

Convention de transcription CIEL-F

Version 4.7. – 28 mars 2012

Anne Catherine Simon (coordination), Françoise Gadet, Ralf Ludwig, Lorenza Mondada, Stefan Pfänder, Marie Skrovec

Ces conventions sont inspirées de celles des équipes ICOR¹, de Freiburg² et VALIBEL³

Principes généraux et logiciels utilisés pour la transcription		
La transcription est alignée sur le son		
On transcrit dans un TextGrid	Un TextGrid est un fichier de texte aligné sur un son. On peut créer facilement un textgrid avec les tires adéquates en utilisant un script (par ex. <code>longsounds_to_Textgrid.praat</code> http://www.fon.hum.uva.nl/praat/	
On transforme le TextGrid en fichier texte (txt ou rtf)	Cette transformation se fait par exemple en utilisant le logiciel Transformer (http://www.oliverehmer.de/transformer/), qui règle la mise en page et la conversion des polices.	

¹ Voir : http://icar.univ-lyon2.fr/projets/corinte/bandeau_droit/convention_icor.htm

² Le document « Conventions appliquées au corpus de Freiburg » est basé sur Selting et al. (1998). Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173, 91-122, consultable en ligne : <http://www.medienprache.net/de/medienanalyse/transcription/gat/gat.pdf> (accédé le 28 mars 2012)

³ Voir : http://www.uclouvain.be/cps/ucl/doc/valibel/documents/conventions_valibel_2004.PDF

		Phénomène	Conventions	Exemples
1	Tires pour la transcription et la traduction / glose (<u>par locuteur</u>)			
1a	Tire de transcription principale CH-ANN-tran	On réserve une tire par locuteur. Cette tire principale contient la transcription en orthographe standard telle qu'elle est décrite aux sections 2 à 6 (infra). Cette tire est nommée par un identifiant de 3 lettres en majuscules, composé à partir du pseudonyme du locuteur (3 premières lettres), auquel on ajoute, pour éviter les doublons, un préfixe correspondant à la zone (le code iso du pays seulement). Par ex. CH-ANN est l'identifiant pour le pseudonyme « Anne », dans la zone Suisse. Dans le cas de Samya et Samir au Liban par ex, on changera une ou deux des 3 premières lettres du prénom: LB-SAM (Samya), LB-SAR (Samir).		
1b	Tire d'affinage de la transcription CH-ANN-affi	On crée facultativement une tire supplémentaire pour une transcription affinée. Cette transcription affinée peut être une duplication de la transcription principale qui note des phénomènes de prononciation non standard en utilisant l'orthographe adaptée ou les symboles phonétiques. La tire d'affinage peut ne porter que sur certains passages de la transcription principale.		
1c	Tire de traduction CH-ANN-trad et / ou CH-ANN-glos	Le transcripteur peut créer une ou deux tires qui contiennent une traduction (-trad) ou une glose (-glos). Le type de traduction ou de glose sera documenté pour chaque transcription.		
	Tires pour la transcription (<u>autres</u>)			
1d	Tire pour les locuteurs incertains ou les groupes LOC1, LOC2	On réserve une tire pour transcrire les paroles dont l'attribution à un locuteur pose problème; attribuées à un groupe de locuteur (« classe », « garçons » etc. On précise alors l'identifiant du locuteur supposé avec un point d'interrogation. (On n'a pas encore décidé si cette tire sert pour transcrire les productions des intervenants dans les enregistrements de programmes radiophoniques).		LOC1 <((ANA?)) si tu veux>
1f	Événement ou silence non attribuables SIL-EVEN	On réserve une tire pour la transcription des silences non attribuables et des événements non langagiers.		SIL-EVEN ((un coup de tonnerre retentit)) SIL-EVEN (0.5)
	Tires pour les commentaires sur la transcription (ces tires sont à usage interne, ne seront pas affichées dans la base)			

1g	Statut de la transcription STATUS	Dans cette tire, le responsable de la transcription fait état de la progression de la transcription (qui transcrit? Qui révise? Quelle partie est déjà transcrite? Etc.)	
1h	Passages à anonymiser ANON	Le transcripateur indique les passages à anonymiser (noms propres, propos susceptibles de permettre l'identification d'un locuteur).	
1i	Commentaires COM	Méta-commentaires du transcripateur (par ex. sur un mot de la transcription). (voir 7a)	

2	Typographie et orthographe		
2a	Police courrier	La police à utiliser pour les transcriptions est <code>courrier</code> . Cette police sera appliquée lors de la transformation du TextGrid en fichier texte, elle ne doit pas être utilisée dans Praat.	
2b	Usage des majuscules	On n'utilise pas la majuscule en début d'unité de transcription. Les majuscules sont réservées à l'identification des noms propres (2c-2d) et de la notation de l'emphase (6b).	
2c	Noms propres	Les majuscules sont utilisées à l'initiale des noms propres et pour la notation des pseudonymes des participants à la conversation.	CH-ANN je m'appelle Anne et je viens de Bruxelles CH-MIC j'ai parcouru Le Monde
2d	Sigles et acronymes	Un sigle (prononcé en épelant les lettres) est noté en majuscules avec espaces. Un acronyme (prononcé comme un mot) est noté comme un nom (avec majuscule).	CH-ANN j'ai pris un train S N C F ⁴ CH-ANN je travaille à l'Onu ⁵ CH-ANN je m'appelle Ann sans E ça s'écrit A N N
2e	Chiffres	Les chiffres sont notés en toutes lettres et avec des traits d'union	CH-ANN je suis née en dix-neuf-cent-quatre-vingt-
2f	Orthographe	On utilise pour transcrire l'orthographe standard de la langue ou de la variété de langue ⁶ . On n'utilise pas l'orthographe adaptée. Quand la langue utilisée n'a pas d'écriture latine, on transcrit en utilisant l'alphabet phonétique (entre chevrons) ou un système de translittération. Le choix du transcripteur doit être documenté dans un fichier qui accompagne la transcription ou dans le tire COM du TextGrid.	monsieur il vient je sais pas tu sais tu as ⁷
3	Hésitations et amorces		
3a	Notation exhaustive des productions verbales	On note tout ce qui est produit par les locuteurs (hésitations, répétitions, etc.).	
3b	Amorce	Insertion de « - » après le constituant amorcé (attention : pas d'espace avant, un espace après). On peut noter une amorce pour un mot dont on perçoit qu'il a été interrompu, par exemple par un coup de glotte.	CH-COR je retr- retrouve mon chemin CH-COR tu- tu viendras demain

⁴ Lorsque la prononciation est [ɛsɛnsɛɛf].

⁵ Lorsque la prononciation est [ɔny], c'est-à-dire sans épellation.

⁶ Si un mot est utilisé qui n'est pas référencé dans le dictionnaire, on utilise l'orthographe la plus plausible pour le transcrire (par ex. « le poitrinoise » (pour patrimoine) ou « d gourgettes » (pour courgettes)). On peut aussi ajouter une transcription phonétique dans la tire LOC-affi.

⁷ L'orthographe reste stable même si les prononciations peuvent être variables ; par exemple « je sais pas » [ʒe se pa], [ʃepa], tu sais [ty se] ou [tse], etc.

	3c	Amorce avec continuation immédiate	Insertion de « - » après le constituant amorcé (attention : pas d'espace avant, un espace après).	CH-COR c'est in- croyable
4	Unité de transcription et phénomènes temporels (chevauchements et pauses)			
	4a	Unité de transcription	<p>La transcription sous Praat implique de segmenter les prises de parole des locuteurs intervalles qui contiennent des unités de transcription.</p> <p>Une unité de transcription est un extrait de parole pour lequel on aligne la transcription sur le son. Il ne s'agit pas d'une unité théorique de type « tour de parole » mais d'une unité technique, due au logiciel utilisé pour transcrire et à l'affichage de la transcription</p> <p>On note les productions verbales de chaque locuteur dans la tire qui lui correspond. Chaque production verbale est segmentée en unités dont la longueur ne dépassera pas, si possible, 10 secondes (une durée de 4-5 secondes est optimale pour l'affichage en ligne des transcriptions).</p>	
	4b	Chevauchement	<p>De crochets "[" et "]" encadrent le chevauchement dans chaque tour. Les crochets ouvrant "[" (début du chevauchement) et fermants "]" sont obligatoires (sans espace entre le crochet et le texte prononcé en chevauchement).</p> <p>(Lors de la transformation automatique du TextGrid en fichier Texte, les crochets sont alignés verticalement au moyen d'espaces.)</p>	<p>CH-SAR a:::ie CH-COR c'est pas utile [ça] CH-SAR [ça] mar[che] CH-COR [attr]ape-le CH-SAR c'est bon/</p> <p>CH-SAR salut Corinne CH-COR sa[lut] CH-SAR [tu] sais ce [qu'ils ont dit/] CH-MME [bonjour Sarah] CH-SAR bonjour madame</p>
	4c	Enchaînement immédiat (latching)	Ce type d'enchaînement est un phénomène plutôt rare, intermédiaire entre le chevauchement et la présence d'une micro-pause entre deux tours de parole. On note si les tours de parole de deux locuteurs successifs s'enchaînent sans aucune pause, donnant une impression de chevauchement. On insère le symbole "=" à la fin du premier tour (sans espace avant) et au début du suivant (sans espace après).	<p>CH-ANN d'où vient son père/ où était né son père: CH-CAR CH-ANN =alors s-= CH-CAR =pis il était inscrit là dans le registre/</p>
	4d	Notation des silences	Les silences sont chronométrés à l'aide d'un logiciel au 10 ^e de seconde près. On note la durée du silence en secondes avec un chiffre après la virgule, sauf pour les silences d'une durée inférieure à 0,2 secondes qui sont notés par un point. Par ex. : (1.2) ou (.)	<p>CH-JAC parce que ma cousine m'a appelé mais euh enfin j'ai fait (.) mon travail d'électricien quoi</p>
	4e	Attribution des silences	<p>Un silence est transcrit dans la tire du locuteur à qui on l'attribue. Dans le cas où le silence n'est pas attribuable à un participant, il est noté dans une tire spécifique SIL-EVEN.</p> <p>Un même silence ne doit pas être noté dans deux tires différentes, sous</p>	<p>CH-LOU non c'est ce matin/ j'étais:\ (0.5) j'ai tout le tour là/ SIL-EVEN (4.0) LOU donc je suis arrivé tard/</p>

			peine d'apparaître deux fois dans la version textuelle finale de la transcription.	SIL-EVEN (5.3) CH-JUL tu as immédiatement branché le serveur
	4f	Aspiration (facultatif)	L'aspiration est notée par la lettre « h » précédée d'un point (attention : le point n'est pas suivi d'un espace).	CH-COR .h ah oui
	4g	Expiration, soupir (facultatif)	L'expiration est notée par la lettre « h » seule.	CH-COR h mais non
	4h	Action à valeur de tour	Description de l'action concernée entre doubles parenthèses, dans la tire du participant à qui cette action est attribuée.	CH-ISA (tu as des) devoirs Stéph// SIL-EVEN (0.9) CH-STE ((mouvement du menton vers l'avant)) CH-ISA tu as beaucoup de devoirs? CH-STE ((hoche la tête)) CH-SAR tu as de la chance/ moi j'ai un D S
5	Production vocale			
		Production vocale qui accompagne la parole	La description d'un style vocal est notée entre doubles parenthèses et précède la transcription, l'ensemble est compris entre chevrons : <((DESCRIPTION)) TRANSCRIPTION>.	CH-COR <((en riant)) crois-tu/>
			Note : La notation d'un enchaînement direct, ou celle d'un chevauchement, précède la description de la production vocale.	CH-COR je ne l'ai pas vue hier= CH-SAR =<((en riant)) non::> CH-COR je ne l'ai pas [vue hier] CH-SAR [<((en riant)) non::>]
		Production vocale isolée	On note entre double parenthèse, et dans la tire du locuteur à qui elle est attribuée, toute production vocale isolée : ((rit))	CH-COR je ne l'ai pas vue hier ((tousse)) CH-SAR non moi non plus ((rit))
6	Prosodie			
	6a	Mouvements mélodiques globaux	On note les mouvements mélodiques globaux seulement aux frontières des unités de transcription , afin de mettre en évidence la segmentation interne des tours. Remarque : on ne combine pas les symboles intonatifs au-delà des possibilités présentées ci-dessous ⁸ .	CH-ABC ces gens là\ (0.2) ils sont arrivés l'autre me dit que bon::~ on fait comment tu- on on t'invite~ (.) avec ta blanche on vous emmène en boîte/\
		Montées intonatives	Les montées intonatives sont notées par « / » et « // » sans espace avant. Le symbole « / » correspond à un ton moyen-haut ou un ton suspensif légèrement et le symbole « // » correspond à un ton haut.	CH-COR tu y vas toi//

⁸ Par exemple on ne transcrit pas ~/ pour essayer de rendre un contour « plat-et-montant ».

		Chutes intonatives	Les chutes intonatives sont notées par « \ » et « \ ». Le symbole « \ » correspond à un ton moyen-bas et le symbole « \ » correspond à un ton bas (ou grave).	CH-COR	j'ai décidé tu y vas\ \
		Combinaison d'une montée et d'une chute intonatives	La réalisation d'un contour montant-descendant ou descendant-montant se note en combinant les symboles : « ^ » ou « v ».	CH-COR CH-COR	c'est dingue/ tu crois \ /
		Contour plat	La réalisation d'un contour mélodique plat ou en suspens est indiquée par un symbole tilde ~.	CH-COR	tu y vas~
	6b	Saillance perceptuelle	Les segments caractérisés par une saillance perceptuelle particulière (intensité accrue et autres) sont notés en majuscules, et général sur la syllabe. <i>On utilise cette notation avec parcimonie.</i> Note : Les majuscules doivent être accentuées quand il y a lieu.	CH-COR	je MAN:ge avec STEfan
	6c	Allongement	Le son allongé est noté par des « : » en respectant l'orthographe (attention : pas d'espace avant « : »). Les « : » sont répétés en fonction de la durée perçue de l'allongement.	CH-COR CH-SAR	tu es: où les garÇONS:::
			Note : La notation d'une montée ou d'une chute suit celle de l'allongement.	CH-COR	tu y vas:://
7	Commentaires et multitranscriptions				
	7a	Commentaires du transcripteur	Les commentaires du transcripteur sont notés dans la tire COM .		
	7b	Multitranscriptions	Les segments entre lesquels hésite le transcripteur sont séparés par un point-virgule et notés entre parenthèses. Les segments incertains sont notés entre parenthèses simples.	CH-COR CH-SAR CH-MME	(parce que ; car ce que) (il mange ; ils mangent) (oui)
	7c	Passages inaudibles ou non identifiables par le transcripteur	Les structures incompréhensibles sont représentées au moyen d'une série de caractères « x », chaque caractère ayant la valeur d'une syllabe. Au-delà de deux syllabes, on note (xxx).	CH-COR	j'ai acheté du (xx) hier soir

Liste des onomatopées

ah		hum	(raclement de gorge)	ouille	
aïe		m	[m]	oula, ouh là	
bah		mm	[mm]	p	[p]
bè	[bE]	moui		pf	[pf]
ben	[be~]	mouais		pff [pf:]	
eh	[E]	m'enfin		pt	
enfin	[a~fe~, fe~]	oh		t	(bruit d'ouverture de bouche)
euh		oh la la		wouf, waf	
gnagnagna		ok			
hein		ouais			