



HAL
open science

Enhancing a multi-faceted Breton verb-centered resource to help a language learner

Annie Foret, Erwan Hupel, Pêr Morvan

► **To cite this version:**

Annie Foret, Erwan Hupel, Pêr Morvan. Enhancing a multi-faceted Breton verb-centered resource to help a language learner. NLP4CALL2024: Natural Language Processing for Computer-assisted Language Learning, Oct 2024, Rennes, France. hal-04869508

HAL Id: hal-04869508

<https://hal.science/hal-04869508v1>

Submitted on 7 Jan 2025

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.

Abstract

This article builds on two recent resources for Breton, a verb-centered database and a set of sentences in the universal dependencies (UD) format. Our focus is on Breton, an endangered language in the Celtic family. We provide an analysis of the resource on verbs and show how it can be connected and transformed to a multi-faceted system intended to help a learner in a flexible way. We discuss several scenarios.

Enhancing a multi-faceted Breton verb-centered resource to help a language learner

Annie Foret, Erwan Hupel, Pêr Morvan

annie.foret@irisa.fr IRISA & Univ. Rennes
erwan.hupel@univ-rennes2.fr Univ. Rennes 2 & CELTIC-BLM
per.morvan.bzh29@gmail.com An Drouzig

<https://gitlab.inria.fr/foret/termis/-/tree/main/Breton/Verbs>

Focus and Objectives

Breton : an endangered language in the Celtic family

DVB = a rich numeric resource on Breton Verbs

UD = Universal Dependencies for sentences a first UD treebank for Breton (888 sent)

FCA = Formal Concept Analysis approach to handle multiple facets

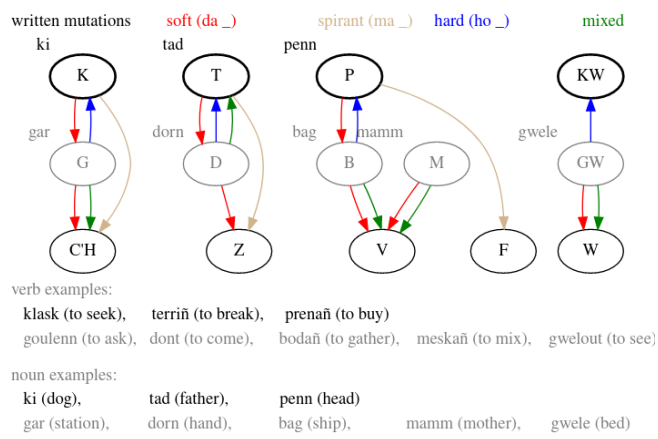
How to help a Breton language learner with DVB as key resource ?

Breton linguistic features - difficulties

- sentence : verb as second constituent (most important first)

- word/verb variation : difficulties

initial consonant mutations



verb without ending

kemer

Verb reizh simple Breton verb, european level A1 verb, verb with infinitive and similar root, Breton verbs that result in French pronominal verbs, 7 French concepts to translate the verb, Other forms : kemerout,

English : to take (all senses), to seize, to capture, to catch, to be offended, to formalize, to take offense

altered/diff. verb base in the infinitive dont

Dont simple Breton verb, déplacement verb, european level A1 verb, Irregular verb, Breton verbs that result in French pronominal verbs, 9 French concepts to translate the verb, Swadesh list verb, childrens, babies, generality, body verbs, Other forms : donet(Gwenedeg)

English : to come, to arrive, to come to, to push, to succeed, to improve, to give, to bring back

dont → deu-, d/z mutation after (a) :

Yann a zeuy hag a ray ur brezegenn.

text[eng]: Yann will come and give a talk.

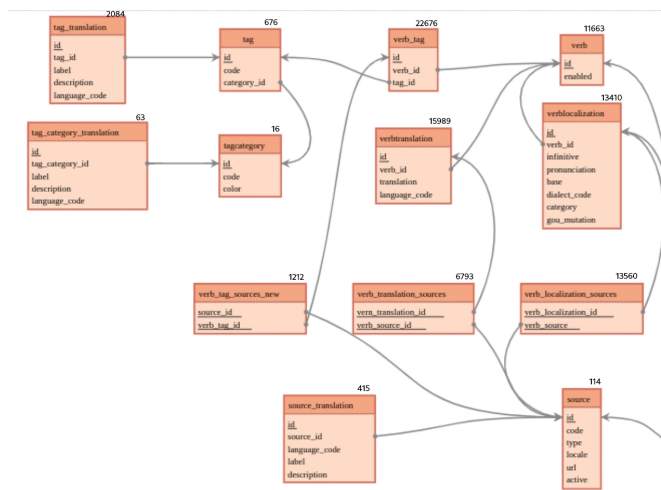
- several verb grammatical categories

To grasp difficulties by facets

A workflow from DVB verbs

DVB site, details on one verb

Our overview of the DVB data

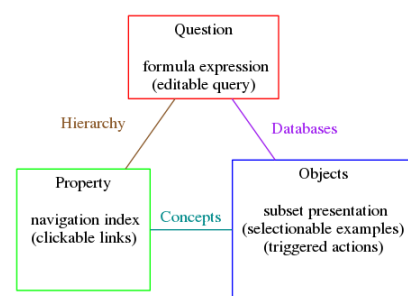


with multiple kinds of tag / facets

Ex. output scenario (1): verbs in hierarchies

Expected benefits of FCA approaches

- linguistic data as formal contexts
- FCA tools, ANR SmartFCA <https://www.smartfca.org/>



Towards easier self-learning processes

... and (structured) sentences

Universal Dependencies (UD) source file

```
# sent_id = apertium.vislcfg.txt:1:0
# text = N'int ket aet war-raok.
# text[eng] = They didn't progress.
# text[fra] = Ils n'ont pas progressé.
# labels = to_check
1 N' ne ADV adv Polarity=Neg 4 advmod _ SpaceAfter=No
2 int bezañ AUX vblex
3 ket ket ADV adv _ 4 advmod _
4 aet mont VERB vblex Tense=Past|VerbForm=Part 0
root _ _
5 war-raok war-raok ADV adv _ 4 advmod _ SpaceAfter=No
6 . PUNCT sent _ 4 punct _ _
```

Ex. output scenario (2): verbs in context

Phrases du corpus UD en breton br_keb_ud

Liste de phrases selon leur verbe. L'infinifit à gauche est lié au site DVB, il est suivi de la phrase du corpus où ce verbe est masqué, sa forme exacte dans la phrase apparaît en passant sur le souligné.

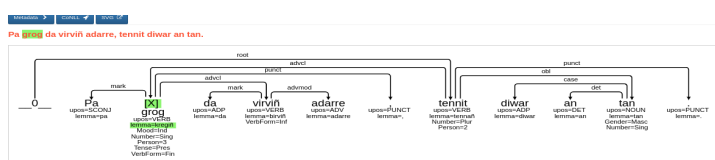
adaozañ (line 13881, sent 803) # text = Prezidant an daou guzul-rannvro ha prezidant kuzul-departamant al Liger-Atlantel hepen a c'hall divizout reiñ lañs da _ ar rannvro. # text[fra] = Seuls les présidents des deux conseils régionaux et le président du conseil général de Loire-Atlantique peuvent décider de lancer la réorganisation de la région.

adkavout (line 13057, sent 762) # text = Muioch-mui e _ ivez ar yezh vrezhon e-liamm gant ar bed ekonomikel. # text[fra] = La langue bretonne est aussi de plus en plus présente dans le monde économique.

selaou (line 7244, sent 402) # text = _ a rit an avel. # text[eng] = You listen to the wind.
selaou (line 8665, sent 511) # text = Va _ ! # text[eng] = Listen to me!
selaou (line 8674, sent 512) # text = Na _ et achanon! # text[eng] = Don't listen to me!

sellout (line 2509, sent 145) # text = Ur stourm ra Ranko kenderc'hel (ha marteze dont da vezañ... nerzhusoc'h) ma fell d'ar Vretoned bezañ doujet evit pezh a _ ouzh skignañ o yezh er radio. # text[eng] = A fight will have to continue (and perhaps take a little nerve) if the Bretons want to be respected with regard to the radio broadcasting of their language.

UD select/view + Grew rewrite to hide/show a form



"When boiling again, draw from the fire". The infinitive "kregin" (EN: to begin) is highlighted in this parse tree (its exact form is "grog", with soft mutation after "Pa"). The second infinitive is "birvñ" (EN: to boil).

Ex. output scenario (3): semantic web RDF

(see gitlab)

Expected outcome and questions

- output features: gathered info in one local file, fewer clicks - less internet ?, flexible navigation, easy-to-use browser-based enabling to train by facets – design ?
- linguistic data input: . verbs : rich breton data - more visible ? . sentences : breton treebank (888 sent) in UD format or its conversion in SUD a new treebank is in preparation UD as a "habitable format" ? (also AMR ?)
- workflow: as a reproducible processing chain easy to adapt (by a developer), which (other) verb and context renderings ?
- preliminary work : no user experiments yet gap filling exercises by a similar workflow.

Work in progress