



OFFICE PUBLIC DE LA LANGUE BRETONNE

# LA LANGUE BRETONNE A L'ERE DU NUMERIQUE

Diagnostic et stratégie de  
développement



Juillet 2025

Langue et innovation numérique

# Table des matières

❖ Introduction.....	5
❖ <b>Partie 1. La Langue bretonne dans notre environnement numérique, état des lieux .....</b>	<b>8</b>
<b>1 Utiliser ses appareils numériques en langue bretonne .....</b>	<b>8</b>
1.1 Sur son ordinateur .....	8
1.2 Sur son smartphone .....	10
1.3 Les jeux vidéo en breton .....	11
1.4 Les paramètres régionaux en langue bretonne .....	12
<b>2 Surfer sur Internet.....</b>	<b>13</b>
2.1 Les moteurs de recherche.....	13
2.2 Les réseaux sociaux.....	14
2.3 Les contenus en langue bretonne.....	15
<b>3 Apprendre la langue à l'aide d'outils numériques .....</b>	<b>18</b>
❖ <b>Partie 2. Inventaire des ressources linguistiques et des outils numériques en langue bretonne .....</b>	<b>21</b>
<b>1 Les ressources linguistiques .....</b>	<b>23</b>
1.1 Ressources lexicales.....	23
1.2 Corpus .....	30
1.2.1 <i>Corpus textuels monolingues</i> .....	32
1.2.2 <i>Corpus parallèles bilingues</i> .....	34
1.2.3 <i>Corpus oraux</i> .....	35
1.3 Grammaire et modèle de langue .....	37
1.4 Ressources sémantiques et bases de connaissances .....	38
<b>2 Technologies de la langue .....</b>	<b>39</b>
2.1 Analyse syntaxique, outils de base.....	39
2.2 Analyse syntaxique .....	40
2.3 Analyse sémantique .....	41
2.4 Extraction d'informations.....	41
2.5 Génération de textes.....	41
2.6 Reconnaissance de l'écriture.....	42

2.7	Traduction automatique.....	43
2.7.1	<i>Traducteur automatique avancé (statistique, neuronal, hybride)</i> .....	43
2.7.2	<i>Traducteurs automatiques de base</i> .....	44
2.8	Traitement de la parole.....	45
2.8.1	<i>Synthèse vocale</i> .....	45
2.8.2	<i>Reconnaissance vocale</i> .....	45
2.9	Autres outils.....	45
<b>3</b>	<b>Conclusion.....</b>	<b>46</b>
<b>❖</b>	<b>Partie 3. Préconisations.....</b>	<b>47</b>
<b>1</b>	<b>Augmenter sensiblement la visibilité du breton dans le monde numérique.....</b>	<b>47</b>
1.1	Les contenus en langue bretonne sur Internet.....	47
1.2	L'utilisation du breton sur les réseaux sociaux.....	47
1.3	Des outils pour l'apprentissage et la mise en valeur du breton.....	48
1.4	La traduction d'applications.....	49
<b>2</b>	<b>Encourager l'innovation numérique au service de la langue bretonne.....</b>	<b>49</b>
2.1	Mettre en réseau les différents acteurs.....	49
2.2	Publier des données de qualité.....	50
2.3	Financer l'innovation en breton.....	51
2.4	Promouvoir les licences libres de droit.....	51
<b>❖</b>	<b>Partie 4. Plan d'action.....</b>	<b>53</b>
<b>1</b>	<b>Ressources linguistiques.....</b>	<b>53</b>
1.1	Corpus.....	54
1.2	Grammaires et modèles de langue.....	54
1.3	Dictionnaires et ressources lexicales.....	55
<b>2</b>	<b>Technologies de la langue.....</b>	<b>56</b>
2.1	Outils d'analyse.....	56
2.2	Traduction automatique.....	57
2.3	Traitement de la parole.....	57
<b>3</b>	<b>Mise en œuvre du plan d'action.....</b>	<b>58</b>
3.1	Ressources matérielles, humaines et financières.....	58

3.2 Acteurs concernés..... 61

## ❖ Introduction

---

Depuis l'apparition de l'outil informatique dans nos vies, la langue bretonne a su s'adapter. D'un point de vue terminologique tout d'abord, dès les années 1980, pour pouvoir parler, jouer, échanger, travailler en langue bretonne sans difficulté. De très nombreux néologismes ont ainsi vu le jour au fil des années et sont rentrés aujourd'hui dans le langage commun : *urzhiataer*, *meziant*, *pladenn galet*, *memor*, *logodenn*, *klikañ*, *postel*, *enrollañ*, *restr*, *Kenrouedad*...

Dans un deuxième temps, c'est la traduction qui a occupé la communauté brittophone afin de proposer des outils aux utilisateurs. De nombreux logiciels libres ont ainsi été traduits en breton et sont mis à jour régulièrement depuis une vingtaine d'années.

La présence de la langue sur Internet afin de proposer de l'information en ligne au sens large, a également fait l'objet d'un important travail, comme le montre le projet d'encyclopédie libre Wikipedia en langue bretonne lancé en 2004.

Enfin, des outils internes spécifiques ont été développés : un traducteur automatique breton-français, des dictionnaires et bases de données linguistiques, des correcteurs orthographiques et grammaticaux.

La plupart de ces avancées ont été le fait d'initiatives individuelles et bénévoles qu'il convient ici de saluer particulièrement. Cela montre que la communauté est vivante et s'adapte en continu aux challenges que représentent les évolutions technologiques incessantes et de plus en plus rapides dans nos vies.

Cependant, force est de noter que ces évolutions positives se sont faites sans coordination et ne peuvent toucher nécessairement qu'un secteur marginal du numérique (le domaine du libre). L'investissement personnel et bénévole, bien que nécessaire, ne sera jamais suffisant à lui seul pour combler les manques importants qui subsistent et lever les obstacles à l'utilisation naturelle de la langue.

C'est pourquoi le conseil d'administration de l'OPLB avait décidé la création d'un poste d'agent dédié à ce domaine en 2019. Le premier travail accompli a été de proposer une stratégie adaptée et progressive à l'ensemble des acteurs, qu'ils soient institutionnels, professionnels ou simplement geek. La première version de ce rapport était le fruit de cette réflexion.

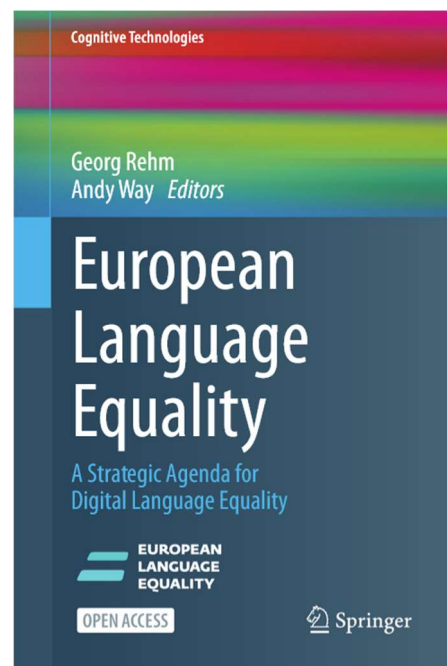
Le Conseil régional a également commencé à fixer des objectifs : le développement de la langue bretonne dans le monde numérique est inscrit dans le *Plan de réappropriation des langues de Bretagne*, proposé par la Région Bretagne en 2023.

Deux dispositifs d'aide *Stlenn* et *Arload* ont été créés pour soutenir des projets numériques.

Une actualisation du rapport 2020 est présentée ici. En moins de 5 ans, il a été possible de proposer de nouveaux outils basés sur des technologies plus récentes, comme celui des traducteurs automatiques neuronaux. Dans la première partie, il sera présenté un état des lieux actualisé de la langue bretonne dans le monde numérique et des nouvelles technologies. Elle aidera à déterminer les besoins futurs pour développer la présence de la langue, son utilisation et sa capacité à s'adapter aux outils numériques dans les prochaines années. L'objectif que nous nous fixons est de continuer à faire de la langue bretonne une langue moderne, parfaitement intégrée dans l'environnement numérique et adaptée aux usages de ses locuteurs. Cet objectif est également une réponse aux nombreuses sollicitations de la part des brittophones de pouvoir utiliser leurs outils numériques en breton dans leur vie quotidienne.

Nous espérons aussi que ce rapport servira à poursuivre l'élan donné au breton dans le monde numérique. Rappelons que la place des langues dans le monde numérique demeure très déséquilibrée. Selon András Kornai<sup>1</sup>, seul 5 % des 7 000 langues existant dans le monde, soit environ 300 à 400 bénéficieraient d'une présence sur Internet. L'anglais étant évidemment la langue la mieux représentée en termes de ressources et d'outils numériques.

En 2022/2023, un groupe de chercheurs avait rédigé un rapport<sup>2</sup> évaluant l'offre d'outils numériques et de données linguistiques de 33 langues officielles ou co-officielles européennes. Seules 4 langues sont répertoriées comme bien ou assez bien représentées : l'anglais, l'allemand, l'espagnol et le français. Toutes les autres langues, bien qu'ayant pro-



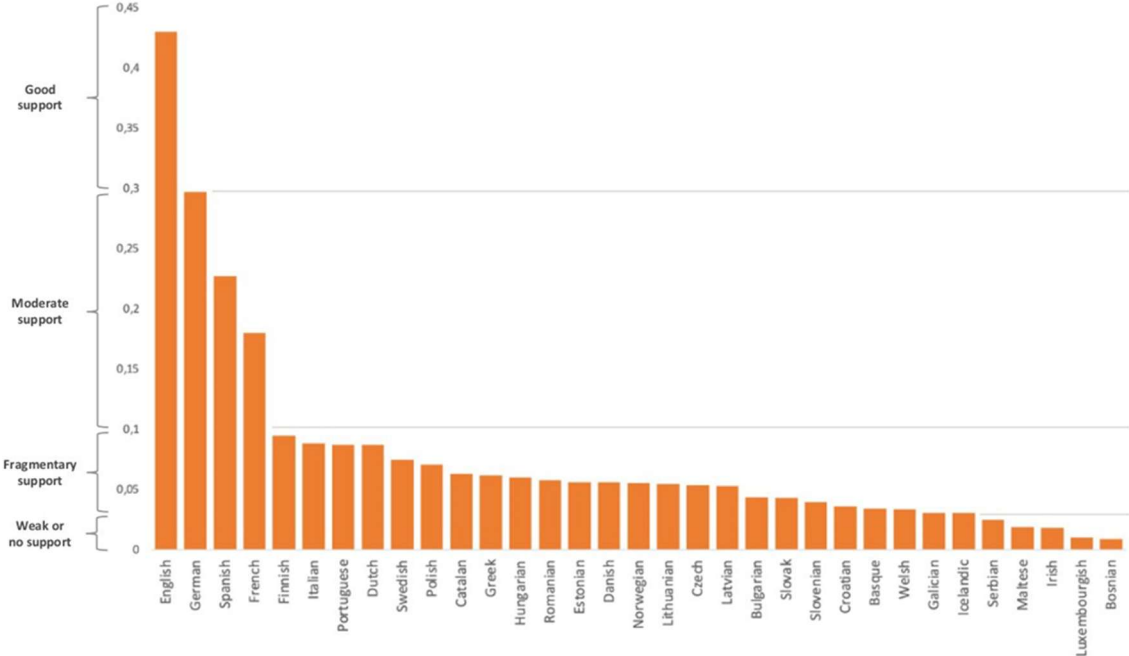
---

<sup>1</sup> Linguiste-mathématicien hongrois : "Only 5% of the world's languages are on the Internet, and the remaining 95% are condemned to death."

[https://en.wikipedia.org/wiki/Andr%C3%A1s\\_Kornai](https://en.wikipedia.org/wiki/Andr%C3%A1s_Kornai)

<sup>2</sup> European Language Equality - A Strategic Agenda for Digital - Language Equality, Giagkou et al. In Rehm & Way, 2023.

gressé dans le domaine du numérique, restent loin du niveau des principales langues. La réflexion sur la nécessité d'intégrer la diversité et les particularités des langues dans les projets de technologies linguistiques est de plus en plus prégnante. Il est à noter que les nouvelles technologies ne sont plus développées seulement pour le bénéfice d'une seule langue, mais qu'elles intègrent une approche multilingue. La langue bretonne devrait tirer profit de ces collaborations et nouvelles techniques. Si le breton figure parmi les langues bénéficiant d'outils et de ressources numériques, leur développement est crucial afin de répondre aux besoins de la Bretagne du XXI<sup>e</sup> et être une langue sociale dans le domaine numérique.



Sarasola, K., Aldabe, I., Díaz de Ilarraza, A., Estarrona, A., Farwell, A., Hernaez, I., eta Navas, E. (2022). *Report on Basque language*. Deliverable 1.4 of the Project ELE (European Language Equality).

## ❖ **Partie 1. La Langue bretonne dans notre environnement numérique, état des lieux**

---

L'un des enjeux importants et probablement le plus prégnant pour la langue bretonne est d'augmenter sa visibilité dans notre environnement numérique. De notre smartphone à notre ordinateur, nous avons répertorié les différentes façons de voir et d'utiliser le breton dans notre vie de tous les jours afin, dans un premier temps, de recenser l'existant pour ensuite pouvoir s'attacher à combler les manques.

### **1 Utiliser ses appareils numériques en langue bretonne**

#### **1.1 Sur son ordinateur**

Il existe de nombreux logiciels libres traduits en langue bretonne, la plupart du temps par des associations comme An Drouizig, par l'Office public de la langue bretonne ou encore par des bénévoles qui travaillent en dehors de toute structure sans malheureusement toujours bien maîtriser la langue, les concepts qu'ils traduisent ou les terminologies adéquates.

L'offre en logiciel libre de droits traduits en breton couvre la plupart des usages actuels d'un ordinateur<sup>3</sup>:

- ◆ Une suite bureautique (**LibreOffice**) ainsi que des correcteurs d'orthographe et de grammaire (disponibles pour Microsoft Office, Chrome, Firefox, Thunderbird...):
  - Le correcteur **Hunspell**, sous licence libre, corrige l'orthographe (<https://github.com/Drouizig/hunspell-br>).
  - Le correcteur **An Drouizig**, disponible sur Microsoft Office, corrige l'orthographe et la grammaire



---

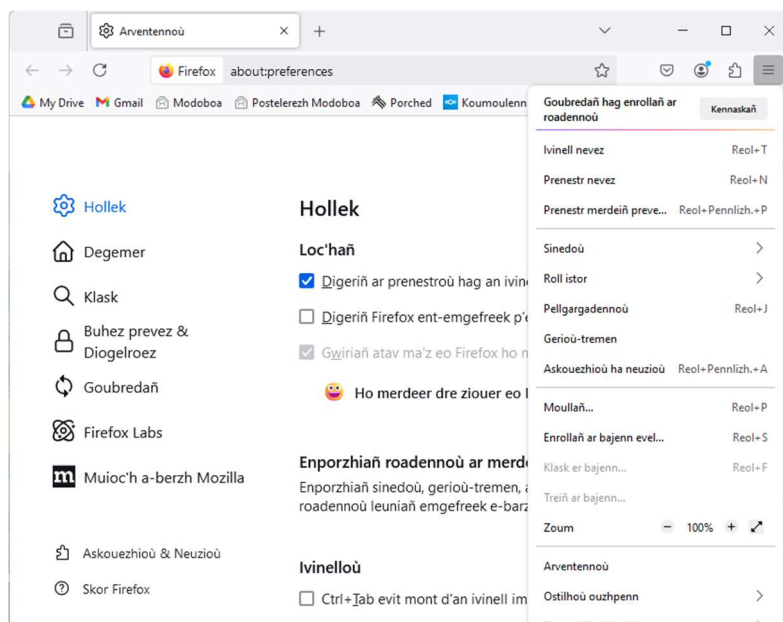
<sup>3</sup> L'annexe présente une liste d'autres logiciels traduits ou partiellement traduits en breton.

(<http://www.drouizig.org/index.php/br/binviou-br/an-drouizig-difazier/172-difazier-an-drouizig-pajenn-nevez>).

- **LanguageTools** corrige l'orthographe et quelques éléments de grammaire (<https://languagetool.org/br/>).
- ◆ Des logiciels pour la navigation sur le web (Mozilla **Firefox**) et l'échange de courriels (Mozilla **Thunderbird**)
- ◆ Des logiciels multimédia (**VLC** pour la vidéo, **Openshot** pour le montage vidéo, **Clementine** pour la musique)
- ◆ Des logiciels d'édition graphique (**Inkscape**, **Gimp**, **Pinta**, **Tuxpaint** logiciel de dessin pour les enfants, **Scribus** logiciel PAO)
- ◆ Des logiciels de gestion de contenu pour monter, gérer et mettre à jour facilement des sites Internet (**SIIP**, **Joomla !**)
- ◆ D'autres logiciels comme **FramaDate** pour créer de petits sondages, **Homebank** pour gérer des dépenses.

En revanche, si des traductions de logiciels propriétaires ont existé par le passé dans les années 1990/2000 (Skype, Facebook...), il n'en existe aujourd'hui plus aucune.

Du côté des systèmes d'exploitation, il existe des versions bretonnes pour certaines interfaces **Linux** parfois incomplètes. Une distribution Linux bilingue (français-breton), à destination des écoles bilingues était auparavant disponible. La traduction d'Ubuntu est entamée (<https://translations.launchpad.net/ubuntu>).



Firefox

La traduction d'Ubuntu est entamée (<https://translations.launchpad.net/ubuntu>).

L'OPLB a également mené à bien une première collaboration avec Microsoft permettant d'intégrer la langue bretonne dans la liste des langues prise en compte par cette entreprise. La locale Windows qui permet de présenter la date et le calendrier en langue bretonne a également été traduite et est aujourd'hui fournie par défaut sur tous les appareils.

Cependant, de manière générale, les systèmes et logiciels grand public les plus populaires et les plus utilisés, souvent privés, ne sont pas disponibles en breton, voire incompatibles avec l'utilisation du breton.

La langue bretonne est davantage présente dans le monde du Libre, qui dispose d'une moindre popularité auprès du public.

Les projets de traduction de logiciels libres peuvent être suivies sur des sites dédiés. Chacun peut participer au travail collaboratif sur des sites tels que **Weblate** (<https://hosted.weblate.org/languages/br/#overview>) ou **Pontoon** (<https://pontoon.mozilla.org/>) pour les outils proposés par Mozilla ou encore **Crowdin** (<https://crowdin.com/projects?q=#showcases>).

## 1.2 Sur son smartphone

La situation est encore plus dégradée en ce qui concerne les smartphones. L'offre d'applications permettant une utilisation de son téléphone en breton est extrêmement réduite. Parmi les applications indispensables (client mail, navigateur Internet, messagerie, appareil photo, GPS...) il n'existe de version bretonne que pour:

- ◆ **Firefox** (iOS et Android), navigateur web
- ◆ **K-9 mail** (Android), client de messagerie
- ◆ **VanilaMusic** (Android), lecteur musical
- ◆ **AntennaPod** (iOS et Android), lecteur de podcasts

On peut néanmoins noter la disponibilité de l'autocorrection et la prédiction de mots en breton sur le clavier virtuel **SwiftKey**. Le clavier **Gboard** développé par Google permet aussi de sélectionner le breton. Il permet aussi de créer des dictionnaires personnalisés pour les langues sélectionnées. L'application **Thumb-key** a également été adaptée pour le breton. Il s'agit d'un clavier conçu pour une utilisation avec les pouces sur les smartphones.

L'application LanguageTool (correcteur) est également disponible pour les appareils IOS. D'autres applications de jeux ou de traduction existent et sont mentionnées par la suite dans ce document.

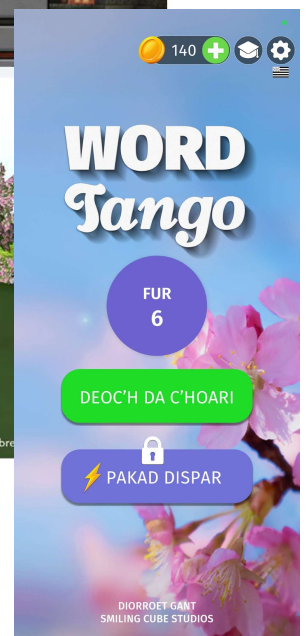
Du côté des systèmes d'exploitation pour smartphone, les principaux produits du marché ne sont pas disponibles en breton (iOS pour Apple et Android pour Google) même s'il existe une version alternative libre d'Android, **Lineage OS**, ouverte à la traduction en breton (<https://crowdin.com/project/lineageos>)<sup>4</sup>. Des projets libres comme Firefox OS et Ubuntu Phone ont été traduits en breton, mais comme ils sont maintenant abandonnés, le travail de bénévoles a été perdu. Le projet Ubuntu Phone a été repris sous le nom UBports et le projet de traduction associé est accessible sur Weblate (<https://hosted.weblate.org/projects/ubports/>).

### 1.3 Les jeux vidéo en breton

En ce qui concerne les jeux vidéo, la situation est la même sur ordinateur ou sur smartphone. Quelques titres ont été traduits, certains propriétaires et d'autres sous licence libre.

On peut mentionner le jeu populaire **Minecraft** par exemple (<https://www.minecraft.net/fr-fr>), ou d'autres jeux comme :

- **FreeCol** : (<https://www.freecol.org/index.html>), licence libre.
- **Knight Bar Fight** : jeu de combat multi-joueur ([https://store.steampowered.com/app/2738440/KBF\\_Knight\\_Bar\\_Fight/](https://store.steampowered.com/app/2738440/KBF_Knight_Bar_Fight/))



- **SuperTuxKart** : jeu de course semblable à Super Mario Kart ([https://supertuxkart.net/Main\\_Page](https://supertuxkart.net/Main_Page))
- **SuperTux** : jeu semblable à Super Mario Bros (<https://www.supertux.org/discover.html>)

---

<sup>4</sup> En janvier 2025, 20 % a été traduit (sans validation).

- **Hyperun** : jeu de course (<https://store.steampowered.com/app/565390/Hyperun/>)

Du côté des smartphones, il existe une poignée de jeux proposant une version bretonne, comme **Steredenn** (<http://steredenn.pixelnest.io/>), **Flipon** ou **SuperTuxKart** et **SuperTux** mentionnés ci-dessus.

Il existe aussi des applications ou des jeux en ligne basés sur la langue comme :

- **Gerdle** : deviner quotidiennement un mot breton de 5 lettres en 6 essais maximum (<https://brezhonadur.com/gerdle.html>).
- **Brezhle** : deviner quotidiennement un mot breton de 5 lettres en 6 essais maximum (<https://brezhle.u2042.com/>).
- **Word Tango** : application puzzle de mots, liste de mots à compléter avec les lettres proposés. iOS et Android (<https://word-tango.fr.ap-toide.com/app>).
- **BombParty** : jeu multijoueur en ligne dont le but est de trouver des mots à partir des lettres proposés (<https://jklm.fun/>).

À noter le site <https://br.boardgamearena.com/> dont l'interface est disponible en breton. Le site propose une sélection de jeux de société et de cartes.

## 1.4 Les paramètres régionaux en langue bretonne

La langue bretonne dispose d'un **code ISO**, qui est une norme internationale de codification des langues, permettant leur utilisation dans un environnement informatique.

Les paramètres régionaux, ou locales en anglais, sont un ensemble d'informations spécifiques à une langue ou à une région, utilisées par les systèmes d'exploitation (d'ordinateur ou de smartphone) pour afficher diverses données, par exemple :

- ♦ Les noms des langues, des territoires, des pays ;
- ♦ Les jours de la semaine, les mois, le format de l'heure ;



- ◆ Les caractères utilisés pour écrire la langue, et l'ordre dans lequel classer les mots ;

L'OPLB a d'abord mené une collaboration avec Microsoft pour rendre disponible les paramètres linguistiques du breton sur Windows. À l'heure actuelle, l'OPLB alimente le **CLDR** (*Common Locale Data Repository*) d'Unicode, qui regroupe l'ensemble des paramètres régionaux à destination des applications informatiques. Lors de la publication de la version 38 du CLDR fin 2020, le breton a atteint le niveau de couverture *Moderate++*, le dernier pallier avant d'atteindre le dernier niveau de couverture *Modern*<sup>5</sup>.

Cependant, ces paramètres régionaux n'ont de réelle utilité que s'ils sont implémentés par les différents éditeurs de logiciels. Par exemple, certains moteurs de recherche ou réseaux sociaux ne considèrent toujours pas « c'h » comme une lettre, et découpent les mots au niveau de l'apostrophe.

## 2 Surfer sur Internet

### 2.1 Les moteurs de recherche

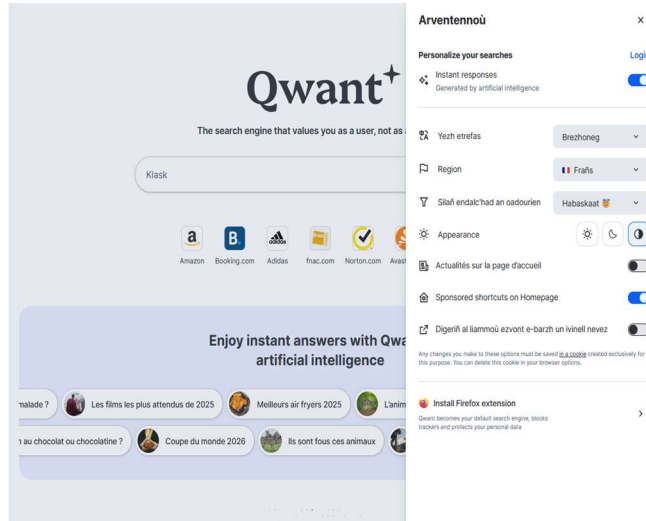
La première chose dont on a besoin pour utiliser Internet, c'est d'un moteur de recherche qui prenne en charge la langue dans laquelle on navigue. Pour le breton, cette première étape présente déjà des obstacles, même si l'on ne part pas de zéro :

- ◆ **Google** propose une interface partiellement disponible en langue bretonne. Il manque par exemples les traductions pour la partie recherche avancée. L'option recherche vocale n'est pas disponible tandis que pour l'option clavier en ligne, celui-ci ne propose qu'un mode par défaut (*Africa Latin*).

---

<sup>5</sup> [https://www.unicode.org/cldr/charts/46/supplemental/locale\\_coverage.html#br](https://www.unicode.org/cldr/charts/46/supplemental/locale_coverage.html#br). Pour la version v46.0 en octobre 2024, le taux atteint est de 99,7% pour le niveau *Moderate* et 84,5% pour le niveau *Modern*.

- ◆ **Qwant**, quant à lui est entièrement traduit en breton pour ce qui est de son interface mais ne prend pas en charge le breton comme langue de recherche.
- ◆ **DuckDuckGo** présente également une interface presque entièrement traduite mais son algorithme de recherche ne prend pas non plus le breton en compte.



## 2.2 Les réseaux sociaux

Même si les principaux réseaux sociaux ne disposent pas d'interface officielle en langue bretonne, ils peuvent compter sur un nombre d'utilisateurs croissant qui s'y expriment en breton.

- ◆ Sur **X (Twitter)**, on comptabilisait, selon indigenoustweets, 713 utilisateurs qui tweetent en breton<sup>6</sup> (au moins un tweet) (17 000 pour le basque, 15 000 pour le gallois ou l'irlandais, 2 000 pour le gaélique écossais, 100 pour l'occitan). Au total 180 824 tweets avaient été rédigés en breton le 27 octobre 2020. Selon les derniers chiffres disponibles (mars 2023), près 273 000 tweets ont été rédigés en breton par 720 utilisateurs. Si le nombre d'utilisateurs n'a pour ainsi dire pas augmenté, le nombre de tweets postés a bien augmenté. Twitter est devenu X en juillet 2023 et l'interface n'est toujours pas disponible en breton.
- ◆ **Facebook** est partiellement disponible en breton. Il s'agit d'une traduction participative de qualité variable. Le groupe « Facebook e brezhoneg », compte plus 12 000 membres quand il y en avait environ 10 000 en 2020.



<sup>6</sup> Source : <http://indigenoustweets.com/br/>, consulté le 27 octobre 2020.

- ◆ **Mastodon**, réseau semblable à Twitter mais libre de droits, fait l'objet d'un projet de traduction participative en breton (<https://crowdin.com/project/mastodon>)<sup>7</sup>.



- ◆ **Instagram**, réseau social pour échanger des photographies et de courtes vidéos. Plusieurs comptes proposent des contenus en langue bretonne comme des posts présentant des mots en breton sur *komzbrezhoneg*, *geriou\_avankdit*, *spered\_prod* ou des posts humoristiques sur *brezhoneghaplijadur*.



- ◆ **BlueSky**, crée par l'un des fondateurs de Twitter a connu une forte augmentation de nombre d'utilisateurs en 2024. L'interface n'est pas disponible en breton mais l'on y trouve de nombreux utilisateurs rédigeant en breton. Il n'est pas encore possible d'estimer correctement la présence du breton sur ce réseau, les données disponibles ne semblant pas détecter correctement la langue.



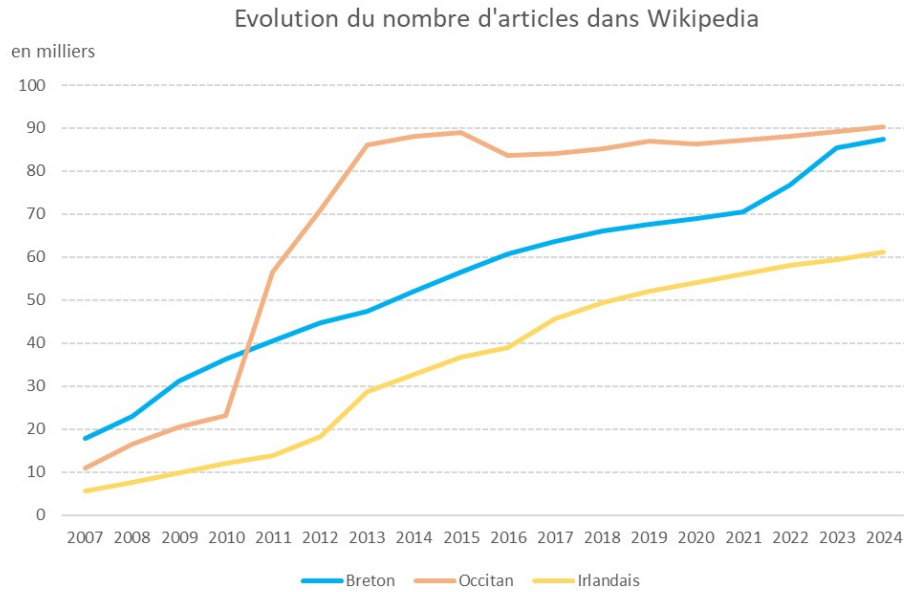
## 2.3 Les contenus en langue bretonne

On trouve de plus en plus de contenu en langue bretonne sur Internet, mais il reste difficile d'évaluer leur développement. On retrouve du breton surtout sur des sites spécialisés dans la culture bretonne ou sur des sites institutionnels. Toutefois, les sites bilingues sont de plus en plus proposés notamment par des mairies, des communautés d'agglomération, des entreprises...

- ◆ Il existe une version en langue bretonne de **Wikipédia** depuis 2004. En novembre 2020, on dénombrait 70 000 articles en breton (86 000 en occitan, 132 000 en gallois, 370 000 en basque contre 2 millions en français et 6 millions en anglais). En novembre 2024, le nombre d'articles s'élève à 87 200 en breton, 90 260 en occitan, 281 380 en gallois, 447 910 en basque tandis qu'il existe 2,6 millions d'articles en français et 6,9 M en anglais. Fin 2024, le breton est classé au 81<sup>e</sup> rang selon le nombre d'articles.

---

<sup>7</sup> En janvier 2025, le taux de traduction atteint 25 % (sans validation).



En 4 ans, le nombre d'articles a progressé d'un quart, un niveau semblable au rythme de progression du basque (+ 20 %). Pour le breton, cela équivaut à un accroissement annuel de + 5,6 %. En gallois, le nombre d'articles a doublé entre 2020 et 2024.

En novembre 2020, 85 utilisateurs actifs écrivaient ou rédigeaient des articles en breton. Ce nombre s'élève à 105 en novembre 2024 (116 en gallois). Wikipédia est le plus grand site de contenus en langue bretonne. Cependant, la qualité et l'exhaustivité des articles restent à améliorer. En 2023, un poste de traducteur de contenus numériques a été créé au sein de l'OPLB. L'une de ses missions consiste à augmenter et compléter le Wikipédia breton. La diversité des articles peut être évaluée grâce à la liste des 100 articles vitaux que chaque Wikipédia devrait avoir.

[https://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias\\_by\\_sample\\_of\\_articles](https://meta.wikimedia.org/wiki/List_of_Wikipedias_by_sample_of_articles)

En novembre 2024, le breton est classé au 83<sup>e</sup> rang. 87 articles de la liste sont manquants pour le breton.

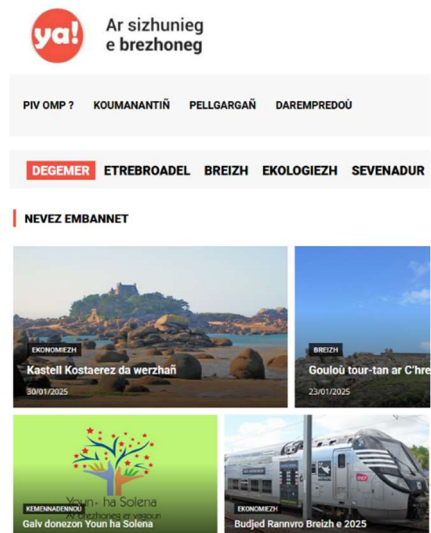
## Les 1000 articles vitaux que chaque Wikipédia devrait proposer

	Breton	Basque	Gallois	Occitan
<b>rang</b>	<b>83</b>	<b>14</b>	<b>55</b>	<b>37</b>
article manquant	87	0	2	0
petit article (<10k caractères)	773	47	662	390
article (>10k et <30k caractères)	105	287	166	256
article long (>30k caractères)	35	666	170	354

*Lecture* : Parmi les 1000 articles vitaux qui devrait proposer chaque Wikipédia, il en manque 87 pour le breton. Une majorité de articles de cette liste sont assez courts (773). Ils contiennent moins de 10 000 caractères. 35 articles rédigés en breton sont longs avec plus de 30 000 caractères. Plus les articles sont exhaustifs, plus le classement de la langue est meilleur. Le basque est classée 14<sup>e</sup> car aucun article vital n'est manquant et les deux tiers d'entre-eux sont longs.

Le fonds de dotation Bretagne numérique propose des ateliers pour développer le contenu relatif à la culture et au patrimoine bretons sur Wikipédia. 14<sup>e</sup>

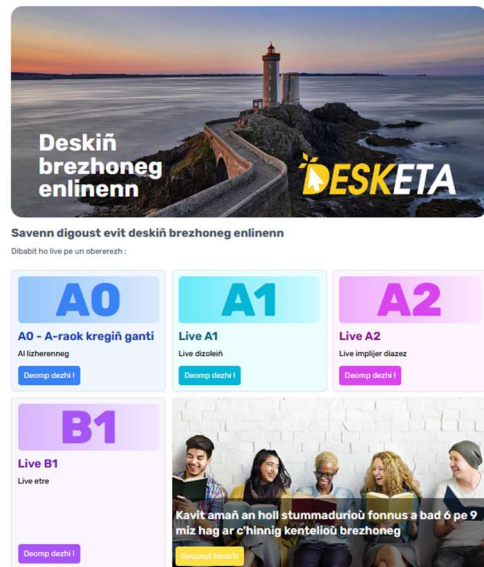
- ◆ De nombreux sites ou blogs existent en breton, pour l'essentiel dédiés à la langue ou la culture bretonne.
- ◆ On trouve également des sites de médias en ligne tels : **Brezhweb** pour la TV, **Radio Breizh** pour la radio. Le site **breizhVOD**, moyennant un abonnement, propose une sélection de films, séries, dessins animés et documentaires doublés en breton. Il existe également quelques chaînes en breton sur la plateforme **Twitch.tv** (Tornoz, geekezig...). À cela s'ajoute les sites regroupant des articles de presse : **Bremañ, Ya, Splann**, etc.
- ◆ Il est également possible d'écouter des **Podcasts** en breton, en plus de ceux produits par les radios, comme *Petra c'hoari ganit*, *Danevelloù Al Liamm*, *Korn ar brezhoneg Ouest-France*, *Heklev...* Le site (<https://podgrab.fedi.bzh/fr/>) en recense une grande partie.
- ◆ Enfin, il existe quelques sites institutionnels traduits en breton. Ils sont plus nombreux qu'en 2020 bien que parfois les sites ne soient pas totalement traduits. Parmi eux, on peut mentionner le site du conseil régional de Bretagne, les sites de certaines mairies : Quimper, Carhaix... ; des sites dédiés au transport : Bibus, Izilo; ainsi que le site de Paysages de Mégalithes portant la candidature à l'inscription des Mégalithes de Carnac au patrimoine mondial de l'Unesco.



### 3 Apprendre la langue à l'aide d'outils numériques

L'apprentissage du breton n'est pas proposé par les outils les plus populaires comme Duolingo, Memrise, Babbel... Il existe cependant quelques sites et applications spécialisées.

- ◆ **Desketa** est une plateforme d'auto-apprentissage gratuite développée par l'OPLB. Elle a été officiellement ouverte en juin 2024 en proposant des cours pour les niveaux A0 et A1. Le niveau A2 a été ajouté début juillet 2024. En janvier 2024, 8 000 personnes étaient inscrites. (<https://desketa.bzh/>)
- ◆ L'application **uTalk** (iOS et Android). L'apprentissage de plus de 150 langues est proposé par cette application payante. Le breton y figure depuis 2022. (<https://utalk.com/en/store/breton>)
- ◆ “**Deskomp brezhoneg**” (depuis juin 2023). Le site propose 5 modules permettant l'apprentissage des bases de la langue bretonne. Il s'agit de l'adaptation d'une méthode galloise réalisée par Mervent et Skol an Emsav. L'inscription est gratuite. (<https://learnwelsh.cymru/supporting-breton/>)
- ◆ L'application gratuite **Kiou** (iOS et Android). Il s'agit d'une adaptation, réalisée en 2021 de l'application galloise Antur Cyw. L'application est destinée aux enfants. Malheureusement, suite à des problèmes techniques, l'application n'est plus téléchargeable (fin 2024).
- ◆ **Loecsen** : site et application proposant des cours à base de tournures et phrases à lire et à écouter, suivis de jeux.(*possibles erreurs*) (<https://www.loecsen.com/en/learn-breton#/en/Essentials>).
- ◆ L'application de la méthode **Assimil** (iOS et Android), payante, qui reprend la version papier.
- ◆ L'application et le site **EduBreizh** lancé en 2011 par Skolanet proposait des cours en ligne payants n'a pas rencontré son public. L'offre n'existe plus.



- ◆ **Kervarker** où l'on trouve quelques leçons et d'autres sites Internet quelque peu démodés qui ne semblent plus être mis à jour. ([http://www.kervarker.org/fr/lessons\\_01\\_toc.html](http://www.kervarker.org/fr/lessons_01_toc.html))

Il existe également des applications ou des sites proposant des outils pour pratiquer la langue : apprendre de nouveaux mots ou de petits jeux.

- ◆ L'application **#bzhg c'hoari** et **#bzhg deskiñ** (iOS et Android). Ces applications sont proposées par Skol an Emsav depuis janvier 2024. La première permet d'apprendre nouveaux mots par thème et la seconde pour retrouver le lexique et les tournures utilisés dans le journal #brezhoneg. L'application est gratuite et peut être utilisée hors-ligne. (<https://skolanemsav.bzh/brezhoneg/>)
- ◆ **Kwizh Brezhoneg** : application pour apprendre des mots de vocabulaire en breton (iOS et Android). (<https://appadvice.com/app/kwizh-brezhoneg/1504003644>).
- ◆ L'application **"kontañ e brezhoneg"** (Android), créé en 2020 permet d'apprendre d'une manière ludique les nombres en breton. (<https://play.google.com/store/apps/details?id=fr.developpeur35.compterenbreton&hl=en>).
- ◆ Le site **educandy propose différents jeux** (mémoire, paires...). Le contenu proposé en breton a été créé par Kentelioù an noz. (<https://www.educandy.com/site/resource.php?activity-code=301b>).
- ◆ Le site **TES** propose de plus en plus de jeux à destination des enfants. (<https://www.reseau-canope.fr/tes/subsites/games/korn-ar-vugale/>)
- ◆ Le site **Mogidell** propose des exercices et du contenu grammatical. <https://mogidell.wordpress.com/>



il existe d'autres outils, certains développés automatiquement pour plusieurs langues. La qualité de la langue proposée n'a pas été évaluée.

- ◆ L'application **Lingue Vive** (iOS et Android) proposant un imagier à destination des enfants. L'application a aussi été développée pour le corse, l'alsacien, le basque et l'occitan. (<https://lingue-vive.com/imagier-breton/>).

- ◆ L'application **Breton M(A)L** (iOS et Android). Multiples jeux basés sur les mots. Application générée automatiquement pour une multitude de langues.  
(<https://www.masteranylanguage.com/c/p/o9ly8m04FI/Breton>).
- ◆ L'application **Vocabulaire Breton-Français** (iOS). Elle permet d'apprendre ou de réviser du lexique à partir de fiches (51 000 mots selon la description).  
(<https://apps.apple.com/fr/app/vocabulaire-breton-fran%C3%A7ais/id1614210335>).
- ◆ L'application "**Clozemaster**" (iOS et Android). Jeux pour apprendre des mots en complétant des phrases (*contient des erreurs*).  
(<https://www.clozemaster.com/languages>).

On peut mentionner le discord Deskiñ brezhoneg, créé en 2019 et comptant près de 800 membres (2025). Sur cette plateforme, les membres peuvent partager des ressources, échanger par écrit et par oral ou organiser des cours.

## ❖ **Partie 2. Inventaire des ressources linguistiques et des outils numériques en langue bretonne**




---

Moins visibles, les ressources et outils linguistiques n'en sont pas moins essentiels. Ils sont destinés avant tout aux linguistes et aux informaticiens. Ils permettent le développement de nouvelles applications, de nouveaux services et de nouvelles technologies en langue bretonne.

Pour établir cet inventaire, nous avons choisi une classification inspirée de celle utilisée par les Livres blancs de META-NET<sup>8</sup>. Sont tout d'abord présentées les ressources linguistiques, qui comprennent les éléments nécessaires à l'élaboration d'outils de traitement de la langue. Nous avons répertorié uniquement les ressources disponibles sur Internet. Ensuite, nous avons listé les outils et technologies de traitement de la langue développés à ce jour pour le breton.

Des tableaux présentent la situation actuelle des ressources et des outils disponibles pour la langue bretonne. Ils sont également inspirés des tableaux utilisés dans les rapports META-NET. Pour chaque type de ressources ou d'outils, 6 indicateurs sont présentés, dont la couleur varie suivant que l'outil est satisfaisant (vert) ou insatisfaisant (rouge). Les 6 indicateurs sont :

- ❖ **Quantité** : Les ressources ou les outils existent-ils pour le breton ? En existe-t-il plusieurs ?

	Il n'existe pas de ressource ou d'outil de ce type pour le breton.
	Les ressources ou outils existent en quantité insuffisante.
	La quantité de ressources ou d'outils de ce type est satisfaisante.

---

<sup>8</sup> META-NET est une organisation européenne qui rassemble des acteurs des technologies de la langue. Elle a écrit plusieurs "Livres blancs" rendant compte de l'état des langues officielles de l'Union Européenne et de plusieurs langues non-officielles de l'UE dans le domaine des technologies de la langue.

- ◆ **Disponibilité** : Les ressources ou les outils sont-ils disponibles en ligne ? Sont-ils libres de droits ? Gratuits ? Compatibles avec plusieurs systèmes d'exploitation ?

Les ressources ou les outils ne sont pas disponibles en ligne, ne sont pas consultables ou ne sont pas téléchargeables.

Les ressources ou outils sont disponibles; mais ne sont pas sous licence libre, ne fonctionnent pas sur tous les systèmes d'exploitation ou sont payants.

Les ressources et outils de ce type sont disponible en ligne gratuitement, téléchargeables et libre de droits.

- ◆ **Qualité** : Les ressources ou outils sont-ils de bonne qualité ? Ont-ils été évalués ? Sont-ils performants ?

Les ressources et outils sont de mauvaise qualité, leurs performances n'ont pas été évaluées.

Les ressources et outils peuvent être de bonne qualité mais leurs performances n'ont pas été évaluées.

Les ressources et outils sont de bonne qualité, leurs performances ont fait l'objet d'évaluations.

- ◆ **Couverture** : Les ressources ou outils assurent-ils une bonne couverture des usages pour lesquels ils ont été conçu (genre, domaine, langues prises en charge...)?

Les ressources ou outils ont une couverture très faible ou ne couvrent que des cas très spécifiques.

Ces ressources ou outils offrent une bonne couverture mais restent spécifiques à certains domaines.

Les ressources ou outils offrent une très bonne couverture et couvrent des domaines variés.

- ♦ **Durabilité** : Les ressources ou les outils sont-ils toujours maintenus ? Sont-ils documentés ? Correspondent-ils aux normes ?

Utilisation de formats propriétaires, peu ou pas de mises à jour, pas de documentation.

Les outils et ressources n'utilisent pas les standards à jour, peu de mise à jour, peu de documentation.

Les ressources et les outils utilisent les dernières normes et sont très bien documentés.

- ♦ **Adaptabilité** : Les ressources ou les outils peuvent-ils être adaptés à de nouvelles tâches, de nouveaux domaines ? Avec quelle facilité ?

Il est impossible d'utiliser les ressources ou les outils pour effectuer d'autres tâches que celles pour lesquelles ils ont été conçus, ou de les transposer à d'autres domaines.

Les ressources ou les outils sont partiellement adaptables à d'autres tâches ou domaines, ou difficiles à adapter.

Les ressources ou les outils peuvent être facilement et entièrement adaptés à de nouvelles tâches ou de nouveaux domaines.

## 1 Les ressources linguistiques

Dans le but de concevoir des outils linguistiques, il est nécessaire en premier lieu de collecter un ensemble de données de différentes natures. Ces données seront ensuite utilisées pour créer, tester ou alimenter différents algorithmes, logiciels et systèmes. Nous pouvons diviser ces ressources en plusieurs catégories : les ressources lexicales qui concernent les mots et leurs caractéristiques, les ressources grammaticales qui offrent des représentations informatiques du fonctionnement de la langue, et les corpus qui donnent des contextes d'utilisation de la langue.

### 1.1 Ressources lexicales

Les **dictionnaires**, **lexiques** et **bases de données terminologiques** sont les principales ressources disponibles pour la langue bretonne.

Ils sont utilisés de la même façon qu'un dictionnaire papier par la majorité des utilisateurs, mais ils sont aussi utilisés par des logiciels ou algorithmes pour effectuer des opérations de traitement de la langue : auto-correction, traduction automatique... à condition qu'ils soient disponibles dans un format numérique et normalisé (TEI, XML...).

Les ressources lexicales sont les plus répandues et usitées. Il est ainsi important de garantir leur durabilité et disponibilité.

Les **dictionnaires monolingues** proposent en général diverses informations : définition, prononciation, exemples, catégorie grammaticale, étymologie, entre autres. Dans les **dictionnaires bilingues**, chaque mot breton correspond au moins à un mot dans une autre langue.

	Quantité	Disponibilité	Qualité	Couverture	Durabilité	Adaptabilité
Dictionnaires monolingues	Jaune	Vert	Vert	Vert	Vert	Jaune
Dictionnaires bilingues	Vert	Vert	Jaune	Vert	Jaune	Jaune
Dictionnaires multilingues	Rose	Rose	Rose	Rose	Rose	Rose
Bases terminologiques	Vert	Vert	Vert	Vert	Vert	Rose
Bases toponymiques	Vert	Vert	Vert	Vert	Vert	Jaune

Une **base de données terminologique** est une base dans laquelle on retrouve des termes spécifiques à certains domaines, qu'on ne retrouve pas habituellement dans les dictionnaires courants.

Comme le montre le tableau ci-dessus, les ressources lexicales pour le breton présentent toutes un **problème d'adaptabilité** : il est difficile, parfois impossible, de les utiliser dans d'autres projets que celui pour lequel elles ont été créées.

Les dictionnaires bilingues existants présentent des **problèmes de qualité** : les informations qui y sont présentées ne sont parfois pas vérifiées ni sourcées ; **de durabilité** : certains sont peu documentés et peu mis à jour ; et **d'adaptabilité** : les dictionnaires les plus complets sont disponibles uniquement sous forme PDF.

C'est un point important à prendre en compte afin que les données perdurent et restent disponibles.

- ◆ Sur Internet, on peut trouver deux **dictionnaires monolingues** :

**Meurgorf** <https://niverel.brezhoneg.bzh/br/meurgorf/>

Dictionnaire historique de la langue bretonne développé par l'OPLB. Fin 2024, le dictionnaire contient 60 802 entrées (+5 000 par rapport à octobre 2020). Il contenait 55 747 entrées en octobre 2020. On y trouve des définitions, des formes dérivées, des exemples historiques...L'outil a été complété par des transcriptions phonétiques, des prononciations enregistrées et la fréquence d'utilisation des termes.

*Disponible en ligne, exemples historiques téléchargeables*

**Wikeriadur** <https://br.wiktionary.org/wiki/Wikeriadur:Degemer>

Version bretonne de Wiktionary, un dictionnaire libre de droits et participatif. Il est incomplet et sa qualité varie suivant les entrées. Fin 2024, il contenait 52 509 articles contre 43 823 fin 2020. (+ 20 %)

*Disponible en ligne*

On peut également mentionner le site <https://geriadur.brezhoneg.world/>. Il propose un complément au dictionnaire monolingue d'An Here avec l'ajout ou la complétion de 320 entrées. Le site propose également des lexiques.

- ◆ Il existe également trois **bases terminologiques** :

**TermOfis** <https://www.fr.brezhoneg.bzh/36-termofis.htm>

Base terminologique de l'OPLB. Base officielle pour la langue bretonne, elle contient 91 000 termes en janvier 2025 (+ 16 500 en quatre ans).

*Disponible en ligne, elle ne peut pas être téléchargée.*

**Brezhoneg21** <http://www.brezhoneg21.com/geriadurGB.php>

Dictionnaire des sciences et techniques. Il a été conçu pour les écoles Diwan, sur les domaines touchant à l'éducation.

*Disponible en ligne et via l'application Brezhoneg21. Il est également téléchargeable.*

**Preder** <https://preder.net/r/geriadur/geriadur.php?locale=fr>

Base de données conçue à partir des dictionnaires publiés par Preder.

*Disponible en ligne, elle ne peut pas être téléchargée.*

Il existe aussi une base de données spécifique à la dénomination des plantes

**PlantKelt** <http://app2.plantkelt.bzh/akp/web/home?2>

Base de données spécifique à la botanique, créé par Roland Mogn, Gwenole ar Menn et Claude Figureau. La base contient près de 40 000 noms en breton (2025).

*Disponible en ligne.*

- ◆ On trouve des **bases toponymiques** :

**KerOfis** <https://niverel.brezhoneg.bzh/br/kerofis/>

Développée par l'OPLB, la base fournit les formes normalisées des noms de lieux, communes, rues, etc. En janvier 2025, la base contient plus de 57 000 entrées.

*Disponible en ligne, les données peuvent être téléchargées.*

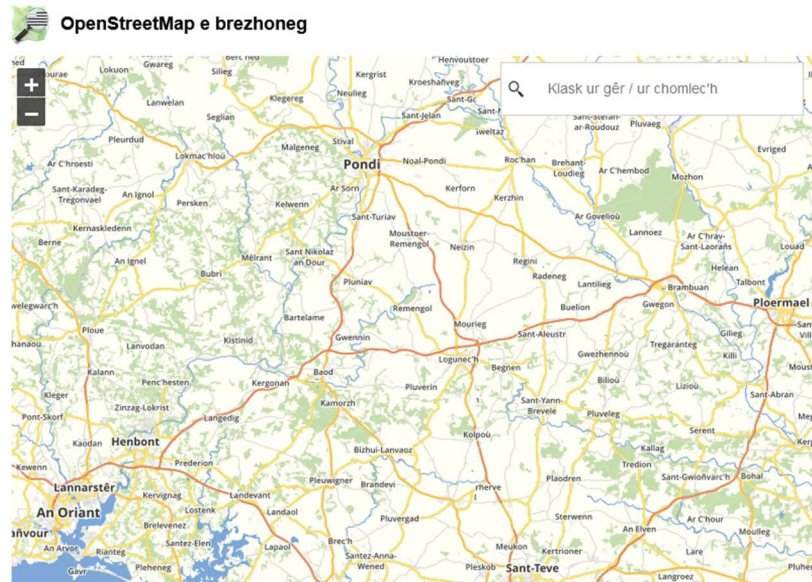
**BAL (Base Adresse Locale)** <https://adresse.data.gouv.fr/commune>

La Base Adresse Nationale, BAN a été créée en 2015. Cette base contient l'ensemble des adresses normalisées ainsi que leurs coordonnées géographiques. La BAN est alimentée à partir des bases BAL (Base Adresse Locale) administrées par chaque commune. Celles-ci ont la possibilité de renseigner la forme bretonne des adresses. L'OPLB les soutient et les conseille dans leurs démarches afin de respecter le patrimoine toponymique breton.

Téléchargeable pour chaque commune (colonne "voie\_nom\_bre" et "commune\_nom\_bre" si renseignée)

**OpenStreetMap** <https://www.openstreetmap.bzh/fr/>

Version en breton du site de cartographie en ligne libre de droits OpenStreetMap.



Disponible en ligne, code disponible sur GitHub.

- ◆ Plusieurs **dictionnaires bilingues**, de qualité variable :

**Glosbe** <https://fr.glosbe.com/br/fr>

Dictionnaire participatif multilingue dans lequel chacun peut ajouter des mots. On y trouve des mémoires de traduction compilées sur Internet, qui peuvent cependant présenter des problèmes d'alignement.

*Disponible en ligne, il ne peut pas être téléchargé.*

**Dictionnaire Cornillet-Ménard** <https://fr.scribd.com/document/794449961/geriadur-brezhoneg-galleg>

Dictionnaire bilingue breton-français, construit par Gérard Cornillet, à partir de plusieurs dictionnaires papier (Geriadur An Here, Dictionnaire français-breton par Martial Ménard...).

*Téléchargeable au format PDF.*

- Favereau** <https://www.skolvreizh.com/geriadur-meur-francis-favereau/>  
Version actualisée du dictionnaire Favereau.  
*Téléchargeable au format PDF.*
- Devri** <http://devri.bzh>  
Dictionnaire diachronique, breton-français, créé par Martial Ménard et qui continue à être complété par Herve Bihan. En juillet 2022, il contenait près de 70 000 entrées complétées d'exemples historiques d'utilisation.  
*Disponible en ligne, ne peut pas être téléchargé.*
- FreeLang** <https://www.freelang.com/enligne/breton.php>  
Dictionnaire créé par Tomaz Jacquet, qui contient 37 800 entrées.  
*Disponible en ligne, téléchargeable (Windows, Android).*
- FreeDict** <https://freedict.org/downloads/>  
Dictionnaire également créé par Tomaz Jacquet, 27 034 mots dans le sens breton-français et 36 017 dans le sens français-breton.  
*Téléchargeable sous formes normalisées (TEI, ...).*
- Apertium** <https://github.com/apertium/apertium-br-fr>  
Dictionnaire inclus dans le programme Apertium, basé sur le dictionnaire de Tomaz Jacquet et complété par Fulup Jakez.  
*Téléchargeable sur le GitHub d'Apertium.*  
Tyers, Francis M. 2009. 'Rule-based augmentation of training data for breton-french statistical machine translation', *Proceedings of the 13th Conference of the European Association for Machine Translation*, 213–218.  
Licence Creative Commons CC BY-SA
- Dictionnaire Jan Deloof** [https://codecs.vanhamel.nl/Jan\\_Deloof/Woordenboek\\_BZH-NL-BZH](https://codecs.vanhamel.nl/Jan_Deloof/Woordenboek_BZH-NL-BZH)  
<https://kevindonnely.org.uk/deloof/index.php>

Dictionnaire bilingue breton-néerlandais et néerlandais-breton.

*Téléchargeable et consultable en ligne sur le site <https://kevindonnelly.org.uk/deloof/index.php>.*

On trouve également un dictionnaire (beta) breton-néerlandais sur le site suivant :

<https://www.woordenboek.eu/vertaal.php?taal=Bretons>

**Geriadur logos** <https://www.logosdictionary.org/index.php>

Dictionnaire multilingue alimenté par un réseau de traducteurs, dont Rhisiart Hincks pour le breton.

**Dictionnaire Rita Williams** <https://bzh.cymru/br/>

Dictionnaire bilingue breton-gallois élaboré par Rita Williams et mis en ligne par l'association Breizh-Cymru.

*Disponible en ligne, ne peut pas être téléchargé.*

**Dictionnaire allemand-breton** <https://github.com/Ofis-publik-ar-brezhoneg/geriadur-DEU-BRE>

Dictionnaire allemand-breton rédigé par Robert Pellen, ancien professeur d'allemand.

*Téléchargeable au format PDF.*

On peut également mentionner plusieurs sites proposant des traductions à partir de plusieurs ressources lexicales :

**Geriafurch** : <https://geriafurch.bzh/br>.

**Lexilogos** : [https://www.lexilogos.com/breton\\_dictionnaire.htm](https://www.lexilogos.com/breton_dictionnaire.htm).

**PanLex** : <https://apps.panlex.org/translator/?langDe=fra-000&langAl=bre-000&txt=commentaire>

D'autres outils sont également disponibles :

- L'application **Dictionnaire Breton-Français** (iOS depuis avril 2020). Il contiendrait 50 000 entrées. (<https://apps.apple.com/fr/app/dictionnaire-breton-fran%C3%A7ais/id1505733566>)



Les corpus sont très **importants** pour le traitement automatique des langues. Ils représentent la manière dont une langue est utilisée en réalité. Quand il est assez complet pour couvrir la plupart des cas d'utilisation de la langue, on parle de **corpus de référence**. Il existe plusieurs types de corpus : les corpus monolingues qui présentent le fonctionnement de la langue, et les corpus **bilingues** ou **multilingues** qui sont utiles pour comparer deux langues entre elles.

Comme on peut le voir dans le tableau ci-dessus, il n'existe pas suffisamment de corpus en breton pour mener à bien le développement d'outils avancés de traitement de la langue bretonne.

	Quantité	Disponibilité	Qualité	Couverture	Durabilité	Adaptabilité
Corpus monolingues	Jaune	Jaune	Rose	Rose	Jaune	Jaune
Corpus monolingues arborés	Jaune	Vert	Vert	Rose	Jaune	Vert
Corpus bilingues parallèles	Jaune	Vert	Jaune	Jaune	Jaune	Vert
Corpus oraux	Jaune	Jaune	Vert	Vert	Vert	Jaune
Corpus multimédia	Rose	Rose	Rose	Rose	Rose	Rose

Par leur nombre et leur taille réduits, les corpus monolingues offrent une couverture très incomplète de la langue bretonne.

Bien que peu nombreux, les corpus de la parole sont ceux qui présentent la meilleure qualité : ils sont accessibles et téléchargeables, sous licence libre de droits, ils sont bien documentés et dans des formats correspondants aux standards en matière de construction de corpus (divisés en plusieurs parties dédiées à l'entraînement, l'évaluation ; emploi de formats standards comme MP3 ou WMA...).

Les corpus sont un enjeu important. Avec le développement de l'intelligence artificielle, les modèles linguistiques sont entraînés sur des données de plus en plus importantes et diverses et sont souvent efficaces pour les langues qui proposent beaucoup de contenu.

## 1.2.1 Corpus textuels monolingues

Les corpus monolingues sont utilisés pour construire des modèles statistiques par exemples, ou pour tester des algorithmes. On peut diviser ces corpus en deux catégories : les **corpus spécialisés**, qui rassemblent des textes présentant un trait particulier de la langue : l'époque, le domaine, le niveau de langue... ; et les **corpus de référence** qui représentent la langue dans son ensemble.

### 1.2.1.1 Corpus spécialisés

Les corpus spécialisés pour la langue bretonne ont principalement été élaborés à partir du contenu disponible dans Wikipedia. Ces corpus peuvent être lemmatisés<sup>9</sup> ou annotés morphosyntaxiquement<sup>10</sup>.

**Leipzig Corpora** <https://wortschatz.uni-leipzig.de/en/download>

Corpus de 100 000 phrases extraites du Wikipédia breton, soit environ 2 M de mots. Domaine encyclopédique/Internet.

*Disponible en ligne, téléchargeable.*

D. Goldhahn, T. Eckart & U. Quasthoff: Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. *In: Proceedings of the 8th International Language Resources and Evaluation (LREC'12), 2012*

Licence Creative Commons CC BY

### 1.2.1.2 Corpus arborés

Un **corpus arboré**, ou **corpus syntaxique**, est un corpus de référence qui, en plus des informations grammaticales, contient des informations sur les relations syntaxiques entre les mots. Ce type de corpus est très utile pour la recherche linguistique ou informatique, mais il requiert une somme de travail très importante. Il existe un corpus arboré du breton, de petite taille.

---

<sup>9</sup> La lemmatisation consiste à remplacer les mots par leur forme non fléchie, par exemple l'infinitif pour les verbes, le masculin singulier pour les noms et adjectifs, etc.

<sup>10</sup> L'annotation morphosyntaxique consiste à étiqueter les mots avec leur classe grammaticale, ou leur partie-du-discours par exemple.

## UD Breton-KEB

[https://github.com/UniversalDependencies/UD\\_Breton-KEB](https://github.com/UniversalDependencies/UD_Breton-KEB)

Petit corpus de 888 phrases (soit environ 10 000 mots) annoté (catégorie grammaticale et arbre syntaxique). Extraits provenant de livres de grammaire, Wikipédia, *Bremaik*, de textes de l'OPLB et de chansons.

*Disponible en ligne et téléchargeable.*

Francis M. Tyers and Vinit Ravishankar, *A prototype dependency treebank for Breton*, Actes de la 25<sup>ème</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN), 2018

Licence Creative Commons CC BY-SA

## UniMorph

<https://unimorph.github.io/>

<https://github.com/unimorph/bre/blob/master/bre>

Corpus annoté de verbes conjugués. (2 339 formes pour le breton en janvier 2025).

*Téléchargeable*

## ARBRES

<http://arbres.iker.cnrs.fr.>

Grammaire descriptive du breton présentant de nombreux exemples annotés au niveau de la catégorie grammaticale et de la structure syntaxique.

ARBRES

rechercher dans Arbres

### Pluriel interne

ce qu'on appelle un pluriel interne est la forme plurielle d'un nom dont la base montre un changement de voyelle entre son singulier et son pluriel. En (1), le pluriel buññ est construit sur la base du singulier buññ. Le pluriel suttial n'a pas de voyelle avant la, elle s'est répandue par anticipation sur la base du nom.

(1) Anññ enññ ar enññ buññ  
buññ est le douzour buññ  
'Le distributeur de billets est là.'  
(2) Anññ enññ ar enññ buññ

Par rapport aux pluriels dérivés uniquement par suffixation, il y a en breton relativement peu de pluriels internes. La plupart des pluriels internes sont visiblement dérivés par une suffixation.

#### Schéma

1. Phonologie

- 1.1. accentuation
- 1.2. mutation diédiale dans le voyelle

2. Morphologie

- 2.1. exemples de finales avec un pluriel interne
  - 2.1.1. final -e: buññ, buññ
  - 2.1.2. final -e: -e, -e
  - 2.1.3. final -e: -e, -e
  - 2.1.4. final -e: -e, -e
  - 2.1.5. final -e: -e, -e
  - 2.1.6. final -e: -e, -e

*Disponible en ligne*

Jouitteau, Mélanie. (éd.). 2009-2023. ARBRES, wikigrammaire des dialectes du breton et centre de ressources pour son étude linguistique formelle, IKER, CNRS

Licence Creative Commons BY-NC-SA.

## 1.2.2 Corpus parallèles bilingues

Un **corpus parallèle** est un corpus bilingue (ou multilingue) où sont mis en relation les segments (phrases, propositions...) et leur traduction. Ils sont souvent constitués à partir de mémoires de traduction. On utilise ce genre de corpus pour développer des traducteurs automatiques, des logiciels d'aide à la traduction, etc.

### **OPUS**

<http://opus.nlpl.eu>

Corpus breton-français, 400 000 phrases, 3M de mots. Un tiers du corpus provient de mémoires de traduction de l'OPLB (<http://opus.nlpl.eu/OfisPublik.php>), le reste de traduction de logiciels ou de sous-titres. Il contient des erreurs d'alignement.

*Téléchargeable dans plusieurs formats.*

Jörg Tiedemann, 2012, Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*

### **Corpus de l'OPLB**

<https://github.com/Ofis-publik-ar-brezhoneg/breton-french-corpus>

Corpus de 62 861 phrases issues de mémoires de traduction de l'OPLB.

*Téléchargeable au format texte.*

### **Tatoeba**

<https://tatoeba.org/eng/downloads>

Fin 2024, près de 8 000 phrases multilingues du breton vers une autre langue.

*Disponible en ligne, téléchargeable.*

Licence Creative Commons CC-BY

On peut également noter l'existence d'autres corpus comme ceux créés à partir de la Déclaration universelle des droits de l'homme, de la bible, etc.

### 1.2.3 Corpus oraux

Pour construire des systèmes de **synthèse de la parole** ou de **reconnaissance de la parole**, il est nécessaire de disposer de données d'entraînement. Un **corpus oral** est un corpus constitué d'enregistrements accompagnés de leurs transcriptions. Il peut être monolingue ou multilingue. Seul le projet Common Voice a été conçu spécialement pour la reconnaissance vocale.

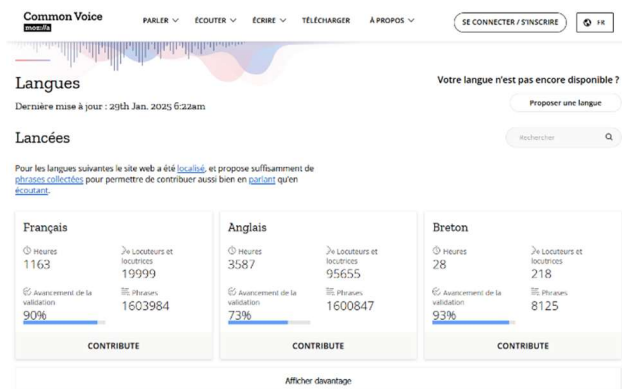
**Mozilla Common Voice** <https://commonvoice.mozilla.org/br>

Corpus de parole participatif. Fin 2024, 25 heures sont enregistrées et validées pour le breton (218 locuteurs différents) contre 12h en 2020.

*Téléchargeable sur le site Common Voice.*

Meyer, J., Morris, R., Saunders, L., Tyers, F. M. and Weber, G. (2020) "Common Voice: A Massively-Multilingual Speech Corpus".

*Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020). pp. 4211—4215*



Licence Creative Commons CC-0

**Corpus synthèse vocale** <https://github.com/Ofis-publik-ar-brezhoneg/audio-breton-corpus>

Corpus oral créé lors du projet de synthèse vocale. 20 heures d'enregistrement.

<b>Corpus oral Meurgorf</b>	<p><a href="https://github.com/Ofis-publik-ar-brezhoneg/phonetic-breton-corpus">https://github.com/Ofis-publik-ar-brezhoneg/phonetic-breton-corpus</a></p> <p>Enregistrement de 6 000 mots accompagnés de leurs transcriptions phonétiques. Les enregistrements peuvent également être écoutés sur la page de Meurgorf.</p>
<b>Banque Sonore des dialectes bretons</b>	<p><a href="http://banque.sonore.breton.free.fr/">http://banque.sonore.breton.free.fr/</a></p> <p>Ensemble d'enregistrements effectués auprès de locuteurs traditionnels du breton. En janvier 2025, il comprend 8 115 fichiers audio pour une durée totale de 18 heures d'enregistrements.</p> <p><i>Disponible en ligne et téléchargeable.</i></p> <p>Licence Creative Commons CC BY-NC-ND</p>
<b>Dictionnaires bretons parlants</b>	<p><a href="http://dico.parlant.breton.free.fr/index.htm">http://dico.parlant.breton.free.fr/index.htm</a></p> <p>Dictionnaire multimédia comprenant des enregistrements effectués auprès de locuteurs traditionnels du breton. Présente une dizaine de points de collecte au 4 novembre 2020.</p> <p><i>Disponible en ligne.</i></p>
<b>Corpus Dastum</b>	<p><a href="http://www.dastumedia.bzh/dyn/portal/index.seam?page=listalo&amp;firstResult=0&amp;fonds=&amp;fi=LANG&amp;va=BR&amp;aloid=53433&amp;cid=4161">http://www.dastumedia.bzh/dyn/portal/index.seam?page=listalo&amp;firstResult=0&amp;fonds=&amp;fi=LANG&amp;va=BR&amp;aloid=53433&amp;cid=4161</a></p> <p>Ensemble de corpus de langue bretonne présents dans la base de données de Dastum. Certains éléments ne sont pas transcrits et ne sont pas destinés à des traitements informatiques.</p> <p><i>Disponible en ligne et téléchargeable, après inscription gratuite.</i></p> <p>Utilisation soumise à autorisation de la part de Dastum.</p>

**Languesdebretagne**

<https://languesdebretagne.bzh/>

Enregistrements courts faits sur le vif par des brittophones. Le projet est participatif. Il est possible d'enregistrer et d'écouter sur le site ou via une application dédiée.

Projet développé par le centre culturel breton de Lannion.

**Lingua Libre**

[https://lingualibre.org/wiki/LinguaLibre:Main\\_Page](https://lingualibre.org/wiki/LinguaLibre:Main_Page)

Corpus oral (mots) participatif. Fin 2024, 3 450 enregistrements sont disponibles à l'écoute et au téléchargement.

Projet développé par l'association Wikimedia.

Licence Creative Commons CC BY-NC

**komzoubrezhoneg**

<https://komzoubrezhoneg.bzh/>

Enregistrements filmés accompagnés de leurs transcriptions. Les conversations ont été menées par Lors Jouin.

### 1.3 Grammaire et modèle de langue

Une langue peut être représentée sous forme numérique de deux façons, avec une grammaire formelle, qui décrit l'ensemble des règles grammaticales de la langue sous forme numérique, ou grâce à un modèle de langue, qui représente de manière statistique la langue.

	Quantité	Disponibilité	Qualité	Couverture	Durabilité	Adaptabilité
Grammaire formelle	Jaune	Jaune	Jaune	Jaune	Rose	Rose
Modèle de langue	Rose	Rose	Rose	Rose	Rose	Rose

Ces deux représentations sont utilisées pour reconnaître la langue, pour vérifier qu'une phrase fait partie de la langue ou encore pour faire de la correction automatique.

Il existe en breton des grammaires incorporées dans d'autres outils :

- ◆ Apertium, pour faire de la désambiguïsation
- ◆ LanguageTools pour l'autocorrection

Ces grammaires sont cependant limitées et difficilement exploitables en dehors des projets pour lesquels elles ont été conçues.

Le site DVB *displeger verboù brezhonek* (conjugaison des verbes bretons) (<https://displeger.bzh/fr>) a été créé en 2020. On y trouve les formes conjuguées des verbes dans l'ensemble des temps (forme standardisée et dialectale).

## 1.4 Ressources sémantiques et bases de connaissances

Cette ressource est de haut niveau, elle est utilisée pour des traitements sémantiques (désambiguïsation, analyse de sentiments...).

Il existe plusieurs types de ressources sémantiques :

- ◆ Les **Word Nets** sont des bases de données qui regroupent les mots par « *synsets* » (*synonym sets*), des groupes de mots interchangeables.
- ◆ Une **ontologie** est une base de données qui contient des relations entre les mots : hyperonymie, hyponymie, synonymie, antonymie, méronymie...
- ◆ Un **thésaurus** est une base de données qui regroupe les mots par champ lexical.

La base **babelNet** (<https://babelnet.org/>) est une base sémantique multilingue, créé par le département NPL de l'université Sapienza à Rome. Elle intègre des données provenant principalement de Wikipédia et Wordnet. Il est possible de rechercher des termes en breton, leurs définitions, leurs traductions et leurs liens avec d'autres termes.

Le site <https://geriadur-ar-sinonimou.net/> propose une base de synonymes en breton. Fin 2024, la base contient 193 entrées liées à 4 402 synonymes. On y trouve également des proverbes et des exemples de phrases.

## 2 Technologies de la langue

### 2.1 Analyse syntaxique, outils de base

Pour effectuer des traitements sur un texte, on commence généralement par utiliser des outils dits de pré-traitement : le **segmenteur** détache les mots les uns des autres et des signes de ponctuation, l'**étiqueteur morphosyntaxique** marque la **catégorie grammaticale** des mots et le **lemmatiseur** remplace les mots par leur **lemme** (infinitif du verbe, masculin singulier du nom...)

	Quantité	Disponibilité	Qualité	Couverture	Durabilité	Adaptabilité
Segmenteur						
Lemmatiseur						
Etiqueteur morphosyntaxique						

#### ◆ Lemmatiseur, segmenteur

Il n'existe pas de segmenteur prenant en compte les spécificités du breton et de son orthographe standard (le *peurunvan*) : prise en compte de « c'h » comme un seul caractère, prise en compte de l'élision des particules verbales...

Un lemmatiseur est utilisé dans le projet Apertium, sans qu'il soit possible de l'extraire pour d'autres usages.

#### ◆ Etiqueteur morphosyntaxique

Les correcteurs orthographiques comme LanguageTools ou encore le traducteur Apertium utilisent des étiqueteurs morphosyntaxiques intégrés. Cependant, il semble difficile de les utiliser pour un autre travail.

Toutefois, LanguageTools propose d'étendre l'utilisation de son outil. Outre LibreOffice mentionné précédemment le correcteur est disponible pour une utilisation dans d'autres éditeurs de textes : Microsoft Word, Google Docs, différents navigateurs et messageries électroniques (GoogleMail, Apple Mail, Thunderbird). Par exemple, l'outil peut être également utilisé pour corriger des textes scannés

via smartphones (iOS) car LanguageTools a développé une application pour ce système.

### **POS tagger Apertium**

<https://beta.apertium.org/index.fra.html#analysis?aLang=bre&aQ=>

Étiqueteur morphosyntaxique intégré dans le traducteur automatique Apertium.

*Disponible en ligne, téléchargeable sur le GitHub d'Apertium.*

Sheikh, Z.M.A.W. and Sánchez-Martínez, F. (2009) "A tri-gram part-of-speech tagger for the Apertium free/open-source machine translation platform". In *Proceedings of the First International Workshop on Free/Open-Source Rule-Based Machine Translation*, p. 67-74, Alacant, Spain.

Licence Creative Commons CC BY-SA

## **2.2 Analyse syntaxique**

Un analyseur syntaxique permet de représenter la structure de la phrase. Un analyseur syntaxique de surface reconnaît les syntagmes<sup>11</sup> nominaux, verbaux, etc. alors qu'un analyseur syntaxique profond peut construire l'arbre syntaxique complet de la phrase et indiquer le rôle de chaque mot.

Notons toutefois les travaux visant à créer des outils applicables à plusieurs langues. Parmi eux, le modèle **LIMA** (Libre Multilingual Analyzer) est un outil libre de droit permettant de réaliser des analyses de type :

- Morphologique (*morphological*) : analyse des morphèmes (forme minimum d'un mot douée de sens)
- Lexicale (*lexical*) : analyse des mots dans une phrase afin de les décomposer en lexème ou radical d'un mot.
- Syntaxique (syntactic) : analyse des phrases complètes ou partielles afin d'étudier la structure et leur sens. Il s'agit d'une analyse grammaticale, à savoir la "structure de la phrase"

---

<sup>11</sup> Un syntagme est un groupe de mots organisés autour d'un noyau. Ce noyau peut-être un verbe, un nom, etc.

- Sémantique (semantic) : analyse afin de comprendre le sens d'une phrase complète ou partielle (intention, message, sentiment exprimé).

LIMA peut être applicable à une soixantaine de langues. L'évaluation du modèle pour le breton a été faite avec le corpus breton d'Universal Dependencies breton-KEB (884 phrases, soit 10 000 mots). Il n'est pas parfait mais l'évaluation globale serait légèrement meilleure que celui du modèle d'Udify. L'outil atteint un score de 61,56 pour l'UAS (*unlabeled attachment score*) à savoir le pourcentage de mots correctement annotés (article, verbe, etc.). Concernant le LAS (*labeled attachment score*) à savoir le pourcentage de mots correctement annotés (article, verbe, etc.) mais aussi correctement associés les uns aux autres, le score s'élève à 41,48.

## 2.3 Analyse sémantique

Les analyseurs sémantiques sont utilisés pour représenter et désambiguïser le sens d'une phrase. Ils prennent le plus souvent la forme d'un arbre sémantique.

Comme mentionné précédemment, le modèle LIMA proposerait ce type d'analyse, mais il n'a pas été évalué pour le breton.

## 2.4 Extraction d'informations

Un système d'extraction d'informations permet de trouver les informations importantes dans un ensemble de données ou de les résumer. On l'utilise aussi pour catégoriser des textes ou pour repérer des comportements dangereux sur les réseaux sociaux.

Il n'existe pas de tel système pour le breton.

## 2.5 Génération de textes

Les systèmes de génération automatique de texte mettent sous forme lisible par les humains des informations ou des données. Ils sont par exemple utilisés pour des rapports météo, des moteurs de recherche, etc. À partir de ces systèmes, il a été créé des agents conversationnels (chatbot) plus efficaces.

Depuis 2022, chacun peut utiliser l'Intelligence artificielle ChatGPT. Il est possible d'interagir en breton et l'agent répond en breton. Il est possible de lui demander de générer des textes en breton, mais la qualité de la réponse demeure aléatoire, tant au niveau de la syntaxe de la langue que du contenu culturel. Depuis lors, différents agents conversationnels sont disponibles.

Il est possible de trouver des intelligences conversationnelles alimentées par des ressources sélectionnées plus spécifiquement. Par exemple, le système est une IA basée sur les écrits de Jean Pierre Le Mat. Il est possible de poser des questions en breton et l'IA répond dans cette langue.

**ChatGPT.bzh** est présenté comme une intelligence artificielle pouvant répondre en breton, mais la qualité de langue semble très faible.

## 2.6 Reconnaissance de l'écriture

Ce domaine est très utile pour la transcription sous forme numérique de textes anciens, manuscrits ou imprimés. Nous pouvons mentionner :

- **Transkribus.** Ce site propose des outils (utilisation gratuite jusqu'à 50 pages par mois) afin de reconnaître des textes. Il est possible d'adapter/entraîner des modèles selon les besoins précis des utilisateurs. Par exemple, Herve Baudry et Kristian Hamon ont entraîné un modèle destiné à lire d'anciens écrits en langue bretonne (XVII<sup>e</sup>-XIX<sup>e</sup> siècle). ([https://www.transkribus.org/model/breton-prints-17<sup>th</sup>-19<sup>th</sup>-centuries](https://www.transkribus.org/model/breton-prints-17th-19th-centuries)).
- **Tesseract** : logiciel libre de reconnaissance d'écriture (formats lus : png; jpeg; tiff; gif, etc.). Le logiciel pourrait reconnaître plus de 100 langues, dont le breton.<sup>12</sup> (<https://github.com/tesseract-ocr/tesseract>) (<https://tesseract-ocr.github.io/tessdoc/Data-Files-in-different-versions.html>). Toutefois, les résultats semblent plus probants lorsque l'on sélectionne une autre langue pour reconnaître des textes en breton.
- L'outil **capture d'écran de Windows** ou l'outil **photos de Windows** : ces outils intégrés à windows proposent une fonction de reconnaissance optique de caractères d'une qualité satisfaisante.

---

<sup>12</sup>Cet OCR est également disponible sur le site <https://abp.bzh/ocr/OCR.php>

## 2.7 Traduction automatique

Il existe aujourd'hui trois types de traducteurs automatiques. Les premiers fonctionnent à base de règles de traduction et de dictionnaires bilingues. Ils nécessitent moins de ressources. Les deux autres types de traducteurs, dits statistiques ou neuronaux, ont besoin d'une grande quantité de textes bilingues alignés ou unilingues pour s'entraîner et apprendre.

### 2.7.1 Traducteur automatique avancé (statistique, neuronal, hybride)

Ces traducteurs sont assez récents. Ils sont de plus en plus utilisés et doivent être améliorés. En effet, les résultats peuvent parfois être mal interprétés par des utilisateurs ne pouvant évaluer la qualité des traductions fournies.

#### Google

<https://translate.google.com/?hl=fr&sl=fr&tl=br&text=Mon%20papa%20est%20italien.%20&op=translate>

En juin 2024, Google a ajouté 110 langues à son traducteur automatique dont le breton.

#### Le traducteur de l'OPLB

<https://niverel.brezhoneg.bzh/br/troer/>

Traducteur français-breton et breton-français disponible sur le site *porched an niverel* depuis octobre 2023.

Ce traducteur est développé par l'OPLB et Alan Entem. Des textes d'une longueur maximum de 5 000 caractères peuvent être traduits.

The screenshot shows the website 'Porched niverel ar brezhoneg' (OPLB Breton Public). The main content area displays a 'Donemat!' (Welcome!) message. Below this, there is a 'Moned prim' (Primary menu) section with three columns of text, each ending with a 'Moned' (Go) button. The columns contain information about the website's mission, the OPLB, and the Meurgor (Breton) language.

MinT est un outil développé par Wikimedia proposant un traducteur basé sur plusieurs modèles de traduction. Le breton figure parmi la liste des langues proposées mais parfois l'outil ne fonctionne pas correctement lorsque l'on cherche à traduire des textes.

Il faut également indiquer les agents conversationnels qui peuvent être utilisés pour réaliser des traductions. La qualité des résultats demeure aléatoire pour le breton. Parmi les agents conversationnels, mentionnons :

- ChatGPT (développé par OpenAI, États-Unis) ;
- Le Chat (développé par Mistral AI, France) ;
- Claude (développé par Anthropic, États-Unis) ;
- Gemini (ex-Bard, développé par Google, États-Unis);
- Perplexity (États-Unis);
- DeepSeek (Chine)

## 2.7.2 Traducteurs automatiques de base

**Apertium** <https://www.apertium.org/index.fra.html?dir=bre-fra#translation>

<https://github.com/apertium/apertium-br-fr>

Traducteur automatique construit à l'aide de règles et dictionnaires. Toujours en développement.

*Disponible en ligne, code téléchargeable sur GitHub*

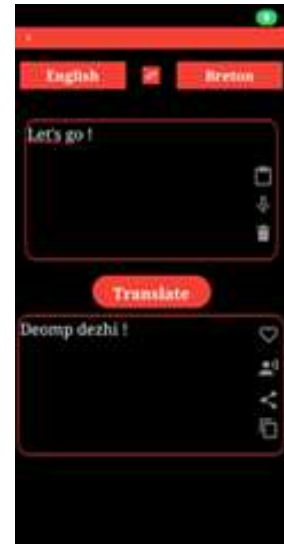
Tyers, F. M. 2010. 'Rule-based Breton to French machine translation', in *Proceedings of the 14th Annual Conference of the European Association of Machine Translation*, 174-181

L'OPLB avait aidé à son développement.

Licence Creative Commons CC BY-SA

Il existe une application (incluant de la publicité) **Breton translator** proposé par World Language Translator. La qualité de l'outil n'a pas été évaluée.

<https://play.google.com/store/apps/details?id=com.bj.translatorswedish>



## 2.8 Traitement de la parole

### 2.8.1 Synthèse vocale

L'outil de synthèse vocale est disponible sur le site *Porched an niverel* de l'OPLB depuis octobre 2023. Il s'agit d'un projet mené par l'IRISA. Il est possible de synthétiser des textes de 1 000 caractères maximum. Cet outil peut aider les personnes ayant des difficultés de lecture (dyslexie, problèmes de vue, etc.). Le département du Finistère par exemple dénombre environ une centaine d'élèves dans les filières bilingues qui pourrait bénéficier de cet outil.

### 2.8.2 Reconnaissance vocale

**Anaouder** est un outil de reconnaissance vocale développé par Gweltaz Duval-Gwennog, à partir de l'outil open-source Kaldi.

(<https://pypi.org/project/anaouder/>)

Anaouder est accessible sur le site <https://abp.bzh/anaouder/>. Il est possible de retranscrire un enregistrement son, vidéo ou un enregistrement réalisé directement sur le site. Un autre outil permet de créer des sous-titres au format adéquat (srt).



**Whisper** est un modèle de reconnaissance automatique de la parole développé par OpenAI et proposé en open-source en 2022. Le modèle a été entraîné sur 98 langues, mais ne serait efficace que pour 57 d'entre elles. (moins de 50% d'erreurs, "Word error rate"). Parmi ces langues se trouve le catalan, le galicien et le gallois, mais pas le breton.

(<https://platform.openai.com/docs/guides/speech-to-text/supported-languages>)

## 2.9 Autres outils

Le site *porched niverel ar brezhoneg* (<https://niverel.brezhoneg.bzh/fr/home/>) propose d'autres outils comme :

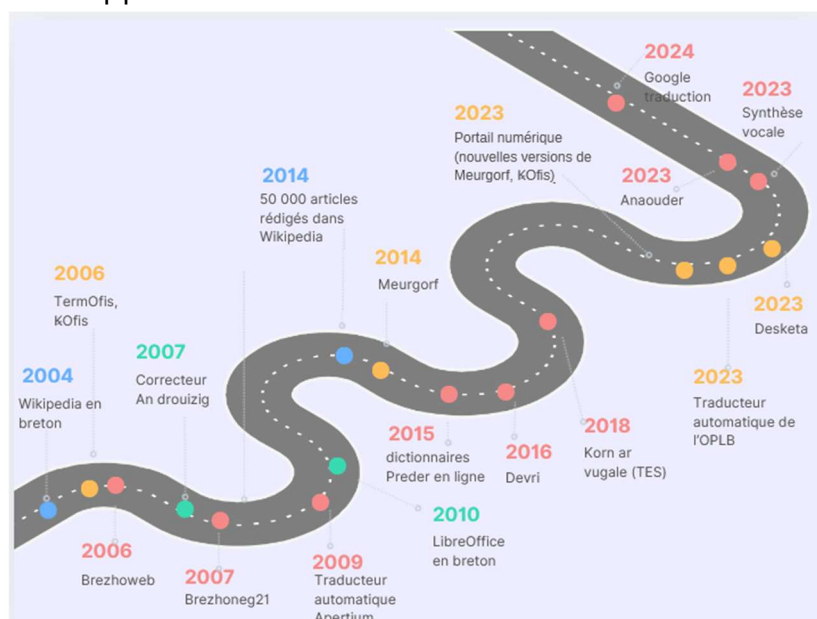
- un **phonétiseur** permettant la transcription de phrases en alphabet phonétique international (API).
- un **syllabifieur** permet la division de phrases en syllabes et aide au placement de l'accent tonique.
- un **normaliseur**. Cet outil permet de retranscrire en breton et en toutes lettres des symboles ou chiffres.
- un **transcripteur Braille** permettant de transcrire des phrases rédigées en breton en écriture Braille. L'outil a été développé grâce aux travaux de Harris Mowbray. Il reste quelques adaptations à faire par rapport à la transcription des chiffres et des dates.

### 3 Conclusion

De nombreux outils et ressources sont disponibles pour le breton, sur Internet. Cependant, beaucoup d'entre eux sont vieillissants et ne sont plus mis à jour. Entre ceux qui ont été abandonnés et ceux dont l'existence reste inconnue du plus grand nombre, il est parfois difficile de trouver l'outil que l'on recherche. Dans ce diagnostic, nous n'avons listé que les outils qui étaient encore en état de fonctionner.

Du point de vue des ressources linguistiques, nous disposons d'une bonne base de départ mais il devient urgent de travailler tant sur l'augmentation de la quantité de ressources disponibles que sur leur qualité et leur diversification. En 2020, il n'existait aucun traducteur automatique avancé, aujourd'hui, il en existe plusieurs. Des outils de synthèse vocale et de reconnaissance vocale sont disponibles depuis 2023. Depuis 2024, il existe une plateforme gratuite d'autoapprentissage du breton.

Schéma du développement du breton dans le domaine du numérique 2004-2024



## ❖ Partie 3. Préconisations

---

À la suite du diagnostic fait précédemment, nous pouvons établir une liste de préconisations à l'ensemble des acteurs engagés dans les technologies de la langue pour le breton. Ces différentes actions seront à répartir entre les différents acteurs suivant leurs compétences et leurs moyens d'action.

### 1 Augmenter sensiblement la visibilité du breton dans le monde numérique

#### 1.1 Les contenus en langue bretonne sur Internet

La visibilité du breton passe d'abord par la disponibilité de contenus utiles et en plus grande quantité sur les principaux sites d'Internet. Parmi eux, l'un des plus consultés est Wikipédia.

Il paraît également nécessaire que les structures qui proposent des sites bilingues français-breton mettent mieux en avant l'utilisation du breton et facilitent la navigation dans un environnement entièrement en langue bretonne.

- ◆ **Sensibiliser le public brittophone** à l'utilisation de Wikipédia en breton et continuer à l'enrichir.
- ◆ Encourager la traduction de **sites culturels** (cinéma, musique, sport...) ou **d'actualité** (presse régionale, portails d'informations, météo, horaires des marées...) qui sont les plus consultés par le public.
- ◆ **Former et coordonner** les différents acteurs qui prennent part à la traduction de contenu et de sites Internet.
- ◆ Veiller à la **qualité de la langue** utilisée (orthographe, syntaxe, terminologies adaptées).
- ◆ Publier un guide des **bonnes pratiques de traduction** pour aider à mettre en place et à mettre en valeur la langue bretonne sur les sites bilingues français/breton. Ce travail est entamé par l'OPLB. Le site de l'office met en ligne les recommandations du conseil scientifique concernant l'orthographe, le style et la typographie. (<https://www.brezhoneg.bzh/177-divizou-hag-erbedadennou-ar-chuzul-skiantel.htm>)

#### 1.2 L'utilisation du breton sur les réseaux sociaux

Même si l'utilisation du breton sur les réseaux sociaux est de plus en plus fréquente, elle reste cependant trop rare, même pour des utilisateurs qui parlent

couramment la langue. La première raison qui pourrait expliquer une si faible utilisation du breton est l'absence d'interface bretonne pour la plupart des sites et applications, disponibles en français et en anglais. L'environnement dans lequel évoluent les utilisateurs n'est pas fait pour les inciter à utiliser d'autres langues.

D'autres hypothèses qui peuvent être avancées sont l'isolement des locuteurs entre eux ainsi que l'insécurité linguistique des locuteurs quant à leurs compétences linguistiques. Sur la plupart des réseaux, il n'existe pas de communauté brittophone ni de moyen de rencontrer d'autres brittophones, ni même d'identifier des brittophones.

- ◆ Continuer à **encourager les utilisateurs** à employer le breton sur les médias sociaux en proposant des termes en lien avec l'actualité.
- ◆ **Participer à la traduction** des plateformes de réseaux sociaux quand c'est possible ou inciter à leur traduction.



### 1.3 Des outils pour l'apprentissage et la mise en valeur du breton

L'une des plus fortes demandes de la part du grand public est le développement d'applications et d'une plateforme web permettant l'apprentissage et le perfectionnement en langue bretonne. C'est une demande qui vient à la fois du public non-brittophone que du public déjà brittophone. L'enquête TMO-Région de 2024 a montré que 33 % des Bretons qui ne parlent pas la langue souhaiterait la connaître (30 % en 2018).

Pour être efficaces, les outils d'apprentissage du breton doivent intégrer une réelle mise en valeur de la langue et utiliser les dernières innovations technologiques, aussi bien pour ce qui est du contenu que de la façon de la présenter :

- ◆ **Continuer à développer la plateforme d'auto-apprentissage Desketa** en proposant des cours de niveau avancé.
- ◆ **Participer à la traduction des logiciels pédagogiques** utilisés dans les écoles (traitement de texte, système d'exploitation bilingues...).

- ◆ **Mener des actions en faveur de la présence du breton** sur DuoLinguo ou d'autres applications d'initiation ou apprentissages de nouvelles langues.
- ◆ **Développer et proposer du contenu pédagogique moderne** accessible à tous pour apprendre ou approfondir ses connaissances sur la langue bretonne. Le projet MediaTES développé par TES, devrait être en ligne en 2025.

## 1.4 La traduction d'applications

À l'heure où les dispositifs numériques occupent une place centrale dans nos vies, que ce soit au travail, à la maison ou lors de nos sorties, il est essentiel de les prendre en compte dans la constitution d'un environnement brittophone.

Il est aujourd'hui difficile, voire impossible, d'activer son GPS en breton, d'utiliser son smartphone en breton, de naviguer sur son ordinateur en breton ou d'utiliser des bornes multimédia en breton, dans les gares ou les banques.

- ◆ Contacter Microsoft pour la **traduction de leur suite logicielle** bureautique Office.
- ◆ Contacter les développeurs des **principaux systèmes d'exploitation** pour ordinateur (**Windows, Mac Os**) et pour smartphone (**iOS, Android**), afin d'étudier la faisabilité de versions bretonnes
- ◆ Participer à la **traduction collaborative en ligne** d'applications majeures.
- ◆ Proposer des solutions pour **utiliser son téléphone en breton** : clavier, messagerie, navigation, GPS, application...
- ◆ **Sensibiliser les grandes entreprises** afin qu'elles proposent leurs services numériques en langue bretonne.

## 2 Encourager l'innovation numérique au service de la langue bretonne

### 2.1 Mettre en réseau les différents acteurs

La Bretagne dispose sur son territoire de nombreux laboratoires et centres de recherche en linguistique et en informatique, dont plusieurs spécialisés en traitement automatique de la langue : les équipes de l'IRISA à Rennes et Lannion, le LS2N à Nantes... Le CNRS travaille également sur le breton dans le monde numérique.

Cependant, les acteurs travaillant sur la langue bretonne dans le domaine numérique sont aujourd'hui peu nombreux. On peut compter parmi eux l'OPLB, qui met

à disposition sur son site dédié au numérique plusieurs outils qu'il a développé, ainsi que des associations comme An Drouzig qui propose des correcteurs orthographiques et des traductions de logiciels, ou encore OpenStreetMap.bzh qui propose un système de cartographie numérique entièrement en breton. Le fonds de dotation Bretagne numérique créé par l'association pik.bzh s'attache à ce que le breton soit intégré parmi les principaux traducteurs automatiques ainsi qu'au développement de contenu breton.

Une première initiative a permis de mettre en relation ces différents acteurs via le débat « Nouvelles technologies et breton » organisé par TermBret en 2006 puis par le réseau *Bed Niverel* en 2017. À l'époque avaient pris part à la rencontre l'OPLB, An Drouzig, OpenStreetMap.bzh, Roued, EduBreizh et des informaticiens d'Arkea. Même si cette rencontre n'a pas donné lieu à de nouvelles coopérations autour de projets numériques, elle reste néanmoins une base intéressante pour la constitution d'un réseau plus actif, en langue bretonne.

Nous proposons qu'une future rencontre sur un modèle similaire soit organisée afin de faire le point sur les projets des différentes structures, leur avancement. Aussi, chacune de ces structures dispose de ressources linguistiques qui pourraient être mises en commun. L'OPLB pourrait avoir un rôle de coordination des projets et des acteurs.

Cette rencontre permettrait d'envisager les bases d'une coordination plus étroite entre les acteurs pour la mise en place de nouvelles ressources et de nouveaux outils, en se basant sur le plan d'action détaillé en partie 3 de ce dossier.

- ◆ **Organiser des rencontres** entre les acteurs du numérique brittophone, en reprenant par exemple l'initiative Bed Niverel.
- ◆ Coordonner la mise en place du **plan d'action pluriannuel**.
- ◆ **Construire des liens** avec le monde universitaire et de la recherche.

## 2.2 Publier des données de qualité

Les données sont la matière première dans les processus de développement de technologies de traitement des langues naturelles. Elles sont utilisées à la fois pour alimenter des algorithmes et des intelligences artificielles, pour étudier des phénomènes linguistiques ou pour tester les performances d'un programme de traitement de la langue.

Les données peuvent être de plusieurs natures : sémantiques, grammaticales, textuelles, audio, vidéo, multimodales... Elles sont organisées différemment et parfois enrichies d'informations supplémentaires (annotations, métadonnées).

Aujourd'hui, la quantité et la qualité des données numérisées disponibles en langue bretonne est problématique. Il n'existe à ce jour pas de corpus de référence pour le breton, constitué de textes annotés lexicalement et syntaxiquement, qui puisse servir de base de référence pour les linguistes et les informaticiens. De plus, il n'existe aucun corpus dont la qualité ait été évaluée.

L'OPLB dispose cependant d'une assez grande quantité de donnée de diverses natures : des corpus de textes bilingues, un dictionnaire historique, une base de données terminologiques ainsi qu'une base de données toponymiques. Les bases KerOfis et Meurgorf ont été mises à jour. Il reste à ordonner les autres données pour une utilisation informatisée.

Pour remédier à la dispersion des données, l'OPLB a ouvert un portail web des technologies numériques pour la langue bretonne. Il accueille les données et les outils de l'OPLB mais a aussi vocation à accueillir les outils et ressources linguistiques numériques de ceux qui le souhaitent.

- ◆ Faire appel à des **spécialistes en linguistique de corpus** pour créer de corpus de haute qualité exploitables de manière informatique.
- ◆ Rendre les données présentes dans les bases Meurgorf, TermOfis et KerOfis **accessibles via une API**.
- ◆ Évaluer la **qualité des données** disponibles pour le breton.

## 2.3 Financer l'innovation en breton

L'un des leviers sur lesquels il est possible d'agir pour encourager le développement de nouvelles technologies de la langue pour le breton est le financement de la recherche. Deux dispositifs *Stlenn* et *Arload* ont été créés par le Conseil régional de Bretagne afin d'encourager la traduction numérique et la création d'applications.

- ◆ **Encourager la mise en place d'une bourse** publique spécifique pour les étudiants ou les doctorants qui mènent un projet liant numérique et langue bretonne, ou permettant la création de nouvelles ressources linguistiques.

## 2.4 Promouvoir les licences libres de droit

Les différentes applications et données développées pour le breton doivent être publiées avec des licences claires qui définissent leurs conditions d'utilisation et de modification. Dans l'optique de leur utilisation par le plus grand nombre et pour qu'elles servent de base à des développements plus poussés par des entreprises,

les licences libres de droits avec droit d'utilisation commerciales sont les plus intéressantes.

- ◆ Placer l'ensemble des outils et ressources de l'OPLB sous **licences libres de droits** ouvrant le droit à la modification pour usage non-commercial ou commercial.
- ◆ Veiller à ce que les projets financés par l'argent public soient **librement accessibles au plus grand nombre** et restent libres de droits.

## ❖ Partie 4. Plan d'action

---

L'amélioration de la place de la langue bretonne dans le monde numérique et le développement de nouvelles applications grand public passent aussi par l'amélioration ou le développement de nouveaux outils et ressources à destination des développeurs, chercheurs et linguistes.

### 1 Ressources linguistiques

Le tableau ci-dessous présente l'état actuel des ressources linguistiques disponibles pour le breton et leur état souhaité dans les 9 prochaines années.

La couleur **verte** signifie que les ressources disponibles sont satisfaisantes, la couleur **orange** signifie que les ressources doivent être améliorées ou enrichies, ou qu'elles sont en cours de développement ; la couleur **rouge** signifie que les ressources sont manquantes ou insuffisantes, ou que leur développement n'est pas encore lancé.

Ressources		2023	2024-2027	2028-2030
Corpus textuels monolingues	Web	Verte	Verte	Verte
	Langue	Orange	Verte	Verte
	Arboré	Rouge	Rouge	Orange
Corpus textuels bilingues		Orange	Verte	Verte
Corpus oraux		Orange	Orange	Verte
Grammaires		Orange	Orange	Verte
Modèles de langue		Orange	Orange	Verte
Lexiques		Verte	Verte	Verte
Dictionnaires		Orange	Verte	Verte
Bases terminologiques		Verte	Verte	Verte
Bases toponymiques		Orange	Verte	Verte
Bases sémantiques		Rouge	Orange	Orange

## 1.1 Corpus

L'OPLB a entamé la création d'un corpus monolingue à partir des articles du journal *Ya!*. Ce corpus pourrait être enrichi par d'autres écrits. Pour alimenter le corpus, on peut envisager de collecter des textes auprès des maisons d'éditions, des associations ou parmi les exemples historiques de la base Meurgorf. Les différents textes de ce corpus devront être catégorisés suivant leur genre, la qualité de la langue et la date de publication. Pour assurer la bonne qualité et la pertinence de ce corpus, il devra être contrôlé par des linguistes. L'objectif est d'atteindre un corpus de **20 millions de mots** courant 2026-2030.

L'OPLB utilise depuis 2023, une mémoire de traduction permettant aux traducteurs d'enregistrer dans une base commune les traductions réalisées. Ce corpus bilingue contient actuellement 60 000 phrases et augmente régulièrement grâce aux nouvelles traductions, mais également par l'ajout d'anciennes traductions. L'objectif est d'atteindre **15-20 millions de mots** d'ici à 2027-2029.

Le **corpus oral** Common Voice doit encore croître sensiblement pour permettre d'améliorer les outils de reconnaissance vocale ainsi que d'accroître les corpus libres de droit pouvant être utilisé par les développeurs ou chercheurs. Des outils de reconnaissance vocale efficaces permettent par la suite de développer des commandes vocales utilisables dans des navigateurs web, des logiciels informatiques, des smartphones... Le corpus Common Voice devra être alimenté par une grande diversité de locuteurs. L'objectif est d'atteindre 100 000 phrases enregistrées d'ici à 2026-2027.

La constitution d'un **corpus de référence** prend plus de temps, car il doit être annoté à la main. Il reste néanmoins très important pour le développement d'outils linguistiques performants, de traduction automatique ou de reconnaissance de la parole par exemple. À l'horizon 2027-2029 nous pouvons fixer l'objectif de **30 000 à 50 000 mots** (actuellement près de 10 000 mots)

## 1.2 Grammaires et modèles de langue

Concernant les modèles de langue, le préalable est la constitution d'un corpus monolingue de taille suffisante. Le travail doit donc d'abord s'orienter vers l'élaboration de ce type de corpus.

De plus en plus d'outils sont élaborés, basés sur des de grand modèle de langage et utilisables pour différentes tâches : questions-réponses; analyse des sentiments, extraire des informations, générer des textes, correction orthographique et grammaticale, synthèse de textes... Si ce type de modèle peut reconnaître le

breton et l'utiliser, la qualité de la langue demeure toutefois limitée. Il est ainsi nécessaire de développer un grand corpus diversifié en langue bretonne. Par la suite ces modèles peuvent être affinés pour le breton grâce aux méthodes de "fine-tuning".

### **1.3 Dictionnaires et ressources lexicales**

Les dictionnaires, lexiques et bases de données terminologiques devront continuer à être enrichis de nouvelles entrées (mots et unités de traduction). Meurgorf intègre maintenant la transcription phonétique, la prononciation enregistrée ainsi que les liens de parentés entre les termes. Il est envisageable d'intégrer des données concernant la terminologie, les synonymes, les antonymes. Ce travail pourra servir de base à la création d'une base sémantique puis une ontologie, un système permettant de classer les mots de la langue suivant leurs sens. Il est aussi nécessaire de créer un dictionnaire bilingue/multilingue de référence pour le breton.

La base terminologique TermOfis continue d'être enrichie et l'outil devra être modernisé afin de faciliter son utilisation.

L'enrichissement des bases toponymiques doit se poursuivre. Une base de données anthroponymique pourrait être créée également, regroupant prénoms, noms et personnages historiques.

## 2 Technologies de la langue

Comme pour les ressources linguistiques, le tableau suivant présente l'état actuel de la disponibilité d'outils linguistiques pour le breton et leur état souhaité dans les 9 prochaines années.

La couleur **verte** signifie que les outils disponibles sont satisfaisants, la couleur **orange** signifie que les outils doivent être améliorés ou enrichis, ou qu'ils sont en cours de développement ; la couleur **rouge** signifie que les outils sont manquants ou insuffisants, ou que leur développement n'est pas encore lancé.

Technologies de la langue	2023	2024-2027	2028-2030
Analyse morphosyntaxique	orange	orange	orange
Analyse syntaxique	rouge	orange	orange
Analyse sémantique	rouge	orange	orange
Traduction automatique	verte	verte	verte
Synthèse de la parole	verte	verte	verte
Reconnaissance de la parole	orange	orange	orange
Extraction d'informations	orange	orange	orange
Génération de textes	rouge	orange	orange

### 2.1 Outils d'analyse

Il existe plusieurs correcteurs orthographiques pour le breton, mais il n'existe aucun outil d'analyse morphosyntaxique (étiqueteurs grammaticaux) autonome et approprié pour étiqueter un texte ou faire des pré-traitements (lemmatisation, tokenisation). Il est difficile de savoir si ces outils sont utiles, mais c'est pourtant une brique de base malgré les avancées technologiques dans le traitement automatique de la langue. Il faut donc prendre en compte ce besoin pour les années à venir (développer un outil d'analyse syntaxique superficiel et/ou un outil d'analyse syntaxique profond, qui peut être créé à partir de données et grâce à l'intelligence artificielle à condition qu'il existe un corpus syntaxique).

## 2.2 Traduction automatique

Il existe au moins deux traducteurs proposés, l'un par l'OPLB et l'autre par Google. Celui de l'OPLB a été entraîné sur un corpus restreint et nécessite d'être entraîné à nouveau sur un corpus plus large et diversifié pour améliorer sa qualité. Cet objectif pourra être atteint en 2025-2026.

## 2.3 Traitement de la parole

La synthèse vocale, projet mené par l'OPLB et l'IRISA est disponible depuis 2023. Elle permet de générer des paroles avec 3 voix différentes.

Concernant la reconnaissance vocale, l'outil "Anaouder" développé par Gweltaz Duval-Gwennog est disponible. L'outil est capable de retranscrire des vidéos ou des enregistrements sonores même si certains accents ne sont pas encore bien reconnus. Ici aussi, l'objectif est d'accroître les corpus oraux.

Il est important de continuer à développer des corpus diversifiés. Cela permettrait de développer des outils plus efficaces dans les années à venir.

### 3 Mise en œuvre du plan d'action

#### 3.1 Ressources matérielles, humaines et financières

Le tableau suivant permet d'évaluer la quantité et le type de ressources nécessaires à réalisation de chaque ressource ou outil. Il est présenté à titre indicatif, et son contenu peut varier suivant les structures en charge du développement de certains outils ou ressources, le recours au bénévolat, l'utilisation de l'existant, les budgets dégagés, etc.

	<b>Ressources matérielles</b>	<b>Ressources humaines</b>	<b>Ressources financières</b>
<b>Constitution de corpus écrits</b>	<p>Des ensembles de textes du domaine public ou dont les droits ont été achetés ou légués.</p> <p>Des mémoires de traductions.</p> <p>Logiciels d'alignements de textes.</p>	<p>Des techniciens ayant été formés ou possédant des connaissances à la fois sur la langue bretonne et la linguistique de corpus.</p> <p>Un travail de constitution de corpus peut être mené par une équipe de deux ou trois personnes.</p>	<p>L'acquisition de droits pour des textes peut s'élever à plusieurs dizaines de milliers d'euros.</p> <p>La constitution d'un corpus nécessite relativement peu de temps, mais dépend du besoin de transcription de texte en version numérique ou dans l'orthographe standard.</p> <p>Le financement doit se faire pour un travail qui peut s'étendre sur une période plus ou moins longue.</p>

<b>Annotation de corpus</b>	<p>Des corpus textuels.</p> <p>Un logiciel d'annotation de texte.</p>	<p>Des techniciens maîtrisant la langue bretonne.</p>	<p>L'annotation de corpus est un travail chronophage et donc coûteux. Il faut compter environ 150 à 200 heures pour l'annotation d'un corpus de 50 000 mots.</p>
<b>Enregistrement d'un corpus de parole</b>	<p>Un studio d'enregistrement ou</p> <p>une plateforme d'enregistrement en ligne comme celle de CommonVoice.</p>	<p>Des locuteurs, qui peuvent être des bénévoles.</p> <p>Un ingénieur capable de paramétrer l'enregistrement et de traiter les données obtenues.</p>	<p>Pour une heure d'enregistrement, il faut compter environ 1h à 1h30 de traitement dans le cas d'un corpus de synthèse de la parole, mais beaucoup moins dans le cas d'un corpus destiné à la reconnaissance vocale.</p>
<b>Rédaction de grammaires et conception de modèle de langue</b>	<p>Ressources grammaticales.</p> <p>Corpus de textes conséquent.</p>	<p>Un spécialiste de la grammaire bretonne, capable de la décrire sous forme de grammaire formelle.</p> <p>Un spécialiste du traitement des langues, capable de transposer ces règles sous forme numérique ou de générer un modèle de langue.</p>	<p>Le développement de grammaires formelles ou de modèles de langue peuvent prendre environ 3 à 6 mois (ou plus s'ils ne sont pas développés par des spécialistes, difficiles à trouver pour le breton)</p>

<p><b>Élaboration d'un dictionnaire bilingue</b></p>	<p>Ressources lexicales.</p>	<p>Un développeur informatique.</p> <p>Une équipe spécialisée pour alimenter le dictionnaire.</p>	<p>Le développement de la structure d'un dictionnaire peut faire l'objet d'un poste de travail pour une durée de 1 à 3 mois.</p> <p>Le développement de bases sémantiques peut faire l'objet d'un travail de thèse ou équivalent.</p> <p>Les dictionnaires et bases sémantiques doivent ensuite faire l'objet d'une alimentation et d'une maintenance régulière sur des années.</p>
<p><b>Des outils d'analyse de texte</b></p>	<p>Des corpus de textes annotés.</p> <p>Des grammaires et modèles de langue.</p>	<p>Chercheurs ou développeurs spécialisés dans le traitement automatique des langues.</p>	<p>Le développement d'outils d'annotation automatique grammatical ou syntaxique peut faire l'objet d'un poste sur environ 6 mois.</p>
<p><b>Système de traduction automatique avancé</b></p>	<p>Corpus bilingue parallèle français-breton.</p> <p>Des ressources informatiques assez puissantes per-</p>	<p>Chercheurs ou développeurs spécialisés dans le traitement automatique des langues.</p>	<p>Un tel système peut prendre plusieurs années (1 à 3 ans) pour une première version, et mobiliser une</p>

	mettant d'entraîner des modèles d'IA.		<p>équipe de recherche ou de développement.</p> <p>Pour l'amélioration d'un modèle pré-existant, quelques mois sont nécessaires pour le mettre à jour.</p> <p>Il est possible d'évaluer à environ 30 000 € le coût pour améliorer un modèle existant.</p>
<b>Reconnaissance vocale</b>	<p>Corpus oraux pour la reconnaissance vocale.</p> <p>Des ressources informatiques assez puissantes permettant d'entraîner des modèles d'IA.</p>	Chercheurs ou développeurs spécialisés dans le traitement automatique des langues et en traitement du signal audio.	<p>Un tel système peut prendre plusieurs années (2 à 3 ans) pour une première version, et mobiliser une équipe de recherche ou de développement.</p> <p>Jusqu'à récemment, il fallait compter entre 150 000 et 200 000 euros par année de développement.</p>

### 3.2 Acteurs concernés

Ce plan d'action repose sur les ressources et outils existants aujourd'hui en langue bretonne. Ils offrent un bon point de départ pour les développements futurs.

Le bon déroulement de cette stratégie demande un effort de coordination entre les différents acteurs du numérique en langue bretonne, de la part de ceux qui

sont déjà impliqués comme de celle de ceux qui sont amenés à s'impliquer à l'avenir.

Il demande en outre un investissement en personnes, temps et crédits de la part des universités et instituts de recherche qui aujourd'hui ne travaillent pas suffisamment dans le domaine du traitement automatique de la langue bretonne.

Nous appelons dès à présent l'ensemble des acteurs développant des compétences et souhaitant participer à un effort commun pour traduire dans les faits cette stratégie à entrer en contact avec l'OPLB.

Enfin la mise en place de ce plan repose également sur un investissement important des pouvoirs publics bretons en vue de développer la place de la langue dans le numérique.

# Annexe

Logiciels ou projets numériques traduits ou partiellement traduits en breton. La liste n'est pas exhaustive.

**Collabora Online** : une suite bureautique en ligne permettant le travail collaboratif sur des dossiers partagés.

**CodevTT** : outil de gestion de projet.

**Blockly** : bibliothèque de codes/scripts aidant au développement de sites.

**Diaspora\*** : réseau social open-source, le réseau n'est la propriété d'aucune personne unique.

**Dissemin** : outil permettant de vérifier que l'on peut retrouver en ligne des articles librement accessibles de chercheurs.

**Drupal** : outil pour gérer des sites Internet.

**Etherpad** : outil de collaboration en ligne.

**FUDforum** : logiciel pour créer des forums en ligne.

**Inaturalist** : site permettant le partage de photographies d'animaux, de plantes.

**Lichess** : site de jeux d'échecs.

**MantisBT** : outil de gestion des bugs pour les développeurs.

**Pidgin** : logiciel de messagerie instantanée.

**Pootle** : logiciel de gestion de projets de traduction en ligne.

**Wordpress** : logiciel pour créer des sites Internet.

**Waymarked Trails** : site d'itinéraires de randonnées à pied, à vélo, etc. Les cartes sont issues d'OpenStreetMap.