - *Francis Bacon*
- (*socce*) To touch the ball with the **hand** or **arm**; to commit **handball**.
 - 2011** February 12, Les Roopanarine, "Birmingham 1 - 0 Stoke", *BBC*:
Robert Huth **handled** a Bentley shot, only for the offence to go unnoticed.

han-dle

- DO WORK** [*transitive*] to do the things that are necessary to complete a job:
 - I handled most of the paperwork.*
 - The case is being handled by a top lawyer.*
 - The finance department handles all the accounts.*
 - Computers can handle huge amounts of data.*
- DEAL WITH A SITUATION** [*transitive*] to deal with a situation or problem by behaving in a particular way and making particular decisions:
 - The headmaster handled the situation very well.*
 - I knew I had handled the matter badly.*
 - Leave it to me. I can handle it.*
 - Most customers were satisfied with the way their complaints were handled.*
 - Opposition leaders will be watching carefully to see how the Prime Minister handles the crisis.*
- DEAL WITH A PERSON** [*transitive*] to deal with a person or behave towards them in a particular way, especially in order to keep them happy.
 - Some customers are quite difficult to handle.*
- NOT BECOME UPSET** [*transitive*] to not become upset in a difficult situation:
 - She can't handle it when people criticize her.*
 - He doesn't handle stress very well.*
- HOLD** [*transitive*] to touch something or pick it up and hold it in your hands:
 - He had never handled a weapon before.*
 - We teach the children to handle the animals gently.*
 - He was roughly handled by the mob.*
- CONTROL A VEHICLE**
 - [*transitive*] to control the movement of a vehicle or an animal:
 - I didn't know if I'd be able to handle such a large vehicle.*
 - [*intransitive*] the way a vehicle handles is how easy it is to control **handles well/badly**
 - The car handles well, even on wet roads.*
- MOVE GOODS** [*transitive*] to move goods from one place to another:
 - The Post Office handles nearly 2 billion letters and parcels over the Christmas period.*
- BUY/SELL GOODS** [*transitive*] to buy or sell goods:
 - Bennet was charged with **handling stolen goods**.*

Figure 7: Entries for the verb HANDLE in the English [Wiktionary](#) (left) and the [Longman Dictionary of Contemporary English](#), free online version (right)

In Figure 7, two entries for the English verb HANDLE are presented side by side: on the left, an entry from the *English Wiktionary*, and on the right, one from the free online version of the *Longman Dictionary of Contemporary English*. The verb HANDLE was chosen at random, only considering relatively common verbs. It is immediately striking that the very first sense covered in *Wiktionary* represents a use that is highly unfamiliar in contemporary English, if ever there indeed was a genuine intransitive pattern, rather than a peripheral absolute use with an elliptical object. Even the *Oxford English Dictionary* does not record such a pattern. In general, senses are not very clearly distinguished, and their ordering appears haphazard. Definitions themselves tend to be made up of lists of rather general words, often used in non-prototypical senses, such as *use* in sense 8, where there is no indication that the object is meant to be human. Examples are mostly citations from the Bible or old literary classics, and are all markedly archaic. Much of this is traceable to the 1913 edition of *Webster's Revised Unabridged Dictionary*, available free of charge on many dictionary aggregator sites (Lew 2011). Although two modern citations have been added from journalistic texts, the one under

sense 7 is much too long: only its very final part is at all relevant here. No attention is given to word combinations.

Compared to this, the entry for the same word from the [Longman Dictionary of Contemporary English](#) is like a breath of fresh air, with its clearly signposted (see e.g. senses 2 and 3) and perspicuously defined senses. Examples of use are contemporary, authentic, and natural-sounding. They illustrate the combinatory potential of HANDLE, and offer usable models for production. All in all, the comparison of the two entries aptly illustrates the significant progress that lexicography has made over the last decades or so. Ironically, it is the collaborative entry, one produced under an innovative Web 2.0 model, which exemplifies outdated ways of lexicographic description. It seems that the web community, while enthusiastically embracing the novelty of online collaboration, propagates the traditional model of lexicographic description perhaps received from their repeated encounters with respectable volumes sitting proudly on their parents' bookshelves.

While a UGC-driven model, perhaps best known from [Wikipedia](#), has been a resounding success in terms of encyclopaedic content (factual data), it does not seem to work so well in lexicographic endeavours, such as [Wiktionary](#). [Wikipedia](#) revolves on the principle that somewhere out there in the world there are experts on every little bit of knowledge, willing to give of their time to freely share their knowledge with other people. While this model works surprisingly well for the reporting of encyclopaedic facts, it is much less robust when it comes to the job of describing words and expressions of a language: their meaning, pronunciation, morphology, syntax, word combination (collocation and colligation), and usage. To put it simply, while it makes good sense that somewhere out there there is an expert on a piece of specialized knowledge (say, a rare species of nettle) who is willing to share their expertise with the world, there is normally no such expert on the meanings of a particular everyday word (say, the word *field*), who would be capable of teasing apart the senses and providing a nuanced treatment of the many combinations and uses of the word. Rather, we could say that there are too many self-proclaimed language experts with a willingness to share, but their best efforts cannot match the output of professional, trained lexicographers. This is so because the quality of lexicographic description heavily depends on a good grasp of lexicographic principles, procedures, and tools of the trade (such as skilled use of corpus data or structural markup). Such appears to be the situation for general language items; things might be looking somewhat better, though, at the interface of encyclopaedic and linguistic description: specialized vocabulary and terminology, where “[e]ach contributor has a certain field of expertise ... [and this] fosters the encoding of a vast amount of domain-specific knowledge” (Meyer/Gurevych 2012: 259).

5.2. Define your world: [Urban Dictionary](#)

Another area of lexis which might benefit from the bottom-up approach are the elusive domains of slang, cant and jargon. This is the focus of another collaborative dictionary project, [Urban Dictionary](#) (Figure 8), which invites site visitors to “define your world”. [Urban Dictionary](#) actually predates [Wiktionary](#): its origins go back to 1999. Mainstream lexicography has traditionally been cautious about including fresh slang terms and fashionable words before they have proven themselves as likely to stay in active use for some time. The idea of [Urban Dictionary](#) is to capture such fleeting words and expressions, and it is an idea that has clearly caught on, given that the resource now holds a staggering 6,90 million definitions (as of 14 January 2013). This does not mean that anywhere near as many terms are defined in [Urban Dictionary](#), as most terms have multiple (sometimes hundreds) alternative definitions from different users. By adding up the numbers of entries as presented for each

respective letter page, I estimated the number of different terms at 1,38 million (as of 14 January 2013).

Figure 8: An [Urban Dictionary](#) entry for BOOTYISM (15 September 2012)

As seen in Figure 8, an [Urban Dictionary](#) entry contains user-supplied definitions, examples of use, and tags. Tags may be useful for searching the dictionary and cross-linking related entries (the dictionary also has an alternative *thesaurus* access interface). Other users can rate the different definitions for quality, and the order in which alternative definitions are displayed depends on those collective ratings. As a result, the ‘best’ definitions are not necessarily, and not usually, those that make a genuine attempt at explaining the meaning, but rather those that are seen as clever or otherwise amusing. Thus, for some of the entries at least, [Urban Dictionary](#) has become a playing field for put-down, scorn, displays of in-group identification, and all sorts of ideology. For example, the top-rated definition of BOOTYISM in [Urban Dictionary](#) (Figure 8) is not so much a comment on the (meaning of the) headword, but rather on Tiger Woods’s character. And the second most popular ‘definition’ of RELIGION runs as follows:

The biggest lie in human history. It has been responsible for more deaths throughout human history than all other unnatural causes combined. For a thousand years the Church was a tyrannical dictatorship that used religion to control the uneducated masses. Free your minds and come into the 21st century.

<http://www.urbandictionary.com/define.php?term=Religion&defid=6179390>

Another user's definition of RELIGION is this: "Apparently what most people on Urban Dictionary hate." — clearly a meta-comment on the ideologization of lexicography in [Urban Dictionary](#). Sometimes, however, a *bona fide* definition will make it into first page, an example being number 3 in Figure 8, or the following one for RELIGION: "An organized group of people with a common belief. Most religions strongly stress ethics and morals along with setting guidelines for people to follow in their day-to-day life."

It is doubtful if users of [Urban Dictionary](#) truly come to the site in search of the basic meaning of such common words as *religion*, so here the strategy of providing something else may actually not be incompatible with the needs of the users of the service. In contrast, for actual slang terms such as *bootyism*, meaning may be what users genuinely seek (especially non-native speakers of English who have failed to find an item of slang in a mainstream dictionary). Well aware that their definitions do not help with this, witty users will sometimes compensate for the lexicographic vacuity of their definitions in the other field of substance, namely *examples of usage*. Interestingly, instances of genuine, or even remotely plausible, examples of usage are even rarer than *bona fide* definitions. It is hard to say whether this is more a matter of misunderstanding the design principle behind this particular data field, flouting it, or else living proof of the claim (by corpus linguists and lexicographers such as John Sinclair or Patrick Hanks) that usage cannot be invented.

5.3. [Wordnik](#)

[Wordnik](#) (Figure 9) is a relatively new (founded in 2009) collaborative resource which offers a rather successful combination of professional lexicographic content from existing dictionaries, its own corpus-based content, content pulled from [Twitter](#) and [Flickr](#), and user-generated content.

The screenshot shows the top of a Wordnik entry for 'rosemary'. At the top, there is a search bar and navigation links: 'Define', 'Relate', 'List', 'Discuss', 'See', 'Hear', and 'Love'. Below this, a prompt says 'Log in or sign up to show 'rosemary' some love.' The 'Definitions' section includes entries from 'American Heritage® Dictionary of the English Language, Fourth Edition' and 'Century Dictionary and Cyclopedia'. The 'Examples' section features quotes from Twitter (dated 07-13-2009 and 07-07-2009) and 'Today's Garden'.

Figure 9: Top of a [Wordnik](#) entry for ROSEMARY

According to the co-founder and CEO of [Wordnik](#) (McKean 2011), its users (calling themselves *Wordniks*) like to think of the service more as a game than a dictionary. Very few users want to write definitions; most prefer to supply metalinguistic information instead. [Wordnik](#)'s users also like to invent new words; so much so, that they have even come up with an in-house adjective to describe such words: *madeupical*. The structure of [Wordnik](#) is user-activity-based, as indicated by the coloured-font options in the top right corner in Figure 9, with the sections stacked on top of one another in one lengthy page. In the default view of a word, definitions from existing works of reference, including out-of-copyright dictionaries, [Wiktionary](#) and [WordNet](#), are accompanied by citations, including some from [Twitter](#). The technology behind [Wordnik](#) prioritizes citations with self-defining contexts, especially when no prior dictionary definitions are available for a term.

Below the definition-and-examples section, a related-words section is presented, which includes hypernyms (more general words), words found in similar contexts, words that contain the headword in their definitions, and semantic tags for the headword. Further down, user-created lists (Figure 10) are presented. Interestingly, word lists are amongst the favourite features of the service.

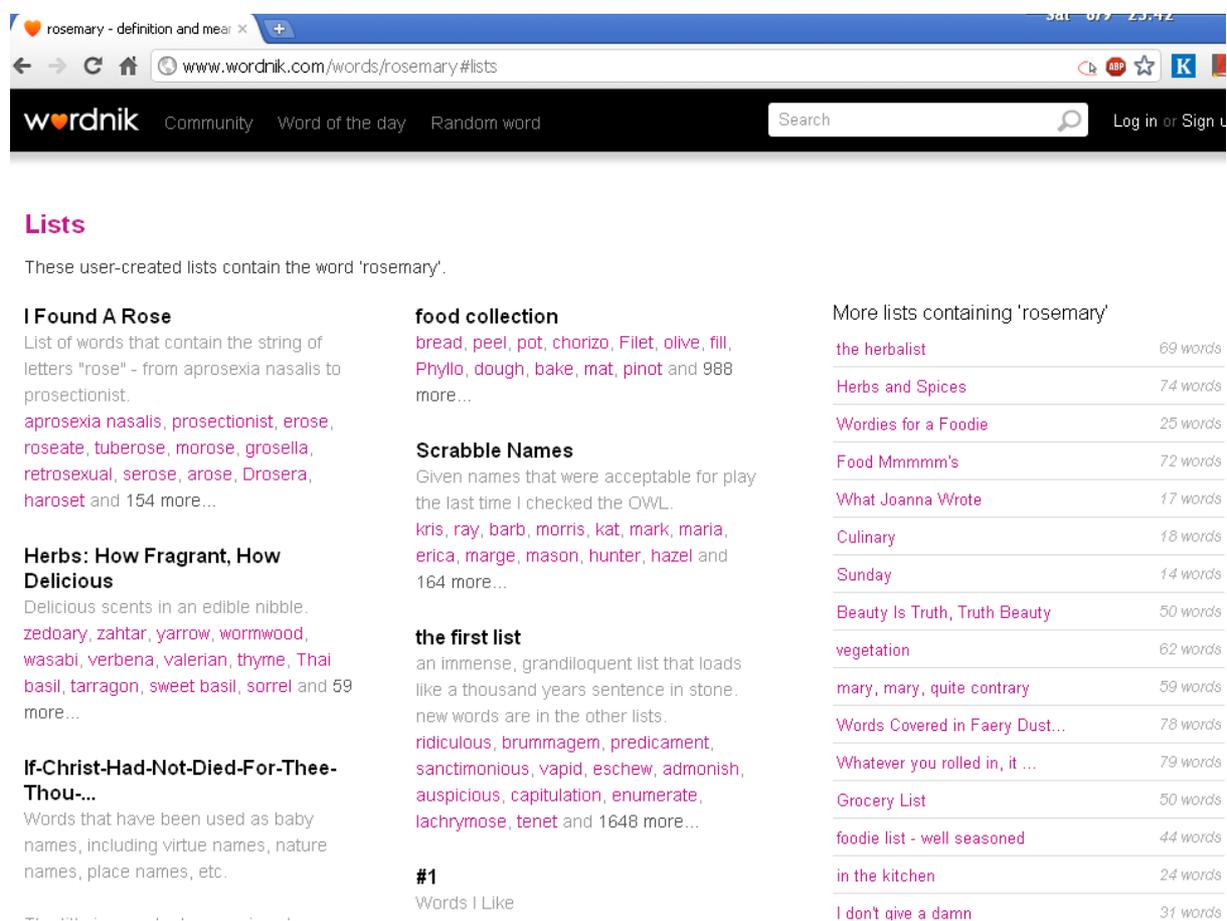


Figure 10: Word lists in [Wordnik](#) containing the word *rosemary*

Not all of the user-created lists would impress a professional lexicographer, such as a list of words containing the string *rose* ("I Found a Rose"), many of which bear no linguistic relationship to *rose* except the mere sequencing of letters (e.g. *aprosexia*). Word lists are not [Wordnik](#)'s original invention: they were inherited from *Wordie* (Figure 11), a former service with which [Wordnik](#) merged, incorporating its user-generated content.

Wordie [wɜːrd · ē]

Like Flickr, but without the photos.

[Login](#) or [Register](#) |

Wordie lets you make lists of words -- practical lists, words you love, words you hate, whatever. You can then see who else has listed the same words, and talk about it. It's more fun than it sounds.

When adding comments and citations wrapping words in [brackets] now links them to their Wordie page. Mad, mad props to [angharad](#) for contributing a [Bookmarklet](#), which lets you select words on any web page and import them directly into Wordie. Works with Safari and Firefox (Mac or PC), does *not* work with IE 6. To install, bookmark the 'Add to Wordie' link below (on Safari and Firefox you can just drag it onto your bookmarks bar).

bookmark this -> [Add to Wordie](#)

Wordie hacking is heartily encouraged; please let me know if I can do anything to facilitate a project. Wordie suggestions and bug reports always appreciated (via [email](#) or my [profile](#)). If you like Wordie, please [dig it](#). [Yahoo!](#) has a short video segment on Wordie. Ridiculous. We're vid #2.

Recent Words [\(rss\)](#)

- [penumbra](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [lilt](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [schadenfreude](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [canter](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [snowth](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [drawl](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [melody](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [splay](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [sprawl](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [damask](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [grin](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [nimbus](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)
- [rain](#) [ninja](#) [words](#) | [tfd](#) | [dictionary.com](#) | [M-W](#) | [UD](#) | [Wikipedia](#) | [Google](#)

Total Words: **29,689**
Unique Words: **18,812**
Word Lists: **1,142**
Wordies: **737**

Wordiest Wordies

- [colleen](#) (1714 words)
- [Kaichi](#) (1330 words)
- [kat](#) (1281 words)
- [robotjohnny](#) (1118 words)
- [thricedotted](#) (1094 words)
- [dienekes](#) (972 words)
- [lanas](#) (930 words)

Figure 11: The service Wordie (no longer operative) as seen on 5 December 2006

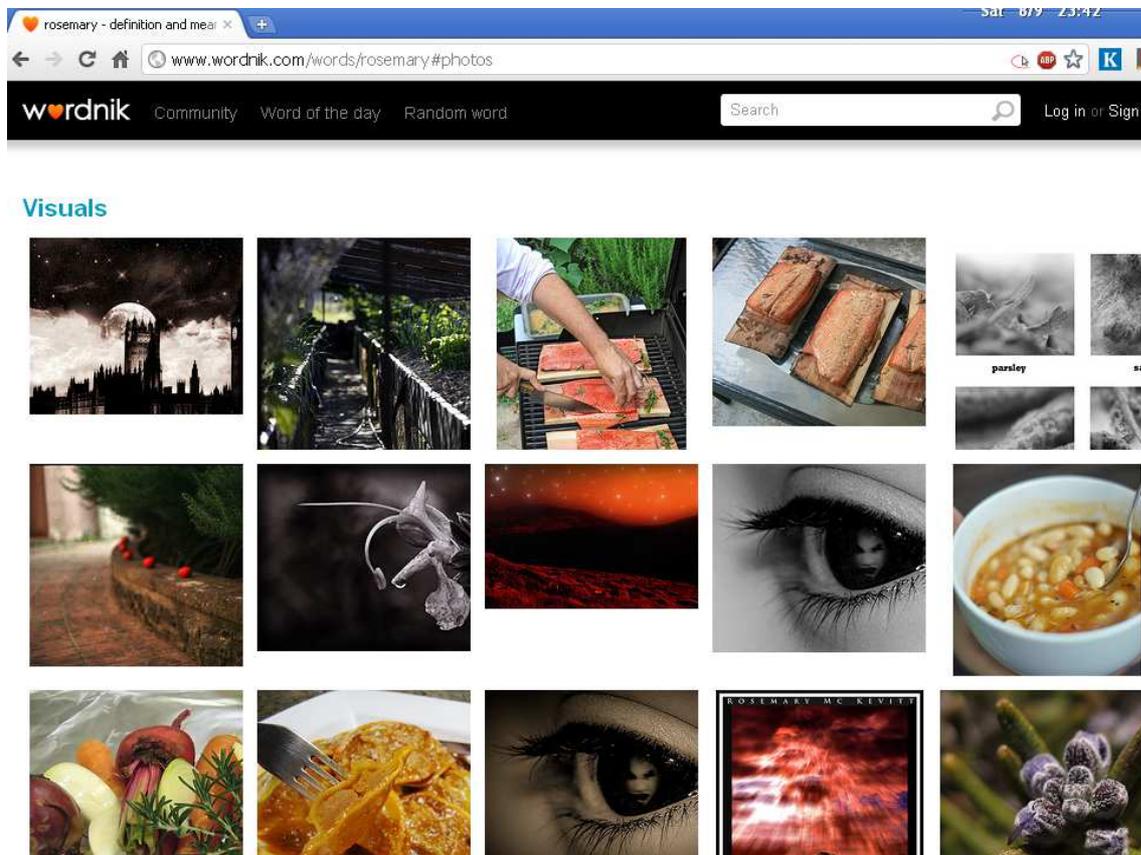


Figure 12: [Flickr](#) images for ROSEMARY in [Wordnik](#)

Wordie's tongue-in-cheek slogan was "Like Flickr, but without the photos" (Figure 11). [Wordnik](#), in contrast, does include photos, pulling them from none other than [Flickr](#). Image selection is based on tags, but this does not guarantee that the pictures so chosen are lexicographically relevant. In the case of ROSEMARY, for instance (Figure 12), how most of the images are related to the headword cannot be appreciated by someone not already familiar with the different uses of the word and its various designates.

6. User-generated content in professional dictionaries

Until recently, major English-language publishers had not engaged users in bottom-up lexicography. Much public attention was drawn to the issue with an announcement made by Collins on 17 July 2012, covered in *The Independent*, *The Guardian*, and *Daily Mail*, and an ensuing discussion in the social media, including the Lex group on [LinkedIn](#). The new feature is called "What's your word?". Alex Brown, Head of Digital at Collins was quoted as saying "Most dictionaries are static. By allowing the public to truly participate, we're ensuring that we stay on top of the evolving English language." (O'Neill 2012). However, in the same article we learn that "Collins will ultimately decide which words are included based in part on its 4.5 billion-word database of language called the Collins Corpus ... based on a number of criteria including frequency of use, number of sources and staying power". Having read this, someone who knows about dictionary-making cannot but wonder what new words users can possibly offer, and how come such words are not *already* in the dictionary. It could be that the primary motivation behind engaging users is different from that stated: the publisher may be counting on the sense of community borne out of users being given a chance to, on the one hand, contribute their own entries, and, on the other hand, enjoy entries edited by their peers. In this way, casual dictionary users may become more meaningfully (pun intended) involved.

In order to suggest an item for inclusion in the [Collins English Dictionary](#), registered users need to fill in a submission form. The form (Figure 13) is very simple and assumes no lexicographic sophistication on the part of the user, with only three text boxes provided for, respectively: (1) the new term, (2) its definition, and (3) additional information.

You are here > [Submit a word to the Collins English dictionary](#)

Submit a word to the Collins English Dictionary

All submissions will be reviewed for inclusion in the next updates of our online and print dictionaries

Your suggested word: *

Definition: *

Additional information (e.g. word origin, example sentences):

Post updates to your facebook wall:

Figure 13: [Collins English Dictionary](#) new word submission form

A/C New Word Suggestion

Submitted By: [sarathbabu_s](#)

Approval Status: Pending Investigation

▶ Definition of A/C

Stands for Air conditioner. It also stands for Account. Also means Assistant Commissioner.

▶ Additional Information

Many origins.

Figure 14: Example entry suggestion to the [Collins English Dictionary](#)

Submitted entries are subject to acceptance by Collins editors. A review of the recent submissions shows many to be mediocre at best, and users' lack of sophistication shows through in accepted entries. A typical submission is shown in Figure 14. Quite often, users will suggest words which are already included in the dictionary: a recent example of suggested addition to the [Collins English Dictionary](#) is *YOUSE*, whereas *YOUS* OR *YOUSE* is already a headword in the dictionary.

Despite the publicity it had received, Collins was not the first authoritative English-language dictionary publisher to introduce bottom-up lexicography as part of its online offerings. A notable case is [Macmillan English Dictionary](#), a dictionary primarily but not exclusively for learners of English. In February 2009, its UGC-model [Open Dictionary](#) was launched, at the same time as the main [Macmillan English Dictionary Online](#). Since then, the dictionary site has invited users to contribute entries to this user-created supplement. In order to submit a new item, a user fills in a special submission form (shown in Figure 15), very similar to the Collins form (though, of course, the Macmillan form was there earlier). Submitted entries are reviewed by Macmillan staff, and a selection are allowed to enter the word list of the [Open Dictionary](#).

When researching entries submitted to Macmillan's [Open Dictionary](#), in September 2012 I located an entry as in Figure 16 containing an ungrammatical definition ("used to parts..."). At the time of this writing (15 January 2013), the defective definition has been corrected to "used to refer to parts..."). This is a positive example of quality control exercised by professional Macmillan editors over user input, which at present amounts to about 1,400 accepted entries. But if this number continues to grow, the challenge of maintaining and updating all the additional entries as well as the main dictionary may become too much of a job. Is it worth the trouble then? Assessing three years of experience with Macmillan's crowd-sourced [Open Dictionary](#), Michael Rundell (2012: 80) writes that "the most fruitful areas where users can make a contribution are neologisms, regional varieties, and technical terminology" (this is sometimes referred to in the industry as the *long tail* of lexis). This observation dovetails with my assessment of the quality of general-language entries in user-constructed dictionaries.

Open Dictionary

New Word

What's your word?

What does it mean?

Could you give us an example sentence? (optional)

Your Details

Name/Nickname

Where in the world are you?

Figure 15: [Macmillan Open Dictionary](#) new word submission form

Open Dictionary

Most recent entries

[Browse all entries](#)

OE

NOUN

original equipment; used to parts made for a car by the car manufacturer itself, or made exclusively for the car manufacturer by a third party

There's good news for anyone called on to repair a Saab. You will still be able to source OE replacement parts, even though Saab is currently going through bankruptcy proceedings and has brought a halt to manufacturing.

Submitted from United Kingdom on 12/07/2012 15:10:00

Figure 16: Entry for OE submitted to [Macmillan Open Dictionary](#) (15 September 2012)

7. Conclusion

Serious lexicography requires specialized skills and expertise; for this reason, a large part of crowdsourced content is of questionable lexicographic value, as we have seen in the entry for HANDLE in [Wiktionary](#) and BOOTYISM in [Urban Dictionary](#). Pulling content from other social sites, though an easy way to enrich content, often produces irrelevant material, and sometimes considerable embarrassment, as in the former *Google English Dictionary* entry for KILT (Figure 17), which for several months in 2009 inadvertently appended a picture of a male wearing a kilt with no underwear, because it happened to be the most popular Google image for the word *kilt*.



Dictionary

Found in dictionary: **English > English**.

★ [Starred words »](#)

★ **kilt** /k'ilt/

Synonyms:

verb: [pleat](#)

kilts plural

1. A **kilt** is a skirt with a lot of vertical folds, traditionally worn by Scottish men. Kilts can also be worn by women and girls. N-COUNT

Images



web.ripnet.com

Figure 17: The *Google English Dictionary* entry for KILT (screenshot taken on 18 December 2009)

However, I have mostly looked at monolingual' dictionaries, and constructing effective definitions is known to be notoriously difficult. For non-native speakers of the language, definitions, however skilfully written, are not usually the best way to convey meaning (Lew 2004; Adamska-Sałaciak 2010). There may be more room for collaboration between professional lexicographers and active users on bilingual (and multilingual) tools, and some bilingual dictionary publishers, for instance in Japan, Malaysia or Germany, are trying to go down that road. One possibility which, to my knowledge, has not been tried yet would be to get users to suggest interlingual equivalents based on professionally supplied raw data, such as automatically-generated concordances illustrating source-language use, or perhaps parallel concordances.

Areas of vocabulary where amateur lexicographers can make the most significant contribution are neologisms, slang, regional varieties, and technical terminology. In contrast, general vocabulary is probably best left to professional lexicographers.

8. References

- Adamska-Sałaciak, Arleta (2010): Why we need bilingual learners' dictionaries. In: Kernerman, Ilan J./Bogaards, Paul (eds.), *English learners' dictionaries at the DSNA 2009*. Tel Aviv, p. 121-137.
- Fuertes-Olivera, Pedro A. (2009): The function theory of lexicography and electronic dictionaries: WIKTIONARY as a prototype of collective free multiple-language internet dictionary. In: Bergenholtz, Henning/Nielsen, Sandro/Tarp, Sven (eds.), *Lexicography at a crossroads: Dictionaries and encyclopedias today, lexicographical tools tomorrow*. Bern, p. 99-134.
- Giles, Jim (2005): Internet encyclopaedias go head to head. In: *Nature* 438, p. 900-901.
- Hanks, Patrick (2012): Corpus evidence and electronic lexicography. In: Granger, Sylviane/Paquot, Magali (eds.), *Electronic lexicography*. Oxford, p. 57-82.
- Krizhanovsky, Andrew (2011): A quantitative analysis of the English lexicon in Wiktionaries and WordNet. In: *Trudy SPIIRAN* 19, p. 87-101.
- Landau, Sidney I. (2001): *Dictionaries: The art and craft of lexicography*, 2nd edition. Cambridge.
- Lew, Robert (2004): Which dictionary for whom? Receptive use of bilingual, monolingual and semi-bilingual dictionaries by Polish learners of English. Poznań.
- Lew, Robert (2011): Online dictionaries of English. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York, p. 230-250.
- McKean, Erin (2011): Wordnik: Notes from an online dictionary project. *Electronic lexicography in the 21st century: New applications for new users (eLex2011)*. Bled, Slovenia.

- Meyer, Christian M./Gurevych, Iryna (2012): Wiktionary: a new rival for expert lexicons? Exploring the possibilities of collaborative lexicography. In: Granger, Sylviane/Paquot, Magali (eds.), *Electronic lexicography*. Oxford, p. 259-291.
- O'Neill, Cordelia. (2012, July 17). Collins dictionary invites word suggestions. *The Independent*. Retrieved from <http://www.independent.co.uk/arts-entertainment/books/news/collins-dictionary-invites-word-suggestions-7952775.html>.
- Ptaszyński, Marcin Overgaard (2009): Function-related secondary user needs and secondary data? A critical examination of some central concepts in the modern theory of lexicographical functions. In: *Lexikos* 19, p. 242-261.
- Rundell, Michael (2012): 'It works in practice but will it work in theory?' The uneasy relationship between lexicography and matters theoretical. In: Fjeld, Ruth Vatvedt/Torjusen, Julie Matilde (eds.), *Proceedings of the 15th EURALEX International Congress*. Oslo, p. 47-92.
- Surowiecki, James (2005): *The wisdom of crowds*. New York.
- Tarp, Sven (2008): *Lexicography in the borderland between knowledge and non-knowledge: General lexicographical theory with particular focus on learner's lexicography*. Tübingen (Lexicographica Series Maior 134).
- Vickery, Graham/Wunsch-Vincent, Sacha (2007): *Participative web and user-created content: web 2.0, wikis and social networking*. Paris.