

# Chapitre 12

## Analyses en parole spontanée

Ce chapitre présente les résultats d'analyse des annotations de pauses et d'accentuation produites par PLSPP sur les trois corpus de parole spontanée conversationnelle. Nous présenterons d'abord les analyses relatives à la distribution des pauses, puis celles concernant les patterns accentuels. Dans chacune des sections, les résultats obtenus avec les locuteurs B1 et B2 du corpus CLES-FR seront mentionnés en premier, et une sous-section finale sera dédiée à la comparaison des groupes de niveau des locuteurs japonophones (CLES-JP) et des locuteurs anglophones natifs (CLES-EN).

### 12.1 Analyse des patterns de pauses

Sur les 10 h20 min de parole continue extraite du corpus CLES-FR (70 locuteurs B1, 99 locuteurs B2), 72 140 intervalles inter-mots ont été analysés. La figure 12.1 présente la distribution de durée de ces intervalles. Leur durée médiane est de 80 ms avec le premier quartile à 40 ms et le troisième à 261 ms, toutefois on voit que la

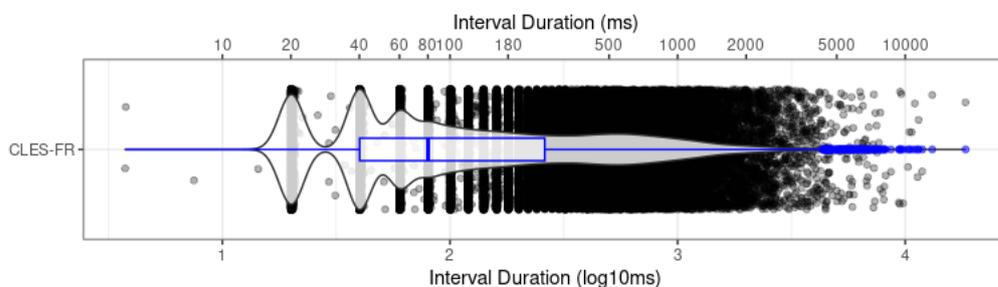


FIG. 12.1 : Distribution de durée des 72 140 intervalles inter-mots du corpus CLES-FR

distribution ne suit pas une loi normale, la grande majorité des intervalles étant de durée inférieure à 200 ms (70,3%), puis un second pic plus étalé autour de 500 ms. On peut voir que l'alignement des mots par Wav2Vec2.0 est effectué sur une fenêtre de 20 ms. La figure ne comprend pas 24 valeurs proches de 0 ( $\log_{10}$  de durée < 0). Parmi ces intervalles, on en compte 22 796 dont la durée est supérieure à 180 ms (32%), 1 085 supérieures à 2 s (1,5%), et 83 à 5 s (0,1%).

## 1.1 Durées et fréquences des pauses

Fixons maintenant un seuil minimum de durée à 180 ms pour ne considérer que les intervalles suffisamment longs pour être perceptibles par l'auditeur, et donc susceptibles d'être perçus comme des pauses. Comme indiqué dans le chapitre 7, nous fixons également un seuil maximum à 2 s pour ignorer les intervalles longs pouvant être dus à des erreurs d'alignement. Restent ainsi 21 710 intervalles (30%), que nous considérerons comme pauses dans la suite des analyses. La valeur médiane se situe maintenant à 481 ms, le premier quartile à 281 ms et le troisième à 782 ms.

La différence de durée de pause en fonction du niveau du locuteur est négligeable, bien que significative, entre les locuteurs B1 et B2 ( $p < 0,05$ , médiane à respectivement 481 ms et 474 ms, cf. figure 12.2a). Que se passe-t-il si l'on choisit des seuils de durée de pause différents ? Si l'on fixe le seuil de durée minimum à 250 ms, la différence entre B1 et B2 n'est plus significative (médianes à 581 ms). La prise en compte de pauses plus longues (jusqu'à 5 s) ne semble pas affecter beaucoup les mesures : la différence B1-B2 reste significative avec un seuil de 180 ms-5 s ( $p < 0,01$ , médianes à 501 ms), et reste non significative avec un seuil de 250 ms-5 s. D'après ces résultats, il semble pertinent de considérer les pauses courtes (inférieures à 250 ms) pour distinguer les niveaux B1 et B2, mais les pauses supérieures à 2 s ne semblent pas discriminantes. Dans tous les cas, la différence entre les deux groupes de locuteurs ne semble pas se situer au niveau de la durée des pauses, toutes pauses confondues, comme l'indique le  $\Delta$  de Cliff toujours proche de 0 (cf. tableau 12.1). Si l'on calcule

Seuils de durée	p-value	$\Delta$ 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	<i>ns</i>	0,009	581	693 – 683	391 – 380
250 ms-5 s	<i>ns</i>	0,014	602 – 601	812 – 791	631 – 610

**TAB. 12.1 :** 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  $\Delta$  de Cliff, et la médiane, la moyenne et l'écart type des deux distributions.

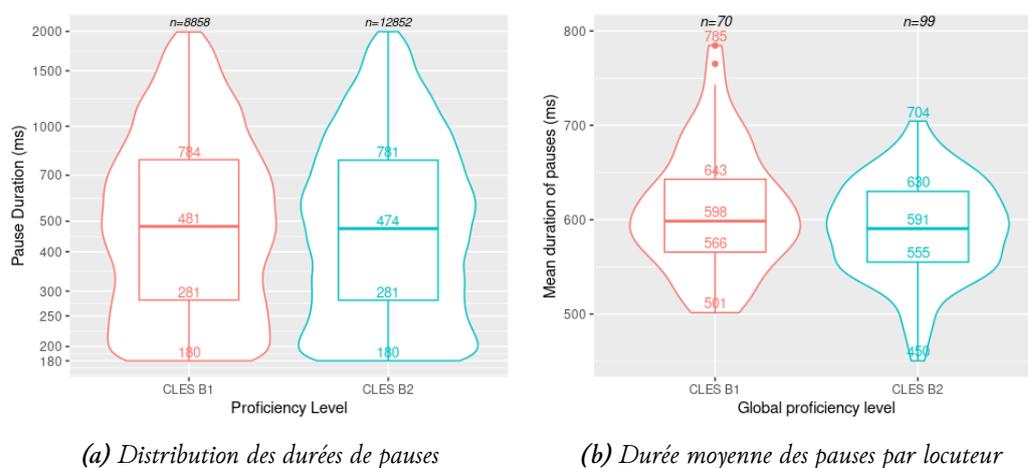


FIG. 12.2 : Durées des pauses dans le corpus CLES-FR (180 ms-2 s)

la durée moyenne des pauses par locuteur (cf. figure 12.2b), il n'y a toujours pas de différence significative entre les locuteurs B1 et B2, mais on se retrouve avec deux distributions aux formes distinctes. Celle des locuteurs B1 pointe vers le haut, indiquant que certains locuteurs font des pauses de durée moyenne particulièrement longue (notamment 6 locuteurs au-delà de 700 ms), tandis que celle des B2 pointe vers le bas, indiquant le cas inverse (5 locuteurs en dessous de 500 ms).

Le débit de parole des locuteurs B1 étant significativement plus lent que celui des B2 ( $p < 0,001$ ,  $\Delta = -0,35$  (medium), médianes respectives 96 et 107 tokens/minute, cf. figure 12.3a), il ne semble pas pertinent de tenir compte du nombre de pauses par minute par locuteur (différence non significative, médianes à 32 pour B1 et 34 pauses/minute pour B2). Le nombre de pauses par token permet quant à lui de mesurer la fréquence d'occurrence des pauses sans être influencé par la vitesse d'élocution. Pour un même nombre de mots, les locuteurs B1 font plus de pauses que les locuteurs B2 ( $p < 0,05$ ,  $\Delta = 0,154$  (small), médianes respectives à 0,32 et 0,29 pauses par token, cf. figure 12.3b).

## 1.2 Distribution syntaxique

S'il est possible de distinguer les locuteurs B1 et B2 à partir des mesures de débit de parole ou de fréquence de pauses, ces deux critères ne sont pas pour autant la cause d'une moins bonne compréhensibilité du locuteurs. Le chapitre 3 met en avant l'importance de la distribution syntaxique des pauses, et plusieurs études ont montré que la fréquence des pauses survenant à l'intérieur des groupes syntaxiques a tendance à être négativement corrélée avec la perception de fluence, tandis que celle des pauses

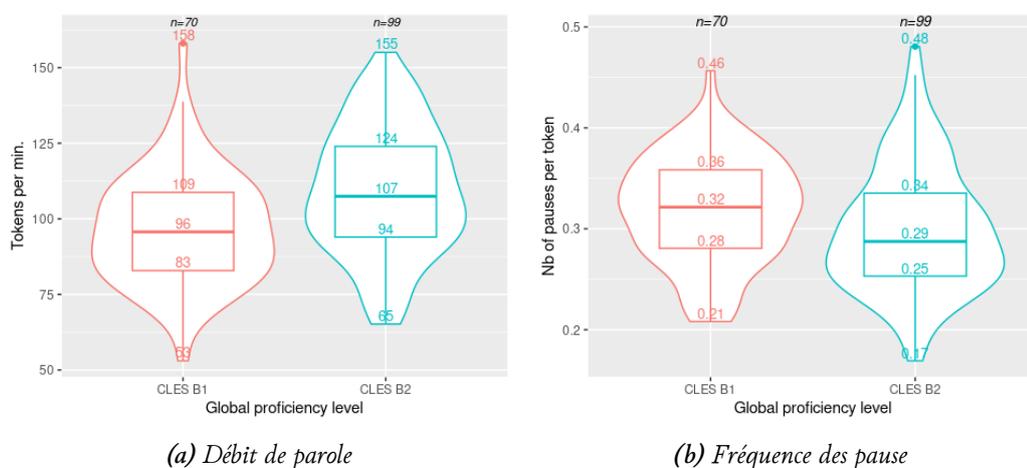


Fig. 12.3 : Débit de parole (gauche) et fréquence des pauses (droite) par locuteur dans le corpus CLES-FR (180 ms-2 s)

survenant entre les groupes semblent moins impacter le jugement de fluence. Inspirés par des études précédentes (Kahng, 2014 ; Kallio et al., 2022 ; Shea & Leonard, 2019 ; Suzuki & Kormos, 2020), nous proposons dans un premier temps de calculer la fréquence des pauses relativement à deux types de groupes syntaxiques : les propositions et les syntagmes. Une pause pourra donc être inter-propositionnelle (*between clauses*, *BC*) si elle se trouve en frontière de proposition, inter-syntagme (*between phrases*, *BP*) si elle se trouve en frontière de syntagme, ou à défaut des deux premiers, intra-syntagme (*within phrases*, *WP*). Pour une meilleure consistance avec les études précédentes, nous donnerons également les proportions de pauses intra-propositionnelles (*Within clauses*, *WC*), qui comprend les pauses inter- et intra-syntagmes.

La figure 12.4 présente la proportion de pauses par type de frontière syntaxique. Pour le même nombre de propositions, il apparaît que les locuteurs B1 ont tendance à faire plus de pauses inter-propositionnelles que les B2 ( $p < 0,001$ , médianes respectives à 47 % et 42 %,  $\Delta = 0,311$  (small)  $IC = [0,132; 0,47]$ ), mais pas significativement plus à l'intérieur de celles-ci (*ns.*, médianes à 28 et 25 %,  $\Delta = 0,172$  (small)  $IC = [0; 0,334]$ ). Tous les locuteurs, même ceux à plus faible niveau, semblent donc privilégier les frontières de haut niveau syntaxique (entre les propositions) pour placer leurs pauses. Descendons maintenant au niveau du syntagme. Là encore, les locuteurs B1 font plus de pauses inter-syntagmes que les B2 mais la différence n'est pas significative (*ns.*, médianes à 29 et 26 %,  $\Delta = 0,149$  (small)  $IC = [-0,027; 0,316]$ ). En revanche, la différence au niveau intra-syntagme est significative, indiquant que les B1 font également plus de pauses que les B2 à l'intérieur des syntagmes ( $p < 0,05$ , médianes à 21 et 18 %,  $\Delta = 0,187$   $IC = [0,009; 0,353]$ ).

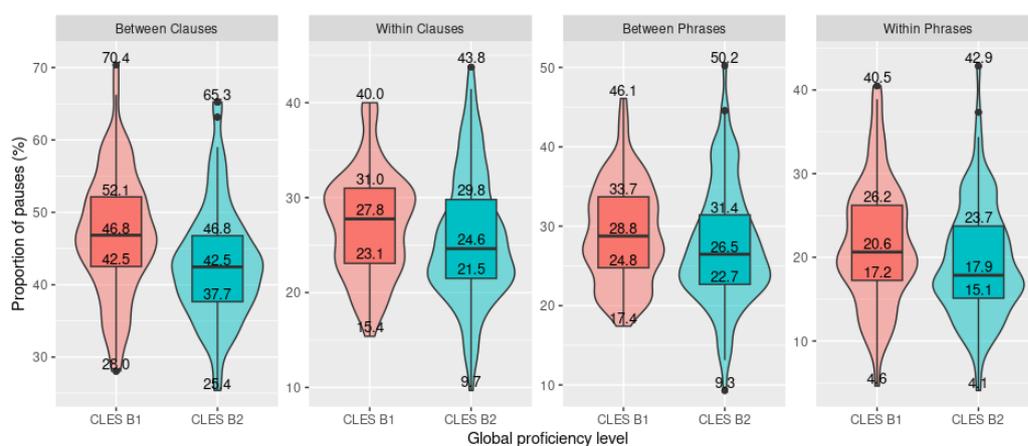


FIG. 12.4 : Proportion de pauses par type de frontière syntaxique

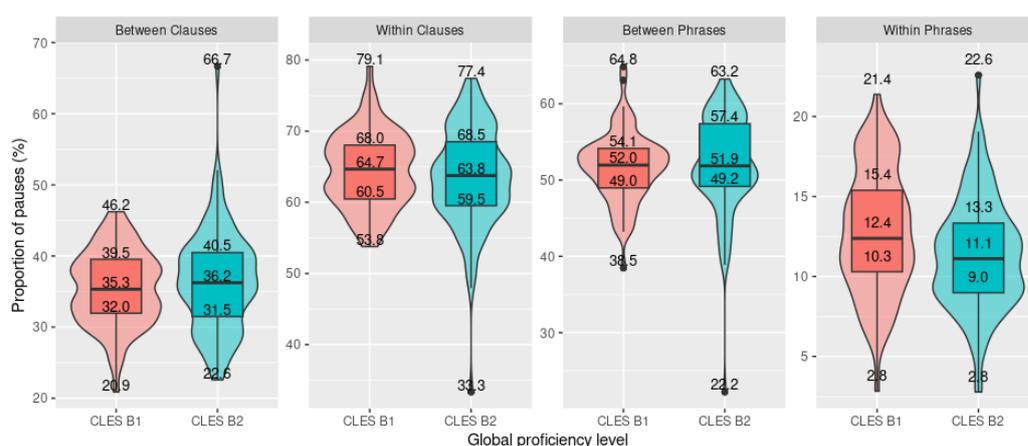


FIG. 12.5 : Proportion de pauses par type de frontière pour 100 pauses

Sachant que les locuteurs B1 ont tendance à faire plus de pauses en général, il n'est pas surprenant d'observer des fréquences plus élevées quelque soit le type de frontière. Pour neutraliser ce facteur, analysons le nombre de pauses de chaque type en fonction du nombre total de pauses par locuteur (cf. figure 12.5). On constate que la proportion de pauses inter-propositions est maintenant légèrement plus faible pour les B1 que pour les B2, mais sans différence significative (*ns.*, médianes à 35 et 36 %,  $\Delta = -0,069$ ), de même pour la proportion de pauses inter-syntagmes (*ns.*, médianes à 52 %,  $\Delta = -0,069$ ). La proportion de pauses intra-syntagmes est quant à elle significativement plus élevée pour les B1, mais toujours avec une taille d'effet relativement limitée ( $p < 0,05$ , médianes à 12 et 11 %,  $\Delta = 0,216$  (small)  $IC = [0,037; 0,382]$ ).

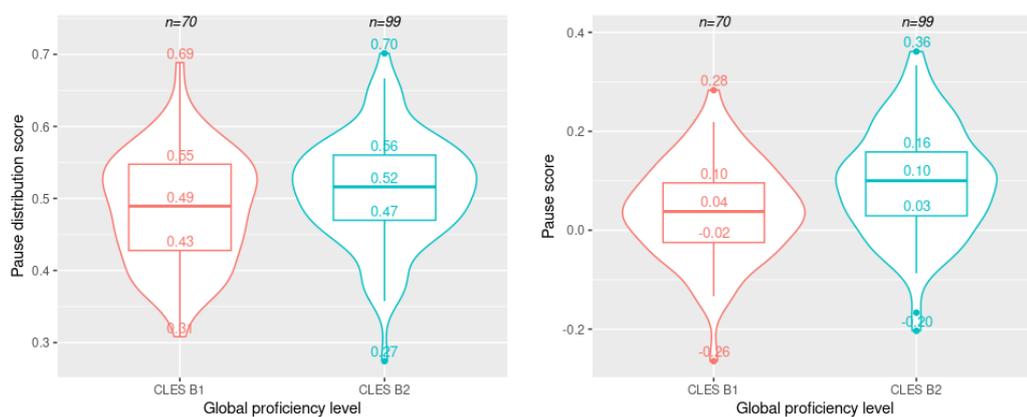
La différence entre les locuteurs B1 et B2 en terme de distribution syntaxique

des pauses semble donc se jouer principalement au niveau de la fréquence des pauses en frontière de bas niveau. Toutefois, la différence que nous observons ici reste limitée.

### 1.3 Score de distribution syntaxique

Calculons maintenant le score de distribution syntaxique des pauses (*SDS*) en fonction des catégories de pauses inter-proposition, inter-syntagme et intra-syntagme. Rappelons que calculer ce score revient à compter le nombre de pauses de chaque type et le normaliser par le nombre total de pause par locuteur, puis en faire un somme pondérée en pénalisant les pauses situées à l'intérieur des syntagmes (-1) et en favorisant les pauses inter-syntagmes (+0,5) et inter-propositions (+1). Le score obtenu permet ainsi de comparer la tendance de distribution syntaxique des pauses par locuteur : plus il est élevé, plus les pauses ont tendance à être placées en frontière de haut niveau. Les résultats indiquent que les locuteurs B1 obtienne en moyenne un score plus faible que les locuteurs B2, mais la taille d'effet reste assez limitée ( $p < 0,05$ , médianes à 0,49 et 0,52,  $\Delta = -0,198$  (small)  $IC = [-0,365; -0,019]$ , cf. figure 12.6a).

On peut également calculer le *SDS* en fonction non plus du niveau des constituants, mais directement du nombre de constituants qui se ferment ou qui s'ouvrent à l'endroit où survient la pause. Il devient ainsi possible de prendre en compte l'imbrication des syntagmes ou des propositions les uns aux autres et permet d'avoir plus de souplesse dans le paramétrage du calcul. La différence entre les locuteurs B1 et B2 apparaît cette fois plus nettement ( $p < 0,001$ , médianes à 0,04 pour B1 et 0,10 pour B2,  $\Delta = -0,301$  (small)  $IC = [-0,455; -0,13]$ , cf. figure 12.6b).



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

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

FIG. 12.6 : Scores de distribution syntaxique des pauses par locuteur (corpus CLES-FR, 180 ms-2 s)

## 1.4 Corpus CLES-JP et CLES-EN

Du côté des corpus CLES-JP et CLES-EN, 21 631 intervalles pour les premiers et 20 486 pour les seconds ont été analysés. Cette fois-ci, la différence de durée des pauses entre les locuteurs B1 et B2 est significative ( $p < 0,001$ ) quelque soit le seuil minimal (180 ou 250 ms) et maximal (2 ou 5 s), avec toutefois une différence faible entre les deux groupes (le  $\Delta$  de Cliff variant 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 compte alors respectivement 6 341 pauses pour les locuteurs japonophones (803 pour les 5 locuteurs B1, 2839 pour les 15 B2, 2699 pour les 9 C1) et 3 785 pauses pour les 15 locuteurs natifs. Comme pour le corpus CLES-FR, plus le niveau du locuteur est élevé plus le débit de parole est élevé également (cf. figure 12.7a), amenant à une quantité de parole, et par extension de pauses observées, plus importante. Ramené au nombre de pauses par mot, la tendance est en revanche inverse, et fortement contrastée entre les niveaux : les locuteurs B1 se situent autour de 43 pauses pour 100 mots, les B2 à 36 pauses, les C1 à 24 et les locuteurs natifs à 18 (cf. figure 12.7b).

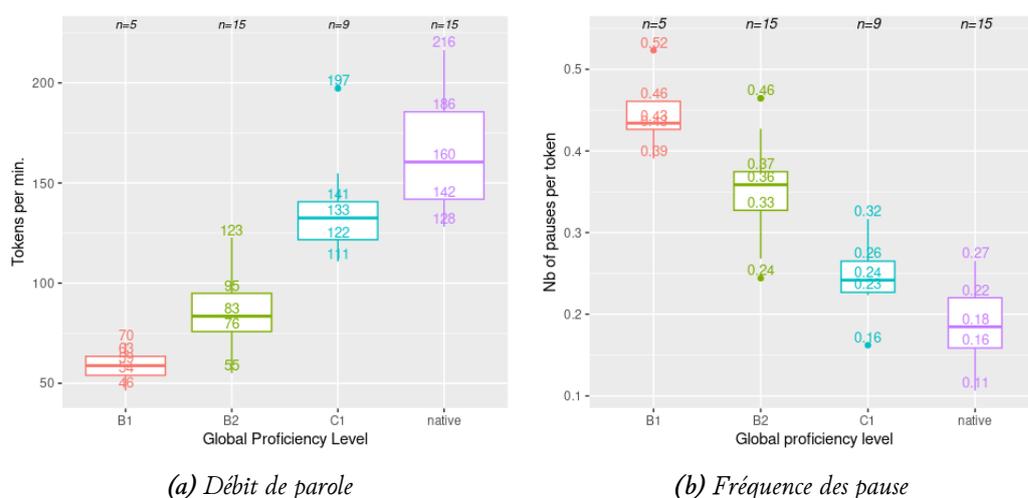
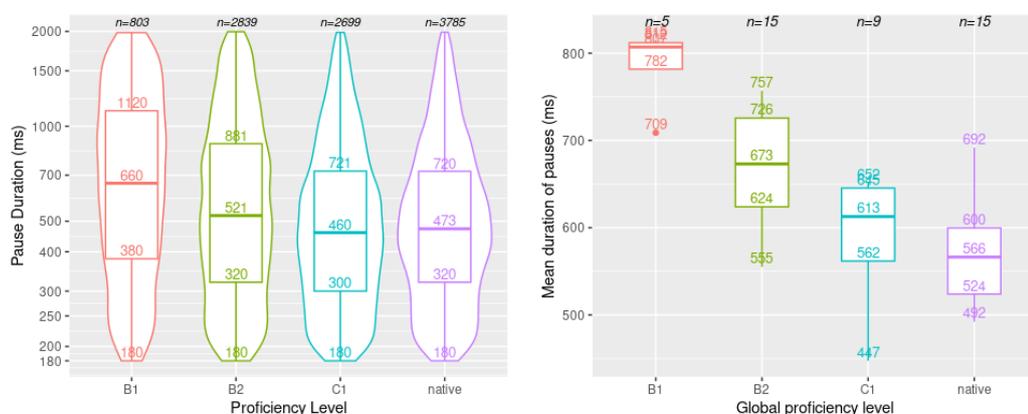


FIG. 12.7 : Débit de parole (gauche) et fréquence des pauses (droite) par locuteur dans les corpus CLES-JP et CLES-EN (180 ms-2 s)

La figure 12.8a montre la distribution des durées de pauses pour chaque groupe. La différence de durée moyenne des pauses par locuteurs apparaît fortement contrastée entre les niveaux. La durée moyenne se situe entre 709 ms et 815 ms, médiane à 807 ms, tandis que la médiane des locuteurs B2 est à 673 ms, celle des C1 à 613 ms et celle des natifs à 566 ms (cf. figure 12.8b). La différence entre les B1 et B2 est significative ( $p < 0,01$ ,  $\Delta = 0,867$  (large)  $IC = [0,338; 0,98]$ ), ainsi qu'entre les locuteurs B1,B2 et les locuteurs natifs ( $p < 0,001$ ,  $\Delta = 0,82$  (large)  $IC = [0,54; 0,937]$ ).



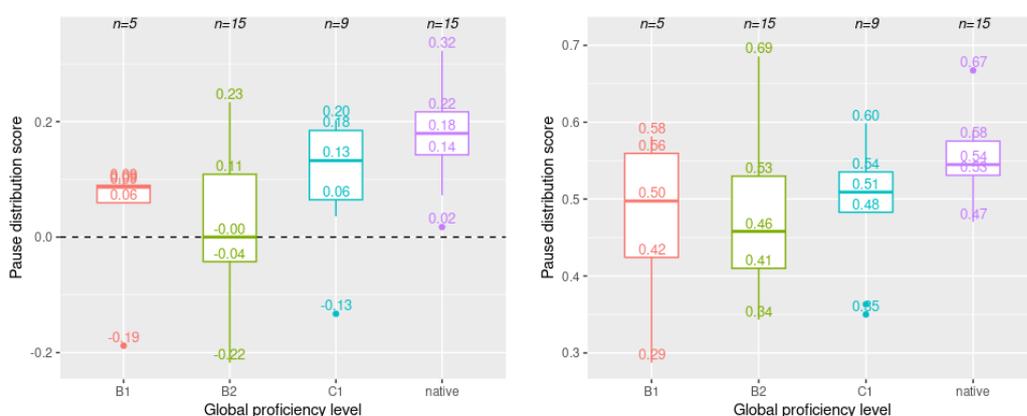
(a) Distribution des durées de pauses

(b) Durée moyenne des pauses par locuteur

FIG. 12.8 : Durées des pauses dans les corpus CLES-JP et CLES-EN (180 ms-2 s)

En termes de distribution syntaxique des pauses, là aussi la différence apparaît plus contrastée que pour les locuteurs francophones, mais suit les mêmes tendances. Les locuteurs B1 font en moyenne significativement plus de pauses que les B2 en frontière de propositions ( $p < 0,01$ , médianes à 58 et 46 %,  $\Delta = 0,840$  (large)  $IC = [0,459; 0,96]$ ), de syntagmes ( $p < 0,05$ , médianes à 43 et 33 %,  $\Delta = 0,657$  (large)  $IC = [0,115; 0,876]$ ) mais la différence n'est pas significative au niveau intra-syntagme, probablement à cause du manque de données pour les B1 ( $ns$ , médianes à 30 et 23 %,  $\Delta = 0,52$  (large)  $IC = [-0,037; 0,83]$ ). Probablement pour les mêmes raisons, la différence de proportion de pauses par type n'est significative à aucun niveau. La comparaison entre les B1+B2 et les locuteurs natifs donne quant à elle des résultats plus fiables : les locuteurs natifs font significativement plus de pauses en frontières de proposition ( $p < 0,05$ , médianes à 35 pour les L2 et 37 % pour les L1,  $\Delta = -0,4$  (medium)  $IC = [-0,687; -0,004]$ ), et moins de pauses à l'intérieur des syntagmes ( $p < 0,05$ , médianes à 13 pour les L2 et 10 pour les L1,  $\Delta = 0,483$  (large)  $IC = [0,087; 0,747]$ ).

Le score de distribution syntaxique des pauses montre également une différence non-significative entre les B1 et B2, qu'il soit basé sur les niveaux de constituants ou sur l'importance des frontières. En revanche, la différence L1/L2 est quant à elle significative :  $p < 0,01$ , médianes à 0,48 pour les L2 et 0,54 pour les L1,  $\Delta = -0,527$  (large)  $IC = [-0,777; -0,131]$  pour le score basé sur les niveaux de constituants, et  $p < 0,001$  médianes à 0,06 et 0,18,  $\Delta = -0,707$  (large)  $IC = [-0,891; -0,32]$  pour le score basé sur la profondeur des frontières syntaxiques (cf. figure 12.9).



(a) Score basé sur le niveau des constituants

(b) Score basé sur la profondeur des frontières

FIG. 12.9 : Score de distribution syntaxique des pauses par locuteur, basé sur le niveau des constituants (propositions et syntagmes, à gauche) ou le niveau de profondeur des frontières syntaxiques (à droite) (corpus CLES-JP et CLES-EN, 180 ms-2 s)

## 12.2 Accentuation lexicale

Nous avons vu dans le chapitre 3 que la précision de position de l'accent lexical est souvent corrélée avec le jugement de compréhensibilité du locuteur. Toutefois, la plupart des études que nous avons citées s'appuient souvent sur de la parole lue ou des annotations manuelles de position de l'accent. Nous avons souhaité savoir si des mesures automatiques permettent de caractériser les patterns accentuels des locuteurs en parole spontanée, et si une différence significative est observable entre les niveaux B1 et B2. En outre, nous nous intéresserons à la qualité de l'accentuation en termes de degré de contraste prosodique entre les syllabes au niveau de la  $F_0$ , de l'intensité et de la durée.

### 2.1 Données analysées

L'ensemble des segments de paroles du corpus CLES-FR analysés par PLSPS compte 68 515 tokens, avec un nombre de tokens par locuteur légèrement inférieur pour les B1 par rapport à celui des B2 (médianes à 376 contre 422, non significatif). Les mesures d'accentuation syllabique que nous avons effectuées ici portent exclusivement sur les mots polysyllabiques lexicaux (noms communs, verbes, adjectifs et adverbes), cependant tous n'ont pas nécessairement été annotés par PLSPS. On compte un total de 14 873 mots polysyllabiques lexicaux, significativement plus nombreux chez les

locuteurs B2<sup>1</sup>. Parmi eux, le nombre de mots annotés par PLSPP n'est que de 6 468, soit seulement 43 % des mots initialement ciblés. Nous reviendrons sur ce constat dans le chapitre 14. L'ensemble des analyses de cette section portent sur ces 6 468 mots, nous les appellerons « mots analysés ».

**Nombre de mots analysés par locuteur** Si le nombre absolu de mots analysés par locuteur est significativement plus élevé pour les B2 ( $p < 0,01$ , médianes à 32 pour B1 contre 41 pour B2), il ne l'est pas ramené au nombre de tokens (médianes à 9 et 10 %) ni au nombre de mots polysyllabiques lexicaux (médianes à 42 et 43 %). On ne peut donc pas dire que les locuteurs B2 utilisent proportionnellement plus de mots polysyllabiques que les B1, ni que les mots qu'ils produisent sont mieux reconnus par PLSPP que ceux des locuteurs B1.

**Caractéristiques des mots analysés** Ce sont en grande partie des noms communs (57% des mots, doublons compris), suivis par des verbes (19%), des adjectifs (12%) et des adverbes (12%). Ils sont majoritairement composés de deux syllabes (73%), mais on trouve également des mots de 3 syllabes (21%), 4 syllabes (5%), 5 et 6 syllabes (moins de 1%). Pour limiter l'influence potentielle de l'accent secondaire, et étant donné que les mots de plus de 3 syllabes représentent moins de 6 % des mots analysés, nous focaliserons les analyses de cette section sur les mots de 2 à 3 syllabes ( $n = 6\ 002$ ). Parmi eux, la syllabe qui porte l'accent primaire théorique (attendu, prescrit) est la syllabe initiale dans 74 % des cas (4 432 mots), médiale dans 13 % (791), et finale dans 13 % (779) des cas (cf. tableau 12.2). Pour les mots à 2 syllabes, 84 % d'entre eux sont accentués à l'initiale contre seulement 16 % en finale. Pour les mots à 3 syllabes, la majorité est accentuée en médiale (58%), puis en initiale (38%) et en finale (4%).

<sup>1</sup> $p < 0,05$ , médianes à 75 pour B1 et 94 pour B2,  $\Delta = -0,232$  (small)  $IC = [-0.396; -0.052]$ ; ramené au nombre de tokens par locuteur :  $p < 0,01$ , médianes à 21 % et 23 %,  $\Delta = -0,238$  (small)  $IC = [-0.397; -0.065]$ .

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

## 2.2 Patterns accentuels observés

Intéressons-nous maintenant aux patterns accentuels produits par les locuteurs. Nous parlerons maintenant de « syllabe proéminente » pour qualifier la syllabe identifiée par PLSPP comme acoustiquement proéminente, et qui sera ainsi perçue, en théorie, comme syllabe accentuée par l'auditeur. Le terme « accent théorique » fait quant à lui référence à l'accent prescrit par le dictionnaire de référence, ici le *CMU Pronouncing Dictionary* (version 0.7b). On observe que 67 % des syllabes proéminentes sont situées en finale, contre 26 % en initiale et 7 % en médiale (cf. tableau 12.2). De manière générale, le pourcentage de mots dont la syllabe proéminente correspond à la position de l'accent primaire théorique est relativement bas (36% à travers le corpus, 32% pour les mots produits par des locuteurs B1, et 38% pour ceux produits par les B2). Comme le nombre de mots analysés varie en fonction des locuteurs, il est plus pertinent de considérer la proportion de mots correctement accentués par locuteur : nous parlerons de « score de position de l'accent ». Les scores individuels sont très variés, ils s'étendent de 0 à 64,8 % (médiane à 34,6 %).

La figure 12.10 présente la distribution des scores par locuteur en fonction du niveau. On constate que les scores se chevauchent largement entre B1 et B2, mais présentent toutefois une différence significative ( $p < 0,01$ , médianes respectives à 30,8 et 36,8 %), avec toutefois une taille d'effet limitée ( $\Delta = -0,275$  (small)  $CI = [-0,432; -0,102]$ ). Par ailleurs, lorsque l'on considère le score obtenu en fonction de la position de l'accent théorique, on constate que le score des locuteurs B2 est effectivement plus élevé pour les mots à accent initial ( $p < 0,01$ , médianes à 25 pour B1 et 32 pour B2,  $\Delta = -0,269$  (small)  $CI = [-0,428; -0,094]$ ) et médial (*ns.*, médianes à 30 et 36,  $\Delta = -0,158$  (small)  $CI = [-0,329; -0,024]$ ),

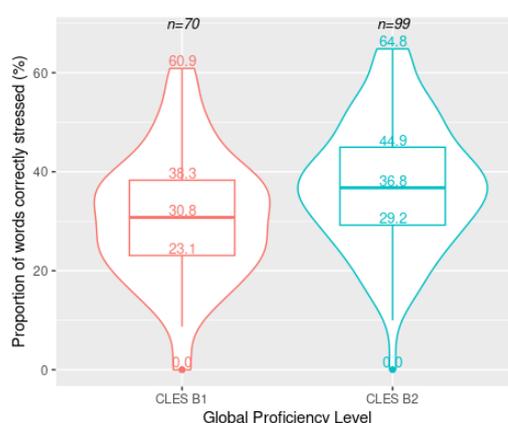


Fig. 12.10 : Score de position de l'accent par locuteur, sur 6 468 mots lexicaux de 2 à 3 syllabes (corpus CLES-FR)

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. 12.3 : Liste des mots de plus de 50 occurrences dans le corpus CLES-FR

mais pas pour les mots à accent final (*ns.*, médianes 77,  $\Delta = -0,032$  (negligible)  $CI = [-0,148; 0,210]$ ). Cela laisse entendre que la différence entre les deux groupes de locuteurs se joue au niveau du taux d’accentuation en initiale et en médiale, plutôt qu’en finale. Le deuxième constat est que les locuteurs B1 semblent présenter des scores plus variés que les B2 (malgré un écart type identique de 12,8 %).

Le tableau 12.3 liste les 18 mots les plus fréquents parmi les mots analysés (plus de 50 occurrences). On peut constater que les mots les plus fréquents sont loin d’être les mieux maîtrisés (“*students*” accentué en initiale seulement 24,6% des fois sur 187 occurrences; “*people*” 20,8% sur 173 occurrences, “*teacher*” 19% (89) et “*teachers*” 10,7% (75) seulement).

### 2.3 Contraste prosodique

Nous nous intéressons dans cette section au degré de contraste entre les syllabes. Il ne s’agit plus simplement de savoir si la proéminence est réalisée sur la syllabe attendue ou non, mais de mesurer à quel point cette syllabe se démarque des autres sur le plan prosodique. Pour cela, nous calculons la différence entre la valeur acoustique normalisée  $P_s$  de la syllabe censée être accentuée et la moyenne  $\bar{P}_u$  des autres syllabes

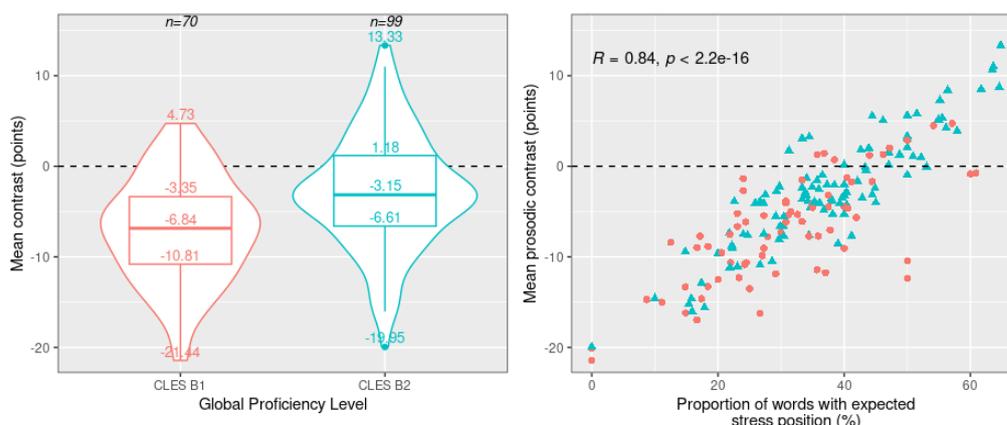


FIG. 12.11 : Contraste prosodique moyen entre  $P_s$  et  $\bar{P}_u$  dans le corpus CLES-FR, à gauche en fonction du niveau du locuteur ; à droite en fonction du score de position de l'accent

du mot. La valeur obtenue varie entre -100 et 100, et indique à quel point le contraste est marqué entre la syllabe censée être accentuée et les autres syllabes. Si la valeur est positive, la syllabe préminente correspond à la syllabe censée être accentuée, plus la valeur est élevée, plus cette syllabe est contrastée par rapport aux autres. Nous appellerons « contraste moyen » la valeur moyenne d'un contraste sur l'ensemble des mots produits par un locuteur. Il pourra s'agir du contraste de  $F_0$ , d'intensité, de durée, ou à défaut la moyenne des trois.

Le contraste moyen par locuteur dans le corpus CLES-FR s'étend de -21,44 à 13,33, avec une médiane à -4,11 indiquant que dans l'ensemble, les locuteurs ont tendance à accentuer la mauvaise syllabe. Si là encore, les locuteurs B1 et B2 se chevauchent largement, la différence entre les deux distributions est significative et plus prononcée qu'avec le simple score de position de la section précédente ( $p < 0,001$ , médianes à -6,84 pour les locuteurs B1 et -3,15 pour les B2,  $\Delta = -0,389$  (medium)  $CI = [-0,534; -0,222]$ , cf. figure 12.11 gauche). On constate également, et sans surprise, que plus le score de position est élevé, plus le contraste moyen est grand, indiquant que les locuteurs qui ont tendance à bien placer l'accent produisent également un contraste plus important entre la syllabe accentuée et les autres syllabes ( $R = 0,84$ ,  $p < 0,001$ , cf. figure 12.11 droite).

Quelle est la dimension prosodique pour laquelle le contraste moyen augmente le plus entre B1 et B2 ? En d'autres mots, quelle est la dimension qui représente le mieux la différence entre les deux groupes de niveau ? Considérons le contraste moyen pour chaque dimension prise séparément (figure 12.12). D'après le delta de Cliff, la plus grande différence entre B1 et B2 se situe au niveau du contraste d'intensité ( $p < 0,001$ , médianes à -5,82 en B1 et 0,77 en B2,  $\Delta = -0,450$  (medium)

$CI = [-0,590; -0,283]$ ), puis vient le contraste de hauteur avec une taille d'effet un peu moins importante ( $p < 0,001$ , médianes à  $-7,38$  et  $-1,44$ ,  $\Delta = -0,317$  (small)  $CI = [-0,475; -0,138]$ ). Quant au contraste de durée de syllabe, il ne présente pas de différence significative entre les deux niveaux (*ns.*, médianes à  $-9,84$  et  $-9,36$ ,  $\Delta = -0,018$  (negligible)  $CI = [-0,196; 0,162]$ ).

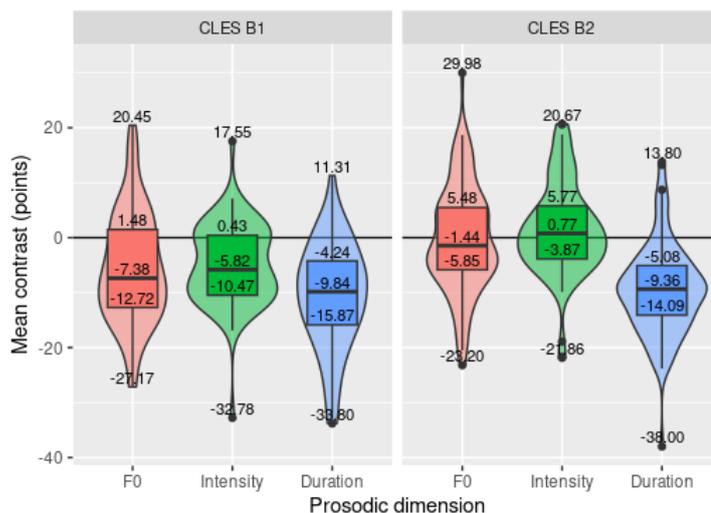


Fig. 12.12 : Contraste moyen par dimension prosodique par locuteur (corpus CLES-FR)

La figure 12.13 présente une visualisation du profil d'accentuation de deux locuteurs SpeakerA et SpeakerB. La figure indique le contraste prosodique moyen du locuteur pour chaque dimension (multidimensionnel,  $F_0$ , intensité et durée ; le premier étant la moyenne des trois suivants). Le contraste est représenté par deux cercles, le premier représente  $P_s$  (la syllabe censée être accentuée), et le second représente  $\bar{P}_u$  (la moyenne des autres syllabes du mots). La valeur inscrite dans les cercles et leur taille représentent la valeur prosodique normalisée, en centile. SpeakerA est représentatif d'un locuteur de haut niveau en termes d'accentuation : son score de position est de 64,8 %, et son contraste moyen (multidimensionnel) est de 13,3. On peut voir que  $P_s$  est bien contrastée au niveau de la  $F_0$  (contraste de 30) et de l'intensité (17), mais pas au niveau de la durée (-5). SpeakerB a quant à lui un score de position de seulement 14,8 %, et un contraste moyen de -9,4. On peut voir que c'est  $\bar{P}_u$  qui a tendance à être accentué (donc la mauvaise syllabe), avec une influence en premier lieu de la durée (-21), puis de la  $F_0$  (-11), tandis que l'intensité ne semble pas mobilisée (+2). Des tendances similaires se retrouvent chez les locuteurs au score et au contraste moyen élevé (prépondérance de la  $F_0$  et de l'intensité, neutralisation de la durée) et chez les locuteurs au score et contraste faible (impact de la durée et de la  $F_0$ , pas ou peu de contraste d'intensité).

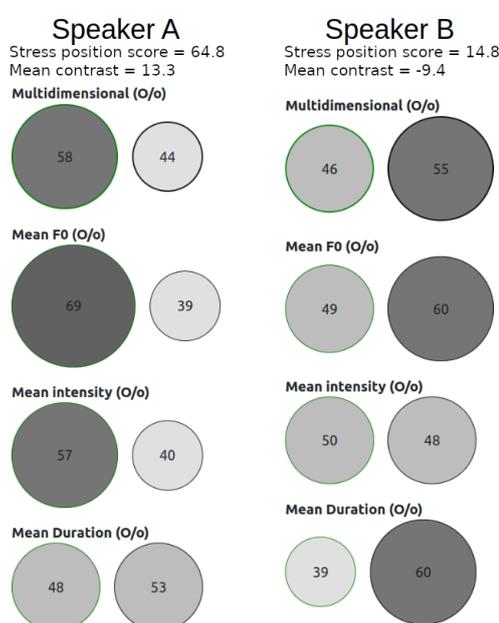


Fig. 12.13 : Contrastes prosodiques de deux locuteurs au profil différent  
Écouter un extrait : *Speaker A* et *Speaker B*.

## 2.4 Corpus CLES-JP et CLES-EN

Quatre-cents-trente segments de parole ont été extraits des corpus CLES-JP et CLES-EN, totalisant 41 714 tokens (21 356 pour les 29 locuteurs japonophones et 20 358 pour les 15 locuteurs natifs). La proportion de mots polysyllabiques lexicaux par token est similaire dans les deux corpus (médianes à 23 % pour les locuteurs B1/B2 du CLES-JP, et 22 % pour les locuteurs natifs), cependant, et contrairement à ce à quoi on aurait pu s'attendre, la proportion de mots annotés par PLSP est plus grande pour les locuteurs japonophones que pour les locuteurs natifs ( $p < 0,001$ , médianes à 41 % pour les locuteurs B1 et B2 ensemble et 30 % pour les natifs,  $\Delta = 0,753$  (large)  $IC = [0,410; 0,910]$ ). Les mots semblent donc en moyenne mieux reconnus quand les locuteurs ne sont pas natifs. Au total, 1 913 mots ont été annotés pour CLES-JP et 1 354 pour CLES-EN. La proportion des catégories grammaticales et des nombres de syllabes est sensiblement similaire entre les deux corpus, comme l'indique le tableau 12.4. Comme pour le corpus CLES-FR, nous nous focaliserons sur les mots de 2 à 3 syllabes, soit environ 96 % des mots annotés.

La figure 12.14 présente les scores de position de l'accent par locuteur. Le premier constat est qu'on n'observe pas d'amélioration significative entre les niveaux des locuteurs japonophones. Les 5 locuteurs de niveau B1 obtiennent un score variant entre 35,3 et 70,6% (médiane à 45,8 %), et les 15 locuteurs B2 entre 31,7 et 72,7 % (médiane

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

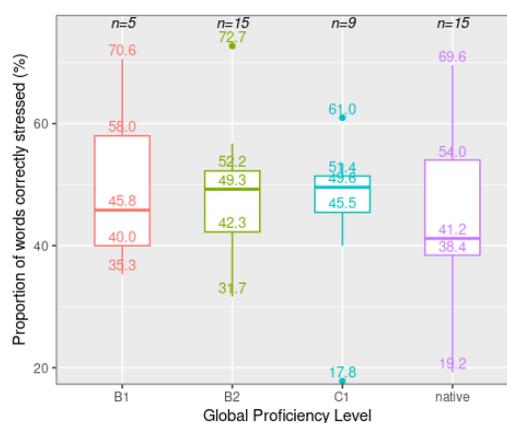


FIG. 12.14 : Scores de position par locuteur, sur 3 014 mots lexicaux de 2 à 3 syllabes (corpus CLES-JP et CLES-EN)

à 49,3 %, différence non significative). Les 9 locuteurs C1 n'obtiennent pas de résultats sensiblement supérieurs. Le deuxième constat est que les locuteurs natifs obtiennent des résultats très variables, s'étalant de 19,2 et 69,6 %, et en moyenne inférieurs à ceux des locuteurs japonophones (médianes à 49,3 % pour CLES-JP et 41,2 % pour CLES-EN, différence non significative,  $\Delta = 0,170$  (small)  $IC = [-0,223; 0,516]$ ). La détection de la syllabe proéminente semble influencée par un facteur indépendant du niveau de compétence en langue.

Le tableau 12.5 présente le nombre de mots par position d'accent théorique (attendu) ou observé (syllabe proéminente détectée par PLSPP). On peut constater que, si on s'attend à avoir environ 80 % de mots accentués en initiale, 7 % de mots en médiale et 10 % de mots en finale dans les deux corpus, PLSPP détecte une proéminence sur l'initiale dans seulement 40 % des mots, 5 % en médiale et presque 60 % en finale, pour les japonophones comme pour les anglophones natifs. Un grand nombre de proéminences sont donc détectées sur la dernière syllabe des mots, même lorsque elle n'est pas censée être accentuée, et ce quelque soit la langue maternelle du locuteur.

Lorsque l'on mesure le contraste prosodique entre la syllabe censée être accentuée  $P_s$  et la moyenne des autres syllabes  $\bar{P}_u$ , on constate que  $P_s$  est systématiquement plus

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

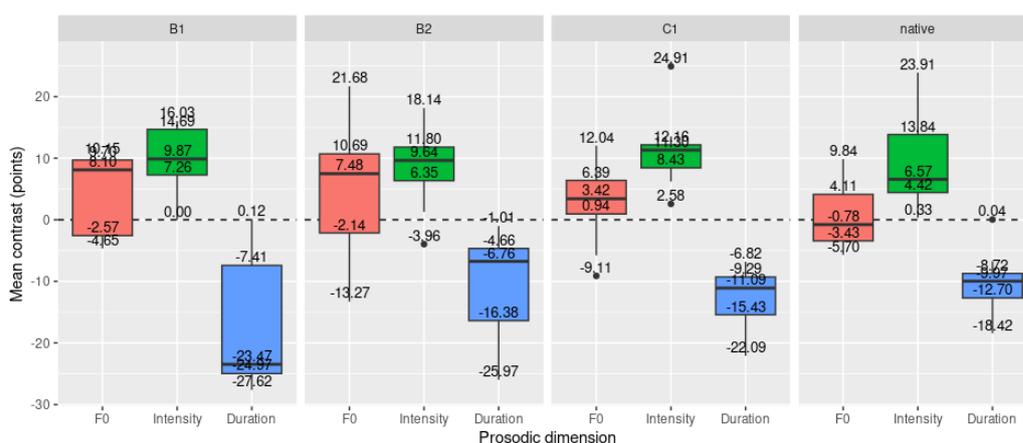


FIG. 12.15 : Contraste moyen par locuteur entre  $P_s$  et  $\bar{P}_u$  (corps CLES-JP, CLES-EN)

haute en intensité, mais plus courte en durée que  $\bar{P}_u$  (cf. figure 12.15). La  $F_0$  a quant à elle tendance à être plus haute mais le contraste est moins fort. On observe les mêmes tendances que pour le corpus CLES-FR (cf. figure 12.11), mais les valeurs apparaissent plus extrêmes : les contrastes d'intensité et de hauteur sont assez marqués et majoritairement en adéquation avec la position de l'accent théorique ; le contraste de durée est quant à lui fortement négatif. Ce contraste négatif indique qu'une syllabe qui n'est pas censée être accentuée est plus longue que les autres, suffisamment pour rendre le contraste moyen des trois dimensions négatif, et induire en erreur la détection de la proéminence. Le problème semble être au niveau de la durée de la syllabe finale, souvent assez longue pour faire pencher la balance pour une proéminence en finale.

En analysant la corrélation entre le contraste par dimension prosodique et le score de position de l'accent des locuteurs japonophones et anglophones, on observe que les deux groupes de locuteurs ne se comportent pas de la même manière : le score des locuteurs natifs apparaît très corrélé avec le contraste d'intensité ( $R = 0,95$ ,  $p < 0,001$ ), tandis qu'il l'est moins pour les locuteurs japonophones ( $R = 0,60$ ,  $p < 0,001$ ). Le constat est inverse pour la  $F_0$  :  $R = 0,54$  ( $p < 0,05$ ) pour les locuteurs natifs mais  $R = 0,65$  ( $p < 0,001$ ) pour les locuteurs du corpus CLES-JP. Le contraste de durée, quant à lui, n'est significatif que pour les locuteurs natifs ( $R = 0,73$ ,  $p < 0,01$  ; contre

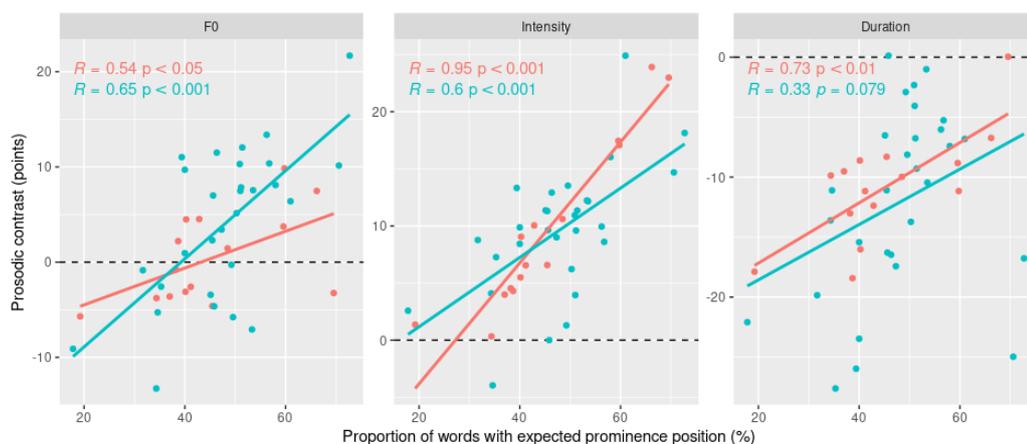


FIG. 12.16 : Corrélation entre le contraste moyen par dimension prosodique et le score de position de l'accent par locuteur (corpus CLES-JP en bleu et CLES-EN en rouge)

$R = 0,33$ ,  $p = 0,08$  pour CLES-JP). Comment interpréter ces résultats ? Il semble possible de formuler l'hypothèse que le paramètre d'intensité est le plus fiable des trois dimensions prosodiques mesurées par PLSP. Les locuteurs natifs qui produisent un contraste d'intensité important entre les syllabes tendent à obtenir un meilleur score global d'accentuation. De leur côté, les locuteurs japonophones semblent plutôt s'appuyer sur le contraste de hauteur : il s'agit de la corrélation la plus forte avec le score de position, mais elle est également plus forte et le contraste est plus élevé que pour les locuteurs natifs. Cela est peut-être dû au fait que les natifs accentuent parfois par une chute de  $F_0$ , ou qu'ils ont tendance à dévoiser certaines voyelles, perturbant les mesures de hauteur. Quant à la durée, elle semble aléatoire chez les locuteurs japonophones (pas de corrélation avec le score de position), ce qui pourrait être dû à une tendance à hésiter en fin de mot et ainsi allonger la dernière syllabe. Le contraste de durée apparaît pourtant fortement corrélé au score de position pour les locuteurs natifs, bien que négatif, indiquant que les locuteurs qui produisent une syllabe accentuée plus longue que les syllabes non accentuées ont bien tendance à obtenir un meilleur score de position de l'accent.

## 12.3 Conclusion

Nous avons vu dans ce chapitre que le nombre de pauses et leur durée moyenne dépendent du niveau de compétence en langue du locuteur, et que, bien que les locuteurs B1 fassent plus de pauses que les B2 à tous les niveaux syntaxiques (inter-proposition, inter-syntagme et intra-syntagme), ils ont tendance à faire proportion-

nellement plus de pauses intra-syntagmes que les locuteurs B2, mais pas significativement plus de pauses inter-propositionnelles – idem pour les locuteurs japonophones vis-à-vis des locuteurs natifs. Les scores de distribution syntaxique des pauses confirment que plus le niveau du locuteur augmente, plus les pauses ont tendance à être placées en frontière de haut niveau syntaxique, avec les locuteurs natifs en tête de classement.

L'analyse des annotations automatiques de proéminence syllabique montrent que les locuteurs B2 ont tendance à mieux positionner l'accent et produire un meilleur contraste acoustique entre la syllabe accentuée et les autres syllabes du mot. On constate toutefois une grande variabilité dans les mesures individuelles, le score de position de l'accent s'étalant de 0 à 65 % selon les locuteurs du corpus CLES-FR et présentant un large chevauchement entre les locuteurs B1 et B2 malgré une différence significative. Les annotations font ressortir une nette influence des patterns accentuels de la langue maternelle des locuteurs. Les francophones ont généralement tendance à augmenter la  $F_0$  et allonger la syllabe finale, tandis que l'intensité reste stable, quelle que soit la position de l'accent théorique. 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. Du côté des locuteurs japonophones, la tendance à accentuer la syllabe finale est beaucoup moins marquée, avec de meilleurs scores de position de l'accent de manière générale (de 17,8 à 72,7 %), et une forte mobilisation de la  $F_0$  et de l'intensité dès le niveau B1. Il n'a malheureusement pas été possible de faire ressortir de différence significative entre les niveaux étant donné le faible nombre de locuteurs du corpus. Enfin, les analyses de locuteurs natifs ont fait ressortir une forte tendance à l'allongement de la syllabe finale et un contraste moyen de  $F_0$  limité, conduisant à des scores de position globalement bas (entre 19,2 et 69,7 %), et en moyenne inférieurs à ceux des locuteurs japonophones. Le contraste d'intensité s'est révélé quant à lui plus important que pour les locuteurs non-natifs et très corrélé avec le score de position de l'accent. Nous revenons sur les limites de l'outil de mesure dans le chapitre 14.