

Approche multi-niveaux de la prosodie du français L2/L3 : analyse temporelle et mélodique chez des apprenants sinophones en lecture non ponctuée

Yufeng YANG^{1*}

¹Université Paris 8, CNRS, Laboratoire Structures Formelles du Langage (SFL), 75017 Paris, France

Abstract. Cet article examine l'acquisition de la prosodie du français L2/L3 chez des apprenants sinophones à travers une investigation multi-niveaux. Basée sur les productions de 28 locuteurs en lecture non ponctuée, l'analyse se structure en trois niveaux : Micro, Méso et Macro. Les résultats révèlent une dissociation fonctionnelle : dans la dimension temporelle (Micro et Macro), l'influence de la L1 (chinois mandarin) est prédominante, les deux groupes d'apprenants affichant des profils similaires mais distincts des natifs. En revanche, dans la dimension mélodique (Méso), si les bilingues font face à une forte instabilité (hésitations), la présence de la L2 (anglais) chez les trilingues amorce une restructuration prosodique, manifestant les prémices d'un système mélodique hybride et autonome.

1 Introduction

Ces dernières années, la recherche sur l'acquisition des langues secondes (L2) s'est progressivement étendue au domaine plus complexe de l'acquisition d'une troisième langue (L3). Toutefois, si le transfert lexical ou morphosyntaxique a fait l'objet de nombreuses études, l'acquisition de la prosodie en L3 reste un domaine relativement sous-exploré [1], laissant en suspens la question des mécanismes qui régissent l'organisation temporelle et mélodique chez les locuteurs plurilingues.

Pour les apprenants sinophones, le français (langue intonative avec accent de phrase) constitue rarement la première langue étrangère (L2). Il s'agit plutôt d'une troisième langue (L3), dont l'acquisition se fonde sur le substrat du chinois mandarin L1 (langue tonale) et de l'anglais L2 (langue intonative à accent lexical) préalablement acquis. Cette configuration complexe du répertoire linguistique des apprenants implique que l'acquisition phonologique de la L3 ne se réduit plus à une simple mise en correspondance entre la L1 et la langue cible, mais relève d'un processus de restructuration dynamique, impliquant de multiples interactions translinguistiques [2, 3].

* Corresponding author: yufeng.yang19@etud.univ-paris8.fr

Pour évaluer la performance réelle à l'interface syntaxe-prosodie, cette étude adopte un protocole expérimental exigeant : la lecture de texte non ponctué. En supprimant les indices visuels explicites, cette tâche contraint l'apprenant à effectuer un traitement syntaxique en temps réel pour segmenter le flux de parole. De plus, pour pallier la subjectivité des analyses auditives, nous utilisons une méthode d'évaluation objective basée sur l'alignement entre les frontières prosodiques produites (IPU) et les constituants syntaxiques prédéfinis (AP et IP).

En s'appuyant sur ce cadre méthodologique, cette recherche propose une investigation systématique structurée en trois niveaux : Micro (six métriques rythmiques), Méso (contours des frontières syntaxiques d'AP et d'IP) et Macro (allongement final ; quantité, durée et ratio des IPU ; distribution des pauses internes et externes). Cette approche multi-niveaux vise à déterminer si les influences de la L1 et/ou de la L2 agissent de manière uniforme sur la prosodie en L3, ou si, comme nous le supposons, chaque niveau prosodique suit sa propre trajectoire d'acquisition. Enfin, une telle étude multi-niveaux de la prosodie chez des locuteurs trilingues (chinois mandarin L1/anglais L2/français L3) fait encore défaut dans la littérature, lacune que la présente recherche tente de combler.

2 Niveau micro : De la classification rythmique traditionnelle aux métriques acoustiques

2.1 Typologie rythmique et métriques acoustiques multidimensionnelles

La dichotomie traditionnelle opposant les langues à rythme syllabique (*Syllable-timed languages*, comme le français) aux langues à rythme accentuel (*Stress-timed languages*, comme l'anglais) repose sur l'hypothèse de l'isochronie [4-6]. Pour les apprenants sinophones, cette classification crée une ambiguïté typologique. Bien que le chinois mandarin (L1) soit une langue tonale, il est globalement classé comme une langue à rythme syllabique [7], à l'instar du français (caractérisé par une relative isochronie syllabique et l'absence de réduction vocalique) [8]. À l'inverse, l'anglais (L2) se distingue par une métrique accentuelle et de forts contrastes de durée. Cependant, fondée sur l'isochronie, cette classification n'a pas résisté aux données acoustiques [9].

Malgré ses limites, cette dichotomie traditionnelle reste fondamentale : elle sous-tend la « psycho-typologie » des apprenants au cœur du *Typological Primacy Model* (cf. section 5.1). Pour la dépasser et objectiver ces propriétés typologiques, nous adoptons une approche multidimensionnelle incluant six métriques acoustiques, ce qui permet de dissocier la complexité syllabique de la variabilité durative :

- Complexité syllabique globale [8] : Le %V (proportion d'intervalles vocaliques) reflète le degré d'ouverture des syllabes. Ainsi, le français et le chinois mandarin (syllabes majoritairement ouvertes) présentent des valeurs de %V élevées, tandis que l'anglais (clusters consonantiques complexes) affiche des valeurs de %V faibles. Quant au ΔC (écart-type des intervalles consonantiques), il mesure la variabilité de la complexité consonantique : le français et le chinois mandarin se caractérisent par des valeurs de ΔC faibles (structure stable), à l'opposé de l'anglais qui présente des valeurs de ΔC élevées.

- Variabilité locale (*Pairwise Variability Index*, PVI) [10] : Ces indices mesurent le contraste de durée entre segments adjacents. Le nPVI-V (PVI vocalique normalisé)

quantifie la réduction vocalique : par exemple, l'anglais (alternance voyelles pleines/réduites) montre des valeurs élevées, mais le français et le chinois mandarin (stabilité vocalique) tendent vers des valeurs faibles. Le rPVI-C (PVI consonantique brut) mesure l'irrégularité des groupes consonantiques successifs, avec des valeurs élevées pour l'anglais et faibles pour la régularité du français et du chinois mandarin.

- Variabilité globale normalisée (Varco) [11] : Afin de neutraliser l'effet du débit de parole, qui varie selon le niveau des apprenants, nous utilisons les coefficients de variation (VarcoV et VarcoC). Ces deux mesures confirment la stabilité temporelle sur l'ensemble de l'énoncé. Comme pour les PVI, on attend des valeurs élevées pour l'anglais (stress-timed) et faibles pour le français et le chinois mandarin (syllable-timed).

Bien que la capacité de ces métriques à valider des catégories rythmiques strictes soit débattue [12], nous les mobilisons ici uniquement comme des indicateurs acoustiques descriptifs pour quantifier la variabilité temporelle et la complexité syllabique des apprenants.

2.2 Études empiriques et lacunes

Sur le plan acoustique, une convergence typologique apparaît entre la L1 (chinois mandarin) et la L3 (français) : toutes deux partagent des valeurs attendues élevées pour le %V et faibles pour les mesures de variabilité. L'anglais (L2) se situe à l'opposé sur tous les plans. Les travaux portant sur l'acquisition du rythme français par des apprenants sinophones sont rares et se concentrent en français L2. Toutefois, ces études suggèrent qu'un transfert positif de la L1 (chinois mandarin) pourrait faciliter l'acquisition du français (deux langues à rythme syllabique) par les apprenants sinophones, notamment en ce qui concerne les mesures de %V et de VarcoV [13, 14]. Par ailleurs, certains travaux [15] ont montré que, parmi six métriques examinées, seul le %V permettait de distinguer significativement les apprenants sinophones de français L2 des locuteurs natifs. De même, d'autres recherches [16] indiquent que les indices rPVI-C, %V, ΔV et ΔC signalent des écarts significatifs entre les locuteurs natifs et les apprenants sinophones bilingues de différents niveaux. Par conséquent, malgré la proximité typologique entre le chinois mandarin et le français, ces recherches révèlent qu'une convergence rythmique totale reste difficile à atteindre sur l'ensemble des plans acoustiques.

3 Niveau méso : interface syntaxe-prosodie

Ce niveau examine la réalisation des frontières prosodiques syntaxiques. Dans le cadre de cette étude, l'intonation est analysée via le système Polytonia [17] (cf. section 6.3).

3.1 Accentual Phrase (AP) : Interférence des langues sources

En français, le Groupe Accentuel (ci-après AP, pour *Accentual Phrase*) est une unité délimitée par ses frontières (*edge-based*). La norme prosodique impose une montée obligatoire de la fréquence fondamentale (H) sur la dernière syllabe du groupe, fonctionnant comme un marqueur de continuité syntaxique [18, 19]. Cette exigence de marquage final se heurte aux systèmes prosodiques des langues sources :

En anglais (L2), l'intonation est régie par l'accent lexical ; ainsi, l'anglais ne possède pas de contrainte prosodique équivalente à la montée frontalière obligatoire du français [20], la cible tonale haute étant alignée sur la syllabe accentuée. Par conséquent, si la syllabe frontière d'AP est inaccentuée (ce qui est fréquent), elle ne porte pas de ton haut structurel et tend à être réalisée avec une fréquence fondamentale basse ou neutre.

Le chinois mandarin (L1), en tant que langue à tons, privilégie la préservation des tons lexicaux sur l'intonation de phrase. Cependant, il n'existe pas de mécanisme intonatif assignant systématiquement un ton haut aux frontières prosodiques mineures comme l'AP [21]. La réalisation mélodique à ces frontières reste entièrement dominée par le ton lexical du dernier caractère (syllabe). Ainsi, le contour aux frontières d'AP tend à rester plat ou à suivre la ligne de déclinaison globale pour préserver le sens lexical.

Par conséquent, aucune des langues sources ne soutient la montée systématique (H) aux frontières d'AP du français. Cette convergence du répertoire linguistique prédit une difficulté à réaliser les montées de continuation du français L2/L3 chez les apprenants.

3.2 Intonation Phrase (IP) : Convergence et variations

Au niveau supérieur (*Intonation Phrase*, ci-après IP), correspondant à l'énoncé complet, les trois langues convergent prototypiquement vers un contour descendant en fin de phrase déclarative (L%) [22]. Cependant, cette convergence de surface masque des mécanismes de production distincts : en français (cible), la chute finale est purement démarcative. Elle se matérialise par un ton de frontière infra-bas (B) associé à la syllabe finale de l'IP, marquant une conclusivité nette en touchant le plancher de la tessiture [22]. En anglais (L2), le schéma intonatif canonique en fin de phrase déclarative est une chute dynamique, initiée par un accent nucléaire descendant, convergeant vers un ton bas [20]. En chinois mandarin (L1), la chute globale correspond principalement à un abaissement progressif du registre (*register lowering* / déclinaison) qui se superpose aux tons lexicaux (T1–T4), plutôt qu'à l'association d'un ton de frontière phonologiquement spécifié [21].

Bien que les apprenants parviennent généralement à réaliser la chute finale, la nature exacte du contour descendant produit demeure une question ouverte.

3.3 Études empiriques et lacunes

Les recherches sur le contour mélodique du français L2 chez les sinophones sont limitées. Toutefois, une étude révèle une déviation majeure au niveau mélodique : influencés par le mécanisme de *pitch reset* du chinois mandarin, les apprenants (notamment aux niveaux intermédiaires) tendent à produire des contours descendants aux frontières d'AP, brisant la continuité mélodique attendue par les locuteurs francophones natifs [23]. À ce jour, le rôle médiateur de l'anglais (L2) chez les trilingues reste inexploré.

4 Niveau macro : Organisation temporelle

4.1 Allongement final

L'allongement final constitue l'indice acoustique primaire de la segmentation en français :

il se manifeste par une augmentation significative de la durée syllabique finale de l'AP ou de l'IP. En français (cible), l'allongement est obligatoire et structurel. Il marque la frontière droite des constituants prosodiques indépendamment de la présence de pauses physiques [18, 19]. Bien que le chinois mandarin (L1) présente un allongement pré-pausal, son amplitude et sa fonction diffèrent du français. Le chinois mandarin s'appuie prioritairement sur des particules syntaxiques (ex : « 的 de ») ou sur la réinitialisation de la hauteur (*pitch reset*) pour délimiter les frontières, rendant l'allongement temporel moins saillant [21]. De plus, l'anglais (L2) possède un allongement final de syntagme [24], mais la présence de l'accent lexical crée une confusion pour les apprenants : il est difficile de distinguer l'allongement induit par l'accent de celui induit par la frontière.

4.2 IPU : Quantité, Durée, Ratio

L'analyse des Unités Inter-Pausales (IPU) permet d'examiner la granularité du flux de parole et la gestion du souffle. En français, la segmentation est principalement définie par la structure syntaxique. Les IPU correspondent idéalement à des groupes de souffle relativement longs, englobant plusieurs groupes accentuels (AP) et alignés sur les frontières syntaxiques majeures [18, 25]. Cette organisation reflète un principe de structuration fondé sur les frontières, où la segmentation du discours est étroitement liée à la hiérarchie syntaxique. Le chinois mandarin (L1), en revanche, privilégie une segmentation plus fréquente. Les réinitialisations de hauteur (*pitch reset*), essentielles à l'organisation tonale, conduisent à une production caractérisée par des unités plus courtes et plus nombreuses, en comparaison avec les langues romanes [21, 26]. Quant à l'anglais (L2), bien qu'il prenne également en compte les frontières syntaxiques majeures (clauses), son organisation prosodique repose davantage sur la structure accentuelle. En tant que langue à rythme accentuel, il structure la parole en pieds métriques centrés sur les syllabes accentuées, ce qui tend à produire une segmentation plus locale et moins dépendante du souffle [6, 27].

Ainsi, la différence entre ces langues ne se limite pas à des préférences de segmentation, mais reflète des principes organisateurs distincts : le français privilégie une structuration par les frontières syntaxiques, tandis que l'anglais est davantage contraint par la distribution des accents, et le mandarin par les exigences du système tonal. Dans cette perspective, les trois métriques retenues capturent des dimensions complémentaires de cette organisation : le nombre total d'IPU reflète le degré de fragmentation du discours ; la durée moyenne des IPU indique la capacité à maintenir le flux articulatoire ; enfin, le ratio de parole normalise ces mesures par rapport au temps total et fournit un indice de densité de segmentation.

4.3 Distribution des pauses : Structuration vs Disfluency

Dans la lecture, la position de la pause est aussi déterminante que sa durée. Le silence remplit deux fonctions distinctes selon son emplacement dans la structure prosodique :

- **Pauses Externes (Structuration) :** Situées aux frontières syntaxiques majeures, elles reflètent une stratégie de découpage du discours réussie. En lecture non ponctuée, leur placement correct témoigne d'une compréhension syntaxique [28].

- **Pauses Internes (Disfluency)** : Situées à l'intérieur des groupes accentuels (AP), elles signalent une rupture involontaire (hésitation, planification) [29].

L'enjeu pour les apprenants n'est pas seulement de réduire le silence, mais de déplacer les pauses des positions internes (non fluentes) vers les positions externes (structurantes), passant ainsi d'une segmentation subie (hésitation) à une segmentation par la syntaxe.

4.4 Études empiriques et lacunes

La littérature sur la prosodie du français L3 chez des locuteurs trilingues reste également limitée, a fortiori sur des tâches de lecture non ponctuée. Concernant les apprenants sinophones en français L2, ils produisent bien un allongement final [23] ; cependant, cet allongement est souvent excessif par rapport à celui des natifs, particulièrement chez les apprenants de niveau intermédiaire. De plus, les apprenants sinophones bilingues tendent à transférer la stratégie de segmentation du chinois mandarin (basée sur la pause) vers le français, insérant des pauses à l'intérieur des AP, contrairement au groupe natif. Notre étude vise à vérifier si l'ajout de l'anglais (L2) modifie cette tendance.

5 Cadre théorique, questions et hypothèses

5.1 *Typological Primacy Model* et *Dynamic Model of Multilingualism*

Dans l'acquisition d'une L3, la question centrale est de déterminer quelle langue source (L1 ou L2) exerce l'influence dominante. Nous adoptons d'abord le *Typological Primacy Model* (TPM) [30, 31], qui postule un transfert initial guidé par la « psycho-typologie » (proximité structurelle perçue). Ce modèle est particulièrement pertinent en prosodie, où l'acquisition repose sur des indices macroscopiques : malgré les limites acoustiques de la dichotomie rythmique traditionnelle (cf. section 2), le chinois mandarin et le français partagent une saillance perceptive réelle (tendance isosyllabique et absence de réduction vocalique). Sur cette base, le TPM prédit logiquement une sélection du mandarin (L1) comme source de transfert, au détriment de l'anglais (L2) [7].

Toutefois, pour rendre compte de la complexité de notre tâche, nous y associons le *Dynamic Model of Multilingualism* (DMM) [32]. En effet, la forte charge cognitive induite par la lecture non ponctuée est susceptible de provoquer des régressions ou des interférences concurrentes de la L2. Cette double focale théorique nous permet ainsi de modéliser non seulement le transfert massif de la L1, mais aussi l'émergence d'un système hybride sous la pression de la L2.

5.2 Questions et hypothèses de recherche

Sur la base du cadre théorique, nous proposons les trois questions et hypothèses suivantes :

- **Niveau micro** : Dans quelle mesure les six métriques rythmiques du français L2/L3 produites par des apprenants sinophones se rapprochent-elles de celles des natifs ?

H1 : Nous postulons que le groupe bilingue (chinois mandarin L1 / français L2) affichera une performance rythmique plus proche de la cible native que le groupe trilingue (chinois mandarin L1 / anglais L2 / français L3). En l'absence de l'anglais dans

leur répertoire, les bilingues devraient bénéficier d’un transfert positif ou neutre de la L1. En revanche, les trilingues subiraient un transfert négatif des propriétés rythmiques de leur anglais L2 (*stress-timed*).

• Niveau méso : Les apprenants du français L2/L3 parviennent-ils à produire les mouvements mélodiques appropriés aux frontières syntaxiques d’AP et d’IP ?

H2 : Conformément au TPM, les apprenants devraient transférer certaines propriétés prosodiques du mandarin (L1), notamment le *register lowering* aux frontières d’IP et le *pitch reset* aux frontières d’AP. Toutefois, une différence est attendue entre les deux groupes : les trilingues pourraient présenter une plus grande variabilité mélodique en raison de l’influence potentielle de l’anglais (L2), tandis que les bilingues devraient manifester un système plus homogène, principalement influencé par la L1.

• Niveau macro : Les apprenants utilisent-ils l’allongement final de manière appropriée pour marquer les frontières d’IPU ? Observe-t-on un morcellement excessif du discours (IPU courtes, pauses fréquentes) ?

H3 : Nous supposons que les deux groupes réaliseront un allongement final, mais d’une durée excessive par rapport aux natifs. Toutefois, les bilingues présenteraient une organisation plus fluide (moins morcelée) que les trilingues, qui devront gérer un « double conflit prosodique » entre les tendances syllabique (L1) et métrique (L2).

6 Méthodologie

6.1 Participants

28 participants, répartis en trois groupes, ont pris part à cette étude :

- Le groupe natif (Natif) est composé de 8 locuteurs natifs du français résidant à Paris.
- Le groupe bilingue (Bilingue, chinois mandarin L1 / français L2) est composé de 10 apprenants sinophones qui, bien qu’exposés à l’anglais scolairement (niveau très bas), ne l’utilisent plus activement au moment de l’étude ni au quotidien. Cette distinction repose donc sur l’usage effectif plutôt que sur l’apprentissage antérieur, le français (niveau C1) constituant leur unique langue étrangère active.
- Le groupe trilingue (Trilingue, chinois mandarin L1/anglais L2/français L3) est composé de 10 étudiants universitaires sinophones qui ont appris l’anglais (niveau B2+) durant leur scolarité en Chine avant de débiter le français (L3, niveau C1).

Tous les participants ont déclaré ne souffrir d’aucun trouble de l’audition ou de la parole. Le Table 1 ci-dessous présente les informations des participants.

Table 1. Informations linguistiques des participants.

Groupes	Âge (ans)	Genre (F/H)	Années d'apprentissage de la L2 (ans)	Années d'apprentissage de la L3 (ans)	Séjour en France (ans)
Natif	36.2 (9.64)	6F/2M	-	-	-
Bilingue	25.7 (2.79)	8F/2M	français 7.9 (3.00)	-	2.2 (1.77)
Trilingue	23.2 (2.17)	9F/1M	anglais 12.6 (3.00)	français 5.89 (2.76)	0.94 (1.51)

Note : Les données sont présentées sous la forme : Moyenne (Écart-type/ *Standard Deviation*). F = Femme ; H = Homme. Le tiret (-) indique une donnée non applicable.

6.2 Corpus et Procédure

Le matériel expérimental est constitué d'un texte narratif continu non ponctué, adapté de la fable « Le Lièvre et la Tortue » (ici intitulée « Le lapin et la tortue »), qui comprend 206 mots répartis en 5 paragraphes, 15 phrases (IP) et 70 AP (cf. Annexe). Ce choix méthodologique vise à neutraliser les indices visuels explicites de segmentation. Les participants sont ainsi contraints de s'appuyer exclusivement sur leur compétence syntaxico-prosodique interne pour structurer le discours. Cela nous permet d'observer les stratégies spontanées de démarcation (pauses, allongements, intonation) et de vérifier si les apprenants identifient correctement les frontières des constituants syntaxiques.

Les enregistrements ont été réalisés individuellement dans une salle insonorisée. La procédure s'est déroulée en deux phases consécutives : une familiarisation avec le texte (2 à 3 minutes de lecture silencieuse), suivie de l'enregistrement à haute voix.

6.3 Traitement et Analyse des données

Dans un premier temps, les fichiers sonores ont fait l'objet d'une segmentation semi-automatique à l'aide du logiciel SPPAS [33], générant un alignement forcé au niveau des phonèmes, des syllabes, des mots et des IPU. Puis, une vérification et une correction manuelles systématiques ont été effectuées sous *Praat* [34].

Pour la segmentation en IPU, nous avons retenu un seuil de 150 ms, inférieur au seuil classique de 200 ms [27]. Ce choix permet, d’une part, d’éviter les faux positifs liés aux occlusives sourdes du français dont la tenue excède rarement 120 ms [35], et d’autre part, de capturer les micro-pauses (150-200 ms) fréquemment utilisées par les apprenants pour marquer la syntaxe en lecture non ponctuée [36], qu’un seuil strict de 200 ms masquerait.

L'analyse de l'intonation a été réalisée via *Prosogram*, en utilisant le module de transcription automatique *Polytonia* [16]. Ce système permet de convertir la courbe de F0 brute en une représentation stylisée basée sur la perception tonale. L'annotation finale comprend sept niveaux (*Tiers*) tels qu'illustrés par la figure 1 ci-dessous :

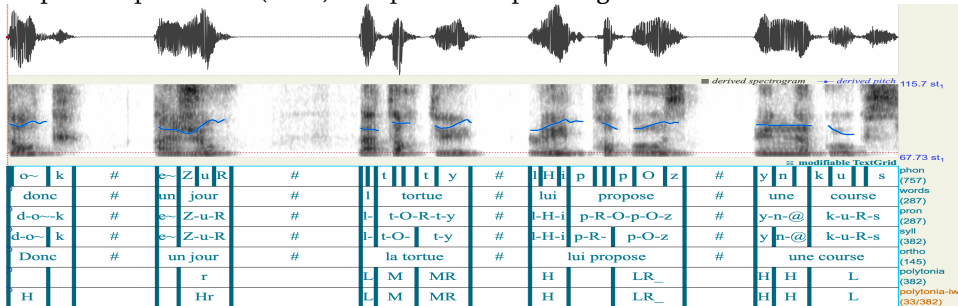


Fig. 1. Exemple d'annotation multi-niveaux et de stylisation mélodique d'un apprenant sous *Praat*

Un point méthodologique crucial concerne la localisation des mesures au niveau méso. L'extraction des symboles générés par Polytonia a été effectuée aux frontières théoriques d'AP et d'IP (définies par la structure syntaxique), et non aux frontières acoustiques

effectivement produites par les apprenants (qui peuvent inclure des pauses erronées ou des disfluences). Ce protocole permet d'évaluer la conformité tonale à l'endroit exact exigé par la structure cible, indépendamment des hésitations ou des pauses erronées. Enfin, l'extraction des données a été automatisée via des scripts *Praat* personnalisés.

7 Résultats

Les analyses statistiques et visualisations ont été réalisées sous *R Studio*. Après traitement des valeurs aberrantes (troncature 5-95 %), les variables continues ont été soumises à des ANOVA à un facteur suivies de tests *post-hoc* de Tukey ($p < .05$). Les données catégorielles (contours intonatifs) ont quant à elles été évaluées par des tests du Chi-deux d'indépendance, complétés par une analyse des résidus standardisés.

7.1 Niveau micro : six métriques rythmiques

La comparaison des performances acoustiques entre les trois groupes (Natifs, Bilingues, Trilingues) selon les six métriques retenues est illustrée dans la Figure 2.

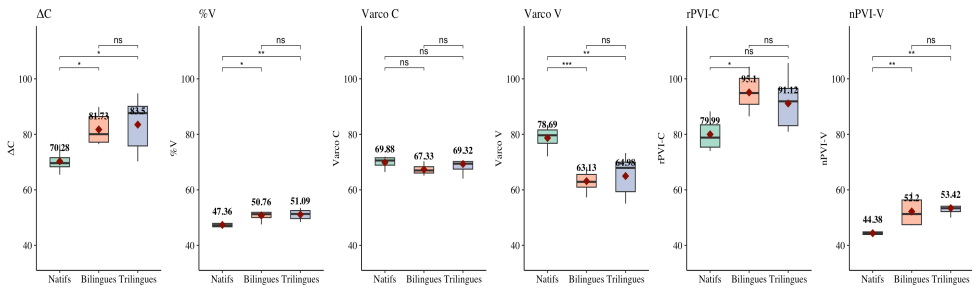


Fig. 2. Comparaison des métriques rythmiques entre les trois groupes

L'examen de la figure 2 révèle une tendance générale claire : les deux groupes d'apprenants (Bilingues et Trilingues) présentent des profils rythmiques très proches l'un de l'autre (aucune différence significative relevée ; $p > .05$). En revanche, ils se distinguent significativement des natifs sur la majorité des métriques (ΔC , $\%V$, VarcoV et nPVI-V).

Plus spécifiquement, les apprenants bilingues et trilingues affichent des valeurs significativement plus élevées que les natifs pour le $\%V$ (50.76 / 51.09 vs 47.36) et le nPVI-V (52.20 / 53.42 vs 44.38). Ces valeurs des natifs se situent dans la fourchette typique des langues à rythme syllabique [8, 10]. Ainsi, les apprenants traduisent une élocution plus lente caractérisée par un allongement vocalique accru. Dans notre corpus, cette tendance peut être directement attribuée à une épenthèse vocalique pour faciliter la production des groupes consonantiques complexes (ex : « vrai » est réalisé comme « [və.ʁe] » chez les apprenants). Ces écarts doivent être interprétés comme des indicateurs de stratégies articulatoires plutôt que comme un strict changement de catégorie rythmique [12].

À l'inverse, les natifs présentent un VarcoV supérieur ($p < .01$). Bien que pour les langues à tendance isosyllabique, ces valeurs des trois groupes soient globalement très élevées et atypiques [11], il est intéressant de noter l'absence de différence significative pour le VarcoC . Cela suggère que les différences significatives observées en ΔC entre apprenants et natifs sont principalement attribuables à l'effet du débit de parole.

7.2 Niveau méso : interface syntaxe-prosodie

La Figure 3 présente la distribution des contours intonatifs observés aux frontières syntaxiques. Les contours dont la fréquence est inférieure ou égale à « 3 » sont regroupés dans la catégorie « Autre », et la catégorie « 0 » correspond à l'absence de tout marquage.

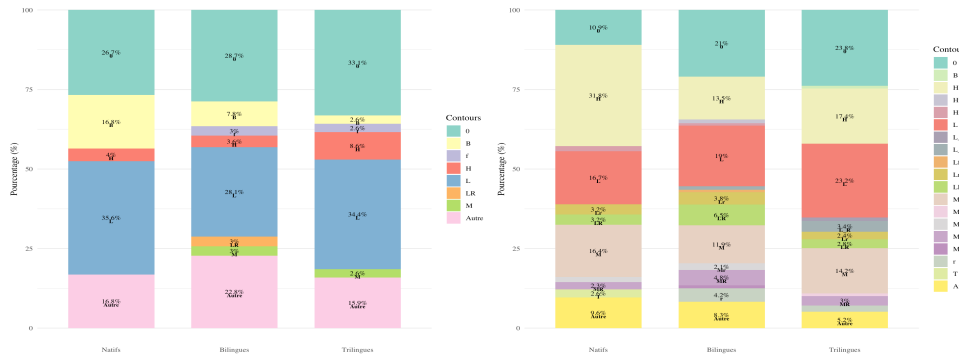


Fig. 3. Distribution des contours intonatifs aux frontières syntaxiques d'IP (à gauche) et d'AP (à droite) pour les trois groupes de locuteurs

La figure 3 (à gauche) révèle que la distribution des contours aux frontières d'IP diffère significativement selon les groupes ($\chi^2(8) = 40.32, p < .001$). Chez les natifs, le système est dominé par le contour descendant « L » (35.6 %) et le contour « B » (16.8 %, résidu = +4.33), tandis que le contour « H » est activement évité (résidu = -2.51), ce qui valide les descriptions standards [22]. Les bilingues manifestent une phase de transition stable : bien qu'ils s'écartent légèrement du « L » natif (28.1 %), aucune déviation statistique n'apparaît sur les contours « B » ou « H ». Les trilingues amorcent une restructuration : s'ils convergent vers le « L » cible (34.4 %), ils s'en distinguent radicalement par une sous-représentation drastique du « B » (2.6 %, résidu = -4.27) et une surreprésentation significative du « H » (8.6 %, résidu = +3.01). D'ailleurs, la catégorie non-marquage « 0 » des trois groupes reste statistiquement neutre.

Aux frontières d'AP (Fig. 3. à droite), la variation est également hautement significative ($\chi^2(24) = 159.39, p < .001$). Chez les natifs, le contour montant « H » prédomine (31.8 %, résidu = +6.44), reflétant la norme de continuation ascendante [22] et un évitement actif du non-marquage « 0 » (résidu = -4.41). Cette dynamique se rompt chez les apprenants. Si les bilingues sous-représentent drastiquement le « H » (13.5 %, résidu = -4.34), c'est au profit de contours atypiques ou d'hésitation : on observe chez eux une surreprésentation significative des patrons « Mr », « r », « LR » et « Hf » (présentant tous des résidus > +3.00). En revanche, ils ne manifestent pas de préférence statistique pour le « 0 » ou le « L » (résidus < 1.96). À l'inverse, les trilingues structurent une véritable stratégie d'évitement en substituant de manière significative des patrons plats « 0 » (résidu = +2.88) ou descendants « L » (23.2 %, résidu = +2.16) aux attentes du français.

En somme, l'acquisition prosodique en L3 dépasse la simple extension de la L2. Si les bilingues conservent une stabilité en IP mais montrent une forte instabilité en AP (contours d'hésitation), les trilingues restructurent l'ensemble du système. Ils systématisent des substitutions en AP (« 0 », « L ») et réorganisent les contrastes en IP (« H »), confirmant ainsi l'émergence d'un système intonatif hybride propre à la L3.

7.3 Niveau macro : Organisation temporelle

7.3.1 Allongement final

Comme l’illustre la Figure 4, l’analyse de la durée syllabique a été menée selon deux axes. La Figure 4 présente les durées syllabiques moyennes, exprimées en millisecondes (ms).

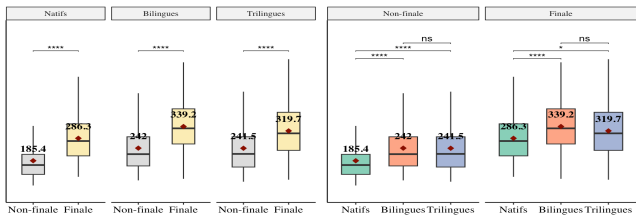


Fig. 4. Analyse de la durée syllabique : comparaisons intra-groupe (gauche) et inter-groupes (droite).

D’une part, la comparaison intra-groupe (à gauche) confirme un allongement final robuste et significatif pour les trois groupes ($p < .0001$). D’autre part, l’analyse inter-groupes (à droite) révèle que les apprenants conservent un débit significativement plus lent que les natifs, quelle que soit la position, sans aucune différence notable entre bilingues et trilingues.

Ces résultats suggèrent que la règle prosodique de l’allongement final est acquise et appliquée, puisque les apprenants respectent la structure rythmique du français malgré un débit plus lent. Cependant, le statut de trilingue ne semble pas conférer d’avantage supplémentaire pour la réalisation de cet allongement.

7.3.2 IPU : Quantité, Durée et Ratio

La Figure 5 présente l’analyse des IPU. Trois indicateurs ont été examinés : le nombre total d’IPU, la durée moyenne par IPU et le ratio de parole.

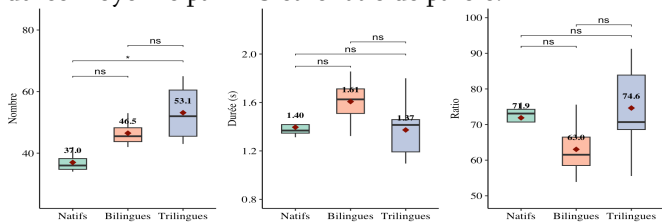


Fig. 5. Analyse des IPU : quantité (à gauche), durée (au centre) et ratio (à droite).

L’analyse du nombre d’IPU (à gauche) met en évidence une fragmentation croissante du discours : natifs (37) < bilingues (46.5) < trilingues (53.1). Cette différence n’est statistiquement significative qu’entre les natifs et les trilingues ($p < .05$). Concernant la durée moyenne de chaque IPU (au centre) et le ratio de parole normalisé pour 100 secondes (à droite), les analyses statistiques ne révèlent aucune différence significative entre les trois groupes ($p > .05$).

Les apprenants, et tout particulièrement les trilingues, tendent à morceler leur discours en un nombre d’IPU, témoignant d’une fragmentation plus marquée. En outre, leur adaptation rythmique ne s’opère pas par la modification de la durée des segments

eux-mêmes (qui reste stable), mais par l'augmentation de la fréquence de segmentation.

7.3.3 Distribution des pauses

La Figure 6 compare la distribution temporelle des pauses silencieuses selon leur position : externe (entre deux énoncés/IP) ou interne (au sein des énoncés/IP).

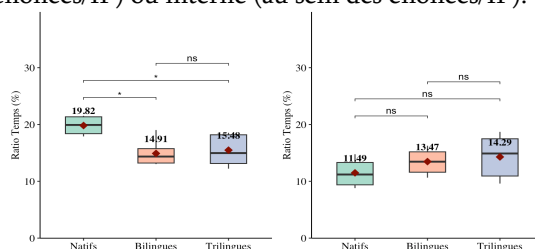


Fig. 6. Analyse des ratios de pauses externes (à gauche) et internes (à droite).

Concernant le ratio de pause externe (à gauche), les natifs y consacrent une part significativement plus importante de leur temps (19.82 %, $p < .05$) que les bilingues (14.91 %) et trilingues (15.48 %), sans distinction entre ces deux groupes d'apprenants ($p > .05$). À l'inverse, le ratio de pause interne (à droite) ne révèle aucune différence inter-groupes significative ($p > .05$).

Les apprenants bilingues et trilingues avancés démontrent une fluidité comparable aux natifs à l'intérieur des phrases (pauses internes), mais se distinguent par une gestion différente des frontières discursives (pauses externes), qu'ils tendent à raccourcir.

8 Conclusion et Discussion

Notre étude visait à déterminer comment les apprenants sinophones (bilingues et trilingues) organisent la prosodie du français dans une tâche de lecture non ponctuée. Nos résultats mettent en lumière une dynamique d'acquisition complexe où le chinois mandarin L1 exerce une influence prépondérante, mais non exclusive. Globalement, nous observons une dissociation fonctionnelle entre les niveaux prosodiques :

- Au niveau temporel (Micro & Macro) : La tendance syllabique du mandarin impose sa structure rythmique au français, semblant bloquer tout transfert potentiel de l'anglais. Ces résultats, interprétés à la lumière des critiques récentes sur les métriques rythmiques [12], soulignent des stratégies articulatoires (comme l'épenthèse) plutôt qu'un changement de catégorie typologique.

- Au niveau mélodique (Mésos) : Face à la difficulté des contours cibles du français, les trajectoires divergent. Si les bilingues pallient ce manque par une forte instabilité (hésitations), c'est précisément chez les trilingues que l'anglais (L2) manifeste son influence latente. L'analyse statistique (Chi-deux et résidus) prouve qu'ils opèrent une véritable restructuration (patrons de substitution et réorganisation), confirmant l'émergence d'un modèle intonatif hybride.

À la lumière de ces observations, il convient à présent de réexaminer nos hypothèses :

Niveau Micro : Prédominance de la L1. Notre hypothèse de départ (H1) est invalidée. L'équivalence des performances entre bilingues et trilingues démontre que l'anglais (L2)

n'impacte pas l'acquisition du rythme français : la L1 demeure la source exclusive de transfert. De plus, bien que le chinois mandarin et le français partagent une tendance syllabique, des écarts significatifs persistent avec les natifs (ΔC , %V, VarcoV, nPVI-V). Cela suggère que la proximité typologique ne garantit pas l'acquisition immédiate des finesses rythmiques, confirmant la nécessité de traiter ces métriques comme des indicateurs descriptifs de la variabilité interlangue.

Niveau Méso : Divergence et Restructuration. Ces résultats nuancent l'hypothèse du TPM. D'une part, les apprenants bilingues parviennent à maintenir une conformité globale aux frontières macro-prosodiques (IP), leur incapacité à réaliser la cible aux frontières internes (AP) ne débouche sur aucune stratégie compensatoire unifiée, mais génère plutôt une forte instabilité articulatoire (prolifération de contours d'hésitation). D'autre part, les apprenants trilingues dépassent ce stade d'instabilité subie pour opérer une véritable restructuration, ils déploient un évitement systématique face à la difficulté en AP (substitution par des patrons simplifiés ou transférés) tout en réorganisant de manière atypique les contrastes en IP. Cette dynamique révèle que l'apprentissage d'une L3 favorise l'émergence d'un système phonologique hybride et autonome, où l'expérience métalinguistique issue de la L2 permet de structurer activement une nouvelle grammaire plutôt que de simplement subir les interférences de la L1.

Niveau Macro : Contraintes cognitives communes. L'hypothèse (H3) est également invalidée. L'absence de différence significative entre les deux groupes d'apprenants (durée syllabique, nombre d'IPU, ratio de pauses) indique que la maîtrise de l'anglais (L2) n'apporte ni avantage ni interférence sur l'organisation temporelle globale. À ce niveau avancé, les contraintes de fluidité et de planification semblent être universelles pour les apprenants sinophones, indépendamment de leur répertoire plurilingue.

En somme, l'acquisition de la prosodie L3 ne suit pas une évolution linéaire de la L2. Elle opère une dissociation fonctionnelle : la dimension temporelle est dominée par la L1 (TPM), tandis que la dimension mélodique subit une amorce de restructuration sous l'influence combinée des répertoires antérieurs (L1 et L2), aboutissant à l'émergence d'un système intonatif hybride et autonome (DMM). Pour l'avenir, il conviendra de comparer les productions des mêmes apprenants dans leurs langues sources (L1 et L2) et d'étendre l'analyse à la parole spontanée pour vérifier si cette configuration hybride persiste.

References

1. I. Mennen, Phonological and phonetic influences in non-native intonation, in *Non-Native Prosody: Phonetic Description and Teaching Practice*, J. Trouvain and U. Gut (Eds.), (De Gruyter Mouton, Berlin, New York, 2007), pp. 53-76.
2. M. Wrembel, Cross-linguistic influence in second vs. third language acquisition of Phonology, in *Universal or Diverse Paths to English Phonology*, U. Gut, R. Fuchs, and E. M. Wunder (Eds.), (Mouton De Gruyter, Berlin, 2015)
3. M. Wrembel, M. Marecka, R. Kopečková, Extending perceptual assimilation model to L3 phonological acquisition, *Int. J. Multiling.* **16**, 513-533 (2019)
4. A. Lloyd James, *Speech signals in telephony*, (London, 1940)
5. K.L. Pike, *The intonation of American English*, (University of Michigan Press, Ann Arbor, 1945)

6. D. Abercrombie, *Elements of general phonetics*, (Edinburgh University Press, Edinburgh, 1967)
7. P.P.K. Mok, On the syllable-timing of Cantonese and Beijing mandarin, *Chin. J. Phon.* **2**, 148–154 (2009)
8. F. Ramus, M. Nespore, J. Mehler, Correlates of linguistic rhythm in the speech signal, *Cognition* **73**, 265-292 (1999)
9. R.M. Dauer, Stress-timing and syllable-timing reanalyzed, *J. Phonetics* **11**, 51-62 (1983)
10. E. Grabe, E.L. Low, Durational variability in speech and the rhythm class hypothesis, in *Papers in Laboratory Phonology 7*, C. Gussenhoven and N. Warner (Eds.), (Mouton de Gruyter, Berlin, 2002), pp. 515-546
11. V. Dellwo, Rhythm and speech rate: A variation coefficient for deltaC, in *Language and language processing*, P. Karnowski and I. Szigeti (Eds.), (Peter Lang, Frankfurt am Main, 2006), pp. 231-241
12. A. Arvaniti, Measuring speech rhythm, in *The Cambridge handbook of phonetics*, R.-A. Knight and J. Setter (Eds.), (Cambridge University Press, Cambridge, 2022), pp. 312-335
13. C. Gabriel, A. Hu, L. Diao, J. Thulke, Transfer, phonological awareness und Mehrsprachigkeitsbewusstsein: Zum Erwerb des französischen Sprachrhythmus durch Schüler/innen mit chinesischem Sprachhintergrund im deutschen Schulkontext, *Z. Fremdsprachenforsch.* **23**, 53–76 (2012)
14. C. Gabriel, J. Stahnke, J. Thulke, On the acquisition of French speech rhythm in a multilingual classroom: Evidence from linguistic and extra-linguistic data, in *Congrès Mondial de Linguistique Française CMLF 2014*, SHS Web Conf. **8**, 1267-1283 (2014)
15. C. Pillot-Loiseau, H. Xie, Transfert rythmique du chinois mandarin au français dans l'apprentissage du Français Langue Étrangère : acoustique et perception, in *Congrès Mondial de Linguistique Française*, SHS Web Conf. **46**, 09001 (2018)
16. L. Xi, R. Ridouane, French Rhythm in Read Speech by French Natives and Chinese L2 Learners, in *Proceedings of the 20th International Congress of Phonetic Sciences*, 1633-1637 (2023)
17. P. Mertens, Polytonia: a system for the automatic transcription of tonal aspects in speech corpora, *J. Speech Sci.* **4**, 17-57 (2014)
18. S.-A. Jun, C. Fougeron, Realizations of accentual phrase in French intonation, *Probus* **14**, 147–172 (2002)
19. E. Delais-Roussarie, Interface Phonologie/Syntaxe : des domaines phonologiques à l'organisation de la Grammaire, in *Phonologie et Phonétique : Approches actuelles*, J. Durand, N. Nguyen, V. Rey, S. Wauquier-Gravelines (Eds.), (Éditions Hermès, Paris, 2005), pp. 159-183
20. D.R. Ladd, *Intonational phonology*, 2nd edn., (Cambridge University Press, 2008)
21. X.S. Shen, *The Prosody of Mandarin Chinese*, (University of California Press, 1990)
22. S.A. Jun, C. Fougeron, A Phonological Model of French Intonation, in *Intonation*, A. Botinis (Ed.), *Text Speech Lang. Technol.* **15**, (Springer, Dordrecht, 2000), pp. 209-242. https://doi.org/10.1007/978-94-011-4317-2_10
23. L. Xi, S. Wachs, R. Ridouane, Production of French final stressed syllables in Accentual Phrase by Chinese learners: A pilot study, in *Proceedings of the 10th International Conference on Speech Prosody (SP2020)*, 895–899 (2020)
24. A.E. Turk, S. Shattuck-Hufnagel, Word-boundary-related duration patterns in English, *J. Phonetics* **28**, 397–440 (2000), <https://doi.org/10.1006/jpho.2000.0123>

25. A. Di Cristo, Intonation in French, in *Intonation Systems: A Survey of twenty languages*, D.J. Hirst, A. Di Cristo (Eds.), (Cambridge University Press, Cambridge, UK, 1998), pp. 195-218
26. C.-y. Tseng, S.-H. Pin, Mandarin Chinese prosodic phrase grouping and modeling - method and implications, in *Proc. First International Symposium on Tonal Aspects of Languages (TAL 2004)*, 193-196 (2004)
27. F. Goldman-Eisler, *Psycholinguistics: Experiments in spontaneous speech*, (Academic Press, London, 1968)
28. F. Grosjean, Linguistic structures and performance structures: Studies in pause distribution, in *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*, H. Dechert, M. Raupach (Eds.), (Mouton, The Hague, 1980), pp. 91-106
29. N.H. de Jong, Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency, *Int. Rev. Appl. Linguist. Lang. Teach.* **54**, 113-132 (2016)
30. J. Rothman, L3 syntactic transfer selectivity and typological determinacy: The Typological Primacy Model, *Second Lang. Res.* **27**, 107-127 (2011)
31. J. Rothman, Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered, *Biling. Lang. Cogn.* **18**, 179-190 (2015), <https://doi.org/10.1017/S136672891300059X>
32. P. Herdina, U. Jessner, *A Dynamic Model of Multilingualism Analysed*, in *A Dynamic Model of Multilingualism: Perspectives of Change in Psycholinguistics*, (Multilingual Matters, Bristol, Blue Ridge Summit, 2002), pp. 111-143
33. B. Bigi, SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech, *Phonetician* **111-112**, 54-69 (2015)
34. P. Boersma, Praat, a system for doing phonetics by computer, *Glott Int.* **5**, 341-345 (2001)
35. D. Duez, Silent and Non-Silent Pauses in Three Speech Styles, *Lang. Speech* **25**, 11-28 (1982), <https://doi.org/10.1177/002383098202500102>
36. M. Candea, Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané. Etude sur un corpus de récits en classe de français., Ph.D. thesis, Université de la Sorbonne nouvelle - Paris III (2000)

Annexe : Texte support de l'expérience

Le lapin est très fier de sa vitesse. Il se moque souvent de la tortue et lui dit T'es si lente Donc un jour la tortue lui propose une course. Le lapin est surpris mais il accepte en riant et dit C'est impossible que tu gagnes

Le lendemain tous les animaux se rassemblent pour voir la course. Au signal de départ le lapin part très vite et laisse la tortue loin derrière. Il pense avec confiance Je suis bien trop rapide pour elle En voyant un arbre le lapin décide de s'y reposer et s'endort.

Pendant ce temps la tortue avance lentement mais ne s'arrête jamais. Elle dépasse le lapin endormi et continue vers la ligne d'arrivée. D'autres animaux sont étonnés et se demandent La tortue va-t-elle vraiment gagner

Quand le lapin se réveille il voit la tortue près de la fin du parcours. Il s'affole et se met à courir mais il est trop tard. La tortue franchit la ligne d'arrivée en premier.

Essoufflé et honteux le lapin demande Comment est-ce possible La tortue souriante répond Rien ne sert de courir il faut partir à point. Tous les animaux applaudissent la tortue et le lapin apprend une leçon importante sur la patience et le travail.