

Visualisation des corpus

Pascal Vaillant

21 avril 2010

1 Introduction

Les corpus de phénomènes de contact de langues, représentés en XML en utilisant le schéma de documents CORPUS-Contacts¹, peuvent être visualisés directement dans un navigateur internet à travers l'utilisation de la « feuille de transformation » XSLT “corpus.xsl”.

Une feuille de transformation XSLT² est un fichier qui décrit des transformations régulières de XML en HTML : en termes génériques, c'est une grammaire de transformation d'arbres.

Un logiciel « navigateur » (Internet Explorer, Firefox, Opera, Konqueror, etc.) est conçu, à la base, pour afficher du HTML. Si on l'utilise pour afficher un document XML, à défaut d'instructions spécifiques sur la manière d'afficher les éléments et attributs, il présente ce document en mettant en évidence les balises, comme dans la figure 1.a : en effet, les éléments d'un schéma XML quelconque n'indiquent pas *a priori* des instructions d'affichage — contrairement aux balises de HTML comme (gras) ou <i> (italique) — mais des éléments de contenu.

En revanche, si le document XML contient une petite ligne d'en-tête (invisible pour l'utilisateur) qui indique au navigateur qu'il doit être visualisé *par l'intermédiaire d'une certaine feuille de transformation XSLT*, alors le navigateur commence par transformer le XML en HTML, en utilisant la feuille XSLT indiquée, avant de l'afficher, comme dans la figure 1.b.

La feuille XSLT “corpus.xsl” contient des indications de visualisation pour les corpus linguistiques du programme Contacts, représentés sous la forme de fichiers XML suivant le schéma de documents CORPUS-Contacts.

¹Le schéma de documents CORPUS-Contacts : http://clapoty.vjf.cnrs.fr/contacts/doc/corpus_xsd.pdf.

²Documentation sur le site du World Wide Web Consortium : <http://www.w3.org/TR/xslt>.

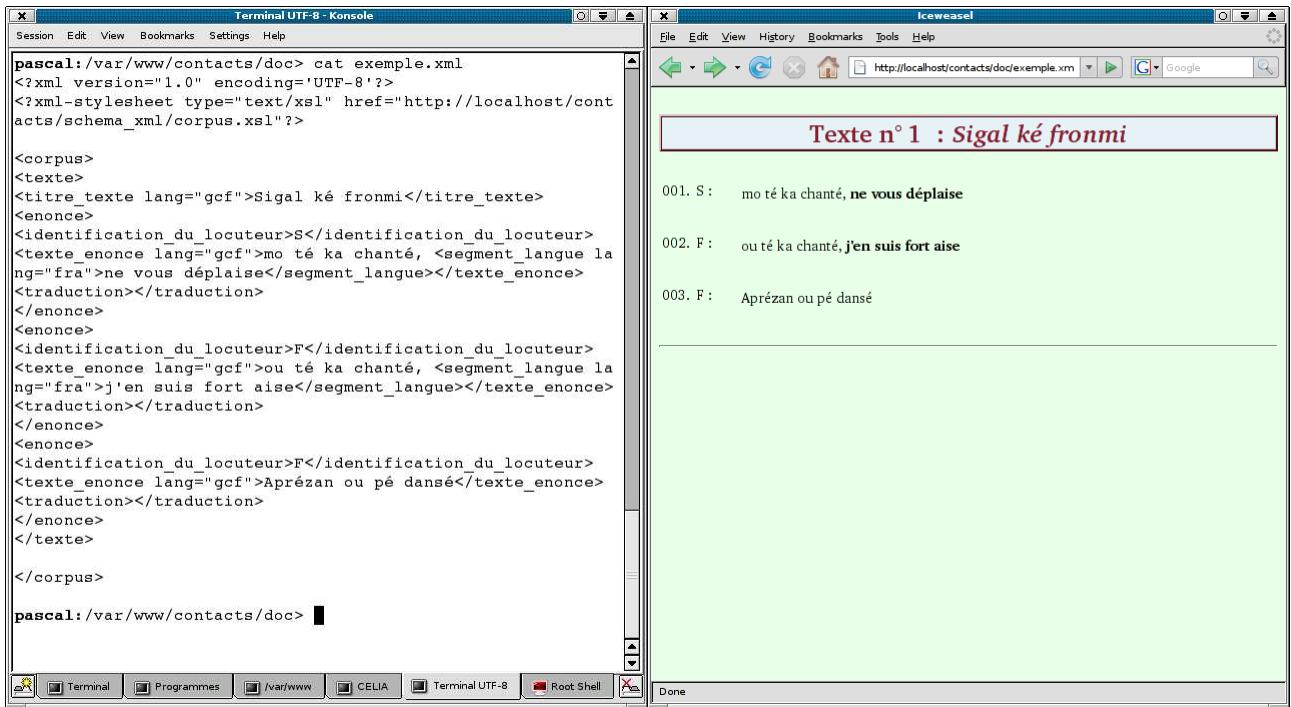
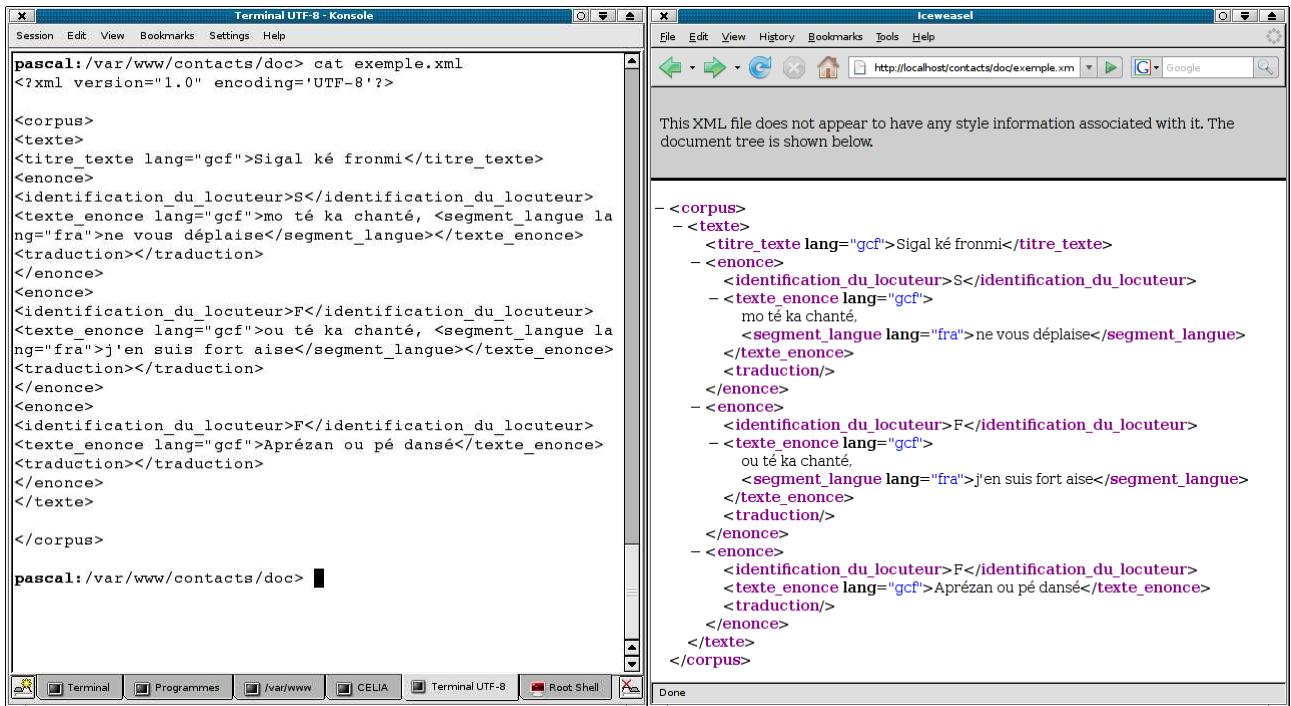


FIG. 1 – Présentation de fichiers balisés XML en absence et en présence d'une feuille de transformation XSLT. En haut (a), le navigateur (ici, *Iceweasel*) tente d'afficher le fichier XML en l'absence de feuille XSLT : il doit se contenter de mettre en évidence la hiérarchie des éléments et attributs du fichier. En bas (b), il dispose d'une feuille XSLT, c'est-à-dire d'une série d'instructions précises lui indiquant comment présenter visuellement chacun des éléments de contenu (la référence à cette feuille est indiquée dans la deuxième ligne du fichier source [à gauche], invisible à l'utilisateur).