

ESTER2

TRANSCRIPTION DETAILLEE ET ENRICHIE

Convention d'annotation version 0.1

08/01/2008

SOMMAIRE

1. Règles de découpage.....	4
1.1. Sections.....	4
1.2. Tours de parole.....	5
1.3. Segments.....	6
2. Règles de transcription.....	8
2.1. Principes.....	8
2.2. Minuscules et majuscules.....	9
2.3. Ponctuation.....	9
2.4. Nombres.....	11
2.5. Adresses Internet.....	12
2.6. Sigles.....	12
2.7. Epellation, abréviations.....	13
2.8. Onomatopées et interjections.....	13
2.9. Hésitations, répétitions.....	14
2.10. Troncation.....	14
2.11. Prononciation erronée.....	14
2.12. Lapsus.....	15
2.13. Elisions et formes réduites.....	15
2.14. Mots inconnus ou absents du lexique.....	15
3. Règles de balisage.....	16
3.1. Bruits.....	16
3.2. Problèmes de prononciation.....	16
3.3. Back Channel.....	17
3.4. Commentaires.....	18
3.5. Disfluences.....	18
3.5.1. <i>Répétitions</i>	18
3.5.2. <i>Hésitations (ou pauses remplies)</i>	19
3.5.3. <i>Révisions</i>	19
3.5.4. <i>Marqueurs de discours</i>	19
3.5.5. <i>Combinaisons de disfluences</i>	20

Avant-propos

Ce document réunit les règles et conventions d'annotation des émissions radiophoniques en français, telles qu'elles ont été remises aux transcribers. En particulier, les exemples fournis sont adaptés au logiciel Transcriber utilisé lors de la transcription. Cette transcription a été effectuée en deux étapes : la transcription détaillée puis, la transcription enrichie. La transcription détaillée est une transcription orthographique synchronisée avec l'audio au niveau de la phrase et prenant en compte les conditions acoustiques. La transcription enrichie est une transcription détaillée à laquelle des informations sont ajoutées. Une fois ces informations prises en compte, le but est de se rapprocher le plus possible d'un texte écrit.

A chacune de ces étapes, une des contraintes principales à laquelle nous devons nous plier concerne l'homogénéité des transcriptions. Cette contrainte sous-entend que chacun des intervenants doit pouvoir avoir la même représentation de la tâche et des règles à appliquer. De ce fait, ce document doit être considéré comme la référence commune, même s'il s'agit d'un document de travail amené à évoluer.

1. REGLES DE DECOUPAGE

1.1. Sections

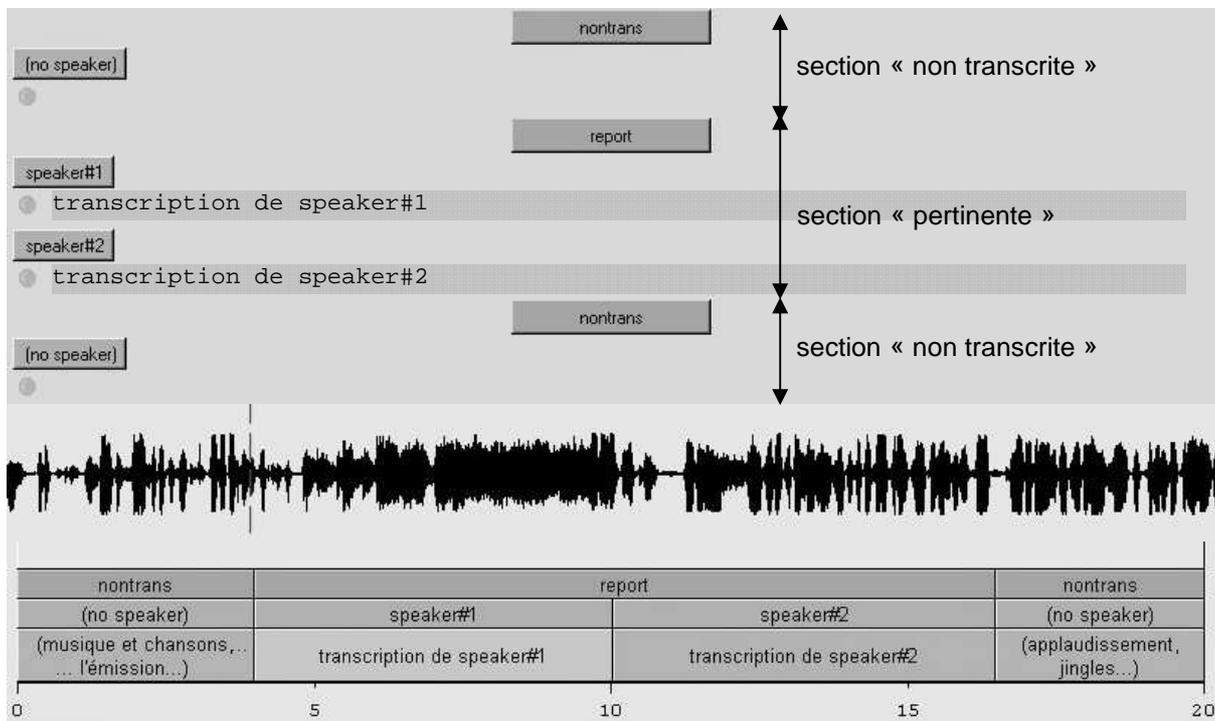
Convention 1.1	Découper le document audio en sections « pertinentes » et en sections « non transcrites ».
<p>Les sections « non transcrites » sont les sections contenant pendant plus de 15s :</p> <ul style="list-style-type: none"> - de la publicité, des passages « météo », des jingles ou des génériques d'émission - des applaudissements - de la musique, des chansons, des extraits de films, de pièces de théâtre... - du début ou de la fin d'une autre émission que l'émission cible - de la parole avec plus de deux locuteurs - de la parole dans une langue autre que le Français - du silence <p>Les autres sections sont des sections « pertinentes ».</p>	

Convention 1.2	Identifier les sections définies par un titre.
<p>Les sections « pertinentes » seront identifiées par le titre « report ».</p> <p>Les sections « non transcrites » seront identifiées par le titre « nontrans ».</p>	

Convention 1.3	Ne segmenter et ne transcrire que les sections « pertinentes ».
<p>Dans la suite du document, ne sont considérées que les sections identifiées comme « pertinentes ».</p>	



Exemple :



1.2. Tours de parole

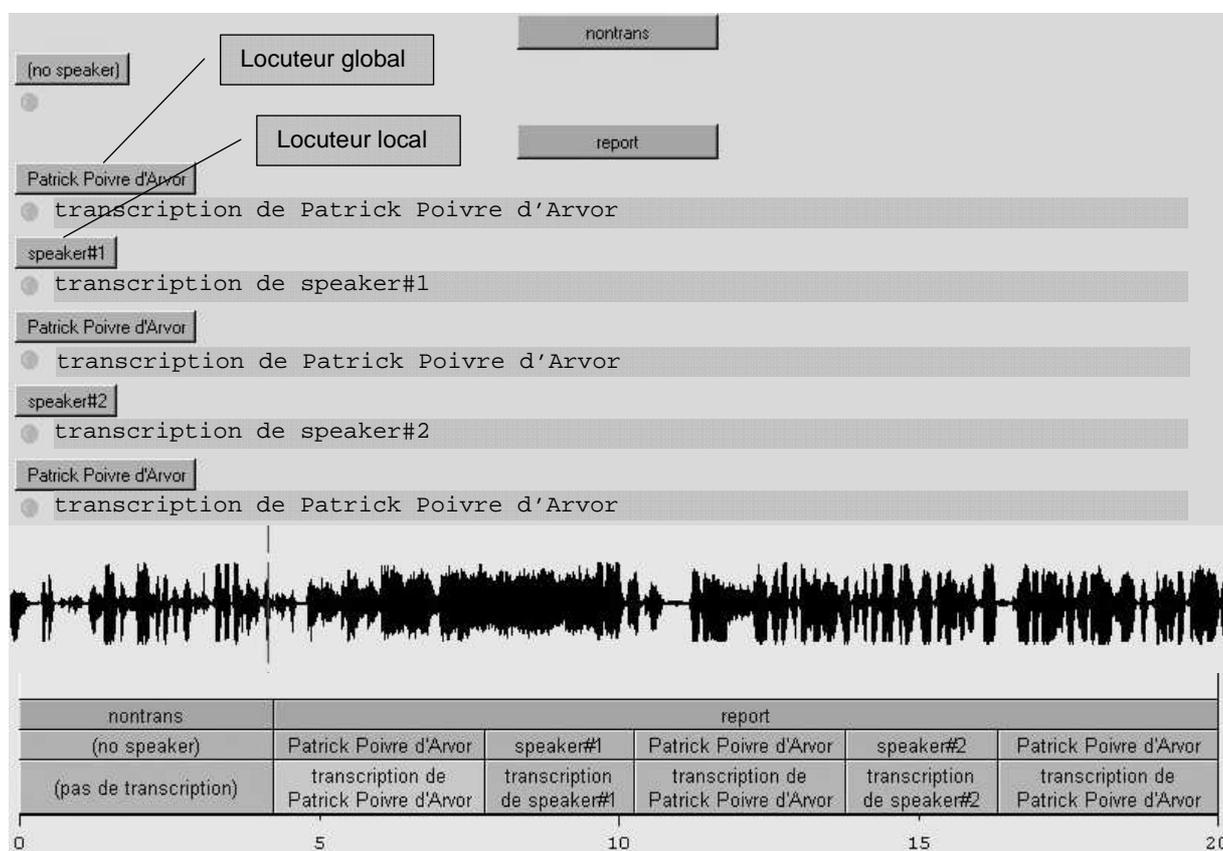
Convention 1.4	Identifier chaque locuteur intervenant dans le document audio.
<p>Les locuteurs globaux¹ sont identifiés par la syntaxe « Prénom Nom ».</p> <p>Les locuteurs locaux² sont identifiés avec la syntaxe « speaker#n », où n est le numéro de 1 à N correspondant à l'ordre d'apparition des locuteurs. S'il est possible de les identifier par leur prénom et leur nom de famille, le faire.</p> <p>Les tours de parole ne faisant intervenir aucun locuteur, sont identifiés par la syntaxe « no speaker ».</p> <p>Tous les locuteurs intervenant dans le document audio doivent être renseignés (sexe, natif...).</p>	

Convention 1.5	Effectuer le suivi des locuteurs intervenant dans le document audio.
Un même locuteur doit toujours apparaître sous le même identifiant.	

Convention 1.6	Ne faire apparaître dans la liste des locuteurs, que les locuteurs intervenant dans le document audio.
-----------------------	--



Exemple :



Convention 1.7	Indiquer la parole superposée de deux locuteurs dans un tour de parole identifié comme tel, sans en effectuer la transcription.
Les deux locuteurs concernés sont identifiés. La parole d'aucun des deux n'est transcrite.	

Convention 1.8	Ne pas inclure de silence et de bruit supérieurs à environ 0,5 seconde dans une intervention de locuteur.
Au-delà de 0,5 seconde, le silence et le bruit sont isolés dans un tour de parole « no speaker ».	

Convention 1.9	Ne pas inclure de silence supérieur à environ 0,2 seconde au début d'un segment
-----------------------	---

¹ Il s'agit de locuteurs communs à plusieurs documents audio, tels que le présentateur, les journalistes...

² Il s'agit de locuteurs propres au document audio, tels que les personnes interviewées inconnues.

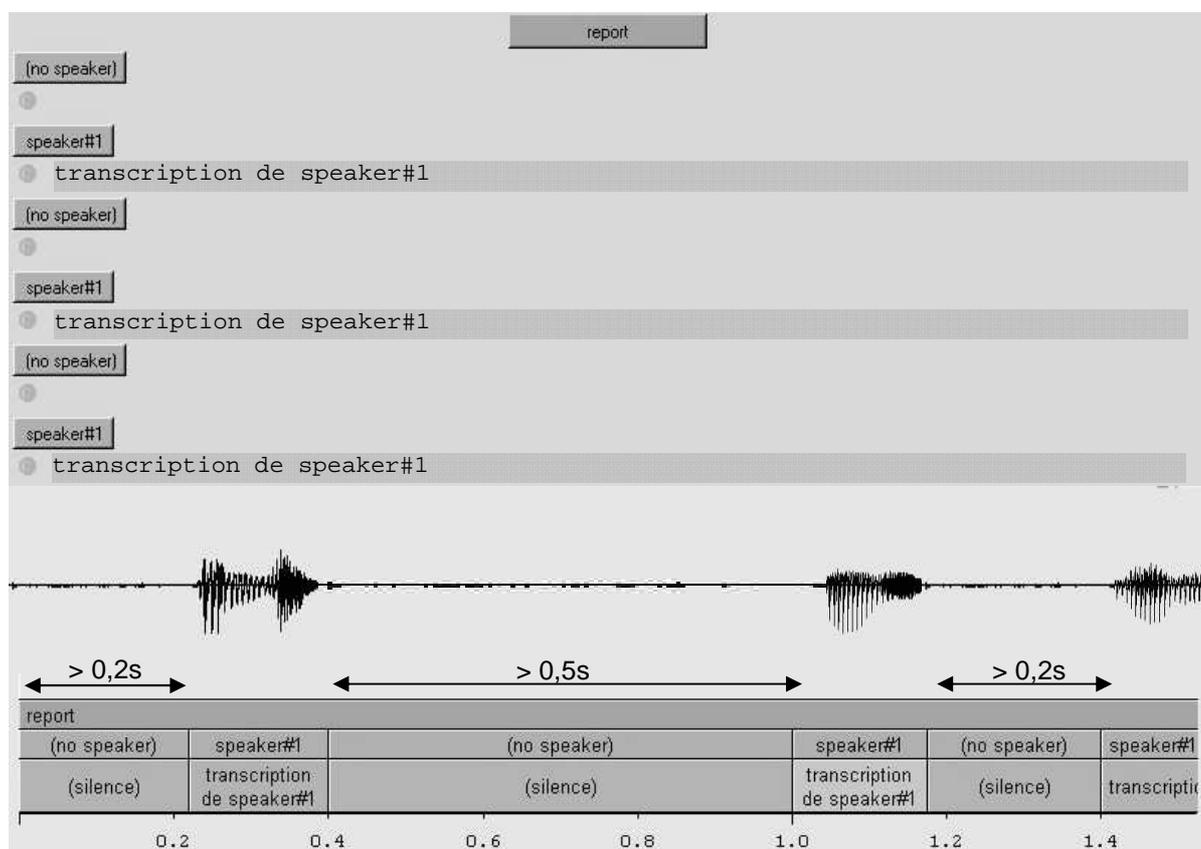
	locuteur.
Au-delà de 0,2 seconde, le silence est isolé dans un segment d'un tour de parole « no speaker » ou intégré au segment du tour de parole « no speaker » précédent.	

Convention 1.10	Ne pas inclure de silence supérieur à environ 0,2 seconde à la fin d'un segment locuteur.
Au-delà de 0,2 seconde, le silence est isolé dans un segment d'un tour de parole « no speaker » ou intégré au segment du tour de parole « no speaker » suivant.	

Convention 1.11	Laisser, dès que possible, une marge d'environ 100 ms au début et à la fin d'un segment locuteur.
------------------------	---



Exemple :



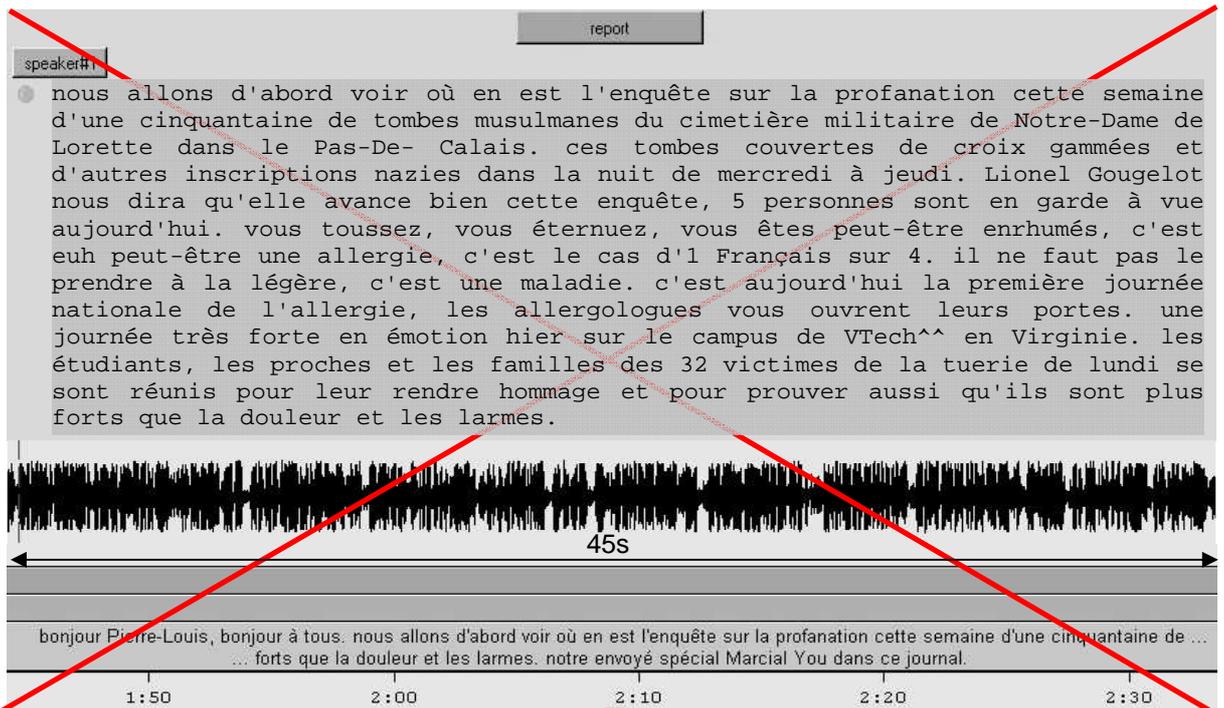
1.3. Segments

Convention 1.12	Limiter un segment d'un tour de parole à une durée maximum de 15s.
Une intervention d'un locuteur supérieure à 15 secondes doit être redécoupée en segments pertinents. Un segment pertinent contient des mots entiers et doit avoir un minimum de cohérence syntaxique et sémantique .	

Remarque : La durée des segments est en moyenne de l'ordre de 5 à 10 secondes.



Exemple :



2. REGLES DE TRANSCRIPTION

2.1. Principes

Convention 2.1 Transcrire tout ce qui est entendu (hésitation, répétition, onomatopée...).
La transcription doit correspondre fidèlement au signal.



Exemple :

 « Euh, c'est quel quel pff nu numéro ? »

~~report~~

speaker#1
euh , c'est quel numéro ?

report

speaker#1
euh , c'est quel quel & pff nu() numéro ?

Convention 2.2 Ecrire les mots en français correct.
La transcription ne doit pas contenir de faute d'orthographe ni de grammaire sur les mots.



Exemple :

 « Je suis souffrante et fatiguée à cause de tous ces événements. »

~~report~~

speaker#1
je suis souffrante et fatigué a cause de tout ses évènements.

report

speaker#1
je suis souffrante et fatiguée à cause de tous ces événements.

Remarques : - Ne pas rendre compte des élisions. (phénomène courant à l'oral).
- Ne pas chercher à corriger les erreurs grammaticales du locuteur. (erreurs courantes à l'oral)



Exemple :

 « J'demande »
« Il faut qu'i soit prêt »
« J'sais pas »
« Il faut qu'on agisse »

report

speaker#1

je demande

il faut qu'il soit prêt

je sais pas

il faut qu'on agisse

Convention 2.3	Transcrire de manière homogène.
L'écriture doit être homogène aussi bien pour les noms propres que pour les noms communs : choisir une écriture et garder la même tout au long de la transcription. Cette règle est particulièrement valable pour les mots dont l'orthographe est incertaine.	



Exemple :



« Copie du cédérom vers la clé, puis de la clé vers le cédérom. »
 « Le train va de Paris à Achères-la-Forêt et d'Achères-la-Forêt à Pau. »

report

speaker#1

~~copie du cédérom vers la clé puis de la clé vers le CDROM.~~
~~le train va de Paris à ^^Achères-la-Forêt et d'^^Achaire-la-Forêt à Pau.~~

report

speaker#1

copie du cédérom vers la clé puis de la clé vers le cédérom.
 le train va de Paris à ^^Achères-la-Forêt et d'^^Achères-la-Forêt à Pau.

2.2. Minuscules et majuscules

Convention 2.4	Ne pas mettre la majuscule en début de segment et après une ponctuation qui en nécessiterait.
L'initiale d'un mot qui est située après un point ou un point d'interrogation ne doit pas être en majuscule.	



Exemple :



« Bonjour, vous allez bien ? Et votre fils ? »

report

speaker#1

~~Bonjour, vous allez bien ? Et votre fils ?~~

report

speaker#1

bonjour, vous allez bien ? et votre fils ?

Convention 2.5	Réserver la majuscule aux sigles, aux épellations, aux noms propres, aux noms d'organisation assimilables à des marques et utilisant des noms communs (journaux, banques...) et aux formules de politesse.
Pour plus d'informations, lire le document : « Utilisation de la majuscule ».	



Exemple :



« Jean Durand vit à Ablon-sur-Seine. »
 « Le Monde et La Nouvelle République paraissent quotidiennement. »
 « Le Crédit Mutuel et La Poste ont une agence à proximité. »
 « Le monsieur qui m'a interpellé est Monsieur Durand, Docteur Durand. »

report

speaker#1

Jean Durand vit à Ablon-sur-Seine.
 Le Monde et La Nouvelle République paraissent quotidiennement.
 Le Crédit Mutuel et La Poste ont une agence à proximité.
 le monsieur qui m'a interpellé est Monsieur Durand, Docteur Durand.

2.3. Ponctuation

Convention 2.6	Ponctuer le texte avec les seules ponctuations « . », « , », « ? », « : », « ; », « ! ».
-----------------------	--

« ... » et « « » » .	
<p>Le point « . » est utilisé pour :</p> <ul style="list-style-type: none"> - marquer une pause relativement longue - signaler l'intonation descendante du locuteur - délimiter une phrase, un syntagme ou un groupe de sens important. <p>La virgule « , » est utilisée pour :</p> <ul style="list-style-type: none"> - marquer de courtes pauses - éviter une confusion sémantique si celle-ci n'est pas présente - marquer l'apposition et l'épithète détaché (fonction de la prosodie) <p><u>Exemple</u> : - les enfants, exténués, dorment malgré le bruit</p> <ul style="list-style-type: none"> - le clown, roi du cirque <ul style="list-style-type: none"> - marquer la juxtaposition <p><u>Exemple</u> : - il court, il s'affole, il crie mais n'avance à rien</p> <ul style="list-style-type: none"> - elle s'émerveille devant les lions, les tigres, les zèbres mais les girafes la laissent indifférente. <p>Le point d'interrogation « ? » est utilisé pour :</p> <ul style="list-style-type: none"> - marquer la fin d'une phrase interrogative. <p>Le double point « : » est utilisé pour :</p> <ul style="list-style-type: none"> - introduire une énumération - introduire un discours indirect <p>Le point virgule « ; » est utilisé pour :</p> <ul style="list-style-type: none"> - marquer des pauses de moyenne durée - remplacer la virgule lorsque celle-ci prêterait à confusion (notamment pour une énumération de chiffres à virgule) <p>Le point d'exclamation « ! » est utilisé pour :</p> <ul style="list-style-type: none"> - marquer la fin d'une phrase exclamative - marquer la fin d'une phrase impérative - souligner une interjection <p>Ne pas abuser de l'utilisation du point d'exclamation. N'utiliser qu'une seule itération du point d'exclamation à chaque fois (pas de « !!!! » pour marquer l'intensité de l'exclamation).</p> <p>Les points de suspension « ... » sont utilisés pour :</p> <ul style="list-style-type: none"> - marquer une interruption par un interlocuteur - marquer une interruption volontaire du locuteur - indiquer la fin d'une énumération non exhaustive <p>Les guillemets « « » » sont utilisés pour :</p> <ul style="list-style-type: none"> - identifier le discours direct rapporté. <p><u>Exemple</u> : - « vous allez être puni » , a dit le directeur.</p> <ul style="list-style-type: none"> - il a dit : « à partir de maintenant, c'est fini ». <p>Ces informations sont fortement liées à la prosodie et à la compréhension.</p>	
Convention 2.7	Ne pas séparer par des espaces les mots liés par un trait d'union ou une apostrophe.
Les mots liés par un trait d'union (mot composé, forme interrogative...) sont écrits normalement, c'est-à-dire sans séparation par un espace, ainsi que les mots reliés par une apostrophe.	

Convention 2.7	Ne pas séparer par des espaces les mots liés par un trait d'union ou une apostrophe.
Les mots liés par un trait d'union (mot composé, forme interrogative...) sont écrits normalement, c'est-à-dire sans séparation par un espace, ainsi que les mots reliés par une apostrophe.	



Exemple :



« C'est aujourd'hui après-midi que vous rencontrerez des Anglo-Saxons. »

report
speaker#1
<ul style="list-style-type: none"> c'est aujourd'hui après-midi que vous rencontrerez des Anglo-Saxons.

Convention 2.8

Ne pas relier par un trait d'union les suites de noms propres.



Exemple :



« Paris Lyon, ça se fait en deux heures. »
« La rencontre OM-Rennes s'est soldée par un match nul. »

report
speaker#1
<ul style="list-style-type: none"> Paris Lyon, ça se fait en deux heures. la rencontre OM Rennes s'est soldée par un match nul.

2.4. Nombres

Convention 2.9

Transcrire les nombres sous forme de chiffres (sans espaces entre les chiffres) ou de lettres (sans faute d'orthographe).

L'usage des lettres peut permettre d'écrire les grands nombres ronds sans oublier de zéro : « 100 millions », « 398 milliards ».

L'usage des chiffres peut permettre d'écrire les nombres plus facilement et sans faute d'orthographe.

Remarque : Ne pas utiliser de chiffres dans l'écriture des adjectifs numéraux ordinaux (premier, deuxième, centième...).



Exemple :



« Il y avait 10.000 médecins pour 40.000.000 d'habitants. »
« En 1778, c'était le 3^{ème} chef du XVIII^{ème} siècle. »

report
speaker#1
<ul style="list-style-type: none"> il y avait 10 mille médecins pour 40 millions d'habitants. en 1778, c'était le troisième chef du dix-huitième siècle.

Convention 2.10

Normaliser l'écriture des nombres et insérer la prononciation entendue dans une balise de prononciation pour toute prononciation particulière ou lorsque des mots sont présents entre les chiffres.

Liste non exhaustive :

- les unités de mesures : mètre, kilo, litre, euro...
- les ponctuations : virgule, point...
- les signes d'opération : fois, plus, moins, égale...
- les fractions : quart, tiers, moitié...
- prononciation particulière : 19 100 78, nonante-six pour 96...

Exception : Le mot « pourcent » est indiqué avec le symbole « % » et est collé au chiffre qui le précède. Ne pas insérer de balise de prononciation pour les pourcentages.



Exemple :



« 1 euro 50, soit 3% »
« 158 virgule 9 litres »
« -10 degrés Celsius »
« entre 2 tiers et 1 quart »
« Henri IV et Louis XIV »

« 29 zéro neuf dix neuf cent soixante seize »
 « je suis né le 29 septembre 1976 »
 « rendez-vous à 17h36 »

report
speaker#1
1,50euro +[pron=1 euro 50], soit 3%
158,9L +[pron=158 virgule 9 litres]
-10°C +[pron=moins 10 degrés Celsius]
entre 2/3 +[pron=2 tiers] et 1/4+[pron=1 quart]
Henri IV +[pron=4] et Louis XIV +[pron=14]
29/09/1976 +[pron=29 0 9 19 100 76]
Je suis né le 29 septembre 1976
rendez-vous à 17h36 +[pron=17 heures 36]

2.5. Adresses Internet

Convention 2.11	Normaliser l'écriture des adresses Internet et insérer la prononciation entendue dans une balise de prononciation.
------------------------	--



Exemple :



« http://www.google.fr »
 « monnom@monadresse.net »
 « prénom_nom@adresse.com »

report
speaker#1
http://www.google.fr +[pron=HTTP 2 points slash slash WWW point google point FR]
monnom@monadresse.net +[pron=monnom at monadresse point net]
prénom_nom@adresse.com +[pron=prénom underscore nom arobase adresse point com]

Aide : Le « @ » constitue un mot à part entière, il est noté suivant sa prononciation :

- at
- arobase

Le « www » et le « http » sont transcrits comme des sigles :

- WWW
- HTTP

La ponctuation (« . », « : », « / », « _ »...) est transcrite sous forme de mots (et non de signes) :

- point
- 2 points
- slash
- underscore

Le « fr », le « com » et le « net » sont transcrits comme suit :

- FR
- Com
- net

2.6. Sigles

Convention 2.12	<ul style="list-style-type: none"> - Ecrire en majuscule sans espace ou point de séparation, tout sigle lu de 4 lettres ou moins. - Ecrire avec seulement la première lettre en majuscule, tout sigle lu de 5 lettres ou plus. - Ecrire en majuscule sans espace ou point de séparation, tout sigle épilé.
------------------------	---



Exemple :



« IFOP, SMIG, SFOR, PACS, ENA »
 « MEDEF, INSEE, STAPS, ASSEDI »
 « FM, ETA, CFDT, EHESS, FNSEA, DGCCRF »

sigles lus
 sigles lus
 sigles épelés

report

speaker#1

- IFOP, SMIG, SFOR, PACS, ENA
- Medef, Insee, Staps, Assedic
- FM, ETA, CFDT, EHESS, FNSEA, DGCCRF

2.7. Epellation, abréviations

Convention 2.13	Ecrire toute épellation en majuscules, avec un espace entre chaque lettre.
------------------------	--

Remarque : Dans le cas d'une épellation où le redoublement de la lettre est indiqué avec un chiffre, accoler le chiffre à la lettre.



Exemple :



« Il va de Paris, P A R I S, à Vannes, V A deux N E S, chez AG2R, A G 2 R. »
 « je m'appelle Ester, É S T E R »
 « je m'appelle Ester, E S T E R »

report

speaker#1

- il va de Paris, P A R I S, à Vannes, V A 2N E S, chez AG2R, A G 2 R.
- je m'appelle Ester, É S T E R.
- je m'appelle Ester, E S T E R.

Convention 2.14	Utiliser les abréviations standard de la langue française, comme « etc » pour « et caetera » ou « OK » pour « okay ».
------------------------	---

2.8. Onomatopées et interjections

Convention 2.15	Transcrire les onomatopées et interjections en les précédant du signe &.
------------------------	--

Choisir les onomatopées et les interjections parmi les listes suivantes :

&ah ; &aïe ; &atchoum ; &baf ; &bah ; &bé ; &ben ; &bof ; &bouh ; &hé ; &hein ; &hm ; &hop ; &hou ; &la ; mah, &mouais ; &moui ; &of ; &oh ; &ouah ; &ouais ; &ouf ; &ouille ; &pff ; &pouh ; &snif ; &wahou ; &yeah ; &zut

La transcription de ces onomatopées ou de ces interjections ne doit pas refléter la manière dont le locuteur prononce le mot :

- ne pas noter « AIE » pour souligner la force de la prononciation
- ne pas transcrire « ooooooooooooooh » pour montrer la durée

Remarque : Si une nouvelle onomatopée ou une nouvelle interjection apparaît dans le document audio, la transcrire le plus simplement possible.



Exemple :



« tadaditadada je sais pas. »

report

speaker#1

- &tadaditadada je sais pas.

2.9. Hésitations, répétitions

Convention 2.16	Transcrire les hésitations grâce à « euh » et « hum ».
Ne pas confondre « hum » et « &hm » :	
- hum est utilisé pour les hésitations	
- &hm est utilisé lors d'acquiescements	

Convention 2.17	Transcrire tous les mots entendus même s'ils sont répétés.
------------------------	--



Exemple :



« Le le le député veut y aller y aller pour se rassurer. »

report

speaker#1

le député veut y aller pour se rassurer.

report

speaker#1

le le le député veut y aller y aller pour se rassurer.

2.10. Troncation

Convention 2.18	Marquer la troncation d'un mot par des parenthèses en début ou fin de mot.
Troncatures finales : vou() pour voula <u>is</u> ou voudra <u>is</u> ...	
Troncatures initiales : ()embre pour <u>se</u> ptembre ou <u>no</u> vembre...	
Un mot complet n'apparaît jamais entre parenthèses .	
Pas de troncature au milieu d'un groupe de lettre correspondant au même son.	

Remarque : Si grâce au contexte, le mot tronqué est connu, noter la partie manquante du mot entre parenthèses.



Exemple :



« Ce magnnn p'tit plateau est pas à ranger dans la car... »

report

speaker#1

ce mag(nifique) p(e)tit plat(e)au (n')est pas à range(r) dans la car().

report

speaker#1

ce magn(ifique) petit plateau est pas à ranger dans la car().

2.11. Prononciation erronée

Convention 2.19	Ecrire correctement le mot identifié et lui accoler une étoile avant la première lettre.
Pour un groupe de mots mal prononcés, chaque mot sera précédé d'une étoile.	

Remarque : Il peut arriver que l'on soit obligé d'associer deux règles pour un même mot : celle pour la troncation () et celle pour la mauvaise prononciation *



Exemple :



« Le pest euh spectacle altistique est présenté par Martin Dupond. »

report
speaker#1
le *spect(acle) euh spectacle *artistique est présenté par *Martin Dupond.

2.12. Lapsus

Convention 2.20	Ne pas corriger les lapsus. Transcrire le mot tel qu'il a été prononcé.
------------------------	---



Exemple :



« C'est une augmentation conséquence euh conséquente. »

report
speaker#1
c'est une augmentation conséquence euh conséquente.

2.13. Elisions et formes réduites

Convention 2.21	Respecter l'orthographe standard même si les mots font l'objet à l'oral d'élosion ou de forme réduite.
La réduction peut porter sur un seul mot ou sur une suite de mots.	

Remarque : Ne pas transcrire le « ne » de la négation s'il n'est pas prononcé.



Exemple :



« T'as l'air embêté que l'développement s'ralentisse. »
 « Dans c'domaine, s'i vous plaît, faites pas bêtises. »
 « J'en sais rien, j'demande à voir. »
 « T'as pas l'air au courant, y s'défile. »

report
speaker#1
tu as l'air embêté que le développement se ralentisse.
dans ce domaine, s'il vous plaît, faites pas de bêtises.
j'en sais rien, je demande à voir.
tu as pas l'air au courant, il se défile.

2.14. Mots inconnus ou absents du lexique

Convention 2.22	Transcrire le mot inconnu avec l'orthographe la plus proche de ce qui est entendu, et précéder le mot de deux accents circonflexes.
Un mot inconnu est un mot absent du lexique. Pour les expressions et mots composés pour lesquels l'orthographe de l'ensemble des mots est incertaine, mettre les accents circonflexes à chacun des mots composant l'expression. Cette règle s'applique en particulier pour l'écriture des noms propres.	

Remarque : L'orthographe des mots inconnus doit rester homogène dans tout le document audio transcrit.



Exemple :



« Dacim Houssila et Francine Mokoko »
« T'es ouf ou quoi, mon appart est vide ! »

report
speaker#1
<ul style="list-style-type: none"> ● ^^Dassim ^^Oussila et Francine ^^Mokoko ● tu es ^^ouf ou quoi, mon ^^appart est vide.

Convention 2.23	Isoler et utiliser une balise [lang=xxx] pour un mot ou quelques mots en langue étrangère « xxx ».
<p>Il est possible de transcrire ces mots en langue étrangère à l'intérieur de la balise [lang=xxx]. Si la langue étrangère est indéterminée, remplacer xxx par « unknown ». Les mots étrangers dont l'usage est courant en Français ne sont pas concernés par cette règle : <i>Exemple : parking, week-end, planning, manager, sandwich...</i></p>	

3. REGLES DE BALISAGE

3.1. Bruits

Convention 3.1	Insérer une balise [r] pour indiquer une respiration, une inspiration ou une expiration du locuteur.
-----------------------	--

Convention 3.2	Insérer une balise [b] pour indiquer un bruit.
<p>Les bruits de l'appareil phonatoire :</p> <ul style="list-style-type: none"> - bruits de bouche (toux, bruit de gorge, rire du locuteurs, sifflement du locuteur, chuchotement...) <p>Les bruits extérieurs :</p> <ul style="list-style-type: none"> - indéterminés - froissement de papiers - bruits de micro 	

Remarque : La balise peut être instantanée pour un bruit isolé dans le signal ou encadrante sur quelques mots pour un bruit qui se superpose à la parole.

Convention 3.3	Dans le cas de parole en français en premier plan, simultanément avec de la parole en fond (langue étrangère pour de la traduction, discours commenté...), transcrire la parole du premier plan et en français. Indiquer la conversation en fond par la balise encadrante [conv].
En général, pour la traduction simultanée, le segment en question est précédé d'un segment en langue étrangère contenant la balise [lang=xxx] où « xxx » est la langue étrangère (voir TR_DET_TRS_2.14_02).	

Convention 3.4	Isoler dans un segment les parties du signal encadrées par des balises décrivant des phénomènes acoustiques.
- Il s'agit des balises [b] et [conv].	

3.2. Problèmes de prononciation

Convention 3.5	Insérer une balise [pron=pi] pour toute parole inintelligible ou inaudible.
-----------------------	---

Convention 3.6	Associer la balise [pron=pi] à un locuteur et l'isoler dans un segment.
-----------------------	---



Exemple :



« L'actualité du jour nous conduit à vous prxxx lxxxx jxxxx qui devrait s'arranger prochainement. »

report
speaker#1
● l'actualité du jour nous conduit à vous
● [pron=pi]
● qui devrait s'arranger prochainement.

3.3. Back Channel

On utilise le back channel lorsque pendant l'intervention du locuteur principal, l'autre locuteur intervient pour acquiescer ou montrer son désaccord avec les propos du premier par des interjections (&hm &hm, OK, oui, d'accord, ...) ou des répétitions. Le back channel a pour fonction de montrer au locuteur principal que son interlocuteur :

- suit bien la conversation,
- comprend ses propos,
- est en désaccord ou non,
- ou veut intervenir.

Convention 3.7	Isoler les interventions de chacun des locuteurs dès que possible.
- Le signal est alors segmenté et les interventions de chacun sont transcrites.	



Exemple :

Agent
● 100 rue de
Client
● oui
Agent
● Richelieu
Client
● Richelieu
Agent
● 75000
Client
● hum
Agent
● 09 Paris

Convention 3.8	Si les acquiescements du locuteur se superposent aux interventions du locuteur principal en rendant la segmentation délicate, transcrire l'intervention du locuteur principal et insérer des balises [bc] pour remplacer les acquiescements de l'autre locuteur.
-----------------------	--



Exemple :

Agent
● 100 rue de + [bc] Richelieu [bc] 75000 + [bc] 09 Paris

3.4. Commentaires

Convention 3.9	Insérer une balise de commentaires [comment] pour indiquer tout événement rencontré dans le document audio, qui ne serait pas décrit dans les règles précédentes.
Il peut s'agir : <ul style="list-style-type: none">- d'une situation non décrite- d'une ambiguïté- d'un problème relatif au document audio- ...	

3.5. Disfluences

On appelle disfluences l'ensemble des phénomènes normaux liés au mode habituel de production de l'oral : hésitations, répétitions, amorces (mots interrompus), autocorrections, inachèvements... Les disfluences sont des phénomènes normaux du langage spontané (non préparé, non lu), qu'il ne faut pas confondre avec les accidents liés à des pathologies (aphasies, bégaiement, etc.). L'intérêt du balisage des disfluences est de signifier que ce sont des mots, parties de mot ou groupes de mots qui peuvent être supprimés à l'écrit sans modifier le sens de l'énoncé. La contrainte est donc de maintenir la lisibilité du texte après suppression des disfluences.

3.5.1. Répétitions

Convention 3.10	Encadrer toute répétition involontaire du locuteur par la balise [dis=rep].
Ne pas baliser la dernière occurrence de répétition. Ne pas oublier les répétitions de mots qui concernent : <ul style="list-style-type: none">- les mots tronqués- les répétitions discontinues (après des incises) Ne pas baliser les répétitions emphatiques et les renforcements d'énoncé. S'aider de la prosodie pour repérer ces répétitions volontaires du locuteur. En général, les répétitions disfluentes se suivent en continu, sans rupture intonative et ne possèdent pas de valeur sémantique ou de valeur syntaxique.	



Exemple :



« Ca veut dire, ça veut dire que le le le député a fait des éco économies importantes ! »
« Pourquoi, je sais bien que c'était très atroce, pourquoi spécialement cette expérience ? »
« D'accord, d'accord, je le reconnais, je suis vraiment très très très heureuse de vous avoir rencontré ! »

~~report~~

speaker#1

- ① [dis=rep-] ça veut dire ça veut dire [-dis=rep] que [dis=rep-] le le [-dis=rep] le député a fait des éco [dis=rep] économies importantes !
- ② pourquoi, je sais bien que c'était très atroce, pourquoi spécialement cette expérience ?
- ③ [dis=rep-] d'accord [-dis=rep] d'accord, je le reconnais, je suis vraiment [dis=rep-] très très [-dis=rep] très heureuse de vous avoir rencontré.



report

speaker#1

- ① [dis=rep-] ça veut dire [-dis=rep] ça veut dire que [dis=rep-] le le [-dis=rep] le député a fait des [dis=rep-] éco(nomies) [-dis=rep] économies
- ② [dis=rep-] pourquoi [-dis=rep] je sais bien que c'était très atroce, pourquoi spécialement cette expérience ?
- ③ d'accord, d'accord, je le reconnais, je suis vraiment très très très heureuse de vous avoir rencontré.

3.5.2. Hésitations (ou pauses remplies)

Convention 3.11

Les seules hésitations sont les mots « euh » et « hum ». Encadrer toute hésitation (ou pause remplies) du locuteur par la balise [dis=hes].

Rappel : Ne pas confondre « hum » et « &hm » : par Convention , le « &hm » a pour fonction d'acquiescer les propos de l'autre locuteur, il ne s'agit pas d'une hésitation alors que le « hum » a pour fonction d'indiquer une hésitation .



Exemple :



« Hum, je sais pas non, euh pourquoi ? »

```
report
speaker#1
[dis=hes-] hum [-dis=hes], je sais pas non, [dis=hes-] euh [-dis=hes],
```

3.5.3. Révisions

Les révisions concernent **des fragments de mots, des mots ou des séries de mots** dont la prononciation est interrompue spontanément par le locuteur pour se corriger par la suite.

Convention 3.12

Encadrer toute révision du locuteur par la balise [dis=rev].

Remarque : Les phrases non terminées à cause d'une interruption par un autre locuteur, ne rentrent pas dans cette catégorie.



Exemple :



« Si on cons si on compte »
« Il va perdre un perdre son revenu. »
« J'ai un une excellente recette ! »
« Si on fait pourquoi fais-tu cela ? »

```
report
speaker#1
[dis=rev-] si on cons(idère) [-dis=rev] si on compte
il va [dis=rev-] perdre un [-dis=rev] perdre son revenu.
j'ai [dis=rev-] un [-dis=rev] une excellente recette !
[dis=rev-] si on fait [-dis=rev] pourquoi fais-tu cela ?
```

3.5.4. Marqueurs de discours

Les marqueurs de discours sont soit des **connecteurs logiques** comme « donc », « alors »... soit des **particules** telles que « &ben », « voilà »... Le marqueur de discours sert à **structurer l'oral**. Il existe deux catégories de marqueurs de discours :

- ceux qui peuvent être omis à l'écrit
- ceux qui ne le peuvent pas

Seuls ceux qui peuvent être omis à l'écrit sont à baliser.

Convention 3.13

Encadrer tout marqueur de discours du locuteur par la balise [dis=dm]. Regrouper les marqueurs de discours accolés dans une même balise.

Aide : Quelques marqueurs de discours courants sont « alors », « bien », « bon &ben », « d'ailleurs », « disons », « donc », « enfin », « et », « &hé bien », « &hein », « je crois », « je pense », « mais » « moi », « voilà », « vous savez »



Exemple :



« Disons qu'il est insupportable. »
 « Et j'en sais rien. »
 « Ben voilà, c'est tout à fait ça ! »
 « Je veux dire, il est charmant. »

report	
speaker#1	
①	[dis=dm-] disons qu' [-dis=dm] il est insupportable.
②	[dis=dm-] et [-dis=dm] j'en sais rien.
③	[dis=dm-] &ben voilà [-dis=dm] c'est tout à fait ça !
④	[dis=dm-] alors [-dis=dm] il est charmant.

3.5.5. Combinaisons de disfluences

Le but est de repérer les disfluences pour pouvoir les supprimer ultérieurement.

Convention 3.14	Isoler, tant que faire ce peut, les quatre phénomènes de disfluences dans leurs balises respectives.
------------------------	--

Convention 3.15	Le cas échéant prendre en considération les exceptions ci-dessous.
Pour les répétitions :	
<ul style="list-style-type: none"> - ne pas indiquer la répétition des hésitations - ne pas indiquer une répétition qui intervient dans une révision - ne pas indiquer la répétition des marqueurs de discours 	
Pour les hésitations (ou pauses remplies) :	
<ul style="list-style-type: none"> - ne pas indiquer une hésitation qui apparaît au milieu d'une répétition - ne pas indiquer une hésitation qui apparaît au milieu d'une révision 	

Convention 3.16	Indiquer seulement par une balise [dis] lorsque l'identification paraît difficile .
------------------------	---



Exemple :



« La la euh la la douche est libre »
 « La la la le jour se lève »
 « Ce n'est ce ce ce n'est c'est un domaine »
 « A ce moment euh à ce dans ce cas »
 « Euh euh je suis pas sûr »
 « Euh nous sommes nous sommes 6 »
 « Bon bah, euh oui pourquoi pas »
 « Bah euh bon c'est ok »

report	
speaker#1	
①	[dis=rep-] la la euh la [-dis=rep] la douche est libre
②	[dis=rev-] la la la [-dis=rev] le jour se lève
③	[dis=rev-] ce n'est ce ce n'est [-dis=rev] c'est un domaine
④	[dis=rev-] à ce moment euh à ce [-dis=rev] dans ce cas
⑤	[dis=hes-] euh euh [-dis=hes] je suis pas sûr
⑥	[dis=hes-] euh [-dis=hes] [dis=rep-] nous sommes [-dis=rep] nous sommes 6
⑦	[dis=dm-] bon &bah [-dis=dm] [dis=hes-] euh [-dis=hes] oui pourquoi pas
⑧	[dis=dm-] &bah [-dis=dm] [dis=hes-] euh [-dis=hes] [dis=dm-] bon [-dis=dm]