



HAL
open science

Comment construire un corpus multimodal cohérent : Etude de la variabilité de la transcription d'un extrait filmé par des transcripateurs de différentes langues maternelles

Pascale Leclercq, Amanda Edmonds, Elisa Sneed German, Emily Christen
Paige, Sandra Heyde, Vladyslav Gilyuk, Katia Makhlouf

► To cite this version:

Pascale Leclercq, Amanda Edmonds, Elisa Sneed German, Emily Christen Paige, Sandra Heyde, et al.. Comment construire un corpus multimodal cohérent : Etude de la variabilité de la transcription d'un extrait filmé par des transcripateurs de différentes langues maternelles. Colloque International du Réseau d'Acquisition des Langues Secondes (Réa2), Jul 2021, Toulouse, France. hal-03303903

HAL Id: hal-03303903

<https://hal-univ-montpellier3-paul-valery.archives-ouvertes.fr/>

hal-03303903

Submitted on 28 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comment construire un corpus multimodal cohérent :

Etude de la variabilité de la transcription d'un extrait filmé par des transcrip-teurs de différentes langues maternelles

P. Leclercq, A. Edmonds, E. Sneed German, E. Paige, S. Heyde, V. Gilyuk, K. Makhlouf

EMMA - Université Paul Valéry Montpellier 3



La transcription

(Ochs 1979, Mondada 2000, 2002, 2007, Saturno 2017, Leclercq 2020, Nagy & Sharma 2013, Watorek & Benazzo 2020)

- Etape cruciale de la constitution d'un corpus oral
- 1^{ère} étape de l'analyse et de l'interprétation des données: processus de filtrage subjectif, informé par des choix théoriques plus ou moins conscients, mais aussi par les possibilités offertes par le logiciel utilisé et par les conventions de transcription choisies.

→ Variabilité inhérente à l'exercice



Le projet PROLINGSA (Linguistic Progress during Study Abroad)

- Etude longitudinale de 12 mois visant à caractériser l'impact du séjour à l'étranger sur le développement linguistique.
- 5 étudiants francophones Erasmus+ en Angleterre ou Irlande + 2 investigatrices (1 américaine, 1 française)
- 5 entretiens semi-guidés filmés au cours des 12 mois. Durée totale: >12h
- Transcription sous CLAN par une équipe de 9 transcrip-teurs, d'expertise variable, et de 4 L1 différentes (anglais, français, kabyle, ukrainien)
- Chaque transcription est relue par un membre expert de l'équipe (vérification/harmonisation)
- Mise en ligne sur [Ortolang](#)

Questions de recherche

- Quelles sont les éléments sujets à variation parmi les transcriptions d'enregistrements en français L2 par une équipe plurilingue?
- Comment faire en sorte que la variabilité dans les transcriptions ne soit pas un obstacle à la construction du corpus?
- Comment valoriser la variation et en faire un moteur pour la recherche et la formation à la recherche?

Dispositif expérimental

- Transcription d'un extrait du corpus (vidéo, NT5, 2mn45)
- Journal de bord rempli par les transcrip-teurs
- 6 transcrip-teurs
- Transcription sous CLAN

Locuteur	Genre	L1	Formation	Durée transcription	Expertise CLAN
LN1	F	Anglais US (stress-timed)	Enseignant-chercheur	43mn	++
LN2	F	Anglais US (stress-timed)	Enseignant-chercheur	1h25	++
LN3	F	Anglais US (stress-timed)	Etudiante en Master	2h30	+
LNN1	F	Kabyle	Etudiante en Master	5h	-
LNN2	F	Français (syllable-timed)	Enseignante	1h34	+
LNN3	M	Ukrainien	Etudiant en Master	1h55	+

Résultats

1. Variation dans les transcriptions en fonction des centres d'intérêts et de l'expertise des transcrip-teurs

- Longueur de la transcription (nombre de mots moyen = 376,5 [361-384])
- Segmentation en énoncés & types de lignes (tiers des locuteurs, commentaires phonologiques, situationnels...)
- Transcription +/- détaillée des *gap fillers*, *backchannels*...
- Utilisation des codes de transcription de CLAN (notamment répétitions, auto-corrrections, annotation des gestes...)



2. Désaccords entre transcriptions

- Au total, 31 cas de désaccords entre les transcriptions
 - 1 réinterprétation syntaxique / 8 désaccords lexicaux / 21 désaccords dans l'interprétation des formes faibles

Locuteur	Transcrip-teur					
	LN1	LN2	LN3	LNN1	LNN2	LNN3
N	to go back in [*] school	for me to go back to school	for me to go back Op rep school	to go ba:ck (.) school	to go back to school	it's cool
INT2	this has given you perspective	this has given you a new perspective	this has given you perspective:e	this is getting you a perspective	this is giving you perspective	this is giving you perspective
N	compar ed to others to actually help	compar ed to others to actually help people	compar ing [*] to others to actually help ?	compare it to others	compar ing to others	compar ing to others to actually helped[*]
INT2	what your plans are for after that	what your plans are for after that	what your plans are for after that	what are your plans for after that	what your plans are for after that	what your plans are for after that

Discussion

La transcription collaborative appelle:

- Une formation des transcrip-teurs aux enjeux de la transcription;
- Une sensibilisation des membres de l'équipe à l'impact des choix de transcription sur les possibilités d'analyse des données;
- L'explicitation transparente des choix effectués;
- Une prise de conscience de la nécessité de prêter une attention particulière aux formes inaccentuées.



ÉTUDES
MONTPELLIÉRAINES
DU MONDE
ANGLOPHONE



Références

Benazzo, Sandra & Watorek, Marzena (2020). Transcriptions de corpus oraux d'apprenants débutants en français L2 : quelques enjeux théoriques. Dans Lorenzo Spreafico, Giuliano Bernini, Ada Valentini & Jacopo Saturno (Eds.), *L'evanescenza del parlato. Un vademecum per il trattamento digitale di dati linguistiche*. Bergamo: Sestante.

Leclercq, P. (2020). Transcribing interlanguage: The case of verb-final [e] in L2 French. In Edmonds, A., Leclercq, P. & Gudemestad, A. (Eds), *Interpreting language-learning data*, pp. 169-197. Zenodo. <https://doi.org/10.5281/ZENODO.4032298>

Nagy, N. & Sharma, D. (2013). *Transcription*. In Robert J. Podesva & Devyani Sharma (eds), *Research Methods in Linguistics*. Cambridge University Press.

Mondada, L. (2000). Les effets théoriques des pratiques de transcription. *Linx*, 42, 131–146. <https://doi.org/10.4000/linx.902>

Mondada, L. (2002). Pratiques de transcription et effets de catégorisation. *Cahiers de praxématique*, 39, 45-75.

Mondada, L. (2007). Commentary: Transcript variations and the indexicality of transcribing practices. *Discourse Studies*, 9(6), 809–821. <https://doi.org/10.1177/1461445607082581>

Edmonds, Amanda, Leclercq, Pascale, & Gudmestad, Aarnes. (2020). *Interpreting language-learning data*. Zenodo. <https://doi.org/10.5281/ZENODO.4032298>

Ochs, Elinor. (1979). Transcription as theory. *Developmental pragmatics* 10(1). 43–72.

Saturno, Jacopo. (2017). *Inflectional morphology and utterance structure in initial varieties of Polish L2*. Bergamo / Paris: Università degli studi di Bergamo & University of Paris 8. (Doctoral dissertation).

Contact: pascale.leclercq@univ-montp3.fr