Official terminology in the internet age: prescription, dissemination, uptake

Merryn Davies-Deacon, Queen's University Belfast

This study considers the dissemination and uptake official terminology in three European languages:

- French: widely used and official in many contexts across the world; corpus planning in France dates back several centuries to the publications of early modern grammarians and the foundation of the *Académie française* in 1635. A perceived threat from English as the dominant language of western society, noted in public discourse in France since at least the mid-20th century (e.g. Etiemble 1964), and now especially relevant in online contexts. Legislation e.g. the *loi Toubon* has sought to protect the status of French in France.
- Breton: a minoritised language of France with around 110,000 speakers (Région Bretagne 2025;20) – this population is in rapid decline. While national language policy protecting the minoritised languages of France is largely absent, Breton was recognised as a language of Brittany by the regional government in 2004 and has since benefited from financial support and the establishment of an official language planning body. A growing number of state schools offer bilingual French/Breton education, alongside an independent network of immersion schools that teach mostly through Breton and encourage a Breton-speaking ethos.
- Cornish: a revived minoritised language of the UK with numbers of fully competent speakers in the hundreds (Humberstone and Broadhurst 2023). Recognised by the UK under the European Charter for Regional or Minority Languages, but no official implementation in education. Mostly learnt as a second language by adult new speakers, and traditionally reliant on volunteer labour for teaching and promotion, although efforts at officialisation since 2000 have led to a standardised orthography, a language office within Cornwall Council, display on public signage, and taster sessions in primary schools.

For each of the three languages, a body develops and prescribes terminology, with varying levels of officiality; all make use of the internet to disseminate their prescriptions.

For French, the *Commission d'enrichissement de la langue française* within the French government works with subcommittees and other bodies including the *Académie française* to disseminate official vocabulary. This is published in the French government's *Journal officiel* and made available to the public through its website, *FranceTerme*, which provides a searchable database, documents with thematic collections of terms, and the full list of terms in XML format. Data are available under an open licence. https://www.culture.fr/franceterme

For Breton, the local government-funded language office, *Ofis publik ar brezhoneg*, is responsible for all aspects of language planning. The TermBret division within the *Ofis* is responsible for corpus planning and works with panels of Breton speakers and professionals in specific domains to develop terminology. It provides a searchable (but not downloadable) list of official terms on its website. https://www.brezhoneg.bzh/87-termofis.htm

For Cornish, *Akademi Kernewek*, a charitable organisation staffed by volunteers, maintains a dictionary in the Standard Written Form of the language. This is searchable online and as a mobile app, and downloadable in PDF format. Unlike the French and Breton examples, which focus on specialist vocabulary (often scientific or technical, or related to public administration), the Cornish dictionary contains general terms. https://cornishdictionary.org.uk


Crowdsourcing

Preliminary research questions for this part of the study: How do bodies prescribing terminology make use of crowdsourcing through their online platforms? To what extent is this successful?

On *FranceTerme*, the *Boîte à idées* allows members of the public to contribute suggestions. This is framed as a way of asking for foreign terms for which the *Commission* can develop French equivalents: users are asked "*Vous avez besoin d'un équivalent français ?*" In 2023, 88 out of 300 new terms published were brought to the *Commission's* attention through this mechanism (DGLFLF 2024:1). The *Commission's* report gives a sample of the terms proposed, with five out of 35 given in French (DGLFLF 2024:23), implying that in these cases, the French suggestion came directly from the user. Regardless of whether users submit French neologisms or terms from other languages, it is clear that this is an important mechanism for monitoring.

Ofis publik ar brezhoneg uses crowdsourcing in the form of a "terminology forum", which gives one or two French words per month and invites users to suggest and then vote on Breton equivalents. Between September 2013 and October 2024, 240 terms were discussed. In 146 of these cases, the winning term has been added to the official database.
Questions that remain to be investigated:
- What sorts of lexical fields tend to be involved in this exercise?
- What kinds of neologism appear among the suggestions, and which tend to attract more votes?
- Is there a correlation between the number of votes each winning term received and whether it was later added to the database?

The *Akademi Kernewek* website, serving the smallest language community, includes a general contact form, but there is no encouragement to contribute to the development of terminology.
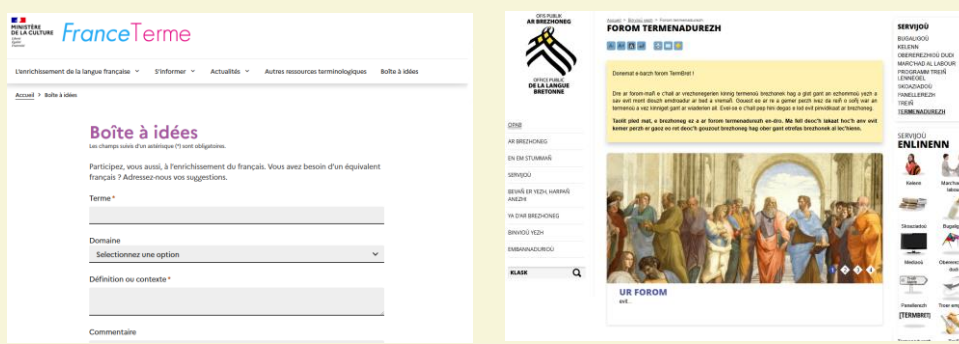
Preliminary findings: Discourse on *OPAB's* terminology forum

Ofis publik ar brezhoneg has provided me with an anonymised dataset containing comments posted on their terminology forum. I am currently annotating these comments as part of a process of thematic analysis.
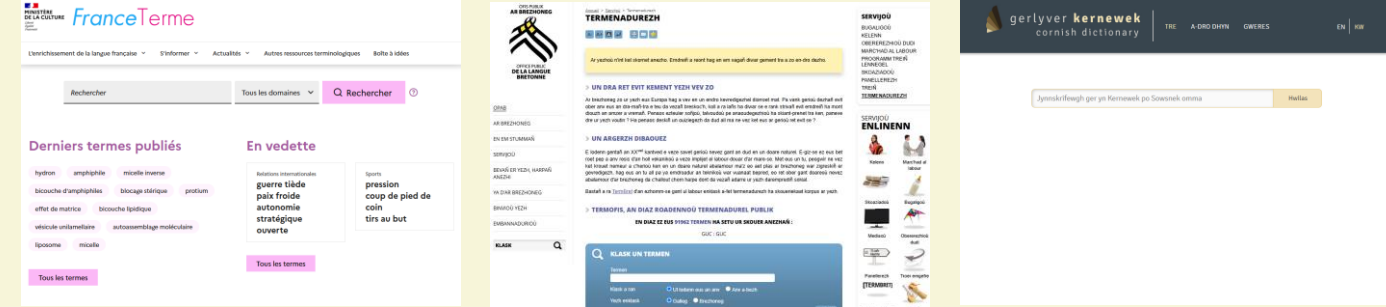
Emerging themes include the following:
- Citing particular sources: dictionaries, online databases of lexis, usage from the participant's own community as well as their own instinct. Is there a sense that some of these sources are considered more authoritative than others?
- Reference to the terms in other languages – sometimes to those closely related to Breton, implying new terms should be consistent with those in related languages. When other languages are mentioned, this may be a way of demonstrating the participant's linguistic skill, legitimating their participation in the task of neology.
- Discussion of Breton's relationship with French as the dominant language: some users note that the influence on French on Breton is inescapable and that all Breton speakers also use French in their daily lives, while others argue that certain proposed terms are unduly close to their French equivalents.
- Certain vocabulary being used to characterise the proposed terms: *aes* (easy), *divalav* (ugly), *resis* (precise), *sklaer* (clear), *iskis* (strange), *poellek* (logical), *chimik* (artificial), *pounner* (heavy), *digoulzet* (old-fashioned) …

Investigating uptake of prescribed terms: Wikipedia and Wiktionary

Preliminary research questions for this part of the study: Do articles on Wikipedia/Wiktionary use officially prescribed terminology? Is the use of terminology openly debated, and what structures enable this? How can we characterise the language policy of Wikipedia/Wiktionary?

Wikipedia and Wiktionary offer contexts for easily exploring the uptake of prescriptions, with freely accessible data including edit histories and "talk pages" where users debate the form and content of articles. Nonetheless, these platforms will be biased and potentially unreflective of reality:
- Contributors tend to belong to dominant sectors of society (male, middle-class, etc.)
- Increasing amounts of content are generated by large language models (perhaps less of a problem for minoritised languages, for which accurate models are less widely available)
- In the case of Breton, those who contribute to Wikipedia are more likely to be new speakers, and may bring a specific perspective on the type of lexis that should be used

Investigating these platforms will nonetheless provide insight into a specific context for language use online, and identify variance among the three languages on platforms that share a common infrastructure but can employ divergent administration practices for each language (Enyedy and Tkacz 2011). Ayers (2020:91) suggests that policy and governance on Wikipedia tend to follow the principle that "the people who show up make the rules" – investigating article histories and discourse around terminology will illustrate whether this holds in the case of language policy.

Previous studies have noticed debates over lexis on Wikipedia, e.g. conflict over how to name the city of Gdańsk (Jemielniak 2014), and have highlighted the impossibility of following Wikipedia's principle of the "neutral point of view" (Lovink and Tkacz 2011). Much research on these platforms concentrates on the English Wikipedia. Baxter (2009) investigates orthographic practices and conflict resolution on the Breton Wikipedia, noting room for linguistic variation, but it is unclear whether this is still the case today, as the form of the platform has evolved.

Methodologies: terms on Wikipedia/Wiktionary

To enable broad-level quantitative analysis of terms and more fine-grained investigation of the discourses present, I have written a Python script to do the following:
- Select a sample of official terms from each language (currently underway with French given the ease of accessing data)
- Query Wikipedia API to discover whether articles exist
- Check whether articles redirect
- Indicate whether there may be discussion of terminology on talk pages

Example initial output:
```
Checking 1 of 50: conservatif
Checking 2 of 50: mutation ponctuelle
Checking 3 of 50: rampe
Checking 4 of 50: espace extra-atmosphérique
Checking 5 of 50: jeu retro […]

29 terms were found: mutation ponctuelle, rampe, espace extra-atmosphérique, notoriété, complexe, ludification, Casablanca, DCC, guerre dissymétrique, disquette, Portugaise, pouvoir d'achat, enjambement, silhouette, Guinéen, Nappyidaw, sourdine, bombe […]
21 terms were not found: conservatif, jeu rétro, la République du Zimbabwé, trichite, commission de direction, refroidissement naturel, baril livré, franco le long du navire […]
```

Emerging themes include:
- Use of talk pages to attempt to find consensus (but discussions not always engaged with)
- Appeal to external authorities including the *Académie française*, despite its general irrelevance to ordinary speakers (Estival and Pennycook 2011)
- Possible assumption that the French of France is the default, despite Wikipedia being organised by language, not by nationality
- Attention to detail and concerns around consistency – difficult to achieve given the heteroglossic nature of the platform

References

Ayers, P. (2020). Wikipedia and libraries. In J. Reagle and J. Koerner (eds.), *Wikipedia @ 20: Stories of an incomplete revolution*. The MIT Press, pp. 89–106.

Baxter, R. N. (2009). New technologies and terminological pressure in lesser-used languages: The Breton Wikipedia, from terminology consumer to potential terminology provider. *Language Problems and Language Planning* 33(1), 70–80.

[DGLFLF] Délégation générale à la langue française et aux langues de France (2024). *Rapport annuel 2023 de la Commission d'enrichissement de la langue française*. Ministère de la culture.

Enyedy, E. and Tkacz, N. (2011). "Good luck with your wikiPAIDia": Reflections on the 2002 fork of the Spanish Wikipedia. In G. Lovink and N. Tkacz (eds.), *Critical point of view: A Wikipedia reader*. Institute of network cultures, pp. 110–118.

Estival, D. and Pennycook, A. (2011). The Académie française and anglophone language ideologies. *Language Policy* 10, 325–341.

Etiemble, R. (1964) *Parlez-vous franglais ?* Gallimard.

Humberstone, K. and Broadhurst, K. (2023). Cornish. *Linguistic Minorities in Europe Online*, DOI: 10.1515/lme.22556693.

Lovink, G. and Tkacz, N. (2011). The "C" in CPOV: Introduction to the CPOV reader. In G. Lovink and N. Tkacz (eds.), *Critical point of view: A Wikipedia reader*. Institute of network cultures, pp. 9–13.

Région Bretagne (2025). *Enquête sociolinguistique 2024 sur les langues de Bretagne : présentation des principaux résultats*. https://www.bretagne.bzh/app/uploads/8460-REGION-BRETAGNE-Enquete-sociolinguistique-R20.01-version-longue.pdf