Recherche Zen
Séance 2 : Contextualisation et bibliographie

Par Carlos Ramisch et Manon Scholivet
Basé en partie sur les cours d'Adeline Paiement

20 mars 2023

Plan

J'ai ma question de recherche ! ... Et maintenant ?

Il est temps de s'intéresser aux travaux des pairs.
C'est l'heure de faire :

- Une recherche bibliographique !
- Une revue de la littérature !

Oui c'est la même chose... Non ?

Quelle est la différence ?

- Recherche bibliographique :

  $\longrightarrow$ **Acquisition, approfondissement** des connaissances sur un sujet précis

- Revue de la littérature (survey) :

  $\longrightarrow$ **Synthèse de l'état de l'art** et des connaissances dans le domaine

Quand faire l'un ou l'autre ?

- Recherche bibliographique :
  - $\longrightarrow$ Avant tout travail de recherche
  - $\longrightarrow$ Pendant l'affinage de la question de recherche
- Revue de la littérature (survey) :
  - $\longrightarrow$ La dernière revue est trop ancienne OU
  - $\longrightarrow$ Une telle revue n'a jamais été faite ET
  - $\longrightarrow$ J'ai le temps de me poser et faire le point

Comment faire l'une ou l'autre ?

- Recherche bibliographique :

  $\longrightarrow$ Pas de méthode stricte

  $\longrightarrow$ Idée de point de départ : un article amène à lire une autre article, tel un dictionnaire dont une première définition mène à une seconde, une troisième, etc...

- Revue de la littérature :

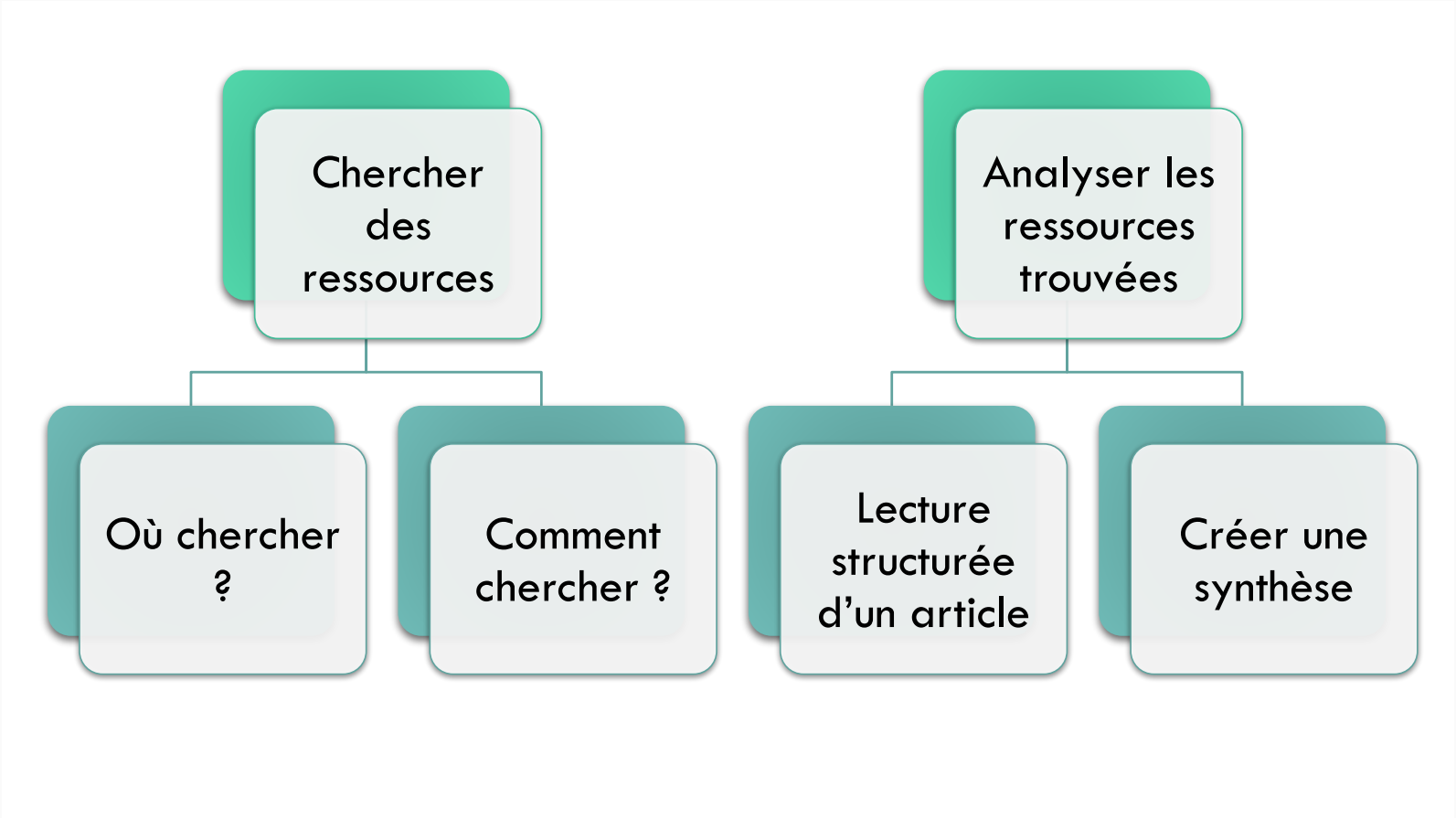
  $\longrightarrow$ Revue **narrative** : pas de méthode stricte

  $\longrightarrow$ Revue **systématique** : méthodologie existante (p.ex. PRISMA)

  $\longrightarrow$ Recherche en priorité des revues déjà existantes, des méta analyses

  $\longrightarrow$ Définir (à priori ou au fur et à mesure) la portée de la revue

La démarche


Chercher des ressources
Où chercher ?
Comment chercher ?
Analyser les ressources trouvées
Lecture structurée d'un article
Créer une synthèse

Plan

Chercher des ressources

Où chercher des ressources ?

- Journaux et actes de conférences (proceedings)
  - → Les BU ont souvent des abonnements aux revues payantes
- arXiv
- Archives
  - Nationales : HAL en France
  - Internationales : en général gérées par les universités
  - Spécialisées : DBLP (informatique), ACL Anthology (TAL)...
- Pages web des labos, équipes, chercheur.se.s du domaine
  - → Parfois on trouve les "preprint" sur des sites perso
- Twitter, réseaux sociaux (selon les domaines)

Chercher des ressources

Comment chercher des ressources ?

- Google Scholar & SSRN
  - Choix judicieux des mots clé
  - Identifier les articles de référence
    - → Nombre de citations et téléchargements
- Newsletters des journaux et éditeurs
  - IEEE, Elsevier, etc.
- Mailing lists thématiques et des sociétés savantes
  - → GDR, projets ANR, associations, etc.
  - → Annonces d'actes de conférences et numéros de revues
- Suivi de liens
  - Trouvés dans les revues de littérature de travaux/articles antérieurs

Chercher des ressources

Comment chercher des ressources utiles ?

- Attention aux dates !
  - Recherche historique : toutes
  - Recherche des dernières avancées : $<$ $\sim$5 ou 10 ans
- Garder en tête qu'on ne pourra pas tout lire !
  - Priorité par pertinence et impact
  - Ordre de grandeur : $\sim$50-100 articles pour un doctorat
  - On s'arrête quand les nouveaux articles semblent prévisibles

    $\rightarrow$ connaissance suffisante du domaine

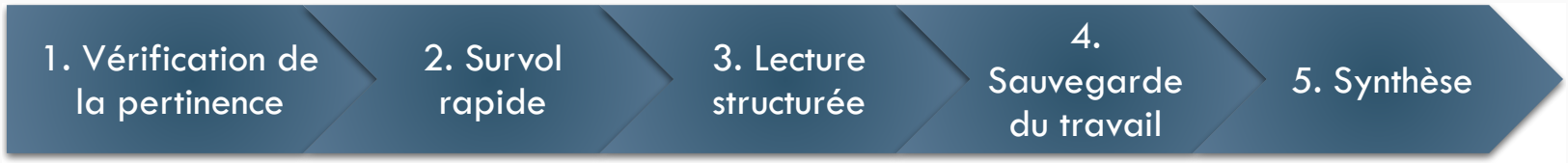
Chercher des ressources

Comment chercher des ressources originales ?

- Pas toujours les mêmes auteurs et labos
  - Couvrir des méthodes et courants de pensée différents
- Rester curieux.se
  - Des idées inspirantes peuvent venir de là où on ne s'attend pas
  - Aller en conférence, aux groupes de lecture, séminaires
- Petit à petit, constituer une base de connaissances générales

Plan

Analyser les ressources trouvées



La structure d'un article

Résumé

- Condensé de l'article, se suffisant à lui-même
- Présente les informations principales

1. Introduction

- Contexte et problématique
- Question de recherche

2. Travaux précédents

- Etat de l'art

3. Méthodologie

- Description de la méthode

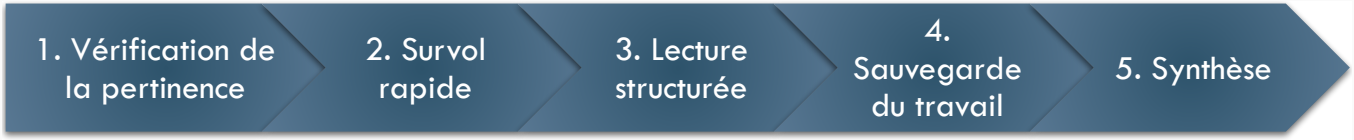
4. Expérimentation & résultats

- Description des expériences réalisées
- Présentation des résultats
- Discussions des résultats et comparaison avec méthodes précédentes

5. Conclusion

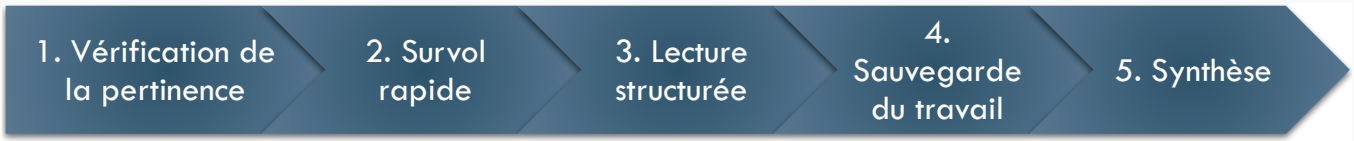
- Récapitulatif des travaux présentés et des résultats principaux

1. Vérification de la pertinence de l'article



- Lecture des sections
  - Résumé
  - Introduction
  - Conclusion
- Survol des autres sections
- Le thème est il pertinent pour mon étude ?
- Les objectifs ont-ils été atteints ?

2. Survol rapide de l'article

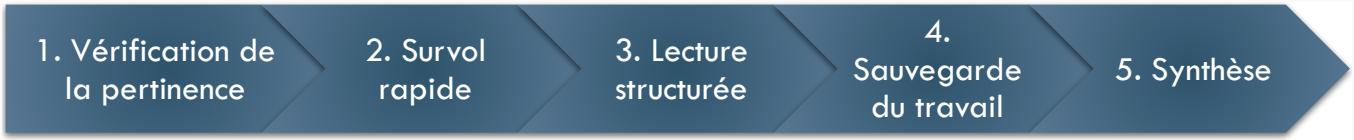


- Lecture rapide de l'article sans s'attarder sur les points difficiles

  Se concentrer sur :

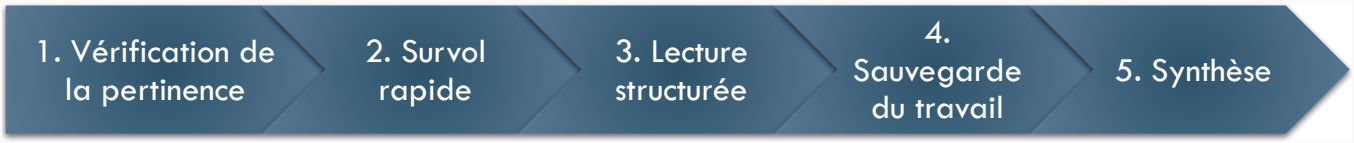
- Résumé, introduction
  - → Bien comprendre la problématique
- Figures (images), tableaux de résultats (chiffres) et conclusion
  - → Première idée des résultats atteints
- But recherché :
  - → Compréhension globale de l'article
  - → Peut on le lire en détail tout de suite, ou préparation nécessaire ?

3. Lecture structurée de l'article



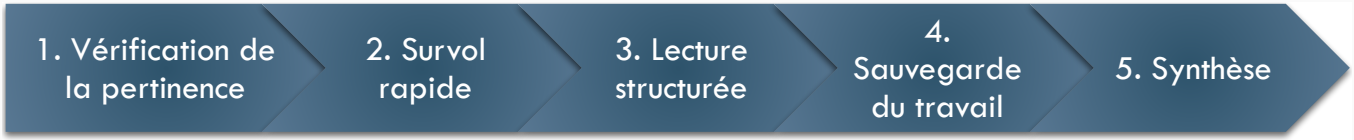
- Lecture détaillée, complète le survol global
- Recherche de ressources annexes si besoin :
  - Articles cités
  - Méthodes citées
  - Travaux précédents des auteur.ice.s
  - ...
- Identifier :
  - But / problématique et question(s) de recherche
  - Méthodes utilisées
  - Résultats comparatifs : points forts et faibles
  - Prise de notes

4. Sauvegarde du travail



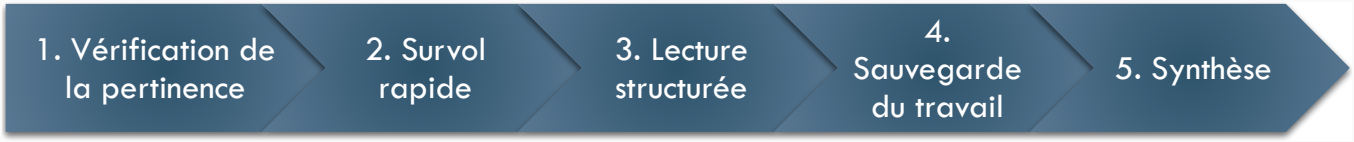
- Liste de ressources
  - Outils de gestion de bibliographie : Zotero, Mendeley, JabRef...
- Notes de lectures
  - Pour chaque article / ressource
  - Format au choix : fichier texte, tableur, papier/crayon...
  - ...
- Rapports synthétiques

4. Sauvegarde du travail



- Citations avec BiBTeX
  - → Mise en forme et tri automatiques
  - → Disponibles sur la plupart des plate-formes
- Outils de "nettoyage" : `bibclean`
- Conventions pour les identifieurs
- Autocomplete sur overleaf et autres éditeurs LaTeX

5. Synthèse

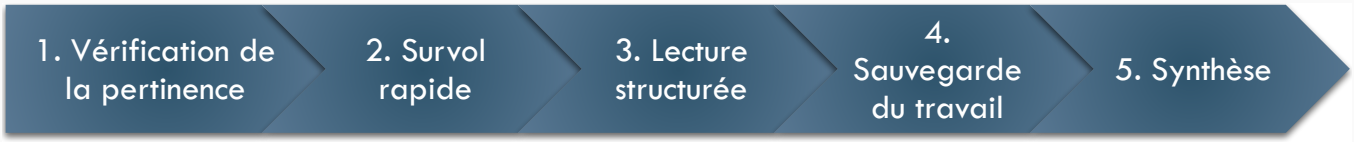


But de la synthèse :

- ~~Empiler articles, les lister indépendamment les uns des autres~~
- Identifier différentes approches pour étudier la question
- Grouper les articles en fonction de caractéristiques communes
  - Théories
  - Méthodologies utilisées
  - Modèles ou algorithmes utilisés
  - Données, datasets
  - Courants de pensée
  - Conclusions obtenues
  - ...

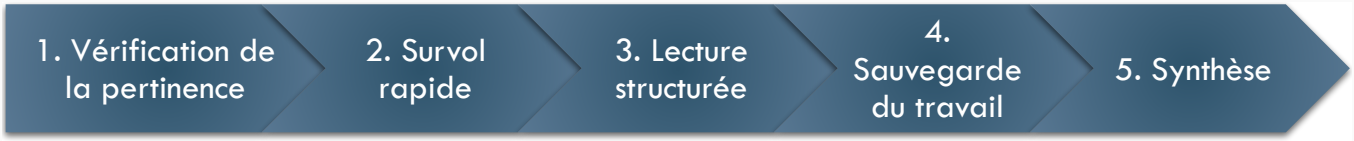
A vous d'identifier la/les caractéristiques pertinentes !

5. Synthèse



- Quoi dire dans une synthèse ?
- Les différentes catégories identifiées
- Les articles qui rentrent dans ces catégories
  - Pourquoi ces articles appartiennent (ou pas) aux catégories
  - Quelles variantes il apportent
- Analyse critique
  - Éléments de comparaison des différentes catégories
  - Montrer les limites des travaux antérieurs : aspects manquants ou insatisfaisants pour notre problématique

5. Synthèse



Méthode :

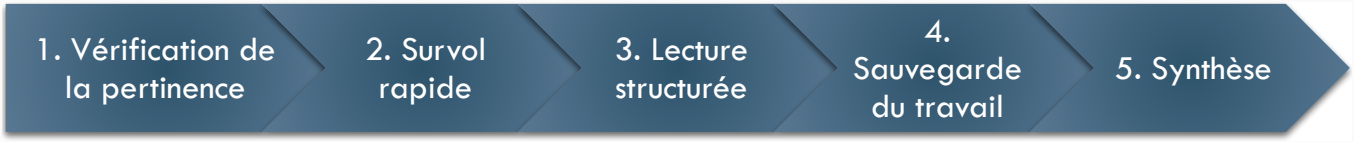
1. Lire les articles en prenant des notes
2. Choisir la ou les catégorisation(s) à utiliser
   - Possibilité de les organiser par progression logique (ex. deep learning, analyse syntaxique en dépendance, puis multilinguisme, ...)
3. Identifier les sous-catégories / variantes principales (ex. réseaux récurrents, réseaux bidirectionnels, réseaux adversaires, ...)
4. Présenter les articles de chaque catégorie et sous-catégorie
   - Avec citation '[x]' ou '(Untel 2016)'

5. Synthèse

Pour écrire une bonne synthèse :

- Pas un catalogue d'articles sans liens entre eux !
- Rester centré sur la problématique !

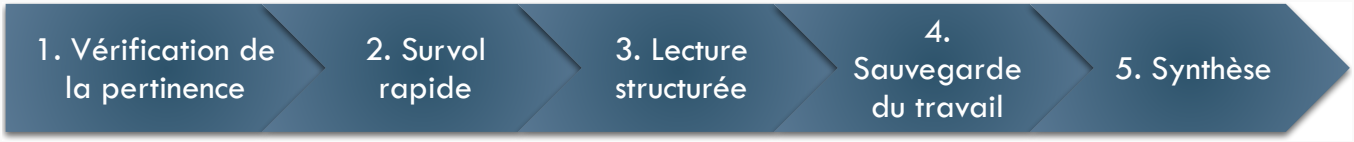
5. Synthèse



Pour écrire une bonne synthèse :

- Pas un catalogue d'articles sans liens entre eux !
- Rester centré sur la problématique !
- Savoir à qui on s'adresse
  - → En général : collègues du même domaine de recherche
- Donner l'essentiel de l'information
  - → Pas trop de détails sur chaque article
  - → Ce n'est pas un cours et nous parlons à des expert.e.s

5. Synthèse



Pour écrire une bonne synthèse :

- Pas un catalogue d'articles sans liens entre eux !
- Rester centré sur la problématique !
- Savoir à qui on s'adresse
  - → En général : collègues du même domaine de recherche
- Donner l'essentiel de l'information
  - → Pas trop de détails sur chaque article
  - → Ce n'est pas un cours et nous parlons à des expert.e.s
- Toujours indiquer ses sources
- Citer des sources "respectées"
  - → Publications scientifiques, rapports officiels...
  - → Attention aux articles trop récents, non relus (arXiv)

Revue de la littérature : exemple

Question : *Quelles choix méthodologiques sont faits dans les expériences en identification automatique d'expressions ?*

Portée

- Articles après 2016 utilisant les datasets PARSEME ou DiMSUM
- Articles disponibles sur ACL Anthology

Grille d'analyse

- Section 2 - corpora for train/eval
  - → 2.1 Domain of the corpora ?
  - → 2.2 Languages of the corpora ?
  - → 2.3 Split of the corpora (standard, cross-validation, etc.)
- Section 3 - pre/post processing
  - → 3.1 What preprocessing before learning/inference ?
  - → 3.2 What post-processing after inference ?
- Section 4 - Evaluation results
  - → 4.1 Which are the metric used (exact match, fuzzy match, link-based) ?
  - → 4.2 Is p-value reported ? Which test is applied ?
  - → 4.3 Which phenomena are looked at in error analysis ?

Revue de la littérature : exemple

1		2 Languages	3 Split of the corpora	3.4 Category of	4.1 Preprocessi	4.2 How are MW
2	**PARSEME 1.0**					
3	The PARSEME Shared Task on Autom	18: BG, CS, DE, EL,	train/test, no dev		N/A	N/A
4	Parsing and MWE Detection: Fips at th	8: FR, EN, DE, IT, E	Not mentioned	VID, LVC, VPC,	Transformation t	N/A
5	The ATILF-LLF System for Parseme Sh	18: BG, CS, DE, EL,	PARSEME data	PARSEME categ	Not mentioned	Binary-lexical tre
6	Detection of Verbal Multi-Word Express	15: CS, DE, EL, ES,	PARSEME data	VPC, LVC, VID,	Not mentioned	Not mentioned
7	USzeged: Identifying Verbal Multiword	9: DE, EL, ES, FR, H	PARSEME 1.0 (no dev	PARSEME 1.0 c	Remove long se	Single-token: rep
8	A data-driven approach to verbal multiv	12: RO, FR, CS, DE,	PARSEME 1.0 - cross	PARSEME 1.0 c	Not mentioned	Two steps: Heac
9	Neural Networks for Multi-Word Expres	15: BG, CS, DE, EL,	80% train, 10% dev, 1(	PARSEME 1.0	Not mentioned	MWE category'
10	**PARSEME 1.1**					
11	Edition 1.1 of the PARSEME Shared Ta	19: BG, DE, EL, EN,	3 languages had no de	LVC, VID, IRV, V	N/A	N/A
12	CRF-Seq and CRF-DepTree at PARSE	19: BG, DE, EL, EN,	PARSEME 1.1 data	PARSEME 1.1	Converting to XM	BI, BIO, and BIL
13	Deep-BGT at PARSEME Shared Task 2	10: BG, DE, ES, FR,	PARSEME 1.1 data	All PARSEME 1.	Merging labels, r	gappy 1-level
14	GBD-NER at PARSEME Shared Task 2	19: BG, DE, EL, EN,	PARSEME 1.1 (no me	All PARSEME 1.	Not mentioned	sub-graphs, usin
15	Mumpitz at PARSEME Shared Task 20	7: BG, DE, EL, ES, F	PARSEME 1.1 (they m	PARSEME 1.1, t	Categories ignor	Binary, whether
16	TRAPACC and TRAPACCS at PARSEM	19: BG, DE, EL, EN,	PARSEME 1.1 (param	PARSEME 1.1	Not mentioned	Similar to ATILF
17	TRAVERSAL at PARSEME Shared Tas	19: BG, DE, EL, EN,	PARSEME 1.1 (develc	PARSEME 1.1	Case lifting (cha	Keep only categ
18	VarIDE at PARSEME Shared Task 201	19: BG, DE, EL, EN,	PARSEME 1.1 (no me	All PARSEME 1.	Ignore categorie	IDIOMATIC vs L
19	Veyn at PARSEME Shared Task 2018:	19: BG, DE, EL, EN,	PARSEME 1.1 (no tun	All PARSEME 1.	Duplicate senter	BIOG (Gaps), IC
20	SHOMA at Parseme Shared Task on A	19: BG, DE, EL, EN,	PARSEME 1.1 data (n	All PARSEME 1.	label conversion	Labels converte
21	**PARSEME 1.2**					
22	Edition 1.2 of the PARSEME Shared Ta	14: DE, EL, EU, FR,	train/dev/test for all lar	LVC, VID, IRV, V	N/A	N/A
23	MultiVitaminBooster at PARSEME Shar	*7: DE, EU, GA, HI, I*	PARSEME 1.2	All PARSEME 1.	N/A	Only 'MWE cate
24	MTLB-STRUCT @Parseme 2020: Cap	14: DE, EL, EU, FR,	PARSEME 1.2	All PARSEME 1.	label conversion	The begining tok

Lecture critique

- Faire attention aux potentiels biais de l'étude
- Discuter avec les collègues si quelque chose semble "bizarre"
  - $\rightarrow$ Si on n'a pas compris, c'est peut-être juste pas clair
  - $\rightarrow$ Un langage peu clair peut cacher une méthodologie pas nette
  - $\rightarrow$ Ne pas avoir peur d'avoir l'air bête – ça permet d'avancer
  - $\rightarrow$ Écrire aux auteur.ice.s, demander leur code, données si possible
- Publish or perish : des relecteur.ice.s pressé.e.s peuvent laisser passer des articles problématiques

Lecture critique : exemple

Exemple

Ganley, Mingle, Ryan, Ryan, Vasilyeva, Perry (2013). Developmental Psychology `https://psycnet.apa.org/record/2013-02693-001`

- *[. . .] no evidence that the mathematics performance of school-age girls was impacted by stereotype threat.*
- Condition expérimentale "stéréotype" : *This is very important, as boys have done much better than girls on this test in the past.*
- Condition expérimentale "non stéréotype" : on ne dit rien

  $\rightarrow$ Or, le stéréotype est aussi activé en l'absence de la consigne !

Source : formation "inégalités de genre" - Isabelle Régner et Magali Putero

“Notre” framework : bibliographie

Une proposition de structuration du travail :

- **Titre, auteur**
- **Résumé**
- **Contributions**
- **Similarités** avec vos travaux
- **Différences** avec vos travaux
- **Remarques**

“Notre” framework : bibliographie

- **Surface Statistics of an Unknown Language Indicate How to Parse It Wang and Eisner (2018)**
- **Résumé** Cet article parle de l’utilisation de statistiques de surface pour faire du parsing sur une langue inconnue....
- **Contributions**
  - Les features apprises sur des corpus annotés en POS aident au parsing.
  - L’utilisation de langues “synthétiques” au moment de l’entraînement augmente les résultats
  - ...
- **Similarités** avec nos travaux
  - Delexicalisé
  - Zero-shot
  - ...
- **Différences** avec nos travaux
  - Non supervisé
  - Ils n’utilisent aucune donnée parallèle
  - ...
- **Remarques**
  - Leur système dépend beaucoup des POS tags gold
  - Ils font des critiques intéressantes du WALS :
    - *The unknown language might not be in WALS.*
    - *Some typological features are missing for some languages.*
    - ...
  - ...

Recherche bibliographique : résumé

- Première étape de tout projet de recherche
- Processus itératif - définition de la question de recherche
- Objectifs
  - Construire un argumentaire scientifique
  - Justifier la question de recherche
    - $\longrightarrow$ Pertinente - combler une lacune
    - $\longrightarrow$ Faisable - s'inspirer de ce qui existe
    - $\longrightarrow$ Intéressante - impact potentiel dans le domaine

Revue de la littérature : résumé

- Synthèse de l'état de l'art dans le domaine
- Nécessite une lecture structurée d'un grand nombre d'articles
  - → Capacité de synthèse
  - → Comparaison, mise en perspective
  - → Analyse et structuration des connaissances
  - → Identification des défis et problèmes ouverts du domaine
- Valorisée / publiée sous la forme d'un survey ou meta-analyse

Sources

- Cours d'Adeline Paiement
- Cours de Damien Driot
- Google images