

Chapitre 4

Collecte de données de parole

Notre objectif est d'observer les schémas de pauses et d'accentuation lexicale chez des locuteurs francophones et japonophones d'anglais, situés à la frontière entre deux niveaux de compétence : « intelligible mais difficile à comprendre » (niveau B1) et « intelligible et facile à comprendre » (niveau B2). Pour mener cette étude, plusieurs contraintes méthodologiques doivent être respectées. Premièrement, les enregistrements de parole doivent être d'une qualité audio suffisante et d'une durée adéquate pour permettre des analyses automatiques. Deuxièmement, la parole recueillie doit être spontanée, de manière à ce que les pauses reflètent la capacité du locuteur à planifier et structurer son discours. Enfin, les locuteurs doivent appartenir aux niveaux B1 ou B2 afin de comparer les schémas de pauses et d'accentuation entre ces deux groupes de compétence.

L'épreuve d'interaction orale du CLES B2 s'est révélée être une opportunité particulièrement adaptée à la constitution de ce type de corpus. Elle propose en effet une mise en situation originale sous la forme d'un jeu de rôles d'une dizaine de minutes entre deux ou trois candidats, évalués *in situ* par un ou deux examinateurs accrédités.

Avec le soutien de la Direction du CLES, des coordinateurs et des évaluateurs sur le terrain, nous avons obtenu l'autorisation d'enregistrer plusieurs sessions du CLES B2 à Grenoble et à Valence. Ces enregistrements ont permis de constituer un premier corpus de locuteurs francophones, ci-après désigné sous le nom de CLES-FR. En parallèle, afin de disposer d'un corpus comparable pour des locuteurs japonophones, nous avons collaboré avec les universités Dōshisha (Kyōto) et Waseda (Tōkyō) au Japon. Ce corpus sera désigné sous le nom de CLES-JP. Enfin, pour observer les comportements de locuteurs anglophones natifs dans une situation de communication similaire, nous avons constitué un troisième corpus, nommé CLES-EN.

Dans ce chapitre, nous présenterons en détail chacun de ces trois corpus : leurs caractéristiques, leurs points communs et leurs spécificités.

4.1 Corpus CLES-FR

La certification CLES évalue les candidats sur quatre habiletés langagières, à trois niveaux différents : B1, B2 et C1. Au niveau B2, l'expression orale est évaluée dans le cadre d'une épreuve d'interaction orale. Lors de cette épreuve, deux ou trois candidats participent à un jeu de rôles, où il leur est demandé de présenter et défendre un point de vue sur un sujet d'actualité, puis d'aboutir à un compromis en une dizaine de minutes. Le rôle attribué à chaque candidat détermine leur position : en faveur ou contre le sujet traité. Ce sujet est directement lié aux épreuves de compréhension orale et écrite réalisées en amont ; les candidats ont donc déjà été exposés à la thématique.

Avant de commencer la discussion, les candidats disposent de 2 minutes de préparation pour organiser leurs idées et prendre des notes, sans toutefois être autorisés à les lire durant l'échange. La discussion s'achève au bout de 10 minutes, ou lorsque les candidats arrivent à un consensus et que les examinateurs jugent avoir suffisamment de matière pour évaluer les participants.

Le protocole d'évaluation CLES stipule que les candidats peuvent être enregistrés, pour permettre au jury de réécouter certains candidats lors de la délibération, d'utiliser certains enregistrements à des fins de formation d'évaluateurs ou de recherches dans le cadre du CLES. La présence des microphones dans la salle n'a donc pas, a priori, influencé la production des candidats, puisqu'il s'agit d'une situation normale d'épreuve d'interaction orale du CLES. Toutefois, nous sommes conscients que le contexte d'examen impacte largement la production des locuteurs (stress, sentiment de jugement, enjeux), à quoi s'ajoute la présence d'un microphone qui peut être une source de stress supplémentaire.

De nouveaux sujets de certification sont régulièrement édités par le CLES. Les sujets utilisés cette fois-ci concernent l'usage de la e-cigarette, la généralisation des caméras de surveillance et les tests cliniques sur les animaux. La formulation des sujets varie légèrement lorsque trois candidats participent au débat, de manière à ce que le troisième prenne un rôle de médiateur, comme présenté table 4.1. L'ensemble des énoncés du CLES B2 sur le thème de la e-cigarette est accessible en ligne¹. Les autres sujets sont encore utilisés par le CLES au moment de l'écriture de ce manuscrit, et ne peuvent donc pas être rendus publics.

¹<https://www.certification-cles.fr/se-preparer/exemples-de-sujets/exemple-de-sujet-cles-b2-anglais-1219069.kjsp?RH=8204107280166102>

Consigne générale	You will participate in a contradictory debate with another candidate. You will be asked to defend a given position. You will use the information from the documents you have been studying to discuss, negotiate, and reach a compromise according to the role that has been assigned to you.
Situation	The journalists you are working with are preparing a documentary film on e-cigarettes. They hesitate on what to insist on. You and your colleague(s) have been asked to suggest the best possible angle for the film. At the end of the discussion, you should reach an agreement based on a compromise between your different perspectives.
Rôle A	You think that the report should concentrate on the concerns about e-cigarettes and the need for regulation.
Rôle B	You think that the report should concentrate on the possible benefits of e-cigarettes.
Rôle C	You think that the report should be objective and integrate all points of views.

TAB. 4.1 : Exemple de sujet du CLES B2 sur l'utilisation de la e-cigarette (situation 2)

L'évaluation est réalisée par un ou deux examinateurs accrédités présents dans la salle, selon une grille standardisée comportant huit critères couvrant tant la qualité linguistique que le respect des consignes (cf. grille CLES B2, section 1.1.5 et [Annexe A.7](#)). Le niveau B2 est attribué uniquement si tous les critères sont validés ; à défaut, le candidat peut valider le niveau B1 si ses compétences sont jugées suffisantes, ou ne rien valider du tout.

Les enregistrements ont été réalisés entre janvier 2020 et janvier 2023, avec des équipements variés (microphones et paramètres d'échantillonnage). Tous les fichiers ont été normalisés pour garantir une qualité homogène (44,1 kHz, stéréo 16-bit PCM, 1411 kb/s).

Données recueillies

Un total de 260 locuteurs ont été enregistrés lors de cinq sessions CLES. Parmi eux, 232 ont participé en binômes, 15 en trinômes, et 13 en monômes. La répartition par genre est parfaitement équilibrée (130 femmes, 130 hommes). Sur l'ensemble, 151 locuteurs (58%) ont obtenu la certification B2, 75 (29%) ont été certifiés B1, et 34 (13%) n'ont validé aucun niveau. Quarante-cinq locuteurs (17%) ont déclaré une langue maternelle autre que le français, incluant notamment l'arabe, le chinois et d'autres langues déclarées par une ou deux personnes chacune.

Pour les besoins de cette thèse, seuls les candidats de langue maternelle française ayant obtenu un niveau B1 ou B2 ont été conservés dans le corpus final, excluant également les monômes. Cela a permis d'obtenir un corpus de 170 locuteurs, désigné par le terme *CLES-FR*. Parmi ces locuteurs, 99 ont été certifiés B2 (58%), 71 B1 (42%). Concernant spécifiquement l'épreuve d'interaction orale, 118 locuteurs (69%) ont obtenu un niveau B2, contre 52 (31%) un niveau B1. La répartition par genre reste équilibrée, avec 89 femmes (52%) et 81 hommes (48%). Enfin, 11 locuteurs (6%) ont participé en trinômes.

+ temps total estimé ; durée moyenne, min max ; durée médiane par locuteur, min max IQR

4.2 Corpus CLES-JP

Pour obtenir un corpus comparable de locuteurs japonais, une seconde collecte a été organisée au Japon, en reproduisant autant que possible les conditions du CLES B2. Les locuteurs ont été recrutés parmi les étudiants des universités Waseda et Dōshisha. Ils remplissaient les critères suivants : être de langue maternelle japonaise, avoir un niveau d'anglais équivalent à B1 ou supérieur (TOEFL iBT Speaking 16, IELTS Speaking 5.5, Eiken English Proficiency test pre-1 level), et accepter les conditions d'utilisation des enregistrements. Chaque candidat a complété un questionnaire détaillé, incluant des informations sur leurs séjours à l'étranger et leur profil linguistique. Les participants ont été répartis en binômes en veillant à ce que chacun des locuteurs ait un niveau proche de celui de son interlocuteur. Grâce au concours des universités locales, une rétribution financière a pu être octroyée aux participants à hauteur de 1 500 ¥ pour les étudiants de Waseda, et 2 000 ¥ pour ceux de Dōshisha (respectivement 9,4 € et 12,5 € au moment de l'expérimentation).

Il n'a pas été possible d'utiliser les mêmes sujets que pour le corpus CLES-FR car ceux-ci étaient toujours en circulation en France. Aussi, deux sujets similaires ont été conçus pour l'occasion : le premier sur l'utilisation des intelligences artificielles génératives en classe de langue et le second sur le travail en parallèle des études. Les deux sujets sont donnés en [Annexe B](#). Créer ces sujets nous a permis d'être plus près des préoccupations actuelles des étudiants et a permis une discussion plus riche, sachant qu'ils n'avaient pas eu l'occasion de travailler ces sujets en amont comme c'est le cas pour les candidats CLES. Par ailleurs, les participants ont bénéficié d'un temps de préparation de durée flexible avant de commencer la discussion, et d'une liste de mots clés écrits qu'ils étaient libres d'utiliser ou non. Pour laisser le temps aux participants de rentrer dans la discussion, la durée de celle-ci n'était pas limitée mais devait durer un minimum de 10 minutes.

L'ensemble des enregistrements a été effectué avec un Zoom Handy Recorder H2n, et échantillonné à 44,1 kHz, stéréo 16-bit PCM, 1411 kb/s.

Données recueillies

Un total de 29 locuteurs de langue maternelle japonaise ont été enregistrés dans le cadre de la constitution du corpus CLES-JP. On compte 17 femmes (59%) et 12 hommes (41%). Leur niveau de compétence en anglais est estimé à partir des résultats obtenus à différentes certifications, sur la base d'une auto-déclaration : 5 d'entre eux sont de niveau équivalent B1 (17%), 15 de niveau équivalent B2 (52%), et 9 de niveau équivalent C1 (31%). Suite à l'absence d'un candidat, l'un des binômes enregistrés est constitué d'un étudiant de niveau C1 et d'un enseignant d'anglais de langue maternelle japonaise qui a dû assurer le remplacement. Les tours de parole de l'enseignant ne sont pas incluses dans le corpus CLES-JP.

+ temps total estimé ; durée moyenne, min max ; durée médiane par locuteur, min max IQR

4.3 Corpus CLES-EN

Afin d'observer les patterns de pauses et d'accentuation par des locuteurs anglophones natifs dans une situation similaire, nous avons constitué un corpus supplémentaire de locuteurs natifs dans les mêmes conditions que le corpus de locuteurs japonophones. Les participants ont été recrutés à l'université de Dōshisha, parmi un groupe d'étudiants en échange originaires des États-Unis, arrivés au Japon quelques jours avant les enregistrements.

L'ensemble des enregistrements a été effectué avec un Zoom Handy Recorder H2n, et échantillonné à 44,1 kHz, stéréo 16-bit PCM, 1411 kb/s.

Données recueillies

Le corpus CLES-EN est constitué de 14 locuteurs, tous originaires des États-Unis, et inscrits en licence dans une université américaine. Ils ont entre 20 et 22 ans ($M = 20,5$), 9 d'entre eux sont des femmes et 5 sont des hommes.

+ temps total estimé ; durée moyenne, min max ; durée médiane par locuteur, min max IQR

4.4 Comparabilité des corpus

Les corpus CLES-FR et CLES-JP partagent un objectif commun : analyser les patterns de pauses et d'accentuation lexicale dans la parole spontanée de locuteurs de niveaux B1 et B2, dans une situation de communication comparable. Le corpus CLES-EN permettra de comparer, à titre indicatif, les résultats obtenus avec ceux de locuteurs anglophones natifs dans une situation de parole similaire.

Ces trois corpus présentent l'avantage d'être comparables à plusieurs niveaux :

- **Format des interactions** : Chaque corpus repose sur des dialogues argumentatifs d'une dizaine de minutes, en binômes ou trinômes, réalisés dans un contexte formel. Les participants ont eu un temps de préparation pour se familiariser avec le sujet avant de commencer la discussion.
- **Contraintes** : Les conversations sont des jeux de rôles, les locuteurs n'ont pas eu le choix du sujet, du rôle qui leur est attribué, ni du ou des partenaires de discussion. Dans la majorité des cas, ils ne connaissent pas leur interlocuteur.
- **Nature des sujets** : Les sujets abordés dans chaque corpus encouragent une prise de position et une argumentation structurée.
- **Enregistrements standardisés** : Les discussions ont été enregistrées dans des conditions acoustiques similaires et rééchantillonnées selon des paramètres uniformes (44,1 kHz, stéréo 16-bit PCM).
- **Niveaux de compétence** : Les participants sont majoritairement de niveaux B1 et B2, estimés à partir de certifications standardisées ou de critères comparables.

Malgré tout, certaines différences importantes demeurent et nous devons en tenir compte lors de l'interprétation des résultats :

- **Contexte et enjeux** : Le corpus CLES-FR a été enregistré dans un cadre réel d'examen, avec le stress associé à une évaluation officielle. Les corpus CLES-JP et CLES-EN, en revanche, ont été constitués dans des conditions expérimentales rétribuée financièrement, avec nécessairement moins d'enjeux.
- **Préparation des participants** : Les locuteurs du CLES-FR ont pu travailler en amont le sujet abordé pendant la conversation, à travers les épreuves précédentes du CLES. Ce n'est pas le cas des locuteurs des deux autres corpus, qui ont découvert le sujet de discussion seulement quelques minutes avant le début de la conversation.

- **Taille des corpus** : Le corpus CLES-FR est le plus important avec 170 locuteurs, suivi par le corpus CLES-JP (29 locuteurs) et le corpus CLES-EN (14 locuteurs). Ces différences influenceront nécessairement la robustesse des analyses statistiques qui suivront.

4.5 Publication des corpus

Un corpus public constitué des enregistrements des sessions CLES pour lesquels l'ensemble des participants ont donné leur accord pour la publication des données a pu être mis à disposition sur la plateforme Ortolang². L'accès au corpus est toutefois restreint à un cadre de recherches académiques. Parmi les 260 locuteurs initialement enregistrés, 162 ont donné leur accord (62%), et 138 d'entre eux font partie d'un binôme ou trinôme où tous les locuteurs impliqués ont donné leur accord de diffusion.

Le corpus publié réunit 62 enregistrements de 128 locuteurs ayant passé l'épreuve du CLES B2, totalisant 10 h de parole. Parmi les locuteurs, 119 ont déclaré avoir le français comme langue maternelle (93%). On compte 61 femmes (48%) et 67 hommes (52%). La durée moyenne des enregistrements est de 9 min 35 s (min.: 5 min 12 s, max.: 14 min 30 s). Le résultat obtenu au CLES est B2 pour 62 d'entre eux (48%), 50 ont validé le niveau B1 (39%) et 16 n'ont rien validé (13%).

Les corpus CLES-JP et CLES-EN ont quant à eux été publiés en l'état, sur la même plateforme (LIENS).

4.6 Annotations *gold standard*

Un échantillon du corpus CLES-FR a été transcrit semi-automatiquement puis corrigé pour constituer un sous-corpus de référence. Vingt enregistrements ont été sélectionnés aléatoirement avec pour contrainte de contenir 20 candidats certifiés B2 et 20 certifiés B1, et un équilibre homme/femme. Le travail d'annotation a été effectué par un stagiaire d'excellence au sein du Laboratoire d'Informatique de Grenoble. Une transcription automatique des enregistrements a d'abord été effectuée à l'aide du logiciel Whisper (Radford et al., 2022, modèle base multilingue), puis manuellement corrigée et segmentée en locuteurs.

Du côté des corpus CLES-JP et CLES-EN, les corpus ont d'abord été automatiquement segmentés en locuteurs avec la pipeline de diarisation pyannotate.audio (Bredin, 2023), puis vérifiés manuellement dans leur intégralité.

²<https://hdl.handle.net/11403/cles-spontaneous-english>

Conclusion

Un total de 304 locuteurs ont pu être enregistrés pendant les deux collectes de données organisées, totalisant 26 h de parole spontanée. Parmi eux, 260 ont été enregistrés lors d'épreuves du test certificatif CLES sur les campus de Grenoble et Valence de l'université Grenoble Alpes ; et 44 locuteurs ont été enregistrés lors de simulations de sessions CLES, organisées dans les universités de Waseda et Dōshisha, au Japon.

Trois corpus ont été créés sur la base de ces enregistrements : CLES-FR, qui contient 170 locuteurs francophones B1 et B2 ; CLES-JP, qui contient 29 locuteurs japonophones B1, B2 et C1 ; et CLES-EN qui contient 14 locuteurs anglophones natifs originaires des États-Unis. CLES-FR constitue notre corpus principal en termes de volume et de représentativité des niveaux cibles de notre recherche, B1 et B2. Les corpus CLES-JP et CLES-EN sont de taille plus modeste mais permettront d'avoir un aperçu des résultats obtenus avec des locuteurs japonophones de différents niveaux, ainsi que des locuteurs natifs de l'anglais. Par ailleurs, ils ont été réalisés dans les mêmes conditions (recrutement des locuteurs, sans enjeu d'examen) et sur les mêmes sujets de conversation. Aussi, dans la partie résultats, nous proposons de comparer les niveaux B1 et B2 du corpus CLES-FR d'une part, et les niveaux B1, B2, C1 et locuteurs natifs d'autre part.

Dans le chapitre suivant, nous présentons les outils utilisés pour segmenter, transcrire et annoter les enregistrements de ces trois corpus, ainsi que la méthodologie adoptée et les outils développés pour analyser les patterns de pauses et d'accentuation lexicale.