**REF2021**

| Institution: University of Wales Trinity Saint David |
|---|

| Unit of Assessment: 26 Modern Languages and Linguistics (Celtic Studies) |
|---|

| Title of case study: GPC: The authoritative dictionary of the Welsh language |
|---|

| Period when the underpinning research was undertaken: 2000–2020 |
|---|

| Details of staff conducting the underpinning research from the submitting unit: |
|---|

| Name(s): | Role(s) (e.g. job title): | Period(s) employed by submitting HEI: |
|---|---|---|
| Andrew Hawke | Managing Editor & Editor-in-Chief | 1983– |
| Dr G. Angharad Fychan | Senior Editor | 1996– |
| Prof. Ann Parry Owen | Senior Editor (0.6 FTE) | 1985– |
| Mary Williams | Assistant Editor | 2005– |
| Brenda Williams | Assistant Editor | 2008– |
| Dr Jenny Day | Assistant Editor (0.4 FTE) | 2010– |
| Sarah Down-Roberts | Assistant Editor (0.6 FTE) | 2008–2020 |
| Gareth A. Bevan | Former Editor & Co-editor | 1976–2006 |
| Dr Manon Roberts | Former Senior Editor | 1987- 2017 |
| M. T. Burdett-Jones | Former Senior Editor | 1978–2001 |
| Patrick J. Donovan | Former Co-editor | 1980–2006 |
| Marian B. Hughes | Former Assistant Editor | 1999–2001 |
| Dr Rhys Tudur | Former Assistant Editor | 1999–2001 |

| Period when the claimed impact occurred: 2014–2020 |
|---|

| Is this case study continued from a case study submitted in 2014? N |
|---|

**1. Summary of the impact**

*Geiriadur Prifysgol Cymru* (*GPC*) is a historical dictionary similar to the *Oxford English Dictionary*, and is the acknowledged authority on the spelling, derivation and meaning of Welsh words. Apart from its multifarious scholarly uses, it is influential in many areas of the Welsh public sphere, providing the lexical infrastructure of the language necessary to produce terminology for bilingual documentation in fields such as government, education, health, law, and business. Since being made freely available online in 2014, and as mobile apps in 2016, *GPC*'s impact has increased dramatically leading to the Dictionary's funding by the Welsh Government from 2016 as an essential part of its *Cymraeg 2050* strategy to increase the number of Welsh speakers to one million by 2050.

**2. Underpinning research**

The project began a century ago. A small team of staff was established at the National Library of Wales in Aberystwyth by the University of Wales in 1921 and with the help of volunteer readers some 1.75 million citation slips were amassed during the 27-year reading programme. The first part of the Dictionary was published in 1950 and the completed work was launched by the First Minister of Wales, Rhodri Morgan, at the National Assembly for Wales in 2002. Subsequently the *A–B* section was revised comprehensively to the latest standards of scholarship, more than doubling its length. The first edition contains a total of 7.3 million words of running text in 3,949 pages, forming four substantial volumes arranged under 84,000 headwords.

Underpinning research on new and updated entries has been undertaken continuously by the research team at the Centre for Advanced Welsh and Celtic Studies, and the parts of the Dictionary published since 1 January 2000 account for 30% of the entire text including the entries *streic*[3] to *Zwinglïaidd*, the second edition of *A–B*, and all the additions to *GPC Online*, amounting to 26,690 articles containing a total of 3,169,112 words of text, each article varying in length from a few dozen words to 3,888 words (*blodau*). The Dictionary is a team effort, benefiting from the skills, knowledge, and experience of the seven (latterly six) editorial staff, an IT specialist, external readers, and consultants. It is a historical dictionary in the mould of the great *Oxford English Dictionary* and like that work is entirely evidence-based. The editors' research task is to assemble as much evidence as practicable, analyse that evidence thoroughly, and present an intelligible synopsis of the findings, supported by a representative selection of citation evidence.

The essential foundation of the work is the citation collection (now numbering about 2.5 million paper slips), but this has been augmented by a far more extensive corpus of electronic texts of

all periods (currently numbering 9,576 items containing 71.34GB of data) and a 169-million-word Web corpus, which is supplemented by advanced online searching of the Web and specific Web resources (as detailed in Part 4). Each article is a distillation of extensive research into the word's orthography, its etymology or morphology, derivative forms (plurals, singulatives, diminutives, etc.), variant forms (including their morphology and etymology where relevant), collocations, and especially semantics. The definitions are based on a thorough analysis of the available evidence, garnered from *GPC*'s citation collection, an extensive library of electronic texts, online research, library research, and guidance from specialist consultants.

The selection of illustrative quotations is particularly important, supporting the semantic analysis presented in the article, showing the use of the word over its recorded history (potentially over a 1,300-year period), and illustrating the various attested forms of the word. Articles may be subdivided into multiple sections based on semantic or syntactical criteria, although the editors strive to make the work as concise as is practicable in order to improve its usability by what is now a far more diverse range of potential users since going online. Hawke's interest in the theory and practice of historical lexicography has led to a number of invitations to contribute to conferences, journals (e.g. sect. 3. ref. 6), and reference works (e.g. sect. 3. ref. 5), as well as roles on advisory panels and as a consultant which has influenced the development of other historical dictionaries, such as the new historical dictionary of Scottish Gaelic, *Faclair na Gàidhlig*. Since the online version of *GPC* (sect. 3, ref. 3) was launched in 2014, 3,071 new articles have been added to it **[REF2 26-AHAF1]**. Of these 422 were in the printed edition but have been completely re-edited. Many of the new entries are either for neologisms or compound words of which many were treated originally as cross-references in order to save space.

*GPC* has long been accepted as the authoritative work on the Welsh lexicon. It is highly influential in the academic sphere, as shown by the 1,100 references to it on Google Scholar, 321 of them from the period 2014–2020, and the 272 papers citing *GPC* on academia.edu. It is considered the standard reference for the Welsh lexicon in the scholarly world and was acknowledged by Universities UK's 2006 publication *Eureka UK!*, as one of only a handful of humanities projects out of '100 discoveries and developments in UK universities that have changed the world over the last fifty years'. This influence continues undiminished.

## 3. References to the research

1. *Geiriadur Prifysgol Cymru / A Dictionary of the Welsh Language*, first edition, Cardiff: University of Wales Press, 1950–2002, ed. R. J. Thomas (1950–1975), Gareth A. Bevan (1975–1998), Gareth A. Bevan and Patrick J. Donovan (1998–2002) [= *GPC1*] (esp. Volume IV (*s–Zwinglïaidd*), published 2002) **[submitted to REF 2014, supplied on request]**

2. *Geiriadur Prifysgol Cymru / A Dictionary of the Welsh Language*, second edition, Cardiff: University of Wales Press, 2003–, ed. Gareth A. Bevan and Patrick J. Donovan (2003–7), and Andrew Hawke (2008–present) [= *GPC2*] **[submitted to REF 2014, supplied on request]**

3. *GPC Ar Lein / GPC Online*, CAWCS, 2014–present, edited by Andrew Hawke (http://geiriadur.ac.uk/gpc/gpc.html); also encompassing Android, iOS, and Amazon Fire mobile apps, 2016–present, designed and edited by Andrew Hawke (http://www.welsh-dictionary.ac.uk/android-ios-apps/) [= *GPC3*] **[REF2 submitted (additions 2014–)]**

4. Website: https://www.geiriadur.ac.uk/ (Welsh); https://www.welsh-dictionary.ac.uk/ (English)

5. Andrew Hawke, 'Defining and Quotation Evidence in Historical Dictionaries' in Philip Durkin (ed.), *The Oxford Handbook of Lexicography*, Oxford: Oxford University Press, 2015, pp. [176]–202. Print publication date: November 2015, ISBN: 9780199691630; published online: March 2016, DOI: https://doi.org/10.1093/oxfordhb/9780199691630.013.14

6. Andrew Hawke, 'Coping with an expanding vocabulary: the lexicographical contribution to Welsh', *International Journal of Lexicography*, **31**.2 (2018), 229–48. DOI (OA as 'Editor's choice'): https://doi.org/10.1093/ijl/ecy004

2016–17: Welsh Government: £160,000
2017–18: Welsh Government: £160,000; Welsh Government supplemental grant: £70,000
2018–19: Welsh Government: £300,000; Welsh Government 'Cymraeg 2050' Innovation Grant for *GPC+*: £20,000
2019–20: Welsh Government: £300,000
2020–21: Welsh Government: £300,000          **TOTAL: £1,432,673**

## 4. Details of the impact

**Welsh Government language policy: *Cymraeg 2050***
*GPC*'s impact is concerned principally with the Welsh language which is spoken by over 562,000 people aged three and over in Wales (2011 Census) and by around 112,000 in the rest of the UK (ONS estimate 2007). The language has equal legal status to English in Wales, is a compulsory school subject from age 5–16, and Welsh provision is a legal requirement in all public services in Wales. The Welsh Government's *Cymraeg 2050: A million Welsh speakers* strategy aims to almost double the number of Welsh speakers to a million by 2050. Whilst a long-term strategy for the period 2017–50, with evolving work programmes, the *Cymraeg 2050* annual report for 2019–20 **(Source 4)** reports that the emerging trend is that numbers seem to be increasing gradually based on recent evidence from the Annual Population Survey: Welsh Language.

As the only standard historical dictionary of the Welsh language, *GPC* is the definitive record of Welsh vocabulary, orthography, and the gender of nouns, and therefore forms the basis for other dictionaries, thesauri, terminology lists and reference works, as well as being a very widely consulted reference for general use. Welsh Government has invested in *GPC* in recognition of its critical role for the future of Welsh language linguistic infrastructure, and how this is contributing to achieving the aims of government policy: notably *Taking Wales Forward 2016–2021,* the *Well-being of Future Generations Act (Wales) 2015* (indicators 36: Percentage of people who speak Welsh daily, and 37: Percentage of people who can speak Welsh), the *Welsh language technology action plan* 2018 (support terminology, lexicography and corpora resources and other elements of Welsh language infrastructure) and *Cymraeg 2050*. The latter sets out the Welsh Government's long-term approach to achieving the target of a million Welsh speakers by 2050 and identifies three strategic themes to achieve this vision: **1)** increasing the number of Welsh speakers, **2)** increasing the use of Welsh, and **3)** creating favourable conditions through the provision of high-quality infrastructure and context. While *Cymraeg 2050* recognizes the multiple interventions required in meeting its targets (most notably through formal education) *GPC* has made a distinct contribution to the strategy in two interdependent areas:
*   **Linguistic infrastructure:** ensure the continued development of Welsh language infrastructure (dictionaries, terminology, the translation profession) as integral to the delivery of this strategy.
*   **Digital technology**: ensure that the Welsh language is at the heart of innovation in digital technology to enable the use of Welsh in all digital contexts.

**Linguistic infrastructure**
Long-term investment in the linguistic infrastructure has been provided by Welsh Government to fund the digital development and continued evolution of *GPC* so that it is the authoritative national lexicographical resource. As noted above, these infrastructure costs are currently funded at £300k annually through a Welsh Government grant (£1.35m since 2015), covering most of the costs of the endeavour so that it generates further positive outcomes for the language (e.g. increasing visibility and enabling the production of high-quality resources for learners and fluent speakers). A ministerial press release (dated 18 July 2018) on Eluned Morgan's (Minister for Mental Health, Wellbeing and the Welsh Language) visit to the *GPC* team at CAWCS quotes the Minister as saying, 'As one of the most important works in the Welsh language, it is very pleasing to see that Welsh Government funding is playing a vital role in its future and therefore the future of the language.' The 2018–19 and 2019–20 annual reports on the *Cymraeg 2050* strategy likewise confirm funding for the next funding period and note that *GPC* has become a linguistic cornerstone of the Welsh language and contributes to many of the

strategic goals, including creating favourable conditions for the language, digital technology and promoting the use of the Welsh language **(Sources 4, 5).** Similarly, a written Cabinet Statement by Eluned Morgan on the future of Welsh language linguistic infrastructure (dated 5 Dec 2018) reaffirmed the funding, considering the Dictionary to be a 'vital resource, comparable to the *Oxford English Dictionary*' **(Source 6)**.

**Language standardization**
*GPC* is the foundation for all contemporary Welsh lexicography, as acknowledged, for instance, by the editor of *Geiriadur Cymraeg Gomer* [Gomer Welsh Dictionary] (Gwasg Gomer, 2016) which was sponsored by the Welsh Joint Education Committee for use as the standard Welsh–English dictionary in all secondary schools in Wales **(Source 1)**. According to StatsWales the total number of secondary schools in Wales in 2019 was 187 with 188,475 pupils. The Welsh medium HE/FE institution Coleg Cymraeg Cenedlaethol encourages its staff and students to use *GPC*, with a page dedicated to it in its online resources library, Y Porth, and the College has commissioned Hawke to produce training videos.

The ongoing revision of the *Oxford English Dictionary* likewise draws on *GPC,* and Hawke's advice as a language consultant has been sought regularly during the drafting (or revision) of the etymologies of 54 *OED* entries since 2014 **(Source 2)**. Two members of the current staff, Fychan and Parry Owen, and the former Editor, Bevan, represent *GPC* on the Welsh Language Commissioner's Place-name Standardization Panel (constituting half of the panel), advising local government and the Welsh Government on the standard spelling of Welsh place-names, based on *GPC*'s evidence for the elements contained therein and the editors' interpretation of the standard orthographical guidelines **(Source 9)**. (*GPC* itself does not include place-names although it does use them as evidence for the existence of certain lexical items.)

**Bilingualism, legislative drafting, and translation services**
*GPC* is also one of the primary linguistic resources used in the drafting of legislation in Welsh, which has the same status in law in Wales as English-language legislation and therefore must be used with great semantic precision. The Welsh Government Translation Service's Legislative Style Guide which 'offers advice and guidance on the legislative translators' house style to those who translate legislative materials on behalf of the Welsh Government' makes frequent reference to *GPC* in regard to preferred spellings and the gender and plural forms of nouns **(Source 7)**. The two teams of translators who work for Senedd Cymru (the Welsh Parliament) and the Welsh Government also use *GPC*, as stipulated, for example, in *Yr Arddulliadur*, a 'Welsh-language style guide … developed to give advice and guidance to translators of general Welsh Government texts' which recommends the use of *GPC* for determining the preferred gender and plural form of nouns **(Source 3)**. Hawke addressed Welsh Government translators from across Wales at a training session in 2019 and is consulted regularly on lexical matters. The *GPC* staff have in this regard re-examined some entries where Welsh Government practice is at variance with *GPC*, and a number of entries have been revised in the light of more recent evidence (e.g. gender changes to *tudalen*, *proses*, *fersiwn*). Professional commercial translators, who are mostly either freelance translators or work for small companies, also make extensive use of *GPC* in their daily work **(Source 8)**.

**Digital technology and public language use**
The *Cymraeg 2050* strategy recognizes that while education is central to the vision for the growth of Welsh language use, it is vital that 'young people come out of the education system ready and proud to use the language in all contexts', and that 'digital technologies will be central to the vision within education, helping workplaces become bilingual and supporting social use'. *GPC* in this regard pioneered the use of microcomputers in Welsh lexicography in 1983 with much software written in-house and the entire text had been digitized by 2002. A new computer-based editing system, iLex, was introduced in 2010 and the text converted to XML in preparation for an online version, which was launched within the current REF cycle, by the First Minister, Carwyn Jones at the Senedd in Cardiff in 2014, receiving extensive media attention (sect. 3, ref. 4). Subsequently, in 2015/16, the Welsh Government funded the development of app versions of *GPC*, which are available at no cost on the Apple iOS, Google Android, and Amazon Fire mobile platforms. *GPC* is, to the best of our knowledge, the only full historical dictionary for any

language that is available as an app, including as it does the entire nine million word text. The apps allow for full or partial download of all entries for off-line access, and include 124,500 headwords, 7,150 variant forms, 34,000 collocations, 475,000 illustrative quotations with source references, over 60,000 cross-references, and are searchable by word or phrase in both Welsh and English. Two language learning games are also included for pedagogical purposes.

The use of *GPC* and its non-academic impact has increased enormously following its availability online since 2014, further establishing its position as a cornerstone of the linguistic infrastructure. The availability of *GPC* on the Web and in app form, freely available to all, has made the research accessible to a vastly greater number of users. The search interface has been designed to facilitate access by searching for all mutated forms of a word, suggesting partial matches, and suggesting similar words if the search term cannot be found, making it more accessible to those less familiar with the language. Pop-ups explain both grammatical and bibliographic abbreviations to aid rapid understanding of the text.

There were fewer than 400 (mostly institutional) subscriptions to the final printed parts of the Dictionary (2013), whereas over the period June 2014–2020, 14,419,515 entries were consulted online (excluding much additional use on mobile devices which cannot be measured if users have downloaded the data to their devices). The number of entries consulted has risen dramatically: **2014:** 216,418; **2015:** 2,088,367; **2016:** 1,812,421; **2017:** 2,226,353; **2018:** 2,402,887; **2019:** 2,436,526; with a significant increase in **2020** to 3,236,543 **(Source 10)**.

Mobile usage has increased even more rapidly, from 303,239 uses in the launch year (Feb. 2016) to 1,263,384 in 2020. The iOS app has been downloaded 12,123 times since 2016, and the Android app is currently installed on 2,984 active devices in 51 countries **(Source 11)**. Although mobile usage has increased steadily as a percentage of total use, PC usage still accounts for over 79% of all transactions, reflecting heavy use in education and by professional translators. Google Analytics data (2016–present) show 78,324 unique users from 130 countries visiting the Welsh-language website and 130,996 unique users from 175 countries visiting the English-language website, indicating worldwide interest in the project as does the fact that *GPC*'s 'Word of the Day' is followed by 4,605 people on Twitter and 2,310 on Facebook.

Hawke was one of two UK members of the EU-funded COST Action ENeL (European Network of e-Lexicography, 2013–17) which led to the establishment of the European Dictionary Portal. *GPC* staff were responsible for the Welsh-language interface of the portal and for providing access to the major Welsh online dictionaries. CAWCS/*GPC* is also an observer and data provider on the Horizon 2020 ELEXIS (European Lexicographic Infrastructure) Project (2018–), which is pioneering the development of digital lexicography globally.

## 5. Sources to corroborate the impact

1. Editor, *Geiriadur Cymraeg Gomer*, Llandysul 2016.
2. Deputy Chief Editor, *Oxford English Dictionary*.
3. Chief Legal Translator in the Office of the Counsel General of the Welsh Government.
4. Chief Executive of Cymdeithas Cyfieithwyr Cymru (the association of Welsh translators and interpreters), the national body which leads, develops, and promotes the profession in Wales.
5. Senior Infrastructure and Research Officer, Office of the Welsh Language Commissioner.
6. *Yr Arddulliadur* (Word version, 06/01/2020): https://gov.wales/bydtermcymru/yr-arddulliadur/word-version-yr-arddulliadur, see under 'cenedl enwau', 'lluosog', 'orgraff'.
7. https://gov.wales/welsh-language-strategy-annual-report-2018-2019.
8. https://gov.wales/welsh-language-strategy-annual-report-2019-2020.
9. https://gov.wales/written-statement-future-welsh-language-linguistic-infrastructure-development-and-coordination.
10. Internal daily logfiles for 2014–20 generated by *GPC Online* (supplied on request).
11. App developer consoles (both accessed 26/1/21): Apple App Store Connect (https://appstoreconnect.apple.com/analytics/ and Google Developer Console https://console.developers.google.com/ – access can be provided to both on request).