

« Peut-on revitaliser la langue picarde grâce aux nouvelles technologies ? »

Christophe REY

Lexiques, Textes, Discours, Dictionnaires : centre Jean Pruvost (EA 7518)

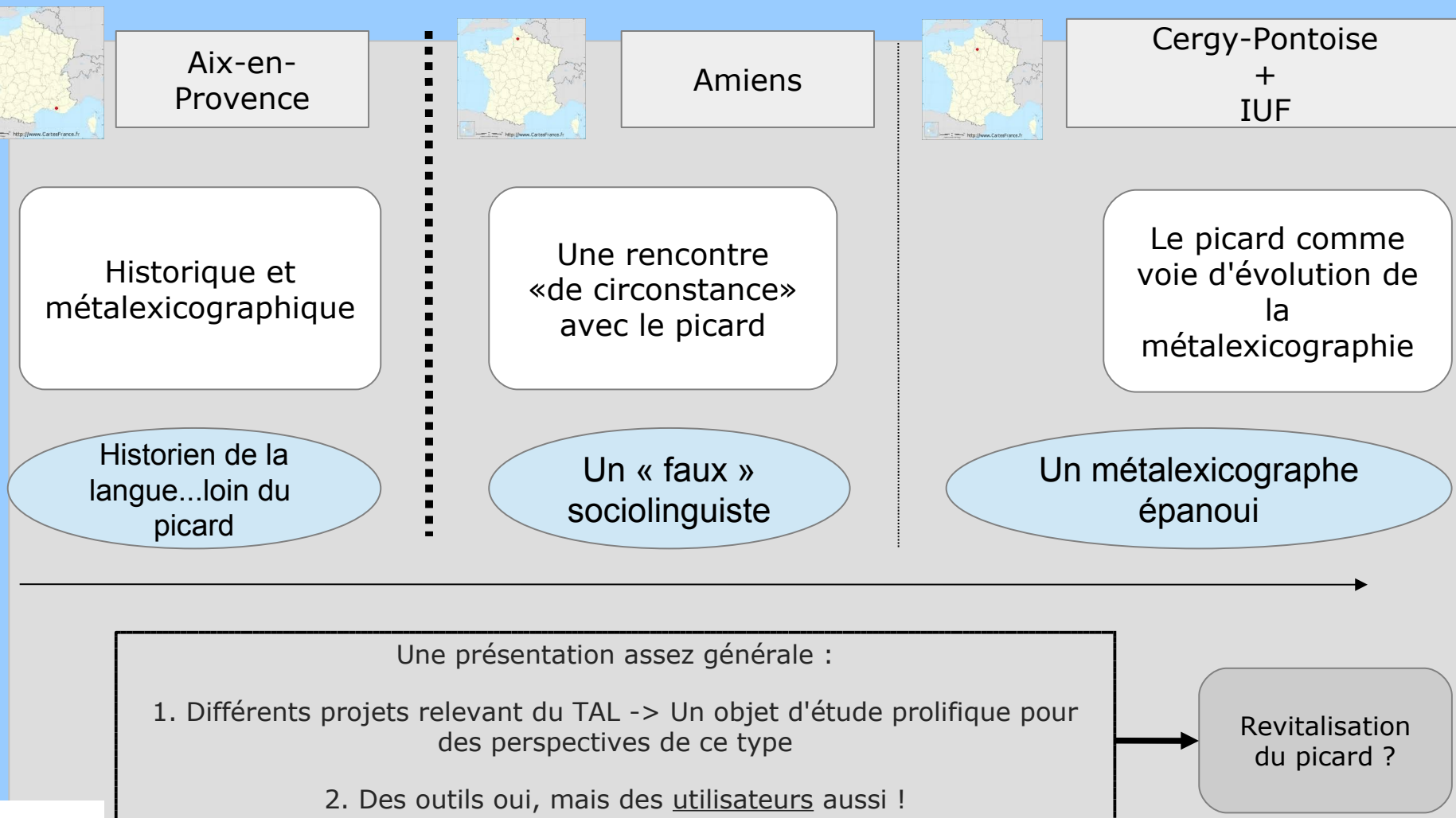
Université de Cergy-Pontoise

Institut Universitaire de France

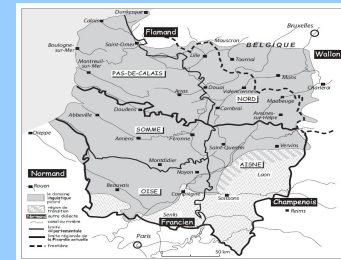
christophe.rey@u-cergy.fr

En guise d'introduction

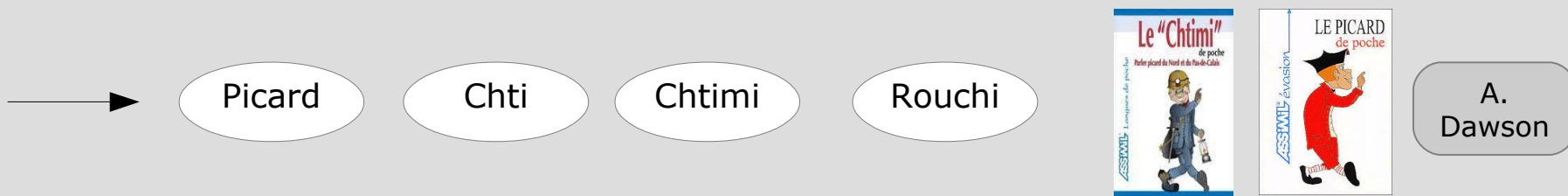
* Le picard comme corpus de travail au sein d'un parcours scientifique au milieu des dictionnaires.



Quelques éléments de présentation de la langue



* Une langue d'oïl existant sur un territoire linguistique très large (depuis l'Aisne jusqu'au Hainaut belge)



* Une variation importante (graphique, phonique, lexicale, etc.) et pas de standardisation véritable (même si tentatives : cf. Système graphique Feller-Carton)

=> **Extrait de *Ele Danme blanke* - François Beauvy. Texte en picard du Beauvaisis (Région située au nord de Beauvais) :** Ché jîns d'èche donmène pi ché fermés d'chés invirons il éteute évnus erbéyé leu mohétrèse pi li dire adé.

=> **« Traduction » du texte en picard de l'Amiénois :** Ché jin dèche donmène pi ché sinsyé dèche z environ il étwète t'évnu erbéyé leu métrèse épi li dire adé.

* Une langue régionale de France richement dotée en ressources lexicographiques

* Une langue en quête d'une meilleure visibilité (Commission de néologie et de terminologie, dictionnaire « fondamental ») -> Des avancées récentes pour l'enseignement...

L'expérience PICARTEXT (Dir : J-M. Eloy et C. Rey) - 2008-2011



- Une base de données littéraire et panchronique
- Environ 10 millions de mots (à étendre)
- Une interface d'interrogation mixte
- <https://www.u-picardie.fr/LESCLaP/PICARTEXT/Public/>



Un terreau fertile pour mieux mesurer les possibilités du TAL pour la langue

Bilan

Une première étape



II. RESTAURE

Des capacités de recherche intéressantes

Recherche d'un mot dans le corpus

Module expérimental de recherche de mots dans le corpus (concordancier).

Mot recherché (exemple : "tchair"):

Méthode de recherche :

- ☒ Chaîne littérale (ex.: trouve uniquement "tchair")
- ☐ Correspondance phonétique (ex.: trouve "tcherre", "tchère", "tcher"...)
- ☐ Correspondance dialectale (ex.: trouve aussi "querre", "queure"...)
- ☐ Expression rationnelle étendue : voir [cette page](#)

Lieu de référence des auteurs :

- ☒ Nord ☒ Pas-de-Calais
- ☒ Aisne ☒ Oise ☒ Somme
- ☒ Hainaut belge

Année de naissance des auteurs : Après Avant

Genres (plusieurs choix possibles) :

BD
 Chanson
 Chronique
 Correspondance

Valider

Annuler

Possibilités
diverses

L'ensemble
du
domaine
linguistique

Diachronie

Multiplicité
de genres

RESTAURE (Dir: D. Bernhard) - 2015-2018



Dates

- > Début du projet : 1er janvier 2015
- > Durée : 42 mois

Financement

Projet financé par l'ANR, référence ANR-14-CE24-0003



Equipe amiénoise :

F. Martin
J-M Eloy
P. Reynés
C. Rey

Les langues régionales au service du Traitement Automatique des Langues (TAL)

Financement
ANR

Une perspective
contrastive
salutaire

Un vrai partage de
compétences

Des avancées
importantes

- Tokenisation
- Lemmatisation
- Etiquetage morpho-syntaxique
- Réflexion sur la constitution de lexiques (de verbes, de noms, de mots grammaticaux, etc.)
- Réflexion sur l'élaboration de ressources lexicographiques

1

Qu'est-ce qu'un mot en picard ?



Le mot
en
TAL ?

2

Un lemme
en picard ?

**Une question déjà posée : cf. « lemme, supralemmme,
dilemme »
(Brun-Trigaud et Carton, 2003)**

Variation graphique avec apostrophes, tirets, points pouvant appartenir au mot), et les consonnes d'appui :

- LES : che, ché, chés, ch'
- UNE : un', un'n, unhn', une, eune, eun', in-n, inne, un.ne, un.n'
- CONSONNES D'APPUI : z, t, n, etc.

Variation diatopique :

- ARGENT : argen, arjin
- CHEMIN : c'min, k'min, q'min, kmin
- JAMAIS : jin.mais, janmais, jamwé, jamé, jama, etc.

Tentative pour définir des règles de prédiction de la variation/des différentes variations:

aer = ar
oer = or
aèr = ar
oèr = or
ons\$ = on\$
ez\$ = e\$
eu\$ = yeu\$
on = o

oait\$ = é\$
oait\$ = è\$
oait\$ = wé\$
oait\$ = wè\$
é\$ = è\$
oais\$ = ouos\$
oait\$ = ouot\$
oai\$ = oao\$
oit\$ = oèt\$

^che = ^c'
^che = ^k'
^che = ^k
^c' = ^k'
^c' = ^'k
^k' = ^k

^re = ^er
er\$ =
eu\$
^ch = ^c
ent\$ =
int\$
lem = l'm
ge\$ =
che\$

L'étude des flexions verbales est une « mine d'or linguistique »

- Avoait, avait, avé, avot, awét, awè't'
- étot, étoué, ètoait, été, étoi, étoait

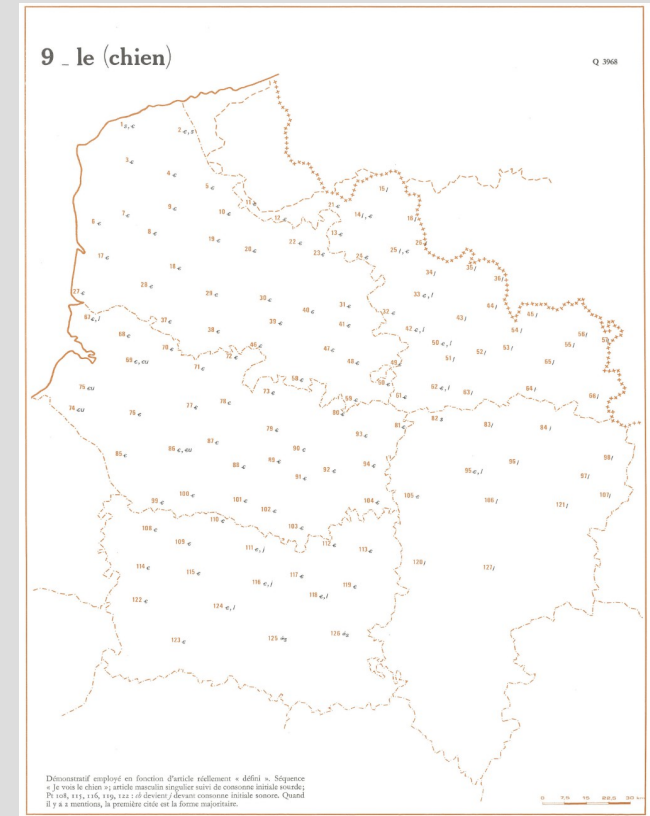
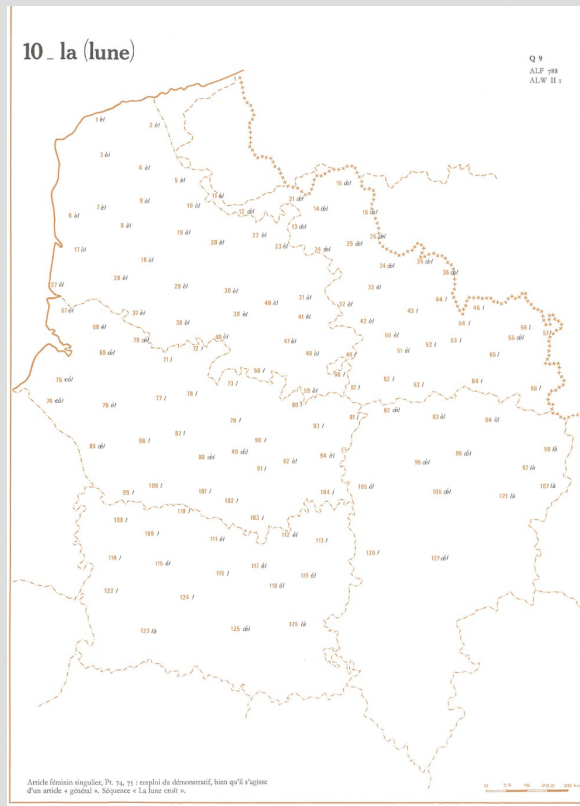
Bilan

RESTAURE : Une avancée formidable pour le picard du point de vue de son exploitation par les outils du TAL.



- * Un projet financé par l'ANR et porté par Esther Baiwir (Laboratoire ALITHILA, Université de Lille)
- * Un projet à orientation « dialectologique » qui offre un prolongement aux travaux de constitution d'atlas linguistiques et notamment à l'*Atlas linguistique et ethnographique picard (ALPic.)* et l'*Atlas linguistique de la Wallonie (ALW)*.

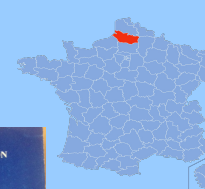
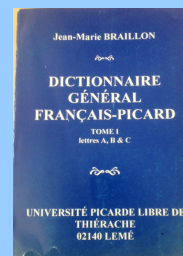
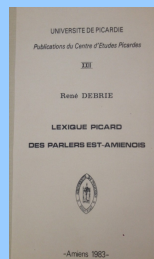
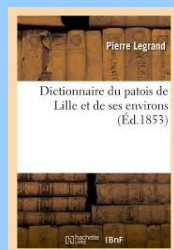
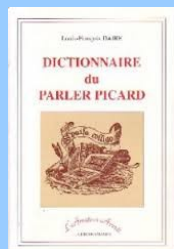
Cartes



Bilan

APPI : Un outil moderne et intéressant

METALPIC (Dir : C. Rey) - IUF 2017-2022



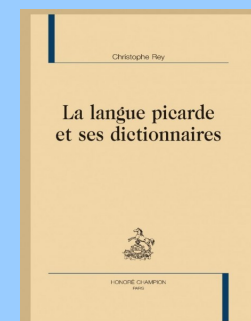
Projet de description métalexographique de la tradition lexicographique picarde.

Une tradition lexicographique très riche et ancienne marquée par une extrême variation.

Un corpus lexicographique à la fois dense, épars et sociolinguistiquement marqué (nombreuses variétés de picard, des productions lexicographiques dans des dictionnaires, des glossaires, des lexiques, etc., un positionnement de langue « minorée », etc.).

Aucune étude de cette envergure et une réinterrogation du concept de « grammatisation » (Auroux, 1992).

Constitution de ressources électroniques (rétroconversion de dictionnaires, bibliographie informatisée, visualisation cartographique, etc.)



2021

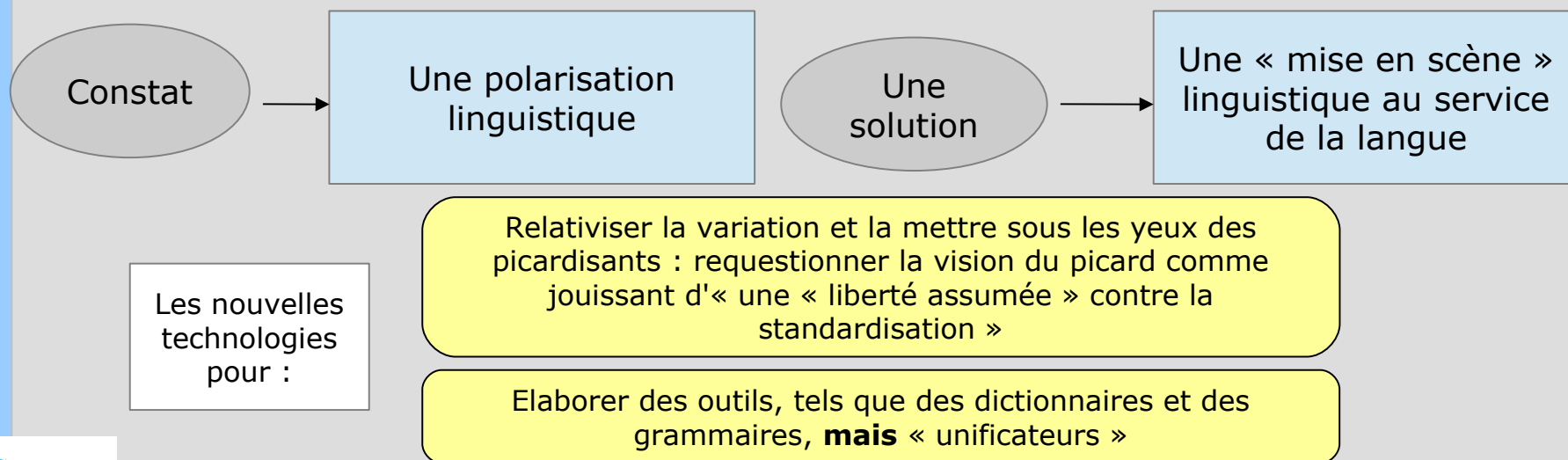
Oui...mais

La mise au point de ressources linguistiques pour une langue donnée sans une exploitation de ces dernières est un investissement incomplet

Une question essentielle soulevée :

Des outils du TAL mais pour quelles finalités autres que celle de l'amélioration des outils rompus à la manipulation de langues bien documentées ?

!> Dans le cas du picard ces outils doivent être conçus et exploités dans la perspective d'une forme de « grammatisation » de la langue (pourtant déjà existante)



Des évolutions récentes qui découlent justement des projets avec orientation TAL

Des équations à résoudre

Des chercheurs pour valoriser et exploiter les données

Mieux cibler les apports pour les usagers eux-mêmes ?

Les choses évoluent

Principes d'une forme de Koïné
(Cf. Albums d'*Astérix et Obélix*)

Création de la Commission de néologie et de terminologie pour la langue picarde (réunions régulières depuis presque 5 ans)

Publication d'un *Dictionnaire fondamental français-picard* (2020) – Liudmila Smirnova et Alain Dawson -> Utilisation de PICARTEXT pour des contextes

Permettre à la langue de sortir de l'ornière d'une variation peut-être trop visible, valorisée et sans doute handicapante

Pas encore d'outils pour le grand public élaborés grâce aux nouvelles technologies



Éléments de conclusion

- * La langue picarde fait partie des langues régionales de France qui ont eu la chance et le courage d'investir le champ du Traitement Automatique des Langues
 - * Grâce aux outils du TAL, le picard est une langue aujourd'hui partiellement outillée
 - * Se pose néanmoins la question de la valorisation de ces ressources existantes
 - * Le TAL a constitué une voie de développement non pas forcément de la langue mais au moins des travaux portant sur la langue picarde
 - * Cette approche est à l'origine d'une évolution dans la conception des outils à mettre en place pour permettre le développement et la valorisation de la langue et de la culture picarde :
 - > Création de la « Commission de néologie et de terminologie pour la langue picarde ».
 - > Elaboration d'un *Dictionnaire fondamental français-picard* (2020- une suite est en route)
- Une vraie avancée pour l'application d'une vraie « harmonisation » orthographique et donc UN PAS EN AVANT pour valoriser la langue.
- * Beaucoup de choses restent à faire et l'entrée du picard dans le domaine des nouvelles technologies a ouvert des portes à ne pas laisser se refermer.