~~Infiltrer~~ Intégrer le CNRS après un Master LTTAC

Alexander Delaporte\ Centre de recherches linguistiques sur l'Asie orientale

alexander.delaporte@cnrs.fr\ tekipaki.hypotheses.org

Mon parcours

In [ ]:
```
import csv
cv_data_path = "./cv_data.tsv"
with open(cv_data_path, "r", encoding="utf-8") as cv:
    cv_reader = csv.DictReader(cv, delimiter="\t")
    print("FORMATION")
    for row in cv_reader:
        if row['TYPE'] == "form":
            print(f"{row['DATE']} : {row['DESCRIPTION']}")
```

Le métier d'ingénieur d'études

> *Les ingénieures et ingénieurs, les techniciennes et techniciens (IT) du CNRS sont chargés des missions d'accompagnement et d'appui des projets de recherche du CNRS. Ils sont plus de 18 000 au sein du CNRS et exercent plus de 200 métiers différents. Ces métiers sont regroupés par branche d'activités professionnelles (BAP).*

Le métier d'ingénieur d'études

- Métier d'appui à la recherche
  - Développement/amélioration technique
  - Valorisation
  - Administration
  - Encadrement, enseignement
- Activité (très) variable selon BAP, fiche de poste, affectation...
- Accessible sans doctorat
- Contractuel ou titulaire

Chercher un (premier) emploi

⚠ Les sites d'emploi "classiques" sont généralement peu adaptés pour les métiers de la recherche

- Newsletters / listes de diffusion
- Réseaux et consortiums
- Concours externe CNRS

Exemples de listes et réseaux à suivre

En français\ ATALA (liste LN), réseau MATE-SHS, liste de diffusion Digital Humanities (DH), consortiums Huma-Num (CAHIER, CORLI)

En anglais\ Corpora-List, LINGUIST List

Passer le concours externe CNRS

Procédure relativement longue (~ 6 mois)

Trois grandes étapes :

1. Choisir le(s) concours
2. Préparer et soumettre son dossier de candidature
3. Passer les épreuves finales d'admission

→ Guide candidat(e) concours externe CNRS 2021

Choix du concours

- Concours annuel
- Période d'inscription limitée (site du concours ITA)
- Postes proposés : BAP, emploi-type, affectation(s)
- Lecture de la fiche de poste
- Un poste = un (numéro de) concours

Passation du concours

Avant d'intégrer le concours :

- Phase d'inscription (inscription en ligne, soumission du dossier de candidature)

Concours en 2 phases :

1. Admissibilité (validation du dossier par le jury)
2. Admission (épreuve orale et/ou écrite)

Dossier de candidature

- CV
  - 1 à 2 pages (exemple)
- Lettre(s) de motivation
  - Une lettre par candidature à un poste
  - 12 000 caractères maximum
- Ordre de préférence des concours à préciser
- Dossier modifiable durant la période d'inscription

Validation du dossier → entrée dans la phase d'admissibilité

Admissibilité

- Étude des dossiers de candidature par le jury
- Établissement d'un classement
- Convocation des candidat(e)s retenu(e)s pour passer l'épreuve d'admission

Dossier retenu par le jury → entrée dans la phase d'admission

Épreuve d'admission

Ingénieurs (IE et IR) = épreuve écrite facultative

Audition :

- Exposé (10 minutes)
  - Respecter strictement le temps imparti
  - Montrer l'adéquation avec le poste mis au concours
  - Avec ou sans support visuel
- Entretien avec le jury (20 minutes)
  - Anticiper les questions
  - Si support visuel : prévoir des diapos annexes

Et après le concours ?

- Résultats publics (pour toutes les phases)
- Notification par courrier postal
- Si lauréat(e) :
  - Proposition d'affectation à accepter/refuser sous 15 jours
  - Prise de fonction au 1er décembre (négociable)
  - Intégration en tant que fonctionnaire stagiaire
  - Titularisation possible après un an

Un exemple concret d'IE en activité (aka, moi-même)

Le CRLAO

Centre de recherches linguistiques sur l'Asie orientale\ UMR 8569 CNRS - INALCO - EHESS\ [site web] (http://crlao.ehess.fr/)

- Unité créée en 1960 (*Centre de linguistique chinoise*)
- Domaine de recherche = linguistique d'Asie orientale, quel que soit le cadre méthodologique ou théorique
- Seul laboratoire européen dont l'activité est entièrement consacrée à la linguistique d'Asie orientale
- Dirigé par Françoise Bottéro depuis janvier 2019

Le CRLAO

4 axes de recherche :

1. *Typologie et langues en contact dans l'aire Asie orientale*

2. *Syntaxe et sémantique formelles des langues contemporaines d'Asie orientale : chinois, coréen, japonais et vietnamien*
3. *Linguistique descriptive des langues d'Asie, phonologie, morphosyntaxe et comparatisme*
4. *Écriture chinoise, classification et cognition*

Le CRLAO

30 membres statutaires :

- 13 chercheurs CNRS
- 14 enseignants-chercheurs
- 3 ITA

36 membres non-statutaires (23 membres associés, 2 postdocs, 13 doctorants)

([Organigramme](http://crlao.ehess.fr/docannexe/file/2164/organigramme_crlao_27_octobre_2021_1_.pdf))

Mon activité

Emploi type = E2C45, *Ingénieur-e en ingénierie logicielle*

Mon activité

Missions

> Conception/réalisation de bases de données et corpus, conception et développement d'outils et logiciels de consultation et/ou de manipulation des données, rédaction de la documentation technique, maintenance des développements réalisés, formation des utilisateurs

Mon activité

Activités

> Recueil des besoins en développement informatique des membres du laboratoire, développement informatique, rédaction de documentation technique, diffusion des développements réalisés et de leur documentation, formation des utilisateurs, veille technologique

Mon activité

Compétences attendues

> Développement informatique, connaissance des techniques de programmation et des systèmes de gestion de bases de données, connaissance des technologies web, XML, XSL, familiarité avec l'encodage Unicode, maîtrise de langages informatiques (Python, Java...), familiarité avec les systèmes d'exploitation Windows, Linux et MacOS, rédaction technique, connaissance de l'anglais technique du domaine

Mon activité

Hors fiche de poste (ou non explicité)

- *Correspondant Sécurité des Systèmes d'Information* (CSSI)
  - Intermédiaire entre membres de l'unité et DSI des tutelles
  - Inventaire du matériel informatique
  - Liste de diffusion
- Membre de différents réseaux, notamment :
  - Mate-SHS, Mate Condorcet
  - Consortium DISTAM
  - GIS Asie

Quelques travaux

DOI-Mercy

DOI-Mercy

Ajout automatique des DOI (*Digital Object Identifiers*) pour les exemples de *A grammar of Japhug* de Guilaume Jacques.

- Programme constitué de 2 sous-programmes développés en Python (3.8)
- DOI issus du corpus Japhug sur Pangloss (LACITO)
- Collaboration avec Séverine Guillaume (coordinatrice Pangloss)

DOI-Mercy

Some finite clauses not followed by postpositions or relator nouns still have clues of a subordinate status. In particular, periphrastic TAME constructions involving an Imperfective verb (§21.2) and a copula (§21.2.2), when they occur in chains, tend to elide the auxiliary in the non-final clauses (§25.1.3), resulting in constructions like (47), where the first clause *ɕʏr tɕe tu-rɤma* 'it is active during the night' is incomplete (lacking the auxiliary *ɲɯ-ŋu*), and the copula *ɲɯ-ŋu* has scope over the two clauses preceding it (see also 10, §21.2.2, for a example of the same type with more than ten clauses sharing a single auxiliary).

(47) *lɯlu nɯ, [ɕʏr tɕe tu-rɤma], sŋi tɕe ku-nɯ-rŋgɯ ɲɯ-ŋu*
cat DEM night LOC IPFV-work day LOC IPFV-AUTO-lie.down SENS-be
'The cat, active during the during the night, sleeps during the day.'
{0003576#S33}

In [ ]:
```
latex = """
\\begin{exe}
    \ex \label{ex:tWZo.tWkWpe}
    \gll  kɯ-pe tú-wɣ-nɤma tɕe li tɯʑo tɯ-kɯ-pe tu \\
    \textsc{sbj}:\textsc{pcp}-be.good \textsc{ipfv}-\textsc{inv}-make \textsc{lnk} again
    \glt  `If one does good things, one will also have good things.' (140518 mao he laos
```

```
\end{exe}
"""
```

In [ ]:
```
Code(latex, language="latex")
```

In [ ]:
```
import re
def nettoyage_extreme(api):
    """Supprimer les balises, signes de ponctuation, chiffres et double-espaces
    Paramètre = api, chaîne de caractères (str)"""
    api = api.replace("$_i$", "")
    api = api.replace("$_j$", "")
    api = api.replace("$_k$", "")
    api = api.replace("ʰ", "h")
    # supprimer balises LaTeX
    balise = re.compile("(\\\\|\r|\n).+?\{", re.IGNORECASE | re.MULTILINE | re.DOTALL)
    api = re.sub(balise, "", api)
    api = api.replace("}", "")
    api = api.replace('\r', "")
    api = api.replace('\n', " ")
    mini_api = api
    # /!\ si retour à une comparaison en nombre de caractères, remettre " " dans la list
    # ajouter "-" à la liste si besoin (ici pertinent de ne pas le supprimer)
    caracteres = ['"', '.', ",", ";", "(", ")", "/", "@", "<", ">",
                  "?", "!", "_", "~", "…", '\t', "`", "'", "[", "]"]
    for c in caracteres:
        mini_api = mini_api.replace(c, "")
    # supprimer chiffres
    mini_api = re.sub(r'\d+', '', mini_api)
    # supprimer espaces multiples
    mini_api = re.sub(r'\s+', ' ', mini_api)
    return mini_api
```

In [ ]:
```
exemple1 = "ɯ-mɤ-ku-kɯ-tsʰi"
exemple2 = "[tɤ-rmi ɯ-kɯ-rɤt] tɤ-pɤtso nɯ ɯ-rkɯ ʑo, [...] pjɤ-zɣɯt tɕe,"
exemple3 = "nɯ ma iʑo ji-kɤ-rku me,  atu spɣi tɤ-ɕe qʰe, laχtɕʰa ŋotɕu nɤ-kɤ-sɯso ʑo nɯn
exemple4 = "tɕe nɯnɯ ɯ-taʁ jɯ\redp{}jɤ-kɯ-ɣe nɯ ku-ndɤm ɲɯ-ŋu."
```

In [ ]:
```
nettoyage_extreme(exemple1)
```

In [ ]:
```
from IPython.display import Code
xml_s109 = """<S id="S109">
    <AUDIO end="301.79" start="299.42" />
    <FORM kindOf="phono">nɯ ɕɯŋgɯ tɕe, ɯ-kɯ-rga pɯ-dɤn.</FORM>
    <DOI>https://doi.org/10.24397/pangloss-0003490#S109</DOI>
</S>"""
```

In [ ]:
```
Code(xml_s109, language='xml')
```

In [ ]:
```
# Augmenter la largeur du notebook pour confort de lecture du tableau
"""from IPython.core.display import display, HTML
display(HTML("<style>.container { width:100% !important; }</style>"))"""
```

Fichier de sortie 1 = table de correspondance

Flag	Index exemple	Exemple grammaire	Modification grammaire	Fichier grammaire	Phrase corpus	Modification corpus	Fichier corpus	Identifiant
Correspondance	7	nɯ ɕɯŋgɯ tɕe, ɯ-kɯ-rga pɯ-dɤn.	nɯ ɕɯŋgɯ tɕe ɯ-kɯ-rga pɯ-dɤn	4-04.tex	nɯ ɕɯŋgɯ tɕe, ɯ-kɯ-rga	nɯ ɕɯŋgɯ tɕe ɯ-kɯ-rga pɯ-dɤn	crdo-JYA_HIST-12-ZMBROKO.xml	S109

					pɯ-dɤn.			
Imbrication	3	tɤ-mtʰɯm ʁɟa ʑo ma nɯ ma, nɤki, tɯjpu mɤ-kɯ-ndza ci tu tɕe,	tɤ-mthɯm ʁɟa ʑo ma nɯ ma nɤki tɯjpu mɤ-kɯ-ndza ci tu tɕe	4-04.tex	tɤ-mthɯm ʁɟa ʑo ma nɯ ma, nɤki,	tɤ-mthɯm ʁɟa ʑo ma nɯ ma nɤki	crdo-JYA_HIST-27-SPJANKW.xml	S190
	5	a-kɯ-tʰu		4-04.tex				

(ordre des lignes modifié pour l'exemple)

Pourquoi 2 sous-programmes ?

- Vérifier qu'il n'y a pas de problème
- Quantifier l'automatisation possible
- Apporter des corrections manuelles
  - Ajouts manuels pris en compte par le second sous-programme

In [ ]:
```
latex_doi = """
\\begin{exe}
    \ex \label{ex:tWZo.tWkWpe}
    \gll  kɯ-pe tú-wɣ-nɤma tɕe li tɯʑo tɯ-kɯ-pe tu \\
    \textsc{sbj}:\textsc{pcp}-be.good \textsc{ipfv}-\textsc{inv}-make \textsc{lnk} again
    \glt  `If one does good things, one will also have good things.' (140518 mao he laos
(https://doi.org/10.24397/pangloss-0004030#S111)
\end{exe}
"""
```

Fichier(s) de sortie 2 = grammaire avec insertion des DOI

In [ ]:
```
Code(latex_doi, language="latex")
```

Le programme duplique les fichiers d'entrée (fichiers d'origine non altérés).

Pa-Hng and Hm Nai dialects from Wiktionary

Pa-Hng and Hm Nai dialects from Wiktionary

Extraction et restructuration d'un tableau de données issu de la page Appendix:Pa-Hng comparative vocabulary list du Wiktionary.

Chinese gloss	Language (location)	Word	Page no.
—	Pa Hng (Guangxi Wenjie)	jɦʉ^{31}	
—	Pa Hng (Guizhou Gundong)	ji^{22}	
—	Hm Nai (Longhui Huxingshan)	i^{13}	
—	Hm Nai (Tongdao Chuansuo)	je^{31}	237
—	Hm Nai (Longhui Xiaoshajiang)	i^{325}	560
一定	Pa Hng (Guangxi Wenjie)	je^{31}tĩ^{35}	
一定	Pa Hng (Guizhou Gundong)		
一定	Hm Nai (Longhui Huxingshan)		
一月	Pa Hng (Guangxi Wenjie)	l̥a^{55}je^{31}	
一月	Pa Hng (Guizhou Gundong)	l̥a^{55}jei^{22}	
一月	Hm Nai (Longhui Huxingshan)	tɕa^{13}ŋ̊ei^{31}	
一百零一	Pa Hng (Guangxi Wenjie)	jɦʉ^{31}pe^{55}ta^{31}jɦʉ^{31}	
一百零一	Pa Hng (Guizhou Gundong)	ji^{22}pe^{55}ta^{22}ji^{22}	
一百零一	Hm Nai (Longhui Huxingshan)	i^{13}pe^{55}lɯŋ^{31}i^{13}	

Pa-Hng and Hm Nai dialects from Wiktionary

- Scripts et fichiers disponibles sur Github
- Fichiers de sortie :
  - Tableau extrait non modifié
  - Tableau restructuré sans numéros de pages
  - Tableau restructuré avec numéros de pages

Sélecteur langues du monde

Sélecteur HTML avec une liste de 1193 langues (vivantes, mortes, construites), issue de la page Liste de langues de Wikipédia.

- Tester en ligne
- Sélecteur avec ou sans catégories
- Code source disponible sur Github

Tekipaki

Tekipaki

Carnet Hypothèses pour la diffusion et la valorisation des travaux d'ingénierie informatique réalisés pour le CRLAO.\ Adossé à des dépôts [Github](https://github.com/alxdrdelaporte?tab=repositories) et [Gitlab](https://gitlab.com/dashboard/groups) + une [page de démo](https://alxdrdelaporte.github.io/)

- Partage de scripts :
  - Issus de projets du laboratoire
  - "Allégés" et calibrés pour une tâche particulière
  - Commentés (construction et adaptation)
- Présentation d'outils d'automatisation
- Vulgarisation (lexicographie, informatique)
- Apprentissage de langues asiatiques

Exemple : cartographie linguistique avec Python

Projets du CRLAO :

- Cartes pour la représentation de données de recherche de membres du laboratoire
- Notebook Jupyter pour que l'équipe de recherche puisse modifier/créer ses propres cartes sans assistance
- Cartographie des langues étudiées au CRLAO dans un but de valorisation (projet en cours)

Exemple : cartographie linguistique avec Python

Sur *Tekipaki*:

1. Créer une carte à partir de données au format CSV
2. Représenter plusieurs séries de données sur une même carte
3. Cartographier des données linguistiques sous forme de surface (polygone)
4. Automatiser la capture d'écran pour explorer une vue PNG d'une carte HTML
5. Créer une carte à partir de données au format GeoJSON

+ notebooks partagés sur Gitlab : cartographie_linguistique_Python\ + cartes partagées sur page Github : une série de points, deux séries de points, choix de fond de carte, polygone