

Projet Shtooka – Interview de Nicolas Vion

Dans l'expression [Rencontres Mondiales du Logiciel Libre](#), il y a le mot rencontre. Et c'est ainsi que nous avons fait connaissance avec Nicolas Vion, fondateur du fort intéressant [Projet Shtooka](#) qu'il nous a semblé judicieux de mettre en avant sur ce blog à la faveur de cette courte interview.



L'objectif du projet vise la création de matériel audio libre pour l'apprentissage des langues étrangères en mettant en place un ensemble de collections sonores de mots et expressions telle la [base audio libre de mots français](#) (cf [démonstration en ligne](#)) ainsi que les outils logiciels capables d'en tirer parti. Outre le français, on trouve aussi des corpus pour d'autres langues telles que le russe, l'ukrainien, le biélorusse, le wolof, le tchèque ou le chinois^[1].

Un entretien réalisé par Emmanuel Poirier (alias LeBidibule) pour Framasoft.

Bonjour Nicolas, je vais te demander de te présenter.

Je m'appelle Nicolas Vion, je m'intéresse depuis quelques années aux langues slaves. Après une licence de russe, je suis actuellement en licence d'ukrainien et travaille comme développeur système. Je m'occupe du [Projet Shtooka](#) depuis son origine.

En quoi consiste ton projet ?

Le projet Shtooka consiste à enregistrer des collections audio libres de mots, d'expressions idiomatiques et d'éléments du

langage avec des locuteurs natifs pour toutes les langues.

C'est aussi un ensemble de logiciels permettant d'enregistrer, d'organiser et d'exploiter ces collections. Cela permet, par exemple, d'ajouter l'audio dans des dictionnaires électroniques (les fichiers sont d'ores et déjà utilisés pour le [Wiktionnaire](#)), mais aussi de faciliter l'apprentissage du vocabulaire d'une langue étrangère en se faisant des fiches de vocabulaire audio à l'aide d'outils comme le [Kit Shtooka](#) (ou bientôt les [SWAC Tools](#)) et ainsi de réviser son vocabulaire à l'aide d'un baladeur MP3 ou Ogg Vorbis en baladodiffusion. Il est également possible de créer à partir de ces fiches de vocabulaire des activités pédagogiques au format [JClic](#) pour les plus petits...

Au delà de l'aspect pédagogique, ces collections représentent, à mon sens, un intérêt linguistiques et scientifique en regroupant en un ensemble cohérent un tel matériel audio qui, je pense, comptera à l'avenir.

Depuis quelque temps, le Projet Shtooka s'est monté en association, ce qui a permis au projet de s'agrandir un peu. Nous sommes aujourd'hui environ quatre personnes actives dans le projet, de plus cela va nous permettre de contacter différents organismes en tant qu'association.

Quelles est la différence entre le Kit Shtooka et les SWAC Tools ?

Le Kit Shtooka, qui est un ensemble de programmes pour Microsoft Windows (mais qui marche aussi sous GNU/Linux avec [Wine](#)), est en quelque sorte l'héritier du petit programme à l'origine du projet. Ce petit programme était une sorte de «bidouille» que j'avais écrit pour mes besoins personnels alors que je me trouvais en Russie, où les connexions Internet ne sont pas forcément bon marché. Aussi, j'ai trouvé un CD de [Delphi](#) et ai donc commencé à coder pour cet environnement.

Le Kit Shtooka est en fait issu d'une refonte de ce programme

originel. Il est mûr et utilisable, l'ayant moi-même utilisé au cours de mes études de Russe, mais des limitations sont apparues liées au choix techniques initiaux... Il y a maintenant un an j'ai commencé un travail de ré-écriture du Kit Shtooka en C++ et utilisant des libraires standards telles que GTK+ et [QT](#), ce qui permet une meilleure portabilité entre plate-formes et une plus grande modularité.

A l'heure actuelle, si le Kit Shtooka est pleinement utilisable, les SWAC Tools sont encore assez jeunes et d'utilisation plutôt difficile (*NdR : il manque notamment le logiciel d'enregistrement, le logiciel exerciceur « Repeat »*) mais devraient à terme remplacer les vieux outils. C'est en quelque sorte la nouvelle génération du Kit Shtooka.

Quand as-tu pensé à utiliser des licences libres pour vos productions et vos logiciels ?

A l'origine, je n'imaginai pas qu'il serait aussi facile d'enregistrer la prononciation d'autant de mots. Tout a commencé par intérêt personnel. Je faisais enregistrer la prononciation de mots russes par mes amis afin de me faciliter l'apprentissage du vocabulaire, sans penser à la licence de ces enregistrements.

Au bout d'un an, je me suis rendu compte que je disposais d'un corpus important (*NdR : mais inutilisable pour le futur projet, faute d'accord avec les locuteurs de l'époque*), et j'ai pensé faire une base de mots audio en français afin d'aider des professeurs de français FLE (Français Langue Étrangère), ou encore pour les étudiants en auto-apprentissage.

C'est au cours de ces enregistrements que j'ai amélioré la technique d'enregistrement (grâce à quelques modifications de l'interface, je pouvais désormais enregistrer plus de 1500 mots de suite sans me fatiguer). J'ai alors décidé de diffuser cette première collection qui regroupait plus de 7000

enregistrements sur une site Internet sous une licence libre (la [Creative Commons By](#)) et un ami m'a soufflé l'idée de faire [une dépêche](#) sur le site d'information LinuxFr.

Très vite, j'ai été contacté par Eric Streit qui était enthousiasmé par le projet et qui a rapidement enregistré une collection de mots russes avec sa femme originaire du Kazakhstan. Un ami chinois m'a également permis de réaliser le corpus chinois : le projet Shtooka était bel et bien lancé.

À mon sens, la licence libre s'imposait car c'était la seule façon de permettre la créativité. En effet, on ne peut imaginer soi-même toutes les applications possibles à partir d'une telle « matière première » sonore.

Enfin en voulant proposer mon outil d'enregistrement à des contributeurs du projet Wiktionnaire (entre autre Betsy Megas) j'ai eu à ré-écrire cet outil, pour en faire un programme séparé, plus facile à utiliser. Cela m'a permis par la même occasion de formaliser la façon de stocker les informations sur les mots enregistrés et les locuteurs, et ai proposé l'adoption des « [SWAC Metatags](#) ». En effet, s'il est possible d'enregistrer dans les fichiers audio Ogg Vorbis des informations pour les collections musicales (nom de l'album, nom de la chanson, style de musique) les « SWAC Metatags » permettent de faire la même chose avec les collections linguistiques grâce à une quarantaine de champs d'information : nom du locuteur, sa langue maternelle, région d'origine...

Avez-vous eu beaucoup de retours jusqu'alors ?

Jusqu'à présent, les retours d'expérience ont été peu nombreux; mais je remercie les personnes, notamment les professeurs de Langues Vivantes qui ont permis de faire évoluer le logiciel, et faire remonter les problèmes liés à la formation de l'outil.

Il est vrai que l'outil a été plus pensé pour des apprenants

autonomes que pour des professeurs. Les [RMLL 2009](#) ont d'ailleurs été l'occasion de rencontrer des gens d'autres projets qui utilisent nos collections. Nous avons également pris contact avec des libristes marocains qui ont l'intention de développer le Projet Shtooka dans leur pays (nous avons d'ailleurs pu enregistrer avec eux une petite collection de mots arabes durant les RMLL).

Quelles sont les aides que l'on peut vous apporter si l'on souhaite participer au projet ?

Les aides sont les bienvenues !

Tout d'abord, si l'on a des compétences pour une langue en particulier, il est possible de créer des listes de mots et expressions qui serviront lors de sessions d'enregistrements avec un locuteur natif. Pour cela il suffit de rassembler les éléments à enregistrer dans un simple fichier texte. Pour ce travail il faut quand même avoir une bonne connaissance de la langue en question : par exemple, pour le corpus français, se pose la question : faut-il prononcer les substantifs avec ou sans article ? Si oui, lesquels : définis ou indéfinis ? Pour d'autres langues se pose le problème des déclinaisons, du déplacement des accents toniques, des mutations, ou encore des cas d'homographie... Ce travail préparatoire peut être fait par un professeur, mais aussi par un apprenant.

Ensuite vous pouvez participer au projet en tant que « locuteur natif » de votre langue maternelle en enregistrant des mots ou expressions. Ceci-dit, il est important d'avoir un matériel d'enregistrement de qualité convenable, quelques connaissances sur la prise de son avec un ordinateur, et de disposer d'un endroit calme. Le logiciel d'enregistrement est très pratique et permet d'enregistrer plus de 1000 mots par heure. Ici, la communauté Shtooka peut aider (dans la mesure du possible) en trouvant le matériel, en organisant l'enregistrement...

Les professeurs de langues peuvent également nous aider en proposant des listes de vocabulaire bilingues organisées par thèmes ou par leçons, car en fait chaque collection audio est indépendante. De même, tous les travaux dérivés (exercices interactifs, imagiers pour les écoles primaires...) sont intéressants et bienvenus, et peuvent donner des idées pour compléter les collections audio.

Si vous êtes développeur C++ (ou autre) vous pouvez aider au développement de nos logiciels. Les développeurs Web et les graphistes peuvent nous donner un coup de main sur le site Internet...

Globalement, nous cherchons des aides ponctuelles mais aussi dans la durée. Il nous faudrait, par exemple, des personnes pour s'occuper du développement de telle ou telle langue (au passage, nous aurions besoin de quelqu'un qui puisse s'occuper de la collection espagnole), alors n'hésitez pas !

Notes

[1] Crédit photo : [Cesarastudillo](#) (Creative Commons By)