



Questions autour des *corpus* en didactique des langues-cultures

*Linguistique de corpus à la croisée de
questionnements théoriques, méthodologiques et
empiriques*

Journée Thématique Transversale de l'Atilf - 22 mars 2024

Organisation : Hélène Vinckel-Roisin et Yvon Keromnes

Dominique Macaire et Virginie Privas-Bréauté

Atilf, Équipe *Didactique des langues et sociolinguistique*



Plan de la présentation

1. Qu'est-ce qu'un *corpus* en didactique des langues-cultures ?
 2. Des *focales* contrastées au sein de la DDLC
 3. Types de centrations favorables au transfert de connaissances
 4. Principes pour le *choix des corpus* en DDLC
 5. Un exemple de corpus d'interactions à l'école maternelle
 6. Conclusion
-

1. Qu'est-ce qu'un *corpus* en *DDL*?

En linguistique de corpus

Des données/“faits”
linguistiques de grande taille
rassemblés et assortis d'outils
d'exploration puissants

=> pour

1. observer finement des phénomènes
2. rechercher des régularités
3. développer des ressources : dictionnaires, grammaires descriptives...
4. tester des hypothèses; confronter des modèles postulés aux réalisations effectives

En didactique des langues-cultures (DDL)

Des données utilisables à un double niveau :

- en recherche en DDL (objet)
- dans les pratiques d'appropriation (moyen)

⇒ pour comprendre et pour agir

1. étudier les conceptions des langues-cultures (invisible)
2. analyser l'activité individuelle et sociale des apprenants dans une écologie du milieu (situations formelles vs informelles ; à distance vs en présentiel; etc.), mesurer des effets (linguistiques, etc.)
3. comprendre les processus d'appropriation et identifier
4. les obstacles/dilemmes potentiels (linguistiques, pragmatiques, cognitifs, culturels, etc.), les incidents
5. Développer, combiner, répliquer des tâches, en mesurer pertinence et/ou influence (ergonomie didactique)

2. Des *focales* contrastées au sein de la DDLC

Les langues en tant que langue-objet ou langue-outil

systèmes lexico-syntaxiques que l'élève apprend à maîtriser essentiellement pour communiquer

« Communiquer c'est utiliser un code linguistique rapporté à une action dans un contexte socio-culturel et linguistique donné » (CECRL, 2001, 2008).

- ⇒ corpus linguistiques (liés à des savoirs, des aspects de la langue, du discours) et à leurs difficultés d'apprentissage
- ⇒ corpus liés à des variations de la langue (accents, lexique)
- ⇒ corpus liés à des compétences

Les langues en tant que langue-sujet dynamique

organismes complexes faisant partie intégrante du vivant, dynamiques (Trocme-Fabre, 1996)

Répertoires plurilingues (Lüdi, 2021, Molinié, 2011)/hybrides (Piccoli, 2021) et interlangue (Selinker 1972) => corpus de productions d'apprenants

Vision compréhensive et non cumulative des langues-cultures (Behra et Macaire, 2020) => corpus de comparaisons entre langues ; représentations

Divers langages expressifs pour « incorporer » les langues (Varela, 1993), notamment théâtral,

2.1 Effets des deux focales en DDLC

1. Typologies de corpus en DDLC

- **Données très variées, souvent combinées et hiérarchisées:** orales / écrites / vidéo / audio / (non) verbales / questionnaires / entretiens / focus groups / visuels / Albums de jeunesse / cahiers de bord / portfolios/ tâches collaboratives /traces ... Ex.: Projet pré-professionnalisant interculturel MIPFE: Dessiner la mobilité d'étudiants Hongkongais; données centrales ou périphériques
- **Temporalité relative à l'appropriation:** corpus constitués en amont (input), (co-)élaborés en situation de pratiques (output), fortuits (imprévus, obstacles). Ex. : projet interculturel INFRAN (France-Allemagne, blog et multimodalité)

=> **Corpus = des construits de divers acteurs de la recherche, et organisés en vue de leur échange entre chercheurs (recherche) et avec les partenaires (enseignement).**

2.2 Effets des deux focales en DDLC

2. Constitution et analyse des *corpus* en DDLC

- **taille des corpus** : généralement assez limitée, mais description précise indispensable
 - **outils d'analyse** : peu nombreux, peu spécialisés, maîtrisés de manière variable, avec un codage souvent manuel
 - **analyses souvent croisées** : quanti, quali, mixtes (encore appelées “triangulation”). Ex.: corpus de laboratoire et pour la classe de langues sur le parler des adolescents de plusieurs pays avec plusieurs types de données (Catcod, 2008, réseau RISC/CNRS)
 - **questions éthiques** : Ressources du projet **MULCE** en open source, Ressources du Consortium **CORLI**, y compris de sa charte éthique; méthodologie **CORINTE** (ICAR, Lyon); Mondada, 2005, http://icar.cnrs.fr/projets/corvis/PDF/Mondada05_ethiqueTerrain.pdf ; Guide **DGLFLF**, <http://www.culture.gouv.fr/culture/dglf/cotelecharger.htm>)
-

2.3 Effets des deux focales en DDLC

3. Choix des logiciels de traitement en fonction des objectifs (recherche ou enseignement-apprentissage)

- Logiciel *Voyant-tools* (<https://voyant-tools.org>)
- Les plus utilisés : 5 logiciels parmi les 24 plus fréquents pour l'oral et les 17 pour l'écrit (TIGR Huma-Num):
 - *MaxQda* (*quali: analyse conversationnelle ou de contenu etxtuel*),
 - *IRaMuTeQ* (*quanti: questionnaires*)
 - *Nvivo* et/ou *antcont* (analyse textuelle, avec *transcriber*)
 - *Elan* (*corpus vidéo*)

En DDLC, voir par exemple Chanier et Ciekanski, 2010; Boré et Elalouf, 2017.

3. Types de centrations en DDLC favorables au transfert de connaissances et issues de la recherche

1. le **savoir, de nouvelles connaissances** de diverses natures (*knowledge push*)
 2. l'**usager** (apprenant) face à un **problème à résoudre** (*demand pull model* ou *problem solving model*)
 3. des **besoins pour l'apprenant** et des **contextes de choix et mise en oeuvre des résultats** (*dissemination model*)
 4. la **co-construction par les acteurs, le plus souvent des apprenants** (*modèle de l'interaction sociale*) qui mobilise une synergie des approches sur les corpus.
- ⇒ Identifier ces focales comme des **filtres préalables** aux *constitutions de corpus en DDLC*
- ⇒ **Analyse des fréquences concordante en français**: à partir de synthèses de colloques; d'1 corpus d'ouvrage: Boulton, 2024; d'un corpus sur articles en français: de Chambers-Lebaron, 2007.

4. Principes pour le *choix des corpus* en recherche en DDLC

1. **Principe d'intentionnalité** : Qu'ai-je besoin de savoir? Les corpus sont-ils les plus pertinents? Quelle est leur place à côté d'autres données?
2. **Délimitation des corpus** : niveaux, ouverture, contrôle, consentement éclairé
3. **Élaboration d'une méthodologie spécifique questions/hypothèses**
4. **Posture de recherche** : tension entre positionnement épistémologique, les données (terrains) et la façon dont on les traite, choix conscientisés
 1. **paradigme de l'écologie des milieux** (Bronfenbrenner, 1979)
 2. **paradigme de la complexité** (Morin, 1999; Le Moigne, 2005) **et des systèmes dynamiques** (Larsen-Freeman, 2000)

(Behra et al., 2024: 22-24)

5. Un exemple de corpus (Corpus 1)

Interactions à l'école maternelle

Corpus 1 : enseignante P + élèves E + allophone A

Les enfants (E), l'allophone (A) et l'enseignant (P) : Aujourd'hui, on est mardi. Nous n'avons pas d'absents mais il fait gris.

E1 : Aujourd'hui, on est lun...

P montre par un geste des mains que ce n'est pas lundi et montre l'étiquette du mardi.

E2 : Mardi

E3 : Aujourd'hui, on est mardi, il n'y a pas d'absents et il fait gris.

E4 : Aujourd'hui, on est...

Plusieurs E : On est mardi. On n'a pas d'absents et...

P et plusieurs E : Il fait gris

E5 à E9 chacun leur tour : Aujourd'hui, on est mardi, il n'y a pas d'absents et il fait gris.

P : Attention, A !

A : Aujourd'hui on est lundi.

P hoche la tête pour dire non

Plusieurs E : Mardi

A : Aujourd'hui, on est mardi. Il y a un absent.

P : Mais non ! Tu répètes ce qui s'est passé hier! Au-jour-d'hui A !

A : Aujourd'hui, on est.... *P abandonne et oublie A. Il fait un geste d'énervement et poursuit avec le reste de la classe...*

(Carol, Macaire et Behra, 2016)

5. Un exemple de corpus (Corpus 2 et Corpus 3) Interactions à l'école maternelle

CORPUS 2 entre pairs (4ans) dont Serge, allophone

Selim: T'as vu ? Y a du sable.

Vera :Voilà moi j'ai fini.

Voilà tiens ! Ça c'est pour toi ! Ça c'est pour toi !

Serge :

Regarde là ! *pointant la lettre R*. Ça veut dire t'es Rayana.

Oh je dépasse encore. Ça ne veut pas arrêter.

CORPUS 3

Enseignant : C'est moi qui vais vous mettre le point.

Pierre : Parce que elle, elle ne sait pas, elle.

Rire de Flora

P : Non moi qui, moi je sais, moi. C'est la maitresse qui met le point.

Flora et Pierre *ensemble* : Ouais.

(Carol, Macaire et Behra,2016)

5. Conclusion

- Les *corpus* en DDLC ne sont **ni neutres, ni objectifs** en DDLC, puisque liés à l'humain (les divers acteurs, dont les langues-cultures).
- Questionner **en amont** de la constitution du corpus en DDLC
 - les épistémologies, les notions, les objets, les méthodes en recherche;
 - les démarches privilégiées pour l'appropriation des langues-cultures;
 - la prise en compte des expérientiels des acteurs et de leurs contextes.
- Pluridisciplinarité **et** pluricatégorialité des chercheurs sont **indispensables en DDLC : linguistes, ingénieurs, informaticiens, psychologues, sociolinguistes, didacticiens, praticiens de l'École**

Bibliographie sélective 1/4

<https://bu.univ-lorraine.fr/ressources-en-ligne>

- Aden, J. (2017). Langues et langage dans un paradigme énonctif, RDLC, 14.1. URL : <http://journals.openedition.org/rdlc/1085> ; DOI : <https://doi.org/10.4000/rdlc.108>
 - Behra, S., Ciekanski, M., Nassau, G. & Macaire, D. (2024). Cherchez le corpus! Pratiques d'utilisation de corpus en didactique des langues dans la formation par la recherche de futurs enseignants. Dans Virginie Privas-Bréauté (dir.). *Du recueil de données à l'analyse des corpus en didactique des langues*, pp.19-38. PUR.
 - Boré, C., Elalouf, M.-L. (2017). Deux étapes dans la construction de corpus scolaires : problèmes récurrents et perspectives nouvelles, *Corpus* [Online], 16 |. <http://journals.openedition.org/corpus/2731> ; DOI: <https://doi.org/10.4000/corpus.2731>
 - Cadet, L. & Carlo, C. (2015). Questions autour des corpus de recherche en didactique des langues, *Recherches en didactique des langues et des cultures* [En ligne], 12-2 | 2015. <http://journals.openedition.org/rdlc/611>. Eschenauer, S., Tellier, M., Zappa, A. (2022). Incorporer les langues vivantes : reconnaître la place du corps pour enseigner et pour apprendre. *Travaux interdisciplinaires sur la Parole et le Langage. Panorama des recherches au Laboratoire Parole et Langage*, 38.
 - Carol, R., Macaire, D., Behra, S. (2016). « Du quotidien communicatif d'enfants allophones en classe maternelle ». Dans : Cadet, Lucile, Pégaz Paquet, Anne (dir.). *Les langues à l'école, la langue de l'école*. Arras, Artois Presses Université, collection « Études linguistiques », série « Didactique des langues », 89-103.
-

Bibliographie sélective 2/4

- Chanier, T. & Ciekanski, M. (2010). Utilité du partage des corpus pour l'analyse des interactions en ligne en situation d'apprentissage: un exemple d'approche méthodologique autour d'une base de corpus d'apprentissage. *ALSIC - Apprentissage des Langues et Systèmes d'Information et de Communication*, 2010, 13. <http://alsic.revues.org/1666>
- Chanier, T. & Wigham, C.R. (2011). (Dir.) Learning and Teaching Corpus ARCHI21. Mulce.org : Clermont Université. <http://repository.mulce.org>
- Dagenais, C. (2006). Vers une utilisation accrue des résultats de la recherche par les intervenants sociaux? Quels modèles de transfert des connaissances privilégiés? *Les Sciences de l'éducation-Pour l'ère nouvelle*, vol. 30/3, 23-35.
- Gagnon, R. & Bourhis, V. (2020). Les corpus parlés et leur didactisation : quelle parole (ap)prise dans l'espace de la classe ? Introduction. *Éla. Études de linguistique appliquée*, 198, 133-136. <https://doi.org/10.3917/ela.198.0005>
- Garric, N. & Longhi, J. (2012). L'analyse de corpus face à l'hétérogénéité des données : d'une difficulté méthodologique à une nécessité épistémologique. *Langages*, 187, 3-11. <https://doi.org/10.3917/lang.187.0003>
- Granger S., Vandeventer A. & Hamel M.-J. (2001). Analyse des corpus d'apprenants pour l'ELAO basé sur le TAL. *Traitement automatique des langues* 42(2). Hermès, 609-621. (voir Revue volume 1, 1998. <https://cahiersforell.edel.univ-poitiers.fr/index.php?id=595#ftn2>)
- Granger, S. (2018). Didactique des langues étrangères, linguistique de corpus et traitement automatique des langues, *Cahiers FoReLLIS - Formes et Représentations en Linguistique, Littérature et dans les arts de l'Image et de la Scène* [En ligne], Revue papier (Archives 1993-2001), Questions d'épistémologie en didactique du français, AXE 2 : RELATIONS AVEC LES DISCIPLINES CONNEXES, mis à jour le : 13/09/2018, URL : <https://cahiersforell.edel.univ-poitiers.fr:443/cahiersforell/index.php?id=595>.
- Larsen Freeman, D. (2000 2nd Edition). *Techniques and Principles in Language learning*. OUP. Disponible en ligne: https://www.academia.edu/34135093/Diane_Larsen_Freeman_SECOND_EDITION
- Lüdi, G. (2021). Genèse et représentations de répertoires plurilingues, *Recherches en didactique des langues et des cultures*, 18-2 | 2021. <http://journals.openedition.org/rdlc/9485>

Bibliographie sélective 3/4

- Macaire, D. & Behra, S. (2015). *Les captations numériques (vidéo) au service de l'analyse des usages de la langue et des pratiques de classe à l'école maternelle : le cas des enfants dont le français n'est pas la langue de la maison*, Rapport d'orientation scientifique du projet de recherche Kidil-L Atilf (non publié).
- Macaire, D. & Boulton, A. (2014). Notion de corpus et didactique des langues - Éditorial, *Recherches en didactique des langues et des cultures* [En ligne], 11-1 | 2014. <https://journals.openedition.org/rdlc/1665>
- Moiragh, É. (2018, 16 avril). Explorer des corpus d'images. L'IA au service du patrimoine. *Carnet de la recherche à la Bibliothèque nationale de France*. Consulté le 20 mars 2024, à l'adresse <https://doi.org/10.58079/m3cp>; <https://bnf.hypotheses.org/2809>
- Molinié, M. (2011). La méthode biographique : de l'écoute de l'apprenant de langues à l'herméneutique du sujet plurilingue. Agence Universitaire de la Francophonie. Guide pour la recherche en Didactique des langues. Approches contextualisées, Éditions des Archives Contemporaines, pp.Ps 144-154. <https://univ-sorbonne-nouvelle.hal.science/hal-01448684/file/MethodeBiographique2010%20Molinié.pdf>
- Py, B. (1996). Les données et leur rôle dans l'acquisition d'une langue non maternelle, *Les Carnets du Cediscor*. <https://doi.org/10.4000/cediscor.402>
- Schaeffer-Lacroix, É. (2018). *Corpus et didactique de l'allemand- La langue à bras le corps*. Lambert-Lucas.
- Selinker, L. (1972). Interlanguage, *International Review of Applied Linguistics* X-3, 209-231.
- Trocme-Fabre, H. (2022). *L'arbre du savoir-apprendre. Vers un référentiel cognitif. The Art of Learning and The Knowledge Tree: Toward a Cognitive Framework*. Édition bilingue. Éditions Le Manuscrit.
- Trocme-Fabre, H. (1996, 2000) *L'arbre du savoir-apprendre*. Éditions Être et connaître. (épuisé).

Bibliographie sélective 4/4

- Varela, F. , Thomson, E. & Rosch, É. (1993). *L'Inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*. Seuil.
- Véronique, D. (2005). Les interrelations entre la recherche sur l'acquisition du français langue étrangère et la didactique du français langue étrangère, *Acquisition et interaction en langue étrangère* [En ligne], 23 | 2005. <https://doi.org/10.4000/aile.1707>
- Wagener, A. (2022). Du corpus comme ressource au corpus comme sample : de l'exploitation néolibérale de la donnée à une écosystémique du discours. *Éla. Études de linguistique appliquée*, 205, 19-33. <https://doi.org/10.3917/ela.205.0023>
- Wigham, C. & Bayle, A. (2012). Enjeux, outils et méthodologie de constitution de corpus d'apprentissage. *Coldoc 2012 : Traitements de corpus : outils et méthodes*, Oct 2012. edutice-00710698

Ressources institutionnelles européennes

- Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues vivantes*, CECRL. <https://rm.coe.int/16802fc3a8>
- Conseil de l'Europe (2018). *Cadre européen commun de référence pour les langues vivantes, Volume complémentaire*. <https://rm.coe.int/cecr-volume-complementaire-avec-de-nouveaux-descripteurs/16807875d5>



Annexe : Logiciels d'analyse de corpus en DDLC

CORLI <http://explorationdecorpus.corpusecrits.huma-num.fr/>;
[http://explorationdecorpus.corpusecrits.huma-num.fr/outils-
logiciels-corpus-oraux-et-multimodaux/](http://explorationdecorpus.corpusecrits.huma-num.fr/outils-logiciels-corpus-oraux-et-multimodaux/) (Les auteurs consultés
respectent la charte CORLI)

CORINTE (pour l'analyse conversationnelle en relation avec les
travaux de Bange)

<http://icar.cnrs.fr/corinte/enregistrement-des-corpus-dinteractions/>

Merci de votre attention!

