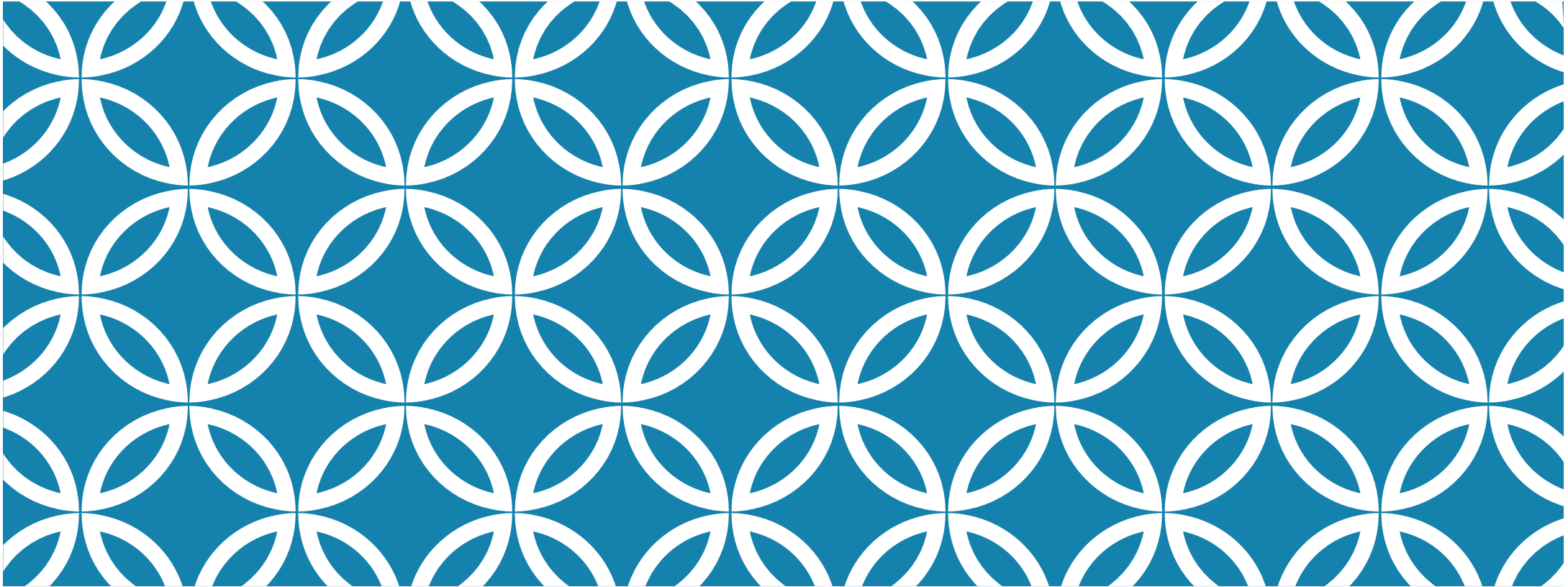




# APPROCHE(S) EN CORPUS POUR UNE LINGUISTIQUE "IMPLIQUÉE" : entre langues de spécialité et traduction outillées

*Unige-UZH Joint Seed Meeting  
17 janvier 2019*

[aurelie.picton@unige.ch](mailto:aurelie.picton@unige.ch)

# PLAN

1. Parcours et intérêts

2. Axes de recherche

- 1. Linguistique outillée pour l'étude de la variation
- 2. Mesure de l'apport de données issues de corpus pour la traduction spécialisée

3. Perspectives pour le projet?

# 1. PARCOURS ET INTÉRÊTS

Formation en linguistique informatique > ingénierie linguistique

Thèse en entreprise (Centre national d'études spatiales)

- Développement d'une approche en corpus pour repérer l'évolution des connaissances dans les projets spatiaux

Post-docs et arrivée à Genève en 2011 en tant que Professeure

- Lexicologie, terminologie et technologies de la traduction (en particulier corpus)

Intérêts

Langues de spécialité (terminologie) – Variation(s) – Corpus – outils

Linguistique « impliquée » - travail avec les experts/professionnels



## 2. Axes de recherche

2.1. Linguistique outillée

2.2. Projet CRISTAL

## 2.2. LINGUISTIQUE OUTILLÉE

Approche corpus : complémentarité possible avec approche équipe Gert de Sutter ?

- Langue de spécialité > corpus plus réduits (< 5 M occurrences)
- Articulation quantitatif/qualitatif

### **Terminologie textuelle**

- (Bourigault et Slodzian (1999), Condamines 2005, Condamines et Picton 2015)
- Demandes appliquées (de terrain)
  - Construction de bases de connaissances terminologique (BCT)
  - Gestion de l'évolution des connaissances, maintenance des BCT, pluridisciplinarité, transfert de connaissances, etc.

# EXEMPLES DE DEMANDES

1. Repérage de l'évolution des connaissances dans les projets spatiaux dits « de longue durée »

- Variation diachronique

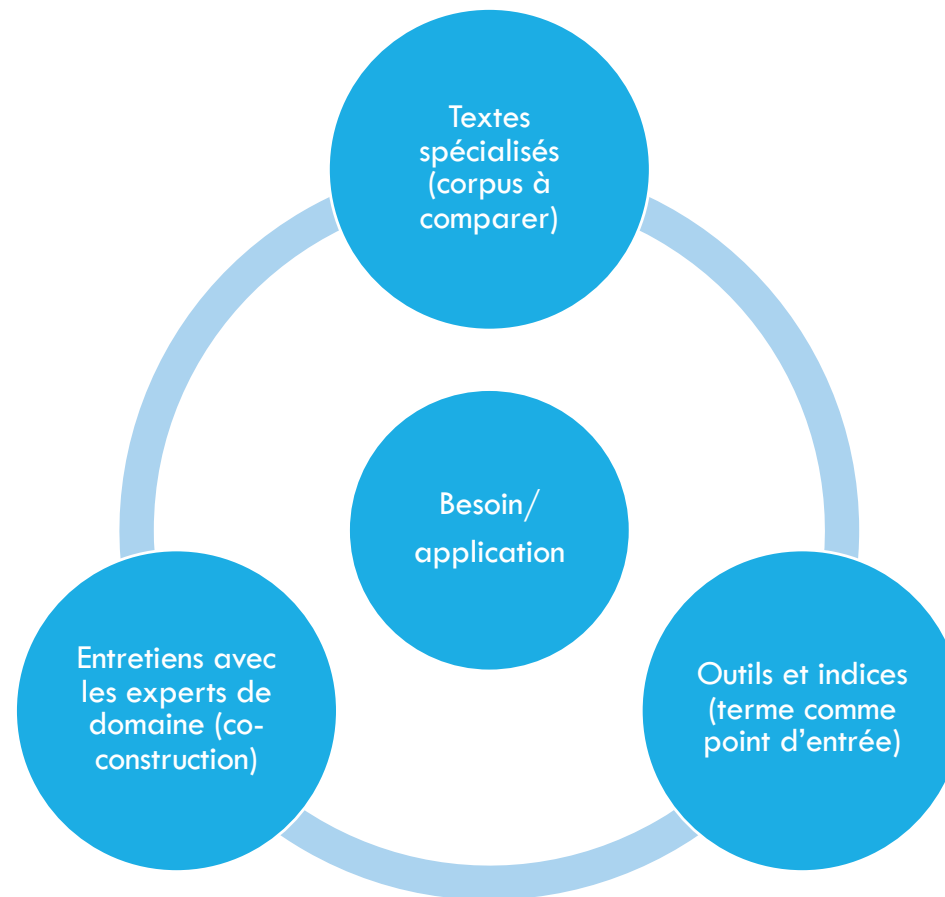
2. Repérage des influences de la langue du spatial dans la langue générale dans le temps (50<sup>e</sup> anniversaire du Cnes)

- Contribuer à définir le phénomène de déterminologisation, comme processus et résultat
- Variation diachronique et diastratique

3. Création d'une ressource dans le domaine de l'humanitaire : Humanitarian Encyclopaedia

- Analyser les différences et convergences d'usages selon différentes perspectives
- À l'attention de différents groupes d'experts à travers le monde
- Variation selon le types d'organisations, la discipline d'origine, diatopique

# APPROCHE GLOBALE



# APPROCHE GLOBALE - COLLABORATION

| Tâches                                                                          | Protagonistes                                      |
|---------------------------------------------------------------------------------|----------------------------------------------------|
| Identifier un besoin<br>(évent. le « traduire » en questionnement linguistique) | Utilisateurs/Experts du domaine/Experts linguistes |
| Construire un corpus                                                            | Experts linguistes/Experts du domaine              |
| Identifier et mettre en œuvre les outils pertinents                             | Experts linguistes/(experts informaticiens)        |
| Interpréter les indices fournis par les outils                                  | Experts linguistes                                 |
| Construire une interprétation finale                                            | Experts linguistes/Experts du domaine              |
| Théoriser le résultat (apport linguistique et reproductibilité)                 | Experts linguistes                                 |

# OUTILS ET INDICES

## Outils

- Extracteurs de termes/de variantes, analyseurs syntaxiques, programmes *ad hoc*, scripts (R, deep-learning\*), concordanciers

## 4 types d'indices

- Quantitatifs (fréquences, statistiques)
- Variations de forme
- Distribution
  - sans interprétation *a priori* (contextes riches en connaissances (Meyer 2001))
  - avec interprétation *a priori* (cooccurrence, en particulier syntaxique)

➤ faisceaux d'indices pour guider l'interprétation

# PRINCIPAUX APPORTS

## Cadre théorique et méthodologique pour observer la variation

- Sur des corpus relativement petits

## Portraits plus riche de fonctionnements linguistiques

- Différents contextes et variables
- Légitimation et description des liens éventuels entre types de variations

## Systématisation d'une approche quali/quantitative en corpus comparables

## Réflexion sur la dimension interprétative en corpus

- Conditions de collaborations avec les experts/acteurs du domaine
- Place et rôle des experts et des linguistes
- Prise en compte des objectifs/besoins pour l'analyse



## 2. Axes de recherche

2.1. Linguistique outillée

2.2. Projet CRISTAL

# PROJET CRISTAL



Contextes Riches en connaissanceS pour la TrAduction TerminoLogique

Projet ANR de 3 ans (2012-2015)

4 partenaires :

- LINA, Univ. de Nantes
- CLLE-ERSS, Univ. Toulouse Jean-Jaurès
- FTI, Univ. de Genève
- Entreprise Lingua et Machina

# KRC

« By knowledge-rich context, we designate a context indicating at least one item of domain knowledge that could be useful for conceptual analysis » (Meyer, 2001: 281)

KRCs (et corpus) sont utiles pour la traduction spécialisée

- “instead of learning to craft intensional definitions or to winnow down a selection of possible contexts to find one ideal context, translators may be better off learning to identify the characteristics of a knowledge-rich context (Meyer 2001) - one that contains useful information - rather than a context that simply attests to the existence of a term. [...] [they] need to learn strategies for isolating such [KRCs] in corpora” (e.g. Bowker 2011: 230)
- Mais rarement disponibles dans les ressources utilisées par les traducteurs (Marshman 2014)

# OBJECTIFS DU PROJET

Extraire automatiquement des Contextes Riches en Connaissances de corpus bilingues comparables

Les intégrer dans un/des outils de TAO

## **Mesurer leur utilité/utilisation : observational study**

- Mieux cerner la notion de KRC (et leur utilisabilité) en se basant sur les opinions exprimées par les traducteurs concernant une situation de travail
- Améliorer l'ergonomie de l'intégration des KRC dans un outil de TAO ("allier efficacité et bien-être", Lavault 2011 : 6)

# EXPÉRIMENTATION

Pré-test à Genève en déc. 2013 : 7 étudiants en M2 et Doctorat en traduction

Mars 2014 : 18 et 24 étudiants M2 traduction d'Angers et Toulouse

Avril 2015 : 11 traducteurs et traductrices professionnels en France

## Tâche

- Traduction d'un texte
  - EN > FR, 150 mots
  - Vulgarisation (volcanologie)
- Validation de KRC utiles pour la traduction (en langue-source et en langue-cible)
- Réponse à un questionnaire en ligne
- Debrief oral (en groupe ou individuel avec visionnage vidéo)

# PROTOCOLE

- Mixed Methods Approach
  - “used when several methods are used to collect or analyze data [and] often understood to mean using both qualitative and quantitative approaches” (Saldanha & O’Brien 2013: 23)
- Triangulation
  - the combined use of different methods of data elicitation and analyses (Alves 2003, Denzin 1978)

# RÉCOLTE DE DONNÉES

## 1. KRC conceptuels

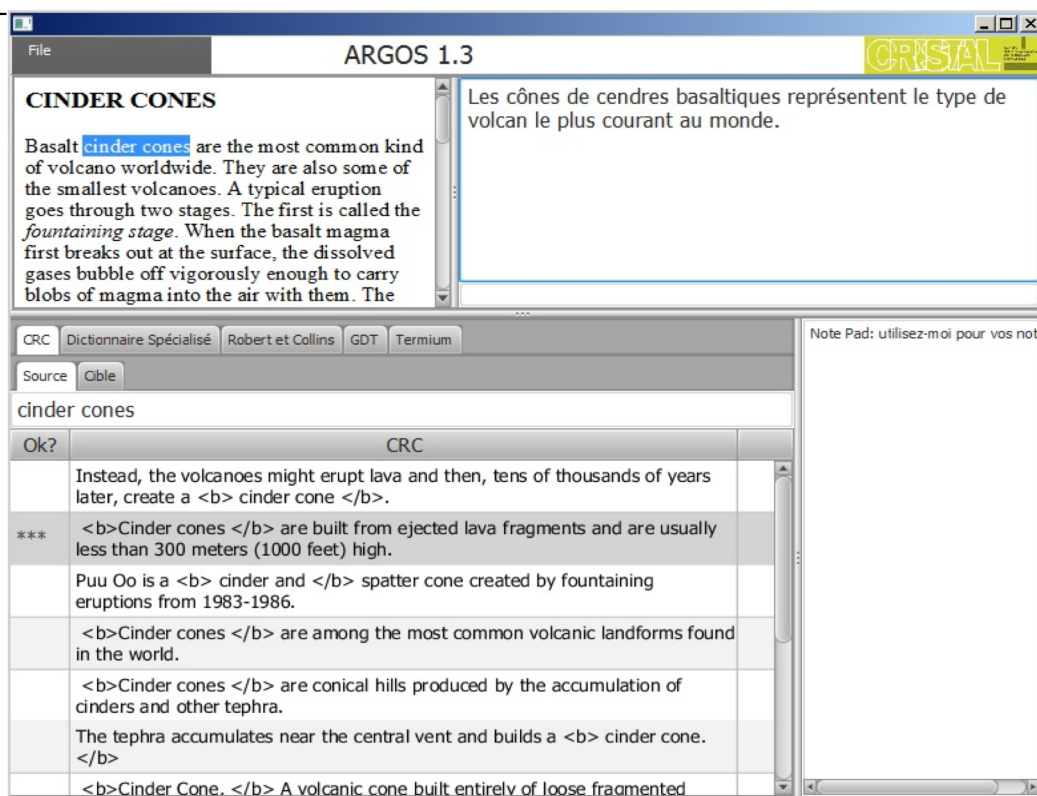
- Approche classique en terminologie par extraction de patrons (Condamines 2003, Marshman 2006, Auger & Barrière 2010)
  - “Classic” approach in terminology: pattern-based (Condamines 2003, Marshman 2006, Auger & Barrière 2010)
    - MAR-REL: [http://redac.univ-tlse2.fr/misc/mar-rel\\_fr.html](http://redac.univ-tlse2.fr/misc/mar-rel_fr.html)

## 2. KRC linguistiques

- Collocations (N, ADJ, V liés au terme)
- Méthodologie automatisée par (Hmida 2016)

## INTERFACE ET ENVIRONNEMENT

Une interface « écologique » (Argos)



42 participants

Étudiants en Ma (Angers & Toulouse)  
Tâche : texte de 150 mots (EN>FR)

**Enregistrement**

Interface logguée(Argos)  
Enregistrement des actions/écran (Bbflashback)

**Questionnaires et interviews**

Focus groups+ cue-based retrospective  
verbalization  
Questionnaire en ligne (Limesurvey)

# DONNÉES OBTENUES POUR L'ANALYSE

## ■ Logs

- Nombre de KRCs consultés
- Liste de KRCs consultés
- Nombre de KRCs validés
- Liste de KRCs validés
- Moment et durée de la consultation
- Terme travaillé
- Autres ressources consultées (avant/après, complémentarité)
- Traduction finale

# DONNÉES OBTENUES POUR L'ANALYSE

- **Questionnaire en ligne**

- Questions ouvertes et fermées
- Opinions des traducteurs sur les KRC, leur utilisation et leur utilité + ergonomie interface

- **Focus groups**

- Discussion et entretiens semi-dirigés
- Approfondissement des opinions des participants sur les atouts/limites, conditions utilisation, etc.

# DONNÉES OBTENUES POUR L'ANALYSE

## ■ Vidéos

- Aide à l'explication d'erreurs (logs)
- Vérifications/complétion d'éléments divers sur les KRC consultés et validés

## Entretiens rétrospectifs avec visionnage des vidéos propres

- Explications de certains choix et validations
- Explications d'actions de recherche
- Explications des attentes du traducteur
- Explications sur l'équivalent final choisi

# PRINCIPAUX APPORTS

Mesure de l'utilité de ressources issues de corpus comparables pour la traduction

- Caractérisation et définition d'un « bon » KRC pour la traduction spécialisée

Mise en place d'un protocole d'observation de « professionnels en action », en traduction

Réflexion sur la notion d'écologie (par ex. Ehrensberger-Dow et Massey 2013) et de la (non) influence du linguiste qui observe



### **3. Réflexions pour le projet ?**



# RÉFLEXIONS

## Approche en corpus (complémentaire?)

- Articulation quanti/quali > échantillonnage pour observation de variables moins volumineuses

## Analyse de la variation (lexico-syntaxique) dans différents genres textuels et contextes

- Domaines, degrés de spécialité, groupes professionnels, région, etc.
- Robuste et adaptable

## Mise en place de protocoles valides « observational studies »

- Environnements écologiques
- Triangulation

➤ Pistes pour les questions 1 et 3 + approche méthodologique

# RÉFLEXIONS

Linguistique « impliquée » (par ex. Leglise 2000; Berthoud & Burger 2014)

- En lien avec un besoin appliqué
  - Sur le terrain
  - Allers-retours théorie/pratique
- 
- Réflexion sur le statut du linguiste comme observateur/conseiller
    - Méthodes, rôle et apports (pour la science et pour le terrain)
  - Réflexion sur le rôle « social » des travaux et recherches



**MERCI!**

[aurelie.picton@unige.ch](mailto:aurelie.picton@unige.ch)





## Annexes

1. Contextes d'application
2. Indices : exemples
3. CRISTAL compléments

