
Article

FOLKMUN.SE: A STUDY OF A USER-GENERATED DICTIONARY OF SWEDISH

Emma Sköldberg

Department of Swedish, University of Gothenburg (emma.skoeldberg@svenska.gu.se)

Lena Wenner

Institute for Language and Folklore, Gothenburg (lena.wenner@sprakochfolkminnen.se)

Abstract

This study examines the Swedish user-generated web dictionary *Folkmun.se*, encompassing roughly 5,000 entries. Initially a general overview of the website is presented, followed by an analysis of how the content of *Folkmun.se* has developed, with particular focus on 190 usernames. The contributors can be grouped together based on two distinct factors: 1) by number of contributions, and 2) by the kind of words they prefer to submit. One conclusion of this study is that a great majority of contributors only publish one or two entries. This entails that a large number of users have contributed to the dictionary. Furthermore, a majority chooses to focus on dialect words or slang words. Many of these entries are not represented in traditional Swedish dictionaries. The advantages of having people of different backgrounds and skills working with word collections are obvious, and their work is an important, albeit often neglected, contribution to general linguistics.

Key words: Swedish, online dictionary, user study, slang, dialect

1. Introduction

Amateur lexicographers compiling Swedish words and expressions into lists of varying scopes is hardly a new phenomenon. Historically, many local heritage societies have recorded dialectal vocabulary. In some notable cases these results have also been published. Thanks to the internet, the work of amateur lexicographers and dialectologists can be made available and distributed to an extent unsurpassed in history.

In recent times there has been some scholarly interest in the results of the dictionary work currently taking place outside of professional lexicographical environments (see Section 2 below). There are, however, very few studies that have attempted to obtain information on the people behind the user-generated content of online dictionaries (cf. [Wolfer and Müller-Spitzer 2016](#)). The questions we seek to address in the present study include the

number of people who have submitted words to a given website, how the dictionary articles are revised, and whether or not it is possible to discern patterns in how and on what subject matters contributors write their articles.

The purpose of this paper is twofold. Firstly, we give a general assessment of the contents of the user-generated Swedish dictionary Folkmun.se (a rough translation of the name of the site would be ‘vernacular speech’). Secondly, we present a form of typology of the contributors to the dictionary. Before rendering the results of our findings, a short summary of recurring questions in the study of user-generated dictionaries in general is given.

2. Previous studies regarding user-generated websites

Dictionaries built on user-generated content have been the subject of increased research in recent years (see for instance [Fuentes-Olivera 2009](#); [Lew 2014](#) on content in English language resources; [Meyer and Gurevych 2012](#) and [Müller-Spitzer et al. 2015](#) for studies on German resources; see also [Rundell 2012](#); [Meyer and Abel 2018](#)).

Previous studies indicate that there are differing opinions on the value and quality of user-generated dictionary websites (see e.g. [Hanks 2012](#): 77–82 and [Meyer and Gurevych 2012](#)). For example, Wiktionary is gaining ground in a number of areas (in linguistics as well as language technology), and according to [Wolfer and Müller-Spitzer \(2016](#): 347), it has ‘great potential to become a serious competitor for publisher-based and academic dictionaries’.

Furthermore, [Lew \(2014](#): 17) notes that sites such as Urban Dictionary play an important role in documenting new words, while their staying power is still uncertain. The websites can therefore potentially have a codifying function compared to more established dictionaries. Websites also have a complementary function when it comes to documenting slang, dialects, and technical terminology ([Rundell 2012](#): 80; [Lew 2014](#): 25; [Sköldberg et al. 2019](#)). The description of a certain type of vocabulary, e.g. words of a sexually explicit nature, is also often insufficient in traditional (particularly normative) dictionaries, since many such dictionaries exclude non-socially acceptable, vulgar language (see Holm’s contribution from 2000 on the contents of some Swedish etymological handbooks). The users also provide opportunities for lexicographers to learn more about how to make dictionaries more user-friendly and to obtain information on how to make lexicographical content more accessible.

Most researchers seem to agree that user-generated websites have their limitations. Criticism has primarily been directed at the fundamental assumption that cooperation between multiple contributors automatically leads to improved content ([Mattus 2009](#): 183; [Abel and Meyer 2013](#): 179; [Mederake 2015](#): 329). Another criticism levelled against user-generated websites is that rather than cooperating to create more thorough descriptions, some contributors seem to prefer to air their prejudices or suchlike things ([Lew 2011](#)). Many user-generated websites such as Wikipedia also suffer from vandalism and ‘edit wars’, as users can change and even delete other users’ articles ([Mattus 2009](#): 192; [Abel and Meyer 2013](#)).

Moreover, [Meyer and Gurevych \(2012](#): 262) point out that the English language Wiktionary has been criticized for certain definitions being insufficient. [Gao \(2012](#): 427–428) notes that many amateur dictionary contributors appear to struggle when it comes to choice of headwords. Contributors also fail to check existing entries beforehand, which

leads to the same meanings reoccurring, albeit worded differently. According to [Lew \(2014: 17\)](#), models where users as a joint venture help to develop and improve an existing resource seem to function better when it comes to encyclopaedias (such as Wikipedia). Wikipedia is based on an expert being willing to share their knowledge with other people, free of charge. The wiki model is less robust when it comes to the task of giving traditional, conventionalized descriptions of words and expressions, their pronunciation, morphology, syntax and meaning.

This paper focuses on a Swedish lexicographical online resource. When it comes to Swedish websites, Folkmun.se has been the subject of some study (see [Sköldbberg and Wenner 2017, 2018, and 2019](#)). [Sköldbberg and Wenner \(2017\)](#) have compared the contents of Folkmun with a similar Swedish site, Luxikon.se. Other user-generated websites include Slangopedia.se, which focuses on slang words and expressions, and – of course – the Swedish part of Wiktionary. The latter is fairly limited in scope, however. [Törnqvist \(2010: 389\)](#) notes that in order for a wiki to grow and increase in quality, a certain number (a critical mass) of contributors is required. The language versions of Wiktionary which have a large user base – such as English and German – have more entries and more detailed articles than the language versions with fewer users. Languages with few speakers can have trouble reaching the critical mass required for sustainable work to take place. In other words, it is plausible that the limited size of Swedish Wiktionary is due to the relatively limited number of Swedish speakers – who also need to be interested in compiling dictionary articles.

Nordic online dictionaries with user-generated content obviously include many different types of words. According to [Törnqvist \(2015: 155\)](#), dialect and/or slang appear to dominate most of them. The more explicitly dialect-oriented websites often have a simple structure and focus on a single dialect. The scope varies between websites, from ten words to several thousand, but most sites contain a few hundred words. The dictionaries are also generally compiled by individuals who are familiar with the dialect in question and targeted at readers without specialized linguistic knowledge. According to Törnqvist, the purpose of these lists is generally to promote local identity. When it comes to slang dictionaries based on user contributions, there are first of all more extensive databases built on contributions by a larger number of users. Secondly, there are dictionaries of geographically limited slang, mainly from major cities. Thirdly, there are dictionaries of special slang. This category includes for example a dictionary of railway slang, compiled by railway enthusiasts. According to Törnqvist, these dictionaries can, despite generally containing fairly few entries, still achieve decent coverage in their particular areas.

One important motivating factor behind launching sites like Folkmun.se seems to be an interest in language in general and dialects in particular. The websites are often founded by people who have never worked professionally with lexicography, but who are knowledgeable for example in programming ([Sköldbberg and Wenner 2017: 26](#))

As noted, we know very little about the contributors and their possible motives for helping expand online dictionaries, despite previous research efforts. [Lew \(2014: 9–10\)](#) discusses three types of motivation: psychological, social, and financial. Many contributors find it psychologically satisfying to add content; their work can be viewed as an expression of altruism. The online resources also allow users to express themselves, give language advice (which happens quite frequently) and act as teachers. Of course, many contributors have a genuine interest in words, their meaning and uses. It can also be noted that those who help create a website of this kind are part of a social context. It is common for the

most committed contributors to gain a certain prestige within the group. Finally, there can be financial benefits to contributing (through various reward systems, bonus material etc.).

The fact that the contributor can often quickly see the fruits of their labour can be another important incentive. As a comparison, proposals for new words can also be sent to the editors of *Svenska Akademiens ordlista* ('The Swedish Academy Glossary'). Of course, it can be perceived as more prestigious to have a proposal accepted into the Swedish Academy's works, but even if the proposed word is included in the database, it can take years before it is publicly included in the Glossary, as new editions are published infrequently. At Folkmun.se, contributors receive recognition right away, which incentivizes them to keep contributing (Sköldbberg and Wenner 2019).

One question raised by previous studies is the factual number of individual contributors there are in any given example. Wolfer and Müller-Spitzer (2016) have conducted quantitative analyses of the English and German versions of Wiktionary. They conclude that 'Concerning the distribution of revisions over users [...] – compared to the overall user base – only very few authors are responsible for the vast majority of revisions in the two Wiktionary editions' (Wolfer and Müller-Spitzer 2016: 347). Specifically, they show that almost half (44.3 %) of all "registered authors" only make a single revision to the English version of the website. The corresponding number for the German version is 42.3 %.

3. Contents of Folkmun.se

The website Folkmun.se was initiated and is run by David Eriksson as a private initiative. He launched the website in 2007. In total, the Folkmun.se database contains 5,629 entries, of which 4,764 entries are visible on the website (December 2017). Statistics from Google Analytics show that the page has around 500 unique visitors per month.

The majority of headwords in Folkmun.se appear to consist of more or less established Swedish words and expressions, but some of the words in the dictionary appear to be neologisms stemming from the contributors themselves. The dictionary primarily consists of regional dialect and slang words (including derogatory terms). Focus is primarily on everyday vernacular language. As there is a limited range of both traditional dialect dictionaries and modern slang dictionaries available in Swedish, the website has an important function for researchers as well as members of the public with an interest in recording and disseminating different kinds of lexical entities.

Eriksson appears to have intended for the dictionary to be primarily descriptive. Many of the contributors to Folkmun.se also appear to want to document a certain use of language, but there are still some normative elements to the dictionary. This is mostly on a user level as there are contributors that mainly comment on or correct other user submissions.

Folkmun.se's start page, shown in Figure 1, shows new entries. Among the latest words added (as of late November 2018) is the dialect term *sogel* which was added in October of 2018 by the user 'Mattis'. According to the contributor, *sogel* means 'the meat in a dish'. To clarify, an example phrase is provided, with the regional pronunciation illustrated: 'ska du bara hau saus å inget sogel?' ('are you just having gravy and no meat?'). Finally, *sogel* is listed as a variation of *sovel*. The form *sovel* can be found in dialects throughout Sweden.

The users can access the content on Folkmun.se either by typing a word into the search field at the top right of the start page, or by clicking a letter of the alphabet. At the bottom of the page there are links to websites with more information about the dictionary. This

The screenshot shows the homepage of Folkmun.se. At the top, there is a search bar with the text "Sök på folkmun.se" and navigation links: "Senaste", "Lägg till ord", "Kontakt", and "Blogg". Below the search bar is a large green heading: "Senast inlagda definitioner". Underneath, a question is posed: "Är det Svenska Akademien som bestämmer över det svenska språket eller är det du?". The main content is organized into a grid of word entries, each with a title, a definition, and the contributor's name and date.

Word	Definition	Contributor
bonka Bonka - knacka/slå	Är anledningen som gör resan värt att resa dit. flera resanledningar skapar tillsammans en destination	av Stephanie C 2018-11-22 - Rapportera
grötvila Att gå och vila sig efter att ha inmundigat tex frukost, middag eller kvällsmat	En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	av Björni Englund 2018-11-01 - Rapportera
le Att vara trött på något.	Är anledningen som gör resan värt att resa dit. flera resanledningar skapar tillsammans en destination	av Mikael Olsson 2018-10-27 - Rapportera
Sogel Köttet i en maträtt. Ska du bara hau saus å inget sogel? Variant av sovel.	En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	av Mattis 2018-10-26 - Rapportera
resanledning Är anledningen som gör resan värt att resa dit. flera resanledningar skapar tillsammans en destination	En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	av Ess Group 2018-10-23 - Rapportera
Halaltarian En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	av Aphrodite 2018-10-19 - Rapportera
Iglug/ Iglu Enveten/envis som en igel. På svenska "iglig". Dialektalt iglug. Uttalas med tjockt första g och tjockt l och utan sista g - iglu	En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	av HBE 2018-10-09 - Rapportera
Uttjimmad Ett annat ord för utmattad, utarbetad, slutkörd, utsjasad (Vaggerys trakter)	En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	av Therese A 2018-10-05 - Rapportera
Vattensvält Uttorkning, tärst	En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	av Manfred 2018-10-03 - Rapportera
Klapp Tex. "Vi klapp den rakt av!" Meningen med detta uttryck är att man "klippte/snodde" den/det personen i fråga pratar om.	En person som äter bara halalskiaktat mat, ej gris, om det inte finns så äter de det vegetariska alternativet.	av Linzhy 2018-09-27 - Rapportera

Figure 1: Folkmun.se (start page, the image has been slightly cropped, accessed 26/11/2018).

includes a top ten-list of the most active contributors to the site. Without an ad blocker, the site also displays advertisements and banners which may affect the overall impression and credibility of the dictionary and the overall effectiveness of the dictionary (Dziemianko 2019).

The entries in Folkmun.se are of varying length, but all contributions contain, in addition to the headword itself, a description of the word's meaning, a username, and a date. Many entries also include example phrases figuring the word in context, as well as information on the regional usage of the headword, its etymology and stylistic level.

The material found in Folkmun.se is not systematically normalized. Users are nonetheless encouraged to check that the information is not already available on the website when submitting a new word or meaning. Different spellings, inflections and meanings of the same word often make up different entries. (cf. Gao 2012 in Section 2).

Below, a few typical example entries listed under the letter K are shown. The article contents are presented unedited, i.e. including possible deviations from standard language. An approximate English translation is provided for each example.

- (1) **Krabbig** Svår, jobbig Gör det inte svårare än vad det är. Ex: krabba inte té at. av Rebecca 2018-08-15
Krabbig Difficult, annoying Do not make something more difficult than it is. Ex: don't complicate things. by Rebecca 15/08/2018
- (2) **Knulrufs** Att vara rufsig i håret, framförallt bakhuvudet, så att det lätt går att misstänka att personen i fråga nyss haft sex. av Gambleputti 2009-08-12

Knullrufs Having messy hair, especially at the back of the head, implying that the person recently has had sex. by Gambleputti 12/08/2009

- (3) **Kånkelbår** Avföring som fastnat i stjärthåret! Hela Sverige. av Oxñ 2007-07-16
Kånkelbår Faeces stuck in the arse hair! Throughout Sweden. by Oxñ 16/07/2007
Kånkelbår Med 'kånkelbår' avses hemorrojder i analen. av Hasse 2011-03-03
Kånkelbår 'Kånkelbår' refers to anal haemorrhoids. by Hasse 03/03/2011

The adjective *krabbig* is one of many dialect words from southern Sweden that can be found in Folkmun.se but not in traditional dictionaries. Both *knullrufs* and the polysemous *kånkelbår* are very informal words, which may explain their absence in more general language dictionaries. These types of words are also common on the website. Pertaining to *krabbig*, it is noteworthy that the word is explained using two synonyms ('difficult, annoying'). The word is then used in an example sentence.

It should also be noted that the contributor uses the adjective *krabbig* as well as the verb *krabba*. It is common for contributors to move between word classes in their articles, which goes against what is usually an important basic principle of traditional Swedish lexicographical work (Svensén 2009: 226, 228). The basic tenet of this principle is that the headword and the definition should be interchangeable without addition/loss of meaning. The present findings indicate, perhaps unsurprisingly, that this principle seems to be more important to professional lexicographers than to the user-generated content contributors.

Concerning the noun *kånkelbår* in (3), previous research by (among others) Aaltonen and Seiler (2016) has shown that contributors appear to be more motivated to build on existing dictionary entries than to create brand new ones. Another example is the entry *datanörd* ('computer nerd'):

- (4) **datanörd** Människa som har en livsstil som innebär att han/hon sitter mycket vid datorn. Bl. a. surfas det på Internet, skrivs dikter och annat. av Anonym 2011-04-02
datanörd Person with a lifestyle where he or she spends a lot of time in front of the computer. They can surf the internet, write poetry etc. by Anonymous 02/04/2011
- (5) **datanörd** Entusiastisk datatekniker eller liknande person, som går in för liv och lust i det senaste med datorer, kopplar till ovanlig utrustning, installerar osannolika program och rentav programmerar datorer. av Dagobert 2012-01-24
datanörd Enthusiastic computer technician or similar person, who is passionate about the latest in computers, adds unusual equipment, installs unlikely software, and might even program computers. by Dagobert 24/01/2012

Perhaps 'Dagobert' did not find the description provided by 'Anonymous' of *datanörd* to be completely satisfactory, as he or she chose to give a more elaborate description. It should also be noted that both of the descriptions are positive (cf. *Svensk ordbok utgiven av Svenska Akademien* 2009, which states that the term *nörd* is 'somewhat derogatory').

Yet another example concerns the homonyms *pucko* and *Pucko*. In the first case, it is an adjective meaning 'idiot', in the second case it signifies the name of a chocolate drink. This particular case also shows that users are open to submitting more encyclopaedic information to the website. The traditional division of labour between dictionaries and encyclopaedias is thus upended here, when users are put in charge (cf. Svensén 2009).

The start page of Folkmun.se also poses the rhetorical question: 'Who owns the Swedish language, the Swedish Academy or you?' (see Figure 1). This question highlights the innate

power dimension of language and language use. It can be interpreted as an aspiration by Eriksson to raise the status of vernacular language. The same ambition is displayed by the founder of Urban Dictionary (Peckham 2007: VI), who also emphasizes the inherent knowledge of language found among normal language users. Folkmun.se, like many other Swedish user-generated dictionaries, is thus characterized by rebellion against the establishment's (description of) language.

Furthermore, when the function 'Lägg till ord' ('add words') is activated, the user is shown editorial guidelines for prospective contributors. For example, he or she is encouraged to write a definition as well as to provide information on word class and regional usage. This implies that the authors of entries are expected to possess certain basic grammatical knowledge. Contributors must also provide a valid e-mail address. Finally, the user, as a means of verification, must confirm the entry by clicking a link sent to the e-mail address. Only then is the entry made public on the site.

Which words the website should include – and how the words should be presented – is to some extent governed by the instructions to the contributors. Sköldbberg and Wenner (2017) show that the simpler, down to earth instructions found on the user-generated website Luxikon lead to more informal and extensive entries than on Folkmun.se. On the other hand, a number of contributors deviate from the editorial instructions. This may be due to the fact that they find instructions difficult to understand, but it is equally likely that they have not read the instructions or checked which words have already been included in the dictionary. To some extent, these problems have been described in the previous-research-section (see above).

Contributors occasionally add socially unacceptable words or offensive expressions. Folkmun.se attempts to resolve this issue by allowing anyone to report entries that are deemed unsuitable. Contributors can also, when submitting a headword, check a box to indicate that the provided data is 'adult content or politically incorrect'. These functions simplify to some extent the editorial process, by flagging words that other users may find offensive. However, this does not automatically lead to the word or definition being deleted.

Eriksson is the sole moderator and he solely determines what entries are published on the website (cf. the rhetorical question about the Swedish Academy above). He reviews all submitted and reported words. In that sense the site could be categorized as a semi-collaborative dictionary (cf. Abel and Meyer 2013: 185, 190; Meyer and Abel 2018: 749–750). Although he, as stated in the user directions, retains the right to edit or delete submitted definitions, he is restrictive about deleting or censoring material. Eriksson justifies this by noting that if a derogatory word exists, it is useful that language users can google the definition and hopefully choose not to use it. Certain, especially highly derogatory, entries remain unpublished, but the most common reason for not publishing an entry is that the contributor has not consented to the article being published on the website, or that they have accidentally submitted the same article multiple times.

4. Contributions to Folkmun.se

4.1. Description of material and method

Eriksson has given us access to the part of the Folkmun.se database that contains all entries submitted to the website from its launch in 2007 to December 2017.¹ We have chosen to

examine the digital traces left by users, hoping to learn more about the types of contributors to sites such as Folkmun.se. Our material shows that the total of over 5,500 entries have 1,764 unique contributors. The spread between different contributors must in this respect be considered to be fairly large (cf. [Wolfer and Müller-Spitzer 2016](#) above).

[Sköldb​erg and Wenner \(2018 and 2019\)](#), which can be seen as pilot studies for the present study, examine the traces left by a limited group of randomly selected contributors (between 2007 and 2016). In total, the two studies cover almost 470 submissions. The method used in the studies has similarities to lexicographical user studies by, for example, [Almind \(2008\)](#) and [Hult \(2017\)](#). Almind discusses the possibility of classifying user types and usage situations based on log files of a certain dictionary. Hult is interested in individual search behaviour and she attempts to get closer to individual users of a learner dictionary by linking log file data with questionnaire data.

The findings of the pilot studies indicate that many contributors appear to want to express something specific through their choice of username, for example their geographical background. Patterns can also be discerned in the contributors' methods, for example the number of entries they compile per sitting, and the tempo of their work. There are also patterns to what types of words they submit (single words or multi-word expressions), the meaning of the submitted headwords, and their stylistic level. Interestingly, many submissions also turn out to be systematically structured, indicating a high level of knowledge of grammar and dialectology among the contributors.

One example of a contributor in the pilot studies is 'Go4it', who made a total of five submissions in 2013, and all entries were published on the website. The submissions – four single words and one expression – are all fairly established colloquial expressions, and according to Go4it they are specific to the city of Karlstad region (in the province of Värmland). In addition to the headword and the information given about the dialect, the articles include a number of other information categories, such as pronunciation, variations in meaning, connotations, and examples. In this way, Go4it's submissions are fairly rich in information, but some of the articles also contain a number of typos and deviations from Swedish writing conventions.

Another example from the pilot studies is 'Chooniiii'. The username might be linked to the multi-ethnic word *shonolshonne* (of unknown origin) meaning 'guy' and given the nature of the submissions this does not seem entirely improbable. The username has submitted eight words from multilingual environments. The entries, concerning female appearance and criminal acts, were submitted in 2009.

In the investigation presented in this study, a larger number of submissions have been analysed in order to confirm or reject the findings from the pilot studies. More specifically, we have sorted our material alphabetically by the contributors' usernames, and to delimit the study, we picked every tenth username in the list. This gave us 190 unique usernames with a total of 1,512 submitted articles.² First of all we have studied the *number* of submissions by each respective username and divided the contributors into groups based on the number of submissions. We have also studied the *contents* of the submissions, and divided the contributors into groups based on the primary nature of their submission. In these categorisations, we have mainly considered the types of words and expressions, whether they have anything in common geographically, stylistically, or semantically.

4.2. Similarities and differences in number of submissions

The contributors in our data set can be divided into three groups, 1) testers 2) medium producers, and 3) major producers, based on the number of submissions. All contributors with up to five unique submissions have been included in the group of testers. The data set comprises a total of 163 testers (of a total of 190 studied contributors), constituting 86 % of the contributors that we have chosen to study. The testers thus only make up 15 % of the total submissions. Further, 21 contributors made between 6 and 29 submissions and we have chosen to call these users medium producers; they constitute 11 % of the contributors in our data. Together, they have made 265 submissions, which is 17 % of the total number of submissions. Finally, there are individuals who have made 30 or more submissions, henceforth called major producers. This group includes 6 contributors, corresponding to 3 % of the total number of contributors. The major producers have made a total of 1,022 submissions, which comprises 68 % of the data (see [Table 1](#)).

4.2.1. Testers

Thus, all contributors with up to five unique submissions have been included in the group of testers. It is worth noting that the majority (121 persons) of these have only made a single submission (cf. [Wolfer and Müller-Spitzer 2016](#)). In this group, there is a contributor with the username ‘Pia Larsson’. He or she has added the following entry, which was subsequently published on the website:

- (6) **hydd** Den första tunna isskorpan på vatten (oavsett typ; vattenpöl, tjärn, sjö, havet, åar osv). av Pia Larsson 2017-10-31
hydd The first, thin covering of ice on water (regardless of type; puddle, pond, lake, sea, brooks etc). by Pia Larsson 31/10/2017

The fairly extensive description of the dialect word has at the time of submission also been tagged with ‘hälsingemål’ (i.e. language from the province Hälsingland) and ‘noun’ – tags which only the moderator can see. The given definition of *hydd* has not previously appeared in writing and it thus supplements the existing dialect descriptions.

Another tester with only one submission is ‘Nils’. The submission in question (from 2016) is the verb *kurka*, meaning ‘to get stuck in on an upwards incline, for example with a lorry or freight train’. A search for the word on Folkmun.se shows that the word *kurka* already had two entries when this definition was submitted, with different meanings: ‘to stop working, to give up’ (submitted by ‘roger.c’ in 2010) and ‘exhaust someone’ (submitted by ‘Håkan Stattin’ in 2011). The three meanings are semantically linked.

Another tester has chosen to call himself ‘Micke Hell’. This contributor has added three words: *grada* (‘to check the temperature of, for example, water’), *trö* (‘to tread on someone

Table 1. Different types of contributors based on their number of submissions.

Types of contributors	Number of contributors		Number of submissions/group	
Testers (1–5 submissions)	163	86 %	225	15 %
Medium producers (6–29 submissions)	21	11 %	265	17 %
Major producers (> 30 submissions)	6	3 %	1,022	68 %
Total	190	100 %	1,512	100 %

or something’) and *fryna* (which is included in *fryna på näsan*, which according to the contributor means ‘to reject or decline something, for example food in some form’ – possibly related to the English expression ‘to frown upon something’). A particular feature of Micke Hell is that he or she, in addition to providing meanings and examples, attempts to explain how the words are pronounced. For example, *trö* is pronounced with ‘a rolling r and a very long ö’, which ought to correspond to a south Swedish pronunciation.

4.2.2. Medium producers

As already mentioned, 21 contributors made between 6 and 29 submissions. One example of the medium producer group is ‘Evert’. He or she has in 2010 added six established words from the province Skåne (Scania), accompanied with synonyms. The headwords, which belong to different word classes, are *töj* (‘clothes’), *lär* (‘leather’), *grisk* (‘gluttonous, greedy, eager for something’), *arri*, *arrig* (‘angry, upset’), *skamma* (‘admonish’) and *skamma sig* (‘to be ashamed’).

Other medium producer has the username ‘Olsson’. He or she has, in 2011, added 16 single headwords that he or she has categorized as ‘bohuslänska (‘Bohuslän dialect’), island language’. Among the headwords are common words with non-standardized spelling intended to represent dialectal pronunciation, i.e. words that are normally not included in traditional dialect collections. But Olsson also submitted more purely dialectal words. All submissions have very rudimentary definitions, and it can be noted that they are almost in alphabetical order. Considering the speed at which the entries were submitted, perhaps ‘Olsson’ was using some form of written source. This can be compared with the discussion regarding the submissions by ‘Hetrometrus’ in Sköldbberg and Wenner (2019).

4.2.3. Major producers

In our data set there are 6 contributors who have made 30 or more submissions each. Thanks to their numerous submissions, many of them are listed on the website’s top ten list of most frequent contributions. An example of a major producer is the user ‘Arla’. Over the course of roughly two years (August 2007–December 2009), he or she added around 70 words, all of which are publicly shown on the website. After this period of activity, he or she has only made one more submission, in August of 2013. Among the submitted headwords are e.g. *player* (‘person with many relationships’), *lack* (‘annoyed, angry’) and *faila* (‘to fail’).

Among Arla’s submissions there are a fairly high number of relatively recent slang terms, submitted as early as 10 years ago. Other words have roots in multi-ethnic youth language. Among the contributions there are also acronyms which have made their way into written Swedish, not least through social media (e.g. *OMG*).

4.3. Similarities and differences in the nature of submissions

In the same way as users can be divided into groups based on their number of submissions, they can be grouped by the type of words they primarily submit. We have identified five main types of entry authors: 1) dialect focused (76 persons), 2) slang focused (71 persons), 3) omnivores (13 persons), 4) provocateurs (7 persons) and 5) other (23 persons). Below these groups are presented with examples. As shown by the numbers in parenthesis above, the first two groups are by far the largest (see also Table 2).

Table 2. Different types of contributors based on the nature of their submissions.

Types of contributors	Number of contributors
Dialect focused	76
Slang focused	71
Omnivores	13
Provocateurs	7
Other	23
Total	190

4.3.1. Dialect focused

A large number of the contributors have almost exclusively submitted dialectal words. Contributors who focus on dialect words include the aforementioned ‘Pia Larsson’ (Hälsingland), ‘Micke Hell’ (Skåne), ‘Evert’ (Skåne) and ‘Olsson’ (Bohuslän).

Another contributor who has chosen to focus on dialect words and expressions is ‘Mikaelystrand’. In 2011, he or she added the word *tvesula* combined with the explanation ‘when you put double toppings on a sandwich, e.g. ham and a slice of cheese’. The headword appears in dialects in several parts of Sweden, especially in the west and south. It is also worth noting that the word is etymologically related to *sogel*, which was discussed in section 3. Another contributor who focuses on dialects is ‘Kristina’. In February 2016 he or she added the adjective *nåk* and in connection with this he or she added the comment ‘Among farmers from the Piteå region it means being mean’. The description indicates that he or she, as many others contributors, refers to an existing article. In the dictionary you do indeed find two already existing articles for *nåk*. First, ‘Hanna Löfdahl’ added an extensive article in 2009, then ‘Lefte HÅ’ almost one year later.

- (7) *Nåk* *nåk*, *nåk* -a, *nåk* -t. Ångermanländskt adjektiv med negativ ton. dålig; sjuk; underlig; smutsig; äcklig, osv. “Vilken *nåk* film”, “Jag mår *nåkt*”. av Hanna Löfdahl 2009-03-03
Nåk *nåk*, *nåk* -a, *nåk* -t. Ångermanland adjective with negative connotations: poor; ill; strange; filthy; disgusting, etc. “Vilken *nåk* film” (What a *nåk* film), “Jag mår *nåkt*” (I feel *nåk*). by Hanna Löfdahl 03/03/2009
- (8) *Nåk* = Tapig, intetsägande, utan utstrålning, torftig (Skellefteå västra). av Lefte HÅ 2010-02-13
Nåk = Drab, uninteresting, without charisma, dry (Skellefteå west). by Lefte HÅ 13/02/2013

The adjective *nåk*, which is used (at least) in different parts of northern Sweden, appears to have a few different albeit related meanings. A common factor, however, is that they have negative associations.

4.3.2. Slang focused

As previously mentioned, there is a group of contributors who primarily add slang words and expressions. One such contributor is ‘Jones’. He or she made four submissions within minutes of each other in March of 2012. The submissions, which are all publicly visible on the website, are *bröta* (‘to make noise’), *svina* (‘throw a ball very hard’), *stret* (‘idiot’), and the fairly archaic expression *rulla batt* which is defined as ‘partying’.

Another contributor who primarily has added slang words chose the username ‘John D’. He or she belongs to the major producer category, and has made over 190 unique

submissions, which puts him/her in fourth place on the website's top ten most frequent contributors. 'John D' turned up in 2010 and has compiled about 100 dictionary articles. This contributor has mainly submitted single headwords, but also a number of established multi-word expressions of a clear slang nature. Unlike the words submitted by 'Arla', 'John D' primarily adds words from an older vocabulary, much like 'Jones'. All submissions are publicly visible, despite 'John D' is using the option to classify them as 'adult content' (see Section 3). The moderator has chosen to publish the entries anyway.

4.3.3. Omnivores

The omnivores group includes users who have submitted words of fairly varying (omnivorous) nature. One example is 'Rudolf 1922', who in 2009 made four submissions. The words were *bladneger* (older slang), *puma* (modern slang), *milf* (modern slang) and *dorving* (dialect). *Milf* is explained in depth:

- (9) **Milf** är ett ord som har sitt ursprung i den amerikanska porrbranschen där det är en förkortning av "a mother I would like to fuck". Ordet har kommit i användning i svenskan som benämning för en medelålders kvinna som är sexuellt attraktiv inte bara för sina jämnåriga utan jämväl för yngre män.
Milf is a word originating in the American porn industry where it is an abbreviation of "a mother I would like to fuck". The word has entered into Swedish usage as a term for a middle-aged woman who is sexually attractive not only to persons of her own age but also to younger men.

Another contributor classified as an omnivore goes by 'Danne'. He or she submitted 22 words and expressions in June, 2008. Among the submissions, there are fairly common – albeit colloquial – general language words, dialect words, old slang words, urban ethnolects, etc.

4.3.4. Provocateurs

Vandalizing Folkmun.se is fairly difficult as you cannot edit or delete other users' submissions (see Sköldbberg and Wenner 2018: 244). However, it is clear that some contributors seem to be motivated by something else than, for example, describing their own dialect. We have chosen to call these persons provocateurs, as they mainly appear to want to annoy the website's owner and users. Our data set only contains a few such contributors, and among them is one with the username 'Fröjd'. He or she has only made one submission, the headword *folkmun* (cf. the name of the dictionary), defined as 'A useless garbage page on the internet, with ridiculous definitions'. The entry is not publicly visible.

Another example from this group of contributors is 'General Fist' who has also only made one submission. The word is the nonsensical compound *dubbeltelefonkukmacka* (lit. 'double-phone-cock-sandwich'), which is defined as 'When you tape two phones to each side of your penis, and put a condom over them, in order to increase the girth of your manhood.' This entry is not publicly visible on the website, but we have had no means of ascertaining if this is due to the contributor not completing all steps of the submission process, or if the moderator stopped the entry.

4.3.5. Others

Finally, there are contributors who do not fit into any of the groups described above. These include users who primarily sent in new words that they themselves had created, often because they see a gap in the existing vocabulary. One contributor which we have included in

this group is ‘Ingelöw’, who in 2009 submitted the, for us unfamiliar, and most likely made up word *skvada*. The word means ‘many unnecessary words’, i.e. ‘when someone writes or speaks too much or uses too many unnecessary words and explanations’. The entry is not publicly visible on the website.

This group of contributors also includes ‘Victor’, who in 2007 added the word *punsch* (‘Swedish arrak punch’) followed by encyclopaedic information about this type of liqueur, and a brief etymological explanation. Another contributor in this group is ‘Tjeder’ who only made one submission in 2008, namely a lengthy explanation of the term *kuttersmycke* (originally ‘a figurehead adorning a boat’) and the semantic change of the noun over time. Both submissions are visible on the website.

5. Concluding discussion

There are several user-generated Swedish language dictionary websites, and one of them is Folkmun.se, which was launched around ten years ago. By reviewing the organic process of how users have acted when given more freedom, a new perspective can be gained on traditional solutions and ways of working in lexicography (with regard to information categories, sorting and presentation of data). Our material shows among other things that contributors do not draw a strict line between proper nouns and common nouns, and the traditional requirement for substitutability between dictionary words and definitions does not appear to be very important.

For linguists and lexicographers, when working with materials such as Folkmun.se it may be difficult to accept the prevalent lack of information compared to traditional dictionaries. What motivates Eriksson and many of his colleagues is a non-academic interest in (or love of) language. From Eriksson’s perspective, the website’s functionality is probably completely satisfactory. The fact that there happen to be duplicate entries etc. appears to be less important – and as can be seen by some of the multiple entry-headwords, there is an undeniable value in documenting meaning and usage in this collage technique.

An interesting question is what drives users to spend time and energy on developing these user-generated resources. In an attempt to get closer to the contributors, we have looked at the traces they left in the database Folkmun.se. More specifically, we investigated a data set of 1,512 dictionary entries submitted by 190 users.

We have studied the number of submissions made by each user. An in-depth examination of the data has shown that contributors can be divided into three groups, testers, medium producers and major producers, based on the number of submissions. We have also studied the content and nature of the submissions made by each selected user. In this study, we have identified the following five main types: dialect focused, slang focused, omnivores, provocateurs, and others.

In the data we studied, the tester category is by far the largest. Many users have only submitted a single word. These persons appear to somehow have found the website, and ‘tested’ the site by making a submission. The fact that the testers are so many could be interpreted to indicate that a larger, more varied vocabulary is described (including words from different parts of the country). At the same time, there are plenty of users who choose to allocate a fair amount of time to help expand the dictionary. Some of them have submitted more than 100 words. From our perspective it is fruitful that many people are involved in describing the Swedish language, particularly such language use that traditional corpus-based lexicography has had difficulties documenting. The users’ contributions should thus

be viewed in a wider context, where many dictionary publishers are struggling with rapidly declining sales, and where dictionary production is limited to a few specialist environments.

In terms of content, Folkmun.se mainly contains dialect words and slang words, which might – at least to some extent – be related to the name of the site. As access to up-to-date Swedish dialect and slang dictionaries is limited, it is important that vernacular speech is described on Folkmun.se. Other parts of the vocabulary are covered to a substantial extent by traditional dictionaries, such as contemporary works produced by the Swedish Academy (cf. Sköldberg et al. 2019).

The study shows that many different Swedish dialects are represented, but as many of the major producers describe dialect words from southern Sweden, especially Scania, southern Swedish words and meanings are highlighted. The slang words included are varied. Some words are associated with older Stockholm slang, while others are representative of urban ethnolects in the suburbs of today.

It is clear that the digital format has many advantages when it comes to describing the more ephemeral and fluctuating parts of the vocabulary. There is no space limitation, which allows for more entries and (potentially) longer articles, and the contributors (as opposed to professional lexicographers) do not need to decide if the word in question has had any longevity in modern colloquial language, since the website is updated continuously. This process entails that more ephemeral language use is documented. The words and the language use documented by sites such as Folkmun.se can henceforth be regarded as a potentially important source for future research.

One caveat pertaining to the fact that many of the websites are created and owned by private individuals is that there is no guarantee that the contents will remain available to future researchers. Legislation on the preservation of digital materials is still in development in Sweden. Ideally, websites of this nature could be maintained by an institution, guaranteeing that the material will be preserved for the future.

Notes

1. A heartfelt thank you to David Eriksson who graciously has shared material and answered our questions. With regards to the study's research ethics aspects, the contributors have not been informed of their participation in the study. The data we have received is anonymized to make linking to physical persons impossible. We thus consider ourselves to have complied with existing research ethics guidelines regarding informant anonymity and protection of informants.
2. The same contributor could of course have used multiple usernames, and various contributors could share the same username. However, for the sake of simplicity we have decided to disregard this in our study.

References

A. Dictionaries

Folkmun.se. Accessed on 5 December 2018. <http://www.folkmun.se/>.

Slangopedia.se. Accessed on 5 December 2018. <http://www.slangopedia.se/>.

Svenska Akademiens ordlista. 2015. (*The Swedish Academy Glossary*). Stockholm: Norstedts. <http://svenska.se/>.

Svensk ordbok utgiven av Svenska Akademien. 2009. ('The Contemporary Dictionary of the Swedish Academy'). Stockholm: Norstedts. <http://svenska.se/>.
Wiktionary. Accessed on 5 December 2018. <http://sv.wiktionary.org/wiki/Wiktionary/>.

B. Other literature

- Aaltonen, A. and S. Seiler. 2016. 'Cumulative Growth in User-Generated Content Production: Evidence from Wikipedia.' *Management Science* 62.7: 2054–2069.
- Abel, A. and C. Meyer. 2013. 'The Dynamics Outside the Paper: User Contributions to Online Dictionaries' In Kosem, I, J. Kallas, P. Gantar, S. Krek, M. Langemets and M. Tuulik (eds), *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the ELex 2013 Conference, 17–19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 179–194.
- Almind, R. 2008. 'Søgemønstre i logfiler.' *LexicoNordica* 15: 33–55.
- Dziemianko, A. 2019. 'The role of online dictionary advertisements in language reception, production, and retention.' *ReCALL* 31(1): 5–22.
- Fuertes-Olivera, P. A. 2009. 'The Function Theory of Lexicography and Electronic Dictionaries: WIKTIONARY as a Prototype of Collective Free Multiple-Language Internet Dictionary' In Bergenholtz, H., S. Nielsen and S. Tarp (eds), *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, 99–134.
- Gao, Y. 2012. 'Online English Dictionaries: Friend or Foe?' In Fjeld, R. V. and J. M. Torjusen (eds), *Proceedings of the 15th EURALEX International Congress 7–11 August 2012*. Oslo. Department of Linguistics and Scandinavian Studies, University of Oslo, 422–433.
- Hanks, P. 2012. 'Corpus Evidence and Electronic Lexicography' In Granger, S. and M. Paquot (eds), *Electronic Lexicography*. Oxford: OUP, 57–82.
- Holm, G. 2000. 'Fitta, fucka, knulla, kuk – en grupp mycket anrika ord, som undviks i bildat samtal' In Jönsson, A and A. Piltz (eds), *Språkets speglingar. Festskrift till Birger Bergh*. Lund, 368–372.
- Hult, A-K. 2017. *Ordboksanvändning på nätet. En undersökning av användningen av Lexins svenska lexikon*. (Göteborgsstudier i nordisk språkvetenskap 27.). Göteborg.
- Lew, R. 2011. 'Online Dictionaries of English' In Fuertes-Olivera, P. A. and H. Bergenholtz (eds), *E-lexicography. The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum, 230–250.
- Lew, R. 2014. 'User-Generated Content (UGC) in Online English Dictionaries.' *OPAL – Online publizierte Arbeiten zur Linguistik* 2014.4: 8–26.
- Mattus, M. 2009. 'Wikipedia – Free and Reliable? Aspects of a Collaboratively Shaped Encyclopaedia.' *Nordicom Review* 30: 183–199.
- Mederake, N. 2015. 'Overwriting knowledge: Analyzing the dynamics of Wikipedia articles.' In Jakubiček, M., J. Kallas and S. Krek (eds), *Electronic Lexicography in the 21st Century: Linking Lexical Data in the Digital Age. Proceedings of the eLex 2015 Conference, 11–13 August 2015, Herstonceux Castle, United Kingdom*. Kosem. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., 327–341.
- Meyer, C. M. and A. Abel. 2018. 'User Participation in the Internet era.' In Fuertes Olivera P. A. (ed), *The Routledge Handbook of Lexicography*. Abingdon/New York: Routledge, 735–753.
- Meyer, C. M. and I. Gurevych. 2012. 'Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography' In Granger S. and M. Paquot (eds), *Electronic Lexicography*. Oxford: OUP, 259–291.
- Müller-Spitzer, C., S. Wolfer and A. Koplenig. 2015. 'Observing Online Dictionary Users: Studies Using Wiktionary Log Files.' *International Journal of Lexicography* 28.1: 1–26.

- Peckham, A. 2007. *Mo' Urban Dictionary: Ridonkulous Street Slang Defined*. Kansas City: Andrews McMeel.
- Rundell, M. 2012. 'It Works in Practice but Will it Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical' In Fjeld, R. V. and J. M. Torjusen (eds), *Proceedings of the 15th EURALEX International Congress 7–11 August 2012*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 47–92.
- Sköldberg, E., L. Holmer, E. Volodina and I. Pilán. 2019. 'State-of-the-art on Monolingual Lexicography for Sweden.' *Slovenščina* 2.0, 7 (1), 13–24.
- Sköldberg, E. and L. Wenner. 2017. 'Folkmun och Luxikon. En jämförelse mellan två användargenererade ordbokssajter.' *Språk och stil, Tidskrift för svensk språkforskning* 27, 21–48.
- Sköldberg, E. and L. Wenner. 2018. 'Amatörlexikografiska insatser på sajten Folkmun.se' In Svavarsdóttir, Á., H. Jónsdóttir, H. Hilmisdóttir and T. Úlfarsdóttir (eds), *Nordiske Studier i Leksikografi* 14. Reykjavík, 237–245.
- Sköldberg, E. and L. Wenner. 2019. 'Amatörlexikografer – vilka, hur och varför? Om insatser på den användargenererade sajten Folkmun.se' In Bianchi, M., D. Håkansson, B. Melander, M. Westman and Östman Carin (eds), *Svenskans beskrivning* 36. Uppsala: Institutionen för nordiska språk vid Uppsala universitet, 295–306.
- Svensén, B. 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Törnqvist, L. 2010. 'Brukarmedverkan i webbordböcker' In Jóhannesson, K., I. Larsson, E. Magnusson Petzell, S-G. Malmgren, L. Rogström and E. Sköldberg (eds), *Bo65. Festskrift till Bo Ralph. (Meijerbergs arkiv för svensk ordforskning 39.)*. Göteborg, 383–390.
- Törnqvist, L., 2015. 'Nordiska dialekt- och slangordböcker på Internet.' *LexicoNordica* 22: 57–75.
- Wolfer, S. and C. Müller-Spitzer. 2016. 'How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions.' *Lexikos* 26: 347–371.