

VECSYS - ZA de Courtaboeuf - Les Ulis

3, rue de la Terre de Feu - 91952 Courtaboeuf Cedex

Tél. : + 33 (0)1 69 29 87 87

Fax : + 33 (0)1 69 07 58 58

[http : // www.vecsys.fr](http://www.vecsys.fr)

vecsys@vecsys.fr

S.A. au capital de 1.011.750 € - R.C. Evry B 316676741



Projet : ERNESTUS

TRANSCRIPTION DETAILLEE

CONVENTION

Rédigé par :	
DARDON, Soledad	
Date :	
Visa :	

Vérifié par :	
GARNIER, Martine	
Date :	
Visa :	

Approuvé par :	
GARNIER, Martine	
Date :	
Visa :	

Fichier de base : transcription_detaillée_A0.doc

Ce document est la propriété de VECSYS S.A. Il ne peut être communiqué à des tiers sans autorisation écrite.			Diffusion :		Large
				x	Restreinte
					Confidentielle
Référence :	Indice :	Date d'édition :			Page :
	001	17/12/2007			1 / 25

RÉVISION DU DOCUMENT

INDICE	DATE	PAGES MODIFIÉES	OBJET	AUTEUR
001	17/12/2007	Toutes	Création du document	S. DARDON

SOMMAIRE

1. PRESENTATION GENERALE	4
2. REGLES DE DECOUPAGE.....	5
2.1. SECTIONS	5
2.2. TOURS DE PAROLE.....	5
2.3. SEGMENTS	5
3. REGLES DE TRANSCRIPTION	5
3.1. PRINCIPES.....	5
3.2. REGLES D'ECRITURE	5
3.2.1. Minuscules et majuscules.....	5
3.2.2. Ponctuation.....	5
3.2.3. Nombres	5
3.2.4. Adresses Internet.....	5
3.2.5. Sigles	5
3.2.6. Epellation	5
3.2.7. Abréviations	5
3.2.8. Onomatopées et interjections	5
3.2.9. Hésitations	5
3.2.10. Répétitions	5
3.2.11. Troncation.....	5
3.2.12. Prononciation erronée	5
3.2.13. Lapsus	5
3.2.14. Elisions et formes réduites.....	5
3.2.15. Mots inconnus ou absents du lexique	5
4. REGLES DE BALISAGE	5
4.1. CONDITIONS ACOUSTIQUES	5
4.2. RESPIRATION.....	5
4.3. RIRES	5
4.4. BRUITS.....	5
4.5. PAROLE EN BRUIT DE FOND	5
4.6. PROBLEME DE PRONONCIATION	5
4.7. REGLES DE SEGMENTATION LIEES AUX BALISES	5
4.8. COMMENTAIRES	5

1. PRESENTATION GENERALE

Dans le cadre du projet Ernestus, il s'agit de définir les conventions de transcription détaillée pour le corpus de documents audio.

Ce document est constitué de quatre chapitres :

- Chapitre 1 : le présent chapitre de présentation
- Chapitre 2 : les règles de découpage
- Chapitre 3 : les règles de transcription
- Chapitre 4 : les règles de balisage

2. REGLES DE DECOUPAGE

2.1. SECTIONS

TR_DET_DEC_1_01-a	Découper le document audio en sections « pertinentes » et en sections « non pertinentes ».
Les sections « non pertinentes » correspondent : <ul style="list-style-type: none">- aux échanges en langue étrangère d'une durée supérieure à 15 secondes,- du bruit ou du silence supérieurs à 15 secondes. Les autres sections sont des sections « pertinentes ».	
TR_DET_DEC_1_02-a	Identifier les sections définies par un titre.
Les sections « pertinentes » seront identifiées par le titre « report ». Les sections « non pertinentes » seront identifiées par le titre « non trans ».	
TR_DET_DEC_1_03-a	Ne segmenter et ne transcrire que les sections « pertinentes ».
Dans la suite du document, ne sont considérées que les sections identifiées comme « pertinentes ».	

2.2. TOURS DE PAROLE

TR_DET_DEC_2_01-a	Segmenter et identifier chaque locuteur intervenant dans le document audio.
Identifier le locuteur par son prénom. Les tours de parole ne faisant intervenir aucun locuteur, sont identifiés par la syntaxe « no speaker ». Tous les locuteurs intervenant dans le document audio doivent être renseignés (sexe, natif...).	

TR_DET_DEC_2_02-a	Effectuer le suivi des locuteurs intervenant dans le document audio.
Un même locuteur doit toujours apparaître sous le même identifiant.	

TR_DET_DEC_2_03-a	Ne faire apparaître dans le dictionnaire local des locuteurs, que les locuteurs intervenant dans le document audio.
-	



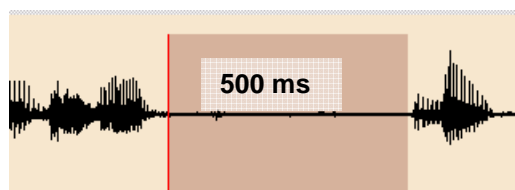
Exemple :



TR_DET_DEC_2_04-a	Ne pas inclure de silence et de bruit supérieurs à environ 500 ms dans une intervention de locuteur.
Au-delà de 500 ms, le silence et/ou le bruit sont isolés dans un tour de parole « no speaker ».	



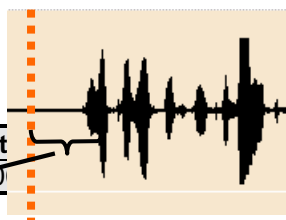
Exemple :



TR_DET_DEC_2_05-a	Ne pas inclure de silence et de bruit supérieurs à environ 200 ms au début d'un segment locuteur.
Au-delà de 200 ms, 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.	



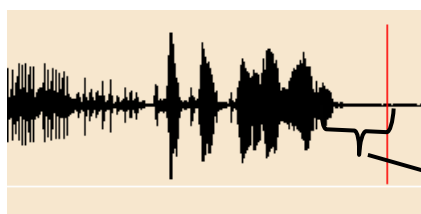
Exemple :



TR_DET_DEC_2_06-a

Ne pas inclure de silence et de bruit supérieurs à environ 200 ms à la fin d'un segment locuteur.

Au-delà de 200 ms, 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.

*Exemple :*

Moins de 200 ms de
silence on accepte la
frontière.

TR_DET_DEC_2_07-a

Laisser, si possible, une « marge » d'environ 100 ms au début et à la fin d'un segment locuteur.

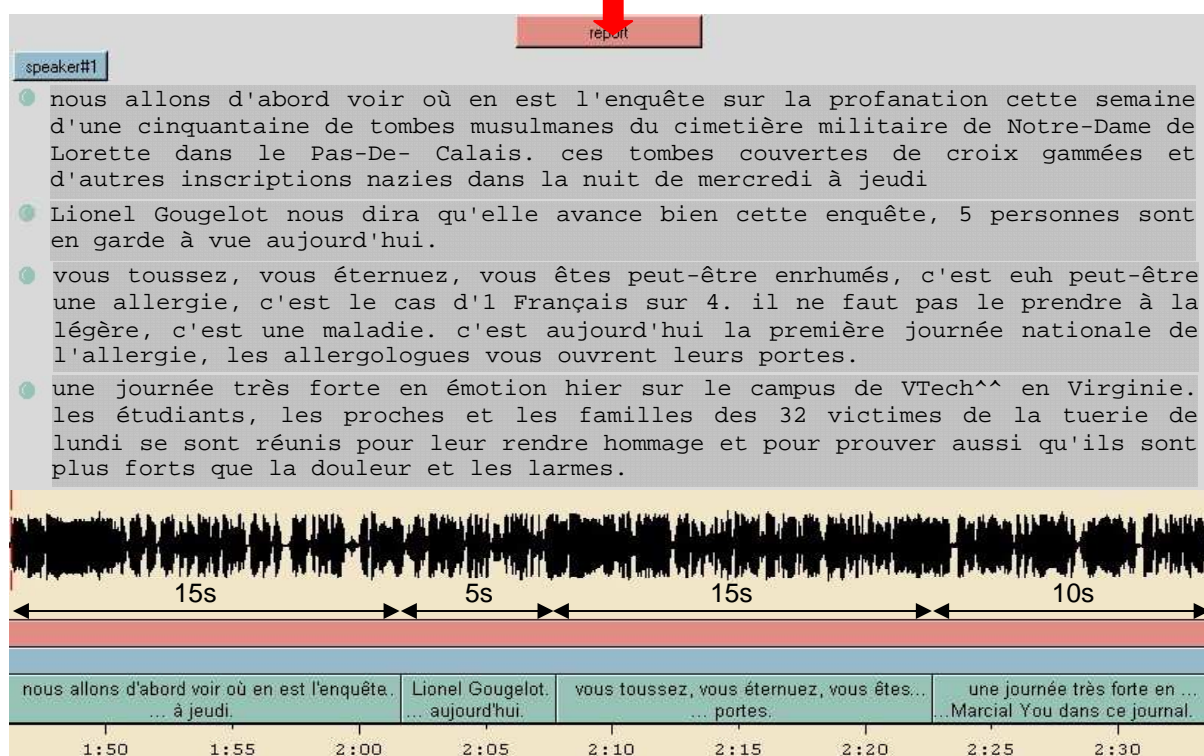
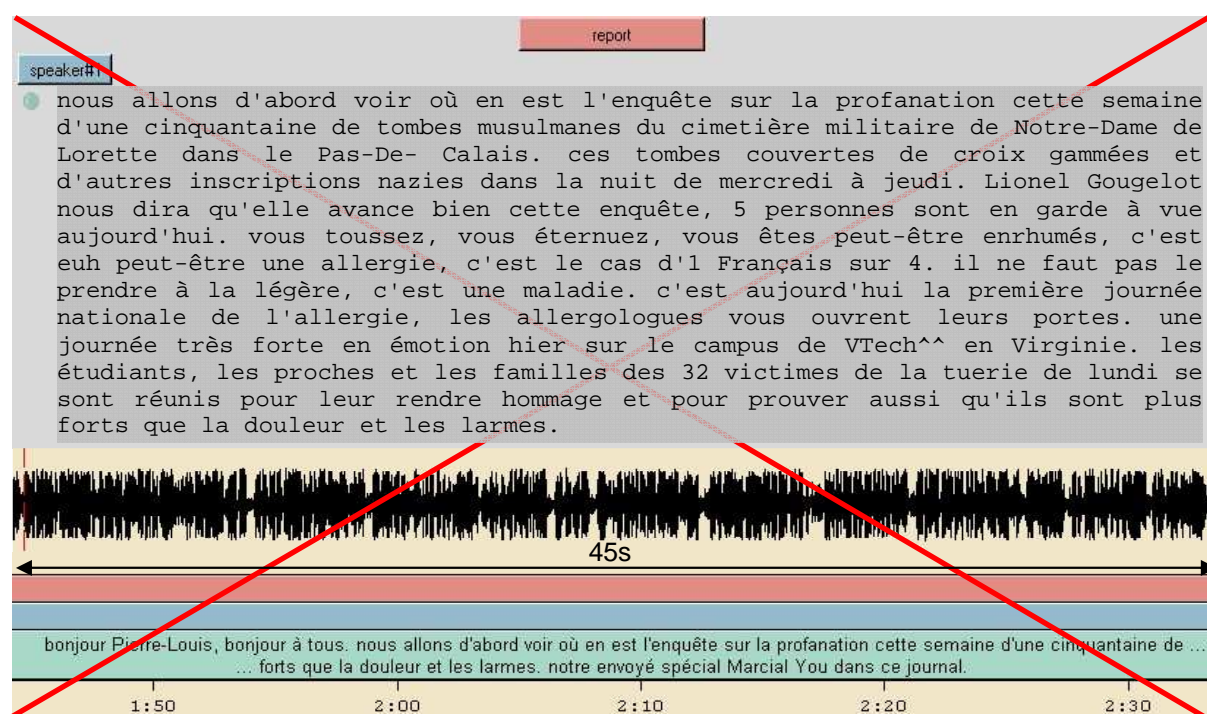
2.3. SEGMENTS

TR_DET_DEC_3_01-a 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**.



Exemple :



Référence :

Indice :

Date d'édition :

Page :

001

17/12/2007

8 / 25

3. REGLES DE TRANSCRIPTION

3.1. PRINCIPES

TR_DET_TRS_1_01-a	Transcrire tout ce qui est prononcé.
La transcription doit correspondre fidèlement au signal.	
Chaque mot entendu (hésitation, répétition, onomatopée...) doit être transcrit.	



Exemple :



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

Diagram illustrating the transcription rule for capturing all heard words, including hesitations and repetitions.

The top part shows a transcription of the audio input: "euh , c'est quel numéro ?". This transcription is crossed out with a large red 'X', indicating it is incorrect because it does not capture all the heard words (hesitations and repetitions).

The bottom part shows the correct transcription: "euh , c'est quel quel &pff nu() numéro ?". This transcription is not crossed out, indicating it is correct because it captures all the heard words, including the hesitation "euh", the repetition "quel quel", the onomatopoeia "&pff", and the repetition "nu()".

A red arrow points from the top (incorrect) transcription to the bottom (correct) transcription.

TR_DET_TRS_1_02-a	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. »

Diagram illustrating the transcription rule for writing words in correct French, including spelling and grammar.

The top part shows a transcription of the audio input: "je suis souffrante et fatigué a cause de tout ses évènements.". This transcription is crossed out with a large red 'X', indicating it is incorrect because it contains spelling and grammar errors (e.g., "souffrante" instead of "souffrante", "fatigué" instead of "fatiguée", "a" instead of "à", "tout" instead of "tous", "ses" instead of "ces", "évènements" instead of "événements").

The bottom part shows the correct transcription: "je suis souffrante et fatiguée à cause de tous ces événements.". This transcription is not crossed out, indicating it is correct because it contains no spelling or grammar errors.

A red arrow points from the top (incorrect) transcription to the bottom (correct) transcription.

REMARQUE TR_DET_TRS_1_02

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 :

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

Before correction:

report

speaker#1

- 1 y a pas
- 2 il faut qui soit prêt
- 3 je ne sais pas
- 4 il faut que l'on agisse

After correction:

report

speaker#1

- 1 il y a pas
- 2 il faut qu'il soit prêt
- 3 je sais pas
- 4 il faut qu'on agisse

TR_DET_TRS_1_03-a

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. »

Before correction:

report

speaker#1

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

After correction:

report

speaker#1

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

3.2. REGLES D'ECRITURE

3.2.1. Minuscules et majuscules

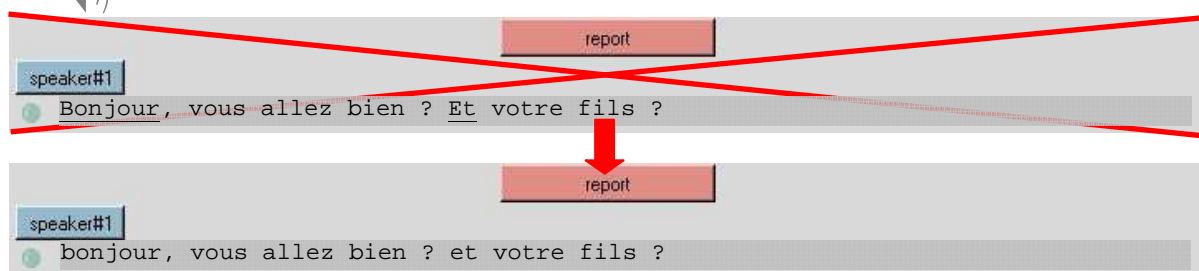
TR_DET_TRS_2.1_01-a	Ne pas mettre la majuscule en début de segment et après une ponctuation qui en nécessiterait.
L'initiale d'un mot qui est située après un point ou un point d'interrogation ne doit pas être en majuscule.	



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 ?

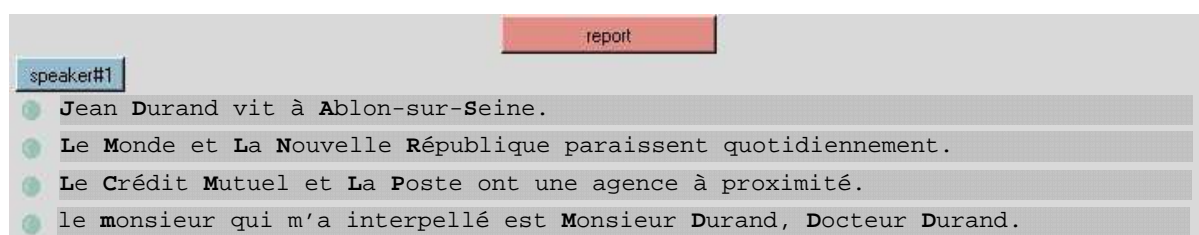
TR_DET_TRS_2.1_02-a	Réserver la majuscule aux usages ci-dessous.
<ul style="list-style-type: none"> - Les sigles (voir TR_DET_TRS_2.5_01) - Les épellations (voir TR_DET_TRS_2.6_01) - Les noms propres - Les noms propres d'organisation assimilables à des marques et utilisant des noms communs (journaux, banques...) - Les formules de politesse 	



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.

3.2.2. Ponctuation

TR_DET_TRS_2.2_01-a	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 <p><u>Exemple :</u></p> <ul style="list-style-type: none"> - Les élèves, qui avaient oublié leurs livres, ont été punis. => (Tous les élèves ont oublié leurs livres et tous ont été punis.) - Les élèves qui avaient oublié leurs livres ont été punis. => (Seuls les élèves qui ont oublié leurs livres sont punis, les autres les avaient avec eux.) - marquer l'apposition et l'épithète détaché (fonction de la prosodie) <p><u>Exemple :</u></p> <ul style="list-style-type: none"> - les enfants, exténués, dorment malgré le bruit - le clown, roi du cirque - marquer la juxtaposition <p><u>Exemple :</u></p> <ul style="list-style-type: none"> - il court, il s'affole, il crie mais n'avance à rien - 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>Ces informations sont fortement liées à la prosodie et à la compréhension.</p>	

TR_DET_TRS_2.2_02-a	Ne pas séparer par des espaces les mots liés par un trait d'union ou une apostrophe.
<p>Les mots liés par un trait d'union (mot composé, forme interrogative...) sont écrits normalement, c'est-à-dire sans séparation par une espace, ainsi que les mots reliés par une apostrophe.</p>	



Exemple :



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

		report
speaker#1		
●	c'est aujourd'hui après-midi que vous rencontrerez des Anglo-Saxons.	

TR_DET_TRS_2.2_03-a	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	
1	Paris Lyon, ça se fait en deux heures.
2	la rencontre OM Rennes s'est soldée par un match nul.

3.2.3. Nombres

TR_DET_TRS_2.3_01-a	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.

Dans le cas de nombres prononcés de manière particulière (comme pour les années), il faut introduire une balise de prononciation [pron= ?]

REMARQUE	TR_DET_TRS_2.3_01
----------	-------------------

Ne pas utiliser les chiffres romains dans l'écriture des nombres.

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 au temps de Louis XIV. »
 « En dix-sept cent soixante-dix-huit, c'était le 3^{ème} chef du XVIIIème siècle. »

report	
speaker#1	
1	il y avait 10 mille médecins pour 40 millions d'habitants au temps de Louis 14.
2	en 1718 +[pron= ?], c'était le troisième chef du dix-huitième siècle.

TR_DET_TRS_2.3_02-a	Introduire une balise de prononciation pour toute ponctuation verbalisée.
---------------------	---

- Lorsqu'un locuteur énonce la ponctuation, comme s'il faisait une dictée, on écrit la forme normalisée suivi d'une balise de prononciation +[pron= ?].

Exemple :

« 105 point 5 »
 « 1 virgule 2 »

report	
speaker#1	
1	105.5 +[pron= ?]
2	1,2 +[pron= ?]

TR_DET_TRS_2.3_03-a

Transcrire en toutes lettres les mots prononcés au milieu des chiffres sauf pour la ponctuation.

Liste non exhaustive :

- les unités de mesures : mètre, kilo, litre, euro...
- les signes d'opération : fois, plus, moins, égale...
- les fractions : quart, tiers, moitié...

Exception pour le mot « pourcent » qui est indiqué avec le symbole « % » collé au chiffre qui le précède.



Exemple :



- « Les 40,8L d'essence ont coûté 56,50€. »
- « A -10°, les 2/3 des animaux n'ont pas survécu. »
- « $3 \times 2 / 100 = 6\%$. »

report

speaker#1

- les 40,8 +[pron= ?] litres d'essence ont coûté 56 euros 50.
- à moins 10 degrés, les 2 tiers des animaux n'ont pas survécu.
- 3 fois 2 divisé par 100 égale 6%.

3.2.4. Adresses Internet

TR_DET_TRS_2.4_01-a	Transcrire la forme normalisée puis insérer une balise de prononciation +[pron= ?].
<p>Le « @ » constitue un mot à part entière, il est noté suivant sa prononciation :</p> <ul style="list-style-type: none"> - at - arobase <p>Le « www » et le « http » sont transcrits comme dessigle :</p> <ul style="list-style-type: none"> - WWW - 3W - HTTP <p>La ponctuation (« . », « : », « / », « _ »...) est transcrite sous forme de mots (et non de signes) :</p> <ul style="list-style-type: none"> - point - 2 points - slash - underscore <p>Le « fr », le « com » et le « net » sont transcrits comme suit :</p> <ul style="list-style-type: none"> - FR - com - net 	

Exemple :

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

report	
speaker#1	
http://www.google.fr	+ [pron= ?]
monnom@monadresse.net	+ [pron= ?]
abc_nom@adresse.com	+ [pron= ?]

TR_DET_TRS_2.4_02-a	Ne pas insérer de balise de prononciation et de troncation lorsque le site Internet ou l'adresse mail est donné sur plusieurs échanges.
-	

Exemple :

Client : c'est sdupont
 Agent : oui
 Client : arobase
 Agent : oui
 Client : wanadoo.fr
 Agent : okay

report	
speaker#1	
1	c'est S dupont
speaker#2	
1	oui
speaker#1	
1	arobase
speaker#2	
1	oui
speaker#1	
1	wanadoo point fr
speaker#2	
1	OK

3.2.5. Sigles

TR_DET_TRS_2.5_01-a	Ecrire en majuscule sans espace ou point de séparation, tout sigle lu de 4 lettres ou moins.
-	

TR_DET_TRS_2.5_02-a	Ecrire avec seulement la première lettre en majuscule, tout sigle lu de 5 lettres ou plus.
-	

TR_DET_TRS_2.5_03-a	Ecrire en majuscule sans espace ou point de séparation, tout sigle épilé.
-	

*Exemple :*

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

	report
speaker#1	
IFOP, SMIG, SFOR, PACS, ENA	
Medef, Insee, Staps, Assedic	
FM, ETA, CFDT, EHESS, FNSEA, DGCCRF	

3.2.6. Epellation

TR_DET_TRS_2.6_01-a	Ecrire avec toutes les lettres en majuscule séparées par des espaces, toute épellation.
Toutes les lettres sont notées en majuscule, séparées les unes des autres par une espace.	

REMARQUE	TR_DET_TRS_2.6_01
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. »

	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.	

3.2.7. Abréviations

TR_DET_TRS_2.7_01-a	Utiliser les abréviations standard de la langue française. Il existe dans la langue française des abréviations standard comme « etc » pour « et caetera » ou « OK » pour « okay ».
TR_DET_TRS_2.7_02-a	Transcrire les abréviations non standards de la langue française mais lexicalisées telles qu'elles sont prononcées. -
TR_DET_TRS_2.7_03-a	Transcrire les abréviations non standards de la langue française et non lexicalisées en ajoutant le symbole \$ à la fin. - Le symbole \$ doit être collé à la fin du mot. - En général, il s'agit d'un lexique propre à EDF donc récurrent dans les dialogues Agent-Agent.



Exemple :



« pour l'appart 10B couper l'élec dans 3 jours car demande de résil. »

	report
speaker#1	Pour l'appart 10B couper l'élec dans 3 jours car demande de résil\$

3.2.8. Onomatopées et interjections

TR_DET_TRS_2.8_01-a	Transcrire les onomatopées et les interjections en les précédant du signe &. Choisir les onomatopées et les interjections parmi la liste suivante : &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	TR_DET_TRS_2.8_01
Si une nouvelle onomatopée ou une nouvelle interjection apparaît dans le document audio, la transcrire le plus simplement possible, et accoler un & avant la première lettre.	



Exemple :



« pam pam pam je sais pas. »

	report
speaker#1	&pam &pam &pam je sais pas.

3.2.9. Hésitations

TR_DET_TRS_2.9_01-a	Transcrire les hésitations grâce à « euh » et « hum ». Ne pas confondre « hum » et « &hm » : - hum est utilisé pour les hésitations
----------------------------	---

Référence :	Indice :	Date d'édition :	Page :
	001	17/12/2007	18 / 25

- &hm est utilisé lors d'acquiescements

3.2.10. Répétitions

TR_DET_TRS_2.10_01-a Transcrire tous les mots entendus même s'ils sont répétés.

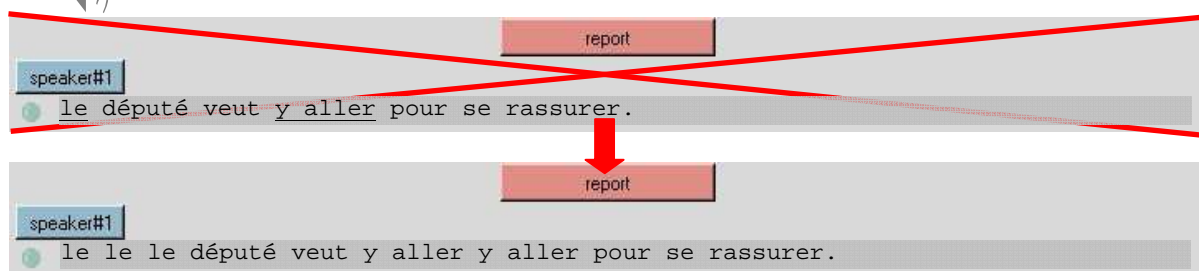
-



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.

3.2.11. Troncation

TR_DET_TRS_2.11_01-a	Marquer la troncation d'un mot par des parenthèses en début ou fin de mot.
Troncations finales : vou() pour voulais ou voudrais...	
Troncations initiales : ()embre pour septembre ou novembre...	
Un mot complet n'apparaît jamais entre parenthèses .	
Pas de troncation au milieu d'un groupe de lettres correspondant au même son.	

REMARQUE	TR_DET_TRS_2.11_01
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... »

Diagram illustrating the correction of a truncated transcription. The top part shows a transcription with a red 'X' over it, indicating it is incorrect. The transcription is: `ce mag(nifique) p(e)tit plat(e)au (n')est pas à range(r) dans la car().` A red arrow points down to the corrected transcription below.

The corrected transcription is: `ce magn(ifique) petit plateau est pas à ranger dans la car().`

3.2.12. Prononciation erronée

TR_DET_TRS_2.12_01-a	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	TR_DET_TRS_2.12_01
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. »

Diagram illustrating the correction of a mispronounced transcription. The top part shows a transcription with a red 'X' over it, indicating it is incorrect. The transcription is: `le *spect(acle) euh spectacle *artistique est présenté par *Martin Dupond.` A red arrow points down to the corrected transcription below.

The corrected transcription is: `le *spect(acle) euh spectacle *artistique est présenté par *Martin Dupond.`

3.2.13. Lapsus

TR_DET_TRS_2.13_01-a	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.

3.2.14. Elisions et formes réduites

TR_DET_TRS_2.14_01-a	Respecter l'orthographe standard même si les mots font l'objet à l'oral d'élision ou de forme réduite.
----------------------	--

La réduction peut porter sur un seul mot ou sur une suite de mots.
--

REMARQUE

TR_DET_TRS_2.14_01

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

3.2.15. Mots inconnus ou absents du lexique

TR_DET_TRS_2.15_01-a	Transcrire le mot inconnu avec l'orthographe la plus proche de ce qui est entendu, et précéder le mot de deux accents circonflexes.
<p>Un mot inconnu est un mot absent du lexique.</p> <p>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.</p> <p>Cette règle s'applique en particulier pour l'écriture des noms propres.</p>	

REMARQUE	TR_DET_TRS_2.15_01
L'orthographe des mots inconnus doit rester homogène dans tout le document audio transcrit.	



Exemple :



« Dacim Houssila et Francine Mokokoh ont vu un ornithorynque. »
 « T'es ouf ou quoi, mon appart est vide ! »

		report
speaker#1		
•	^^Dassim ^^Oussila et Francine ^^Mokoko ont vu un ^^ornitorinque.	
•	tu es ^^ouf ou quoi, mon ^^appart est vide.	

TR_DET_TRS_2.15_02-a	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].</p> <p>Si la langue étrangère est indéterminée, remplacer xxx par « inconnue ».</p> <p>Les mots étrangers dont l'usage est courant en français ne sont pas concernés par cette règle :</p> <p><i>Exemple :</i> parking, week-end, planning, manager, sandwich...</p>	

4. REGLES DE BALISAGE

4.1. CONDITIONS ACOUSTIQUES

TR_DET_BAL_1_01-a	Indiquer les conditions acoustiques de chaque appel.
<ul style="list-style-type: none"> - music pour la musique de fond - speech pour l'ambiance centre d'appel - shh pour une dégradation de la communication - other pour toute autre condition acoustique. 	
Il est possible de combiner plusieurs conditions acoustiques.	

4.2. RESPIRATION

TR_DET_BAL_2_01-a	Insérer une balise [r] pour indiquer une respiration, une inspiration ou une expiration du locuteur.
-	

4.3. RIRES

TR_DET_BAL_3_01-a	Insérer une balise [rire] pour indiquer des rires du locuteur.
-	

4.4. BRUITS

TR_DET_BAL_4_01-a	Insérer une balise [b] pour indiquer un bruit.
Les bruits de l'appareil phonatoire : <ul style="list-style-type: none"> - bruits de bouche (toux, bruit de gorge, sifflement du locuteur, chuchotement...) Les bruits extérieurs : <ul style="list-style-type: none"> - indéterminés - froissement de papiers - bruits de micro 	

REMARQUE	TR_DET_BAL_4_01
<ul style="list-style-type: none"> - La balise peut être instantanée pour un bruit isolé dans le signal ou encadrante pour un bruit court qui se superpose à la parole. 	

4.5. IMITATION

TR_DET_BAL_5_01-a	Insérer une balise [pron=imitation] pour indiquer une imitation.
-	

4.6. PAROLE EN BRUIT DE FOND

TR_DET_BAL_6_01-a	Indiquer la conversation en fond par la balise encadrante [conv].
<ul style="list-style-type: none"> - Les parties du signal encadrées par la balise [conv] doivent être isolées dans un segment. 	

4.7. PROBLEME DE PRONONCIATION

TR_DET_BAL_7_01-a	Insérer une balise [pron=pi] pour toute parole inintelligible ou inaudible.
-	Cette balise peut être insérée dans une balise de parole superposée.

TR_DET_BAL_7_02-a	Associer la balise [pron=pi] à un locuteur et l'isoler dans un segment.
-	



Exemple :



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

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

TR_DET_BAL_7_03-a	Insérer une balise +[pron= ?] pour écrire la forme prononcée du mot précédent écrit de façon normalisée.
-	

4.8. REGLES DE SEGMENTATION LIEES AUX BALISES

TR_DET_BAL_8_01-a	Au sein de l'intervention d'un locuteur, les balises [r] et [b], d'une durée supérieure à 500 ms sont isolées dans un segment.
-	

Exemples :

(pas de locuteur)	
1	
speaker#1	speaker#1
1	
[r]	
oui oui bien sûr	oui oui bien sûr
	[r]
	c'est difficile

TR_DET_BAL_8_03-a	Une balise encadrante ne peut débuter dans un segment et finir dans un autre. Encadrer chaque segment si cela est nécessaire.
-	

4.9. COMMENTAIRES

TR_DET_BAL_4.9_01-a	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- ...	