Évaluation des Systèmes de Transcription enrichie d'Émissions Radiophoniques (ESTER)

Plan d'Évaluation

phase 2

Version 1.1 du 7 janvier 2005

Table des matières

1 Préambule

Ce document décrit le *plan d'évaluation de la phase 2* de la campagne d'Évaluation des Systèmes de Transcription enrichie d'Émissions Radiophoniques (Ester).

Ce document présente tout d'abord très brièvement les objectifs scientifiques de la campagne. Il décrit ensuite l'ensemble des tâches qui seront évaluées dans cette deuxième phase, puis la description des données de développement et de test, ainsi que les règles régissant la réalisation des différentes tâches et le calendrier de l'évaluation. Les formats de soumission des résultats sont décrits en annexe.

2 Objectifs

La campagne Ester, organisée dans le cadre du projet Evalda du programme Technolangue, a pour buts principaux de promouvoir une dynamique de l'évaluation en France, autour du traitement de la parole de langue française, de mettre en place une structure pérenne d'évaluation et de diffuser le plus largement possible les informations et les ressources concernées par ces évaluations. L'axe prioritaire sera d'assurer un accès aux évaluations à un nombre aussi large que possible de participants.

Sur le plan scientifique, les résultats attendus sont bien évidemment de mesurer objectivement et de faire progresser les performances des systèmes de transcriptions enrichies en français, et d'encourager la fédération des efforts de recherche dans ce domaine.

L'objectif est également d'améliorer la visibilité du secteur de recherche concerné, par la mise en évidence du niveau de performance atteint par l'état de l'art, par la constitution d'un « club » d'acteurs identifiés et pouvant prouver leur niveau de compétence, et par la publicité assurée au projet.

3 Tâches de l'évaluation

La phase 2 de l'évaluation reprend les deux thèmes principaux de la phase 1, à savoir la transcription orthographique (T) et la segmentation (S), et y ajoute la tâche de détection des entités nommées du thème extraction d'information (E)[1].

Le tableau 1 résume les différentes tâches dans les trois thèmes. Les participants qui souhaitent ne s'engager que sur le thème de la segmentation doivent fournir des résultats pour au moins une des tâches orientées locuteur (SRL ou SVL).

[1] Une phase 3 est à l'étude pour les autres tâches du thème extraction d'information.

Tab. 1 – Récapitulatif des thèmes et taches.

thème	tâche	description
T	TRS	transcription orthographique
T	TTR	transcription temps réel
S	SES	suivi d'événements sonores
S	SRL	segmentation et regroupement de locuteurs
S	SVL	suivi de locuteurs
E	EN	détection d'entités nommées

3.1 Transcription

Ce thème consiste à évaluer la transcription orthographique en sortie des systèmes de reconnaissance automatique de la parole, en terme de taux d'erreur de mots. Deux tâches sont définies pour la transcription. La tâche de transcription orthographique sans contrainte de temps de traitement (TRS), et la tâche de transcription orthographique en temps réel (TTR).

Les participants s'engageant sur ce thème doivent obligatoirement participer à la tâche "transcription orthographique" (TRS) et sont encouragés à soumettre un maximum de résultats contrastifs. Un même système peut bien évidemment participer dans plusieurs catégories.

3.1.1 Transcription orthographique (TRS)

Cette tâche consiste à produire une transcription orthographique à partir du signal de parole, sans contrainte de temps de traitement autre que le délai global de la période de test.

La normalisation (adaptation au locuteur, normalisation de scores, etc) ne peut se faire que sur la base d'un document. Un contraste utilisant une normalisation sur plusieurs documents est toujours possible mais ne pourra être présenté comme système principal.

Les participants doivent fournir une description du ou des système(s) utilisés en spécifiant clairement les ressources linguistiques utilisées, les algorithmes et méthodes mis en œuvre ainsi que le temps de traitement et la taille mémoire pour le décodage.

3.1.2 Transcription temps réel (TTR)

Cette tâche est similaire à celle de la tâche TRS, mais vise des systèmes pour lequel le temps de calcul pour le système complet est limité à une fois le temps réel sur un mono-processeur standard.

Le temps de calcul comprend l'ensemble des opérations (segmentation, E/S, décodage, etc.) et se mesure entre l'instant où le traitement est lancé et l'instant où il finit (commande `date` sous Unix). C'est le temps perçu par un utilisateur

du système, et non le seul temps CPU (tel que le mesurerait la commande `time` sous Unix). Le traitement doit être lancé à froid (pas de lancement partiel pour mettre des informations en cache). Le ratio temps réel est le résultat de la division du temps de traitement par la durée du fichier, noté XxTR.

Se qualifie pour la tâche TTR un système capable de traiter le corpus de développement en moins de 1xTR en moyenne sur l'ensemble des fichiers. Le temps de traitement effectif sur les données de test pourra éventuellement dépasser 1xTR. On rapportera les temps de traitement non seulement pour le test mais aussi pour le corpus de développement. On publiera les caractéristiques de la machine (Marque/type, CPU, bus, mémoire vive, disque dur).

Un programme de calibrage de la puissance de la machine utilisée sera fourni afin de faciliter les comparaisons.

Pour les site le désirant, une machine commune sera mise à disposition par la DGA (au minimum 2GHz de vitesse d'horloge et 1Go de mémoire vive).

3.1.3 Mesure des performances

La mesure des performances pour les tâches de transcription est le taux d'erreur après alignement entre la sortie du système et la transcription manuelle au format STM (Segment Time-Mark), après normalisation des textes.

Ce taux d'erreur est mesuré par le script `score-trs.v`*x.x* du package d'évaluation qui met en œuvre l'outil `sclite` du NIST, et les règles de normalisation sont décrites dans l'annexe C.

3.2 Segmentation

Ce thème vise à évaluer les systèmes de segmentation du flux sonore et de suivi d'événements dans le flux. Les tâches considérées sont

- le suivi d'événements sonores (parole/musique)
- le suivi de locuteur
- la détection des tours de parole et regroupement en locuteur

Les participants s'engageant sur ce thème sans participer au thème transcription doivent fournir des résultats dans au moins une des tâches orientées locuteur (SRL ou SVL) et sont invités à participer à l'ensemble des tâches. Les participants s'engageant sur le thème de la transcription sont invités à soumettre des résultats pour la tâche SES afin d'évaluer la qualité du module de détection de parole dans le système de transcription.

Les participants doivent fournir une description du ou des système(s) utilisés en spécifiant clairement les ressources utilisées, les algorithmes et méthodes mis en œuvre ainsi que le temps de traitement et la taille mémoire nécessaires.

3.2.1 Suivi d'événements sonores (SES)

Cette tâche consiste à détecter les plages du flux audio pour lesquelles un événement sonore particulier est présent. Dans le cadre de l'évaluation, les deux événements étudiés seront la présence de parole et la présence de musique.

Pour chaque événement, le système doit déterminer les plages contenant cet événement sur l'ensemble des documents du corpus de test.

Nous soulignons que cette tâche de suivi d'événement est légèrement différente de la tâche de classification classique qui consiste à segmenter en classes génériques silence, parole, musique, parole et musique. Dans cette évaluation, il s'agit d'une détection parole/non-parole puis musique/non-musique.

3.2.2 Segmentation et regroupement de locuteurs (SRL)

Cette tâche a pour but d'évaluer les algorithmes permettant de découper le flux audio en tours de parole et de regrouper les plages associées à un même locuteur. L'identification des locuteurs concernés n'est pas requise. Le système retourne une segmentation du document spécifiant les plages de silence et un identifiant arbitraire de locuteur pour chaque plage contenant de la parole. Pour chaque document du corpus de test, la segmentation et le regroupement devront être établis sur la seule base de ce document.

L'utilisation de connaissances a priori, comme l'utilisation de modèles de locuteurs connus a priori, est autorisée. Cependant, l'objectif principal de cette tâche du point de vue applicatif étant la détection des tours de parole, les participants utilisant des connaissances a priori sont invités à soumettre un contraste où aucune connaissance a priori n'est utilisée.

3.2.3 Suivi de locuteurs (SVL)

Cette tâche vise à évaluer les systèmes de suivi de locuteur qui permettent l'enrichissement de la transcription ou de la description du document. L'objectif est de détecter les plages correspond à un locuteur donné, connu à l'avance.

La liste des locuteurs à suivre est constituée de tous les locuteurs ayant parlé au moins deux minutes dans les 82h du corpus d'apprentissage et identifiables à leur nom [2]. Pour chaque locuteur de la liste, le système doit identifier les plages correspondant à ce locuteur dans l'ensemble des données de test.

L'utilisation de données autres que le corpus d'apprentissage, notamment pour le(s) modèle(s) de normalisation (modèle(s) du monde), est autorisée. Comme pour la tâche de transcription, les sites utilisant d'autres données sont invités à soumettre un contraste en n'utilisant que les données du corpus d'apprentissage.

Si possible, à des fins de contrastes, un (ou plusieurs) segment(s) d'une minute sera identifié pour chacun des locuteurs de la liste des locuteurs à suivre. Les participants seront alors invités à présenter un contraste où les données d'apprentissage pour les locuteurs cibles sont limitées aux segments identifiés.

[2] fichier locuteurs-train.tgz disponible sur le site ESTER.

3.2.4 Mesure des performances

Pour les tâches de suivi d'événements (SES et SVL), les performances seront mesurées sur la base du rappel et de la précision. Le rappel est défini par

$$R = \frac{\sum_i t(c_i; c_i)}{\sum_i t(c_i; c_i) + t(\overline{c}_i; c_i)}$$

tandis que la précision est donnée par

$$P = \frac{\sum_i t(c_i; c_i)}{\sum_i t(c_i; c_i) + t(c_i; \overline{c}_i)}$$

où $t(c_i; c_i)$ correspond au temps où l'événement i a été détecté correctement, $t(c_i; \overline{c}_i)$ au temps où l'événement i a été détecté à tord (fausse acceptation) et $t(\overline{c}_i; c_i)$ au temps où l'événement i n'a pas été détecté à tort (faux rejet). Les événements considérés sont, pour la tâche SES, parole et musique, et, pour la tâche SVL, l'ensemble des locuteurs à suivre. Les performances des systèmes seront comparées sur la base de la F-mesure définie par

$$F = \frac{2RP}{R+P} \ .$$

Les temps seront mesurés en secondes. Les événements i sont parole et musique pour la tâche SES et les locuteurs cibles pour la tâche SVL.

Pour la tâche de segmentation et regroupement de locuteur (SRL), la métrique est le taux d'erreur défini comme la somme des taux de parole non détectée, de fausse détection de parole et mauvaise détection. Le taux de parole non détectée correspond aux portions de parole détectée comme silence. Inversement, le taux de fausse détection de parole correspond aux portions de silence pour lesquelles un locuteur a été détecté. Le taux de mauvaise détection correspond aux erreurs sur les identités (arbitraires) des locuteurs. Une correspondance entre noms de locuteurs et noms arbitraires fournies par le système est établie par appariement. Ces taux seront calculées sur l'ensemble des documents.

Pour l'ensemble des tâches de segmentation, les références seront établies à partir des transcriptions manuelles et d'une détection automatique des plages de silence. L'usage d'un détecteur automatique de silence permet de marquer les plages de silence d'une durée supérieure à une seconde qui n'ont pas été marquées comme telles lors de la transcription manuelle. La détection des silences se fera par l'outil ssad[3].

3.3 Extraction d'information

Les tâches d'extraction d'information sont considérées comme expérimentales dans ESTER.

[3] http ://www.afcp-parole.org/ester/private/audioseg-1.1.tar.gz.

3.3.1 Détection d'entités nommées (EN)

Cette tâche consiste à détecter dans les transcriptions orthographiques automatiques et manuelles les mentions d'entités nommées de plusieurs types (personnes, organisations, etc.). Les données de référence ont été annotées selon le document de conventions diffusé aux participants.

Les transcriptions à traiter sont les d'une part la transcription manuelle, et d'autre part toutes les transcriptions automatiques que les participants au thème transcription accepteront de fournir (plusieurs ont confirmé leur intention de le faire).

3.3.2 Mesure des performances

L'outil de mesure de performance est en cours de développement et sera diffusé au plus vite.

4 Ressources autorisées

L'ensemble des ressources de la phase 2 de la campagne ESTER sont mises à la disposition des participants par la DGA (ressources acoustiques) et par ELDA (ressources acoustiques et textuelles). Avant d'avoir accès aux données, les participants doivent signer un contrat d'engagement avec ELDA[4]. Ce contrat stipule en particulier que les participants soumettant des résultats lors de la campagne de test pourront conserver les données gratuitement à des fins de recherche à l'issue de la campagne.

Toute ressource antérieure au 01/05/2004, distribuée dans le cadre de la campagne ou pas, ainsi que les données non transcrites distribuées dans le cadre de la campagne peuvent être utilisées dans la phase de développement. Les ressources postérieures à cette date, à l'exception des données non transcrites distribuées, ne sont pas autorisées.

Les systèmes utilisant des ressources autres que celles fournies pour la campagne devront identifier ces ressources et fournir, dans la mesure du possible, des résultats contrastifs illustrant l'apport de ces ressources. Dans un souci de comparaison scientifique entre les différents constituants d'un système, les participants à cette tâche sont invités à faire une évaluation en n'utilisant que les ressources distribuées pour la campagne. Un classement des systèmes respectant ces conditions sera effectué en plus de l'évaluation officielle.

4.1 Ressources acoustiques

Les ressources acoustiques sont constituées d'émissions radiophoniques transcrites manuellement. Les émissions enregistrées sont des émissions d'information comportant le journal ainsi que des dossiers liés à l'actualité du moment. Pour

[4] http ://www.afcp-parole.org/ester/download/contrat_utilisateur_6.pdf.

TAB. 2 – Resources acoustiques

source	développement		test
	transcrit	non transcrit	
France Inter	33/2	337	2
France Info	8/2	643	2
RFI	23/2	445	2
RTM	18/2	–	2
France Culture	–	252	1
"surprise"	–	–	1
total	82/8	1677	10

la phase 2, les données d'apprentissage proviennent de quatre sources : France-Inter, France-Info, Radio France International (RFI) et Radio Télévision Marocaine (RTM). Les données de développement ont été enregistrée sur trois périodes : 1998, 2000 et 2003.

Les données de tests comprendront 2h de chacune de ces sources, plus 1h de France Culture et 1h d'une source inconnue. Les données de test correspondent à une période s'étalant du 01/10/2004 au 31/12/2004.

Un ensemble de données non transcrites est également fourni aux participants qui en font spécifiquement la demande. Ce corpus a pour objectif l'étude de l'utilisation de grand volume de données non transcrites pour améliorer les performances dans les tâches de transcription. Les données non transcrites proviendront des sources France-Inter, France-Info, RFI et France Culture. Les données non transcrites ont été enregistrées entre le dernier trimestre 2003 et le septembre 2004.

Le tableau 2 récapitule la répartition de l'ensemble des données. Une liste de fichiers désignés comme "corpus de développement" est donnée en annexe D.

4.2 Ressources textuelles

En plus des transcriptions manuelles des données acoustiques, les ressources textuelles pour la phase 2 de la campagne correspondent aux années 1987 à 2003 du journal "Le Monde" augmentés du corpus MLCC contenant des transcriptions des débats du Conseil Européen.

4.3 Ressources lexicales

Aucune ressource lexicale n'est spécifiée. Il existe cependant des phonétiseurs libres qui permettent la phonétisation du corpus d'apprentissage et du lexique[5]. En particulier, le Laboratoire Informatique d'Avignon met à la disposition des participants son phonétiseur, accessible depuis le site ESTER.

[5] Cf. `http ://tcts.fpms.ac.be/synthesis/mbrola.html`

5 Règles de participation

Pour l'ensemble des tâches, les règles suivantes s'appliquent :

- l'origine du document (c-à-d la chaîne de radio correspondant à un enregistrement) ainsi que la tranche horaire de l'enregistrement est une information qui peut être utilisée ; cependant, les données de test pourront provenir d'une tranche horaire pour laquelle aucune donnée d'apprentissage n'est disponible.
- les données utilisées doivent respecter les contraintes décrites dans la section 4 du présent document.
- Pour chaque tâche, les participants soumettant plusieurs systèmes devront identifier un système principal qui servira pour établir le classement officiel des participants. Les autres soumissions seront considérées à titre de contrastes.
- les résultats retournés après la date de clôture du test (cf. calendrier) ne seront pas considérés dans la classification des systèmes.

Par ailleurs, quelques règles essentielles sont rappelées ici :

- Les données audio ne peuvent être examinées ou écoutées avant ou pendant le test.
- Les systèmes évalués ne peuvent être modifiés une fois le traitement commencé. Un système ne peut être testé qu'une seule fois.
- Le traitement des données doit être entièrement automatique. Le résultat de ce traitement ne peut en aucun cas être modifié. Les seules interventions manuelles autorisées sont limitées aux opération de lancement des traitements, aux vérifications de bon fonctionnement et aux opérations de relance éventuelles en cas de problème informatique.
- Si plusieurs systèmes sont évalués pour une tâche, aucun résultat ne peut être examiné avant la fin du dernier traitement à soumettre. Un système et un seul doit être identifié comme système primaire.

6 Calendrier

10/01/05 distribution des données de test
31/01/05 date limite de soumission des tâches T et S
01/02/05 distribution des données pour la tâche EN
(transcription de reférence sans EN et sorties de systèmes)
10/02/05 date limite de soumission de la tâche EN
11/02/05 distribution de l'ensemble des transcriptions

Les dates limites de soumission s'entendent à minuit (heure de France métropolitaine).

Un atelier de présentation des résultats et des systèmes aura lieu les 30-31 mars 2005.

7 Format des soumissions

Le format des soumissions sera similaire à celui de la phase 1. Un document compémentaire sera fourni.

Les soumissions se font par envoi d'un mail[6]. Un accusé de réception du message sera retourné sous 48h.

8 Contacts

Pour plus de renseignements, contactez l'un des organisateurs :

Guillaume Gravier (AFCP), ggravier@irisa.fr, 02 99 84 72 39
Edouard Geoffrois (DGA), Edouard.Geoffrois@etca.fr, 01 42 31 96 68
Sylvain Galliano (DGA), Sylvain.Galliano@etca.fr, 01 42 31 97 59
Djamel Mostefa (ELDA), mostefa@elda.fr, 01 43 13 33 33

A Format et métrique pour la transcription

Le format de soumission des résultats pour les tâches de transcription est le format CTM (Conversation Time-Mark). Chaque ligne de la soumission correspond à un mot avec une spécification de temps et un identifiant de fichier, suivant la syntaxe

```
source A début durée mot confiance
```

où `source` correspond au nom du fichier (sans extension, sans chemin), `début` au temps de début du mot en secondes par rapport au début du fichier, `durée` à la durée du mot en secondes et `confiance` à une mesure de confiance normalisée dans l'intervalle [0,1]. Un seul fichier CTM (encodage iso-8859-1) contenant les mots pour l'ensemble des fichiers du corpus de test sera retourné par système. De plus, ce fichier doit être trié par ordre croissant selon les trois premières colonnes : les deux premières par ordre alphabétique, la troisième par ordre numérique. La commande Unix sort +0 -1 +1 -2 +2nb -3 permet d'effectuer ce tri. Pour plus de détails, voir la documentation du logiciel sctk 1.2c[7].

Les soumissions seront évaluées à l'aide du script `score-trs.v`*x.x* fourni dans le package d'évaluation. Celui-ci effectue des normalisations pour ne pas compter comme erreur des variantes orthographiques autorisées. Le dictionnaire de normalisation sera augmenté de nouvelles équivalences par les organisateurs pour prendre en compte les données de test.

L'alignement des soumissions à la transcription de référence se fait en deux temps : un premier alignement temporel permet d'affecter les mots (CTM) aux segments de la transcription de référence (au format STM), sur la base des temps des instants d'occurrences des mots. Dans un deuxième temps, un algorithme

[6] L'adresse mail pour la soumission est `ester-soumission@etca.fr`.

[7] http ://www.nist.gov/speech/tools.

d'alignement dynamique est utilisé indépendamment pour chaque segment de la référence.

Certains phénomènes nécessitent un traitement particulier et sont optionnels dans la transcription. Dans ce cas, aucune erreur n'est comptée si le mot est absent de la transcription. Cependant, les mots optionnels sont pris en compte dans le calcul du nombre total de mots dans la transcription de référence. Les mots optionnels correspondent aux phénomènes suivants :

- mots partiellement prononcés : ces mots sont indiqués dans la référence en mettant entre parenthèse la partie manquante du mot. Un mot reconnu à la place d'un mot partiel sera considéré comme correct si la partie transcrite du mot correspond au début du mot inséré.
- mots d'origine étrangère autre que noms propres et noms couramment utilisés en français (par exemple, sandwich)

Les segments de parole vérifiant les conditions suivantes (dans la transcription de référence) sont ignorés dans la mesure des performances :

- segment contenant de la parole superposée
- segment contenant plus d'un mot prononcé dans une langue autre que le français, sans compter les noms propres, les acronymes et les mots couramment utilisés en français
- segment correspondant à de la publicité (non transcrit dans la référence)
- segment contenant plus de deux mots d'origine étrangère autre noms propres et noms couramment utilisés en français

B Format et métriques pour la segmentation

B.1 Suivi d'événements

Pour les tâches de détection d'événements (SES et SVL), les résultats seront retournés au format ETF (Event Tracking File). Chaque ligne de la soumission correspond à un segment et un événement, indiqués selon le format suivant

```
source A début durée type sous-type événement score décision
```

où la signification des champs est

- source : nom du fichier sans extension ni chemin.
- début : temps de début du segment, en secondes par rapport au début du fichier
- durée : durée du segment, en secondes
- type : type d'événement (spk pour la tâche SVL, sc pour la tâche SES). Ce champ n'est pas utilisé pour la mesure des performances.
- sous-type : pour la tâche SVL, le sous-type peut-être male ou female ou encore unknown. Ce champ n'est pas utilisé pour la mesure des performances.
- événement : pour la tâche SVL, les événements correspondent aux noms des locuteurs ; pour la tâche SES, les événements sont music et speech.
- score : score associé à la décision ; plus le score est élevé, plus la décision est sûre. Ce champ n'est pas directement utilisé pour l'évaluation des

performance mais permettra d'établir des courbes DET pour une meilleure comparaison des systèmes sur l'ensemble des points de fonctionnement. Ce champ peut être remplacé par un tiret (-) si aucun score n'est disponible.
- décision : décision de présence (true) ou absence (false) de l'événement recherché. Si ce champ n'est pas renseigné, l'événement est réputé présent.

Les lignes débutant par un point virgule seront traitées comme des lignes de commentaires.

Pour une tâche (SES ou SVL), les résultats d'un système seront soumis sous la forme d'un unique fichier ETF (encodage iso-8859-1). Le fichier de soumission spécifie l'ensemble des segments qui contiennent les événements considérés.

B.2 Segmentation et regroupement de locuteurs

Pour la tâche SRL, les résultats seront retournées au format MDTM (Meta Data Time-Mark). Chaque ligne du fichier identifie un segment selon le format suivant

```
source A début durée type confiance sous-type id
```

où la signification des champs est
- source : nom du fichier sans extension ni chemin.
- début : temps de début du segment, en secondes par rapport au début du fichier
- durée : durée du segment, en secondes
- type : type d'événement ('speaker')
- confiance : mesure de confiance associée à la décision, dans l'intervalle [0,1]. Ce champ est optionnel et peut-prendre la valeur NA lorsqu'il n'est pas spécifié.
- sous-type : sous-catégorie parmi 'adult_male', 'adult_female', 'child' ou 'unknown' (champ non utilisé)
- id : identifiant arbitraire de locuteur (par exemple, loc1, loc2, etc.)

Les lignes débutant par un point virgule seront traitées comme des lignes de commentaires.

Un seul fichier MDTM (encodage iso-8859-1) par système sera soumis. Pour des raisons pratiques et afin de favoriser les systèmes produisant des segmentations réalistes, le nombre total de segment par fichier considéré sera limité à 5 000 par soumission. Si une soumission contient plus de 5 000 segments, seuls les 5 000 premiers seront considérés.

Le taux d'erreur de classification est établi en cherchant le meilleur appariement entre les locuteurs de la segmentation de référence et les identifiants arbitraires de la soumission. Le taux d'erreur est ensuite calculé à partir de cet appariement par comptage du temps total de segments (in)correctement classifiés.

B.3 Segmentation de référence

Les règles suivantes seront appliqués pour toutes les tâches de segmentation (SES, SRL et SVL).

Les segmentations de référence seront établies à partir des transcriptions manuelles au format Transcriber. De plus, une détection automatique des zones de silence d'une durée supérieure à une seconde sera utilisée afin de corriger les segmentations de référence. L'objectif de la détection automatique des silences est de d'étiqueter correctement les pauses longues marqués comme parole dans la transcription de référence. Le détecteur utilisé est la commande ssad de audioseg 1.1 avec un seuil de 1 seconde, disponible sur le site de la campagne. De plus, le résultat de la détection de silence sera fourni aux participants au format UEM.

Les segments non transcrits dans la référence (publicité) seront éliminés pour la mesure des performances. Une tolérance de 0.25 secondes sera appliqués aux frontières des segments afin de ne pas pénaliser un léger décalage des frontières.

Des scripts permettant de générer les segmentations de référence aux formats ETF et MDTM à partir des transcriptions de référence au format transcriber et de mesurer les performances seront mis à la disposition des participants sur le site de la campagne.

C Règles de normalisation des transcriptions

Les règles de normalisations sont décrites en détail dans la documentation du package de scoring. Elles sont rappelées de manière synthétique ici :

- la casse n'est pas prise en compte (tout les mots sont en minuscule)
- la ponctuation est supprimée
- un espace est inséré après les apostrophes liées à des élisions (l'importance, quoiqu'il, losrqu'on, jusqu'à, etc), l'apostrophe étant maintenue dans le constituant de gauche. Les prefixes pouvant donner lieu à élision sont données dans le script perl ci-dessous.
- les mots composés (séparation par un tiret) sont divisés en leur constituants, le tiret étant supprimé.
- les expressions numériques sont récrites sous forme littérale
- les sigles sont laissés dans leur forme compacte (séquence de sans espace ni point). Cependant, pour les sigles contenant des chiffres (comme les noms de routes et d'autoroutes), l'équivalence dans laquelle la partie numérique est développée sous forme littérale est produite.
- les mots d'hésitations (euh, hum, huhum, mm) sont remplacés par le symbole %hésitation (aucune erreur n'est comptée pour une hésitation non reconnue)
- la partie non prononcée des mots partiellement prononcés est supprimée (aucune erreur n'est comptée lorsqu'un mot partiellement prononcé n'est pas reconnu ; un mot dont le début correspond orthographiquement à la partie prononcée du mot partiellement prononcé sera compté comme correct à l'alignement)

– les mots mal prononcés (mais pas tronqués) sont laissés tels quels (forme du dictionnaire).
– les mots marqués à orthographe incertaine, qui ne représentent qu'une proportion très faible des données, ne reçoivent pas de traitement particulier même s'ils sont des candidats privilégiés à des entrées dans le dictionnaire déquivalence.
– un dictionnaire d'équivalence (par exemple, événement et évènement, clé et clef ou encore Hong Kong et Hongkong) sera fourni aux participants peu avant la campagne ; toute forme graphique apparaissant dans le dictionnaire est récrite en une forme unifiée.

L'orthographe des mots dans la transcription de référence a été validée par *aspell* et sert d'orthographe de référence pour les mots concernés dans les transcriptions. Les variantes othographiques ou grammaticales attestées dans les dictionaires Larousse et Robert, ou dans le Grévisse, sont acceptées. Les variantes fréquemment rencontrées sur internet peuvent également être acceptées.

Les participants peuvent proposer des mises à jour du dictionnaire jusqu'au démarrage de l'évaluation. Après soumission des résultats, les organisateurs produiront une proposition de mise à jour du dictionnaire pour tenir compte des données de test et des soumissions. Les participants auront 48h pour réagir à cette proposition. Les modifications n'ayant pas fait l'objet d'opposition seront intégrées pour la mesure des résultats officiels.

D Liste des fichiers de développement

```
20030418_0700_0800_FRANCEINTER_DGA
20030418_0800_0900_FRANCEINTER_DGA
20030418_1200_1300_FRANCEINFO_DGA
20030418_1700_1800_FRANCEINFO_DGA
20030508_1400_1500_RFI_ELDA
20030509_1400_1500_RFI_ELDA
20030717_0700_0715_RTM_ELDA
20030717_1300_1320_RTM_ELDA
20030717_2000_2020_RTM_ELDA
20030717_2300_2315_RTM_ELDA
20030719_0700_0715_RTM_ELDA
20030719_1300_1320_RTM_ELDA
20030719_2000_2015_RTM_ELDA
20030719_2300_2310_RTM_ELDA
```