วารสารสมาคมครูภาษาฝรั่งเศสแห่งประเทศไทย
ในพระราชูปถัมภ์สมเด็จพระกนิษฐาธิราชเจ้า
กรมสมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี
Journal homepage: https://www.tci-thaijo.org/index.php/bulletinatpf/index

Oralité et pédagogie énactive : un antidote face aux dangers de l'intelligence artificielle

Jean-Philippe BABU *

Université Naresuan, Thaïlande

Orality and Enactive Pedagogy: An Antidote to the Dangers of Artificial Intelligence

Jean-Philippe Babu *

Naresuan University, Thailand

Article Info

Academic Article

Article History:

Received 2 February 2026

Revised 22 April 2026

Accepted 7 May 2026

Mots clés :

Oralité, Pédagogie énactive, IA, Prosodie

Keywords:

Orality, Enactive pedagogy, Artificial intelligence, Prosody

Résumé

Face aux transformations induites par l'intelligence artificielle dans l'enseignement des langues, cet article défend une approche énactive de l'oralité en FLE comme antidote aux risques de désincarnation et de prolétarisation cognitive. S'appuyant sur l'énaction de Varela et le modèle PAS de Masciotra (se disposer, se situer, se positionner, (se) réaliser), nous proposons de recentrer la didactique sur une parole incarnée et située.

L'article expose d'abord les fondements de la pédagogie énactive où la compétence émerge de l'interaction dynamique entre l'apprenant et son environnement. Il développe ensuite une méthodologie de l'oralité articulant travail vocal, prosodie structurante et expressive, gestes co-verbaux, partitions orales multiniveaux et exercices rythmiques. Les outils d'IA (transcription automatique, visualisation prosodique) sont repositionnés comme supports d'observation plutôt que substituts de l'apprentissage.

Dans une perspective pharmacologique inspirée de Stiegler, nous analysons l'ambivalence de l'IA : poison lorsqu'elle court-circuite l'appropriation subjective, remède si elle soutient la réflexivité. Cette réflexion, ancrée dans la pratique de classe, vise à former des sujets parlants autonomes et non de simples utilisateurs de technologies langagières.

* Corresponding author

E-mail address: jphbabu@gmail.com ; jeanph@nu.ac.th

Abstract

In response to the transformations brought about by artificial intelligence in language teaching, this article advocates an enactive approach to orality in French as a Foreign Language as an antidote to the risks of disembodiment and cognitive proletarianization. Drawing on Varela's theory of enaction and Masciotra's PAS model — se disposer, se situer, se positionner, (se) réaliser — this article proposes to refocus language pedagogy on embodied and situated speech.

The article first outlines the foundations of enactive pedagogy, in which competence emerges from the dynamic interaction between the learner and their environment. It then develops a methodology of orality combining vocal work, structuring and expressive prosody, co-verbal gestures, multi-level oral scores, and rhythmic exercises. Artificial intelligence tools, such as automatic transcription and prosodic visualization, are repositioned as supports for observation rather than substitutes for learning.

From a pharmacological perspective inspired by Stiegler, the article analyzes the ambivalence of artificial intelligence: it acts as a poison when it short-circuits subjective appropriation, but as a remedy when it supports reflexivity. Grounded in classroom practice, this reflection seeks to foster autonomous speaking subjects rather than mere users of language technologies.

1. Introduction

À l'heure où l'intelligence artificielle reconfigure les modes d'apprentissage et de production langagière, une question centrale s'impose : comment préserver l'autonomie des apprenants dans un contexte éducatif en pleine mutation ? Depuis 2022, avec la diffusion massive de ChatGPT, l'intelligence artificielle s'est installée dans les usages quotidiens, bouleversant en profondeur les pratiques d'enseignement-apprentissage. Elle interfère désormais directement avec les processus d'apprentissage eux-mêmes, à l'écrit comme à l'oral, en transformant la manière dont les apprenants s'approprient les savoirs, sollicitent leur mémoire, structurent leur pensée et formulent leurs réponses. Ces outils ne se limitent plus à des fonctions d'assistance ponctuelle :

les IA génératives peuvent désormais produire des textes complets à la demande, dialoguer en temps réel et, dans le domaine spécifique de l'apprentissage des langues, fournir un accompagnement continu sur la prononciation et l'intonation (Duolingo, ELSA Speak).

Or, comme toute technologie, l'IA se prête à deux usages opposés : un usage individuant, au sens de Bernard Stiegler, qui renforce l'attention, la réflexivité et l'appropriation subjective du savoir, et un usage désindividuant, qui remplace l'effort d'apprentissage par des réponses toutes faites, réduisant l'engagement personnel de l'apprenant. Les travaux récents montrent des effets ambivalents : l'IA peut améliorer l'efficacité perçue et certaines performances d'apprentissage (méta-analyse sur 51 études) (Wang & Fan, 2025), mais la surdépendance s'associe à une fragilisation de la pensée critique (Gerlich, 2025), effet aggravé lorsque la validation des sources et des critères fait défaut (Zhai, Wibowo et Li, 2024). Ces dérives rappellent que la technique est, fondamentalement, un *pharmakon* — au sens de Stiegler, à la fois remède et poison — et qu'elle appelle une pharmacologie éducative. Nous appelons "pharmacologie de l'oralité" l'antidote face aux usages de substitution de l'IA : ralentir volontairement le rythme de l'apprentissage, donner aux apprenants le temps de se disposer et de s'ajuster corporellement et vocalement (respiration, posture, regard, intention) plutôt que de se précipiter vers une réponse immédiate. Dès lors, nous recentrons l'attention sur l'oralité : l'oralité rééquilibre vers l'humain en mobilisant corps, voix et espace, l'écran n'y jouant qu'un rôle auxiliaire. En pratique, nous mettons en œuvre un agir vocal et corporel centré sur la prosodie — rythme, mélodie, pauses, souffle, timbre — en nous appuyant sur Michel Billières (2002), qui montre que la parole engage le corps entier, ainsi que sur Lhote et Llorca (2001), qui confirment le lien geste–expressivité.

Pour donner à ce recentrage sa cohérence théorique, nous l'inscrivons dans la théorie de l'énaction, élaborée par Francisco Varela : la cognition y est un processus incarné et dynamique où perception, action et pensée se co-déterminent. Cette perspective renverse le paradigme computationnel qui domine aujourd'hui l'intelligence artificielle : penser n'est pas calculer, mais faire émerger du sens dans l'action. Sur le plan pédagogique, Masciotra et al. (2008) en proposent une traduction à l'aide du modèle PAS (Personne en action et en situation), qui conçoit la compétence comme un

agir situé, co-régulé par quatre fonctions interdépendantes. Dès lors, l'évaluation ne porte plus sur des « compétences » hors sol, mais sur cet agir en contexte, où les critères classiques (précision morphosyntaxique, contrôle phonologique) sont recontextualisés dans une dynamique d'ajustement visant la transformation de la situation.

Cette perspective rejoint l'orientation actionnelle et l'approche par compétences lorsqu'elles valorisent la performance en contexte, tout en s'en distinguant par l'attention portée à la co-régulation et à la qualité des choix en temps réel — dimension que des listes de descripteurs saisissent mal.

Dans cet article, nous suivrons trois étapes. D'abord, nous présenterons l'approche énactive, de Varela à Masciotra, en montrant comment elle permet de penser la compétence comme agir incarné et situé, structuré par quatre fonctions : se disposer, se situer, se positionner, (se) réaliser. Nous montrerons également en quoi cette épistémologie se distingue des modèles d'évaluation dominants, trop souvent centrés sur des descripteurs décontextualisés. Ensuite, nous explorerons la place de l'oralité dans l'enseignement, en montrant comment le travail sur la prosodie, la gestuelle et le temps réel soutient un apprentissage incarné, à rebours des logiques de standardisation et d'automatisation portées par l'IA. Enfin, nous conclurons par une réflexion philosophique, dans le sillage de Stiegler et de la phénoménologie, sur la nécessité d'adosser la technique à une pratique réflexive de la parole, afin de maintenir le langage comme expérience temporelle et humaine, au-delà des logiques de substitution technologique.

2. Fondements : de l'énaction à la pédagogie (modèle PAS)

2.1 L'énaction : une rupture fondatrice, d'une brûlante actualité

Lorsqu'elle apparaît à la fin des années 1980 sous l'impulsion du neurobiologiste chilien Francisco Varela, l'approche énactive marque une rupture profonde avec les paradigmes dominants des sciences cognitives. Contre le modèle computationnel hérité de la cybernétique — qui conçoit le cerveau comme un processeur d'informations et la pensée comme manipulation de symboles — Varela propose une

conception radicalement différente de la cognition : non comme représentation passive d'un monde extérieur, mais comme action incarnée et création située. Dès 1988, dans *Cognitive Science: A Cartography of Current Ideas*, puis dans *The Embodied Mind* (1991 ; trad. fr. *L'inscription corporelle de l'esprit*, 1993), il affirme que la cognition est « l'avènement conjoint d'un monde et d'un esprit à travers l'histoire des actions d'un être vivant dans ce monde » (Varela, Thompson, & Rosch, 1993, p. 35). Reprise et développée notamment par Evan Thompson dans *Mind in Life* (2007), cette idée prend une portée nouvelle à l'heure où l'intelligence artificielle et les discours post-humanistes annoncent l'avènement d'un humain augmenté, optimisable par la machine.

À l'ère de l'intelligence artificielle, il devient plus pressant que jamais de distinguer performance symbolique et intelligence vivante. Comme le souligne Varela, « en posant des règles pour exprimer l'activité mentale et des symboles pour exprimer les représentations, on s'isole justement du pivot sur lequel repose la cognition dans sa dimension vraiment vivante » (Varela, 1996, p. 98). Selon lui, ce type de modélisation n'est possible que dans un cadre restreint, où presque tout est statique (*ceteris paribus*, autrement dit en supposant que les autres facteurs restent constants). Or, poursuit-il, « le contexte et le sens commun ne sont pas des artefacts résiduels pouvant être progressivement éliminés grâce à des règles plus sophistiquées. Ils sont en fait l'essence même de la cognition créatrice. » Autrement dit, réduire l'intelligence à un traitement formel revient à méconnaître son ancrage corporel, historique et évolutif.

C'est également ce que rappelle le biologiste Bernard Claverie (2019). Selon lui, les systèmes artificiels, aussi puissants soient-ils, demeurent désincarnés : leur architecture est figée, non évolutive, dépend d'un programme symbolique et ne participe à aucun véritable couplage avec un monde. Ils restent ainsi étrangers à cette dimension poïétique de la cognition, où le sens se construit dans l'interaction avec un environnement mouvant.

En définitive, la cognition n'est pas une opération sur des symboles, mais une dynamique de sens indissociable de l'histoire vécue du sujet. Elle ne se déploie pas à partir de représentations préalables, mais se trace dans l'expérience même, comme le montre Varela : « les facultés cognitives sont inextricablement liées à l'historique de ce qui est vécu, de la même manière qu'un sentier au préalable inexistant apparaît en

marchant » (Varela, 1989, p. 111). Penser ne consiste donc pas à résoudre un problème par calcul, mais à faire émerger du sens dans l'action, en assurant la continuité du vécu. Sur ce socle, la question didactique devient : comment traduire cette intelligibilité de l'action en une pédagogie de la compétence attentive aux conditions de l'activité réelle en contexte ?

2.2 La pédagogie énactive : le modèle PAS

2.2.1 L'agir au cœur de l'apprentissage : la vision de Masciotra

Dans la continuité de cette compréhension incarnée et située de la cognition, Domenico Masciotra s'impose comme l'un des représentants majeurs de la pédagogie énactive, notamment avec *Énaction. Apprendre et enseigner en situation* (Masciotra et al., 2008). Ses travaux, mobilisés par plusieurs auteurs en didactique des langues (notamment en FLE, voir par ex. Aden, 2017 ; Pelissero, 2019), soulignent que la compétence ne se réduit ni à un savoir ni à une procédure : elle émerge de l'interaction constante entre la personne et la situation.

Dans cette perspective, Masciotra (2017) propose le modèle PAS (Personne en Action et en Situation), qui conçoit la compétence comme dynamique de l'agir en contexte. Cette dynamique se déplie selon quatre fonctions interdépendantes — se disposer, se situer, se positionner, (se) réaliser — entendues non comme étapes séquentielles mais comme régulations conjointes par lesquelles la PAS s'organise, s'ajuste et s'accomplit dans l'action.

Ce modèle rompt avec une conception additive de la compétence (somme de savoirs + savoir-faire + savoir-être) pour lui substituer une vision systémique : la compétence n'est ni dans la personne ni dans la situation, mais dans leur couplage dynamique. Elle ne préexiste pas à l'action : elle s'actualise dans et par l'action, en réponse aux exigences émergentes de la situation.

Cette approche a des implications majeures pour l'enseignement des langues. Elle invite à dépasser une logique d'accumulation (apprendre d'abord le vocabulaire, puis la grammaire, puis parler) pour privilégier une logique d'immersion située : l'apprenant se forme en agissant dans des situations authentiques, où les quatre fonctions se co-régulent en temps réel. Apprendre une langue, ce n'est pas accumuler

des ressources pour agir plus tard ; c'est développer progressivement sa capacité à se coupler avec des situations langagières de plus en plus complexes, en mobilisant conjointement corps, attention, diagnostic situationnel et stratégies d'ajustement.

2.2.2 Les quatre fonctions de l'agir

Se disposer

« Se disposer » désigne l'entrée en présence et en énergie de l'apprenant : maîtrise de soi, vigilance attentionnelle, mobilisation affective et disponibilité corporelle. Masciotra rattache cette fonction à l'intelligence dispositionnelle, en l'articulant aux trois autres fonctions « se situer, se positionner, (se) réaliser ». Au cœur de ce travail, la disponibilité fait figure de disposition maîtresse : elle est « la clé pour se disposer mentalement de façon à affronter les situations » ; être disponible, c'est « être présent, et plus profondément, être une Présence », en canalisant ses ressources vers la tâche. En son principe, « se disposer » constitue une mise au point préalable : sans réglage initial du corps et de l'attention (posture, souffle, focalisation), l'engagement et la performance qui suivent demeurent aléatoires.

En classe de langue étrangère, la prise de parole en public constitue une difficulté spécifique, parfois renforcée par des normes culturelles d'exposition de soi ; dans ce cadre, le travail de disponibilité aide à réduire l'anxiété et à abaisser le seuil d'entrée en interaction par de courts rituels (respiration lente, bref silence, posture d'ancrage). Cet appui est particulièrement pertinent en Thaïlande, où, pour des raisons culturelles, parler en public génère du stress et de l'anxiété chez les apprenants (Thaksanan, 2024 ; Suratin & Sribayak, 2025). Par ailleurs, à l'ère de l'hyperconnexion, marquée par la surcharge attentionnelle, « se disposer » prend une acuité nouvelle : recréer des temps brefs de centration et d'écoute de soi constitue une réponse pédagogique simple aux dispersions contemporaines.

Se situer

« Se situer » constitue la fonction par laquelle la personne appréhende et comprend activement la dynamique situationnelle en cours. Se situer engage une intelligence situationnelle qui opère selon une dialectique fondamentale : la personne est à la fois situante — elle génère en partie la situation par sa présence active — et située — elle est affectée par les circonstances. Dans la perspective de l’énaction,

Masciotra rappelle que le situant et le situé forment un « couplage structurel : l'un n'existe pas sans l'autre » (Masciotra, 2017, p. 32). La situation émerge de l'interaction entre modalités internes (ressources cognitives, conatives, corporelles) et modalités externes (circonstances, ressources matérielles et humaines).

Fondamentalement, une personne se situe selon les ressources dont elle dispose (Masciotra & Medzo, 2009). Se situer implique une prise de conscience tacite de l'inventaire de ses moyens — ce sentiment incarné de pouvoir ou d'impuissance face aux exigences perçues. Face à une piste de ski abrupte, le débutant perçoit une « situation impossible », le skieur expérimenté un « défi », le professionnel un simple « terrain d'entraînement ». Cette intelligence situationnelle établit le diagnostic ressources/situation, mais ne décide pas encore de leur mobilisation : c'est l'intelligence positionnelle (fonction suivante) qui opérera ces choix stratégiques.

Cette fonction répond à un questionnement continu : Où suis-je ? Qu'est-ce qui se passe ? De quoi je dispose ? Qu'est-ce qui m'est accessible ? Se situer n'est pas une opération unique, mais un processus itératif de re-situation perpétuelle qui accompagne l'agir.

Dans le domaine des langues, « se situer » implique de reconnaître en temps réel ses moyens linguistiques disponibles : est-ce que je comprends globalement ? ai-je accès au vocabulaire nécessaire ? quelles stratégies compensatoires est-ce que je connais ? Cette conscience située des ressources — distincte de leur mobilisation effective — permet d'évaluer l'écart entre exigences communicatives et répertoire activable. L'apprenant qui se situe perçoit ce qui est présent ou absent, établit un état des lieux dynamique avant de décider (fonction positionnelle) comment les articuler stratégiquement.

Pédagogiquement, favoriser cette intelligence situationnelle suppose des dispositifs qui rendent visible le diagnostic ressources : observation participante avec verbalisation (« De quoi disposiez-vous dans cette situation ? »), rituels métacognitifs (« Faites l'inventaire de vos moyens avant d'agir »), et variation situationnelle pour apprendre à reconnaître l'évolution de ses ressources. Apprendre, c'est affiner progressivement cette intelligence diagnostique qui précède et informe l'action stratégique.

Se positionner

« Se positionner » relève de l'intelligence positionnelle, c'est-à-dire de la maîtrise des situations. À ce titre, elle correspond à la notion de compétence : une intelligence des situations dont la finalité est le fonctionnement compétent, et qui opère pré-activement (préparer les conditions de l'action) et pro-activement (anticiper, orienter, corriger en cours d'action).

Concrètement, se positionner ne se réduit pas à une posture : c'est générer et superviser un réseau de possibilités d'action, se donner un but, anticiper, évaluer, réorienter — bref, la dimension profonde de la compétence, par contraste avec la gestualité plus « en surface ».

Dans l'économie des quatre formes d'intelligence qui co-régulent l'agir — dispositionnelle (maîtrise de soi, « se disposer »), situationnelle (lecture et intelligence de la situation, « se situer »), positionnelle (compétence : maîtrise des situations, « se positionner ») et gestuelle (dextérité/performance, « (se) réaliser ») — se positionner occupe une place médiatrice : c'est à partir d'un bon état dispositionnel et d'une appréciation situationnelle juste que l'on positionne des possibles, lesquels s'actualisent gestuellement en performance.

Transposée à l'oral en langue étrangère, la fonction « se positionner » éclaire la compétence de communication comme maîtrise des situations : mobiliser prioritairement la compétence linguistique disponible et, le cas échéant, recourir à des voies compensatoires (gestes, prosodie, regards, supports contextuels) pour atteindre l'intention visée. L'essentiel est le fonctionnement compétent : préparer/ajuster un réseau de possibles (pré-/pro-activité) puis en assurer la réalisation — par les mots ou par les gestes selon les exigences de la situation — en solidarité avec se disposer, se situer et (se) réaliser.

(Se) réaliser

« (Se) Réaliser » est la quatrième fonction de l'agir. Elle complète « se disposer », « se situer » et « se positionner » en rendant compte de ce que la personne fait effectivement, ici et maintenant, pour faire progresser l'activité. Masciotra précise que « réaliser » est synonyme d'exécuter, d'effectuer, d'accomplir, et qu'on peut, selon les domaines, parler aussi d'interpréter ou de performer (arts, sports) : l'essentiel est

d'expliciter les actes ponctuels (mouvements, gestes, pauses, micro-actions) par lesquels l'activité se déploie.

L'exemple du musicien est éclairant : chaque contact avec la corde produit une ou plusieurs notes, puis une phrase, puis une pièce entière. « Réaliser » renvoie alors au fini, au raffiné, à la précision, bref à la performance de la gestuelle par laquelle l'œuvre advient.

Sur le plan fonctionnel, « (se) réaliser » est l'actualisation de l'agir : ce qui a été préparé (se disposer), lu (se situer) et supervisé/orienté (se positionner) prend corps dans des gestes et énoncés précis, sous les aléas du réel. Les quatre fonctions co-régulent ainsi l'agir compétent de la PAS (Personne en Action et en Situation).

Transposée à l'oral en langue étrangère, « (se) réaliser » renvoie à l'effectuation précise : articulation, accentuation, contours prosodiques, enchaînements, gestes co-verbaux, autant de micro-actions par lesquelles l'intention communicative devient performance observable — en continuité directe avec les trois autres fonctions.

2.2.3 Implications épistémologiques pour l'évaluation

Le modèle PAS appelle une refonte de nos dispositifs d'évaluation. Si la compétence est un agir situé et co-régulé, elle ne peut être réduite à une liste de descripteurs isolés, mesurés hors contexte. C'est précisément ce que souligne Glenn Fulcher (1996), qui critique les échelles intuitives, peu ancrées dans des performances réelles, et prône des échelles empiriques construites à partir de données observées. Côté francophone, Patrick Chardenet (1999) conçoit également l'évaluation comme une activité située plutôt que comme une addition d'items décontextualisés.

Cette critique ne vise pas à écarter les critères classiques — précision morphosyntaxique et lexicale, contrôle phonologique, fluidité — mais à les recontextualiser dans une dynamique d'ajustement. L'évaluateur ne se demande plus seulement « L'apprenant maîtrise-t-il tel point de grammaire ? », mais « Comment mobilise-t-il ses ressources pour transformer la situation ? Comment se dispose-t-il face à l'imprévu ? Comment ajuste-t-il sa stratégie en temps réel ? »

Cette perspective rejoint l'orientation actionnelle du CECRL (Conseil de l'Europe, 2021) et l'approche par compétences lorsqu'elles valorisent la performance en contexte. Elle s'en distingue toutefois par l'attention portée à la co-régulation et à la

qualité des choix en temps réel — dimensions que des listes de descripteurs génériques saisissent mal. Le modèle PAS invite à observer l'agir dans sa globalité systémique : non pas une somme de micro-compétences juxtaposées, mais un couplage dynamique entre personne et situation, où les quatre fonctions s'articulent en permanence.

Notre approche de l'oralité prolonge la notion de médiation mise à l'honneur dans le *Volume complémentaire du CECRL* (Conseil de l'Europe, 2021), entendu ici comme une co-construction du sens plutôt que comme une simple transmission. Nous savons tous qu'en classe, la médiation est décisive : il s'agit de faire circuler la compréhension par des reformulations simples, des relances ciblées (où/quand/pourquoi), la clarification d'un terme, la levée d'un malentendu, l'apaisement d'un désaccord, la répartition des tours de parole et, au besoin, par une courte synthèse pour valider l'accord du groupe. Or, si le Cadre fournit des descripteurs et des repères pour les productions finales, la perspective énactive documente comment les processus corporels (souffle, gestes, prosodie) rendent possible la compétence en structurant l'énonciation de l'intérieur. Cette distinction devient déterminante face à l'IA, capable d'imiter des performances décrites par le CECRL sans engager le substrat corporel où s'enracine la communication authentique.

Concrètement, cela implique :

1. Évaluer en situation authentique : privilégier les tâches ouvertes, non scriptées, où l'apprenant s'ajuste à des imprévus (simulation, jeu de rôle improvisé, interaction réelle).
2. Observer les régulations et actes de médiation : prêter attention au-delà du seul résultat linguistique (correction, fluidité) aux ajustements en cours d'action — reformulations, levée d'un malentendu, relances ciblées, répartition des tours de parole, courtes synthèses.
3. Valoriser l'engagement corporel : reconnaître que la compétence orale ne se limite pas à la production verbale, mais inclut posture, regard, prosodie, gestes co-verbaux — autant de ressources mobilisées dans le couplage avec la situation.
4. Documenter l'évolution : suivre la progression non comme accumulation linéaire de savoirs, mais comme affinement du couplage personne-situation dans des

contextes de plus en plus complexes ; s'appuyer, à titre de repères, sur les descripteurs du CECRL (médiation) sans les réduire à une checklist.

En somme, le modèle PAS nous invite à observer moins une somme de compétences qu'un sujet en action. C'est dans cette perspective que nous abordons maintenant la question de l'oralité en classe de FLE. Il s'agira, dans cette deuxième partie, de montrer comment une approche énactive de l'oralité peut être mise en œuvre à partir d'un travail progressif sur la voix, la prosodie structurante et la prosodie expressive. L'enjeu sera de décrire comment ces dimensions permettent de réancrer la parole dans le corps, dans l'espace, dans le temps de l'énonciation et dans le milieu même de l'interaction.

3. L'oralité énactive en classe de FLE

En Thaïlande, la prise de parole en langue étrangère est souvent vécue comme anxiogène. Des travaux récents (Kaewchoom et al., 2023; Thaksanan, 2024; Suratin & Sribayak, 2025) montrent que des facteurs culturels et psychologiques spécifiques exacerbent cette anxiété lors des présentations orales. Ces constats résonnent avec ma propre pratique, nourrie par près de vingt ans d'enseignement du FLE en Thaïlande : si le contexte évolue — essor de l'intelligence artificielle, développement des réseaux sociaux qui exposent davantage les apprenants à des échanges avec des locuteurs étrangers, et multiplication des chatbots de conversation — le stress de la prise de parole persiste, avec le risque corrélatif de retrait ou de sur-dépendance aux outils.

C'est face à ces conditions que j'ai élaboré des stratégies corporelles et vocales pour restaurer la disponibilité, abaisser le seuil d'entrée en parole et soutenir une prise de parole plus confiante. Ces stratégies s'adossent au modèle PAS de Masciotra (se disposer, se situer, se positionner, (se) réaliser) et à la conception incarnée de la parole défendue par Billières (2002) — « la parole est mouvement ». Nous allons maintenant voir comment ces principes se traduisent concrètement dans le travail sur la voix et la prosodie, en montrant comment l'oralité incarnée engage progressivement le corps, l'attention et l'intention communicative.

Comme le souligne Meunier (2015, p. 13), « si l'enseignement de l'intonation a pris tellement de retard, c'est parce que les linguistes, ne sachant trop qu'en faire, et n'en ayant pas trouvé les règles, ont décidé une fois pour toutes de la mettre dans la catégorie des phénomènes extralinguistiques, comme les gestes que l'on fait en parlant. » C'est précisément ce statut marginal que nous contestons ici : loin d'être ornementale ou accessoire, la prosodie est constitutive de l'intelligibilité et de la présence vocale. Dans la perspective énactive qui est la nôtre, elle engage le corps entier et s'articule aux quatre fonctions du modèle PAS pour soutenir un agir vocal incarné et situé.

3.1 La voix

L'ancrage corporel (se disposer)

Le protocole d'initialisation est minimal : se lever, fermer les yeux, respirer profondément, puis émettre un son vocalique simple. Travailler en deçà de la langue — dans le registre du souffle et de la résonance — désamorce l'appréhension liée à la correction linguistique et installe une disponibilité corporelle hors de tout enjeu évaluatif. Cette phase engage prioritairement « se disposer » : l'entrée en présence et en énergie, la maîtrise de soi, la vigilance attentionnelle. La disponibilité constitue « la clé pour se disposer mentalement de façon à affronter les situations » ; être disponible, c'est « être une Présence », en canalisant ses ressources vers la tâche (Masciotra, 2017).

La station debout libère le diaphragme, ouvre la cage thoracique et permet une amplitude respiratoire que la posture assise comprime. La vocalisation initiale — un simple « a » ou « o » tenu — établit le contact avec sa propre voix, teste la résonance dans l'espace, éprouve la durée du souffle. Sans ce réglage initial du corps et de l'attention, l'engagement qui suit demeure aléatoire.

La projection dans l'espace (se situer / se positionner)

La mise en voix engage ensuite l'espace et la distance d'adresse. Cette phase mobilise conjointement « se situer » et « se positionner ».

« Se situer » renvoie à la conscience de l'espace et du positionnement : où suis-je ? à quelle distance se trouvent mes interlocuteurs ? de quel volume vocal je dispose ? L'apprenant perçoit tacitement l'étendue de la salle, la proximité ou l'éloignement de ses pairs, et établit un diagnostic.

« Se positionner » désigne l'ajustement stratégique qui en découle : moduler le volume, orienter le regard, adopter une posture d'adresse. Cette fonction relève de l'intelligence positionnelle, qui opère de manière pré-active et pro-active (Masciotra, 2017).

Concrètement, je demande à un apprenant de se placer à différentes distances et d'adresser la même phrase simple à un camarade. Cette variation contraint naturellement à ajuster le volume : l'apprenant perçoit physiquement la nécessité de porter davantage sa voix. Cette « pédagogie des distances » rend sensible la dimension spatiale de la parole.

L'aménagement spatial joue ici un rôle déterminant. Ouvrir l'espace — en disposant les apprenants en cercles, diagonales, ou travées dégagées — crée des interactions multidirectionnelles et matérialise ce que Delhay nomme un « espace du collectif, de l'écoute et de la parole » (2022).

L'entrée dans le collectif : (se) réaliser

La vocalisation passe au mot signifiant (« bonjour », « merci », ou tout autre mot choisi). Ce franchissement du pré-linguistique vers l'énonciation conserve la spontanéité du souffle tout en sollicitant un contrôle articulatoire minimal et l'inscription dans un code partagé.

Cette phase engage pleinement (se) réaliser, au sens de Masciotra — « ce que la personne fait effectivement, ici et maintenant, pour progresser dans une activité » (2017). Sur le plan de l'agir langagier, cette fonction prend corps dans des micro-actions observables (articulation, accentuation, contours prosodiques, enchaînements brefs, regards, mimiques, gestes co-verbaux) qui rendent l'intention immédiatement opératoire et adressée. Masciotra précise que « l'art de réalisation réfère à la qualité […] avec laquelle des transformations successives sont produites, dans la mesure où elles permettent de progresser vers le but fixé, puis de l'atteindre ». Et surtout, à mesure qu'une personne développe son pouvoir de "réaliser", elle progresse vers une réalisation de soi comme finalité — c'est pour cette raison que la fonction est formulée « (se) réaliser ». Dans l'orientation de ce travail, nous lisons cette « réalisation de soi » à la lumière de l'individuation (psychique et collective) chez Simondon, réélaborée par

Stiegler comme processus de formation du sujet en couplage avec son milieu technique et social ; ce point sera approfondi en partie 3 (discussion philosophique).

Dans le prolongement de ces micro-actions, un dispositif d'échauffement peut être mobilisé lorsque les objectifs du cours le justifient, afin de mettre en voix les apprenants et d'installer une disponibilité expressive avant des tâches orales plus exigeantes (jeu de rôle, présentation, interaction guidée). À titre d'illustration concrète, on peut recourir au dispositif qui suit, simple et rapide à mettre en place.

Dispositif. La classe est d'abord organisée en petits cercles de quatre ou cinq. Une fois les groupes constitués, la consigne initiale est donnée : dire son mot préféré. À tour de rôle, sans préparation, chaque apprenant énonce rapidement un mot. Le rythme reste bref — quelques secondes pour inspirer et oser — tandis que l'enseignant veille à prévenir toute autocensure. Les consignes varient ensuite (animal, fleur, fruit, personnage célèbre) et peuvent être combinées avec une couleur d'intonation (joie, surprise, fatigue), de sorte que chaque mot peut être porté par une intention affective. Le tour de mots devient alors un échauffement expressif : la voix se délie, l'adresse s'affirme, l'énergie circule. Souple et modulable, ce dispositif favorise l'ancrage de l'expressivité et de la disponibilité vocales, préparant efficacement des tâches orales plus exigeantes.

Co-régulation des quatre fonctions

Si cette progression privilégie successivement se disposer, se situer / se positionner, puis (se) réaliser, les quatre fonctions co-régulent constamment l'agir. Masciotra insiste sur leur interdépendance : elles ne sont pas des étapes séquentielles mais des dimensions concomitantes de l'action située.

Dès le premier temps, l'apprenant se situe déjà minimalement et se positionne en choisissant tacitement son niveau d'engagement vocal. Inversement, dans le troisième temps, la fonction « se disposer » demeure active : maintenir sa présence attentionnelle, gérer son énergie corporelle, rester disponible à l'écoute.

Cette progression permet un déploiement gradué : on stabilise d'abord la disponibilité corporelle avant d'introduire les ajustements spatiaux, puis on actualise ces dispositions dans une performance interactionnelle. Cette gradation respecte la logique

énactive selon laquelle la compétence s'actualise dans et par l'action, en réponse aux exigences émergentes de la situation.

En définitive, ce protocole incarne la conception de Billières : « La parole est mouvement » (Billières, 2002). La mise en voix devient un acte de présence où les quatre fonctions du modèle PAS s'articulent pour soutenir un agir vocal incarné et situé.

3.2 La prosodie structurante (intonation non marquée)

Définir la « prosodie structurante »

À partir de la mise en voix, la parole s'organise dans le temps : elle progresse par unités, marque des continuités et des clôtures, et structure l'information. Nous retenons, avec Billières (2002), la prosodie comme la partie de la phonétique qui traite de l'organisation des structures suprasegmentales — au premier chef l'intonation et le rythme. Concrètement, les paramètres de durée, intensité et hauteur s'ancrent sur les syllabes et forment des unités perceptibles (accents, groupes rythmiques, syntagmes/contours intonatifs) qui rendent audible la structure de l'énoncé. De fait, « la parole est mouvement » et « une courbe mélodique correspond à un mouvement intonatif » (Billières, 2002), soit un mouvement réglé dont les variations créent les repères auditifs nécessaires pour segmenter et hiérarchiser l'énoncé. Dès lors, la prosodie n'est pas un ornement : elle conditionne l'intelligibilité.

Dans ce cadre, nous nommons prosodie structurante le profil non marqué, orienté prioritairement vers la clarté syntaxique : découpage en mots phoniques, gestion des poursuites et des chutes, hiérarchisation discrète de l'information. Cette prosodie structurante repose sur ce que Meunier appelle l'intonation non marquée, c'est-à-dire le patron mélodique de base à fonction syntaxique qui organise l'énoncé en mots phoniques et en assure la lisibilité (Meunier, 2020). Tout ce qui relève d'une mise en relief intentionnelle ou affective — contrastes, surprise, évidence — appartient à la prosodie marquée et sera examiné en section 3.3.

Visualiser l'intonation non marquée : la « partition » de Meunier

Dans la lignée de la phonétique expérimentale, Christian Meunier propose une partition prosodique qui donne à voir le profil intonatif et fournit un code commun à l'enseignant et à l'apprenant. L'outil se présente comme une portée à cinq niveaux (1–

5) servant de schéma de hauteur relative : chaque syllabe est représentée par un rectangle, l'accent final du mot phonique par un triangle posé sur sa dernière syllabe, et les frontières prosodiques par les signes « / » (fin de mot phonique) et « // » (fin d'énoncé).

Dans cette échelle, le niveau 2 correspond au niveau habituel de la voix (fondamental usuel) : c'est le plan de base sur lequel se déroulent les syllabes non accentuées et le point d'attaque des énoncés affirmatifs. Le niveau 3 marque la poursuite (fin de groupe non final), le niveau 1 réalise la clôture déclarative et le niveau 4 signale la montée finale des interrogatives totales. Le niveau 5, quant à lui, ne relève pas du registre neutre : il signale les reliefs expressifs (surprise, évidence, insistance emphatique, exclamation, contraste) et sera analysé dans la section consacrée à la prosodie marquée.

Pour la lecture et la production, quelques principes suffisent. L'accent porte sur la dernière syllabe de chaque mot phonique ; les segments non accentués se déroulent au niveau 2 et l'attaque de l'énoncé affirmatif s'y place ; la fin d'un groupe non final se pose au niveau 3, ce qui maintient la continuité informationnelle ; la déclarative se clôt au niveau 1 avec un léger allongement de la syllabe finale ; l'interrogative totale se reconnaît à l'élévation vers le niveau 4 sur la dernière syllabe du groupe verbal. Ces repères minimaux rendent immédiatement perceptible la « grammaire audible » de l'intonation non marquée.

Figure 1 *: Visualisation de la prosodie d'une phrase affirmative simple (Meunier, 2020)*

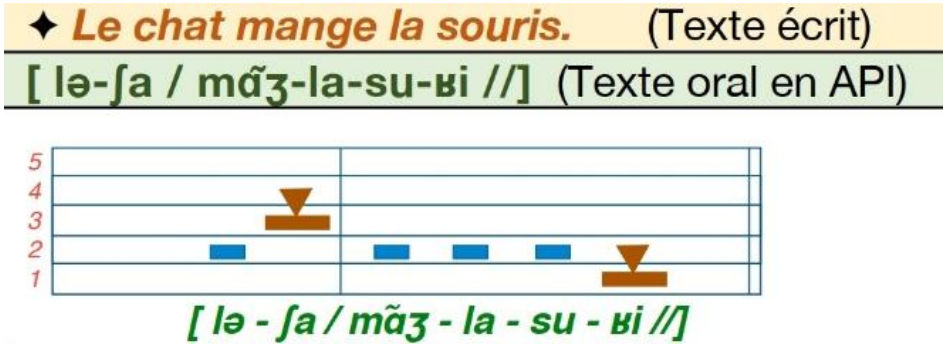
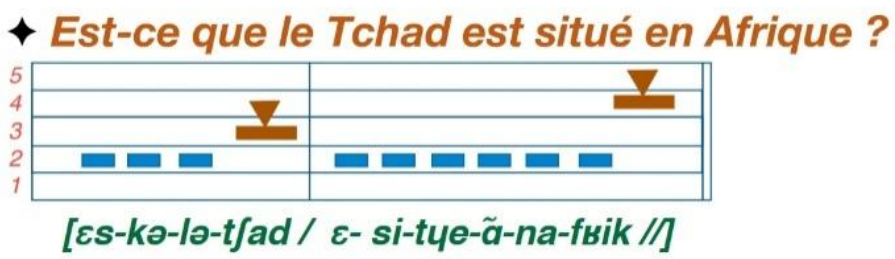


Figure 2 *: Visualisation de la prosodie d'une phrase interrogative (Meunier, 2020)*



Incarner la prosodie par le modèle PAS

Si la mise en voix prépare l'énonciation, la prosodie en façonne la forme. Dans l'esprit de Billières, l'ancrage corporel et la gestualité servent de médiateurs pour installer rythme et ligne mélodique : « la parole est mouvement », la mélodie est un mouvement intonatif, et la prise de parole engage le corps dans son ensemble (Billières, 2002).

Comme le souligne Billières (2002, p. 43), « la micro-motricité est en synergie avec la macro-motricité ». Il précise : la micro-motricité correspond à la « gestualité fine correspondant à l'activité des organes de la phonation », tandis que la macro-motricité renvoie à la « gestualité corporelle globale ». Pratiquées par les apprenants, ces deux gestualités stabilisent la performance prosodique : elles spatialisent le mouvement mélodique, régulent le débit et facilitent l'allongement vocalique et les chutes.

Côté enseignant, la gestualité a un rôle de guidage. Mais, surtout, côté apprenants, elle doit être exercée et intériorisée : tracer de la main la ligne intonative, marquer les paliers, suspendre le geste aux pauses, « tomber » physiquement sur la chute. Cette pratique corporelle n'est pas un simple doublage visuel : elle soutient l'intonation, aide à la tenir et la rend observable et mémorisable (Billières, 2002).

La partition de Meunier externalise l'architecture temporelle (paliers 1–5, mots phoniques, chutes). Combinée à la gestualité des apprenants, elle devient un véritable outil de calibration : la main suit les paliers (montées 2⟶3, descente à la chute), s'accorde au débit, s'immobilise aux silences ; le corps, quant à lui, ancre le souffle et cadre l'énergie. Ainsi, la partition ne vaut pleinement qu'« en corps » : elle complète la gestualité fine et la gestualité corporelle globale en offrant un support conjoint de

visualisation et d'incorporation, stabilisateur et facilitateur de la performance prosodique (Meunier, 2020 ; Billières, 2002).

Cette pratique mobilise directement les quatre fonctions du modèle PAS. Se disposer : ancrage corporel (posture stable, respiration basse, disponibilité attentionnelle) qui conditionne la tenue des niveaux intonatifs et la stabilité du geste. Se situer : la partition rend visible l'architecture de l'énoncé et permet à l'apprenant de se repérer (où suis-je ? quel niveau ? où est la pause ?), le geste matérialisant cette conscience en temps réel. Se positionner : ajustements stratégiques (amplitude, tempo, synchronisation) pour anticiper et réguler la production — par exemple, ralentir avant la chute finale, garder un niveau de hauteur constant sur la portion centrale de la phrase (geste horizontal pour éviter de "monter" ou "descendre" par inadvertance), ou mettre en relief le dernier mot du groupe rythmique. (Se) réaliser : concrètement, la voix exécute les actions prévues et synchronisées avec le geste : lancer une légère montée sur la syllabe porteuse, allonger la voyelle finale, marquer une pause courte (300–500 ms) au silence indiqué, et coordonner l'attaque consonantique avec l'impulsion de la main. Chaque mouvement de la main devient alors une micro-action qui guide la voix ici-et-maintenant (Masciotra, 2017 ; Billières, 2002).

Par ailleurs, dans un contexte multilingue, la gestualité facilite le débrayage de schémas d'accentuation concurrents — notamment lorsque des apprenants passent d'un accent lexical (anglais) à la saillance finale de groupe (français). Exemple : en anglais, music, chocolate, cinema, hospital portent l'accent interne (MU-sic, CHO-co-late, CI-ne-ma, HOS-pi-tal) ; en français, leurs équivalents musique, chocolat, cinéma, hôpital réalisent la saillance sur la dernière syllabe (mu-SIQUE, cho-co-LAT, ci-né-MA, hô-pi-TAL). Le geste-repère rend ce basculement immédiatement perceptible : la main reste stable pendant le groupe, puis effectue une impulsion brève haut⟶bas sur la syllabe finale au moment de son émission. Ce médiateur prosodique stabilise l'accent final et l'allongement concomitant, en aidant simplement l'oreille et le corps à adopter le profil français (Billières, 2002).

En définitive, incarner la prosodie revient à concevoir l'énoncé comme un système dynamique où se co-régulent souffle, gestes et courbes mélodiques : le geste agit comme médiateur prosodique qui stabilise l'accent et les frontières, la partition

fournit un modèle de lisibilité (durée × hauteur), et l'ensemble s'articule aux quatre fonctions de l'agir — se disposer, se situer, se positionner, (se) réaliser. Le couplage geste–voix n'ajoute pas qu'un effet expressif : il assure l'intelligibilité (unités et reliefs sont audibles) et stabilise la production en continu.

3.3 La prosodie expressive : coloration intentionnelle et incarnation

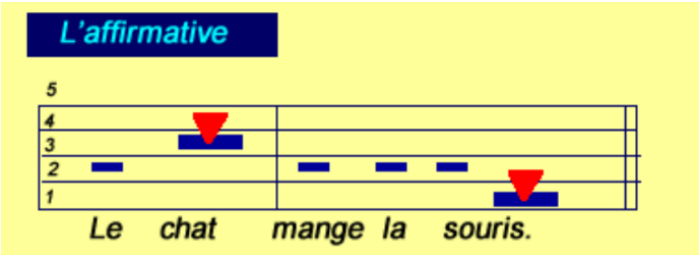
3.3.1 Les 3 modulations expressives de Meunier

Prolongeant la prosodie structurante, la prosodie expressive règle des cibles de hauteur et de durée pour des intentions usuelles — par exemple la mise en doute, l'évidence, la surprise ou l'indignation (le gabarit neutre servant de référence). En pratique, on effectue des déplacements calibrés : ligne plutôt basse avec allongement final et moue pour le doute ; montée progressive 2⟶4⟶5 en finale pour l'évidence (haussement d'épaules discret) ; surélévation des pics d'une interrogative totale et énergie accentuelle accrue pour la surprise/indignation. Ces réglages restent ancrés dans la charpente (niveaux 1–5, groupes rythmiques, poursuite/chute) et se coordonnent à des gestes co-verbaux prescrits, afin de produire une parole claire et incarnée.

Gabarit de base avec « Le chat mange la souris »

Dans le profil neutre, l'énoncé se réalise en deux mots phoniques : [lə-ʃa /], dont la dernière syllabe est au palier médian (≈ 3) signalant la poursuite, puis [/ mɑ̃ʒ-la-su-ʁi //], qui se clôt au palier bas (1). C'est ce « gabarit » qui sert de référence à toute modulation expressive.

Figure 3 *: Prosodie de la phrase affirmative neutre (Meunier, 2015)*

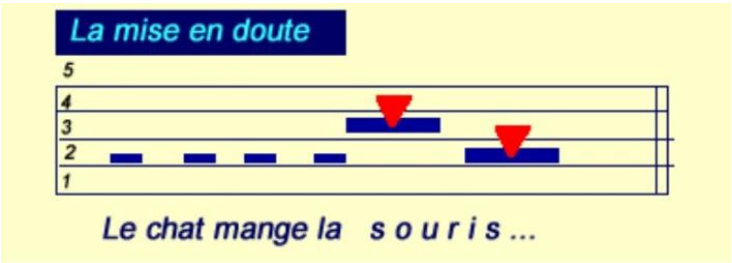


Meunier formalise trois intentions usuelles — mise en doute, évidence, surprise/indignation — qui se décrivent comme des déplacements calibrés par rapport au gabarit précédent. (Meunier, 2015, pp. 12–13.)

La mise en doute

Pour montrer à quelqu'un qu'on ne le croit pas, on répète ses propos en modifiant complètement la façon de parler. Dans « Le chat mange la souris ! », les deux dernières syllabes [su-ʁi] sont étirées longuement : elles durent deux à trois fois plus longtemps que les syllabes atones [lə-ʃa] et [mɑ̃ʒ-la]. La voix reste au niveau 2 pour l'essentiel, mais monte au niveau 3 sur l'avant-dernière syllabe [su].

Figure 4 *: Prosodie de la mise en doute (Meunier, 2015)*

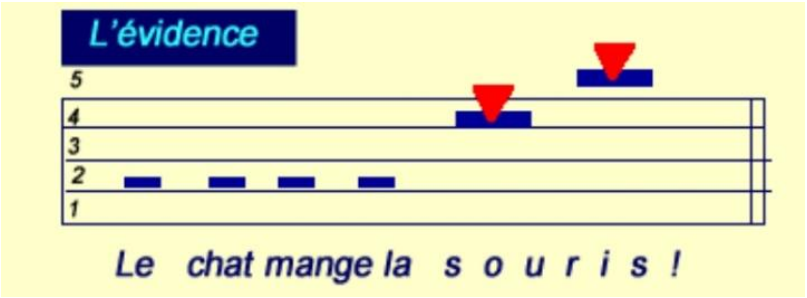


Le plus important, c'est la position des lèvres : on les projette en avant sur [su-ʁi], souvent avec une moue de dégoût. Ce geste allonge le passage de l'air dans la bouche et rend la voix plus grave et plus sourde. C'est précisément cet effet vocal qui donne l'impression de mépris ou de scepticisme.

L'évidence

Pour exprimer que quelque chose va de soi, on modifie radicalement l'intonation. Dans « Le chat mange la souris ! », les deux dernières syllabes [su-ʁi] sont accentuées, mais d'une manière particulière. La mélodie monte progressivement : l'avant-dernière syllabe [su] atteint le niveau 4 et sert de tremplin pour que la dernière syllabe [ʁi] culmine au niveau 5. Les syllabes précédentes [lə-ʃa-mɑ̃ʒ-la] restent au niveau 2.

Figure 5 *: Prosodie de l'évidence (Meunier, 2015)*



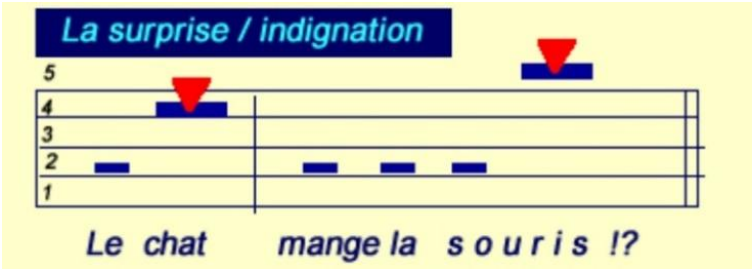
Contrairement à ce qu'on pourrait penser, il faut aborder cette montée avec calme et accentuer le moins possible les deux dernières syllabes. On peut accompagner le geste d'un haussement d'épaules sur [su-ʁi], qui renforce l'impression que c'est une évidence.

Cette intonation s'emploie pour montrer qu'on trouve quelque chose naturel ou logique. Par exemple, si un ami demande de l'aide, on répondra : « Mais bien sûr ! C'est évident ! ».

La surprise ou l'indignation

Pour exprimer la surprise ou l'indignation, on amplifie l'intonation interrogative. Dans « Le chat mange la souris ! », l'énoncé se décompose en deux parties distinctes. La première partie [lə-ʃa] commence au niveau 2 et monte brusquement au niveau 4 sur la syllabe accentuée [ʃa]. Après cette première montée, la voix redescend au niveau 2 pour les syllabes [mɑ̃ʒ-la-su-ʁi], avant de remonter progressivement vers le niveau 5 sur la dernière syllabe [ʁi].

Figure 6 *: Prosodie de la surprise ou de l'indignation (Meunier, 2015)*



Concrètement, il suffit de prendre l'intonation d'une question normale (sans mot interrogatif comme "où" ou "quand") et de hausser tous les pics mélodiques d'un niveau : le niveau 3 devient 4, et le niveau 4 devient 5. Il faut aussi mettre plus d'énergie que d'habitude dans l'accentuation des syllabes toniques.

3.3.2 De la structure à l'expression : une table de mixage

L'articulation de la prosodie expressive au modèle PAS ne diffère pas fondamentalement de celle décrite pour la prosodie structurante : les quatre fonctions (se disposer, se situer, se positionner, (se) réaliser) demeurent les mêmes régulateurs de l'agir vocal. Toutefois, comme sur une table de mixage, ce ne sont pas les canaux qui changent, mais l'intensité et le dosage de chacun.

Là où la prosodie structurante privilégie la clarté syntaxique avec une énergie mesurée et des gestes directifs, la prosodie expressive engage une intensité expressive accrue qui sollicite davantage le corps entier. Les paramètres prosodiques (durée, hauteur, intensité) ne servent plus seulement à structurer l'énoncé, mais à incarner des attitudes : le doute, l'évidence, la surprise ou l'indignation. Cette incarnation mobilise des ressources corporelles supplémentaires — mimiques faciales (moue, sourcils levés), gestes co-verbaux amplifiés (haussement d'épaules, ouverture des paumes), modulations d'énergie (accentuation plus ou moins marquée) — qui transforment la voix en présence expressive.

Les curseurs de l'expressivité

Si l'on file la métaphore de la table de mixage, trois curseurs principaux s'ajustent pour passer de la structure à l'expression :

L'énergie vocale : La mise en doute demande une énergie retenue, presque sourde, renforcée par la projection labiale qui assombrit le timbre. L'évidence requiert une montée calme mais assurée. La surprise appelle une accentuation maximale, avec des pics d'intensité marqués.

L'amplitude gestuelle : Le geste de la main, directif et sobre dans la prosodie structurante, devient signifiant. Pour le doute, il reste suspendu ou hésite. Pour l'évidence, il s'ouvre (paumes vers le haut) en accompagnement du haussement d'épaules. Pour la surprise, il amplifie brusquement les montées mélodiques.

Les mimiques faciales : La moue labiale du doute crée un rétrécissement du conduit vocal qui rend la voix plus grave. Le haussement d'épaules de l'évidence s'accompagne d'un regard franc. La surprise mobilise l'ensemble du visage : sourcils levés, yeux écarquillés, autant de marques qui rendent l'intention lisible avant même que l'énoncé ne soit achevé.

Vers l'art dramatique : élargir la palette expressive

Les trois modulations de Meunier — mise en doute, évidence, surprise — constituent un socle de base, mais le champ de la modulation expressive est bien plus vaste et confine à l'art dramatique (Meunier, 2015). Comme les musiciens qui commencent par la mélodie, le rythme et les silences avant d'interpréter une œuvre, les apprenants gagnent à stabiliser d'abord ces curseurs fondamentaux avant d'élargir progressivement leur palette.

3.3.3 Exemple appliqué : Baudelaire, « L'Invitation au voyage » (v. 1–3).

La lecture expressive de « Mon enfant, ma sœur, / Songe à la douceur / D'aller là-bas vivre ensemble ! » illustre cette multiplication des curseurs (Baudelaire, 2020, « L'Invitation au voyage », v. 1–3). Au-delà des trois paramètres de base (énergie vocale, amplitude gestuelle, mimiques), entrent ici en jeu la hauteur (focalisations douces par montées de 2 à 3 ou 4), la durée (allongements finaux sur —fant, sœur, —ceur), l'intensité (accents d'insistance brefs) et le découpage (pauses signifiantes, notamment après « Songe »). Chaque paramètre trouve son appui gestuel : mains qui s'élancent et se posent pour la hauteur, retenue au point d'atterrissage pour la durée, arrêt net pour la pause.

Quelques réglages concrets

« Mon enfant, ma sœur » : montée au niveau 3 sur les finales -fant et sœur, avec allongement légèrement emphatique ; mains projetées vers l'avant, paumes ouvertes (comme offertes), qui retombent doucement après enfant puis sœur. Les apprenants stabilisent ainsi l'accent final tout en incarnant l'adresse affectueuse de l'invitation (comme une prière).

« Songe » : le ton monte jusqu'à 4 (élan enthousiaste), puis pause brève — silence porteur d'imaginaire. En parallèle, les deux mains (paumes vers l'intérieur,

près des tempes) s'ouvrent au loin, figurant l'essor de la pensée. Les apprenants perçoivent mieux l'appel et le silence signifiant.

« à la douceur » : dire doucement, avec allongements ; doux monte à 3, la finale retombe à 2 (clôture calme). Cette inversion (pic sur doux, chute finale) met en relief "douceur" et installe un recueillement dans la douceur. Marquer ensuite une pause avant la reprise (« d'aller… »). Gestes lents et fluides, iconiques de la douceur (au choix : effleurement, enveloppement, ouverture souple des paumes) ; mimique de plaisir tranquille, à inventer par chacun.

« d'aller là-bas » : une main désigne le lointain et « atterrit » sur -bas (accentuation marquée). Le geste facilite l'accent final et clarifie la visée.

« vivre // ensemble » : dissociation rythmique (petite pause après vivre) ; allongement de vivre (mains qui s'expansent depuis la poitrine), puis allongement de la finale de ensemble (mains qui se rapprochent), symbolisant l'union.

Pour mémoire, nous sommes partis du gabarit neutre de Meunier et des trois intentions (doute, évidence, surprise), puis de la « table de mixage » (énergie, gestes, mimiques) jusqu'à la mise en œuvre gestuelle sur Baudelaire. Les observations en classe révèlent que l'articulation de la partition à cinq niveaux et de la macro-gestualité stabilise la tenue des finales et la visée des cibles de hauteur ; elles montrent aussi que ce que l'on nomme « talent » relève en réalité d'un ensemble de techniques entraînables, accessibles par l'apprentissage incarné.

Dans ma pratique d'enseignement–apprentissage de la prosodie, trois dimensions se sont imposées : (1) le déplacement dans l'espace-classe, parce que la parole s'y ancre dans un corps qui fait et habite l'espace ; (2) l'improvisation théâtrale, indissociable de la création d'un espace où la parole se situe, s'incarne et se libère — une dynamique intrinsèquement énactive et poïétique ; (3) la création collective, car l'apprenant exerce sa parole avec les autres : regards, tours de parole et co-régulations qui soutiennent l'individuation et le développement de compétences socio-interactionnelles, dans une perspective que l'on peut rapprocher de Vygotski (1978). Ces trois appuis ouvrent désormais sur la réflexion de la Partie 3 : resituer l'oralité dans une généalogie des techniques du langage (parole, écriture, IA), examiner la temporalité formatrice (rétentions husserliennes, durée bergsonienne) et proposer une

pharmacologie des usages numériques — afin de distinguer ce qui outille l'attention, la mémoire et le jugement de ce qui les court-circuite et prolétarise. Autrement dit, comprendre comment l'IA, envisagée comme *pharmakon*, peut soutenir une individuation par l'oral plutôt que s'y substituer.

4. Généalogie des techniques du langage et pharmacologie de l'oralité Pourquoi l'oralité ? Recentrer l'apprentissage face à l'IA

Dans la logique de notre pharmacologie de l'oralité — l'antidote que nous proposons face aux usages de substitution de l'IA —, il s'agit de ralentir volontairement le rythme de l'apprentissage et de rééquilibrer vers l'humain (corps, souffle, adresse), l'outil numérique ne jouant qu'un rôle auxiliaire (mesurer, visualiser, relancer). Pour situer l'enjeu technique, nous replaçons d'abord l'oral dans une généalogie des techniques du langage : parole, écriture, imprimerie, IA.

4.1 Une généalogie technique du langage

Pour comprendre ce que l'IA change dans notre rapport à la parole, il faut situer l'oral dans l'histoire des techniques du langage : la parole incarnée d'abord, puis l'écriture qui extériorise et spatialise, l'imprimerie qui industrialise la reproduction, l'IA qui automatise, à l'échelle industrielle, la production de phrases et de textes. Chaque stade reconfigure le précédent sans l'abolir.

4.2 L'oralité : premier stade anthropologique

Du point de vue anthropologique (histoire de l'espèce) comme du point de vue ontogénétique (développement de l'individu), l'oralité précède l'écriture. Elle s'enracine dans une expérience incarnée et temporelle : voix, regard, gestes, alternance des tours. Comme l'écrivait Valéry, « longtemps, longtemps, la voix humaine fut base et condition de la littérature […] Un jour vint où l'on sut lire sans épeler, sans entendre, et la littérature en fut tout altérée » (1941).

Dès l'Antiquité, Socrate — dans le *Phèdre* de Platon — met en garde contre l'écriture comme *technè* qui affaiblit la mémoire vivante (anamnèse) au profit d'un aide-

mémoire externe (hypomnèse) et produit l'apparence du savoir sans sa maîtrise réelle. Bernard Stiegler (2001) y voit la première analyse d'un problème qui traverse toute l'histoire des techniques intellectuelles : quand un support de mémoire (rétention tertiaire) se substitue au geste d'apprentissage au lieu de le soutenir, il risque de prolétariser — c'est-à-dire de déposséder l'individu de son savoir-faire.

De fait, l'écriture transforme profondément le rapport au savoir. Comme l'ont montré Goody (1979) et Ong (2014), elle spatialise le langage : elle le fixe sur une surface, permet sa manipulation (listes, tableaux, indexation), sa stabilisation et son accumulation (archivage, commentaire, recombinaison). À l'oral, en revanche, on agit avec son corps. Chez Billières (2002), la parole est mouvement, en ce qu'elle engage à la fois des micro-mouvements articulatoires et des macro-gestes corporels. Dans le prolongement de cette conception incarnée, Cyril Delhay, dans un entretien publié par Réseau Canopé (2019), définit l'oral comme une parole debout, physiquement engagée, qui mobilise le corps, la respiration et la voix.

4.3 L'oralité comme objet temporel : durée et rétentions

La différence décisive se joue dans le temps. La parole relève d'une durée vécue, au sens bergsonien : elle ne se donne jamais d'un seul bloc, mais comme un flux continu. Bergson (1889) l'illustre par la mélodie : une note isolée n'a pas de valeur ; elle ne prend sens qu'à partir de celles qui la précèdent et de celles qui la suivent. Husserl (1964 [1928]) approfondit cette intuition en montrant que le présent vivant se constitue dans la rétention de ce qui vient de s'écouler et dans la protention de ce qui vient. Comprendre la parole, ce n'est donc pas additionner des éléments successifs, mais habiter un flux, retenir, anticiper, relier.

Avec Stiegler, cette analyse s'élargit : la mémoire humaine ne se forme pas seulement dans le vécu immédiat, mais aussi à travers des supports extérieurs de conservation — écriture, livres, archives, enregistrements, dispositifs numériques. Il nomme rétentions tertiaires ces formes de mémoire extériorisée. Loin d'être de simples traces déposées hors de nous, elles orientent déjà nos souvenirs, notre attention et nos manières d'interpréter.

Dans cette histoire, l'IA marque toutefois un seuil. Le livre est déjà une mémoire extérieure, mais il laisse encore au sujet la charge du parcours interprétatif : lire, revenir en arrière, traverser une difficulté, construire peu à peu un sens. Il conserve la trace, sans dispenser du trajet. L'IA, au contraire, ne se contente plus de conserver : elle traite, sélectionne, résume, reformule, anticipe. Elle peut même, d'une certaine manière, lire pour nous. La rétention tertiaire devient alors mémoire opératoire, capable de préformater l'accès au sens.

Dès lors, une question s'impose : lorsque nous laissons l'IA lire pour nous, que perdons-nous ? Nous ne perdons pas seulement du temps de lecture ; nous risquons de perdre l'épreuve même par laquelle se forment l'attention, la mémoire longue et le jugement. L'algorithme ne modifie pas seulement notre accès aux contenus : il tend à prendre en charge le travail de reprise, d'hésitation, de sélection et de reformulation par lequel la mémoire extériorisée formait encore la mémoire vécue. Le problème n'est donc pas l'extériorisation de la mémoire en elle-même, mais la délégation croissante des opérations qui formaient le sujet interprétant.

C'est pourquoi l'oralité retrouve ici une importance décisive. Dans la parole vivante, le sujet ne peut déléguer entièrement ce travail d'appropriation : il lui faut encore habiter une durée, retenir et anticiper, ajuster sa voix, son rythme, son regard et ses gestes à la situation. Là où l'algorithme tend à livrer un sens déjà préparé, la parole réengage le corps parlant dans l'épreuve même de l'énonciation. Elle réinscrit le langage dans une temporalité formatrice, où comprendre n'est pas consommer un résultat, mais se transformer dans et par l'acte même de parler. En ce sens, l'oralité redevient un lieu d'individuation.

4.4 Face à l'IA : préserver la temporalité formatrice

L'enjeu éducatif devient alors clair. Comme le suggère Stiegler lorsqu'il oppose, dès 1990, le « temps du savoir » — qui est aussi celui de la lecture — au « temps-lumière » de l'information, la tâche pédagogique consiste à préserver les temporalités lentes de la formation (Stiegler, 1990). Mais cette opposition ne vaut pas seulement pour les médias de l'information d'alors : Stiegler la prolonge en 2013 en décrivant le web comme un espace de rétentions tertiaires numériques fonctionnant à

la vitesse de la lumière, dont l'automatisation constitue précisément le risque majeur (Stiegler, 2013). Dans le contexte actuel de l'IA, le problème n'est donc pas seulement l'accélération, mais la substitution progressive de résultats calculés aux opérations mêmes qui formaient le sujet apprenant.

Cette exigence est d'autant plus urgente que les technologies numériques contemporaines, désormais renforcées par l'IA, ne sont pas seulement des outils : elles relèvent aussi, comme l'a montré Stiegler, de formes de pouvoir qu'il a pensées dans le prolongement du biopouvoir foucaldien, sous les noms de psychopouvoir puis de neuropouvoir (Stiegler, 2008, 2015). Elles interfèrent ainsi avec les milieux traditionnels de transmission — familiaux, scolaires, culturels, politiques — en imposant des rythmes, des sollicitations et des automatismes qui déstabilisent les formes longues de l'attention. Dans la classe, l'enjeu est alors d'ouvrir un autre temps, un autre rythme, un autre espace de présence.

C'est en ce sens que Stiegler peut écrire que l'éducation consiste à cultiver le cerveau « comme un lopin de terre » (Stiegler, 2015, para. 39). L'image est décisive : cultiver, ce n'est pas produire instantanément un résultat, mais former patiemment un milieu, entretenir une mémoire, sélectionner, laisser mûrir. Dans le domaine du langage, cela signifie cultiver ce que de Saussure appelait le « trésor de la langue » : non un dépôt inerte, mais un fonds de formes partagées, déposé par la pratique de la parole. Et, selon une formule transmise par les manuscrits saussuriens, « tout ce qui entre dans la langue a d'abord été essayé dans la parole » (de Saussure, 2002, ms. E.1.383.2560.1R, 2.23). Éduquer, ce n'est donc pas déléguer ce travail à des prêts-à-parler automatiques, mais former des sujets capables de reprendre, d'ajuster et de faire vivre cette matière linguistique dans l'acte même de parler. C'est pourquoi l'oralité retrouve ici une valeur décisive : elle oblige le sujet à habiter le temps de l'énonciation, et ainsi à se former dans l'acte même de parler. Parler ne consiste plus seulement à produire un énoncé, mais à se former soi-même en formant sa parole.

5. Conclusion

Nous entrons dans une automatisation quasi générale où, portée par le développement de l'intelligence artificielle, la technique ne se contente plus d'assister certaines tâches : elle tend à se substituer à un nombre croissant d'activités de la vie professionnelle et privée, jusque dans les actes les plus ordinaires de l'existence. Elle s'insinue désormais dans la sphère quotidienne comme instance de réponse, de conseil et d'orientation, non seulement pour fournir des informations, mais pour suggérer quoi faire, quoi dire et comment agir. Chez Stiegler, le risque majeur est alors la prolétarisation : la perte des savoirs — savoir-faire, savoir-vivre, savoir-penser — lorsque la technique se substitue aux exercices par lesquels se forment les gestes, l'attention, la mémoire longue, le jugement et les manières d'agir. D'où la question décisive pour l'éducation : comment préserver l'autonomie des apprenants et cultiver le temps de l'apprendre au lieu de le court-circuiter ? « Cultiver le cerveau comme un lopin de terre » signifie alors qu'il ne suffit pas de disposer d'outils puissants : encore faut-il des techniques qui respectent le milieu qu'elles travaillent. De même qu'on ne cultive pas bien une terre lorsqu'on la réduit à un simple support d'intrants, au lieu de la nourrir comme un milieu vivant, on ne forme pas un esprit en le soumettant à des automatismes qui le dispensent d'attention et d'appropriation. L'enjeu est donc d'organiser des milieux techniques qui nourrissent le travail d'appropriation au lieu de s'y substituer.

Recentrer l'apprentissage sur l'oralité réancre la transmission dans son socle anthropologique — corps, voix, espace, durée, interactions — et rééquilibre ainsi l'apprentissage vers l'humain. À cette condition, la technique demeure auxiliaire de l'individuation, et non une machine entropique qui standardise et dépossède.

Le travail de la prosodie en donne une illustration claire : l'oral n'est jamais séparé de toute technique. Dès que la parole est spatialisée — par l'écriture, par la partition intonative, par la visualisation graphique des hauteurs, des pauses, des accents ou des longueurs de syllabes — elle devient observable, comparable et plus aisément retravaillable. En ce sens, la raison graphique (Goody, 1979) ne s'oppose pas à l'oralité : elle peut au contraire lui donner des prises nouvelles. Dans le modèle

PAS (se disposer, se situer, se positionner, (se) réaliser), l'apprenant fabrique sa parole : posture et souffle, visée et adresse, choix d'une amplitude, coordination du geste et de la voix, puis retour au profil de base. C'est là la poïésis : non l'application d'une règle, mais l'invention réglée d'une forme, un art d'habiter la langue plutôt que de la consommer.

Les technologies numériques et, désormais, certains outils fondés sur l'IA prolongent cette possibilité. Ils permettent à l'apprenant de disposer d'un retour automatique et quasi instantané sur sa production orale : prononciation, et, dans certains cas, dimensions plus proprement prosodiques comme le rythme, l'accentuation ou l'intonation. L'efficacité de tels outils n'a donc pas à être niée : ils peuvent au contraire constituer des appuis puissants pour éclairer l'apprenant sur sa propre parole et soutenir son travail d'écoute, de reprise et d'ajustement. Mais ils ne parlent pas à sa place : ils ne respirent pas, ne tiennent pas une finale, ne choisissent pas une intention. Leur place est celle d'un outillage : rendre perceptible, objectiver, accompagner le travail d'appropriation, sans s'y substituer.

D'où l'alternative stieglerienne : prolétarisation ou outillage formateur. La première survient lorsque l'outil remplace les gestes, externalise la délibération et substitue l'immédiateté de la réponse aux temporalités lentes de la mémoire longue. Le second désigne au contraire des usages techniques qui soutiennent l'attention, la reprise et l'ajustement sans déposséder le sujet de son effort. Dans l'apprentissage des langues, cela concerne aussi la matière linguistique elle-même : mots, groupes de mots, ce « trésor de la langue » déposé par la pratique de la parole, dont les usages nouveaux sont d'abord essayés dans la parole avant d'entrer dans la langue. L'enjeu n'est donc pas d'opposer la technique à l'humain, mais de distinguer les techniques qui forment de celles qui dépossèdent. C'est là que se joue l'individuation.

Ainsi, recentrer l'apprentissage sur l'oralité n'est ni repli ni nostalgie : c'est assumer une politique de l'humain incarné et situé — une éthique du temps vécu, d'une attention orientée plutôt que captée, et du jugement au sens kantien du *Sapere aude !* : « Aie le courage de te servir de ton propre entendement ! Voilà la devise des Lumières » (Kant, 2006). Une telle orientation ne vaut pas rejet de la technique, mais exigence de la réinscrire dans une pratique formatrice. Nous ne parlons pas contre la

technique, mais depuis elle : le "lopin de terre" de Stiegler rappelle que l'éducation ne forme pas hors sol, mais cultive ce qui doit l'être avec soin, dans la durée et par des médiations appropriées. C'est aussi un choix phénoménologique et énactif, au sens de Varela : reconnaître que la cognition n'est pas un calcul désincarné, mais une action située et incarnée, où le sujet se forme en traçant son propre chemin dans l'expérience. L'intelligence humaine n'est pas une donnée hors sol : elle est vivante, incarnée, située, et se forme dans l'action. C'est pourquoi il faut aussi accueillir ce qui déborde la répétition — l'erreur, l'intuition, la surprise, l'imprévu —, sans quoi il n'y aurait ni invention, ni découverte, ni véritable appropriation. C'est une évidence : on n'échappe pas à la technique, et il serait illusoire de vouloir le faire, car elle est constitutive de notre humanité. Des premiers outils et des proto-langages humains aux technologies contemporaines de traitement automatique du langage, l'histoire humaine s'est construite avec elle. Il ne s'agit donc pas de lutter contre la technique en tant que telle, mais contre une technique qui nous deviendrait néfaste en se substituant à nos capacités d'attention, de jugement et de parole. Comme l'écrit Christian Bobin, « une voix, ce n'est pas que le souffle, ni les paroles, ni même les silences. Une voix, c'est le monde entier repeint par la personne » (Bobin, 2019, p. 9). L'enjeu est que cette voix vivante continue de conduire notre formation comme sujets, de préserver notre autonomie et notre pouvoir de choix, et, dans la situation éducative, de laisser à l'apprenant la part décisive qui lui revient dans son propre développement. C'est à cette condition seulement que notre devenir collectif peut demeurer humain.

Références bibliographiques

Aden, J. (2017). Langues et langage dans un paradigme énactif. *Recherches en didactique des langues et des cultures, 14*(1). https://doi.org/10.4000/rdlc.1085

Baudelaire, C. (2020). *Les Fleurs du mal* (C. Pichois, Ed.). Gallimard.

Bergson, H. (1889). *Essai sur les données immédiates de la conscience*. PUF.

Billières, M. (2002). Le corps en phonétique corrective. In R. Renard (Ed.), *Apprentissage d'une langue étrangère/seconde. Vol. 2 : La phonétique verbotonale* (pp. 35–70). De Boeck Supérieur. https://doi.org/10.3917/dbu.renar.2002.01.0035

Bobin, C. (2019). *Pierre*. Gallimard.

Claverie, B. (2019). Dynamique exponentielle et naturalité de l'intelligence artificielle. *Hermès, La Revue, 85*, 187–200. https://doi.org/10.3917/herm.085.0187

Conseil de l'Europe. (2021). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer – Volume complémentaire*. Éditions du Conseil de l'Europe.

Delhay, C. (2022, 13 juin). Salle de classe : « Des espaces modulables, des tables en îlots, des espaces vides favorisent l'échange. Ils créent l'espace du collectif, de l'écoute et de la parole ». *Le Monde*.

de Saussure, F. (2002). *Écrits de linguistique générale* (S. Bouquet & R. Engler, Éds., avec la collab. d'A. Weil). Gallimard.

Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing, 13*(2), 208–238. https://doi.org/10.1177/026553229601300205

Gerlich, M. (2025). AI tools in society: Cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006

Goody, J. (1979). *La Raison graphique. La domestication de la pensée sauvage*. Minuit.

Husserl, E. (1964). *Leçons pour une phénoménologie de la conscience intime du temps* (H. Dussort, Trans.). PUF.

Kaewchoom, N., Thamthangsat, S., Langu, T., Rungsang, N., Meejit, C., Kawinkoonset, P., & Suebsook, K. (2023). Investigating factors causing anxiety in speaking English of Thai university students: The case of students in English program at HCU. Huachiew Chalermprakiet University. https://has.hcu.ac.th/jspui/handle/ 123456789/2816

Lhote, É., & Llorca, R. (2001). Le geste, outil d'écoute. *Le Français dans le monde. Recherches et applications*, numéro spécial, Oral : variabilité et apprentissages, 160–164.

Masciotra, D., Roth, W.-M., & Morel, D. (2007). *Enaction: Toward a Zen Mind in Learning and Teaching* (Vol. 14). Sense Publishers.

Masciotra, D., Roth, W.-M., & Morel, D. (2008). *Énaction. Apprendre et enseigner en situation*. De Boeck Supérieur.

Masciotra, D., & Medzo, F. (2009). *Développer un agir compétent : Vers un curriculum pour la vie*. De Boeck Supérieur.

Masciotra, D. (2017). La compétence : entre le savoir agir et l'agir réel. Perspective de l'énaction. *Éthique publique. Revue internationale d'éthique sociétale et gouvernementale, 19*(1). https://doi.org/10.4000/ethiquepublique.2888

Meunier, C. (2015). *Petit guide pratique de la phonétique corrective du FLE*. Éditions du FLE.

Meunier, C. (2020). *Initiation à la phonétique corrective du FLE : Phonèmes, prosodie et intonation*. BoD – Books on Demand.

Ong, W. J. (2014). *Oralité et écriture : La technologie de la parole*. Les Belles Lettres.

Pelissero, C. (2019). Énaction : principe et moyens dans le cadre du français langue étrangère. *Synergies Europe, 14*, 159–174. https://gerflint.fr/Base/Europe14 /pelissero.pdf

Réseau Canopé. (2019, septembre). L'apprentissage de l'oral, un enjeu de société : Entretien avec Cyril Delhay. *Magazine « Oral » – Nouveaux Programmes*. https://www.reseau-canope.fr/nouveaux-programmes/magazine/ oral/lapprentis sage -de-loral-un-enjeu-de-societe.html

Simondon, G. (1989). *L'individuation psychique et collective*. Aubier.

Stiegler, B. (1990). Les temps de la lecture et les nouveaux instruments de la mémoire. Ars Industrialis. https://arsindustrialis.org/node/1935

Stiegler, B. (1994). *La technique et le temps, 1 : La faute d'Épiméthée*. Galilée.

Stiegler, B. (2001). *La technique et le temps, tome 3 : Le temps du cinéma et la question du mal-être*. Galilée.

Stiegler, B. (2008). *Économie de l'hypermatériel et psychopouvoir : Entretiens avec Philippe Petit et Vincent Bontems*. Mille et une nuits.

Stiegler, B. (2013). L'esprit des Lumières à l'époque du philosophical engineering. *Intellectica. Revue de l'Association pour la Recherche Cognitive, 59*(1), 29–40. https://doi.org/10.3406/intel.2013.1084

Stiegler, B. (2015). Du psychopouvoir au neuropouvoir. In C. Larsonneur, A. Regnauld, P. Cassou-Noguès, & S. Touiza (Éds.), *Le Sujet Digital* (pp. 42–60). Presses universitaires de Paris Nanterre. https://doi.org/10.4000/books.pupo.25602

Suratin, N., & Sribayak, V. (2025). Factors contributing to speaking anxiety and anxiety reduction techniques in Thai adult EFL learners. *LEARN Journal: Language Education and Acquisition Research Network, 18*(1), 294–319. https://doi.org/10.70730/MYTS3987

Thaksanan, P. (2024). The study of EFL students' public speaking anxiety and academic oral presentation strategies at a tertiary level in Thailand. *Journal of Education and Innovation, 26*(4), 48–59. https://so06.tci-thaijo.org/index.php/edujournal_nu/article/view/270538

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.

Valéry, P. (1941). *Tel quel*. Gallimard.

Varela, F. J. (1989). *Invitation aux sciences cognitives* (1re éd.). Seuil.

Varela, F. J. (1996). *Invitation aux sciences cognitives* (rééd.). Seuil.

Varela, F., Thompson, E., & Rosch, E. (1993). *L'inscription corporelle de l'esprit : Sciences cognitives et expérience humaine*. Seuil.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications, 12*, Article 621. https://doi.org/10.1057/s41599-025-04787-y

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*, Article 28. https://doi.org/10.1186/s40561-024-00316-7