

**ENTRE
PRÉSENCE ET
DISTANCE**

Enseigner et apprendre
les langues à l'université
à l'ère numérique

2ÈME COLLOQUE DU PERL
12 - 13 - 14 DÉCEMBRE 2019

PERL
PÔLE D'ÉLABORATION
DE RESSOURCES LINGUISTIQUES

u^spc
Université Sorbonne
Paris Cité



Corpus numériques : critères pour l'enseignement des langues

CRISTELLE CAVALLA



Présentation

Corpus numériques



Ça fait peur !!

$$\square \overline{18} \square \frac{1}{10} \square \overline{.5} \frac{1}{100} \square \frac{1}{100} \overline{3} \overline{500.5} \ 2) \overline{21} \overline{3.00} \overline{3.5} \square \overline{3} \overline{18} \square$$

$$\times \square \square \overline{3.5} \overline{3} \frac{1}{100} \overline{3} \overline{500.5} \square \square \frac{1}{10} \square \overline{500184} \overline{)12} \frac{1}{12} \frac{1}{10} \square \overline{.5} \ 2) \overline{21} \overline{3.5} \square \frac{1}{12} \overline{3} \frac{1}{10} \frac{1}{100} \overline{3} \overline{6} \overline{)39} \frac{1}{12} \square$$

$$\square \square \square \overline{500184} \overline{)12} \frac{1}{12} \frac{1}{10}$$

$$\square \overline{500} \frac{1}{12} \overline{186} \overline{)39} \frac{1}{12} \overline{500} \overline{3} \ \overline{3.5} \frac{1}{100} \square \square \overline{18} \square \overline{18} \ 2) \overline{21} \square \frac{1}{10} \square \overline{500184} \overline{)12} \frac{1}{12} \frac{1}{10} \square \overline{.5} \square \ 2) \overline{21} \square \frac{1}{10} \frac{1}{10} \square$$

$$\square \square \ 2) \overline{21} \square \overline{.5} \square \frac{1}{12} \square \ ' \square$$

$$\underbrace{\square \overline{500.00.00} \square \overline{.5} \frac{1}{100} \ \overline{3.5} \frac{1}{100} \square \square \overline{18} \square \overline{18} \ 2) \overline{21} \square \frac{1}{10} \square \overline{500184} \overline{)12} \frac{1}{12} \frac{1}{10} \square \overline{.5} \square \ 2) \overline{21} \square \frac{1}{10} \frac{1}{10} \square}$$

$$\square \square \ 2) \overline{21} \square \overline{.5} \square \frac{1}{12} \square \ ' \square$$

$$\underbrace{\square \overline{500.00.00} \square \overline{.5} \frac{1}{100} \square \square \overline{500} \overline{3} \frac{1}{10} \overline{3} \overline{18} \ 2) \overline{21} \square \frac{1}{10} \square \overline{500184} \overline{)12} \frac{1}{12} \frac{1}{10} \square \square \overline{3.5} \frac{1}{100} \square \square \overline{18} \square \overline{18} \ ' \square}$$

Présentation

Corpus numériques



Mais non... ;)

1. Présentation liminaire
2. Qu'est-ce qu'un corpus en linguistique de corpus ?
3. Pourquoi intégrer les corpus en classe de langue ?
4. Comment intégrer les corpus en classe de langue ?
5. Comment choisir les corpus à intégrer ?
6. Conclusion

Présentation

Corpus numériques



1. Présentation liminaire
2. Qu'est-ce qu'un corpus en linguistique de corpus ?
3. Pourquoi intégrer les corpus en classe de langue ?
4. Comment intégrer les corpus en classe de langue ?
5. Comment choisir les corpus à intégrer ?
6. Conclusion

Michel SERRES

**Les nouvelles technologies :
révolution culturelle et
cognitive, 2007, Conférence
INRIA**

*La petite Poucette, M. Serres,
2012, Le Pommier.*

« Les nouvelles technologies
nous ont condamnés à
devenir intelligents ! »



Conférence sur Youtube, 1mn à partir de 51'30 :
<https://www.youtube.com/watch?v=ZCBB0QEmT5g>

Critères de choix des corpus numériques pour l'enseignement des langues

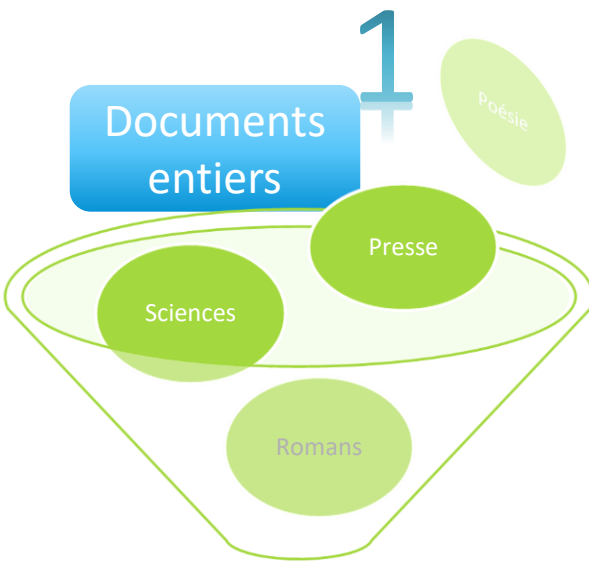
CRITÈRES CORPUS ÉCRITS	CRITÈRES CORPUS ORAUX
Types de documents : presse, littérature, articles scientifiques...	Types d'enregistrements : audio, vidéo / transcription ou pas
Variation : région, authenticité	
Taille : fréquence, représentativité	
Date : synchronie, diachronie	
Pour qui et pour quoi la création du corpus	
Scripteurs / Locuteurs : natif ou allophone / âge / sexe / mono-multilingue	

Présentation

Corpus numériques



1. Présentation liminaire
2. Qu'est-ce qu'un corpus en linguistique de corpus ?
3. Pourquoi intégrer les corpus en classe de langue ?
4. Comment intégrer les corpus en classe de langue ?
5. Comment choisir les corpus à intégrer ?
6. Conclusion



Classer les documents



Discours journalistique

Genres



Économie



Politique



Faits divers



Culture...

2bis



Extraits

Extraction des données

Corpus de presse française (Le Monde, Libération, Le Figaro...)
2010-2015
63 M de mots
Lexicoscope

Num	Identifiant	Contexte gauche	Pivot	Contexte droit
1	s12478	Il a rappelé avoir mis les axes prioritaires et permis le cursus empo-	retraite	sans plafond de revenu
2	s108485	SNCF-RATP	retraite	d'office plus tardive SOCIAL. Ce n'est pas un alignement des régimes de la RATP et de la SNCF avec ceux des salariés du privé, mais l'avancée est significative.
3	s109407	Pendant que d'autres puissances économiques, comme les États-Unis ou la Grande-Bretagne, sortent de la crise par la mise en oeuvre de politiques budgétaires, le nouveau président a sapé l'appareil productif de la France en augmentant, dès 1981, le smic de 10 %, en abaissant la durée du temps de travail (cinquième semaine de congés,	retraite	à 60 ans, semaine de 30 heures) et en créant une société d'assurance, avec l'invention du RSI et l'augmentation taraboussée de toutes les allocations.
4	s189189	Mères de 3 enfants et parents d'enfants handicapés	retraite	à leur plein restera définitivement fixée à 65 ans pour les parents d'un enfant handicapé
5	s1187	Celles sur la branche	retraite	découleront de la réforme en cours.
6	s15919	D'abord parce que Betty a sorti de la tête des riches, 1,2 milliard d'euros d'avantages fiscaux (notamment la déduction de cotisation	retraite	ou de prévoyance).
7	s22644	" On est prêt mais on attend que la séquence	retraite	soit terminée car on ne voudrait pas que les syndicats le prennent comme une stratégie de diversion " avoue un ministre.
8	s20844	Pour conclure la nécessité de réviser les déficits et l'impulsion d'une gestion participative dans l'entreprise, il y a des solutions, comme, par exemple, celle de lier uniquement les sommes qui seraient perçues immédiatement, et non celles qui sont épargnées pour le plus long terme dans les plans d'épargne d'entreprise et les plans d'épargne.	retraite	
9	s44054	Les formules pures d'épargne	retraite	telles que les Perp, les Perco et les contrats Madelin progressent certes, elles touchent désormais près de 30 % des actifs soit 7,6 millions de cotisants.
10	s44071	En 2003 avait été acté le principe d'une hausse des cotisations	retraite	compensée par la baisse des cotisations chômage

Showing 1 to 10 of 1,420 entries

Nombre total d'occurrences 1420

Élaboration d'un corpus

Rastier, F. (2002). Enjeux épistémologiques de la linguistique de corpus. URL : http://www.revue-texto.net/inedits/Rastier/Rastier_Enjeux.html

Sinclair, J.M. (1996). Preliminary Recommendations on Corpus Typology. Rapport technique. *Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.*

<http://www.ilc.cnr.it/EAGLES96/corpus/corpus.html>

Num	Identifiant	Contexte gauche	Pivot	Contexte droit
1	s53478	Il a rappelé avoir mis fin aux préretraites et permis le cumul emploi-	retraite	sans plafond de revenu.
2	s109485	SNCF-RATP :	retraite	d'office plus tardive SOCIAL Ce n'est pas un alignement des régimes de la RATP et de la SNCF avec ceux des salariés du privé, mais l'avancée est significative.
3	s159407	Pendant que d'autres puissances économiques, comme les États-Unis ou la Grande-Bretagne, sortaient de la crise par la mise en oeuvre de politiques libérales, le nouveau président a sapé l'appareil productif de la France en augmentant, dès 1981, le smic de 10 %, en abaissant la durée du temps de travail (cinquième semaine de congés,	retraite	à 60 ans, semaine de 39 heures) et en créant une société d'assistanat, avec l'invention du RMI et l'augmentation faramineuse de toutes les allocations.
4	s189159	Mères de 3 enfants et parents d'enfants handicapésLa	retraite	à taux plein restera définitivement fixée à 65 ans pour les parents d'un enfant handicapé.
5	s1187	Celles sur la branche	retraite	découleront de la réforme en cours.
6	s15519	D'abord parce que Bercy a sorti de la liste des niches, 1,2 milliard d'euros d'avantages fiscaux (notamment la déduction de cotisation	retraite	ou de prévoyance).
7	s22646	" On est prêt mais on attend que la séquence	retraite	soit terminée car on ne voudrait pas que les syndicats le prennent comme une stratégie de diversion ", avoue un ministre.
8	s24844	Pour concilier la nécessité de réduire les déficits et l'impulsion d'une gestion participative dans l'entreprise, il y a des solutions, comme, par exemple, celle de taxer uniquement les sommes qui seraient perçues immédiatement, et non celles qui sont épargnées pour le plus long terme dans les plans d'épargne d'entreprise et les plans d'épargne-	retraite	.
9	s44054	Les formules pures d'épargne	retraite	telles que les Perp, les Perco et les contrats Madelin progressent certes, elles touchent désormais près de 30 % des actifs soit 7,6 millions de cotisants.
10	s64571	En 2003 avait été acté le principe d'une hausse des cotisations	retraite	compensée par la baisse des cotisations chômage.

Corpus de presse française (Le Monde, Libération, Le Figaro...)

2010-2015

63 M de mots

Lexicoscope

Présentation

Corpus numériques



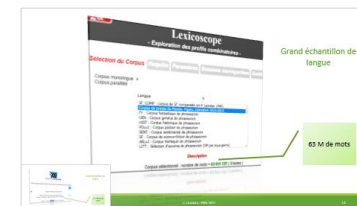
1. Présentation liminaire
2. Qu'est-ce qu'un corpus en linguistique de corpus ?
3. Pourquoi intégrer les corpus en classe de langue ?
4. Comment intégrer les corpus en classe de langue ?
5. Comment choisir les corpus à intégrer ?
6. Conclusion

3. Pourquoi les corpus ?

PLAN MATÉRIEL

- Outils gratuit en ligne
 - Grand échantillon (Sinclair 1996)

- Défi l'intuition des natifs



Lexicoscope

- Exploration des profils combinatoires -

Sélection du Corpus Requête Paramètres Sessions sauvegardées Guide

Corpus monolingue
 Corpus parallèle

Langue

- SF_COMP : corpus de SF comparable en-fr (années 1990)
- Corpus de presse du Monde, Figaro, Libération 2010-2015**
- FY : Corpus fantastique de phraseorom
- GEN : Corpus général de phraseorom
- HIST : Corpus historique de phraseorom
- POLv2 : Corpus policier de phraseorom
- SENT : Corpus sentimental de phraseorom
- SF : Corpus de science-fiction de phraseorom
- ARLv2 : Corpus Harlequin de phraseorom
- LITT : Sélection d'oeuvres de phraseorom (5M par sous-genre)

Description

Corpus sélectionné : nombre de mots = **63 011 137** (3 textes)

Grand échantillon de langue

63 M de mots

CORPORA COLLECTION
LEIPZIG UNIVERSITY

428 corpora corpora basé dictionnaires monolingues pour 252 langues.

Corpus: French (fr_mixed_2012)

French mixed corpus based on material from 2012
 Phrases: 74.823.426 - Types: 7.873.935 - Mots: 1.488.768.604

1,5 milliard de mots



428 corpora corpora basé dictionnaires monolingues pour 252 langues.



Corpus: French (fra_mixed_2012)

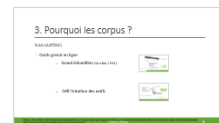
French mixed corpus based on material from 2012
Phrases: 74,823,426 · Types: 7,873,935 · Mots: 1,468,766,604



Allemand Anglais Français Arabe Russe tous...

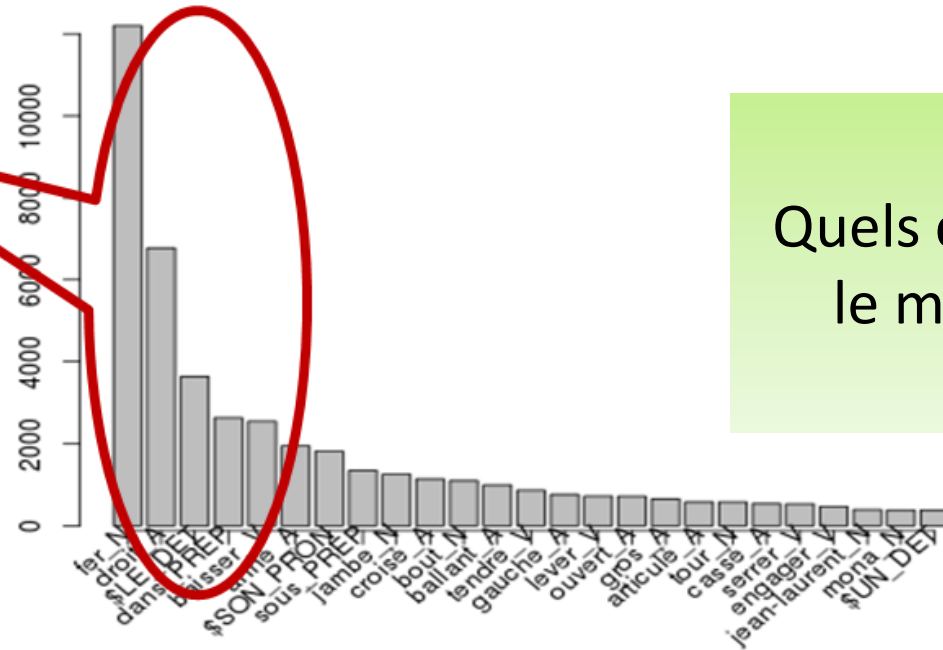
Grand échantillon de langue

1,5 milliard de mots



- Bras de fer
- Bras droit
- Baisser les bras

Lexicogramme relatif au pivot
bras_N



Quels contextes pour le mot « bras » ?

Lexicoscope, corpus journalistique 112M de mots

Intuition des locuteurs

3. Pourquoi les corpus ?

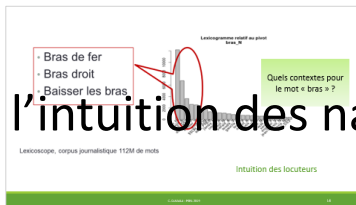
PLAN MATÉRIEL

- Outils gratuit en ligne

Grand échantillon (Sinclair 1996)



Défi l'intuition des natifs



PLAN DIDACTIQUE

- Aide à la découverte de phénomènes langagiers écrits ou oraux
- Conduire les apprenants à l'utilisation autonome (Chambers, 2011)

Sinclair, J.M. (1996). Preliminary Recommendations on Corpus Typology. Rapport technique. *Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.*

<http://www.ilc.cnr.it/EAGLES96/corpus/corpus.html>

Chambers, A., Farr, F., & O'Riordan, S. (2011). Language teachers with corpora in mind: from starting steps to walking tall. *The Language Learning Journal*, 39:1, 85-104.

https://ulir.ul.ie/bitstream/handle/10344/4583/Farr_2011_corpora.pdf;sequence=1

Présentation

Corpus numériques



1. Présentation liminaire
2. Qu'est-ce qu'un corpus en linguistique de corpus ?
3. Pourquoi intégrer les corpus en classe de langue ?
4. **Comment intégrer les corpus en classe de langue ?**
5. Comment choisir les corpus à intégrer ?
6. Conclusion

Éléments en présence

Données authentiques en classe de langue

Corpus numériques et autres outils en ligne

Bureautique
Abord de nouveaux logiciels (comme app. tél)

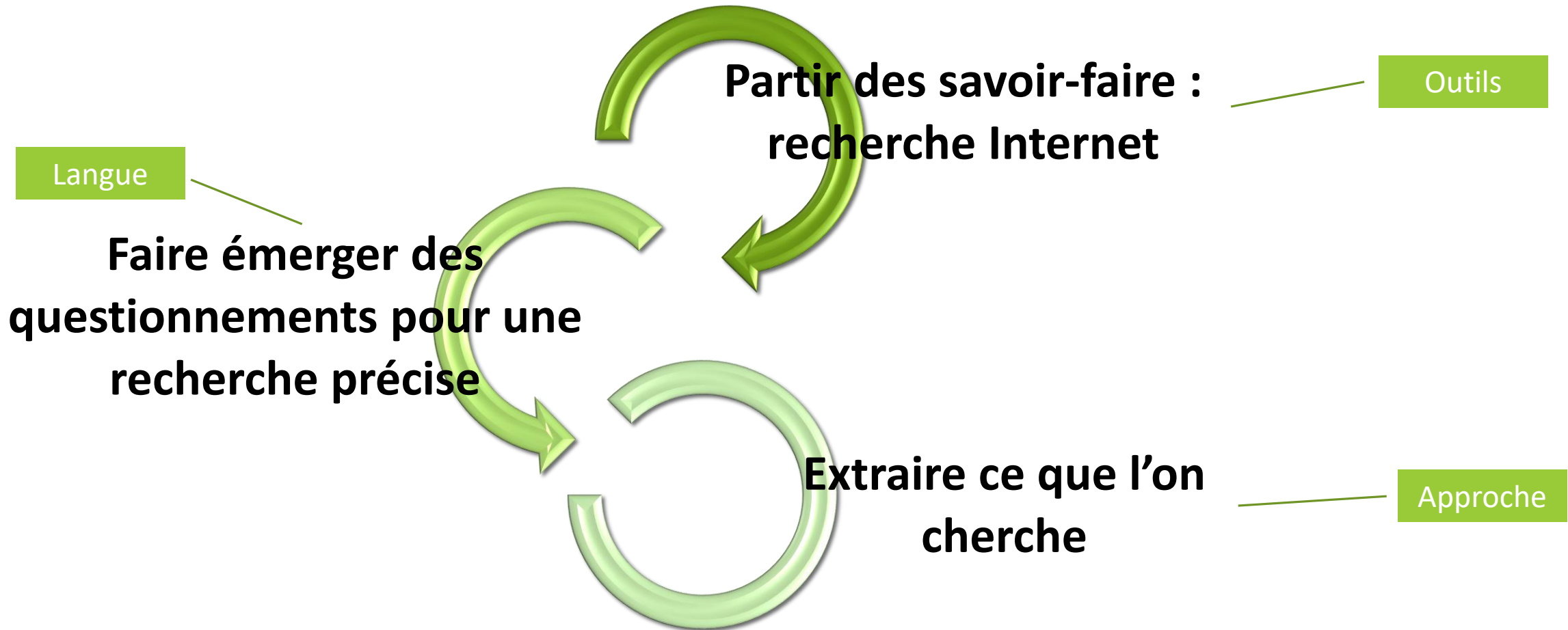
Analyse de la langue

Langue immédiatement réinvestie

Approche didactique

Méthode de travail vers l'autonomie

4. Comment faire découvrir ?



Approches didactiques

CORPUS DRIVEN (Firth 1957)

L'utilisateur cherche ce que le corpus contient

DDL « *Data Driven Learning* » ou « apprentissage guidé par les données »

Inductif

CORPUS BASED (Leech 1992)

L'utilisateur sait ce qu'il cherche et utilise le corpus comme un moyen de vérification

Déductif

APPROCHE ASC - APPRENTISSAGE SUR CORPUS

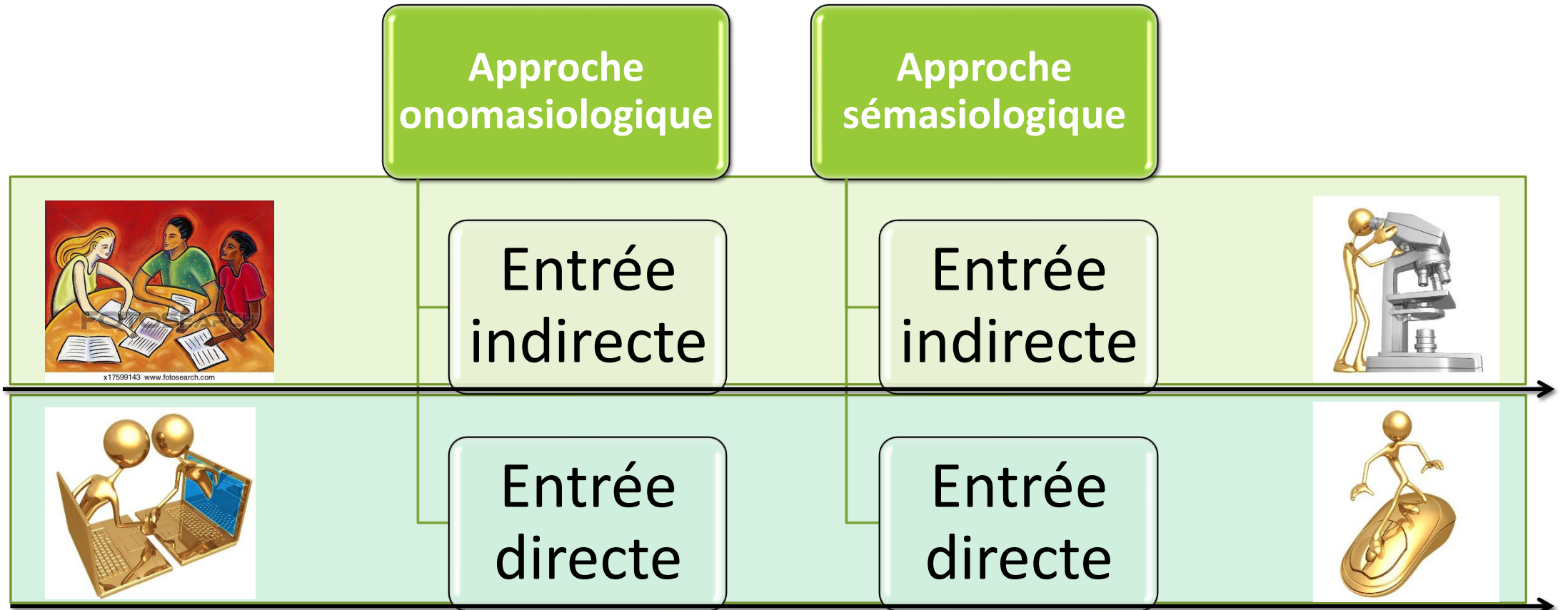
→ toute utilisation de données tirées de façon explicite d'un corpus pour exploitation par des non natifs à des fins d'apprentissage ou d'aide à l'expression écrite, la traduction, etc. (Boulton *et al.* 2014: 145)

Leech, G. (1992). « Corpora and Theories of Linguistic Performance », in J. Svartik (éds), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium, 4-8 August 1991*, Berlin-New York : Mouton de Gruyter, p. 105-122.

Firth, J. R. (1957). *Modes of Meaning. Papers in Linguistics 1934-1951*, 190-215.

Boulton, A. & Tynes, H. (2014). *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*. Paris: Didier.

En classe de langue



Onomasiologie	Sémasiologie
Entrée par un thème	Entrée par un mot
Déclassement de mots	Déclassement dans une
Il n'y a pas de besoin pour dire le thème ?	Il n'y a pas de besoin pour dire le mot ?

Onomasiologie



Entrée par un thème

Didactique : de quels mots a-t-on besoin pour dire ce thème ?

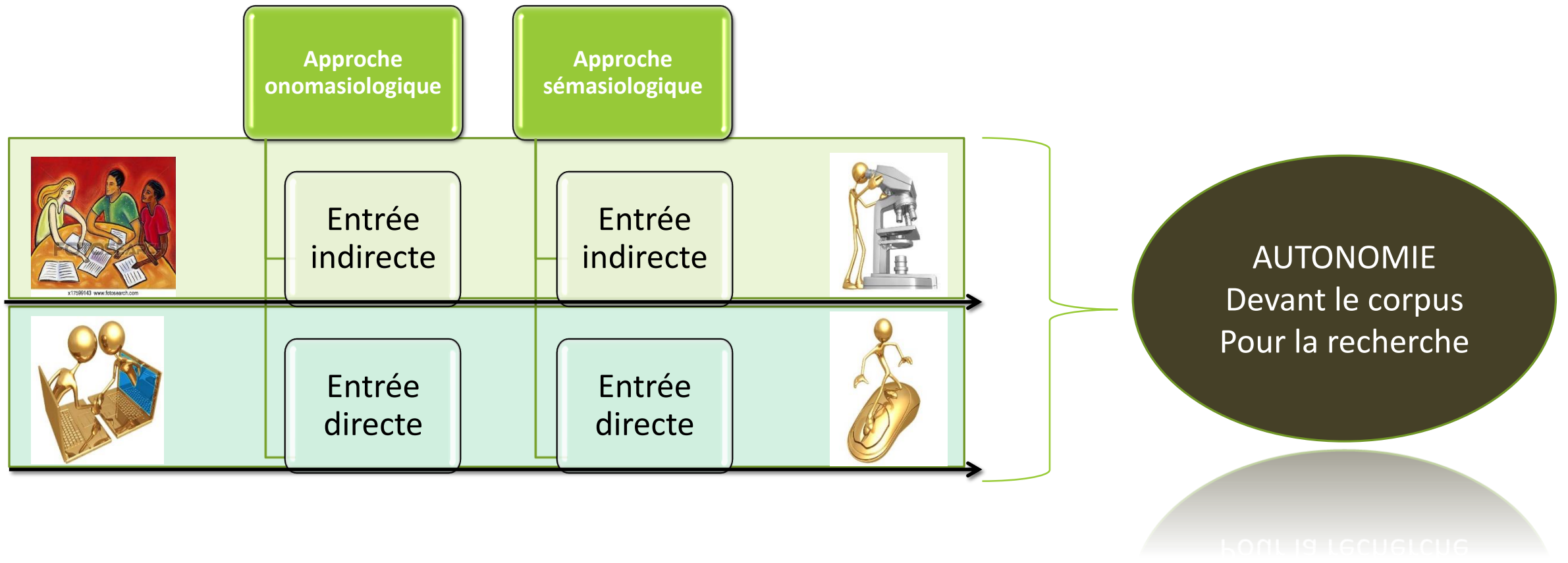
Sémasiologie



Entrée par un mot

Didactique : dans quel thème peut-on utiliser ce mot ?

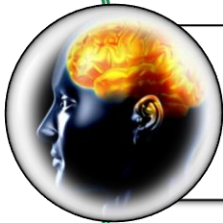
En classe de langue



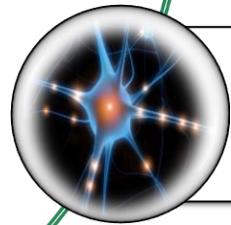
4. Objectifs de la démarche



Laisser le temps de la prise de conscience par manipulation



Poser ses propres questions en fonction de ses propres besoins



Instrumentaliser l'outil pour ses propres besoins

Trouver les réponses dans les corpus
Instrumentalisation

Présentation

Corpus numériques



1. Présentation liminaire
2. Qu'est-ce qu'un corpus en linguistique de corpus ?
3. Pourquoi intégrer les corpus en classe de langue ?
4. Comment intégrer les corpus en classe de langue ?
5. Comment choisir les corpus à intégrer ?
6. Conclusion

Liste de corpus gratuits en ligne en français

ÉCRIT

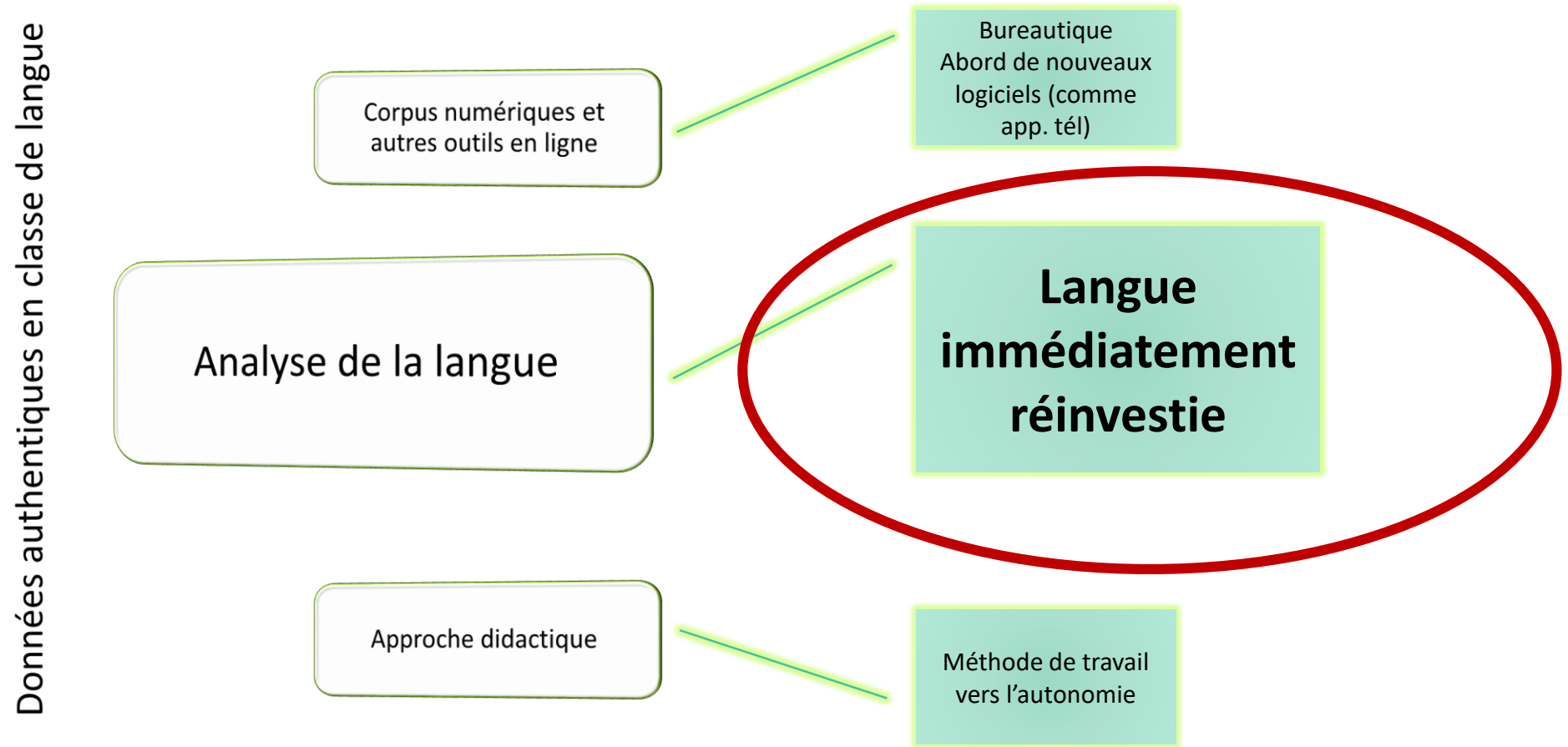
CORLI : plusieurs corpus dont Scientext
Lexicoscope
Lexiquum
Lextutor
Ortolang : plusieurs corpus à partir du CNRTL
Valibel : Orfeo
Varitext
Webcorp
Wortschatz : Leipzig Corpora Collection

ORAL

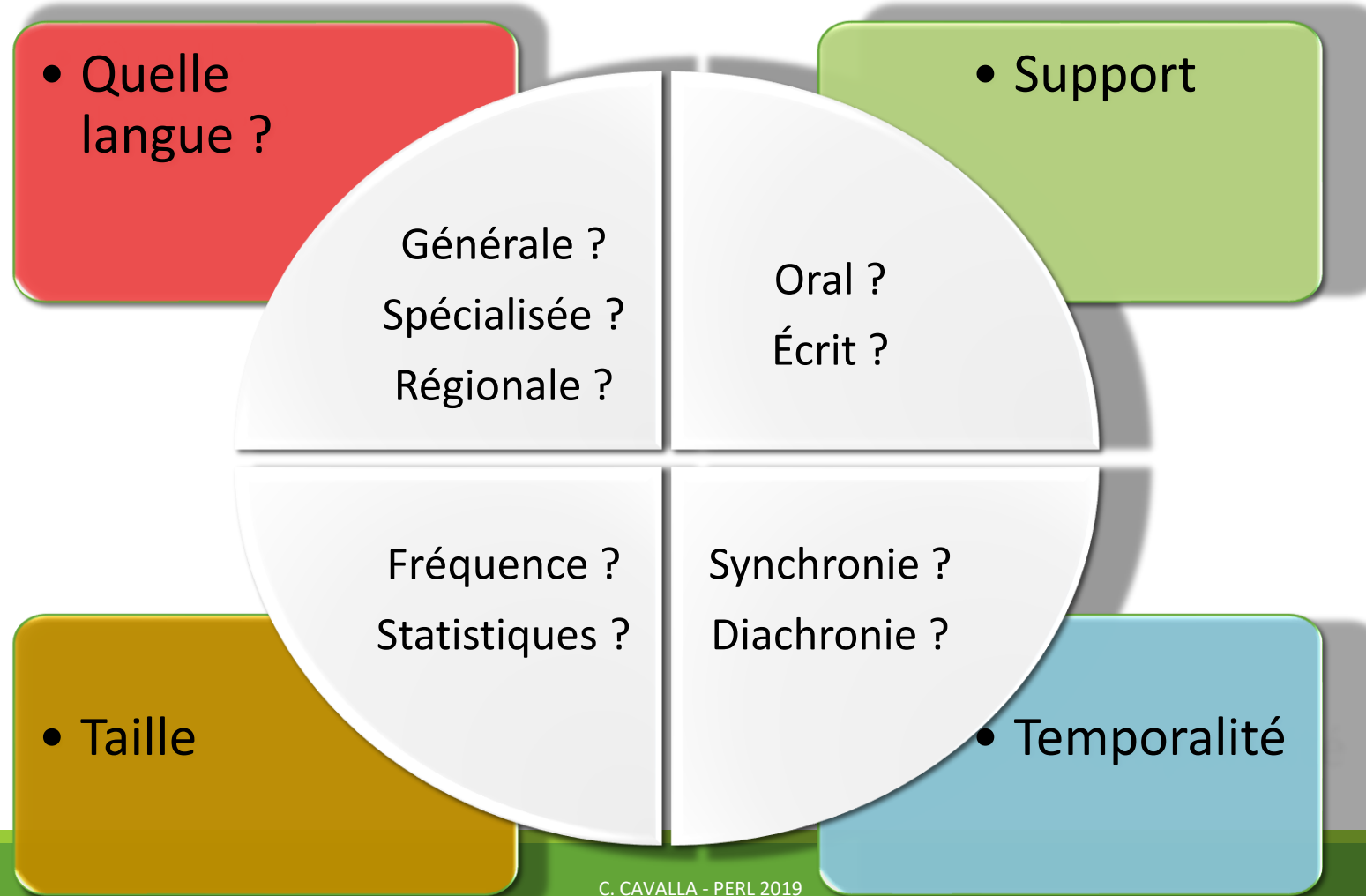
ESLO2
IPFC: Floral-PFC
Florale
ORFEO : plusieurs corpus dont CFPP2000, CLAPI (lien avec CIEL-F), CRFP, C-ORAL-ROM, Fleuron, French oral narrative, OFROM, TCOF

Corpus d'apprenants : en France et à l'étranger

Éléments en présence



5. Comment choisir les corpus ?



Critères de choix des corpus numériques pour l'enseignement des langues

CRITÈRES CORPUS ÉCRITS	CRITÈRES CORPUS ORAUX
Types de documents : presse, littérature, articles scientifiques...	Types d'enregistrements : audio, vidéo / transcription ou pas
Variation : région, authenticité	
Taille : fréquence, représentativité	
Date : synchronie, diachronie	
Pour qui et pour quoi la création du corpus	
Scripteurs / Locuteurs : natif ou allophone / âge / sexe / mono-multilingue	

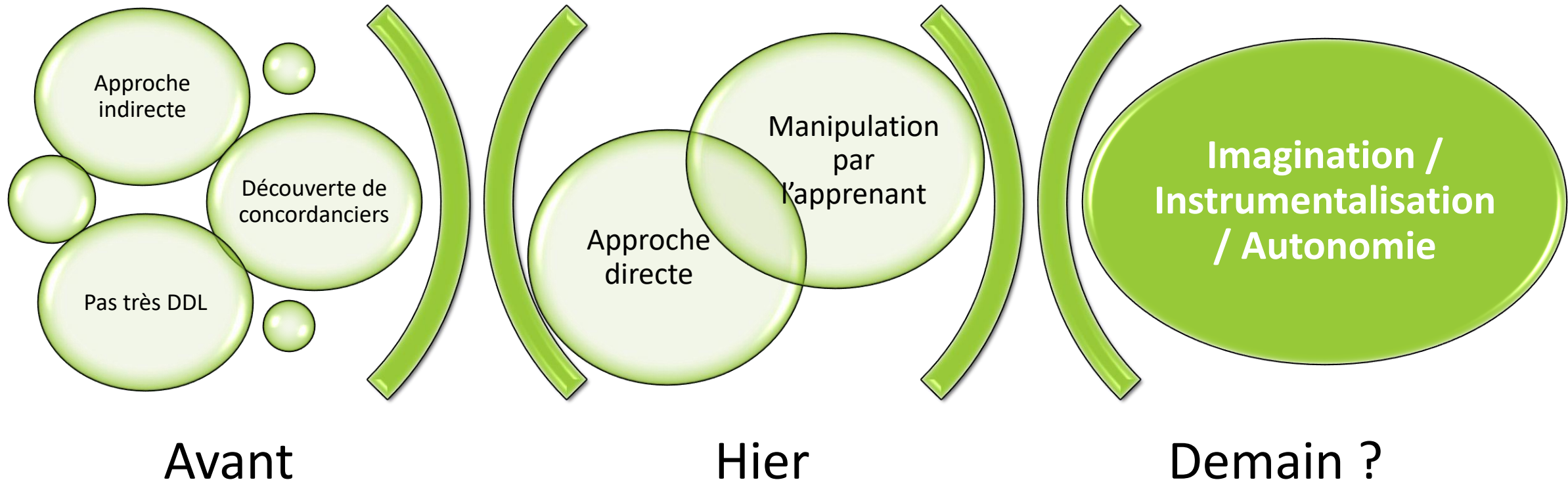
Présentation

Corpus numériques



1. Présentation liminaire
2. Qu'est-ce qu'un corpus en linguistique de corpus ?
3. Pourquoi intégrer les corpus en classe de langue ?
4. Comment intégrer les corpus en classe de langue ?
5. Comment choisir les corpus à intégrer ?
6. Conclusion

Pour conclure



**ENTRE
PRÉSENCE ET
DISTANCE**

Enseigner et apprendre
les langues à l'université
à l'ère numérique

2ÈME COLLOQUE DU PERL
12 - 13 - 14 DÉCEMBRE 2019

PERL
PÔLE D'ÉLABORATION
DE RESSOURCES LINGUISTIQUES

U^{SPC}
Université Sorbonne
Paris Cité



Merci

Corpus numériques : critères pour l'enseignement des langues

CRISTELLE CAVALLA

cristelle.cavalla@sorbonne-nouvelle.fr

