

Historique

Le corpus FrOMi a pu être collecté grâce à la rencontre de N. Dherbey Chapuis et de V. Milan lors du forum du 13 novembre 2018 organisé par L. Pasquier responsable de [Paysages éducatifs](#) à la ville de Fribourg. Mme Milan était alors directrice de l'école Villa Thérèse au quartier du Schönberg.

Description

FrOMi est un corpus longitudinal constitué de productions orales spontanées de 10 élèves allophones qui ne parlaient pas un mot de français en entrant à l'école, et de 3 élèves de français langue seconde (FLS) qui avaient une maîtrise variable du français au moment de leur entrée à l'école. Chaque participant a été enregistré durant un matin d'école (3h15) tous les 3 mois pendant 4 années scolaires. FrOMi est composé de 445 heures d'enregistrements et de leurs transcriptions en français standard.

FrOMi offre l'opportunité d'étudier le développement longitudinal du français langue seconde sur une période de 36 mois dans un groupe d'apprenants très jeunes (de 5 à 9 ans) avec des langues premières sous-étudiées (Amharic, Tigrinya, kurde...).

Projets de recherche portant sur le corpus FrOMi

2025-2028 Demande de financement **en attente**, *Influence de l'usage oral des chunks sur le développement de la grammaire en français langue seconde*. Ce projet est une collaboration de la Dr. Dherbey Chapuis et de la Prof. Thomas. La mise en perspective de la phonologie et de la grammaire vise à mieux comprendre le développement grammatical tant d'un point de vue cognitif, que de ses possibles applications didactiques.

2024-2027 *Transcription du corpus FrOMi* (Sept.2024-Dec.2027); Financement POOL de recherche (Ref. PO2490, 30'000CHF; N. Dherbey Chapuis). Ce projet vise à réaliser une transcription automatique des enregistrements grâce à l'intelligence artificielle avec un taux d'erreurs proche de 20%. Ce projet est une collaboration du Dr. Madikeri (Computerlinguistik, UZH) et la Dr. Dherbey Chapuis.

2024-2027 *Production de ressources pédagogiques pour les élèves allophones de 1H et 2H*, (Sept.2024-Dec.2027); Financement OFC-BAK Art.10 (Ref. 95123, 193'528CHF, N. Dherbey Chapuis). Ce projet est une recherche collaborative avec les enseignants du Quartier du Schönberg, Fribourg (cofinancement OFC, SEnOF, CIIP). Il vise à produire des ressources adaptées aux besoins des élèves FLS en 1-2H à partir des analyses linguistiques du corpus FrOMi.

2023-2028 Thèse de N. Fèlix sous la Direction du Prof. R. Berthele.

2021-2023 *Sons de la langue de scolarisation : du défi des premières heures à la lecture*, Financement Institut de Plurilinguisme ; 18'000CHF ; Responsable du projet N. Dherbey Chapuis <https://institut-plurilinguisme.ch/fr/personnel/dherbey-chapuis-nathalie>

2021-2022 *Observer et analyser le processus d'acquisition du français oral à l'école par des élèves allophones* ; Financement Fond du Centenaire (Ref. 835) ; 6'587CHF ; N. Dherbey Chapuis.

Communications

Colloques

2024 07/09 **CLARIN-CH Day**, Université de Neuchâtel, Suisse. *Challenges in Utilizing Automatic Speech Recognition for a Corpus of Young L2 French Learners (FrOMi)*. Oral communication. Dherbey-Chapuis, N.

2023 09-10/11 **Tagung Mehrsprachigkeit, PHGR**, Davos, Suisse. *Code-switching during oral learning L2 French at school*. Oral communication. Dherbey-Chapuis, N.

- 2022** 7/11 **12^e Journées de Linguistique Suisse**, Lausanne, Suisse. Projet : *Apprentissage oral du français langue seconde*. Projet de Master : *Acquisition de l'accord sujet-verbe chez des élèves (1H) scolarisés sans connaissance du français*. Poster. N. Félix.
- 2022** 7/11 **12^e Journées de Linguistique Suisse**, Lausanne, Suisse. *Oral development of L2 French at school by 5 year old pupils*. Oral communication. Dherbey-Chapuis, N.

Travaux d'étudiant-e-s

- 2023** *Morphosyntaxe de l'accord verbal chez les élèves allophones de cinq ans scolarisés sans connaissance du français*. Thèse de master soutenue par Nicolas Félix, Université de Fribourg, Suisse
- 2023** *La (non) réalisation du schwa dans les clitiques ce, de, je, le, me, ne, se, te et que : clitiques seuls et doubles*, Gabriela Luthi, travail de séminaire « Français parlé » sous la direction de Sandra Schwab, Master en langues et littérature françaises FLE, Université de Bern

Intervenants

- ❖ **Conception du corpus, demandes d'autorisation, demandes de fonds, direction du projet** : N, Dherbey Chapuis
- ❖ Ont participé pour contribuer à la **recherche de fonds** :
 - Ecriture des demandes de fonds : N. Dherbey Chapuis
 - Etude des sources de financement via les fondations : N. Félix
 - Relecture des demandes de fonds : I. Affolter, R. Berthele, E. Dutton, N. Félix, A. Thomas, S. Schwab.
- ❖ **Recueil des données sur le terrain**
 - Contact avec les intervenants en milieu scolaire : N. Dherbey Chapuis
 - Autorisations SEnOF : N. Dherbey Chapuis
 - Recueil des données :
 - N. Dherbey Chapuis : 133 BS
 - N. Félix : 4 BS
- ❖ **Transcriptions**
 - C. Gilgien, N.Félix (2021-2022)
 - Lisa Schneider (2021-2025)
- ❖ Ont participé au titre de leur recherche personnelle :
 - *Transcriptions* :
 - Gabriela Luthi : Lena8 et Lena10 pour le T1 (2BS)
 - Nicolas Félix : Lena8 pour le T2 (1BS)
 - Nathalie Dherbey Chapuis : Lena2, Lena3, Lena4 pour le T1 (3BS)
 - *Vérifications* :
 - Nathalie Dherbey Chapuis : Lena1, Lena2, Lena3, Lena4, Lena5, Lena8, Lena9, Lena10 pour le T1 (8BS) + Lena7, Lena3 pour le T2 (2BS)

Méthode

Une transcription par réplique d'interaction

Comme les apprenant-e-s étaient de grands débutant-e-s au premier enregistrement les phrase n'étaient pas structurées correctement et difficilement reconnaissables. Pour vérifier que tout ce qui faisait sens dans l'échange communicatif était transcrit, les séquences d'interaction ont été identifiées.

Au premier enregistrement, les ouvertures de séquences d'interaction ne sont pas représentées par des éléments linguistiques mais peuvent être repérées par l'adresse à une personne, une question posée (identifiée soit par sa prosodie, soit par l'attente d'une réponse). La partie centrale de la séquence correspond au développement du thème ou à la répétition d'une demande ou d'une remarque. La fin de la séquence peut être liée à l'intervention d'un tiers ou le changement de thème. Il n'y a pas de fermeture de la séquence (p.ex. « aurevoir »...). Cette méthode sera complétée au fil de l'acquisition par le repérage des indicateurs d'ouverture et de fermeture de séquence d'interaction.

Codage des transcriptions

On utilise l'orthographe canonique du mot le plus probable

- ajout **1** après un phonème **mal prononcé**:
Ex réalisation de [bouton] ; pour mouton -> on écrit <m1outon>
- ajout **2** après un phonème **tronqué**
Ex réalisation de [mout] ; pour mouton -> on écrit mout**on2**
Ex réalisation de [mou-on] ; pour mouton -> on écrit mout**2on**
- Pour la syllabe **tronquée**:
 - Si le contexte permet de la deviner Ex réalisation de [mou] ; pour mouton -> on écrit mou<t2<on2>
 - Si le contexte ne permet pas de la deviner mais qu'on sent que le locuteur voulait dire quelque chose à suivre : Ex réalisation de [mou] -> on écrit mouXXX
- Mot ou syllabe non reconnaissable : XXX
- **Discours en langue étrangère** : XXX09 XXX09 XXX09 ; mots isolé en LE : XXX09
- Mots de langue étrangère : **09** ; Ex go09 casa09
- **Les accents** sont écrits normalement
- **Pas de majuscules**
- Si vous mettez de la ponctuation **laissez un espace** avec le mot qui précède et celui qui suit
- AJOUT d'un phonème :
 - Tentez de vous rapprocher d'une forme existante : EX réalisation de /atãd/ : si un interlocuteur =attendl1s ; si deux interlocuteurs =attendez2
 - Si c'est impossible de se rapprocher d'un mot existant comme « une z-ailes » -> z&ailes
« fileur » -> fi&leur
- Lorsque l'enfant répète on colle une parenthèse à la fin de chaque mot répété: (répète : un) deux)
- Lettre E à n'utiliser que quand on ne peut pas savoir si le e correspond à un infinitif ou non et é **éviter le plus possible**
- Ne pas coder la prononciation de la maîtresse
- Attention /ija/ s'écrit 'il2 y a' ; mais /ja/ s'écrit 'y a'. On **n'ajoute pas** des mots pour être plus correct.

Anonymisation

Mettre la dernière syllabe du prénom en verlan et les faire suivre de ~

Tableau récapitulatif

Phénomène	Correct ou incorrect	Code	Exemple
phonème mal prononcé	incorrect	1	m1outon
phonème tronqué	incorrect	2	/mut/ mout on2 -> mouton 2 ; /muõ/ mout 2on
syllabe tronquée	incorrect	8 changement	Mout 2on2
Mot non reconnaissable		XXX	
Mot de langue étrangère		09	Come09

Si 2 codes à suivre : mettre l'indication après le phonème concerné (p.ex. '2') puis le code mot étranger (p.ex.09).

Ex wha(t) this_> what209

Phrases, pauses, hésitation

Ponctuations autorisées : « . ? ! , » **toujours mettre un espace avant et après**

- **Mettre un point** avec un espace avant et un espace après : quand l'intonation ne laisse aucun doute que la phrase est terminée et quand la pause entre les mots (fin d'une phrase et début de l'autre) est supérieure à 600ms.
- « ? » pour une intonation montante de question (le mot interrogatif comme « **QUI** c'est qui ... » ne rentre **pas** dans le critère de décision).
- « ! » pour une intonation d'un ordre donné
- « , » marque une pause réalisée par le locuteur dont la durée est supérieure à la pause réalisée entre les mots. **Ne marquer les virgules que quand la pause est réalisée et si elle facilite la compréhension du message.**

Traits d'union et apostrophe et autres particularités

On respecte l'orthographe française :

Là-bas, parce que, aujourd'hui

Qu'est-ce que c'est ; vas-y ; va-t'en

C'est

1H = première Harmos ; 2H= deuxième Harmos ; 3H= troisième Harmos

VT = Villa Thérèse

S.~ = Schoenberg

Participants

Voir Tableau page suivante

ID Participant	Genre	Age at recording (start, Months)	Classroom	French before School	Length in Switzerland (Months)	L1	L2	L3
Lena1	F	62	1	No	8	Arabic	Turkish	
Lena2	M	65	1	Yes	24	French	Krio	English
Lena3	M	60	1	No	60	Kurdish		
Lena4	M	67	1	No	8	Romanian	English	
Lena5	F	54	1	No	48	Tigrinya		
Lena12	F	64	2	Intermediate	64	Portuguese	French	
Lena6	F	54	2	No	54	Tigrinya	English	
Lena7	M	67	2	No	67	Kurdish	Turkish	
Lena8	F	61	2	No	61	Portuguese		
Lena9	M	67	2	Some words	55	Turkish	French	
Lena10	M	61	3	No	49	Portuguese		
Lena11	F	60	3	No	48	Amharic	Italian	
Lena13	M	55	3	Yes	55	Albanian	French	

