

CONVENTIONS DE TRANSCRIPTION EN VUE D'UN ALIGNEMENT TEXTE-SON AVEC TRANSCRIBER

Virginie André, Christophe Benzitoun, Emmanuelle Canut, Jeanne-Marie Debaisieux, Bertrand Gaiffe, Evelyne Jacquey

Préambule

La transcription ne peut pas refléter l'enregistrement avec une fidélité parfaite : la prononciation d'un même sujet/terme, par un enfant comme par un adulte est variable. Le but de notre travail sur corpus n'est pas une analyse fine de la prononciation. C'est une étude approfondie de la sémantique et de la syntaxe en fonction de l'objectif précis que nous fixons à l'analyse. Cette analyse doit être rigoureuse, chaque corpus présentant des données particulières. Les éléments transcrits sur lesquels subsistera un doute ne seront ni interprétés ni analysés, seules des hypothèses pourront éventuellement être formulées. (On ne formulera éventuellement que des hypothèses.)

Tours de parole

Chaque tour de parole fait l'objet d'une identification du locuteur dans TRANSCRIBER. Cette identification est représentée sur l'écran par un encadré. *Pour les corpus adulte-enfant, l'adulte est identifié par « adulte » et l'enfant par son prénom. Pour les corpus adulte-adulte les locuteurs sont identifiés comme « L » et sont numérotés dans l'ordre de leur prise de parole : L1, L2, etc.* Sur l'écran on obtient la configuration suivante :

Corpus adulte-enfant :

adulte
• ta maman vient te chercher
David
• non
adulte
• ah bon

Corpus entre adultes :

L1
• et concrètement qu'est ce qu'il a fait pour euh
L2
• euh plein de choses euh en fait euh grâce à lui euh j'ai pu rencontrer Jackie Chan euh la semaine de mon arrivée
L1
• hum hum

Pour les corpus adulte-enfant, les énoncés doivent être numérotés mais la numérotation n'est pas à réaliser au moment de la transcription. Tous les énoncés seront numérotés automatiquement dans des balises « commentaire » après la première phase de transcription et avant la phase de vérification via un programme informatique.

Principes généraux de transcription

1. Transcrire tout ce qui est dit par l'adulte et par l'enfant, y compris les hésitations et les répétitions, selon l'orthographe usuelle.

Ne pas ajouter de morphèmes non verbalisés. Par exemple, ne pas écrire « ne » lorsque cette partie de la négation n'est pas réalisée.

2. Ne pas ponctuer les énoncés.

Ne pas employer les majuscules sauf pour les noms propres.

Mettre des guillemets droits pour les titres, avec majuscule sur le premier mot (ex : "Arnaque, crime et botanique").

Exception pour les corpus adulte-enfant : indiquer « ? » pour exprimer l'intonation interrogative¹.

3- Ne pas anonymiser le corpus (indiquer tous les noms propres, chiffres etc. entendus, même ceux permettant l'identification de locuteurs). Le balisage des éléments susceptibles d'être anonymisés et l'anonymisation (dans le son et le texte) se fera ultérieurement selon un codage spécifique.

¹ Une ponctuation dans les énoncés de l'adulte pourra être réalisée ultérieurement en fonction des besoins de l'analyse.

Orthographe et prononciation : spécificités

- Les nombres doivent respecter les normes habituelles de l'écrit. Tous les nombres doivent être écrits en lettres.
- Phatiques et onomatopées sont codifiées et doivent être transcrits selon l'orthographe fournie :

ah, aïe, areu, atchoum, badaboum, baf, bah, bam, bang, bé, bêêê, beurk, ben, bing, bon, boum, brom, cataclap, clap clap, coa coa, cocorico, coin coin, crac, croa croa, cuicui, ding, ding deng dong, ding dong, dring, eh, eh ben, eh bien, euh, flic flac, flip flop, frou frou, glouglou, glou glou, groin groin, grr, hé, hep, hi han, hip hip hip hurra, houla, hurra, hum, mêêê, meuh, miam, miam miam, miaou, oh, O.K., ouah, ouah ouah, ouais, ouf, ouh, paf, pan, patatras, pchhh, pchit, pff, pif-paf, pin pon, pioupiou, plouf, pof, pouet, pouet pouet, pouf, psst, ron ron, schlaf, sniff, splaf, splatch, sss, tacatac, tagada, tchac, teuf teuf, tic tac, toc, tut tut, vlan, vroum, vrrr, wouah, zip.

- Les sigles sont ponctués quand on lit les lettres isolément (S.N.C.F.), non ponctués lorsqu'il s'agit d'acronyme (CROUS). Le sens des sigles peut être précisé lors de la première apparition avec une balise TRANSCRIBER : A.E.E. {sigle=Agence Européenne de l'Environnement}
- Lorsque l'orthographe est incertaine (cas de certains noms de marques, toponymes, etc.) on pourra utiliser une orthographe approximative si elle est plausible et le signalant avec des balises TRANSCRIBER, soit à la suite du mot isolé : un certain Dupont+[lex=?] ; soit en englobant la séquence : [lex=?-] Sous la Voivre[-lex=?]
- Les accords non standards sont suivis de l'indication {sic} : « tu as vu des chevaux {sic} », « ils croient {sic} que c'est vrai ». Utiliser la balise « commentaire » {sic} répertoriée dans TRANSCRIBER.
- Respecter les règles d'accord sauf si on a une réalisation phonique particulière. Par exemple, « on est **parti** avec maman » mais « on s'est **mises** à dormir »
- Quand il y a hésitation entre plusieurs possibilités pour la transcription, noter les mots entre barres obliques, séparées par une virgule : /ça, chat/ ; /va, vois/
- Quand on a des mots indistincts (incompréhensibles, inaudibles, inconnu...), mettre * si cela ne s'applique qu'à une syllabe, *** si cela s'applique à plusieurs syllabes. On pourra éventuellement ajouter une balise prononciation : c'est une * [pron=pul ?]
- Les liaisons particulières sont indiquées entre deux signes = : « le =n= ours », « donne-moi =z= en ».
- On utilisera les parenthèses pour les variantes morphologiques indécidables (non réalisées à l'oral) : « il(s) disai(en)t... » ; « on (n') est pas là »
- Ne pas rétablir les élisions non réalisées : « parce que il est pas là ».
- Dans le cas où sont verbalisés des mots en langue étrangère, transcrire selon la norme de la langue d'origine et insérer une balise langue dans TRANSCRIBER. On obtient par exemple la configuration : speed+[lang=anglais]

SPECIFICITES POUR LES CORPUS ADULTE-ENFANT

- Dans les corpus adulte-enfant, on pourra utiliser {sic} quand il y a reprise par l'adulte d'un mot « tronqué » verbalisé préalablement par l'enfant :

enfant

- non c'est un éléphant [pron=efa~]

adulte

- c'est un oui c'est un éphant {sic} c'est un éléphant

- Dans les corpus adulte-enfant, quand certains éléments (syllabes) ne sont pas réalisés à l'intérieur d'un mot ou quand le mot est parfaitement reconnaissable mais que ponctuellement il est verbalisé avec une prononciation un peu différente (inversion de syllabes par ex) : écrire le mot correctement orthographié puis dans TRANSCRIBER insérer une balise de prononciation et écrire la prononciation exacte entendue avec l'alphabet SAMPA. Sur l'écran de transcrire, on obtient la configuration suivante : il frappe [pron=ifap] ; elle court [pron=Ecu] ; parce que [pron=pak2] ; tu as [pron=ta], je sais [pron=SE], spectacle [pron=pEstakl], chat [pron=sa], pomme [pron=pEm], etc.

En revanche, ne pas mettre de balises de prononciation pour la prononciation élidée du schwa (mots comportant un « e » susceptible d'être prononcé et qui ne l'est pas) et conserver uniquement la forme orthographique usuelle : transcrire petit pour [pti].

Dans le cas d'une prononciation particulière récurrente dans le corpus, par exemple l'enfant prononce [z] tous les [Z] (« ze » pour « je »), elle ne sera pas reportée dans la transcription mais indiquée en commentaire dans la fiche signalétique.

- Si ce que dit un enfant n'est pas compréhensible ou s'éloigne assez d'une prononciation standard, ne pas écrire de mots ou segments en orthographe standard mais utiliser directement dans TRANSCRIBER une balise de prononciation [pron=pap2mut] suivie d'un commentaire {interprétation=pamplemousse}.

Dans le cas d'un doute ou d'une impossibilité d'interprétation, mettre un point d'interrogation dans l'accolade : [pron=iai] {interprétation=parti ?} ; [pron=emusEbele] {interprétation=la mouche s'est envolée ?}.

Dans les énoncés de l'enfant, quand il y a un doute sur l'interprétation du morphème réalisé, transcrire en alphabet SAMPA dans une balise de prononciation et mettre à la suite, dans une balise commentaire, les deux interprétations possibles : [pron=ki] {/qui, qu'il} ; [pron=i] {/il, qui}

Phénomènes propres à l'oral

Pour les corpus adulte-enfant :

=> indiquer les chevauchements entre les énoncés de l'adulte et l'enfant via la balise « locuteur superposé » dans TRANSCRIBER Sur l'écran on obtient la configuration suivante :

adulte

- ta maman range les

adulte + enfant

- 1- yaourts dans le frigo
- 2- ben non c'est pas le frigo

enfant

- c'est pas le frigo

Dans ce cas, 1- correspond à l'adulte et 2- à l'enfant.

Pour les corpus adulte-adulte :

=> Les chevauchements de paroles sont notés entre chevrons (non collés au mot).

- L1 pardon < toujours lié à l'ordinateur
- L2 et euh sou- souvent souvent > je veux pas dire

- Noter les amorces de mots par un tiret (collé) :

« il a il a p- il a pris »

- Les pauses sont notées par une croix (espace avant et après) : +

- Les pauses très longues (silence) sont notées /// et peuvent être accompagnées d'un commentaire (*balise commentaire dans TRANSCRIBER*) : /// {l'enfant regarde attentivement l'image} ; /// {l'enfant va chercher un jouet dans sa chambre} ; /// {L1 se sert un verre d'eau}

- Préciser les détails de la situation (rires, bruits, etc.) avec une balise commentaire de TRANSCRIBER : Si l'événement ou le « bruit » est répertorié dans TRANSCRIBER, insérer cette balise (ex : [rire]). Si l'événement ou le « bruit » n'est pas répertorié comme balise dans TRANSCRIBER, le décrire : {bruit d'une sirène dans la rue} ; {éternuement de l'adulte} ; {l'enfant saute sur son lit} ; {une porte qui claque}

Si un événement est récurrent (bruits de fond, etc.), il ne sera pas reporté dans la transcription mais indiqué en commentaire dans la fiche signalétique.

- Les parties enregistrées mais non transcrites sont notées ### suivi d'un commentaire (*balise commentaire dans TRANSCRIBER*) :

{la mère entre dans la chambre et pose une question à l'enfant}

- Les coupures dans l'enregistrement sont indiquées par \$\$\$ suivi d'un commentaire (*balise commentaire dans TRANSCRIBER*) :

\$\$\$ {l'enfant éteint le magnétophone par inadvertance} ; \$\$\$ {sonnerie du téléphone : l'adulte arrête l'enregistrement pour répondre} ; \$\$\$ {partie confidentielle, les locuteurs ont demandé l'arrêt du magnétophone}

RECAPITULATIF DES SYMBOLES DE TRANSCRIPTION

{...	
Commentaire	
<hr/>	
Variante	
graphique	
indécidable	
Pauses	
Pauses	très
longues	
Liaison	non
standard	
remarquable	
Hésitations	entre
transcription	
Amorces	
Syllabe	
incompréhensible	
Suite de syllabes	
incompréhensible	
s	
Passage	
enregistré	non
transcrit	
Coupure	de
l'enregistrement	

ANNEXE 1 – ALPHABET SAMPA

CONSONNES	Symbole	Exemple	Transcription
Plosives	p	pont	po~
	b	bon	bo~
	t	temps	ta~
	d	dans	da~
	k	quand	ka~
	g	gant	ga~
Fricatives	f	femme	fam
	v	vent	va~
	s	sans	sa~
	z	zone	zon
	S	champ	Sa~
	Z	gens	Za~
Nasales	m	mont	mo~
	n	nom	no~
	J	oignon	oJo~
	N	camping	ka~piN
Liquides	l	long	lo~
	R	rond	Ro~
Semi-consonnes	w	coin	kwe~
	H	juin	ZHe~
	j	pierre	pjER

VOYELLES	Symbole	Exemple	Transcription
Orales	i	si	si
	e	ses	se
	E	seize	sEz
	a	patte	pat
	A	pâte	pAt
	O	comme	kOm
	o	gros	gRo
	u	doux	du
	y	du	dy
	2	deux	d2
	9	neuf	n9f
	@	justement	Zyst@ma~
	Nasales	e~	vin
a~		vent	va~
o~		bon	bo~
9~		brun	bR9~
E/			= e ou E
Indéterminées	A/		= a ou A
	&/		= 2 ou 9
	O/		= o ou O
	U~/		= e~ ou 9~

ANNEXE 2– EXEMPLE D’UN EXTRAIT DE CORPUS ADULTE-ENFANT TRANSCRIT AVEC TRANSCRIBER

Véronique

- là je vois la poupée avec un nounours là

adulte

- tu crois que c’est un nounours ça ?
- c’est bien plus gros qu’un nounours

Véronique

- non c’est un Babar [*pron=baba*]

adulte

- ah c’est un Babar

Véronique

- non c’est un éléphant [*pron=efa~*]

adulte

- c’est un oui c’est un éphant {*sic*}

adulte + Véronique

- 1- c’est un éléphant
- 2- *** [*pron=sasepu ?*]

adulte + Véronique

- 1- attends
- 2- c’est une * [*pron=pul ?*]

adulte

- attends ne va pas si vite

Véronique

- c’est une * [*pron=pul ?*]

adulte

- attends attends attends moi j’ai pas eu le temps de tout voir encore
- alors je vois la poupée qui est assise

Véronique

- sur [*pron=sy*] les dessins

adulte

- elle est assise sur les dessins oh drôle d’idée +
- je vois le gros éléphant bleu qui la regarde

Véronique

- moi je vois ça

adulte

- tu vois ça ? qu’est-ce que c’est ça ?

Véronique

- / moi, non/ /va, vois/ ça

adulte

- *** oh c’est difficile ça c’est l’excavatrice
- c’est pour faire des trous pour faire des grands trous + et puis aussi de petits

Véronique

- et puis à la da- à la télé j’ai vu [*pron=ki*] {/qui, qu’il/} pince la dame +
- il [*pron=i*] pince + /pincé, pincer, pinçait/ la dame

adulte

- tu as vu qu’on pinçait la dame ?

Véronique

- dans ma télé à moi

adulte

- dans ta télé à toi parce que dans la mienne j’ai pas vu ça {*rire*}
- on pinçait une dame ou c’était la dame qui pinçait quelque chose ?

Véronique

- non [*pron=na-*]
- adulte
- non alors explique-moi bien
- adulte + Véronique
- 1- parce que je
 - 2- non c'est pas
- Véronique
- la da(me) qui +
- Véronique + adulte
- 1- qui
 - 2- bon
- adulte
- bon alors explique-moi bien ce que tu as vu parce que moi je ne l'ai pas vu dans ma télé ça +
 - qu'est-ce que tu as vu dans ta télé ?
- Véronique
- j'ai vu euh des des lits
- adulte
- ah non tu as [*pron=ta*] pas vu des lits dans ta télé
- Véronique
- si
- adulte
- si ? et qui il y avait [*pron=javE*] dans ces lits ?
- Véronique
- [*pron=lavE*] avait des beaucoup petits enfants
- adulte
- il y avait beaucoup de petits enfants ?
 - c'était comme à l'école alors ? + {*V. acquiesce*}
 - oui ? c'était comme à l'école ?
 - et qu'est-ce qu'ils faisaient ces petits enfants ?
- Véronique
- il a dormi [*pron=ladomi*] pour + il a dormi [*pron=ladomi*] + beau- beaucoup dans les lits
- adulte
- ils ont dormi beaucoup dans les lits ah bon

TRANSCRIBER

L1

- ben tu conduirais

L2

- {rires} en voiture quelle voiture

L1

- je sais pas elle a dit qu'on trouverait un moyen

L2

- ah bon sûrement parce que moi je peux pas prendre de voiture

L1

- oui je sais mais euh peut-être que Jean aura une bagnole

L2

- c'est combien de temps jusqu'à Troyes

L1

- pff ça fait trois heures je crois c'est long hein

L2

- mh oh c- ça va

L1

- le truc c'est que tu pars avec 300 euros tu te rhabilles pour l'année quoi

L2

- c'est clair

L1

- parce que c'est vraiment trop pas cher on regardait euh alors moi j'en parlais déjà avec des amis
- le jean Levi's que tu paies euh pff 70 euros un truc du genre même plus un jean Levi's tu le paies plus

L2

- c'est clair

L1

- tu paies 10 euros là-bas ben

L2

- < ben non

L1

- un truc > de fou quoi enfin

L2

- oui oui ça vaut le coup je pense

L1

- hum