



**« The role of collaborative web projects for lexicography :  
the Wiktionary »**

**Bence Damokos**  
Vice president – Wikimedia Hungary  
bence.damokos@wikimedia.hu

**Adrienne Charmet – Alix**  
Directrice des programmes – Wikimédia France  
adrienne.alix@wikimedia.fr

Good morning!

Thank you for the opportunity to present here. My name is Bence Damokos, and I was asked to step in and make this presentation instead of Adrienne Alix who could not be here today. I am using her translated slides and notes today, I hope I can do justice to the topic – any errors and mistakes are my own.

I am the vice president of Wikimedia Hungary, the Hungarian NGO that promotes and supports free knowledge projects such as Wikipedia and Wiktionary itself. I am doing this in a volunteer capacity as an editor more of Wikipedia than Wiktionary itself, my background is in English Studies – where I have taken a course on lexicography - and International Relations, and I am currently starting to do my masters thesis on linguistic rights in Central Eastern Europe.

Today, I am going to speak to you about Wiktionary a collaboratively created free online lexicography community that aims to describe all words of all languages.

I - Le « Wiktionnaire », projet de dictionnaire multilingue libre  
*Wiktionary, a free multilingual dictionary*

II - Le fonctionnement d'un dictionnaire multilingue collaboratif  
*How a multilingual and collaborative dictionary works*

III - Le Wiktionnaire, un outil pour promouvoir et diffuser les langues  
*Wiktionary, tool for promoting and disseminate the languages*

In the first part of my talk, I will present the concept of Wiktionary, what it is, why it was created a year after Wikipedia itself

In the second part I will go into a bit of detail on how this multilingual and collaborative dictionary works

While in the final, third part, I will briefly explain what can be the role of Wiktionary in projects that promote languages and the collaborative development of languages



To start with a bit of history, the first wiki (meaning fast in Hawaiian) was set up by Ward Cunningham in 1995.

Wikipedia was set up in 2001 as a freely editable online encyclopedia that was intended to be the staging ground for articles to be published in the traditionally curated encyclopedia called Nupedia. However, Wikipedia quickly surpassed all expectations in terms of community size, the number of articles and even the quality, so that eventually the idea of Nupedia was abandoned.

The idea of Wiktionary grew out from the community of Wikipedia, the editors wanted a separate space to provide dictionary definitions of ordinary terms that might not have a place in a proper encyclopedia. So the English Wiktionary was established in December 2002.

Wiktionary was the first of a number of so called sister projects that now make up the Wikimedia family, maintained from donations by the non-profit Wikimedia Foundation. The family includes about 10 main projects, including the encyclopedia Wikipedia, Wiktionary, the first sister project to be created, but also a huge multimedia archive called Wikimedia Commons, a site for digitized public domain texts called Wikisource, and a number of others. The common theme among these sites are that they are all free to edit, the content can be freely reused and they all run on the open source Mediawiki wiki software. It should be noted that because the software itself is free and the word „wiki” is not a protected trademark, there are millions of similar looking sites that have nothing to do with the Wikimedia family.

## Wiktionary is a dictionary...

### \* Everyone can edit

- no selection of contributors
- no obligation to create an account
- community of contributors
- openness to everybody

### \* published under a free license

- content freely reusable
- content freely editable
- obligation to cite the authors
- obligation to keep the content free



Wiktionary is a dictionary and as I will explain, it is a bit more than a dictionary, is a multilingual web-based project to create a free content dictionary.

It can be edited by anyone, there is no formal editorial board, or any selection criteria, one doesn't even need to register to participate. This doesn't mean that the content is unreliable or that anyone can add any nonsense to it and no-one would be the wiser. Wiktionary has a community of active participants who care deeply about the project and who follow the changes almost in real time on the projects and try to verify each entry. There are two ways to achieve this: the policies of the project, especially the criteria for inclusion require that wherever possible a proofs, citations or references be provided and that the content cannot infringe other's copyrights. The second principle is that when in doubt the more people see a given problem, the more likely a solution will be found, so Wiktionary has a number of forums to debate the inclusion of certain words, to verify etymologies and translations.

The second very important distinguishing characteristic is that the content is published under a free licence, the Creative Commons Attribution Share Alike licence. This means that the content can be freely reused, edited, with the conditions that any modified versions of the content need to be published under the same conditions of freedom.

The screenshot shows the French Wiktionary page for the word "bateau". The page is in French and includes a navigation menu on the left, a search bar at the top, and a main content area. The main content area features a welcome message, a "Le saviez-vous ?" section, and a detailed entry for "bateau". The entry includes the word's etymology, its singular and plural forms, and a list of definitions. A small image of a sailboat is also visible.

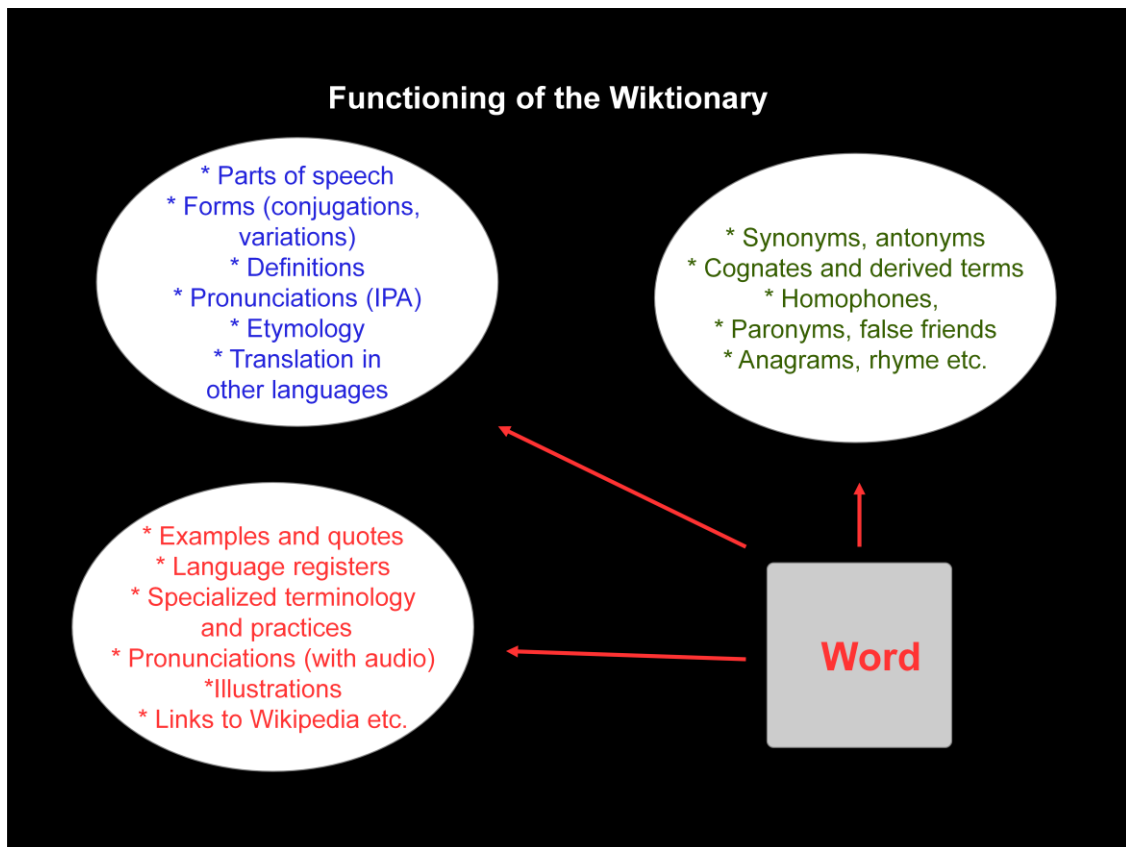
Wiktionary → 170 different languages (Wikipedia : 285) – online and in Android App format  
 14,883,341 pages, including 3 million+ on the English Wiktionary  
 27 languages with more than 100,000 articles ; 3 languages > 1 million (English, French, Malagasy)  
 Wiktionary Fr → words description in 2 200+ languages; around 320 active contributors

The Wiktionary interface and the language of definitions and linguistic terms is available in 170 different languages both online and in the form of a handy smartphone application.

The 170 different editions have about 15 million pages combined, with more than three million pages on the English Wiktionary. This 3 million includes terms in 450 languages and separate pages for the inflected and conjugated word forms linking back to the main entry; it has about 379 thousand lexemes in English (250 thousand if we do not count inflected forms).

There are 27 editions with more than a 100 thousand articles, and 3 languages have more than a 1 million pages (English, French, Malagasy).

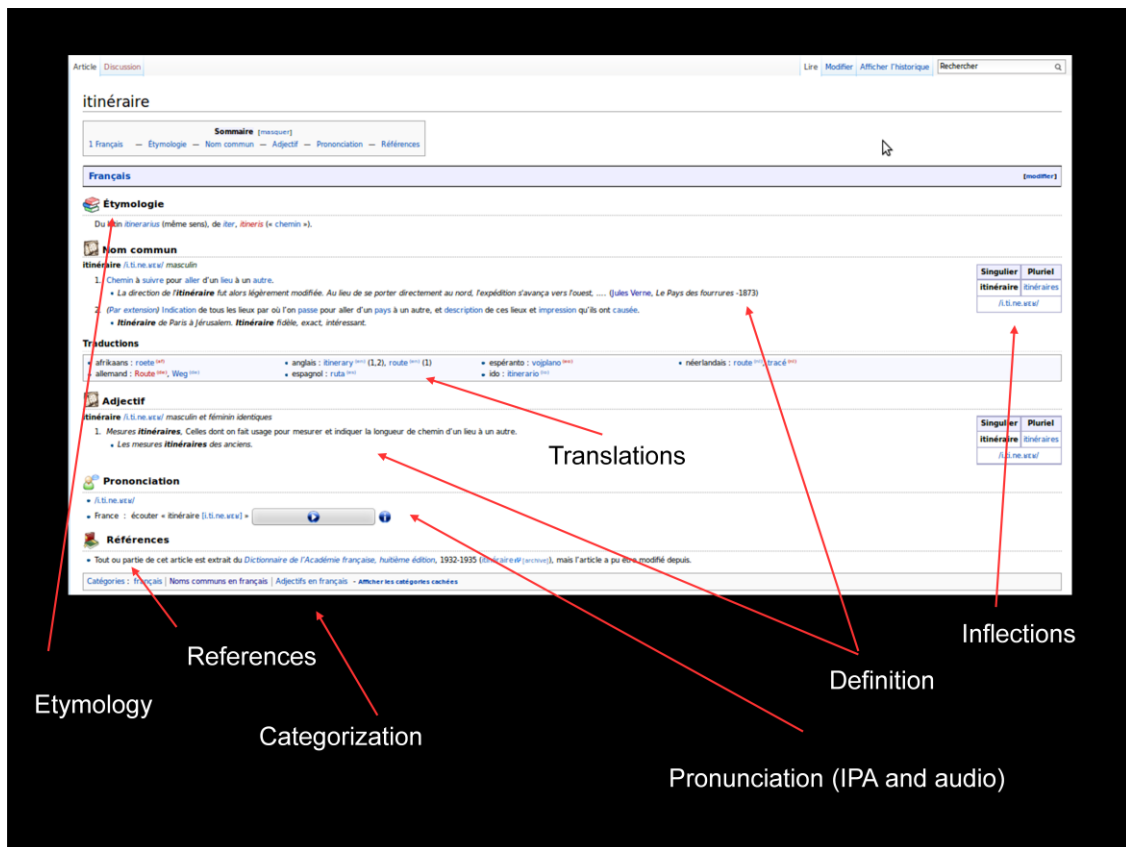
To take an example the French Wiktionary has word descriptions in more than 2000 languages, with about 320 active contributors. Smaller editions have obviously smaller communities.



Wiktionary definitions for a given word or lexeme, include parts of speech, morphology (conjugation and inflection), a definition for terms in the native language of the Wiktionary and translations for foreign language terms, pronunciations usually with the IPA, and sometimes with an audio file, etymology, and translations in other languages.

Wiktionary has grown beyond a standard dictionary and now includes a [thesaurus](#), a rhyme guide, phrase books, language statistics, paronyms (A word derived from the same root, or with the same sound, as another word.), false friends, cognates, anagrams with the aim to provide enough information to really understand a word in its full context.

An important criteria for inclusion on Wiktionary is the attestedness of a given word, and therefore entries include references, quotes and examples, and where possible illustrations, links to Wikipedia with more in depth explanations, and appendices with phrasebooks and glossaries.



To show this in practice, here is a typical French Wiktionary page for the word *itinerer* meaning itinerary.

Here you can find a definition for both the noun and the adjective in French, as you would find in a monolingual dictionary.

You can find the etymology.

There is a pronunciation section with IPA and an audio file. To the right the inflected forms of the word are derived. If one were to click on them, they would get to a similar page that explained that the given word is an inflected form of this word and a link pointing back here. This can be very useful for languages where the root form of a given word is not immediately obvious.

Towards the bottom of the page references are provided to show where the information comes from. In this case from a public domain dictionary.

Finally, the various words are collected in so called categories based on their common characteristics such as certain linguistic properties, allowing readers to find similar words and helping computers in automatically extracting relevant data.

Visualisation of collaborative work: add content, interlanguage links, layout, typography etc..

Contribution history view

Wiktionary is different from traditional dictionaries in that the history of changes is openly available to all by clicking on the history tab at the top of any page.

Here you can see how the previous article developed by adding content, translations, fixing the typography or the layout, in a collaborative manner.



## Multilinguism on Wiktionary

### Traductions

- afrikaans : [roete](#) <sup>(af)</sup>
- allemand : [Route](#) <sup>(de)</sup>, [Weg](#) <sup>(de)</sup>
- anglais : [itinerary](#) <sup>(en)</sup> (1,2), [route](#) <sup>(en)</sup> (1)
- espagnol : [ruta](#) <sup>(es)</sup>
- espéranto : [vojplano](#) <sup>(eo)</sup>
- ido : [itinerario](#) <sup>(io)</sup>
- néerlandais : [route](#) <sup>(nl)</sup>, [tracé](#) <sup>(nl)</sup>

Translations inside  
the Wiktionary

Article Discussion Lire Modifier Affili

# itinerary

**Sommaire** [masquer]

- [-] Anglais
  - Etymologie
  - Nom commun
  - Prononciation

**Anglais**

**Etymologie**

*Etymologie manquante ou incomplète. Si vous la connaissez, vous pouvez l'ajouter en cliquant ici.*

**Nom commun**

**itinerary**

- Itinéraire.

**Prononciation**

→ *Prononciation manquante.* (Ajouter)

• États-Unis : écouter « itinerary »

Catégories : anglais | Noms communs en anglais - [Afficher les catégories cachées](#)

**roete**

**Afrikaans**

**Etymologie**

*Etymologie manquante ou incomplète. Si vous la connaissez, vous pouvez l'ajouter en cliquant ici.*

**Nom commun**

**roete**

- Itinéraire.

Catégories : afrikaans | Noms communs en afrikaans - [Afficher les catégories cachées](#)

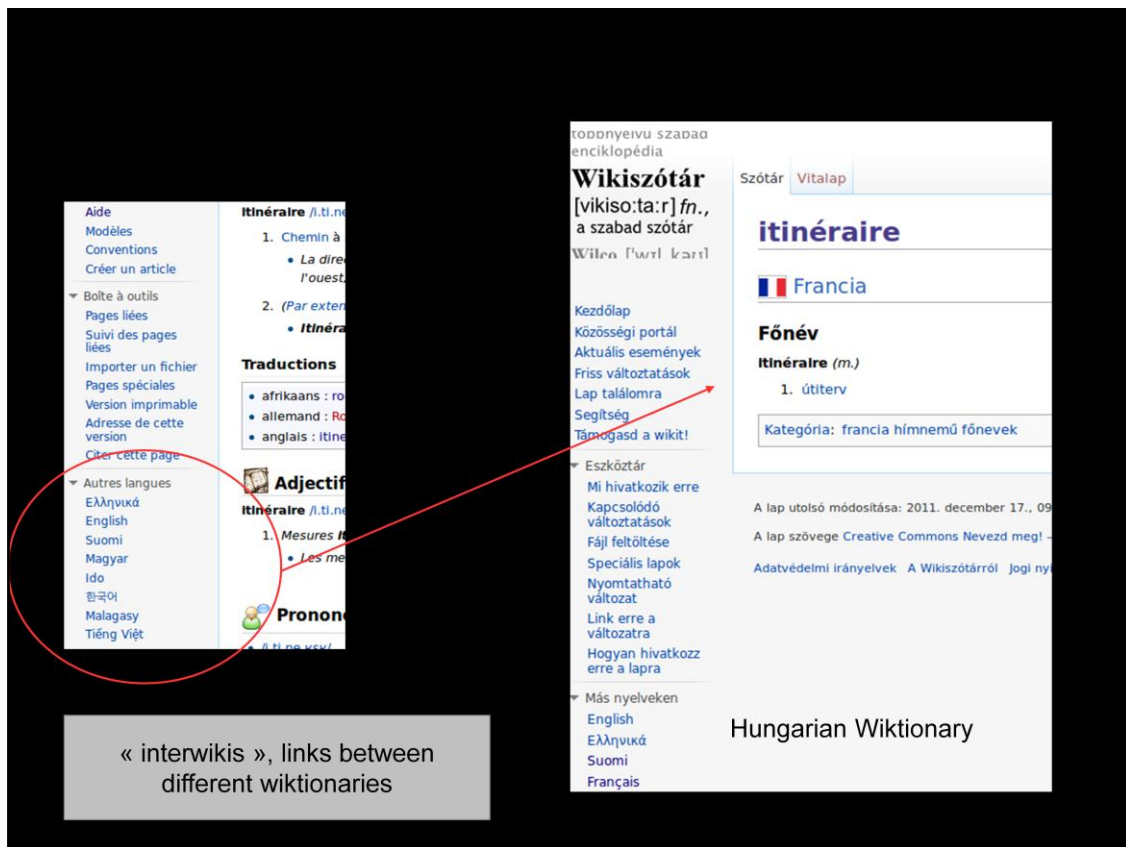
Wiktionary is inherently multilingual. Native language definitions include translations of the word into other languages, often disambiguated by the sense they refer to, and foreign language terms are also defined. Here you can see the same French example for itinerary, from where you can find pages for the same word in English and Afrikaans. As you might see, foreign terms are given a translation, and also include etymology and pronunciation. In this case, the linked articles still need a bit of improvement to add the proper IPA transcription and a better etymology.

The word « bord »  
In different  
languages

Multilingualism inside the  
Wiktionary

If a given written word form has a meaning in a number of languages, then they are collected on one page, starting with the native language variant and going in alphabetical order usually. As can be seen here with the example of the French word *bord*, the page includes the meaning of the same written word in Afrikaans, Catalan, Danish, Gaelic, Dutch, Norwegian and Swedish, including pronunciation, and a French translation.

It is very interesting to compare the different etymologies of the word in different languages.



There are Wiktionary editions in different languages, which provide the interface and the explanations and definitions in the native language of the given edition. You can see the French and Hungarian Wiktionaries side by side for the same term. If the given term exists in more than one language edition of Wiktionary, you can jump to the other version through links provided in the sidebar on the left. For example, if I looked up itinerary on the Hungarian Wiktionary, I would get a Hungarian translation only, however, if I speak French, I can quickly switch to the French version of the page with a more detailed definition and additional information.

## The importance of Wiktionary for language vitality

Wiktionary → 170 different languages versions (Wikipedia : 285)

14,883,341 articles, including + 2-2 million on French wiktionary

27 languages > 100,000 articles ; 3 languages > 1 million (English, French, Malagasy)

\* English:

3,139,000 pages  
definitions in 450 languages  
+ 1000 active contributors

\* French:

2,241,400 pages  
definitions in 2200 languages  
+ 340 active contributors

\* Malagasy:

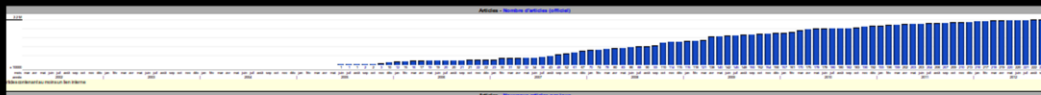
2 057 482 pages  
Definitions in 64 languages  
60 active contributors

In some contexts, a Wiktionary can be an outlet to document and preserve languages less spoken. In this context the French Wiktionary seems to be remarkable in collecting definitions of terms in 2200 languages, or the Malagasy language Wiktionary which has managed to create over 2 million pages in a short time.

## Evolution of the 3 biggest wiktionaries since 2002



English



French



Malagasy

To illustrate, here are some graphs that show the development in size of the three largest Wiktionaries. English Wiktionary has about 90 million pageviews a month and a 1000 contributors, about 400 of them fairly active and has developed quite organically.

The big jumps seen in the case of Malagasy and the French edition show the inclusion of the content older public domain dictionaries, and articles created in a semi-automatic fashion – for example through importing and translating data already available in a different language version of Wiktionary or by creating separate pages for the different inflection forms of a word.

## Enrich the Wiktionary...

### Incorporation of dictionaries in Public Domain

- \* Oxford English Dictionary (1884) – English
- \* Dai Nippon (1910) – Japanese
- \* Dictionnaire de l'Académie Française (1935) – French

### Enrichment with specific vocabularies

- \* vocabulary from different levels : distinguished, familiar etc
- \* specific vocabulary: professions, domains of knowledge (theology, chemistry)

### Projects

- \* Work with Wikisource : put online specific dictionaries
- \* Heritage preservation project : recording and documenting the vocabulary of Ceramic and porcelain professions, in partnership with the Museum of Sèvres (2012-2013)

As noted earlier, Wiktionary is enriched not only through the tireless contributions of volunteers – basically anyone with an internet connection, but also through incorporating dictionaries that have gone out of copyright – of course the content of these usually need to be updated by the editors to include changes in the meanings in the intervening time. For example, the English language corpus contains the 1884 edition of the Oxford English Dictionary, and the French edition a 1935 edition of the Dictionnaire de l'Académie Française.

Wiktionary can also be enriched by creating and incorporating special phrasebooks and thematic dictionaries, such as definitions for phrases used in theology or chemistry.

Outside partners can also help in enriching the content and making lexicological data available on the internet. For example the Museum of Sèvres /sɛvʁ/ is running a heritage preservation project where they are recording and documenting the vocabulary of the ceramic and porcelain /'pɔʁ.sə.lɛn/ professions.

## Wiktionaries in the principal languages of Europe

English (english) : 3,059,407 articles  
French : 2,218,279 articles  
Russian (Русский) : 441,708 articles  
Greek (Ελληνικά) : 403,967 articles  
Polish (Polski) :- 291,481 articles  
Swedish (Svenska) : 252,000 articles  
German (Deutsch) : 219,683 articles  
Dutch (Nederlands) : 211,423 articles  
Finnish (Suomi) : 196,646 articles  
**Hungarian (Magyar) : 184,017 articles**  
Portugese (Português) : 180,936 articles  
Norwegian (Norsk) : 128,030 articles  
Italian (Italiano) : 118,248 articles  
Estonian (Eesti) : 97,549 articles  
Spanish (Español) : 72,481 articles  
Romanian (Română) : 47,003 articles  
Czech (Česká) : 37,919 articles  
Catalan (Català) : 35,075 articles  
Ukrainian (Українська) : 33,032 articles  
Bulgarian (Български) : 26,789 articles  
Croatian (Hrvatski) : 22,551 articles  
Serbian (Српски) : 15,956 articles  
Danish (Dansk) : 7,291 articles  
Slovak (Slovenčina) : 1,462 articles

### What can you do ?

- Be closer with the contributors to organize contribution projects
- Put on the internet and under free licence reusable dictionaries
- Promote contribution on Wiktionay

Finally, I would like to finish my presentation with a number of ways you can be involved either individually and as representatives of your organisations.

You could organize language and heritage presentation projects in collaboration with the contributors of Wiktionary, like in the example of the French Museum and the ceramic industry. Or effort in Poland to document the words of a local minority language online – I believe it is the Kashubian language - in Wiktionary, and recording the pronunciations of the same.

A second way would be to participate in the online ecosystem of creating content that is freely usable, and not just in the sense of free as beer that is no cost, but in the sense of free as in speech that people can build on.

Finally, you can help by promoting contribution to Wiktionary and occasionally adding to it or correcting errors in your free time.

As you can see on the slides, most language editions of Wiktionary could benefit from additional content and editors to become useful dictionaries for the layman, and language learners.

Find the Wiktionary in **your** language

**<http://wiktionary.org/>**

For a brief introduction for linguists see:

Meyer, C. M., & Gurevych, I. (2012).

In S. Granger & M. Paquot (Eds.), *Electronic Lexicography*. Oxford: Oxford University Press

**<http://bit.ly/MeyerWiktionary>**

contact :

[adrienne.alix@wikimedia.fr](mailto:adrienne.alix@wikimedia.fr)  
[bence.damokos@wikimedia.hu](mailto:bence.damokos@wikimedia.hu)



If you would like to find out more about Wiktonary, I recommend you visit the website at Wiktonary.org. Also, Meyer and Gurevych have written a very accesible introduction to Wiktionary aimed at linguists, that is available online.