

Lexique

ISSN : 2804-7397

Éditeur : Université de Lille

37 | 2025

**Le lexique préfabriqué des interactions : des descriptions linguistiques et
outillées à l'application didactique**

Une ressource de phraséologie scientifique orale en français et quelques pistes d'application didactique

*A Resource for French Oral Academic Phraseology and Its Didactic
Applications*

Chaeyoung Lee

 <https://www.peren-revues.fr/lexique/2041>

DOI : 10.54563/lexique.2041

Référence électronique

Chaeyoung Lee, « Une ressource de phraséologie scientifique orale en français et quelques pistes d'application didactique », *Lexique* [En ligne], 37 | 2025, mis en ligne le 31 décembre 2025, consulté le 25 février 2026. URL : <https://www.peren-revues.fr/lexique/2041>

Droits d'auteur

CC-BY

Une ressource de phraséologie scientifique orale en français et quelques pistes d'application didactique

A Resource for French Oral Academic Phraseology and Its Didactic Applications

Chaeyoung Lee

PLAN

1. Introduction
2. Phraséologie et discours scientifiques écrits et oraux
 - 2.1. Formules discursives : des marqueurs discursifs aux routines discursives
 - 2.2. Discours scientifique et genre de la communication scientifique orale
3. Élaboration d'une ressource phraséologique de l'oral scientifique en français
 - 3.1. Le CComSciO, corpus de discours scientifique oral en français
 - 3.2. Typologie des formules de prise en compte de l'allocutaire à l'oral scientifique
4. Un bref examen des ressources didactiques numériques déjà existantes
 - 4.1. Ressources lexicographiques de l'anglais académique
 - 4.2. Outils élaborés à partir d'un corpus d'écrits scientifiques en français
 - 4.3. Ressources didactiques du français parlé en FLE
5. Un exemple de présentation des fonctions et de simplification des étiquettes de notre ressource
6. En guise de conclusion

NOTES DE LA RÉDACTION

Reçu : Mars 2025 / Accepté : Septembre 2025

Mise en ligne : Décembre 2025

TEXTE

1. Introduction

- ¹ Les textes scientifiques, qu'il s'agisse d'articles de revue, de mémoires ou de thèses, recourent à une grande diversité de lexiques préfabri-

qués (Wray, 2002 ; Tutin, 2018), parmi lesquels on peut distinguer des connecteurs (*dans un premier temps, en d'autres termes*), des collocations lexicales (*jouer un rôle, présent article*) ou des locutions (*point de vue, mettre en lumière*). On y observe aussi des phrases stéréotypées à fonction discursive ou rhétorique (*nous avons vu que...*, *nous l'avons vu, comme on l'a vu* pour le rappel d'un élément déjà mentionné), relevant de ce que Tutin et Kraif (2016), puis Ji (2022), appellent des « routines sémantico-rhétoriques »¹. Il s'agit des séquences, récurrentes dans un genre spécifique et sémantiquement prédictibles, qui autorisent des variations à la fois syntagmatiques et paradigmatiques sur le plan énonciatif. Représentant des façons de dire spécifiques, ces expressions servent non seulement à assurer la fluidité du discours, mais également à construire l'identité discursive de l'auteur au sein de sa communauté scientifique, à travers l'emploi d'un langage codé. Elles contribuent ainsi, par leur fréquence, à caractériser le genre scientifique écrit, en mettant en avant ses spécificités structurelles, rhétoriques, énonciatives ou interactionnelles².

- 2 Les premières études sur les routines sémantico-rhétoriques se sont plutôt concentrées sur les écrits scientifiques (Tutin, 2018 ; Wang, 2022 ; Ji, 2022), particulièrement pour l'application didactique de l'écrit universitaire (Yan, Tutin & Tran, 2018). Toutefois, cette routinisation se manifeste également dans l'oral scientifique, où l'on observe des routines exprimant un ensemble de fonctions discursives et rhétoriques comme l'annonce de l'objectif de la présentation (*nous allons vous présenter ...*), l'appui sur des savoirs partagés (*on sait que ...*) ou le rappel (*comme je vous l'ai dit ...*). Par ailleurs, au-delà de ces constructions phrastiques, qui peuvent être réalisées de plusieurs façons sur le plan lexical, on peut aussi identifier un autre type d'énoncé polylexical, si vous voulez, dont le figement morphosyntaxique est plus marqué. Celui-ci se rapproche d'un « marqueur discursif propositionnel » (Andersen, 2007) ou bien d'une « phrase préfabriquée des interactions » au sens de Tutin (2019), fréquents dans la langue parlée, exprimant certaines fonctions dans le contexte d'interactions. Ce sont ces deux types de formules préconstruites qui nous intéressent en particulier, en raison de leur rôle discursif et interactionnel au sein du discours. Leur étude s'inscrit dans le cadre de la recherche de la phraséologie du discours scientifique oral.

- 3 Cet article a pour objectif de présenter une ressource phraséologique portant sur les formules de prise en compte de l'allocutaire dans l'oral scientifique (Lee, 2024), ainsi que de réfléchir à quelques pistes pour son application didactique. Dans une première partie, nous définirons les formules discursives et proposerons un bref état de l'art sur la phraséologie scientifique orale. Dans une seconde partie, nous présenterons notre corpus de discours scientifique oral en français, ainsi que la typologie fonctionnelle de formules que nous en avons dégagée. Enfin, certaines pistes seront étayées par l'examen d'outils numériques existant dans des thématiques voisines, parmi lesquelles nous explorerons particulièrement la simplification des étiquettes de formules.

2. Phraséologie et discours scientifiques écrits et oraux

2.1. Formules discursives : des marqueurs discursifs aux routines discursives

- 4 Afin de mener une étude phraséologique dans un contexte marqué par un foisonnement de terminologies et de notions émergentes (Legallois & Tutin, 2013, pp. 3-4), il est essentiel de choisir un objet adapté aux questions de recherche et d'en définir les contours. Ayant pour objectif de mettre en lumière les routines linguistiques propres à l'oral scientifique en français, faits linguistiques relativement moins explorés que leurs équivalents écrits (Jacques & Tutin, 2018 ; Luodonpää-Manni, Grossmann & Tutin, 2022), nous proposons d'étudier une catégorie que nous désignons par les « formules discursives »³ (Lee, 2024). Cette catégorie regroupe deux types de constructions polylexicales, hétérogènes en termes de degré de préfabrication, mais qui ont pour point commun de remplir certaines fonctions discursives, rhétoriques et/ou pragmatiques dans le contexte. Précisons que notre étude se limite aux routines intégrant un prédicat de sens plein (par ex. *dire, vouloir, savoir*) qui nous paraît jouer un rôle essentiel dans l'actualisation de leur valeur, en excluant la copule *être* et la locution *il y a*.

Formules discursives	Définition	Exemples
Type 1	Suites de mots complètes et peu variables sur le plan morpho-syntaxique, très souvent parenthétiques, qui peuvent apparaître dans un grand nombre de genres oraux avec certaines fonctions	<i>on va dire, si vous voulez, vous savez, vous voyez, ...</i>
Type 2	Patrons lexico-sémantico-syntaxiques, qui intègrent à la fois des éléments fixes et variables sur le plan paradigmatique, ainsi que des éléments obligatoires et/ou facultatifs sur le plan syntagmatique, qui permettent différentes réalisations sur le plan énonciatif, et qui remplissent certaines fonctions caractérisant un genre spécifique	[(comme) + Sujet locuteur de référence + (le en) + Prédicat de communication + (de sur) + (Syntagme nominal complétive que) + (Marqueur temporel)] : <i>on en a parlé ce matin, comme le disait la présentatrice précédente, ...</i>

Tableau 1. Définition et exemples des deux types de la catégorie de formules discursives

- 5 Le premier type se rapproche de la notion de marqueurs discursifs propositionnels d'Andersen (2007) : il est généralement invariable et apparaît dans une position parenthétique. Si ce type de formules peut aussi être observé dans d'autres contextes d'interactions orales, les fonctions du type 2, plus souples du point de vue du figement, sont, quant à elles, étroitement liées aux enjeux discursifs ou rhétoriques du genre scientifique, tels que le guidage du lecteur (Ji, 2022). Ce second type se rapproche ainsi de la notion de routines sémantico-rhétoriques (Section 1), qui a été principalement explorée à l'aide de l'extraction automatique des Arbres Lexico-syntaxiques Récurrents⁴ (Tutin & Kraif, 2016), dans le cadre de l'analyse phraséologique des genres scientifiques écrits. Tout comme les routines, ce type de formules comporte certains éléments facultatifs (par ex. le subordonnant *comme*, le marqueur temporel), et des éléments susceptibles de varier selon le contexte énonciatif : par exemple, un prédicat de communication comme *dire* et *parler*, ou encore l'élément correspondant au locuteur de référence (*on* / *la présentatrice de référence*) (voir les exemples dans le Tableau 1). Nous pensons qu'étudier ces deux types de constructions nous permet une description riche et complète de la phraséologie pragmatique de l'oral scientifique.
- 6 Un autre critère définitoire des formules discursives réside dans leur récurrence significative au sein d'un corpus. Ce critère d'ordre quan-

titatif a déjà été retenu pour la définition d'autres phraséologismes, tels que les motifs (Longrée & Mellet, 2013), les routines (Ji, 2022), ou bien les « *lexical bundles* » en anglais (Biber & Barbieri, 2007). L'interprétation de la significativité de la fréquence peut toutefois varier d'une étude à l'autre. Dans notre cas, bien que le corpus mobilisé soit de taille modeste (Section 3.1), nous avons retenu un seuil de fréquence de 3 occurrences pour l'ensemble du corpus. Ce choix repose sur le fait que les patrons de routines présentent souvent une certaine longueur, composée d'au moins trois éléments lexicaux⁵.

2.2. Discours scientifique et genre de la communication scientifique orale

7 L'objectif principal d'un discours scientifique⁶ est de communiquer les savoirs acquis au cours de la recherche, en respectant les procédures préétablies par la « communauté discursive » (Swales, 1990) de chaque discipline. Relevant des langues de spécialité, la langue du discours scientifique a déjà fait l'objet de nombreux travaux en linguistique française, sous différents angles (Kocourek, 1991 ; Fløttum, Dahl & Kinn, 2006 ; Hatier, 2016 ; Jacques, 2017, entre autres). Plus particulièrement, la plupart des études lexicographiques se sont d'abord intéressées aux genres écrits. L'attention portée à la dimension lexicographique de l'oral est plutôt récente, même dans le domaine de l'*English for Academic Purposes* (EAP), où l'oral a suscité l'intérêt des chercheurs plus tôt qu'en contexte francophone, notamment pour les genres pédagogiques tels que les cours magistraux ou les travaux dirigés (Dang, Coxhead & Webb, 2017 ; Li, Xu & Zhang, 2024). En ce qui concerne le Français sur Objectifs Universitaires, la facette orale du discours scientifique que les étudiants doivent maîtriser reste aussi encore peu explorée, à l'exception d'une étude récente de Tran, Masset-Martin et Mangiante (2023).

8 Le genre des communications scientifiques orales peut être considéré comme représentatif de l'oral scientifique, dans la mesure où il constitue l'un des principaux moyens par lesquels un chercheur présente régulièrement son travail (Carter-Thomas & Jacques, 2017), au même titre que l'article dans la modalité écrite. Ce genre se distingue par plusieurs traits discursifs, parmi lesquels une production du discours organisée, mais généralement spontanée (Jacques,

2017), une interactivité immédiate entre le locuteur et l'allocutaire (Rinck, Fløttum & Poudat, 2017), et une communication à dominante multimodale, entre autres. On peut donc supposer que ces caractéristiques ont un effet plus ou moins important sur la manière dont le locuteur construit son discours à l'oral et sur les expressions déjà toutes faites qu'il utilise, contrairement à l'écrit. Par exemple, certaines routines syntaxiques sont propres à l'oral, quasi absentes à l'écrit, telles que les pseudo-clivées *ce qui/que* - *P c'est X*, les structures en *il y a*, et celles en *si* + pronom personnel (*si on veut être précis*), d'après Carter-Thomas et Jacques (2022). La prédominance de celles-ci est, selon les autrices, liée à la nécessité de hiérarchiser de nombreuses informations ou de capter l'attention de l'auditoire à l'oral. Siepmann (2023) a souligné, quant à lui, la variété de fonctions méta-discursives qu'un même verbe remplit à l'oral : par ex. le verbe *dire* pour le rappel (*ce qui a été dit*), pour l'appui sur des connaissances conventionnelles (*on a souvent dit que...*) ou pour la reformulation (*on va dire*).

3. Élaboration d'une ressource phraséologique de l'oral scientifique en français

3.1. Le CComSciO, corpus de discours scientifique oral en français

- 9 Parmi les corpus oraux en français, les genres académiques ou scientifiques sont relativement rares, hormis quelques corpus comme le corpus ESLO (Baude & Dugua, 2016), qui intègre un module des conférences universitaires, ou le corpus issu du projet EIIDA⁷ (Carter-Thomas & Jacques, 2017). Ce dernier inclut des présentations scientifiques orales de format standard (en moyenne 27 minutes), collectées selon des langues et des disciplines dans une perspective de mise en évidence des particularités discursives propres à chaque langue et discipline, que nous considérons comme relevant du genre des communications scientifiques orales. Une partie de ces données - 15 communications en langue française en linguistique - a été inté-

grée à notre corpus, dans une démarche de constitution d'un corpus équilibré par disciplines⁸.

- 10 Notre corpus d'étude, le CComSciO (Corpus de Communications Scientifiques Orales), réunit 60 communications dans trois disciplines des sciences humaines et sociales (SHS), que sont la linguistique, les sciences de l'information et de la communication, ainsi que la didactique des langues. Nous nous intéressons ici particulièrement à la phraséologie du discours scientifique en SHS, à l'instar d'autres travaux (Yan, 2017 ; Jacques & Tutin, 2018 ; Ji, 2022), en supposant que les discours scientifiques dans ce domaine se distinguent de ceux en sciences dites *dures* (Jacques, 2017), y compris à l'oral. Chaque discipline comprend 20 présentations, issues de différents colloques ou journées d'études, d'une durée de 15 à 42 minutes, ce qui représente un corpus total d'environ 26 heures, soit 262 824 mots⁹.
- 11 Pour l'élaboration du CComSciO, nous avons adopté la transcription orthographique standard, proposé par le groupe VALIBEL (Bachy, Dister, Francard, Geron, Giroul, Hambye, Simon & Wilmet, 2007), puis la segmentation par les constructions verbales, suivant le guide du corpus Gold d'ORFEO, rédigé par A. Valli et J. Deulofeu. Intégré au logiciel ELAN avec un alignement entre le son et le texte¹⁰, le CComSciO est consultable à l'heure actuelle sur le Lexicoscope 2.0 (Kraif & Diwersy, 2012), outil d'exploration de corpus numérique¹¹. Sa diffusion sur d'autres plateformes numériques comme ORTOLANG est aussi en préparation pour le rendre accessible à l'ensemble de la communauté scientifique.

3.2. Typologie des formules de prise en compte de l'allocutaire à l'oral scientifique

- 12 L'étude de la phraséologie du genre scientifique ou académique dans une approche didactique vise généralement à établir une typologie fonctionnelle pour les expressions à maîtriser. Parmi plusieurs typologies déjà élaborées, deux ont particulièrement nourri notre réflexion pour la classification des formules de prise en compte de l'allocutaire à l'oral : celle du métadiscours académique écrit et oral d'Ädel (2010) en anglais, et celle des routines de guidage du lecteur,

développée par Ji (2022) à propos des articles en français. Dans son modèle, Ädel (2010) a souligné des fonctions spécifiques à l'oral telles que la gestion de la compréhension (*You know what I mean?* 'Vous voyez ce que je veux dire') et de la discipline de l'auditoire (*Can I get your attention please?* 'Puis-je avoir votre attention'), qui relève d'une prise en compte explicite de la présence de ce dernier lors de la production orale. Ji (2022), quant à elle, a proposé une typologie fonctionnelle des routines, fondée sur le corpus TermITH-Transdisciplinaire (Hatier, Augustyn, Yan, Tran, Tutin & Jacques, 2016), en combinant les modèles d'Hyland (2005) et d'Ädel (2006). Dans cette typologie, le guidage du lecteur a été considéré comme un moyen essentiel d'assurer une lecture efficace et exacte du contenu et de la structure de l'article. Son modèle distingue trois dimensions : (a) structuration textuelle, (b) métalinguistique, et (c) topicalisation. Notre observable étant proche de celui de Ji (2022), nous avons adapté le cadre de sa typologie à la nôtre, tout en cherchant à élargir le concept de guidage du lecteur à celui de prise en compte de l'auditoire, pour mieux rendre compte des dynamiques interactionnelles inhérentes à l'oral scientifique.

- 13 La typologie des formules de prise en compte de l'allocutaire dans le CComSciO a été élaborée en plusieurs étapes. Après la constitution du corpus (Section 3.1), les formules candidates ont été repérées à la fois manuellement et à l'aide de requêtes dans l'outil *AnaText*, en partant des lemmes verbaux servant de pivot aux formules étudiées. Puis, l'annotation fonctionnelle a été faite en trois phases : (a) pré-annotation manuelle de l'ensemble des formules candidates, ce qui a donné lieu à une pré-typologie, (b) trois sessions d'annotation collaborative sur une partie du corpus pour affiner celle-ci, et (c) annotation finale de l'ensemble du corpus sur la base de notre typologie définitive. En parallèle, sur l'ensemble des formules candidates, nous avons identifié cinq propriétés linguistiques : (1) structure syntaxique, (2) lexique et (3) classe sémantique¹² du prédicat, et (4) réalisation lexicale et (5) rôle énonciatif du sujet (locuteur, allocutaire, ou les deux, etc.). Les formules ont ensuite été regroupées manuellement selon leur fonction principale ainsi que leur patron lexico-sémantico-syntaxique, puis sélectionnées selon des critères de fréquence et de dispersion (pour 3 occurrences au total, au moins 1 sur chacune des trois disciplines).¹³

- 14 Dans cette typologie, nous proposons trois dimensions : (a) métalinguistique et méta-énonciative, (b) structuration discursive, et (c) interactionnelle. Chacune inclut plusieurs sous-catégories fonctionnelles ou fonctions spécifiques, pour un total de 23 fonctions avec 49 types de formules différents, comme le montre le tableau suivant. Parmi ces 49 types de formules, si la plupart correspondent à des patrons grammaticaux, deux constituent des formules figées, sans variations morphologiques et lexicales : *on va dire* et *si vous voulez*.

Dimensions	Fonctions		Exemples ¹⁴		N ^{bre} occ.		
Métalinguistique et métaénonciative	Reformulation		ça veut dire que...		26		
	Réparation		ça veut pas dire que...		5		
	Précision terminologique	Dénomination	N° 1	X qu'on a appelé Y	69	79	
			N° 2	on appelle Y	10		
		Emprunt terminologique		je reprends une expression à...		14	
		Choix terminologique		j'utilise...		11	
	Approximation méta-énonciative		N° 1	<u>on va dire</u>	30	37	
			N° 2	... qu'on peut dire...	7		
	Sous-total					172	

Structura- tion discursive	Introduction /Conclusion de la présentation	Annonce du plan	N° 1	dans un premier temps je vais montrer...		52	68	
			N° 2	on va commencer par...		16		
		Annonce de l'objectif/l'objet	N° 1	je vais vous présenter...		34	81	
			N° 2	nous cherchons à proposer...		31		
			N° 3	ce qu'on va vous présenter aujourd'hui		16		
		Clôture/conclusion	je vais conclure				30	
	Renvoi discursif	Renvoi interdiscursif	Renvoi à la littérature	N° 1	je cite	17	31	
				N° 2	je vous renvoie au...	14		
			Renvoi à d'autres communications	N° 1	on en a parlé ce matin	44*	73	
				N° 2	ce que disait XXX en introduction	10*		
				N° 3	ça a été dit tout à l'heure	7		
				N° 4	On en a entendu parler ce matin de...	7		
				N° 5	ce qu'on a vu...	5		
				Renvoi intradiscursif	Rappel	N° 1		comme je l'ai dit
			N° 2			... dont je vous parlais avant	72	
			N° 3			comme on a vu	27	
		N° 4	je rappelle que...			10*		
		N° 5	... qu'on a vu précédemment			9		
			Anticipation	N° 1	vous allez voir	81	172	
				N° 2	je le dirai tout à l'heure	42		
				N° 3	j'y reviendrai	35		
				N° 4	... dont on va vous parler	14		
		Changement du cadre discursif	Transition	je vais aller à...		42		
			Reprise	je vais revenir sur...		10		
			Passage de tour de parole	je passerai la parole à...		10		
		Contextualisa- tion discursive	Limitation des thèmes abordés	N° 1	je vais pas rentrer dans le détail		54	83
	N° 2			je passe là-dessus		29		
	Organisation							

du support de la présentation	N° 1	<i>je vous ai mis un exemple</i>	30	37		
	N° 2	<i>le tableau que je vous ai mis ici</i>	7			
Sous-total				872		
Interactionnelle	Mise en valeur		N° 1	<i>ce qui nous intéresse c'est que...</i>	29	65
			N° 2	<i>je rappelle...</i>	17	
			N° 3	<i>ce qu'on voit c'est...</i>	16	
			N° 4	<i>... sur lequel j'insiste</i>	3	
	Co-constat		N° 1	<i>vous le voyez ici</i>	369	384
			N° 2	<i>... que vous avez sous les yeux</i>	15	
	Appui sur des connaissances partagées		N° 1	<i>vous savez que...</i>	67	92
			N° 2	<i>... que vous connaissez</i>	25	
	Engagement méta-énonciatif de l'allocutaire		N° 1	<i>si vous voulez</i>	40	48
			N° 2	<i>vous allez me dire</i>	8	
	Gestion de la compréhension de l'allocutaire		N° 1	<i>vous voyez (...)</i>	37	52
			N° 2	<i>vous l'avez compris</i>	15	
	Sous-total					641
Total					1685	

Tableau 2. Typologie des formules de prise en compte de l'allocutaire dans l'oral scientifique (Lee, 2024).

- 15 Les formules métalinguistiques et méta-énonciatives ont pour fonction d'explicitier le contenu et l'acte d'énonciation, pour éviter d'éventuelles difficultés liées au lexique ou à la formulation. Le locuteur peut ainsi reformuler ce qui a été dit, le corriger ou clarifier l'emploi des termes en précisant le processus de dénomination, ou en justifiant un emprunt ou un choix de termes. Les formules de structuration discursive visent, quant à elles, à structurer la présentation d'une façon compréhensible, de l'introduction à la conclusion, en assurant la cohésion et la fluidité de l'énonciation à travers les cadres distincts. Les formules de renvoi y sont prédominantes, traduisant une préoccupation d'établir des connexions entre les éléments discursifs avec ceux d'autres discours, ou dans le cadre du discours en question, et de signaler un changement dans le fil énonciatif. Les formules interactionnelles reflètent, quant à elles, la dynamique directe et immédiate des interactions entre le locuteur et l'auditoire. Elles permettent ainsi de mettre en lumière certains éléments essentiels

ou de renforcer le lien phatique, en s'appuyant sur des processus cognitifs comme le constat, la mobilisation des connaissances ou la vérification de la compréhension. Fondée sur un corpus de taille modeste, cette typologie reste ouverte à des ajustements selon les données ou le type de sous-genres scientifiques oraux analysés. Pour plus de détails, voir Lee (2024).

4. Un bref examen des ressources didactiques numériques déjà existantes

- 16 La typologie phraséologique développée pour le français scientifique parlé présente un intérêt à deux niveaux. D'une part, elle permet de mettre en évidence les spécificités linguistiques du genre des communications scientifiques orales, à travers leur comparaison avec celles d'autres genres, qu'ils relèvent du discours scientifique écrit (Siepmann, 2023 ; Lee, 2024) ou d'autres registres oraux (par ex. celui d'entretiens semi-directifs chez Lee (2022)). D'autre part, elle contribue à la conception de ressources pédagogiques en Français sur Objectifs Universitaires (désormais, FOU), visant à aider les étudiants à maîtriser ces formules en vue de la réalisation d'une présentation plus interactive. C'est dans cette deuxième perspective que nous proposons un aperçu de quelques outils numériques, regroupés en trois catégories. Nous commencerons par des ressources de l'EAP, puis nous examinerons trois plateformes dédiées à la consultation de corpus d'écrits scientifiques français. Enfin, nous finirons par examiner trois outils élaborés pour l'apprentissage du FLE à partir des corpus oraux. Le choix des ressources s'est fondé sur leur reconnaissance, afin d'analyser des outils déjà bien identifiés dans le domaine.

4.1. Ressources lexicographiques de l'anglais académique

- 17 Parmi les nombreuses ressources lexicographiques et phraséologiques élaborées en EAP, trois nous semblent particulièrement pertinentes à examiner : le *Louvain English Academic Purposes Dictionary*

(Granger & Paquot, 2010), l'*Academic Formulas List* (Simpson-Vlach & Ellis, 2010), et l'*Academic Phrasebank* (Davis & Morley, 2018)¹⁵. Voici un tableau synthétique de ces trois plateformes :

	<i>Louvain English Academic Purposes Dictionary</i> ¹⁶	<i>Academic Formulas List</i> ¹⁷	<i>Academic phrasebank</i> ¹⁸
Entrées proposées	1) par mots 2) par fonctions 3) par traductions	Listes simples	1) par fonctions 2) par parties textuelles
Modalités couvertes	Écrit	Écrit et oral	Écrit
Particularités pédagogiques	Présentation d'erreurs typiques, conseils pour l'usage, exercices	Pas de visée pédagogique explicite	Explications détaillées sur chaque fonction

Tableau 3. Récapitulatif des trois ressources lexicographiques de l'anglais académique

- 18 Ces travaux ont en commun de proposer, chacun à leur manière, une forme de classification fonctionnelle des expressions, bien que celle de l'*Academic Formulas List* (désormais, AFL) ne soit pas explicitement présentée sur le site. Malgré son caractère un peu sommaire, le site de l'AFL nous permet d'observer les effets de la modalité sur l'utilisation des formules : certaines fonctions (par ex. le rappel) se réalisent différemment à l'écrit (*as shown in* 'comme il a été montré dans') et à l'oral (*come back to* 'revenons à'), et on y retrouve des traits typiques du langage parlé dans les formules orales comme des formes réduites (*wanna*, *we're*) ou la répétition d'interjections (*no no no no*).
- 19 D'un point de vue pédagogique, le *Louvain English Academic Purposes Dictionary* (désormais, LEAD) se distingue des autres, en proposant des commentaires à destination des utilisateurs, comme on le voit dans la figure ci-dessous, sur un emploi erroné, accompagné de sa correction (à gauche), sur une expression typiquement utilisée par les apprenants, jugée moins naturelle par des chercheurs experts (à droite), ou encore sur des conseils pratiques à travers la rubrique *Tip*.

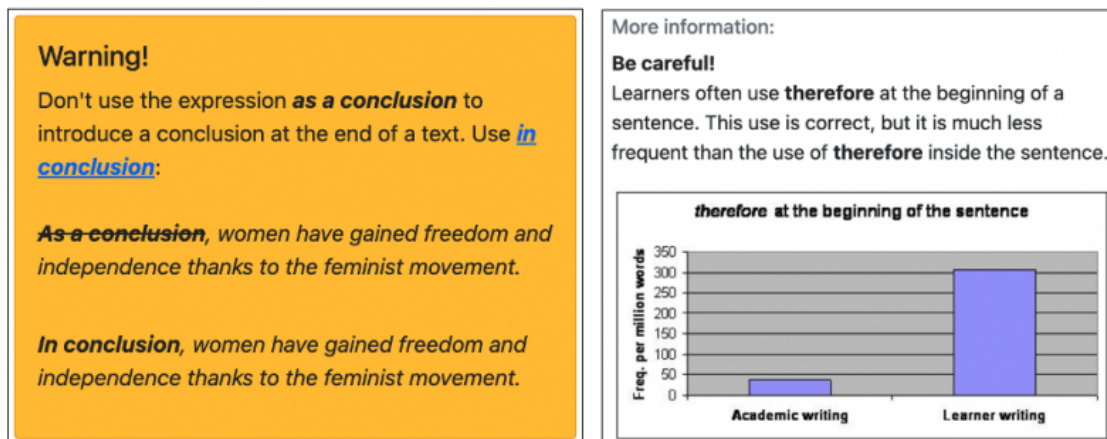


Figure 1. Exemples de commentaires concernant l'usage du lexique sur le LEAD

- 20 Le LEAD présente également des exercices pour certaines fonctions principales : textes à trous, QCM à deux choix lexicaux, sémantiquement proches, ou distinctions entre formules similaires (par ex. *on the contrary* VS *on the other hand*). Ces activités visent à accompagner à la fois l'auto-apprentissage des apprenants et la préparation pédagogique des enseignants pour le cours.
- 21 Contrairement au LEAD et à l'AFL, l'*Academic phrasebank* porte sur des schémas phrastiques typiques, directement réutilisables par les utilisateurs, et présente une classification selon les parties textuelles (introduction, méthodologie, références, conclusion, etc.), où les expressions apparaissent fréquemment. Par exemple, la figure ci-dessous illustre des exemples de phrases visant à signaler une lacune dans la littérature, typiquement utilisées dans l'introduction. Les éléments variables (X ou Y) sont à adapter selon le contexte du rédacteur.

— Identifying the paucity or lack of previous research

There is little published data on ...
No previous study has investigated X.
The use of X has not been investigated.
Data about the efficacy and safety of X are limited.
Up to now, far too little attention has been paid to ...
A search of the literature revealed few studies which ...
The impact of X on Y is understudied, particularly for ...
Few studies have investigated X in any systematic way ...
In addition, no research has been found that surveyed ...
So far, very little attention has been paid to the role of X.
Surprisingly, the effects of X have not been closely examined.
In contrast to X, there is much less information about effects of ...
A systematic understanding of how X contributes to Y is still lacking.
Despite the importance of X, there remains a paucity of evidence on ...
To date, the problem has received scant attention in the research literature.

There is	no little	published work on ... available literature on ... published research on ...
There have been	no few	attempts to investigate ... studies that have assessed the role of ... randomised clinical trials that have investigated ...

Figure 2. Exemple de phrases proposées par l'Academic Phrasebank

4.2. Outils élaborés à partir d'un corpus d'écrits scientifiques en français

22 Concernant la phraséologie des écrits scientifiques en français, trois outils numériques peuvent être particulièrement examinés : l'ARTES (Aide à la rédaction de textes scientifiques) (Pecman & Kübler, 2011), ScienQuest (Falaise, Tutin & Kraif, 2011), et la base LST (Lexique Scientifique Transdisciplinaire) SHS (Li, 2024) (voir le Tableau 4). Ces outils offrent, à l'instar des ressources anglo-saxonnes évoquées, une entrée onomasiologique, bien que cela soit moins développé chez ScienQuest¹⁹.

ARTES ²⁰	ScienQuest ²¹	Base LST SHS ²²
---------------------	--------------------------	----------------------------

Entrées proposées	1) par mots 2) par fonctions 3) par traductions	1) par mots 2) par catégories sémantiques	1) par mots 2) par fonctions 3) par traductions
Particularités pédagogiques	Contextualisation des résultats à l'aide des corpus		
	Navigation complexe, peu d'exemples	Ergonomie simple	Interface intuitive, diverses fonctionnalités

Tableau 4. Récapitulatif des trois outils de corpus d'écrits scientifiques en français

- 23 Les interfaces ARTES et ScienQuest permettent particulièrement d'affiner la recherche : celle de ScienQuest propose à l'utilisateur notamment de choisir une discipline, un sous-genre (HDR, article, thèse, etc.) et une partie textuelle (introduction, développement, conclusion, résumé, etc.), qui l'intéressent en particulier. En ce qui concerne l'ergonomie, la plateforme ARTES est marquée par un processus de recherche relativement complexe, comme l'a déjà mentionné Li (2024, p. 103), et par un nombre d'exemples limité. En revanche, l'interface ScienQuest nous semble plus simple à utiliser grâce à un système de cases à cocher sans menu déroulant.
- 24 Concernant la base LST SHS, élaborée en lien avec le projet ANR TermITH (2012-2016) et les travaux récents sur les routines (Ji, 2022 ; Wang, 2022), il s'agit d'un outil numérique visant à l'auto-apprentissage du lexique et de la phraséologie scientifique. Plusieurs aspects de cette base méritent d'être soulignés sur le plan pédagogique. D'abord, les étiquettes fonctionnelles de l'accès onomasiologique ont été simplifiées à partir de tests réalisés auprès d'étudiants (Li, 2024, pp. 129-135), ce qui les rend plus accessibles. Par exemple, l'étiquette « pour définir la structure du texte » (LST SHS) paraît plus intuitive qu'« organiser son discours » (ARTES), ou encore « renvoyer aux travaux des autres » (LST SHS) plus facile que « renvois métatextuels » (ScienQuest). Le site se distingue aussi par sa clarté : chaque fonction s'accompagne d'une formule prototypique (en italique à droite), ses sous-valeurs sont consultables avec un exemple sans changer de page, et un simple clic sur le symbole de point d'interrogation permet d'afficher une définition plus précise (Figure 3).

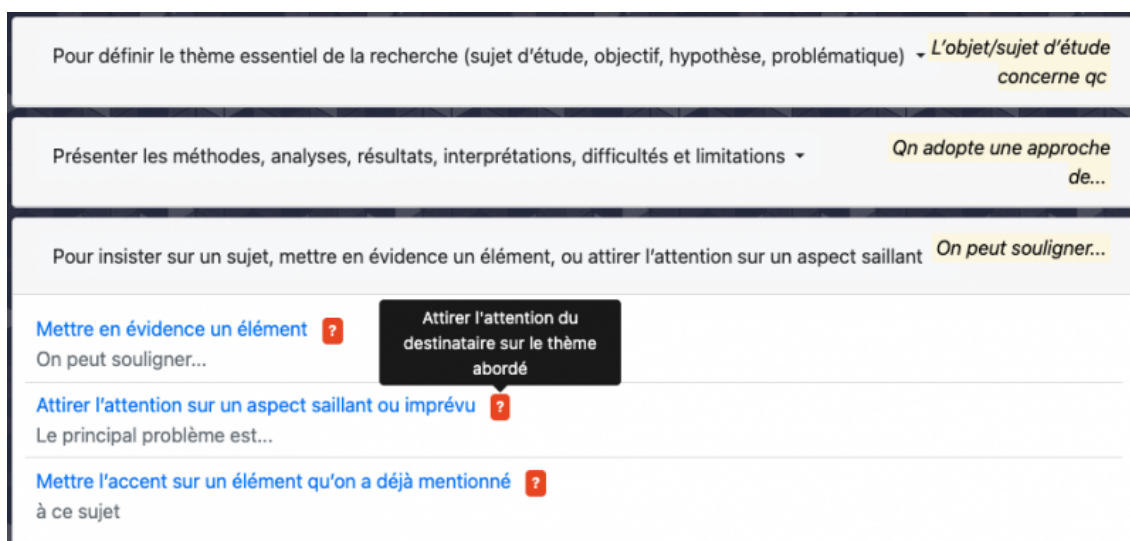


Figure 3. Exemple de la présentation des fonctions dans l'accès onomasiologique sur la base LST SHS

- 25 Parmi les fonctions complémentaires, on note la possibilité de visualiser des articles annotés avec les éléments du LST, offrant ainsi un aperçu global de l'utilité de ces lexiques en contexte. Un module e-learning interactif, encore en développement, proposera également une mise en situation virtuelle, où l'utilisateur acquiert des expressions pour aider d'autres personnages à rédiger leur mémoire.

4.3. Ressources didactiques du français parlé en FLE

- 26 Le dernier ensemble de ressources examinées concerne les interfaces didactiques d'exploration de corpus oraux, déjà connus dans le domaine du FLE : FLORALE (Français langue orale pour le FLE) (Surcouf & Ausoni, 2018), CORAIL (Corpus Oraux Apprenant Interaction Linguistique) (Ravazzolo & Etienne, 2019), et INTERFARE (Interagir plus facilement en réunion) (Ravazzolo, Etienne & André, 2023). Elles se distinguent particulièrement des ressources précédentes (Sections 4.1 et 4.2) par les exigences en termes d'organisation qu'implique la représentation numérique des données orales. Par exemple, chez FLORALE, l'utilisateur a une possibilité d'élargir le contexte des occurrences à l'aide des flèches rouges (Figure 4), d'écouter la bande sonore et d'en ralentir la lecture.

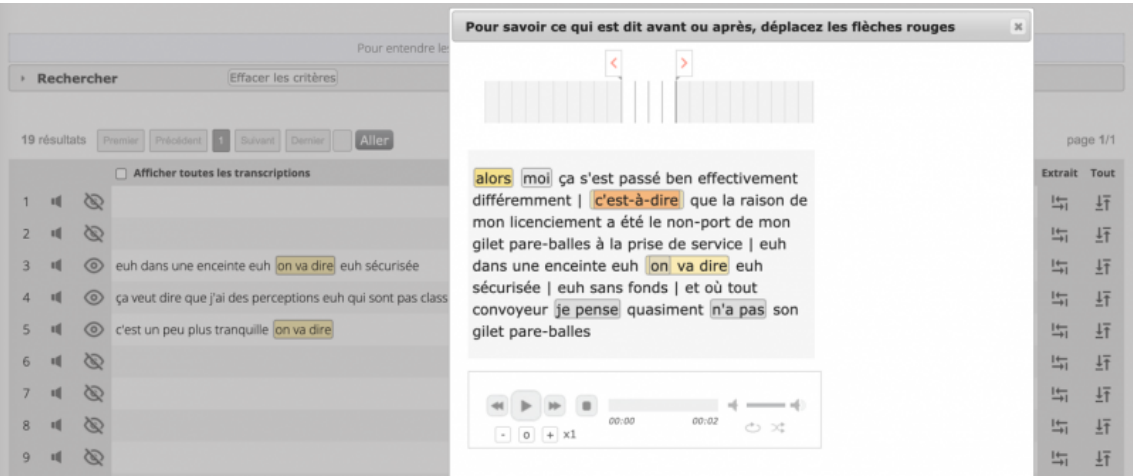


Figure 4. Exemple des résultats de recherche sur la plateforme FLORALE

27 Nous pouvons récapituler les caractéristiques principales de ces trois plateformes, comme suit :

	FLORALE ²³	CORAIL ²⁴	INTERFARE ²⁵
Registres	Émissions de radio	Divers types d'interactions	Interactions en réunion
Entrées proposées	1) par catégories linguistiques 2) par notions linguistiques 3) par fonctions (en partie)	1) par situations 2) par actes de langage 3) par expressions	1) par étapes-clés (rubriques) 2) par expressions
Particularités pédagogiques	Données audio et transcription et usage de métalangage facile		
	Contextualisation des résultats de corpus	Vidéos, explication des séquences d'interaction	Explication détaillée sur chaque étape-clé

Tableau 5. Récapitulatif des trois ressources didactiques du français parlé en FLE

28 L'accès onomasiologique est aussi dominant au sein des outils numériques du français parlé. Par exemple, dans CORAIL, pour l'entrée « Comment dire ? Comment faire ? », les extraits du corpus sont organisés par actes de langage (remercier, refuser, plaisanter, inviter, etc.). Cette approche onomasiologique est également adoptée, bien que partiellement, par FLORALE : les formules méta-discursives, présentées dans le cadre de la catégorie « Outils de conversation », sont classées selon leur fonction (préciser ce qu'on dit, réfléchir à sa manière de dire, nuancer ce qu'on dit, etc.). D'une façon proche, mais distincte, l'interface INTERFARE propose un accès par étapes-clés

que l'on retrouve généralement dans les interactions en réunion : participer activement à la réunion et mettre en œuvre les décisions prises (Figure 5).

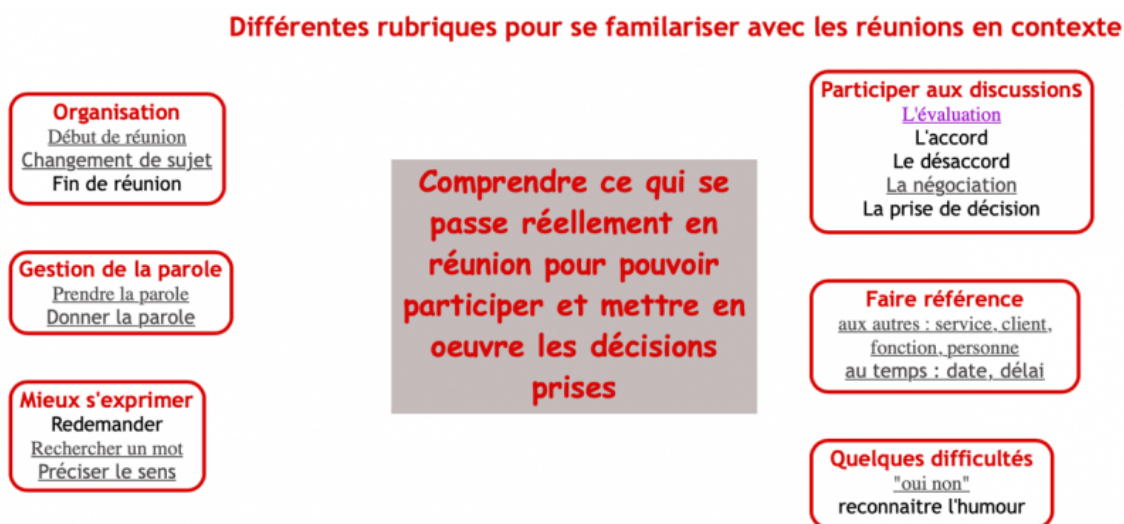


Figure 5. Page d'accueil des rubriques de la plateforme INTERFARE

29 Par ailleurs, il nous semble important de souligner que les trois plateformes du FLE intègrent toutes des éléments phraséologiques, dont certains apparaissent aussi dans notre ressource de phraséologie scientifique orale, comme des marqueurs discursifs *vous savez, vous voyez, on va dire* de FLORALE²⁶.

5. Un exemple de présentation des fonctions et de simplification des étiquettes de notre ressource

30 Parmi les différentes pistes soulevées lors de l'examen des outils numériques sur la langue du discours scientifique ou sur l'oral en FLE²⁷, celle de la formulation des étiquettes fonctionnelles s'avère particulièrement déterminante, puisque la plupart des ressources repose sur une entrée onomasiologique. En effet, la lisibilité des fonctions influence l'accessibilité et l'appropriation de la ressource par les usagers non spécialistes de la linguistique, en particulier les apprenants. La base LST SHS (Li, 2024) a déjà fait, comme on l'a vu plus haut, le choix d'une simplification métalinguistique fondée sur

des tests utilisateurs, afin de rendre les étiquettes plus intuitives. Par conséquent, nous proposons dans cette section de réfléchir à la façon dont les étiquettes de notre typologie des formules de prise en compte de l'allocutaire pourraient être transposées dans la conception d'un futur outil pédagogique du français scientifique oral.

- 31 Pour cela, nous avons d'abord reformulé le concept de la prise en compte de l'allocutaire, à l'instar de l'interface INTERFARE, sous la forme d'une question clé, « Comment rendre sa présentation orale plus interactive ? ». Dans un second temps, nous avons simplifié les intitulés de nos trois dimensions fonctionnelles, qui constitueront les points d'entrée de cette ressource : « expliquer et préciser ce qu'on dit » pour la dimension métalinguistique et méta-énonciative, « guider son auditoire dans l'organisation de la structure » pour la dimension structurelle, et « faire une présentation vivante et dynamique » pour la dimension interactionnelle, comme dans la figure suivante.

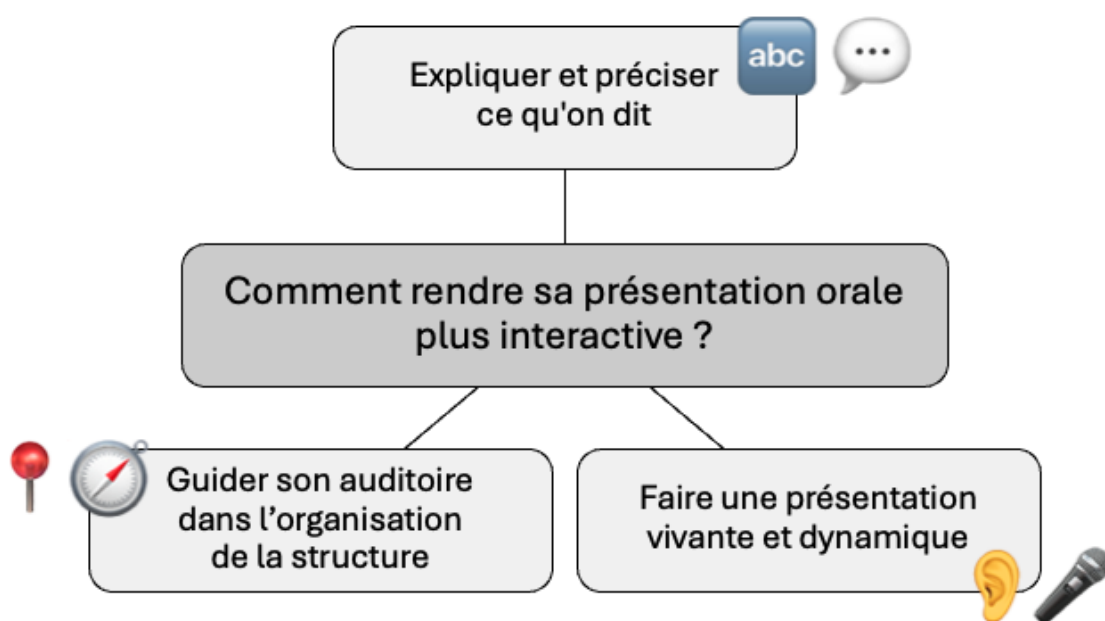


Figure 6. Réalisation possible des trois axes dans une interface d'accueil de notre ressource

- 32 L'usage d'emojis comme appui visuel symbolique nous semble particulièrement utile dans cette interface afin de soutenir l'appropriation intuitive des enjeux de chaque catégorie (Lin & Luo, 2024) : par ex. ceux qui évoquent l'acte de parler sur le langage et de commenter la langue employée, pour la dimension métalinguistique et méta-

énonciative, ceux qui sont liés au repère ou à l'orientation dans la structure, pour la dimension de structuration, enfin, ceux qui renvoient à l'interaction orale immédiate, pour la dimension interactionnelle²⁸.

- 33 Enfin, il est possible de reformuler chaque fonction de notre typologie dans une version moins technique. Pour ce faire, nous nous sommes appuyée sur les étiquettes proposées dans la base LST SHS (Li, 2024), qui intègre la typologie de Ji (2022). Comme on l'a déjà vu (Section 3.2), cette dernière partage plusieurs fonctions spécifiques, notamment celles des dimensions métalinguistiques et structurelles, avec notre typologie.

Catégories	Fonctions	Étiquettes simplifiées
Expliquer et préciser ce qu'on dit	Reformulation	Dire autrement ce qu'on vient de dire
	Réparation	Corriger un malentendu possible
	Dénomination	Appeler un phénomène par une expression
	Emprunt terminologique	Préciser sa reprise d'un terme d'un autre
	Choix terminologique	Préciser son choix d'un terme
	Approximation méta-énonciative	Nuancer l'utilisation d'une expression
Guider son auditoire dans l'organisation de la structure	Annonce du plan	Présenter le plan de sa présentation
	Annonce de l'objectif/l'objet	Dire l'objectif ou l'objet de sa présentation
	Clôture/conclusion	Conclure sa présentation
	Renvoi à la littérature	Renvoyer à des travaux précédents
	Renvoi à d'autres communications	Rappeler une autre présentation orale
	Rappel	Dire qu'on en a déjà parlé
	Anticipation	Dire qu'on en parlera plus tard
	Transition	Passer à un autre élément
	Reprise	Revenir sur un point précédent
	Passage de tour de parole	Donner la parole à son co-présentateur

Guider son auditoire dans l'organisation de la structure	Limitation des thèmes abordés	Préciser ce dont on ne va pas parler
	Organisation du support de la présentation	Expliquer ce qu'on a mis sur le support
Faire une présentation vivante et dynamique	Mise en valeur	Mettre en lumière un élément important
	Co-constat	Inviter les gens à constater quelque chose
	Appui sur des connaissances partagées	Faire appel à ce que le public sait déjà
	Engagement méta-énonciatif de l'allocutaire	Imaginer ce que va dire le public
	Gestion de la compréhension de l'allocutaire	Vérifier que le public suit sa présentation

Tableau 6. Notre typologie fonctionnelle des formules avec des étiquettes simplifiées

- 34 Si une évaluation empirique demeure nécessaire afin d'apprécier la pertinence réelle de cette proposition d'étiquettes, cette étape préliminaire constitue néanmoins un jalon utile en amont d'une possible mise en œuvre didactique de notre ressource phraséologique.

6. En guise de conclusion

- 35 À l'issue de l'examen d'un sous-ensemble d'outils linguistiques numériques à visée didactique, plusieurs pistes ont été identifiées pour guider la conception d'une plateforme dédiée à notre ressource de phraséologie orale dans le contexte de l'enseignement et de l'apprentissage du FOU. Tout d'abord, une approche onomasiologique semble pertinente, voire primordiale, afin de sensibiliser les usagers à la diversité des stratégies de prise en compte de l'allocutaire dans la réalisation d'une communication scientifique orale. Dans cette optique, nous avons proposé une première reformulation simplifiée des étiquettes fonctionnelles de notre typologie en vue de faciliter l'appropriation par des apprenants non spécialistes, peu habitués au métalangage technique.
- 36 Par ailleurs, il serait intéressant de segmenter le CComSciO en différentes parties structurelles de la présentation scientifique orale, à l'instar de l'*Academic phrasebank* et de ScienQuest, afin d'associer certaines fonctions identifiées à chaque partie, dans une logique d'accompagnement progressif de la construction du discours scienti-

fique oral. L'ergonomie de l'interface constitue un autre axe de réflexion essentiel : un outil d'exploration efficace devrait permettre une navigation fluide, en minimisant les efforts de recherche de la part des utilisateurs. En complément de la ressource brute, il serait également souhaitable de proposer des aides à l'interprétation (explications contextuelles, paraphrases) et des supports pédagogiques prêts à l'emploi, facilitant une intégration directe en classe ou en auto-formation des apprenants. Enfin, certaines fonctionnalités propres à la nature orale du corpus devraient être envisagées, telles que la manipulation fine des extraits sonores, la transcription interactive (par ex. sous forme de dictée), ou encore l'accès à des informations prosodiques.

- 37 L'utilisation des formules routinisées se révèle aussi riche dans l'oral scientifique que dans les écrits scientifiques. S'il est désormais établi « un rapport non négligeable entre l'enseignement du lexique (la phraséologie scientifique) et la rédaction d'écrits universitaires » (Cavalla, 2008, p. 149), il semble légitime de postuler un lien similaire entre la maîtrise des formules de prise en compte de l'auditoire et la capacité à produire une présentation orale plus interactive. Dans cette optique, cet article entend contribuer à poser les fondements d'une « phraséodidactique » (González-Rey, 2019) du français parlé scientifique, encore peu explorée dans l'enseignement et l'apprentissage du FOU.

BIBLIOGRAPHIE

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. John Benjamins Publishing.
- Ädel, A. (2010). Just to give you kind of a map of where we are going: A Taxonomy of Metadiscourse in Spoken and Written Academic English. *Nordic Journal of English Studies*, 9(2), 69-97.
- Andersen, H. L. (2007). Marqueurs discursifs propositionnels. *Langue française*, 154, 13-28. <https://doi.org/10.3917/lf.154.0013>.
- Bachy, S., Dister, A., Francard, M., Geron, G., Giroul, V., et al. (2007). Conventions de transcription régissant les corpus de la banque de données VALIBEL. <http://hdl.handle.net/2078.1/165551>
- Baude, O., & Dugua, C. (2016). Les ESLO, du portrait sonore au paysage digital. *Corpus*, 15. <https://doi.org/10.4000/corpus.2924>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and

- written registers. *English for Specific Purposes*, 26, 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Carter-Thomas, S., & Jacques, M.-P. (2017). Interdisciplinary and interlinguistic perspectives on Academic Discourse: the mode variable. *Chimera: Romance Corpora and Linguistic Studies*, 4(1), 1-11.
- Carter-Thomas, S., & Jacques, M.-P. (2022). Mobilisation de routines discursives dans le discours scientifique oral : perspectives croisées. In M. Luodonpää-Manni, F. Grossmann, & A. Tutin (Eds), *Les routines discursives dans le discours scientifique oral et écrit* (pp. 151-177). UGA Éditions.
- Cavalla, C. (2008). Propositions didactiques pour l'enseignement d'éléments phraséologiques en FLE. In O. Galatanu, M. Pierrard, D. Van Raemdonk, M.-E. Damar (Eds), *Enseigner les structures langagières en FLE* (pp. 147-158). Peter Lang.
- Coulmas, F. (1981). *Conversational routines: Explorations in standardized communication situations and prepatterned speech*. Mouton.
- Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, 67(4), 959-997.
- Davis, M., & Morley, J. (2018). Facilitating learning about academic phraseology: Teaching activities for student writers. *Journal of Learning Development in Higher Education*, Special Edition: ALDinHE Conference 2018. <https://doi.org/10.47408/jldhe.v0i0.468>
- Falaise, A., Tutin, A., & Kraif, O. (2012). Une interface pour l'exploitation de corpus arborés par des non-informaticiens : la plate-forme ScienQuest du projet Scientext. *TAL*, 52, 103-128.
- Fløttum, K., Dahl, T., & Kinn, T. (2006). *Academic Voices: Across languages and disciplines*. John Benjamins Publishing.
- González-Rey, M. I. (2019). La Phraséodidactique : État des lieux. *Repères-Dorif*, 18. <https://www.dorif.it/reperes/maria-isabel-gonzalez-rey-la-phraseodidactique-etat-des-lieux/>
- Granger, S., & Paquot, M. (2010). The Louvain EAP Dictionary (LEAD). *Proceedings of the XIV EURALEX International Congress*, 321-326.
- Hatier, S. (2016). *Identification et analyse linguistique du lexique scientifique transdisciplinaire. Approche outillée sur un corpus d'articles de recherche en SHS*. Thèse de doctorat, Université Grenoble Alpes.
- Hatier, S., Augustyn, M., Yan, R., Tran, T. H., Tutin, A., & Jacques, M.-P. (2016). French cross-disciplinary scientific lexicon: Extraction and linguistic analysis. *Proceedings of the XVII EURALEX International congress Lexicography & Linguistic diversity*, 355-365.
- Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. Bloomsbury Publishing.
- Jacques, M.-P. (2017). La structuration textuelle en discours scientifique : Comparaison oral/écrit. *CHIMERA : Romance Corpora and Linguistic Studies*, 4(1), 89-116.
- Jacques, M.-P., & Tutin, A. (2018). *Lexique transversal et formules discursives des sciences humaines*. ISTE Éditions.

- Ji, Y. (2022). *Les routines de guidage du lecteur dans les écrits scientifiques en français*. Thèse de Doctorat, Université Grenoble Alpes.
- Kocourek, R. (1991). *La langue française de la technique et de la science : Vers une linguistique de la langue savante*. Oscar Brandstetter.
- Kraif, O., & Diwersy, S. (2012). Le Lexicoscope : un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques. *Actes de la conférence conjointe JEP-TALN-RECITAL volume 2 : TALN* (pp. 399-406), Grenoble, France.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford University Press.
- Lee, C. (2022). Formules parenthétiques en dire et leur fonctionnement discursif dans les communications scientifiques orales. *Langue française*, 216, 13-28. <https://doi.org/10.3917/lf.216.0013>.
- Lee, C. (2024). *La phraséologie dans le discours scientifique oral en français : formules de prise en compte de l'allocutaire*. Thèse de Doctorat, Université Grenoble Alpes.
- Legallois, D. (2022). Routine, routinisation : à quelles unités et processus a-t-on affaire exactement ? In M. Luodonpää-Manni, F. Grossmann, & A. Tutin (Eds), *Les routines discursives dans le discours scientifique oral et écrit* (pp. 19-35). UGA Éditions.
- Legallois, D., & Tutin, A. (2013). Présentation : Vers une extension du domaine de la phraséologie. *Langages*, 189, 3-25. <https://doi.org/10.3917/lang.189.0003>.
- Li, J. (2024). *Réflexion et mise en œuvre d'une adaptation didactique d'une ressource lexicale pour l'aide à la rédaction scientifique*. Thèse de Doctorat, Université Grenoble Alpes.
- Li, L., Xu, H., & Zhang, X. (2024). Creation and Application of an Academic Spoken Collocation List. *Tesol QUARTERLY*. <https://doi.org/10.1002/tesq.3339>
- Lin, T. S., & Luo, Y. (2024). Visual-textual integration: Emoji as a supplement in health information design. *International Journal of Design*, 18(2), 37-58. <https://www.ijdesign.org/index.php/IJDesign/article/view/5439/1069>
- Loffler-Laurian, A. M. (1983). Typologie des discours scientifiques : Deux approches. *Études de Linguistique Appliquée*, 51, 8-20.
- Longrée, D., & Mellet, S. (2013). Le motif : Une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours. *Langages*, 189, 65-79. <https://doi.org/10.3917/lang.189.0065>.
- Luodonpää-Manni, M., Grossmann, F., & Tutin, A. (2022). *Les routines discursives dans le discours scientifique oral et écrit*. UGA Éditions.
- Pecman, M., & Kübler, N. (2011). ARTES: An online lexical database for research and teaching in specialized translation and communication. In B. Sagot (Ed), *Proceedings of the First International Workshop on Lexical Resources* (pp. 87-93). http://alpage.inria.fr/~sagot/pub/WoLeR_2011_proceedings.pdf
- Ravazzolo, E., & Etienne, C. (2022). Nouvelles ressources pour le FLE à

- partir des études en interaction. *Linx*, 79. <https://doi.org/10.4000/linx.3454>
- Ravazzolo, E., Etienne, C., André, V. (2023). Comment enseigner l'oral en classe ? L'exemple d'un dispositif pour apprendre à interagir plus facilement en réunion de travail. *Repères Dorif*, 28. <https://iris.unitn.it/handle/11572/397900?mode=simple>
- Rinck, F., Fløttum, K., & Poudat, C. (2017). Rôles d'auteur et références à d'autres sources Comparaison entre écrit et oral. *Chimera: Romance Corpora and Linguistic Studies*, 4(1), 117-143. <https://doi.org/10.15366/chimera2017.4.1.005>
- Siepmann, D. (2023). This deserves a brief mention: A multi-corpus comparison of written and spoken academic discourse in English and French with implications for pedagogy and lexicography. *Languages in Contrast*, 23(1), 60-92. <https://doi.org/10.1075/lic.21004.sie>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sitri, F., & Veniard, M. (2017). Routines discursives, variation et normes de genre. *Langage et société*, 159, 99-114. <https://doi.org/10.3917/ls.159.0099>
- Surcouf, C., & Ausoni, A. (2018). Création d'un corpus de français parlé à des fins pédagogiques en FLE : la genèse du projet FLORALE. *Études en didactique des langues*, 31, 71-91. <https://api.unil.ch/iris/server/api/core/bitstreams/f6f3b6ad-face-4910-95fa-229e841be0e6/content>
- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Tran, T. T. H., Masset-Martin, A., & Mangiante, J.-M. (2023). Apprentissage sur corpus et discours oral universitaire. *Corpus*, 24, <https://journals.openedition.org/corpus/8111>
- Tutin, A. (2018). Les expressions polylexicales transdisciplinaires dans les articles de recherche en sciences humaines : Retour d'expérience. In M.-P. Jacques, & A. Tutin (Ed), *Lexique transversal et formules discursives des sciences humaines* (pp. 73-90), ISTE Éditions.
- Tutin, A. (2019). Phrases préfabriquées des interactions : quelques observations sur le corpus CLAPI. *Cahiers de lexicologie*, 114, 63-91.
- Tutin, A., & Kraif, O. (2016). Routines sémantico-rhétoriques dans l'écrit scientifique de sciences humaines : L'apport des arbres lexico-syntaxiques récurrents. *Lidil*, 53, 119-141. <https://doi.org/10.4000/lidil.3966>
- Vlachogianni, P., & Tselios, N. (2021). Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education*, 54(3), 392-440. <https://doi.org/10.1080/15391523.2020.1867938>
- Wang, Y. (2022). *Repérage et annotation des routines sémantico-rhétoriques transdisciplinaires dans les écrits scientifiques*. Mémoire de Master, Université Grenoble Alpes.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Yan, R. (2017). *Étude des constructions verbales scientifiques dans une perspective didactique : Utilisation des corpus pour le diagnostic des besoins langagiers en FLE à l'aide des techniques de TAL*. Thèse de Doctorat, Université Grenoble Alpes.

Yan, R., Tutin, A., & Tran, T. T. H. (2018). Routines verbales pour le français langue étrangère : des corpus d'experts aux corpus d'apprenants. *Lidil*, 58. <http://doi.org/10.4000/lidil.5411>

NOTES

1 Le terme de *routines* renvoie ici à l'usage routinisé d'expressions préconstruites. Il se distingue toutefois de celui de « *conversational routines* » de Coulmas (1981), qui met davantage l'accent sur l'aspect social, pragmatique et interactionnel de la routinisation langagière dans une conversation (*ça va ? je vous en prie, bonne chance !*). Les routines de Tutin et Kraif (2016), quant à elles, relèvent plutôt d'expressions sémantiquement transparentes, jouant des rôles rhétoriques dans un genre codifié comme le genre scientifique. Concernant l'origine du terme de *routines* en linguistique cognitive (Langacker, 1987) ou en analyse conversationnelle, voir Legallois (2022).

2 Il existe aussi d'autres concepts phraséologiques, mobilisés à des fins similaires, tels que les « motifs » (Longrée & Mellet, 2013) ou encore les « routines discursives » (Sitri & Veniard, 2017).

3 Ce terme a déjà été employé par Jacques et Tutin (2018) pour désigner l'ensemble du lexique non terminologique, constitué d'unités lexicales renvoyant à la description des activités de recherche, au raisonnement de l'auteur et à l'écriture scientifique.

4 Il s'agit d'extraire des structures lexico-syntaxiques fréquemment cooccurrentes et statistiquement significatives dans un corpus, et ce, grâce à une méthode itérative fondée sur la cooccurrence syntaxique. Cette méthode permet de modéliser de manière abstraite et moins bruitée les routines, tout en capturant la configuration hiérarchique des mots plutôt que de simples séquences linéaires. Compte tenu de la taille et du caractère oral de notre corpus d'étude (Section 3.1), et en l'absence de pré-traitements morpho-syntaxiques, nous n'avons pas appliqué cette méthode, mais procédé à l'extraction des routines de façon manuelle (Section 3.2).

5 Il s'agit d'un seuil inspiré de Ji (2022), 5 occurrences sur 500 articles (0,01) : dans notre étude, 3 occurrences sur 60 présentations (0,05). Une formule figée ou un patron de formules doit apparaître au moins 1 fois dans

chacune des trois disciplines du corpus (pour la composition de notre corpus, voir Section 3.1).

6 Nous entendons ici le discours scientifique dans le sens du « discours scientifique spécialisé », défini par Loffler-Laurian (1983), dans lequel l'émetteur et le destinataire, en tant que chercheurs disposant d'expertise sur le sujet, appartiennent à la même communauté disciplinaire.

7 Cf. le numéro 4(1) de la revue *CHIMERA : Romance Corpora and Linguistic Studies* auquel a donné lieu le projet EIIDA : <https://revistas.uam.es/chimera/issue/view/679>

8 Le contexte du recueil des données a été particulièrement contraint en raison de la crise sanitaire liée à Covid-19 en 2020.

9 Cf. Lee (2024, pp. 350-352) pour les métadonnées du corpus (sources de données, nombre de locuteurs, sexe).

10 Ce travail a bénéficié du financement accordé par le consortium CORLI (CORpus, Langues et Interactions).

11 Site : http://phraseotext.univ-grenoble-alpes.fr/lexicoscope_2.0/

12 Pour cela, nous nous sommes appuyée sur la catégorisation sémantique des verbes transdisciplinaires de la base LST SHS (Hatier et al., 2016), tout en créant de nouvelles classes, comme « rappel », « passage de tour de parole », « connaissance » ou « compréhension », absentes de cette base.

13 Les formules présentes dans seulement deux disciplines ont été retenues lorsqu'elles ont plus de 10 occurrences. Cela concerne 3 cas sur les 49 types de formules (marqués par un astérisque à côté du nombre d'occurrences dans le Tableau 2). Bien que le seuil minimal ait été fixé à 3 occurrences, il convient de noter que 48 types de formules en comptent plus de 10.

14 Il s'agit des exemples les plus prototypiques de chaque type de formule.

15 Pour plus de détails sur ces ressources, voir les synthèses de Ji (2022), Li (2024) ou Lee (2024).

16 Site : <https://leaddico.uclouvain.be/>

17 Site : <https://www.eapfoundation.com/vocab/academic/afl/>

18 Site : <https://www.phrasebank.manchester.ac.uk/>

19 Il s'agit des catégories liées aux enjeux d'un discours scientifique tels que « formulation d'une hypothèse » et « validation d'une hypothèse » (catégorie Autour des hypothèses), ou « verbes de choix et d'intention » et « verbes de résultats et apports scientifiques » (Propositions propres de l'auteur). On

peut toutefois remarquer une certaine hétérogénéité dans les dimensions des catégories proposées.

20 Site : <https://artes.app.univ-paris-diderot.fr/>

21 Site : <https://corpora.aiakide.net/scientext20/>

22 Site : <https://lst.demarre-shs.fr/>

23 Site : <https://florale.unil.ch/>

24 Site : <http://clapi.icar.cnrs.fr/Corail/>

25 Site : <https://icar.cnrs.fr/interfare>

26 La plateforme CORAIL regroupe, dans l'entrée « Les expressions du quotidien », des formules interactionnelles (*ça marche, laisse tomber*) et méta-discursives (*tu sais, tu vois*), et le site INTERFARE, dans la rubrique « Quelques expressions », les idiomes (*faire le point, mettre au courant*), les marqueurs méta-discursifs (*pour te dire, dis-moi, je dirais*) ou encore les routines (*faut/faudrait* + infinitif, *l'idéal/l'idée c'est*).

27 Il serait toutefois pertinent de donner la parole aux utilisateurs finaux et d'évaluer leur satisfaction ainsi que leur ergonomie à l'aide d'une échelle standardisée, telle que la *System Usability Scale* (Vlachogianni & Tselios, 2021), proposée par un des relecteurs de cet article.

28 Il convient de garder à l'esprit que l'interprétation des emojis peut varier d'une culture à l'autre des apprenants.

RÉSUMÉS

Français

Cet article présente une ressource d'expressions phraséologiques fréquemment observées dans le discours scientifique oral en français, expressions qui rendent l'énonciation du locuteur plus interactive par la prise en compte de l'auditoire. Ces *formules de prise en compte de l'allocutaire* facilitent la compréhension du contenu et de la structure de la présentation pour l'auditoire, tout en l'impliquant dans la construction discursive. En s'appuyant sur des travaux antérieurs en phraséologie et en langue des genres scientifiques, la catégorie de « formules discursives » inclut à la fois les suites de mots figées et les routines fondées sur un patron grammatical plus souple. Un corpus de 60 communications scientifiques orales en sciences humaines et sociales a été constitué, à partir duquel les formules ont été identifiées et classifiées selon une typologie en trois dimensions fonctionnelles : (a) méta-linguistique et méta-énonciative, (b) structurelle, et (c) interactionnelle. En complément, nous proposerons un rapide tour d'horizon des outils numé-

riques didactiques consacrés au discours scientifique ou au français parlé, pour proposer quelques pistes d'application - notamment la simplification des étiquettes fonctionnelles - en vue d'un futur usage pédagogique de notre ressource.

English

This article presents a resource of phraseological expressions frequently observed in spoken French academic discourse, which make the speaker's enunciation more interactive by taking into account the presence of the audience. These *formulas for considering the addressee* help the audience better understand the content and structure of the discourse, while involving them in the construction of the speaker's presentation. Building on previous work in phraseology and the language of academic genres, the category of "discursive formulas" includes both fixed word sequences and routines based on a more flexible grammatical pattern. A corpus of 60 oral academic presentations in the humanities and social sciences was compiled, from which the formulas were identified and classified according to a three-dimensional functional typology: (a) metalinguistic and meta-enunciative, (b) structural, (c) interactional. In addition, we will provide a brief overview of digital didactic platforms for academic discourse or spoken French, and suggest some possible applications - notably the simplification of functional labels - for the future pedagogical use of our resource.

INDEX

Mots-clés

phraséologie, discours scientifique oral, formules discursives, prise en compte de l'allocutaire, Français sur Objectifs Universitaires

Keywords

phraseology, spoken academic discourse, discursive formulas, considering the addressee, French for academic purposes

AUTEUR

Chaeyoung Lee

LiDiLEM, Université Grenoble Alpes
chaeyoung.lee@univ-grenoble-alpes.fr