

Developing the Cornish Dictionary using open source tools and data

Davydh Trethewey

Project Support Assistant, Sodhva Kernewek - Cornish Language Office
Konsel Kernow - Cornwall Council
davidtreth@gmail.com
taklowkernewek.neocities.org

Keskusulyans Wikimedia Kolm Keltek - Wikimedia Celtic Knot
conference,
5th July 2019, Penryn Campus

Previous Online Dictionary

- Standard Written Form of Cornish online dictionary
cornishdictionary.org.uk (Internet Archive Wayback Machine)

cornish language partnership
maga cornish dictionary / gerlyver kernewek

Welcome / Dynmargh

Cornish words

[Apply](#)

English words (exact)

[Apply](#)

English words (include)

[Apply](#)

Search

[Apply](#)

Cornish / Kernewek A-Z

English / Sowsnek A-Z

Karppola / Feedback

**Abbreviations /
Berhewansow**

**'Middle' and 'Late' Cornish
forms / Fomya 'Kies' ha
Owedher**

Pronunciation / Leveryans

**Traditional graphs / Grafys
hengovek**

The team / Kesoberoryon

login

Welcome to the new online, searchable, dictionary of Cornish. The dictionary is written in the Standard Written Form of Cornish and the intention is that it will be constantly updated and extended. As well as adding new terminology and extending the range of the dictionary, over time information about each word and usage will also be made available as will audio files to allow the pronunciations to be heard.

A pdf version is available on the MAGA website and this will be updated on a 6 monthly basis to reflect additions to the online version.

Feedback on this resource is most welcome and a feedback form is included for your use.

MAGA hopes that this will be a useful addition to the wide range of resources now available to help both speakers and learners of Cornish. For further information about Cornish, resources and events, visit the website www.magaherrow.org.uk

or contact the office on cornishlanguage@cornwall.gov.uk

Dynmargh dhe'n gerlyver hwadlow warlenn nowyth a Gemewek. Serfys yw an gerlyver y'n Furm Sowsnek. Serfys a Gemewek ha'n mynas yw y lydh prest nowythys hag ydyrnyys. Keffys ha keworra terminologyeth nowyth hag ydyrnyia elander an gerlyver, dres lernyn kethlow a-dro dhe bub ger ha'y usadow a vydh kavadow, yn herwydh restrinnow sonek dhe asa bos klawys an leveryans.

Vershyon pdf yw kavadow war wasaw MAGA hag y lydh hemma nowythys pub 6 nis dhe dhasowynnyas keworansow dhe'n vershyon warlenn.

Dasly war an asodh ma a vydh meur dhyrnyys ha laven dasly yw komprehensys rag agas us.

MAGA a wayt y lydh hemma keworansow dhe les dhe'n asodhow a les kinda lernyn kavadow dhe'n gweres ha keworyon ha dyskoryon a Gemewek. Rag kethlow pella a-dro dhe Gemewek, asodhow ha hwarowow, vysyyn an wasaw www.magaherrow.org.uk

po kesdava orth an sodhva war cornishlanguage@cornwall.gov.uk

news

[SWF Review](#)

This dictionary has now been updated following the SWF Review. The PDF version will be available shortly. We would be grateful if users would report any accidental errors using the feedback form....
[read more](#)

Features in an ideal online dictionary

- Improve usability on different platforms desktop/tablet/mobile
- Cater for the various users of the language, including users of different varieties of Cornish
- Show personal forms for verbs and prepositions
- Ability to add sound samples
- Disambiguation of translation equivalents

Features in an ideal online dictionary

- Improve usability on different platforms desktop/tablet/mobile
- Cater for the various users of the language, including users of different varieties of Cornish
- Show personal forms for verbs and prepositions
- Ability to add sound samples
- Disambiguation of translation equivalents

Features in an ideal online dictionary

- Improve usability on different platforms desktop/tablet/mobile
- Cater for the various users of the language, including users of different varieties of Cornish
- Show personal forms for verbs and prepositions
- Ability to add sound samples
- Disambiguation of translation equivalents

Features in an ideal online dictionary

- Improve usability on different platforms desktop/tablet/mobile
- Cater for the various users of the language, including users of different varieties of Cornish
- Show personal forms for verbs and prepositions
- Ability to add sound samples
- Disambiguation of translation equivalents

Features in an ideal online dictionary

- Improve usability on different platforms desktop/tablet/mobile
- Cater for the various users of the language, including users of different varieties of Cornish
- Show personal forms for verbs and prepositions
- Ability to add sound samples
- Disambiguation of translation equivalents

Things to improve

- English to Cornish, and Cornish to English had been separately created in previous version
- Some errors and inconsistencies
- Updates according to 2013 review of the Standard Written Form
- Integration of work done by Terminology Panel of Akademi Kernewek
- Provide platform for further work by Akademi Kernewek

Things to improve

- English to Cornish, and Cornish to English had been separately created in previous version
- Some errors and inconsistencies
- Updates according to 2013 review of the Standard Written Form
- Integration of work done by Terminology Panel of Akademi Kernewek
- Provide platform for further work by Akademi Kernewek

Things to improve

- English to Cornish, and Cornish to English had been separately created in previous version
- Some errors and inconsistencies
- Updates according to 2013 review of the Standard Written Form
- Integration of work done by Terminology Panel of Akademi Kernewek
- Provide platform for further work by Akademi Kernewek

Things to improve

- English to Cornish, and Cornish to English had been separately created in previous version
- Some errors and inconsistencies
- Updates according to 2013 review of the Standard Written Form
- Integration of work done by Terminology Panel of Akademi Kernewek
- Provide platform for further work by Akademi Kernewek

Things to improve

- English to Cornish, and Cornish to English had been separately created in previous version
- Some errors and inconsistencies
- Updates according to 2013 review of the Standard Written Form
- Integration of work done by Terminology Panel of Akademi Kernewek
- Provide platform for further work by Akademi Kernewek

Dictionary data

- Exported from the software used in the previous version as an XML file
- Each word or phrase in the dictionary is a <lemma> tag group
- Various information in sub-tags
- e.g. pronunciation, part of speech, plural, English glosses, example sentence, etymology, attestations in the traditional texts

Dictionary data

- Exported from the software used in the previous version as an XML file
- Each word or phrase in the dictionary is a <lemma> tag group
- Various information in sub-tags
- e.g. pronunciation, part of speech, plural, English glosses, example sentence, etymology, attestations in the traditional texts

Dictionary data

- Exported from the software used in the previous version as an XML file
- Each word or phrase in the dictionary is a <lemma> tag group
- Various information in sub-tags
- e.g. pronunciation, part of speech, plural, English glosses, example sentence, etymology, attestations in the traditional texts

Dictionary data

- Exported from the software used in the previous version as an XML file
- Each word or phrase in the dictionary is a <lemma> tag group
- Various information in sub-tags
- e.g. pronunciation, part of speech, plural, English glosses, example sentence, etymology, attestations in the traditional texts

Tidying the XML

- Use Python [Beautiful Soup](#) to analyse the XML
- Allows any errors or inconsistencies to be spotted by looping through the <lemma>s in the dictionary
- Simplify some of the structure, move subentries to their own <lemma>s
- Collaboration with Dewi Bryn Jones (Bangor University) to enable it to be in a format suitable for import into Maes T

Tidying the XML

- Use Python [Beautiful Soup](#) to analyse the XML
- Allows any errors or inconsistencies to be spotted by looping through the <lemma>s in the dictionary
- Simplify some of the structure, move subentries to their own <lemma>s
- Collaboration with Dewi Bryn Jones (Bangor University) to enable it to be in a format suitable for import into Maes T

Tidying the XML

- Use Python [Beautiful Soup](#) to analyse the XML
- Allows any errors or inconsistencies to be spotted by looping through the <lemma>s in the dictionary
- Simplify some of the structure, move subentries to their own <lemma>s
- Collaboration with Dewi Bryn Jones (Bangor University) to enable it to be in a format suitable for import into Maes T

Tidying the XML

- Use Python [Beautiful Soup](#) to analyse the XML
- Allows any errors or inconsistencies to be spotted by looping through the <lemma>s in the dictionary
- Simplify some of the structure, move subentries to their own <lemma>s
- Collaboration with Dewi Bryn Jones (Bangor University) to enable it to be in a format suitable for import into Maes T

Variants within the language

- Cornish as a revived language derives from sources at different time epochs
- Broadly speaking, two time periods of Middle and Late Cornish
- Different groups within the revival have based the revived language primarily on Middle or Late sources
- Orthographic decisions have sometimes made these seem further apart than they really are

Variants within the language

- Cornish as a revived language derives from sources at different time epochs
- Broadly speaking, two time periods of Middle and Late Cornish
- Different groups within the revival have based the revived language primarily on Middle or Late sources
- Orthographic decisions have sometimes made these seem further apart than they really are

Variants within the language

- Cornish as a revived language derives from sources at different time epochs
- Broadly speaking, two time periods of Middle and Late Cornish
- Different groups within the revival have based the revived language primarily on Middle or Late sources
- Orthographic decisions have sometimes made these seem further apart than they really are

Variants within the language

- Cornish as a revived language derives from sources at different time epochs
- Broadly speaking, two time periods of Middle and Late Cornish
- Different groups within the revival have based the revived language primarily on Middle or Late sources
- Orthographic decisions have sometimes made these seem further apart than they really are

Maes T software

- Software that is used for terminology dictionaries in Welsh online at <http://termau.cymru>
- Developed by [Language Technologies Unit](#), Canolfan Bedwyr, Bangor University [1] [2]
- Deployable via an API to the web or apps [4]
- Maes T has also been used for geiriadur.bangor.ac.uk and www.termiaduraddysg.org where sound samples are also presented

Maes T software

- Software that is used for terminology dictionaries in Welsh online at <http://termau.cymru>
- Developed by [Language Technologies Unit](#), Canolfan Bedwyr, Bangor University [1] [2]
- Deployable via an API to the web or apps [4]
- Maes T has also been used for geiriadur.bangor.ac.uk and www.termiaduraddysg.org where sound samples are also presented

Maes T software

- Software that is used for terminology dictionaries in Welsh online at <http://termau.cymru>
- Developed by [Language Technologies Unit](#), Canolfan Bedwyr, Bangor University [1] [2]
- Deployable via an API to the web or apps [4]
- Maes T has also been used for geiriadur.bangor.ac.uk and www.termiaduraddysg.org where sound samples are also presented

Maes T software

- Software that is used for terminology dictionaries in Welsh online at <http://termau.cymru>
- Developed by [Language Technologies Unit](#), Canolfan Bedwyr, Bangor University [1] [2]
- Deployable via an API to the web or apps [4]
- Maes T has also been used for geiriadur.bangor.ac.uk and www.termiaduraddysg.org where sound samples are also presented

Adapting Maes T to better serve Cornish

- Collaboration with Dewi Bryn Jones to adapt the Maes T software for Cornish
- Some relevant grammatical differences between Welsh and Cornish
- Other changes come from using it for a general dictionary website rather than terminology dictionaries that usually have a 1:1 correspondence between Welsh-English in a given context

Adapting Maes T to better serve Cornish

- Collaboration with Dewi Bryn Jones to adapt the Maes T software for Cornish
- Some relevant grammatical differences between Welsh and Cornish
- Other changes come from using it for a general dictionary website rather than terminology dictionaries that usually have a 1:1 correspondence between Welsh-English in a given context

Adapting Maes T to better serve Cornish

- Collaboration with Dewi Bryn Jones to adapt the Maes T software for Cornish
- Some relevant grammatical differences between Welsh and Cornish
- Other changes come from using it for a general dictionary website rather than terminology dictionaries that usually have a 1:1 correspondence between Welsh-English in a given context

Transition to editing within Maes T

- Once the structure is stable, move away from manually editing XML to editing within Maes T
- More practical for a wider range of people e.g. Dictionary Panel members of the Akademi Kernewek to edit
- Manually editing an XML file can be error prone, which was mitigated by the Python scripts validating / analysing it

Transition to editing within Maes T

- Once the structure is stable, move away from manually editing XML to editing within Maes T
- More practical for a wider range of people e.g. Dictionary Panel members of the Akademi Kernewek to edit
- Manually editing an XML file can be error prone, which was mitigated by the Python scripts validating / analysing it

Transition to editing within Maes T

- Once the structure is stable, move away from manually editing XML to editing within Maes T
- More practical for a wider range of people e.g. Dictionary Panel members of the Akademi Kernewek to edit
- Manually editing an XML file can be error prone, which was mitigated by the Python scripts validating / analysing it

Terminology Panel

- [Akademi Kernewek](#), the Cornish language academy has a **terminology panel** to research new terms for the language
- A number of subject areas have been considered so far: plants, insects, mining, minerals, architecture, grammar

Terminology Panel


- [Akademi Kernewek](#), the Cornish language academy has a [terminology panel](#) to research new terms for the language
- A number of subject areas have been considered so far: plants, insects, mining, minerals, architecture, grammar

Using open source data from Wikimedia

sunflower (2)

entries that correspond to 'sunflower'

sunflower bleujen an howl *n.f*
LAE: Helianthus annuus
CY: blodyn y haul, heulfodyn



Terminology: Flowers
sunflower howbleujen *n.f* howbleujennow, H howbleujednow L

www.cornishdictionary.org.uk/#sunflower [3]

New dictionary website

- Demonstrate the new dictionary website (demo of cornishdictionary.org.uk)

Conclusions and future ideas

- We already have some example sentences, but could have many more of these, and audio of them by speakers
- Method of handling Middle and Late variants allows multiple variants to be supported while keeping them semantically as one <lemma> item

Conclusions and future ideas

- We already have some example sentences, but could have many more of these, and audio of them by speakers
- Method of handling Middle and Late variants allows multiple variants to be supported while keeping them semantically as one `<lemma>` item



Tegau Andrews and Gruffudd Prys. "Terminology Standardization in Education and the Construction of Resources: The Welsh Experience". In: *Education Sciences* 6.1 (2016), p. 2.



Tegau Andrews, Gruffudd Prys, and Dewi Bryn Jones. "The Maes T System and its use in the Welsh-Medium Higher Education Terminology Project". In: *Creation, Harmonization and Application of Terminology Resources* (2011), p. 49.



Gruffudd Prys et al. "Crossing between environments: the relationship between terminological dictionaries and Wikipedia". In: *Terminologie(s) et traduction*. Peter Lang, 2018. ISBN: 9783631746431.



Gruffudd Prys et al. "Distributing Terminology Resources Online: Multiple Outlet and Centralized Outlet Distribution Models in Wales". In: *The 2nd Workshop on the Creation; Harmonization and Application of Terminology Resources*. 72. Linköping University Electronic Press. 2012, pp. 37–40.