

Chapitre 8

Analyses en parole spontanée

Ce chapitre présente les résultats d'analyses des annotations de pauses et d'accentuation produites par PLSPP sur les trois corpus de parole spontanée conversationnelle. Nous présentons d'abord les analyses relatives à la distribution des pauses, puis celles concernant les patterns accentuels. Chaque section débute par les résultats obtenus avec les locuteurs B1 et B2 du corpus CLES-FR, suivis d'une comparaison des groupes de niveau de locuteurs japonophones (CLES-JP) et des locuteurs anglophones natifs (CLES-EN).

Pour le corpus CLES-FR, l'extraction automatique des segments de parole par le premier module de PLSPP a permis d'identifier 1 559 segments, représentant un total de 10 h20 min de parole continue produite par 70 locuteurs B1¹ et 99 locuteurs B2. Les corpus CLES-JP et CLES-EN ont respectivement fourni 275 segments (3 h 37 min, 29 locuteurs) et 113 segments (1 h 55 min, 14 locuteurs). Les analyses présentées dans ce chapitre reposent sur ces ensembles de segments.

8.1 Analyse des patterns de pauses

1.1 Corpus CLES-FR

Sur les 10 h20 min de parole continue extraite du corpus CLES-FR, 72 140 intervalles inter-mots ont été analysés. Parmi ces intervalles, 22 796 (32%) présentent une durée supérieure à 180 ms, 1 085 (1,5%) dépassent 2 s, et 83 (0,1%) excèdent 5 s.

¹Un des locuteurs n'a pas généré de segments d'au moins 8 s, ce qui explique un effectif final de 70 locuteurs pour le corpus CLES-FR.

Seuils de durée	p-value	Δ de Cliff	médianes	moyennes	écarts-types
180 ms-2 s	< 0,05	0,021	481 – 474	600 – 585	400 – 390
180 ms-5 s	< 0,01	0,024	501	701 – 675	616 – 594
250 ms-2 s	ns	0,009	581	693 – 683	391 – 380
250 ms-5 s	ns	0,014	602 – 601	812 – 791	631 – 610

TAB. 8.1 : Différence de distribution de durée de pauses entre B1 et B2 selon différents seuils de durée. Avec la p-value du test non-paramétrique Wilcoxon-Mann-Whitney, le Δ de Cliff, et la médiane, la moyenne et l'écart type des deux distributions.

Durées et fréquences des pauses

Pour les analyses, nous avons fixé un seuil minimal de 180 ms, considérant que seuls les intervalles suffisamment longs pour être perceptibles par l'auditeur peuvent être interprétés comme des pauses. Par ailleurs, un seuil maximal de 2 s a été retenu pour exclure les intervalles potentiellement dus à des erreurs d'alignement, comme détaillé dans le chapitre 5. Ainsi, 21 710 intervalles (30%) ont été identifiés comme des pauses. Ces pauses présentent une durée médiane de 481 ms, avec un premier quartile à 281 ms et un troisième quartile à 782 ms.

La comparaison des durées de pauses entre les locuteurs B1 et B2 montre une différence faible, bien que statistiquement significative ($p < 0,05$, médianes respectives de 481 ms et 474 ms, cf. figure 8.1a). Cependant, cette différence devient non significative si l'on élève le seuil minimal à 250 ms (médianes à 581 ms). L'inclusion de pauses plus longues (jusqu'à 5 s) n'affecte pas notablement les résultats : la différence reste significative avec un seuil de 180 ms-5 s ($p < 0,01$, médianes de 501 ms) mais devient non significative avec un seuil de 250 ms-5 s. Ces résultats suggèrent que les pauses courtes (moins de 250 ms) pourraient contribuer à distinguer les niveaux B1 et B2, tandis que les pauses longues (supérieures à 2 s) ne sont pas discriminantes. Dans l'ensemble, la durée des pauses ne semble pas suffire à différencier les deux groupes, comme en témoigne le Δ de Cliff, toujours proche de 0 (cf. tableau 8.1).

En examinant la durée moyenne des pauses par locuteur (cf. figure 8.1b), aucune différence significative n'apparaît entre les B1 et B2. Cependant, la forme des distributions révèle des distinctions : la distribution des B1 présente une queue longue, indiquant que certains locuteurs font des pauses moyennes particulièrement longues (6 locuteurs au-delà de 700 ms), tandis que celle des B2 est plus compacte, et présentant au contraire 5 locuteurs dont la durée moyenne des pauses est particulièrement courte.

Le débit de parole des locuteurs B1, significativement plus lent que celui des B2 ($p < 0,001$, $\Delta = -0,35$ (moyen), médianes respectives de 96 et 107 tokens/minute, cf. figure 8.2a), invalide l'utilisation du nombre de pauses par minute comme indicateur

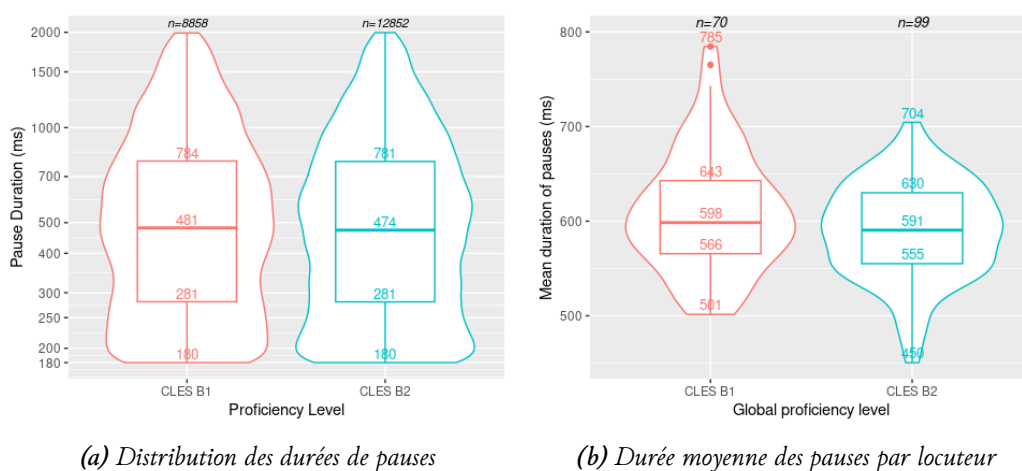


Fig. 8.1 : Durées des pauses dans le corpus CLES-FR (180 ms-2 s)

pertinent (différence non significative, médianes respectives de 32 et 34 pauses par minute). En revanche, le nombre de pauses par token, qui neutralise l'effet de la vitesse d'élocution, révèle une différence significative : les B1 font plus de pauses par token que les B2 ($p < 0,05$, $\Delta = 0,154$ (faible), médianes respectives de 0,32 et 0,29 pauses/token, cf. figure 8.2b).

Distribution syntaxique

Bien que les mesures de débit de parole et de fréquence des pauses permettent de différencier les locuteurs B1 et B2, ces critères ne suffisent pas à expliquer les variations de compréhension. Le chapitre 3 souligne l'importance de la distribution syntaxique des pauses. Des études antérieures ont montré que la fréquence des pauses situées à l'intérieur des groupes syntaxiques est souvent négativement corrélée avec la perception de fluidité, tandis que les pauses situées entre les groupes semblent avoir un impact moindre (Kahng, 2014 ; Kallio et al., 2022 ; Shea & Leonard, 2019 ; Suzuki & Kormos, 2020).

Pour analyser ces patterns, nous avons distingué deux types de groupes syntaxiques : les propositions et les syntagmes. Une pause est classée comme interpropositionnelle (*between clauses*, *BC*) si elle se situe en frontière de proposition, intersyntagme (*between phrases*, *BP*) en frontière de syntagme, ou intra-syntagme (*within phrases*, *WP*) si elle est à l'intérieur d'un syntagme. Par cohérence avec la littérature, nous rapportons aussi les pauses intra-propositionnelles (*within clauses*, *WC*), englobant les pauses inter- et intra-syntagmes.

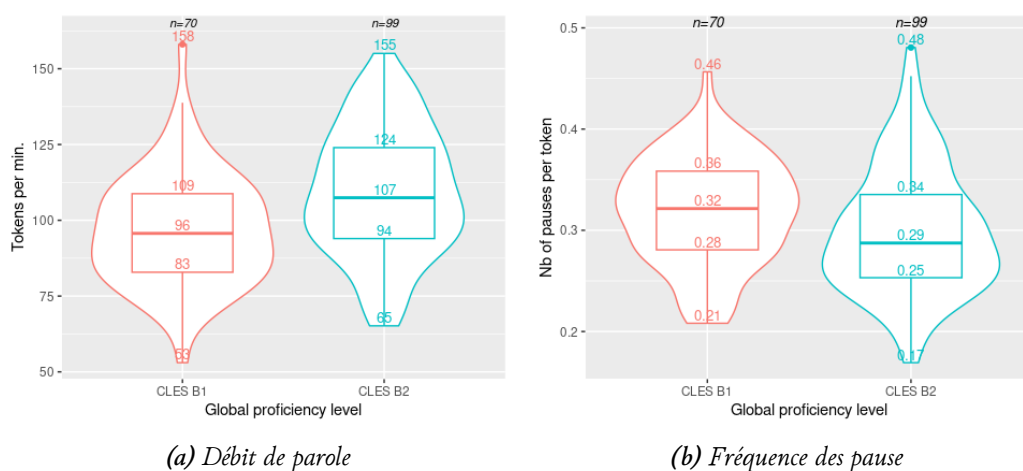


FIG. 8.2 : Débit de parole (gauche) et fréquence des pauses (droite) par locuteur dans le corpus CLES-FR (180 ms-2 s)

La figure 8.3 montre les proportions de pauses selon le type de frontière syntaxique. À nombre égal de propositions, les B1 effectuent davantage de pauses inter-propositionnelles que les B2 ($p < 0,001$, $F_{p,BC}$ médianes : 47 % contre 42 %, $\Delta = 0,311$ (faible), $IC = [0,132; 0,47]$), mais pas significativement plus de pauses intra-propositionnelles ($ns.$, $F_{p,WC}$ médianes : 28 % contre 25 %, $\Delta = 0,172$ (faible), $IC = [0; 0,334]$). Tous les locuteurs semblent donc privilégier les frontières syntaxiques de haut niveau (entre propositions) pour placer leurs pauses.

Au niveau des syntagmes, les B1 réalisent davantage de pauses inter-syntagmes, mais la différence n'est pas significative ($ns.$, $F_{p,BP}$ médianes : 29 % contre 26 %, $\Delta = 0,149$ (faible), $IC = [-0,027; 0,316]$). En revanche, les pauses intra-syntagmes sont significativement plus fréquentes ($p < 0,05$, $F_{p,WP}$ médianes : 21 % contre 18 %, $\Delta = 0,187$ (faible), $IC = [0,009; 0,353]$).

Pour neutraliser l'effet du nombre total de pauses, nous avons calculé la fréquence relative de chaque type ($P_{p,i \in \{BC, WC, BP, WP\}}$, cf. figure 8.4). La proportion de pauses inter-propositionnelles est alors légèrement plus faible chez les B1, sans différence significative ($ns.$, $P_{p,BC}$ médianes : 35 % contre 36 %, $\Delta = -0,069$). Les proportions de pauses inter-syntagmes sont également similaires ($ns.$, $P_{p,BP}$ médianes : 52 %, $\Delta = -0,069$). En revanche, la proportion de pauses intra-syntagmes reste significativement plus élevée chez les B1 ($p < 0,05$, $P_{p,WP}$ médianes : 12 % contre 11 %, $\Delta = 0,216$ (faible), $IC = [0,037; 0,382]$).

Ainsi, la principale différence entre les B1 et B2 réside dans la fréquence des pauses en frontières de bas niveau syntaxique. Cependant, cette différence demeure limitée en ampleur.

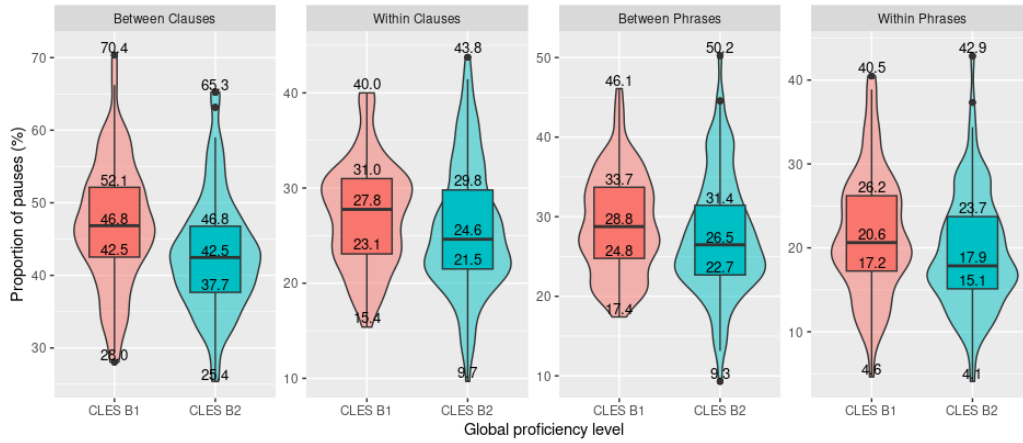


FIG. 8.3 : Proportion de pauses par type de frontière syntaxique ($F_{p,i \in \{BC,WC,BP,WP\}}$)

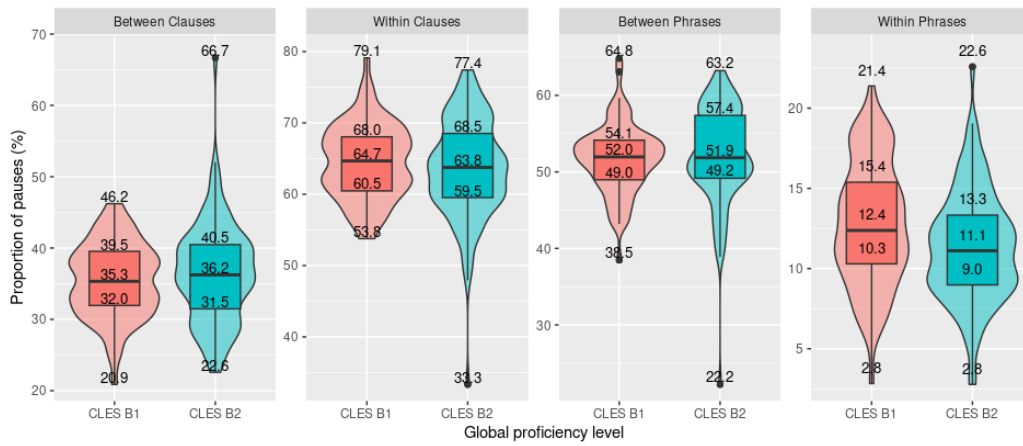
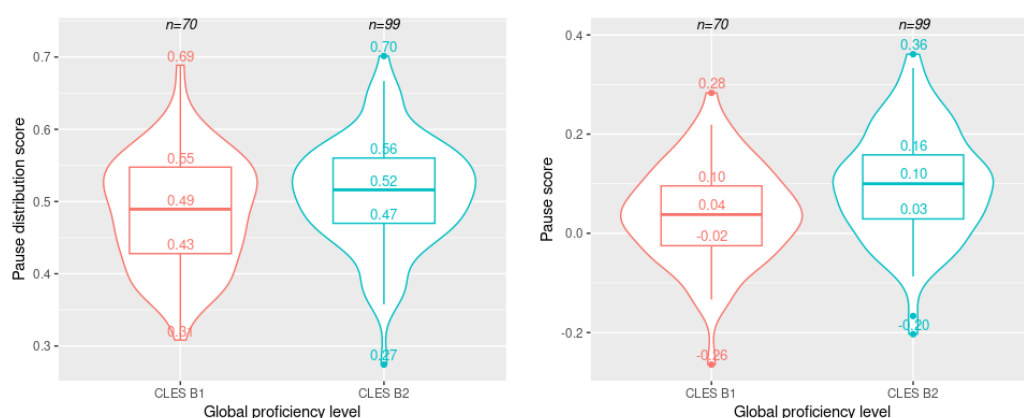


FIG. 8.4 : Proportion de pauses par type de frontière pour 100 pauses ($P_{p,i \in \{BC,WC,BP,WP\}}$)



(a) Score basé sur le niveau des constituants (proposition ou syntagme, DSP_i)

(b) Score basé sur le nombre de constituants qui s'ouvrent ou se ferment (DSP_n)

FIG. 8.5 : Scores de distribution syntaxique des pauses par locuteur (corpus CLES-FR, 180 ms-2 s)

Score de distribution syntaxique des pauses

Le score de distribution syntaxique des pauses (DSP_i) est calculé à partir des pauses inter-propositionnelles, inter-syntagmatiques et intra-syntagmatiques. Ce score est obtenu en normalisant le nombre de pauses de chaque type par le total des pauses par locuteur, puis en effectuant une somme pondérée. Les pauses intra-syntagmatiques sont pénalisées (-1), tandis que celles inter-syntagmatiques et inter-propositionnelles sont respectivement pondérées à +0,5 et +1. Ce calcul reflète la tendance des locuteurs à placer leurs pauses en frontières syntaxiques de haut niveau : plus le score est élevé, plus les pauses respectent ces frontières.

Les résultats montrent que les locuteurs B1 obtiennent en moyenne un score DSP_i plus faible que les locuteurs B2. Cependant, la différence entre les deux groupes reste limitée ($p < 0,05$, médianes : 0,49 pour B1 et 0,52 pour B2, $\Delta = -0,198$ (faible), $IC = [-0,365; -0,019]$, cf. figure 8.5a).

Une variante de ce score peut également être calculée en fonction du nombre de constituants syntaxiques qui se ferment ou s'ouvrent au moment où survient une pause, plutôt que de s'appuyer uniquement sur le niveau des constituants (propositions ou syntagmes). Ce score DSP_n permet de tenir compte de l'imbrication des groupes syntaxiques et offre davantage de flexibilité dans les paramètres de calcul. Avec cette approche, la différence entre les locuteurs B1 et B2 devient plus marquée ($p < 0,001$, médianes : 0,04 pour B1 et 0,10 pour B2, $\Delta = -0,301$ (faible), $IC = [-0,455; -0,13]$, cf. figure 8.5b).

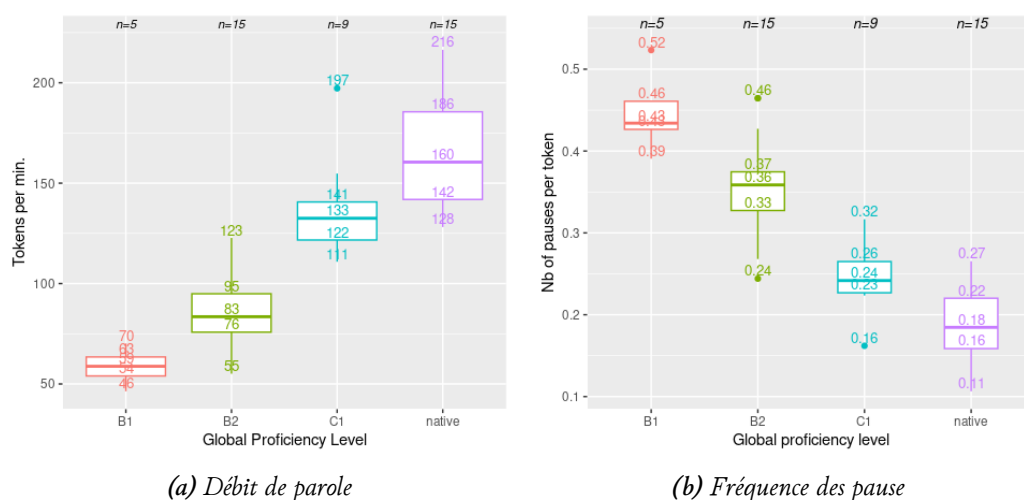


FIG. 8.6 : Débit de parole (gauche) et fréquence des pauses (droite) par locuteur dans les corpus CLES-JP et CLES-EN (180 ms-2 s)

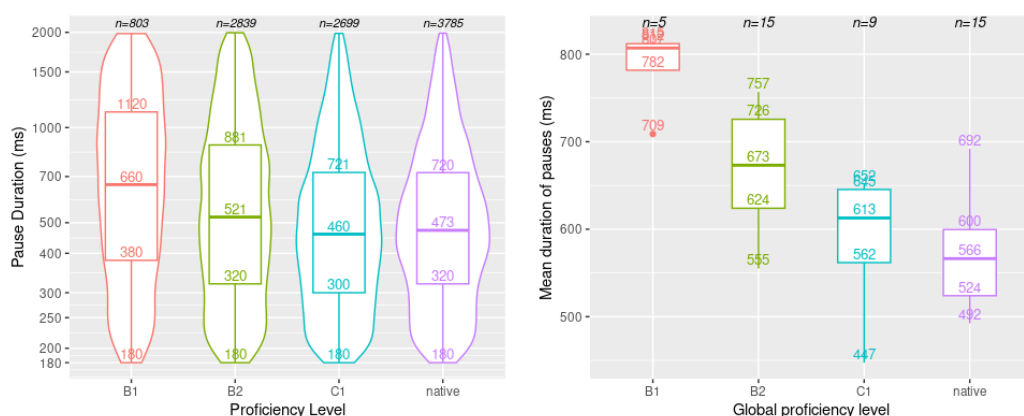
1.2 Corpus CLES-JP et CLES-EN

Pour les corpus CLES-JP et CLES-EN, respectivement 21 631 et 20 486 intervalles ont été analysés. Les différences de durée des pauses entre locuteurs B1 et B2 sont significatives ($p < 0,001$) pour tous les seuils étudiés (durée minimale : 180 ou 250 ms, maximale : 2 ou 5 s), bien que les différences soient faibles (le Δ de Cliff varie de 0,145 pour 250 ms-2 s à 0,171 pour 180 ms-5 s). Avec un seuil de 180 ms à 2 s, le corpus contient 6 341 pauses pour les locuteurs japonophones (803 pour les 5 B1, 2 839 pour les 15 B2, 2 699 pour les 9 C1) et 3 785 pauses pour les 15 locuteurs natifs.

Durées et fréquences des pauses

Comme pour le corpus CLES-FR, le débit de parole augmente avec le niveau du locuteur (cf. figure 8.6a), entraînant une quantité de parole, et donc de pauses observées, plus importante. En revanche, le nombre de pauses par mot suit une tendance inverse, fortement contrastée entre niveaux : les B1 réalisent en moyenne 43 pauses pour 100 mots, les B2 36, les C1 24, et les natifs 18 (cf. figure 8.6b).

La figure 8.7 montre la distribution des durées de pauses selon les niveaux. La durée moyenne des pauses par locuteur est fortement contrastée entre les niveaux : B1 : médiane à 807 ms, B2 : médiane à 673 ms, C1 : médiane à 613 ms, et locuteurs natifs : médiane à 566 ms.



(a) Distribution des durées de pauses

(b) Durée moyenne des pauses par locuteur

FIG. 8.7 : Durées des pauses dans les corpus CLES-JP et CLES-EN (180 ms-2 s)

Les différences sont significatives entre B1 et B2 ($p < 0,01$, $\Delta = 0,867$ (élevé), $IC = [0,338; 0,98]$), et B1/B2 et natifs ($p < 0,001$, $\Delta = 0,82$ (élevé), $IC = [0,54; 0,937]$).

Distribution syntaxique des pauses

La distribution syntaxique des pauses montre des tendances similaires à celles observées pour les locuteurs francophones, mais avec des contrastes plus marqués :

- **Frontières de propositions** : les B1 réalisent significativement plus de pauses que les B2 ($p < 0,01$, $F_{p,BC}$ médianes à 58 % pour B1 et 46 % pour B2, $\Delta = 0,840$ (élevé), $IC = [0,459; 0,96]$).
- **Frontières de syntagmes** : différence significative également ($p < 0,05$, $F_{p,BP}$ médianes à 43 % pour les locuteurs B1 et 33 % pour B2, $\Delta = 0,657$ (élevé), $IC = [0,115; 0,876]$).
- **Intra-syntagmes** : taille d'effet importante mais pas de différence significative, probablement en raison du faible nombre de données pour les B1 (ns , $F_{p,WP}$ médianes à 30 % pour B1 et 23 % pour B2, $\Delta = 0,52$ (élevé), $IC = [-0,037; 0,83]$).

Pour les proportions de pauses par type, aucune différence significative n'est observée entre les groupes de niveau des locuteurs du corpus CLES-JP. Cependant, la comparaison des locuteurs B1+B2 avec les locuteurs natifs du corpus CLES-EN révèle :

- Davantage de pauses en frontières de propositions chez les natifs ($p < 0,05$, $P_{p,BC}$ médianes à 35 % pour L2 et 37 % pour L1, $\Delta = -0,4$ (moyen), $IC = [-0,687; -0,004]$),
- Moins de pauses à l'intérieur des syntagmes ($p < 0,05$, $P_{p,WP}$ médianes à 13 % pour L2 et 10 % pour L1, $\Delta = 0,483$ (élevé), $IC = [0,087; 0,747]$).

Score de distribution syntaxique des pauses

Le score de distribution syntaxique des pauses (DSP) révèle une différence non significative entre B1 et B2, quelle que soit la méthode de calcul (niveaux de constituants ou profondeur des frontières). Cependant, la différence entre locuteurs natifs (L1) et non natifs (L2) est significative :

- **Niveaux de constituants (DSP_i)** : $p < 0,01$, médianes à 0,48 pour L2 et 0,54 pour L1, $\Delta = -0,527$ (élevé), $IC = [-0,777; -0,131]$.
- **Profondeur des frontières syntaxiques (DSP_n)** : $p < 0,001$, médianes à 0,06 pour L2 et 0,18 pour L1, $\Delta = -0,707$ (élevé), $IC = [-0,891; -0,32]$ (cf. figure 8.8).

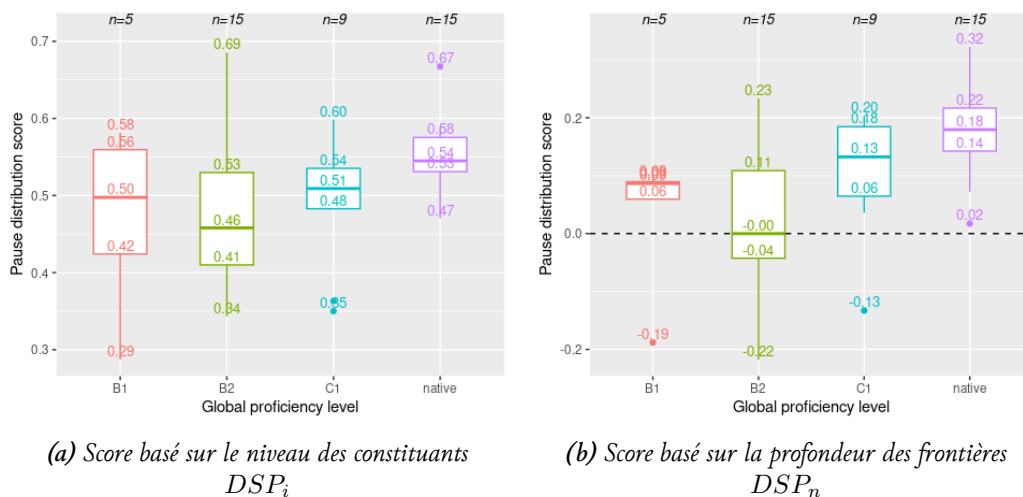


FIG. 8.8 : Score de distribution syntaxique des pauses par locuteur, basé sur le niveau des constituants ou le niveau de profondeur des frontières syntaxiques. Corpus CLES-JP et CLES-EN, 180 ms-2 s

Ces résultats montrent que les pauses sont plus souvent situées en frontières syntaxiques de haut niveau chez les locuteurs natifs du corpus CLES-EN que chez les locuteurs non-natifs japonophones du corpus CLES-JP.

8.2 Accentuation lexicale

Dans le chapitre 3, nous avons montré que la précision de la position de l'accent lexical est souvent corrélée avec le jugement de compréhensibilité des locuteurs. Cependant, la majorité des études mentionnées s'appuient sur des énoncés en parole lue ou sur des annotations manuelles de la position de l'accent. Nous avons cherché ici à déterminer si des mesures automatiques peuvent caractériser les schémas accentuels des locuteurs en parole spontanée, et si une différence significative entre les niveaux B1 et B2 est observable. Par ailleurs, nous avons examiné la qualité de l'accentuation en termes de contraste prosodique entre les syllabes au niveau de la f_0 , de l'intensité et de la durée.

2.1 Corpus CLES-FR

Données analysées

Les segments de parole du corpus CLES-FR analysés par PLSPP comprennent un total de 68 515 tokens, correspondant peu ou prou au nombre de mots. Le nombre de tokens par locuteur est légèrement inférieur pour les B1 par rapport aux B2 (médianes à 376 contre 422), mais cette différence n'est pas significative. Les mesures d'accentuation syllabique réalisées ici portent exclusivement sur les mots polysyllabiques lexicaux (noms communs, verbes, adjectifs et adverbes), bien que tous n'aient pas nécessairement été annotés par PLSPP. On compte un total de 14 873 mots polysyllabiques lexicaux, significativement plus nombreux chez les locuteurs B2². Cependant, parmi ces mots, seulement 6 468 ont été annotés par PLSPP, soit 43 % des mots ciblés initialement. Cette proportion limitée vient du fait que seuls les mots pour lesquels le nombre de pics d'intensité détectés correspond au nombre de syllabes attendues ont été conservés. Nous reviendrons sur ce constat dans le chapitre 10. Toutes les analyses de cette section portent sur ces 6 468 mots, désignés par la suite comme « mots annotés ».

Nombre de mots annotés par locuteur Le nombre absolu de mots annotés par locuteur est significativement plus élevé pour les B2 ($p < 0,01$, médianes à 32 pour B1 contre 41 pour B2). Cependant, cette différence n'est pas significative lorsqu'on rapporte le nombre de mots annotés au nombre total de tokens (médianes à 9 et 10%) ou au nombre de mots polysyllabiques lexicaux (médianes à 42 et 43%). Cela indique

² $p < 0,05$, médianes à 75 pour B1 et 94 pour B2, $\Delta = -0,232$ (faible), $IC = [-0.396; -0.052]$; rapporté au nombre de tokens par locuteur : $p < 0,01$, médianes à 21 % et 23 %, $\Delta = -0,238$ (faible), $IC = [-0.397; -0.065]$.

Position	B1		B2		all	
	Théorique	Observée	Théorique	Observée	Théorique	Observée
Initiale	74 % (1636)	23 % (502)	74 % (2796)	28 % (1059)	74 % (4432)	26 % (1561)
Médiale	14 % (301)	6 % (131)	13 % (490)	7 % (264)	13 % (791)	7 % (395)
Finale	12 % (274)	71 % (1578)	13 % (505)	65 % (2468)	13 % (779)	67 % (4046)

TAB. 8.2 : Position théorique et observée de l'accent lexical dans les mots de 2 à 3 syllabes annotés par PLSPP (corpus CLES-FR, $n = 6\ 002$)

que les locuteurs B2 n'utilisent pas proportionnellement plus de mots polysyllabiques que les B1, et que les mots produits ne sont pas mieux reconnus par PLSPP pour les B2 que pour les B1.

Caractéristiques des mots annotés Les mots annotés sont principalement des noms communs (57% des occurrences, doublons inclus), suivis des verbes (19%), des adjectifs (12%) et des adverbes (12%). La majorité de ces mots est composée de deux syllabes (73%), mais on trouve également des mots de trois syllabes (21%), quatre syllabes (5%) et cinq ou six syllabes (moins de 1%). Pour limiter l'influence potentielle de l'accent secondaire et étant donné que les mots de plus de trois syllabes représentent moins de 6% des mots annotés, nous concentrons nos analyses sur les mots de deux à trois syllabes ($n = 6\ 002$).

Parmi ces mots, l'accent primaire est attendu en initiale dans 74 % des cas (4 432 mots), en médiale dans 13 % (791 mots) et en finale dans 13 % (779 mots) (cf. tableau 8.2). Pour les mots à deux syllabes, 84 % d'entre eux sont accentués en initiale contre 16 % en finale. Pour les mots à trois syllabes, la majorité est accentuée sur la syllabe médiale (58%), suivie de l'initiale (38%) et de la finale (4%).

Patterns accentuels observés

Examinons à présent les patterns accentuels produits par les locuteurs. Nous utiliserons désormais le terme « syllabe proéminente » pour désigner la syllabe identifiée par PLSPP comme étant acoustiquement proéminente au sein du mot annoté, et donc perçue, en théorie, comme syllabe accentuée par l'auditeur. Le terme « accent théorique » fera quant à lui référence à la position de l'accent prescrit par le dictionnaire de référence, ici le *CMU Pronouncing Dictionary* (version 0.7b).

Les résultats montrent que 67 % des syllabes proéminentes se situent en finale, contre 26 % en initiale et seulement 7 % en médiale (cf. tableau 8.2). De manière générale, la proportion de mots dont la syllabe proéminente correspond à la position de l'accent primaire théorique est relativement faible : 36 % sur l'ensemble du corpus,

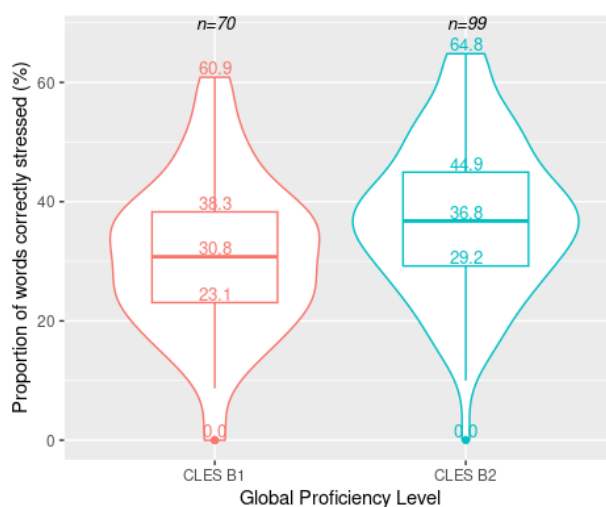


FIG. 8.9 : Score de position de l'accent par locuteur, sur 6 468 mots lexicaux de 2 à 3 syllabes (corpus CLES-FR)

avec 32 % pour les mots produits par les locuteurs B1 et 38 % pour ceux produits par les B2.

Cependant, étant donné que le nombre de mots annotés varie entre les locuteurs, il est plus pertinent d'examiner la proportion de mots correctement accentués pour chaque locuteur, que nous appellerons « score de position de l'accent » (S). Ces scores individuels présentent une grande variabilité, allant de 0 à 64,8 % (médiane à 34,6 %).

La figure 8.9 montre la distribution des scores par locuteur selon leur niveau. Bien que les scores des locuteurs B1 et B2 se chevauchent largement, une différence significative est observée entre les deux groupes ($p < 0,01$, médianes respectives à 30,8 % pour les B1 et 36,8 % pour les B2). Toutefois, cette différence est accompagnée d'une taille d'effet limitée ($\Delta = -0,275$ (faible), $CI = [-0,432; -0,102]$).

Lorsqu'on examine le score obtenu en fonction de la position de l'accent théorique, les résultats montrent que les locuteurs B2 obtiennent des scores plus élevés pour les mots à accent initial ($p < 0,01$, médianes à 25 pour les B1 et 32 pour les B2, $\Delta = -0,269$ (faible), $CI = [-0,428; -0,094]$) et médial (*ns.*, médianes à 30 pour les B1 et 36 pour les B2, $\Delta = -0,158$ (faible), $CI = [-0,329; -0,024]$). En revanche, aucune différence significative n'est observée pour les mots à accent final (*ns.*, médianes à 77 ; $\Delta = -0,032$ (négligeable), $CI = [-0,148; 0,210]$). Ces observations suggèrent que la différence entre les deux groupes de locuteurs réside principalement dans le taux d'accentuation en initiale et en médiale, tandis que la tendance à accentuer la syllabe finale est observée sur les deux niveaux.

Rang	Mot	Gabarit accentuel théorique	Position accent théorique	Catégorie grammaticale	Fréquence	Détection gabarit attendu (%)
1	students	Oo	initial	NOUN	187	24.60
2	maybe	Oo	initial	ADV	186	41.40
3	people	Oo	initial	NOUN	173	20.81
4	computer	oOo	medial	NOUN	155	30.32
5	testing	Oo	initial	VERB	111	26.13
6	also	Oo	initial	ADV	100	28.00
7	really	Oo	initial	ADV	96	33.33
8	computers	oOo	medial	NOUN	95	26.32
9	problem	Oo	initial	NOUN	92	39.13
10	teacher	Oo	initial	NOUN	89	19.10
11	children	Oo	initial	NOUN	87	27.59
12	teachers	Oo	initial	NOUN	75	10.67
13	cameras	Ooo/Oo	initial	NOUN	73	23.29
14	student	Oo	initial	NOUN	59	25.42
15	money	Oo	initial	NOUN	57	21.05
16	very	Oo	initial	ADV	56	44.64
17	paper	Oo	initial	NOUN	53	30.19
18	agree	oO	final	VERB	52	50.00

TAB. 8.3 : Liste des mots annotés de plus de 50 occurrences dans le corpus CLES-FR (“O” représente la syllabe censée porter l’accent primaire)

Enfin, il est intéressant de noter que les locuteurs B1 présentent des scores plus variés que les locuteurs B2, malgré un écart type identique de 12,8 %.

Le tableau 8.3 présente les 18 mots les plus fréquents parmi les mots annotés (plus de 50 occurrences). On observe que les mots les plus fréquents ne sont pas nécessairement ceux dont l’accentuation est la mieux maîtrisée. Par exemple, “students” n’est accentué sur la syllabe initiale que dans 24,6 % des cas (sur 187 occurrences), “people” dans 20,8 % des cas (173 occurrences), “teacher” dans 19 % des cas (89 occurrences), et “teachers” dans seulement 10,7 % des cas (75 occurrences).

Contraste prosodique

Dans cette section, nous examinons le degré de contraste prosodique entre les syllabes d’un mot. Il ne s’agit pas uniquement d’évaluer si la syllabe proéminente correspond à celle qui porte l’accent lexical d’après le dictionnaire de référence, mais de mesurer à quel point cette syllabe se démarque des autres sur le plan prosodique. Pour cela, nous calculons le contraste prosodique C , obtenu par la différence entre la valeur acoustique normalisée P_s de la syllabe censée être accentuée et la moyenne \bar{P}_u des

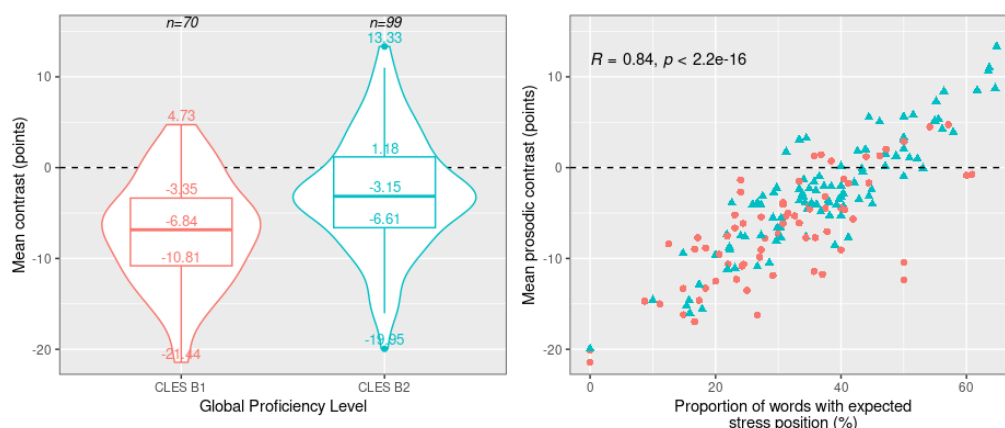


FIG. 8.10 : Contraste prosodique moyen \bar{C} dans le corpus CLES-FR, à gauche en fonction du niveau du locuteur ; à droite en fonction du score de position de l'accent S

autres syllabes du mot. La valeur obtenue est comprise entre -100 et 100, où une valeur positive indique que la syllabe préminente correspond à la position de l'accent primaire. Plus cette valeur est élevée, plus le contraste prosodique avec les autres syllabes est marqué. Nous nommerons \bar{C} la moyenne des contrastes calculés sur l'ensemble des mots annotés pour un locuteur. De plus, le contraste mesuré spécifiquement sur la fréquence fondamentale sera noté C_{f_0} , celui sur l'intensité C_{int} , et celui sur la durée C_{dur} . C est quant à lui obtenu par moyenne des trois dimensions prosodiques pour chaque syllabe.

Le contraste moyen par locuteur varie entre -21,44 et 13,33 dans le corpus CLES-FR, avec une médiane à -4,11. Cela indique une tendance générale des locuteurs à ne pas accentuer la syllabe censée porter l'accent primaire. Bien que les scores des locuteurs B1 et B2 se chevauchent largement, la différence entre les deux groupes est significative et plus prononcée qu'avec le score de position S décrit dans la section précédente ($p < 0,001$, \bar{C} médianes à -6,84 pour les locuteurs B1 et -3,15 pour les B2, $\Delta = -0,389$ (moyen) $CI = [-0,534; -0,222]$, cf. figure 8.10 gauche). De plus, il existe une corrélation forte entre le score de position S et le contraste moyen \bar{C} ($R = 0,84$, $p < 0,001$, cf. figure 8.10 droite), montrant que les locuteurs ayant une meilleure maîtrise de la position accentuelle produisent également un contraste prosodique plus marqué³.

Différences entre les dimensions prosodiques Pour identifier la dimension prosodique différenciant le mieux les locuteurs B1 et B2, nous avons examiné le contraste moyen séparément pour chaque dimension (f_0 , intensité, durée). Les résultats (fi-

³Constat toutefois sans surprise, les deux variables étant dépendantes.

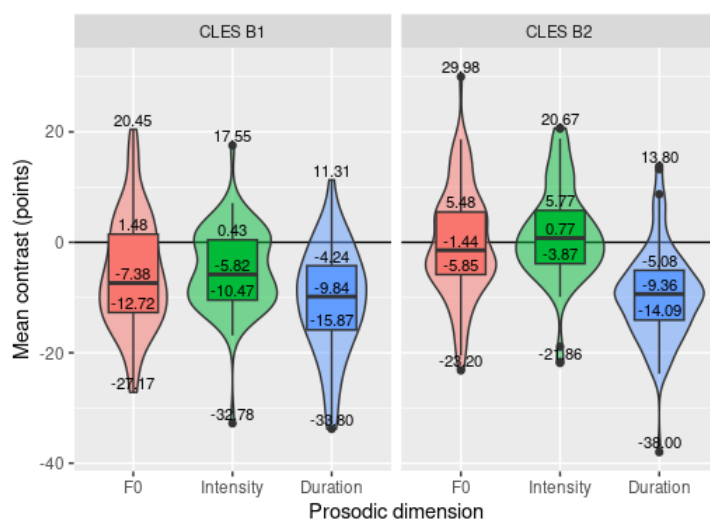


FIG. 8.11 : Contraste moyen par dimension prosodique par locuteur (corpus CLES-FR)

gure 8.11) montrent que la plus grande différence se situe au niveau de l'intensité ($p < 0,001$, \overline{C}_{int} médianes à $-5,82$ pour les B1 et $0,77$ pour les B2, $\Delta = -0,450$ (moyen), $CI = [-0,590; -0,283]$). Vient ensuite la fréquence fondamentale, avec une différence moins marquée mais significative ($p < 0,001$, \overline{C}_{f_0} médianes à $-7,38$ et $-1,44$, $\Delta = -0,317$ (faible), $CI = [-0,475; -0,138]$). En revanche, la durée des syllabes ne présente pas de différence significative entre les deux niveaux (*ns.*, médianes à $-9,84$ et $-9,36$, $\Delta = -0,018$ (négligeable) $CI = [-0,196; 0,162]$).

Profils d'accentuation des locuteurs La figure 8.12 illustre les profils d'accentuation de deux locuteurs représentatifs, SpeakerA et SpeakerB. La figure indique le contraste prosodique moyen du locuteur pour chaque dimension (\overline{C} , \overline{C}_{f_0} , \overline{C}_{int} et \overline{C}_{dur}). Les cercles représentent les valeurs prosodiques normalisées (P_s pour la syllabe censée être accentuée et \overline{P}_u pour la moyenne des autres syllabes). La taille des cercles et les valeurs inscrites reflètent les centiles des valeurs prosodiques.

- **SpeakerA**, avec un score de position de $S = 64,8\%$ et un contraste moyen de $\overline{C} = 13,3$, est représentatif d'un locuteur de haut niveau en termes d'accentuation. On peut voir que la syllabe accentuée P_s se démarque clairement en termes de f_0 (contraste de 30) et d'intensité (17), mais pas en durée (-5).
- **SpeakerB**, en revanche, présente un score de position de seulement $S = 14,8\%$ et un contraste moyen de $\overline{C} = -9,4$. La syllabe censée être accentuée P_s a tendance à être moins marquée que les autres, particulièrement au niveau de la durée (-21) et de la f_0 (-11), tandis que l'intensité reste peu mobilisée (+2).

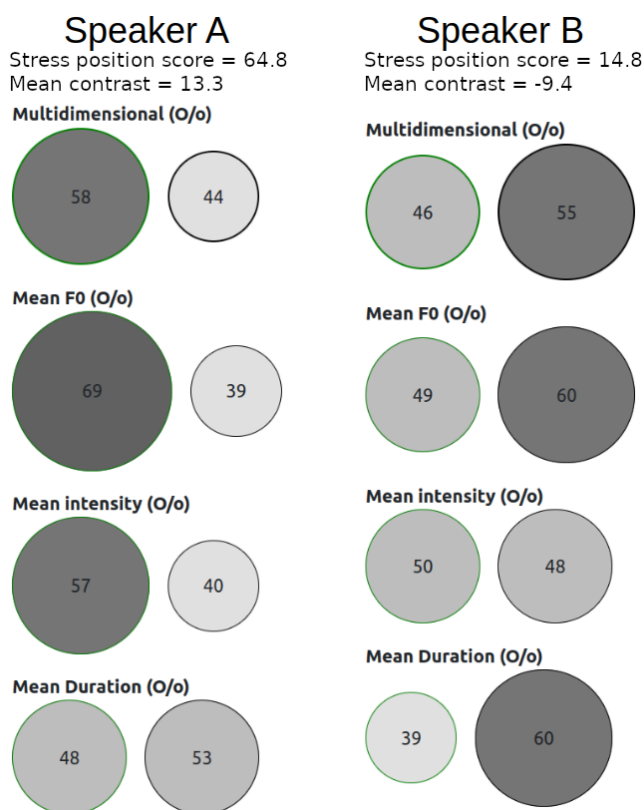


FIG. 8.12 : Contrastes prosodiques de deux locuteurs au profil différent
Écouter un extrait : *Speaker A* et *Speaker B*.

Ces profils montrent des tendances générales : les locuteurs avec un score de position et un contraste élevés mobilisent davantage la f_0 et l'intensité, tandis que ceux avec des scores faibles sont principalement influencés par un contraste négatif de durée et de f_0 , sans exploitation notable de l'intensité.

2.2 Corpus CLES-JP et CLES-EN

Les corpus CLES-JP et CLES-EN comptent respectivement 275 et 113 segments de parole (21 356 et 20 358 tokens). La proportion de mots polysyllabiques lexicaux par token est similaire dans les deux corpus (médianes : 23 % pour les locuteurs B1/B2 du CLES-JP et 22 % pour les locuteurs natifs). Cependant, la proportion de mots annotés par PLSPP est significativement plus élevée chez les locuteurs japonophones que chez les natifs ($p < 0,001$, médianes : 41 % pour les locuteurs B1/B2 et 30 % pour les natifs, $\Delta = 0,753$ (élevé) $IC = [0,410; 0,910]$). Ce résultat inattendu suggère une meilleure reconnaissance des mots pour les locuteurs non natifs. Au total, 1 913 mots

	CLES-JP	CLES-EN		CLES-JP	CLES-EN
Noms	53 % (945)	43 % (567)	2 syll.	82 % (1483)	83 % (1094)
Verbes	24 % (436)	26 % (342)	3 syll.	14 % (247)	14 % (190)
Adjectifs	15 % (264)	16 % (211)	4 syll.	3 % (60)	2 % (30)
Adverbes	9 % (154)	15 % (196)	5+ syll.	0,5 % (9)	0,2 % (8)

TAB. 8.4 : Catégories grammaticales et nombres de syllabes des mots annotés par PLSP dans les corpus CLES-JP ($n=1913$) et CLES-EN ($n=1354$)

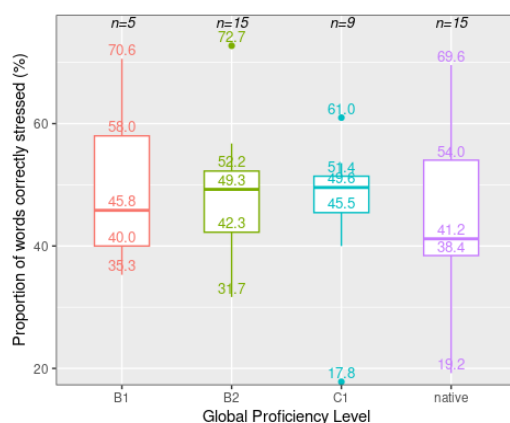


FIG. 8.13 : Scores de position par locuteur, sur 3 014 mots lexicaux de 2 à 3 syllabes (corpus CLES-JP et CLES-EN)

ont été annotés dans le CLES-JP et 1 354 dans le CLES-EN. Les distributions des catégories grammaticales et du nombre de syllabes restent similaires entre les deux corpus, comme l'illustre le tableau 8.4. À l'instar du corpus CLES-FR, l'analyse se focalise sur les mots de 2 à 3 syllabes, représentant environ 96 % des mots annotés.

La figure 8.13 montre les scores de position de l'accent pour chaque locuteur. Deux observations majeures émergent. Premièrement, aucune amélioration significative n'est observée entre les niveaux des locuteurs japonophones : les locuteurs B1 obtiennent des scores variant entre 35,3 % et 70,6 % (médiane à 45,8 %), tandis que les locuteurs B2 atteignent des scores compris entre 31,7 % et 72,7 % (médiane à 49,3 %, différence non significative). Les 9 locuteurs C1 n'affichent pas de résultats sensiblement supérieurs. Deuxièmement, les scores des locuteurs natifs présentent une grande variabilité (19,2 % à 69,6 %) et, en moyenne, restent inférieurs à ceux des locuteurs japonophones (médianes à 49,3 % pour le CLES-JP et 41,2 % pour le CLES-EN, différence non significative, $\Delta = 0,170$ (faible), $IC = [-0,223; 0,516]$). Ces résultats suggèrent que la détection de la syllabe proéminente est influencée par un facteur indépendant du niveau de compétence en langue.

Position	CLES-JP		CLES-EN	
	Théorique	Observée	Théorique	Observée
Initiale	83 % (1430)	40 % (689)	83 % (1071)	38 % (485)
Médiale	7 % (120)	4 % (76)	6 % (79)	6 % (71)
Finale	10 % (180)	56 % (965)	10 % (134)	57 % (728)

TAB. 8.5 : Position théorique et observée de l'accent lexical dans les mots des corpus CLES-JP ($n=1\ 913$) et CLES-EN ($n=1\ 354$)

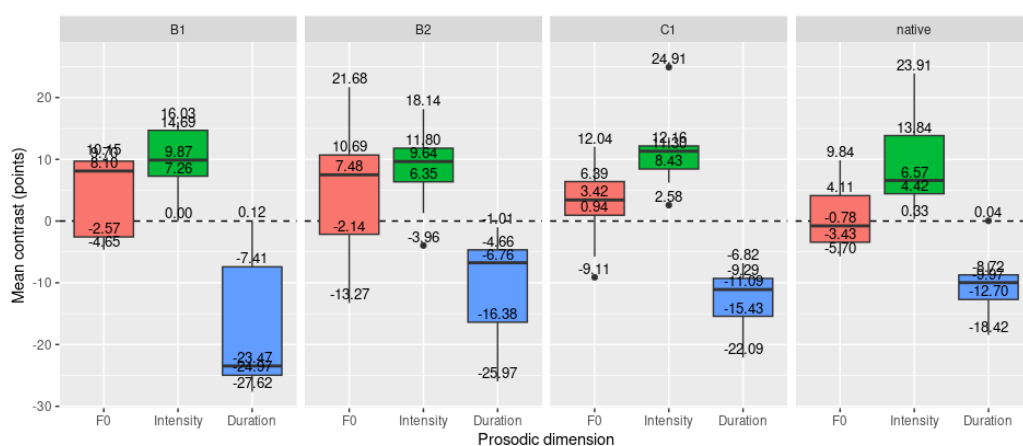


FIG. 8.14 : Contraste moyen par locuteur entre P_s et $\overline{P_u}$ (corps CLES-JP, CLES-EN)

Le tableau 8.5 compare le nombre de mots par position d'accent théorique (attendu) et observée (proéminence détectée par PLSP). Alors que l'accent est attendu en initiale dans 80 % des cas, en médiale dans 7 %, et en finale dans 10 %, PLSP détecte une proéminence en initiale dans seulement 40 % des mots, en médiale dans 5 %, et en finale dans près de 60 %. Cette tendance est similaire pour les locuteurs japonophones et natifs, indiquant que la syllabe finale est souvent considérée comme proéminente, même lorsqu'elle ne devrait pas l'être a priori, et ce indépendamment de la langue maternelle du locuteur.

L'évaluation du contraste prosodique moyen \overline{C} indique que la syllabe censée être accentuée (P_s) est systématiquement plus élevée en intensité, mais plus courte en durée que la moyenne des autres syllabes (cf. figure 8.14). Le contraste de f_0 tend à être positif, mais avec une amplitude moindre. Ces résultats reflètent des tendances similaires à celles observées dans le corpus CLES-FR (cf. figure 8.10), bien que les valeurs soient ici plus extrêmes : les contrastes d'intensité et de hauteur sont généralement marqués et alignés avec la position d'accent théorique, tandis que le contraste de durée est fortement négatif. Ce contraste négatif suggère que la dernière syllabe des mots, souvent allongée, influence la détection de la proéminence, au point de rendre le contraste global négatif.

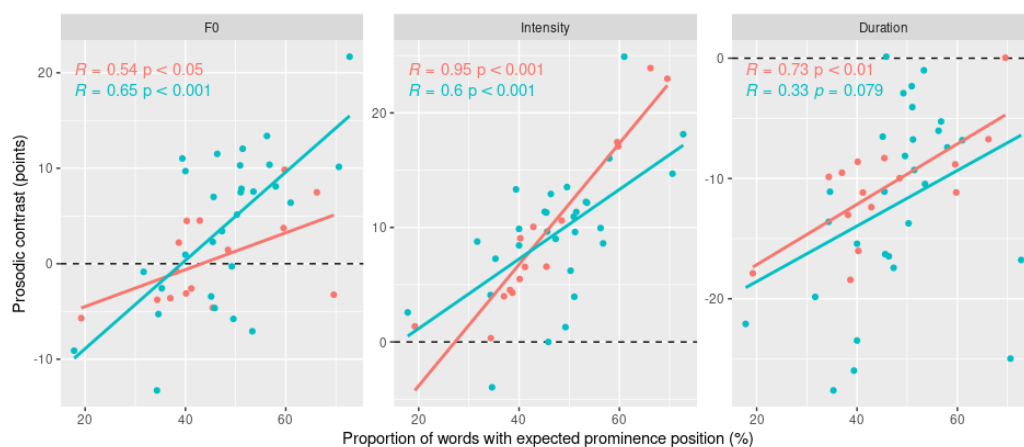


FIG. 8.15 : Corrélation entre le contraste moyen par dimension prosodique et le score de position de l'accent par locuteur (corpus CLES-JP en bleu et CLES-EN en rouge)

En examinant les corrélations entre le contraste prosodique par dimension et le score de position d'accent, des divergences notables apparaissent entre les locuteurs japonophones et natifs. Chez les locuteurs natifs, le score de position est fortement corrélé au contraste d'intensité ($R = 0,95$, $p < 0,001$) et beaucoup moins à celui de la f_0 ($R = 0,54$, $p < 0,05$), tandis que chez les locuteurs japonophones, c'est d'abord le contraste de f_0 qui prime ($R = 0,65$, $p < 0,001$), suivi de près par celui d'intensité ($R = 0,60$, $p < 0,001$). Enfin, le contraste de durée n'est significatif que pour les locuteurs natifs ($R = 0,73$, $p < 0,01$) et non pour les locuteurs japonophones ($R = 0,33$, $p = 0,08$).

Ces résultats suggèrent que l'intensité est le paramètre prosodique le plus fiable mesuré par PLSPP. Les locuteurs natifs avec un contraste d'intensité élevé tendent à obtenir un meilleur score global d'accentuation. Les locuteurs japonophones, quant à eux, semblent s'appuyer davantage sur le contraste de f_0 . Quant à la durée, elle semble influencer les scores des natifs, bien que négativement, indiquant que les syllabes accentuées plus longues contribuent à un meilleur score d'accentuation. Chez les locuteurs japonais, cette dimension paraît plus erratique, probablement en raison d'hésitations plus fréquentes, résultant en allongements de syllabes finales.

Conclusion

Nous avons montré dans ce chapitre que, bien que les locuteurs de niveau B1 produisent plus de pauses que ceux de niveau B2 à tous les niveaux syntaxiques (inter-proposition, inter-syntagme et intra-syntagme), ils ont tendance à produire propor-

tionnellement davantage de pauses intra-syntagmes, mais pas significativement plus de pauses inter-propositionnelles. Une tendance similaire est observée chez les locuteurs japonophones par rapport aux locuteurs natifs. Autrement dit, c'est bien au niveau de la fréquence de pauses aux frontières syntaxiques de bas niveau que s'observe la principale différence entre les locuteurs B1 et B2.

Nous avons proposé un score de distribution syntaxique des pauses qui permet de mesurer la tendance de positionnement des pauses dans l'énoncé d'un locuteur. Qu'il soit basé sur les types de constituants ou le niveau de profondeur des frontières syntaxiques, ce score confirme que plus le niveau de compétence du locuteur est élevé, plus les pauses sont positionnées aux frontières syntaxiques de haut niveau, avec les locuteurs natifs atteignant les scores les plus élevés.

L'analyse des annotations automatiques de proéminence syllabique révèle que les locuteurs de niveau B2 positionnent en général mieux l'accent et produisent un contraste acoustique plus marqué entre la syllabe accentuée et les autres syllabes du mot. Toutefois, les mesures individuelles présentent une grande variabilité : le score de position de l'accent s'étend de 0 à 65 % dans le corpus CLES-FR, avec un chevauchement important entre les locuteurs B1 et B2 malgré une différence significative. Les annotations mettent également en évidence l'influence notable des schémas accentuels de la langue maternelle des locuteurs. Les francophones tendent souvent à augmenter la f_0 et à allonger la syllabe finale des mots, tandis que l'intensité reste stable, quelle que soit la position attendue de l'accent. Cette tendance diminue à mesure que le niveau du locuteur augmente, avec une meilleure maîtrise de la f_0 et de l'intensité chez les locuteurs B2.

Chez les locuteurs japonophones, la tendance à accentuer la syllabe finale est beaucoup moins marquée. Ces locuteurs obtiennent globalement de meilleurs scores de position de l'accent (variant de 17,8 à 72,7 %) et mobilisent fortement la f_0 et l'intensité dès le niveau B1. Toutefois, le faible nombre de locuteurs dans ce corpus n'a pas permis de détecter des différences significatives entre les niveaux. Enfin, l'analyse des locuteurs natifs a révélé une forte tendance à l'allongement de la syllabe finale et un contraste de f_0 limité, entraînant des scores de position globalement faibles (entre 19,2 et 69,7 %) et en moyenne inférieurs à ceux des locuteurs japonophones. En revanche, le contraste d'intensité est plus marqué chez les locuteurs natifs et très corrélé avec le score de position de l'accent.

Nous reviendrons sur les limites de l'outil de mesure dans le chapitre 10.

Metric	Definition	CLES-FR				CLES-JP					CLES-EN		
		B1	B2	B1/B2 (p)	Cliff Δ	B1	B2	C1	B1/B2 (p)	Cliff Δ	Native	JPB1B2/EN (p)	Cliff Δ
N_{tokens}	Number of tokens												
SR	Speech rate (token/min)	96	107	B1<B2 ***	-0,35								
N_p	Number of pauses (180ms-2s)								B1>B2 ***			JP>EN ***	
F_p	Number of pauses per token	0,32	0,29	B1>B2 *	0,154	0,43	0,36	0,24	B1>B2 ***		0,18	JP>EN ***	
\bar{d}_p	Mean duration of pauses	598	591	B1>B2		807	673	613	B1>B2 **	0,867	566	JP>EN ***	0,82
$F_{p,BC}$	Number of BC pauses per BC boundary (%)	47	42	B1>B2 ***	0,311	58	46		B1>B2 **	0,84			
$F_{p,WC}$	Number of WC pauses per WC boundary	28	25	B1>B2	0,172								
$F_{p,BP}$	Number of BP pauses per BP boundary	29	26	B1>B2	0,149	43	33		B1>B2 *	0,657			
$F_{p,WP}$	Number of WP pauses per WC boundary	21	18	B1>B2 *	0,187	30	23		B1>B2	0,52			
$P_{p,BC}$	Number of BC pauses per number pauses	35	36	B1<B2	-0,069						37	JP<EN *	-0,4
$P_{p,WC}$	Number of WC pauses per number pauses												
$P_{p,BP}$	Number of BP pauses per number pauses	52	52	-	-0,069								
$P_{p,WP}$	Number of WP pauses per number pauses	12	11	B1>B2 *	0,216						10	JP>EN *	0,483
DSP_i	Pauses syntactic distrib. based on boundary type	0,49	0,52	B1<B2 *	-0,198	0,5	0,46	0,51	B1>B2		0,54	JP<EN **	-0,527
DSP_n	Pauses syntactic distrib. based on boundary depth	0,04	0,1	B1<B2 ***	-0,301		0	0,13	B1>B2		0,18	JP<EN ***	-0,707
N_{poly}/N_{tokens}	Number of lexical polysyllabic words per token	0,21	0,23	B1<B2 **	-0,238						0,22	JP<EN	
N_{ann}	Number of words annotated with PLSPP	32	41	B1<B2 **									
N_{ann}/N_{tokens}	Number of N_{ann} per token	0,09	0,1										
N_{ann}/N_{poly}	Number of N_{ann} per N_{poly}	0,42	0,43								0,3	JP>EN ***	0,753
S	Stress position score	30,8	36,8	B1<B2 **	-0,275	45,8	49,3	49,6	B1<B2		41,2	JP>EN	0,17
$S_{initial}$	S for words with initial stress	25	32	B1<B2 **	-0,269								
S_{medial}	S for words with medial stress	30	36	B1<B2	-0,158								
S_{final}	S for words with final stress	77	77	-	-0,032								
\bar{C}	Mean prosodic contrast	-6,84	-3,15	B1<B2 ***	-0,389								
\bar{C}_{f_0}	Mean f_0 contrast	-7,38	-1,44	B1<B2 ***	-0,317	8,1	7,48	3,42	B1>B2		-0,78	JP>EN	
\bar{C}_{int}	Mean intensity contrast	-5,82	0,77	B1<B2 ***	-0,45	9,87	9,64	11,3	B1>B2		6,57	JP>EN	
\bar{C}_{dur}	Mean duration contrast	-9,84	-9,36	B1<B2	-0,018	-24,97	-6,76	-11,09	B1<B2		-9,97		

TAB. 8.6 : Tableau de résultats d'analyses en parole spontanée.
Chaque métrique est calculée par locuteur, le tableau indique la valeur médiane par groupe de locuteurs.

À COMPLÉTER