



HAL
open science

Breton-français et numérique, projet LangNum-br-fr (phase conception)

Annie Foret

► **To cite this version:**

Annie Foret. Breton-français et numérique, projet LangNum-br-fr (phase conception). Conférence Langues et numérique 2018, Jul 2018, Paris, France. hal-01959451

HAL Id: hal-01959451

<https://hal.inria.fr/hal-01959451>

Submitted on 18 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Breton-français et numérique projet LangNum-br-fr (phase conception)

Annie Foret
Univ Rennes, CNRS,
IRISA
F-35000 Rennes
annie.foret@irisa.fr

Le projet LangNum-br-fr¹ concerne la paire de langues français-breton et le numérique, dans un cadre pluridisciplinaire. Il regroupe des informaticiens spécialistes en traitement automatique des langues (Univ. Rennes & IRISA, LIG Grenoble, Univ Tours), des linguistes spécialistes des langues celtiques (CRBC) et des spécialistes des usages des TIC² (LOUSTIC Rennes2), en vue de valoriser des ressources pour le breton et avec une préoccupation pédagogique. Une analyse des besoins des apprenants en est la première étape avant la définition de traitements logiciels pour des meilleurs et nouveaux usages.

Comment va la langue bretonne ? Avec 200 000 locuteurs³, la langue bretonne est dans une situation précaire. La filière bilingue breton-français comprenait environ 14000 élèves en Bretagne en 2012 - nombre en augmentation - répartis dans trois types d'école (en immersion à *Diwan*⁴, en bilingue public à *Div yezh*⁵ et en bilingue privé à *Dihun*⁶). Nous souhaitons avec ce projet aider les apprenants, les enseignants et peut-être d'autres intéressés par le breton, et le mener en lien avec la recherche.

— De quels logiciels et ressources pour le breton dispose-t-on ? La notion d'information numérique englobe plusieurs *formes* : texte, son, image, vidéo ; ce type d'information peut être accédé, construit ou exploité de différentes manières. La forme privilégiée par le projet LangNum-br-fr est l'écrit (pas l'aspect vocal) avec des textes en breton ou en français : avant de proposer de nouveaux traitements, nous avons réalisé un état de l'art des ressources et des logiciels existants pour l'écrit [1], complété lors de ce projet : dictionnaires en ligne, logiciel de traduction, de détection de langue, correcteur.

Comment aider les apprenants à l'ère du numérique ? Quel outil serait le plus intéressant ? En début de phase de conception, nous avons considéré qu'une analyse des besoins des apprenants devait être la première étape avant la définition de traitements logiciels pour des

1 page du projet : <http://www.irisa.fr/prive/foret/LangNum-br-fr>.

2 « TIC » signifie « technologies de l'information et de la communication »

3 Le Télégramme, 18 mai 2012 <https://www.letelegramme.fr/ig/generales/regions/morbihan/langue-bretonne-une-situation-contrastee-18-05-2012-1706749.php> (chiffre stable, voir enquête TMO 2018).

4 https://fr.wikipedia.org/wiki/%C3%89cole_Diwan et site <http://www.diwan.bzh/>

5 https://fr.wikipedia.org/wiki/Div_yezh_Breizh et site <http://www.div-yezh.bzh>

6 <https://fr.wikipedia.org/wiki/Dihun> et site <http://www.dihunbreizh.bzh/>

meilleurs et nouveaux usages. Notre état de l'art est proposé comme base pour les enquêtes d'usage (comprenant une partie libre et une partie guidée) avec plusieurs pistes de nouveaux systèmes destinés à un usager (quelques prototypes avaient déjà été réalisés [1] [2] avant le projet.) :

- la lecture augmentée ou pro-active (apportant automatiquement de l'information supplémentaire en contexte, par exemple un élément de traduction, des mots ou concepts reliés etc.) ;

- des systèmes d'aide à la « relecture ou réécriture » pour une meilleure qualité du breton écrit ; ceci comprend les correcteurs orthographiques et grammaticaux, et aussi des systèmes de normalisation (selon les variantes d'écriture) ;

- un système d'aide à la recherche et à l'exploration d'information en breton ;

- une plateforme de discussions.

Du côté technologie de la langue, nous pensons au développement d'analyseurs (morphologiques et syntaxiques), à l'amélioration de réseaux sémantiques et de concepts (comme wordnet ou babelnet) et à établir les liens à ce type de réseau. Il serait aussi intéressant de développer un système de détection d'ambiguïtés pour le breton. Enfin, d'autres développements pourraient consister à adapter au breton des plateformes existantes pour d'autres langues, telles une plateforme de discussions. ».

À l'issue de la phase 1 du projet, nous proposerons de développer un outil choisi selon l'analyse des besoins (réalisée par la laboratoire LOUSTIC). Deux systèmes ressortent : la correction orthographique et la lecture augmentée.

Nous pensons retenir ce dernier type de système pour plusieurs raisons. Un correcteur existe déjà, il faudrait l'améliorer mais il s'agit d'une tâche à assez long terme. Un système de lecture augmentée pourrait être plus rapidement proposé (plus simple à développer et évolutif), et s'appuyer éventuellement sur des outils pour le gallois (comme Vocab ajoutant du vocabulaire par bulles d'information intégrées sur un media internet).

Nous mentionnons cependant deux difficultés liées au breton pour ce type de projet :

- il faudra tenir compte du phénomène des mutations initiales (changement de la première lettre de certains mots selon la nature du mot et son contexte, un phénomène des langues celtiques)

- il serait souhaitable d'intégrer des variétés de breton et tenir compte de différents types d'écriture.

Bibliographie

[1] Foret, Annie, Bellynck, Valérie, Boitet, Christian 2015. Akenou-Breizh, un projet de plate-forme valorisant des ressources et outils informatiques et linguistiques pour le breton. Conférence TALARE (Traitement Automatique des Langues Régionales de France et d'Europe), http://www.atala.org/taln_archives/ateliers/2015/TALaRE/

[2] Foret, Annie 2016. Enrichissement de données en breton avec Wordnet. Conférence CLTW (Traitement automatique des langues celtiques) <http://www.lattice.cnrs.fr/CLTW/>