



PROLINGSA



Corpus PROLINGSA

De la création à l'analyse

**Amanda EDMONDS^a, Pascale
LECLERCQ^b, Elisa SNEED GERMAN^b**

^aUniversité Côte d'Azur

^bUniversité Paul Valéry Montpellier 3

Quelques mots d'introduction

- Dans les deux dernières décennies, de nombreux travaux sur l'influence du contexte d'apprentissage (dont le séjour en immersion) sur l'acquisition d'une langue seconde ont vu le jour (Barbara Freed, Martin Howard, Robert DeKeyser, etc.)
- En même temps, plusieurs voix s'élèvent en faveur des méthodologies longitudinales (Lourdes Ortega, Celeste Kinginger, etc.) et l'intérêt de faire des études de cas détaillées (e.g., Patricia Duff)

PROLINGSA : pourquoi ?

- COST Action 15130 (2016-2020)
- Créer un corpus longitudinal qui permet d'étudier le développement langagier au cours d'un séjour en immersion de plusieurs angles
 - morphologiques, syntaxiques, discursifs, lexicaux, phonologiques, etc.
- Créer un outil qui pourra être exploité par nos étudiants (notamment en Master)
- À terme, rendre le corpus disponible à la communauté scientifique
 - <https://www.ortolang.fr/market/corpora/prolingsa>

Plan de la présentation

1. Recueil des données: présentation du corpus PROLINGSA et remarques techniques
2. Transcription collaborative: Enjeux méthodologiques
3. Mise en ligne sur Ortolang: enjeux éthiques
4. Annotation des données: étude des emplois de *don't* dans le corpus PROLINGSA



Présentation du corpus

Collecte des données

- Participants

Oxford Quick Placement Test

Participant	Gender	Age	Languages		Pre-stay	Post-stay	Country
			(specialization)	L1			
A	F	19	English + Italian	French	Lower intermediate	Upper intermediate	Ireland
C	F	18	English + Chinese	French	Lower intermediate	Upper intermediate	Ireland
M	F	19	English + Spanish	French	Advanced	Advanced	Ireland
N	M	19	English + Chinese	French	Advanced	Advanced	Ireland
Y	M	19	English + Arabic	French + Turkish	Advanced	Advanced	England

Présentation du corpus

Collecte des données

- Protocole

Séance	Tâches
1 (juin 2018)	Entretien semi-guidé Oxford proficiency test Lex30
2 (octobre 2018)	Entretien semi-guidé
3 (février 2019)	Entretien semi-guidé
4 (mars 2019)	Entretien semi-guidé
5 (juin 2019)	Entretien semi-guidé Oxford proficiency test Lex30

Présentation du corpus

Corpus final

- Caractéristiques du corpus final

Partici pant	Séance 1 (juin 2018)		Séance 2 (novembre 2018)		Séance 3 (janvier 2019)		Séance 4 (mars 2019)		Séance 5 (juin 2019)		TOTAL
	durée	# mots	durée	# mots	durée	# mots	durée	# mots	durée	# mots	# mots
A	27:00	1240	35:03	4008	24:38	2686	17:13	1837	34:16	3703	13,512
C	26:49	1093	26:40	1901	12:18	903	12:51	994	23:27	1219	6,118
M	18:34	1048	15:39	1380	13:47	1309	15:20	1433	20:51	1999	7,167
N	23:55	1318	44:24	4397	23:40	2038	22:12	2071	37:21	3394	13,232
Y	44:24	3490	56:11	6899	48:20	6418	43:48	4867	53:47	7107	28,790



Présentation du corpus

Collecte des données

- Quelques remarques techniques
 - Réfléchir au cadrage
 - Bien connaître son matériel !

Présentation du corpus

Collecte des données

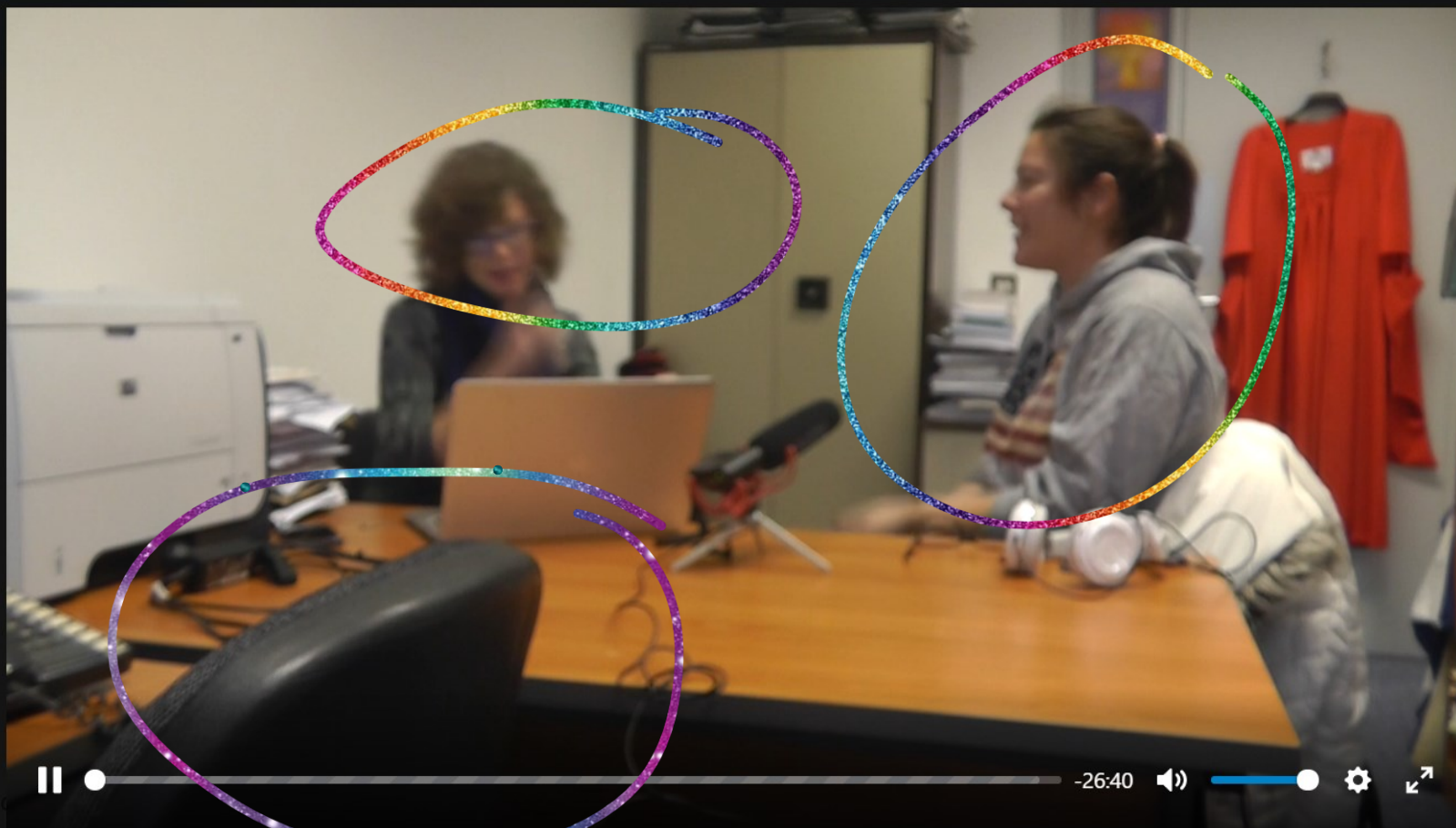
Prolingsa_T2_C.mp4



Présentation du corpus

Collecte des données

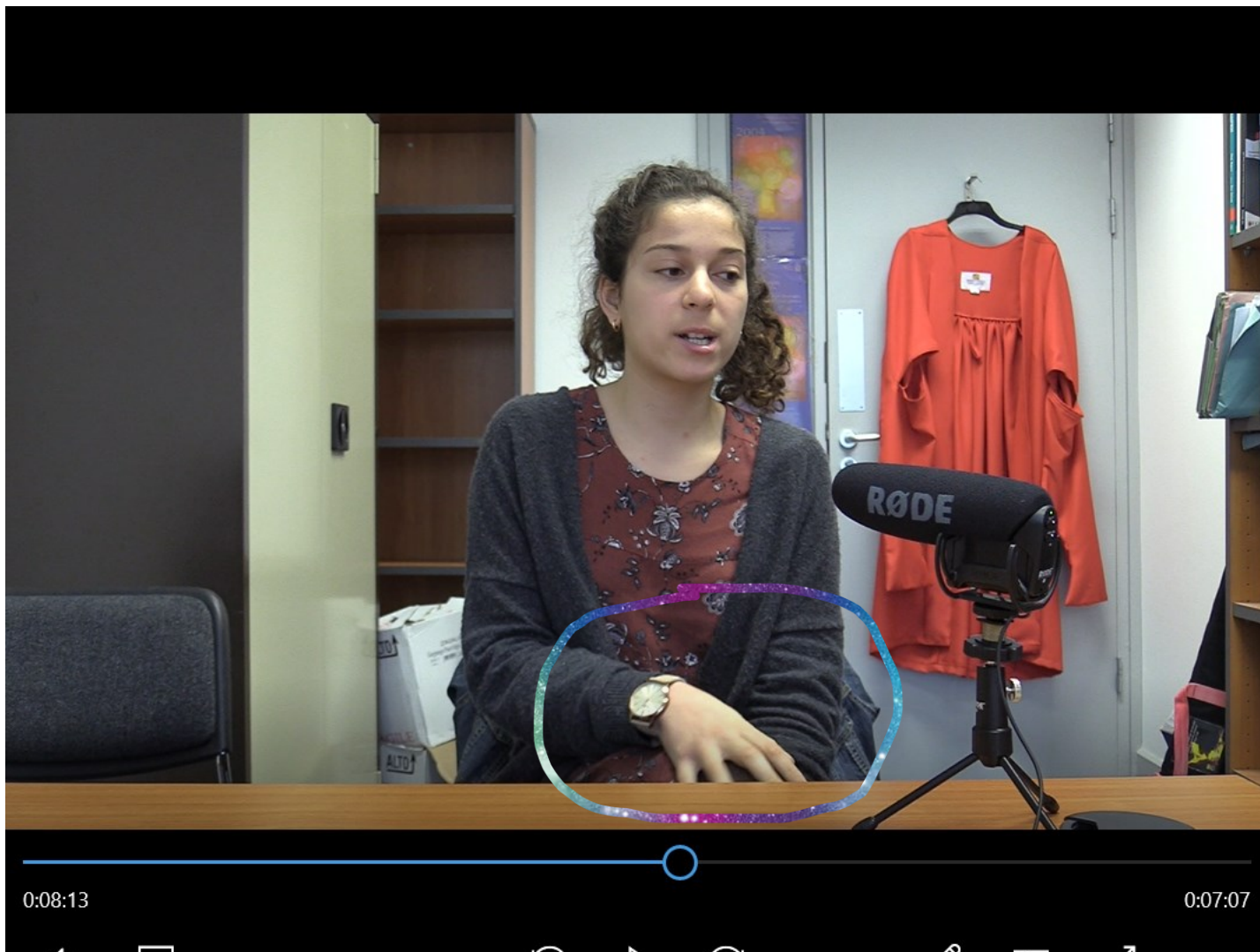
Prolingsa_T2_C.mp4



Size
2.1 GB
1.6 GB
2.7 GB
58.5 MB
29.9 MB
7.7 GB

Présentation du corpus

Collecte des données



Présentation du corpus

Transcription collaborative

- Transcription collaborative
 - Transcription à l'aide de CLAN
 - Chaque entretien a été transcrit par un transcripteur et relu / corrigé par au moins un autre membre de l'équipe

```

clan - [Anouk_T4]
File Edit View Tiers Mode Window Help
@Begin
@Languages:  eng, fra, ita
@Participants:  A Target_Adult, INT1 Pascale Investigator
@ID:  eng, fra, ita|anouk_time 2|A|22;[female]||Target_Adult||
@ID:  eng, fra, ita|change_corpus_later|INT1||||Investigator||
@Media:  A_Time_4, video
@Transcriber:  TJ
*A:  ok so thank anouk for coming here . .
*INT1:  ok so thank you anouk for being here . .
*A:  you 're welcome . .
*INT1:  in this nice guest house . .
*INT1:  so &=laughs this is our last &uh on-site interview . .
*INT1:  and then we 'll have one last one . .
*INT1:  when you come back to montpellier . .
*INT1:  but i 'll talk to you about it later . .
*A:  ok . .
*INT1:  ok so spring is coming . .
*A:  yeah . .
*INT1:  at last . .
*A:  ireland spring . .
*INT1:  yeah . .
*INT1:  so how do you feel after your first winter in a rainy country ? .
*A:  &pff <i> [ / ] &=laughs i feel cold . .
*INT1:  yeah . .
*A:  a lot . .
*A:  but i kind of [ / / ] it 's not raining everyday . .
*A:  it just depends like . .
*A:  it sometimes can rain like . .
*A:  i don 't know like ten times a day . .
*A:  for like ten days <i> 's > [ / ] it 's never ending . .
*INT1:  mmhm@i . .
*A:  and sometimes <we have> [ / ] we have like . .
09oct19[E[CHAT] 1
  
```



Présentation du corpus

Transcription collaborative

- Transcription collaborative
 - Réflexion méthodologique sur la construction d'un corpus multimodal

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. Makhlof
 EMMA - Université Paul Valéry Montpellier 3

La transcription

- Etude longitudinale de 12 mois visant à caractériser l'impact du séjour à l'étranger sur le développement linguistique.
- 5 étudiants francophones. Français en Angleterre ou l'inverse + 2 investigatrices (1 américaine, 1 française).
- 5 entretiens semi-structurés filmés au cours des 12 mois. Durée totale : >1h.
- 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 revue par un membre expert de l'équipe (vérification/harmonisation).
- Mise en ligne sur [CLANbase](#).

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. Français en Angleterre ou l'inverse + 2 investigatrices (1 américaine, 1 française).
- 5 entretiens semi-structurés filmés au cours des 12 mois. Durée totale : >1h.
- 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 revue par un membre expert de l'équipe (vérification/harmonisation).
- Mise en ligne sur [CLANbase](#).

Questions de recherche

- Quelles sont les éléments sujets à variations parmi les transcriptions.
- 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, NTS, Zimac)
- Journal de bord rempli par les transcrip-teurs
- 6 transcrip-teurs
- Transcription sous CLAN

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 moyens > 376,3 (Stk: 384))
- Segmentation en énoncés & types de lignes (lignes des locuteurs, commentaires phonologiques, situationnels...)
- Transcription et/ou détection des gaz (fiers, backchannels...)
- Utilisation des codes de transcription de CLAN (notamment répétitions, auto-corrections, annotation des gestes...)

2. Décalages entre transcriptions

Au total, 31 cas de décalages entre les transcriptions

- 1 interprétation syntaxique / 8 décalages lexicaux / 21 décalages dans l'interprétation des formes faibles

Locuteur	LN1	LN2	LN3	LN1L	LN2L	LN3L
N	to go back [N*]	for me to go back to school	for me to go back to school	to go ba ck () school	to go back to school	it's cool
NT2	this has given you a perspective	this has given you a new perspective	this has given you a perspective	this is getting you a perspective	this is giving you perspective	this is giving you perspetive
N	compared to others to actually help people	compared to others to actually help people	comparing it to others to actually help?	comparing it to others	comparing to others	comparing to others to actually helped[?]
NT2	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	what your plans are for after that

Discussion

La transcription collaborative apporte :

- 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 inaccoutées.

Pascale Leclercq, Amanda Edmonds, Elisa Sneed German, Emily Paige, Sandra Heyde, et al.. 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. Colloque International du Réseau d'Acquisition des Langues Secondes (Réal2), Juillet 2021, Toulouse, France. (hal-03303903)

Présentation du corpus

Transcription collaborative



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

Transcription au sein du projet PROLINGSA

- 5 étudiants francophones Erasmus+ en Angleterre ou Irlande + 2 investigatrices (1 américaine, 1 française)
- Transcription sous CLAN par une équipe de 9 transcripteurs, 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)

Pascale Leclercq, Amanda Edmonds, Elisa Sneed German, Emily Paige, Sandra Heyde, et al.. Comment construire un corpus multimodal cohérent : Etude de la variabilité de la transcription d'un extrait filmé par des transcripteurs de différentes langues maternelles. Colloque International du Réseau d'Acquisition des Langues Secondes (Réal2), Juillet 2021, Toulouse, France. (hal-03303903)

Présentation du corpus

Transcription collaborative

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)	Étudiante en Master	2h30	+
LNN1	F	Kabyle	Étudiante en Master	5h	-
LNN2	F	Français (syllable-timed)	Enseignante	1h34	+
LNN3	M	Ukrainien	Étudiant en Master	1h55	+

Pascale Leclercq, Amanda Edmonds, Elisa Sneed German, Emily Paige, Sandra Heyde, et al.. 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. Colloque International du Réseau d'Acquisition des Langues Secondes (Réal2), Juillet 2021, Toulouse, France. (hal-03303903)

Présentation du corpus

Transcription collaborative

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	Transcripteur					
	LN1	LN2	LN3	LNN1	LNN2	LNN3
N	to go back in[*] school	for me to go back to school	for me to go back Oprep 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 perspectiv:e	this is getting you a perspective	this is giving you perspective	this is giving you perspective
N	compared to others to actually help	compared to others to actually help people	comparing[*] to others to actually help ?	compare it to others	comparing to others	comparing 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

Pascale Leclercq, Amanda Edmonds, Elisa Sneed German, Emily Paige, Sandra Heyde, et al.. 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. Colloque International du Réseau d'Acquisition des Langues Secondes (Réal2), Juillet 2021, Toulouse, France. (hal-03303903)

Transcription collaborative: conclusions

- Besoin d'un manuel de transcription au sein du projet;
- Nécessité de former les transcrip-teurs aux enjeux de la transcription, notamment sur l'impact des choix de transcription au moment de l'analyse;
- Nécessité d'explicitier de façon transparente les choix effectués.

Mise en ligne sur ORTOLANG

<https://www.ortolang.fr/market/corpora/prolingsa>

The screenshot displays the ORTOLANG website interface. At the top, the ORTOLANG logo is on the left, and navigation links for 'Informations', 'Langue', and 'Conn' are on the right. A search bar with the placeholder 'Rechercher un corpus' and a magnifying glass icon is positioned below the header. On the left side, a vertical sidebar contains navigation icons and labels: 'Accueil', 'Corpus', 'Lexiques', 'Terminologies', 'Outils', 'Projets Intégrés', 'Actualités', 'Informations', and 'Producteurs'. The main content area features a grid of ten corpus cards. Each card includes an image, a title, and a date. The cards are: 'Signes en famille' (18 oct. 2021), 'CLeLPC' (4 oct. 2021), 'PROLINGSA' (29 juil. 2021), 'CLM-MOCAP' (9 juin 2021), 'Présences numériques' (7 avr. 2021), 'CorpAGEst' (Speech and gesture in later life), 'Autisme-Dascalu' (Au), 'Vidéos du corpus CID' (Vi), 'Dicta-Sign-LSF-v2' (Dicta-Sign LSF V2), and 'Diners familiaux' (Di).

ORTOLANG

Informations Langue Conn

Rechercher un corpus

Accueil

Corpus

Lexiques

Terminologies

Outils

Projets Intégrés

Actualités

Informations

Producteurs

Signes en famille

18 oct. 2021

CLeLPC

4 oct. 2021

PROLINGSA

29 juil. 2021

CLM-MOCAP

9 juin 2021

Présences numériques

7 avr. 2021

CorpAGEst

Speech and gesture in later life

Au

Autisme-Dascalu

Vi

Vidéos du corpus CID

Dicta-Sign-LSF-v2

Dicta-Sign LSF V2

Di

Diners familiaux

- Accueil
- Corpus**
- Lexiques
- Terminologies
- Outils
- Projets Intégrés
- Actualités
- Informations
- Producteurs

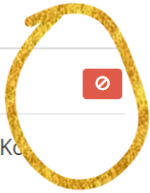


PROLINGSA

[Retourner à la fiche](#)

prolingsa - Grid Sort Search Settings

Nom ↑	Type	Dernière modification	Taille
PROLINGSA_Donnees	collection	29/07/2021 21:47	-
PROLINGSA_Presentation	collection	17/03/2021 10:08	-
PROLinsa_ARCHIVES	collection	09/04/2021 16:51	-
PROLINGSA.png	image/png	24/02/2021 17:02	121,3 Ko





Enjeux liés à la mise en ligne sur Ortolang

- Enjeux éthique: diffusion d'infos personnelles?
- Anonymisation des données (bip!)
- Garder les données pour soi... ou les diffuser???



Étude sur les emplois de *don't*

- Edmonds, A., Sneed German, E. & Leclercq, P. (2021). A phraseological and phonological analysis of *don't*: A stay abroad perspective. *Lexis* [Online], 18.

Étude sur les emplois de *don't*

- Background
 - Ce que nous savons déjà sur l'unité lexicale sous étude : *don't*
 - Bybee and Scheibman (1999)
 - les emplois de *don't* sont fortement restreints à quelques régularités lexicales
 - Sujets : 63,77% **I**
 - Verbes : 28,26% **know** / 14,49 % *think* / 6,52 % *have*
 - les combinaisons fréquentes – dont *I don't know* – montrent généralement de la réduction phonologique
 - Pichler (2009)
 - Vérification de la co-occurrence de la nature phraséologique (avec fonction pragmatique) de l'emploi et son sa réalisation phonologique

Étude sur les emplois de *don't*

- Background
 - Développement de la compétence phraséologique lors d'un séjour en immersion
 - Marqueurs discursifs : “linguistic items or expressions that have little or no referential meaning but serve multiple pragmatic functions in the interpersonal and textual domains of discourse.” (Pichler, 2009, p. 561)
 - Sous-utilisation, en comparaison aux locuteurs natifs (Fuller, 2003; Sankoff et al., 1997)
 - Sur-utilisation d'un nombre restreint de marqueurs discursifs (Fuller, 2003)
 - Effet significatif du séjour en immersion (Tavakoli, 2018)
 - Corrélation significative entre l'utilisation des marqueurs discursifs et les contacts avec des locuteurs natifs (Hellerman & Vergun, 2007; Liao, 2009)

Étude sur les emplois de *don't*

- Questions de recherche
 1. Quelles régularités distributionnelles caractérisent les emplois de *don't* dans le corpus PROLINGSA ?
 2. Pour la séquence *I don't know*, comment son emploi en tant que marqueur discursif et sa réalisation phonétique évoluent-ils au cours d'un séjour en immersion ?

Étude sur les emplois de *don't*

- Méthodologie
 1. Réfléchir à l'analyse à effectuer
 - Trop peu de participants pour des analyses statistiques
 - Proposer des analyses descriptives riches (tradition des études de cas, Duff, 2014)
 2. Identifier toutes les instances de *don't* dans le corpus ($n = 844$)
 - les [coder](#) pour des aspects de l'analyse phraséologique (deux chercheurs)
 - procéder à l'analyse phonologique (limitée à deux participants / 3 entretiens) (une chercheur)

Participant	Interview	Time	Context	Form	Subject	V	function	phraseo	Closure	Burst	Vowel length	Consonant
Y	3	10:56	&uh but &uh yeah i don 't know at the attachment or &uh &uh i don 't know it 's maybe like	don't	I	know	pragmatic	y	y	n	0,018139469	0,031510695
Y	5	30:11	at the same thing it 's contrasted[*] so i just don 't know how to &uhm> [//] i don 't know what the outcome would[*] be	don't	I	know	literal	n	y	n	0,018638062	0,042123839
Y	5	17:34	and <there w(asn)> [//] there [x 2] wasn 't really communication with him &uhm i don 't know if <it was> [///] i think	don't	I	know	pragmatic (meaning "maybe"??)	n	y	n	0,019886991	0,066520468
Y	1	10:21	o like i 'm not like doing one thing <th(irty)> [/] i dunno thirty hours a week	don't	I	know	pragmatic	y	n	n	0,022326713	NO_C
Y	5	20:04	i 'm seventy percent birtish and thirty percent american because i don 't know it 's just maybe> [///] i don 't know i feel like	don't	I	know	pragmatic	y	y	n	0,022728103	0,033826006
Y	5	25:16	sometimes you just kind of don 't know where you are &uhm but yeah	don't	you	know	literal		y	y	0,024449087	0,112910935
AE	2	10:53	o i don 't know how it works at DIT	don't	I	know	literal	n	y	n	0,024505958	
Y	3	43:00	&uh but yeah identity &uh yes [x 2] because &uh i don 't know how to explain	don't	I	know	literal	n	y	y	0,024780072	0,048891838
Y	1	23:15	and <i really liked> [//] i really like &um: &=click tongue i dunno the [x 2] fact	don't	I	know	pragmatic	y	n	n	0,025113819	NO_C

Étude sur les emplois de *don't*

- Quelques résultats (QR1)
 - Régularités distributionnelles
 - Emplois de *don't* avec le sujet *I* : (87.7%, 644/734)
 - Emplois de *don't* avec les trois verbes les plus fréquents :
 - *don't know* ($n = 424$) : 57,77%
 - *don't have* ($n = 66$) : 8,92%
 - *don't think* ($n = 41$) : 5,59%

Participant	Interview	Time	Context	Form	Subject	V	function	phraseo	closure	Burst	Vowel length	Consonant
Y	3	10:56	&uh but &uh yeah i don 't know at the &uh i don 't know it 's maybe like attachment or &uh	don't	I	know	pragmatic	y	y	n	0,018139469	0,031510695
Y	5	30 11	at the same thing it 's contrasted[*] so <i just don 't know how to &uhm> [//] i don 't know what the outcome would[*] be	don't	I	know	literal	n	y	n	0,018638062	0,042123839
Y	5	17:34	and <there w(asn)> [//] there [x 2] wasn 't really communication with him &uhm i don 't know if <it was> [///] i think	don't	I	know	pragmatic (meaning "maybe"??)	n	y	n	0,019886991	0,066520468
Y	1	10:21	so like i 'm not like doing one thing <th(irty)> [/] i dunno thirty hours a week	don't	I	know	pragmatic	y	n	n	0,022326713	NO_C
Y	5	20:04	i 'm seventy percent birtish and thirty percent american because i don 't know <it 's just maybe> [///] i don 't know i feel like	don't	I	know	pragmatic	y	y	n	0,022728103	0,033826006
Y	5	25 16	sometimes you just kind of don 't know where you are &uhm but yeah	don't	you	know	literal		y	y	0,024449087	0,112910935
AE	2	10:53	so i don 't know how it works at DIT	don't	I	know	literal	n	y	n	0,024505958	
Y	3	43 00	&uh but yeah identity &uh yes [x 2] because &uh i don 't know how to explain	don't	I	know	literal	n	y	y	0,024780072	0,048891838
Y	1	23:15	and <i really liked> [//] i really like &um: =&clicks:tongue i dunno the [x 2] fact	don't	I	know	pragmatic	v	n	n	0,025113819	NO_C

Étude sur les emplois de *don't*

- Quelques résultats (QR2)

Tableau 1. Emplois de *I don't know* en tant que marqueur discursif.

	A		C		M		N		Y	
Entretien	#	per 1K	#	per 1K	#	per 1K	#	per 1K	#	per 1K
1	5		1		2		2		13	
2	5		–		1		3		26	
3	4		–		6		2		29	
4	6		1		8		1		29	
5	26		–		6		9		48	

Étude sur les emplois de *don't*

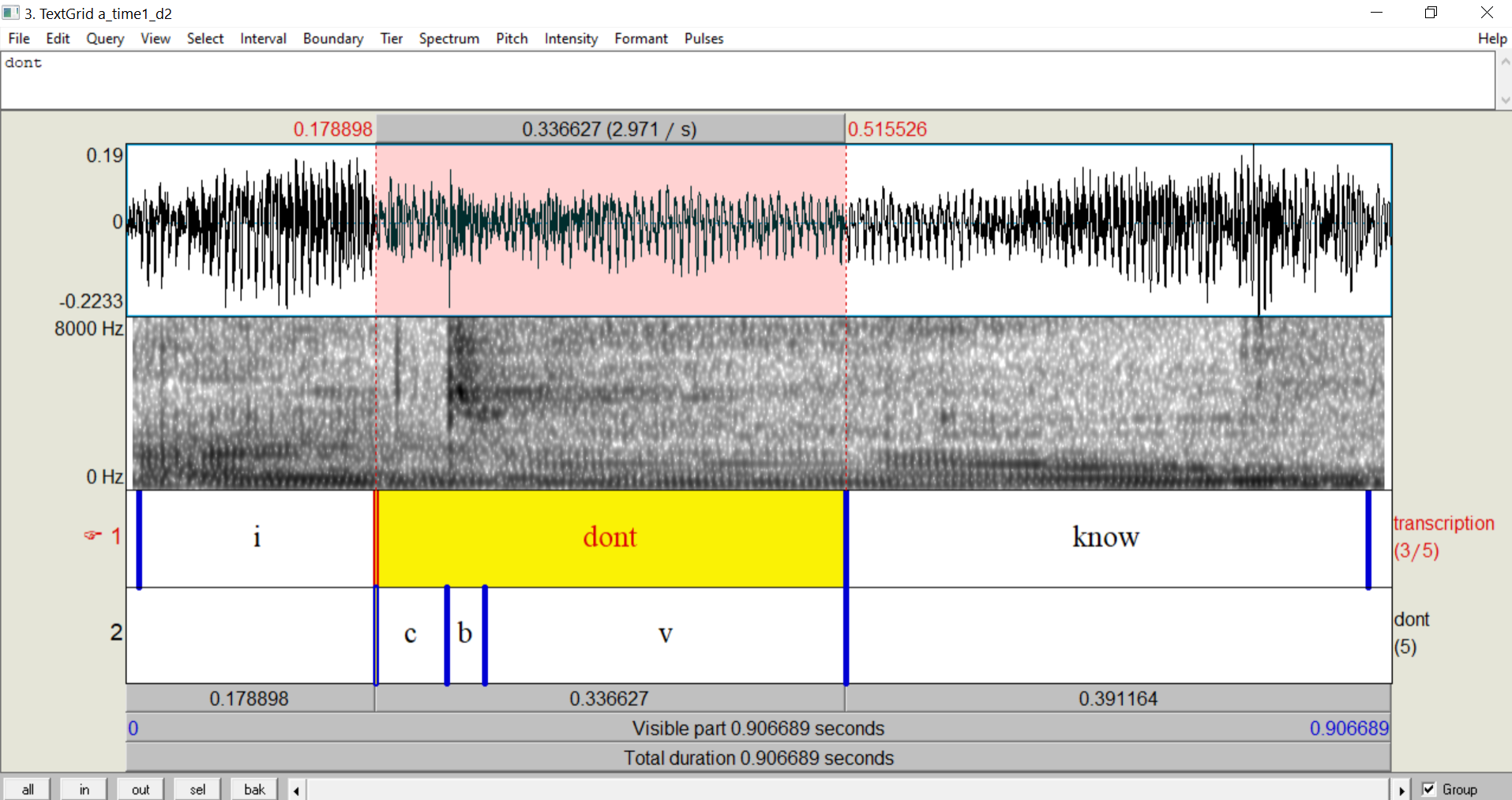
- Quelques résultats (QR2)

Tableau 1. Emplois de *I don't know* en tant que marqueur discursif.

	A		C		M		N		Y	
	#	per 1K	#	per 1K	#	per 1K	#	per 1K	#	per 1K
1	5	3.92	1	0.91	2	1.91	2	1.50	13	3.72
2	5	1.25	–		1	0.72	3	0.68	26	3.77
3	4	1.49	–		6	4.58	2	0.98	29	4.52
4	6	3.26	1	1.01	8	5.58	1	0.48	29	5.95
5	26	7.02	–		6	3.00	9	2.65	48	6.75

Participant	Interview	Time	Context	Form	Subject	V	function	phraseo	Closure	Burs	Vowel length	Consonant
Y	3	10:56	&uh but &uh yeah i don 't know at the &uh i don 't know it 's maybe like attachment or &uh	don't	I	know	pragmatic	y	y	n	0,018139469	0,031510695
Y	5	30 11	at the same thing it 's contrasted[*] so <i just don 't know how to &uhm> [//] i don 't know what the outcome would[*] be	don't	I	know	literal	n	y	n	0,018638062	0,042123835
Y	5	17:34	and <there w(asn)> [//] there [x 2] wasn 't really communication with him &uhm i don 't know if <it was> [///] i think	don't	I	know	pragmatic (meaning "maybe"??)	n	y	n	0,019886991	0,066520468
Y	1	10:21	so like i 'm not like doing one thing <th(irty)> [/] i dunno thirty hours a week	don't	I	know	pragmatic	y	n	n	0,022326713	NO_C
Y	5	20:04	i 'm seventy percent birtish and thirty percent american because i don 't know <it 's just maybe> [///] i don 't know i feel like	don't	I	know	pragmatic	y	y	n	0,022728103	0,033826006
Y	5	25 16	sometimes you just kind of don 't know where you are &uhm but yeah	don't	you	know	literal		y	y	0,024449087	0,112910935
AE	2	10:53	so i don 't know how it works at DIT	don't	I	know	literal	n	y	n	0,024505958	
Y	3	43 00	&uh but yeah identity &uh yes [x 2] because &uh i don 't know how to explain	don't	I	know	literal	n	y	y	0,024780072	0,048891838
Y	1	23:15	and <i really liked> [//] i really like &um: =&clicks:tongue i dunno the [x 2] fact	don't	I	know	pragmatic	y	n	n	0,025113819	NO_C

Étude sur les emplois de *don't*

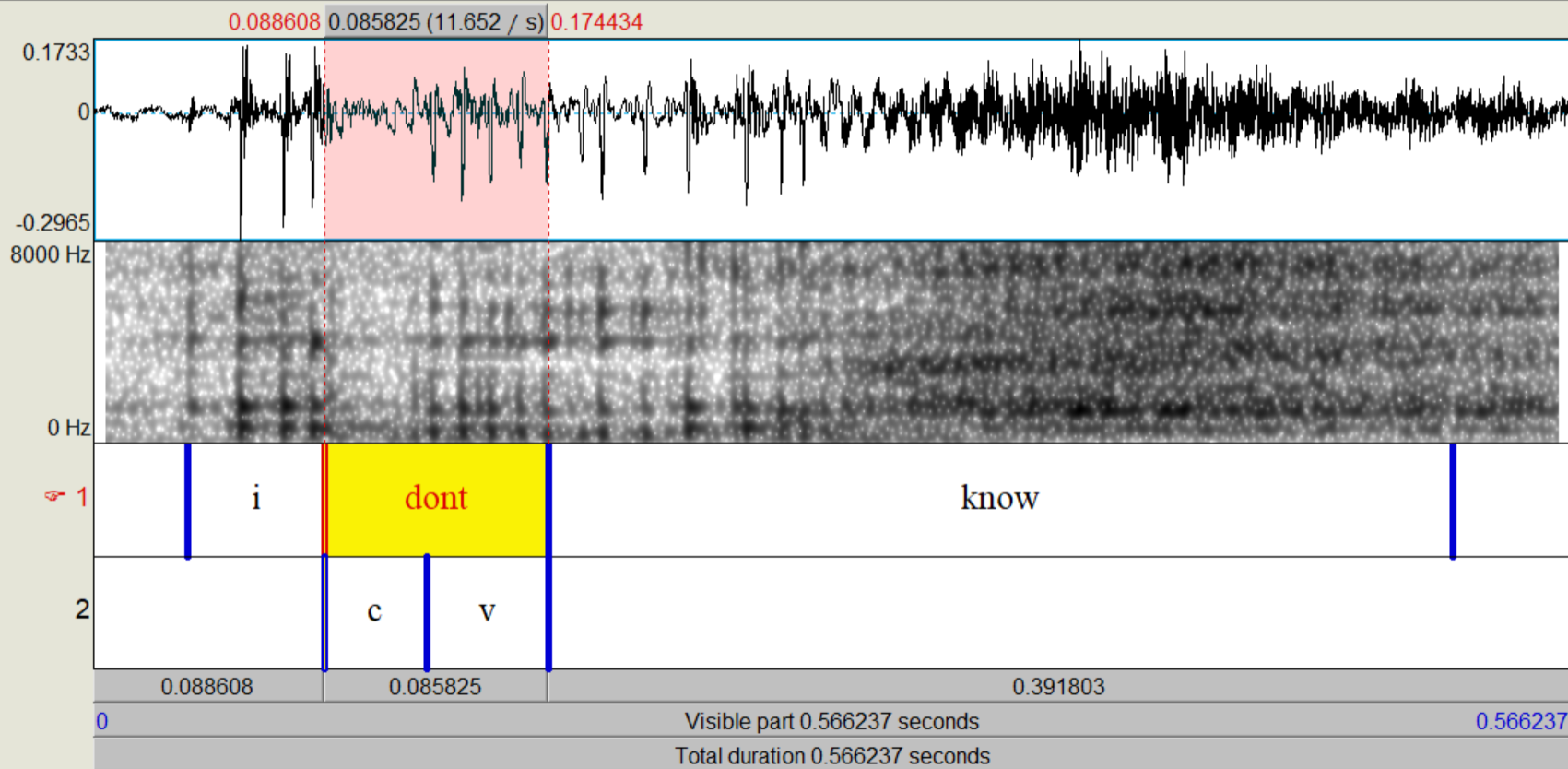


Étude sur les emplois de *don't*

5. TextGrid y_time3_c2

File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help

dont|



transcription
(3/5)

dont
(4)

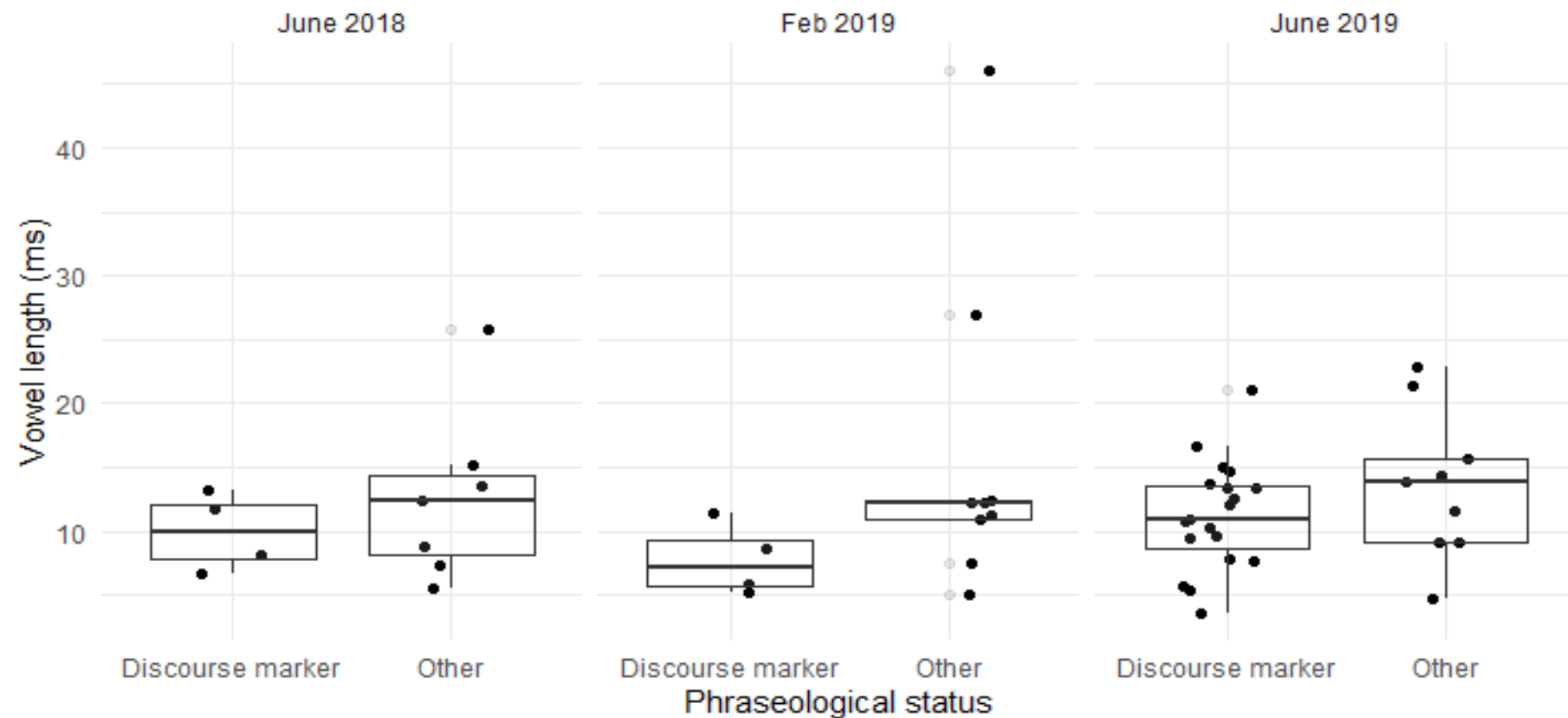
all in out sel bak

Group



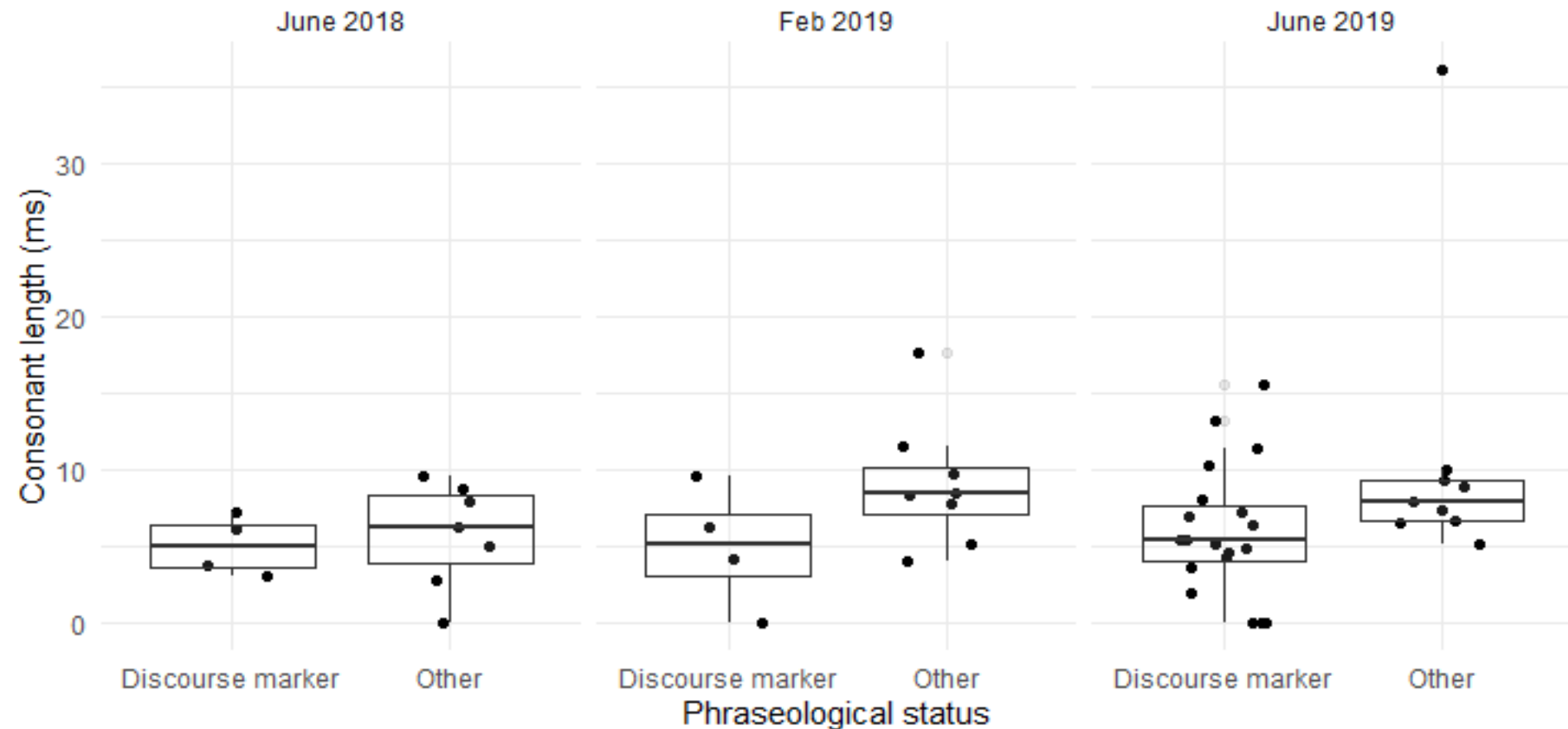
Étude sur les emplois de *don't*

- Quelques résultats (QR2) : voyelles du participant A



Étude sur les emplois de *don't*

- Quelques résultats (QR2) : consonnes du participant A



Étude sur les emplois de *don't*

- Conclusions
 - Choix à faire : on ne peut pas tout avoir !
 - Codage détaillé, quitte à fusionner des catégories par la suite
 - Un corpus comme PROLINGSA offre de nombreuses possibilités d'exploitation
 - Analyses fonctionnalistes
 - Analyses acoustiques (ex. fluence)
 - Analyses gestuelles
 - ...

- 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.
- BYBEE Joan & SCHEIBMAN Joanne, 1999, "The effect of usage on degrees of constituency: The reduction of *don't* in English", *Linguistics*, 37(4), 575-596.
- EDMONDS, A., SNEED GERMAN, E. & LECLERCQ, P. (2021). A phraseological and phonological analysis of *don't*: A stay abroad perspective. *Lexis* [Online], 18.
- FULLER Janet M., 2003, "Discourse marker use across speech contexts: A comparison of native and non-native speaker performance", *Multilingua*, 22, 185-208.
- HELLERMAN John & VERGUN Andrea, 2007, "Language which is not taught: The discourse marker use of beginning adult learners of English", *Journal of Pragmatics*, 39, 157-179.
- LECLERCQ, P. (2020). Transcribing interlanguage: The case of verb-final [e] in L2 French. In Edmonds, A., Leclercq, P. & Gudmestad, A. (Eds), *Interpreting language-learning data*, pp. 169-197. Zenodo. <https://doi.org/10.5281/ZENODO.4032298>
- LIAO Silvie, 2009, "Variation in the use of discourse markers by Chinese teaching assistants in the US", *Journal of Pragmatics*, 41(7), 1313-1328.
- 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>
- Nagy, N. & SHARMA, D. (2013). *Transcription*. In Robert J. Podesva & Devyani Sharma (eds), *Research Methods in Linguistics*. Cambridge University Press.
- OCHS, Elinor. (1979). Transcription as theory. *Developmental pragmatics* 10(1). 43-72.
- PICHLER Heike, 2009, "The functional and social reality of discourse variants in a northern English dialect: I DON'T KNOW and I DON'T THINK compared", *Intercultural Pragmatics*, 6(4), 561-596.
- SANKOFF Gillian, THIBAUT Pierrette, NAGY Naomi, BLONDEAU Hélène, FONOLLOSA Marie-Odile & GAGNON Lucie, 1997, "Variation in the use of discourse markers in a language contact situation", *Language Variation and Change*, 9, 191-217.
- 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).
- TAVAKOLI Parvaneh, 2018, "L2 development in an intensive study abroad EAP context", *System*, 72, 62-74.