

ROCme! : logiciel pour l'enregistrement et la gestion de corpus oraux

Emmanuel Ferragne¹ Sébastien Flavier² Christian Fressard²

(1) CLILLAC-ARP, Université Paris 7

(2) Dynamique Du Langage, UMR 5596, CNRS-Université de Lyon

rocme@ish-lyon.cnrs.fr

RÉSUMÉ

ROCme! permet une gestion rationalisée, autonome et dématérialisée de l'enregistrement de corpus oraux. Il dispose notamment d'une interface pour le recueil de métadonnées sur les locuteurs totalement paramétrable via des balises XML. Les locuteurs peuvent gérer les réponses au questionnaire, l'enregistrement audio, la lecture, la sauvegarde et le défilement des phrases (ou autres types de corpus) en toute autonomie. ROCme! affiche du texte, avec ou sans mise en forme HTML, des images, du son et des vidéos.

ABSTRACT

ROCme!: software for the recording and management of oral corpora

ROCme! has been designed to allow a sensible, autonomous, and dematerialized management of speech recordings. Users can create interfaces for metadata collection thanks to XML tags. Speakers autonomously fill in questionnaires, record, play, and save audio; and browse sentences (or other types of corpora). ROCme! can display text, optionally with HTML formatting, images, sounds, and video.

MOTS-CLÉS : corpus, oral, linguistique, logiciel

KEYWORDS : corpus, oral, linguistics, software

1 Résumé

Le développement du logiciel ROCme!¹ – Recording of Oral Corpora Made Easy – intervient en réponse à un besoin de la communauté des linguistes. Il s'agit de : i) permettre à un locuteur d'enregistrer de façon autonome des phrases qui s'affichent à l'écran, ii) recueillir des métadonnées par le biais d'un questionnaire dématérialisé, iii) assurer la cohérence de la procédure à travers un projet d'enregistrement.

ROCme! s'appuie sur le moteur d'exécution Adobe AIR pour offrir une interface graphique conviviale compatible Windows et Mac. Les locuteurs gèrent l'enregistrement de leurs productions orales via les boutons de l'interface ou les touches du clavier. Les phrases à afficher sont incluses dans un fichier texte UTF-8.

Afin de garantir une collecte cohérente des données orales, il convient de définir un ensemble de contraintes favorisant l'homogénéité d'une série d'enregistrements ; c'est le but de la notion de projet. Ainsi, à la création d'un projet, on pourra choisir de présenter les stimuli en ordre aléatoire, et de faire apparaître un masque entre le déclenchement de l'enregistrement et l'apparition d'un stimulus (pour éviter que le début ne soit coupé).

¹ www.ddl.ish-lyon.cnrs.fr/rocme

On peut aussi interdire la sauvegarde d'un signal écrité, prévoir un mot de passe empêchant le participant de quitter le prompteur, afficher une horloge (utile pour des locuteurs payés à l'heure), passer dans un mode où chaque locuteur ne lit qu'une partie du corpus, ou encore enregistrer plusieurs fois chaque phrase. Concernant la collecte des métadonnées, le projet peut prévoir que l'utilisateur soit invité à remplir lui-même son questionnaire. Enfin, les options audio (taux d'échantillonnage, résolution d'amplitude et nombre de canaux) sont paramétrables.

ROCme! permet la mise en place d'un questionnaire pour le recueil de métadonnées à partir de balises XML. À la création, le code XML peut être directement entré dans une boîte de l'interface pour pré-visualiser la version graphique du questionnaire. Un exemple est présenté dans la Figure 1, qui inclut les 6 valeurs de types de données prises en compte par ROCme ! (age, date, number, single, choice, multiple). Spécifier un type de données est optionnel, mais vivement recommandé car ROCme! dispose d'un module qui traite les valeurs de type pour proposer en temps réel des statistiques descriptives concernant le projet, permettant ainsi à l'investigateur de repérer rapidement certains déséquilibres et donc d'ajuster sa stratégie de recrutement.

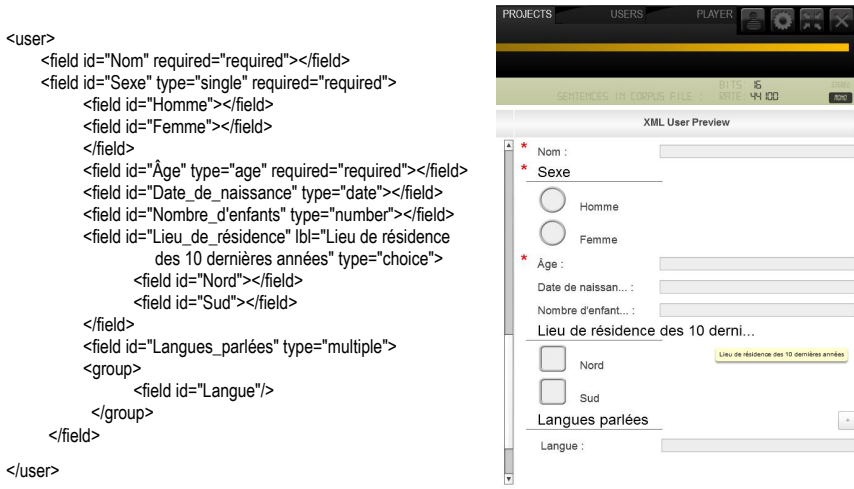


FIGURE 1 – Questionnaire en version XML et interface graphique correspondante

Remerciements

Le projet ROCme! a reçu le soutien financier du Bonus Qualité Recherche de l'Université Paris 7 Diderot et le soutien technique du Laboratoire Dynamique Du Langage.