

Utilisation des documents CORPUS-Contacts

Pascal Vaillant

30 avril 2012

1 Introduction

Ce document donne des indications pour la manipulation, sur votre ordinateur individuel, des fichiers de corpus de contacts de langues, avec les outils qui ont été développés initialement pour le programme « Contacts » du CELIA¹, puis dans le cadre du projet ANR CLAPOTY².

La gestion numérique des corpus de contacts de langues fait intervenir plusieurs composantes :

- Un « **schéma de documents** » **XML**, c'est-à-dire la définition de la structure des données qu'on manipule. Le schéma de documents défini pour cet objectif s'appelle "**corpus.xsd**".
- Un **logiciel d'édition** de XML, permettant de créer et d'éditer les documents dont le type est défini par le schéma précédemment cité. N'importe quel éditeur XML peut faire l'affaire (il en existe plusieurs), mais on suggère ici d'utiliser JAXE, logiciel libre pour lequel un fichier de configuration *ad hoc* a été créé pour le programme Contacts.
- une « **feuille de transformations** » **XSLT**, qui permet la visualisation des documents XML selon des conventions typographiques qui les rendent aisément lisibles. La feuille de transformations définie pour ce projet s'appelle "**corpus.xsl**".
- diverses autres composantes superficielles, dont une feuille de style CSS (qui définit des paramètres de visualisation comme la taille de la police de caractères, la couleur du fond d'écran ...)

Les différents éléments de cette chaîne de traitement sont disponibles sur un serveur web³ qui est installé dans les locaux du SEDYL (CELIA) à Villejuif : **clapoty**. Son adresse internet (visible de l'extérieur) est : **clapoty.vjf.cnrs.fr**.

N.B. Ce serveur est le serveur dédié au projet ANR CLAPOTY, qui poursuit et étend les efforts du programme « Contacts de langues » de l'ex-CELIA ; une présentation générale du projet CLAPOTY est accessible à partir de l'URL du site web « vitrine » :

<http://clapoty.vjf.cnrs.fr>

et l'accès au site interne de l'équipe se fait à partir de l'URL :

<http://clapoty.vjf.cnrs.fr/contacts>

1. CELIA : *Centre d'Étude des Langues Indigènes d'Amérique*, équipe aujourd'hui intégrée à l'UMR SEDYL : *Structure et Dynamique des Langues*.

2. CLAPOTY : *Contacts de Langues : Analyses Plurifactorielles assistées par Ordinateur et conséquences Typologiques*.

3. Il s'agit d'un serveur Apache 2 sur système GNU/Linux Debian, installé sur un Dell PowerEdge T610.

Le présent document décrit brièvement comment l'utilisateur de PC (Windows) peut installer le minimum sur son ordinateur personnel⁴ pour commencer à utiliser cette chaîne de gestion de documents numériques.

Pour des utilisateurs d'autres configurations (MacOS, Linux), il n'y a aucune difficulté pour utiliser les mêmes outils ; ce document ne prévoit pas le cas pour l'instant, mais les différences ne portent que sur des détails mineurs⁵.

2 Installation de JAXE

Bien qu'un document XML puisse être créé et modifié avec n'importe quel éditeur (spécialisé dans la manipulation de fichiers XML, comme *Oxygen* ou de nombreux autres ; ou même simple et générique, comme *WordPad* ...), je suggère l'utilisation de JAXE⁶, qui présente les avantages suivants :

- Développé en Java, JAXE fonctionne indifféremment sur toutes les plates-formes informatiques ;
- il est configurable, pour adapter l'interface à l'édition de fichiers XML devant se conformer à des schémas de documents particuliers ;
- JAXE est un logiciel libre, ce qui signifie qu'il n'est pas « verrouillé », mais que le programme lui-même est ouvert à des modifications⁷ ;
- et il est distribué gratuitement sous licence GPL.

Voici les instructions pour télécharger et installer sur un PC fonctionnant sous Windows le logiciel JAXE.

2.1 Téléchargement

Le paquetage JAXE pour l'édition de corpus de contacts de langues (CORPUS-Contacts) comprend le programme lui-même (légèrement modifié par rapport à sa version publique de référence) et un fichier de configuration spécifique pour les documents du type CORPUS-Contacts.

Une version à jour de ce paquetage, avec les versions les plus récentes des « *plug-ins* » et des fichiers de configuration spécifiques, est disponible à l'adresse :

<http://clapoty.vjf.cnrs.fr/contacts/jaxe.zip>.

4. On suppose ici que l'ordinateur personnel de l'utilisateur est connecté à internet : ceci permet de n'installer sur le disque dur local que le programme d'édition JAXE (les autres ressources restant disponibles en ligne à travers internet). Si l'on souhaite utiliser le système d'éditions de corpus en configuration hors-ligne, il faut également installer localement les autres composantes (« *corpus.xsd* » et « *corpus.xsl* »), ce qui peut s'envisager si nécessaire, mais est un peu plus compliqué.

5. Dans l'intervalle, consulter l'auteur : vaillant@vjf.cnrs.fr.

6. Éditeur XML configurable développé par Damien Guillaume, de l'Observatoire de Paris-Meudon. Disponible sous licence GPL (*GNU General Public License*) à partir de : <http://jaxe.sourceforge.net/Jaxe.html>.

7. Des adaptations ont par exemple été faites à JAXE pour son utilisation dans le cadre du programme Contacts ; le détail — dont la lecture n'est pas nécessaire pour l'utilisateur ordinaire — en est fourni dans le fichier « *README_PV.txt* », livré avec le paquet logiciel dont l'adresse est donnée dans ce document.

2.2 Copie sur le disque local

Il faut tout d'abord installer ce fichier, “jaxe.zip”, dans le répertoire “C:\Program Files” de l'ordinateur⁸.

Selon le navigateur que vous employez et ses configurations par défaut, lorsque vous tenterez d'ouvrir le fichier “jaxe.zip” mentionné à l'adresse ci-dessus, soit on vous proposera de choisir un emplacement de destination, soit le fichier sera copié d'office, par défaut, sur le Bureau. Si vous êtes dans le premier cas, choisissez le disque “C:” (Système), puis le répertoire “Program Files”; si vous êtes dans le second, laissez le téléchargement se terminer, puis copiez le fichier “jaxe.zip” (icône sur le bureau) dans le répertoire “C:\Program Files” en le faisant glisser dessus dans l'interface de l'explorateur de fichiers.

2.3 Décompression et installation

Ouvrez l'explorateur de fichiers (soit en tapant la combinaison de touches “touche Windows” + “E”, soit en choisissant “Explorateur” dans le menu Démarrer, sous-menu Programmes).

Ouvrez le répertoire “C:\Program Files”.

Double-cliquez sur le fichier “jaxe.zip”.

Cliquez (en haut à gauche) sur l'option “Extraire tous les fichiers”. Le choix d'emplacement proposé par défaut (“C:\Program Files\jaxe”) est bon. Laissez la décompression et l'installation des fichiers se dérouler jusqu'au bout.

À la fin de cette opération, vous pouvez supprimer le fichier compressé “jaxe.zip”.

L'éditeur JAXE est maintenant installé dans le sous-répertoire “jaxe” du répertoire des programmes (“C:\Program Files”). Si une mise à jour était nécessaire, il suffirait de supprimer ce répertoire “jaxe”, et de répéter les étapes décrites jusqu'à maintenant à partir de la nouvelle version du packaging dont l'adresse est donnée en §2.1.

3 Création d'un répertoire de travail

Créez, à l'emplacement de votre choix sur votre disque dur, un répertoire de travail dans lequel vous stockerez localement vos corpus en cours d'édition (avant stockage sur le serveur).

Si vous avez un disque de données (p.ex. “D:”) distinct du disque dur où est installé Windows (“C:”), il est préférable de l'utiliser. En effet une réinstallation de Windows suite à un hypothétique incident informatique (p.ex. une infection due à un virus) écrasera tout ce qui se trouve sur le disque “C:”⁹.

Admettons par exemple que vous créez un répertoire de travail “D:\clapoty”.

8. Sur certaines versions françaises de Windows (c'est très pénible, d'ailleurs), le nom réel du répertoire “Program Files” est masqué, dans l'interface de l'explorateur de fichiers, par un alias en français : “Programmes”. Ce qui serait merveilleux (vive la langue française) si ce n'était pas, justement, qu'un alias, qui entre parfois en conflit avec le véritable nom sous-jacent qui reste “Program Files”. Si, sur votre ordinateur, vos programmes ont l'air d'être stockés dans un répertoire qui s'appelle “Programmes”, merci de considérer, en poursuivant la lecture de ce document, que tout ce qui est dit ici au sujet de “Program Files” s'applique chez vous à “Programmes”.

9. Notez que les répertoires par défaut que propose Windows (dans la série “Mes Documents”, “Mes Images”, etc.) se trouvent physiquement sur “C:”, et ne sont donc pas la solution idéale de ce point de vue ... mais vous pouvez bien sûr les utiliser pour profiter du « pré-câblage » imposé par Windows (qui fait que ces répertoires sont toujours disponibles par défaut avant les autres lorsque vous ouvrez un nouveau fichier).

Commencez par télécharger, dans ce répertoire, le fichier “squelette_corpus.xml” situé à l’adresse :

http://clapoty.vjf.cnrs.fr/contacts/corpus/squelette_corpus.xml

Il s’agit d’un fichier (presque) vide, qui contient simplement l’ossature essentielle du schéma de documents XML CORPUS-Contacts.

Important : À l’avenir, pour entamer un nouveau corpus, commencez toujours par recopier ce fichier “squelette_corpus.xml” en lui donnant un autre nom, puis repartez de cette ossature vide.

En effet, si au lieu de procéder comme cela, vous utilisez la fonction “Nouveau” de JAXE, il manquera, dans le document créé, la ligne indiquant quelle feuille de transformations XSLT doit être utilisée, et le document ne s’affichera pas correctement par la suite¹⁰.

3.1 Édition d’un nouveau corpus

Pour travailler sur un nouveau corpus, commencer par recopier le fichier “squelette_corpus.xml” en lui donnant un autre nom. Par exemple, si l’on utilise l’explorateur de fichiers Windows, on peut procéder comme suit :

- Ouvrir le répertoire de travail contenant les corpus ;
- cliquer (une seule fois) sur l’icône du fichier “squelette_corpus.xml” avec le bouton droit de la souris (“clic droit”) ;
- choisir “Copier” dans le menu ;
- désélectionner le fichier “squelette_corpus.xml” en cliquant une fois (avec le bouton gauche de la souris) dans le fond vide de la partie droite de la fenêtre (“Volet fichiers”) ;
- cliquer une fois, avec le bouton droit de la souris (“clic droit”) dans le fond de la fenêtre ;
- choisir “Coller” dans le menu : un nouveau fichier apparaît, avec par défaut le nom “Copie de squelette_corpus.xml” (et il est par défaut sélectionné) ;
- cliquer à nouveau une fois avec le bouton gauche de la souris, *en cliquant précisément sur le nom du fichier* : celui-ci devrait à présent apparaître en surbrillance, et encadré d’un petit rectangle indiquant que le nom peut-être modifié ;
- taper un nouveau nom à la place de “Copie de squelette_corpus.xml” (par exemple, si vos initiales sont XY, “xy_awala_2006_01.xml”).

Tout corpus situé dans votre répertoire de travail peut être ouvert de deux manières :

- soit en lecture seule, à travers un navigateur web ordinaire (Internet Explorer, Firefox ...) : c’est ce qui se produit par exemple quand vous faites un “double-clic” sur le fichier “xy_awala_2006_01.xml” ; dans ce cas, c’est la feuille de transformations “corpus.xsl” qui détermine les paramètres d’affichage du texte¹¹.
- soit en édition, à travers JAXE.

10. Une explication plus claire est donnée dans http://clapoty.vjf.cnrs.fr/contacts/doc/corpus_xsl.pdf.

11. Cf. http://clapoty.vjf.cnrs.fr/contacts/doc/corpus_xsl.pdf.

4 Utilisation de JAXE pour éditer un corpus

Pour éditer un corpus avec l'aide de l'éditeur JAXE, il faut d'abord commencer par lancer ce programme :

- Dans une fenêtre de l'explorateur de fichiers, ouvrir le dossier “C:\Program Files\jaxe” ;
- double-cliquer sur l'icône du fichier “jaxe.bat”¹².

N.B. Pour accélérer cette étape à l'avenir, on peut créer un raccourci vers le fichier “jaxe.bat” sur le bureau ...

Lorsque le programme JAXE est lancé, il affiche au départ seulement une petite fenêtre avec les trois choix “Nouveau”, “Ouvrir” et “Quitter” (ou “New”, “Open” et “Quit” si vous utilisez une version anglaise de Windows).

Cliquer sur “Ouvrir”.

Puis choisir le fichier XML que vous voulez éditer dans la fenêtre de choix du fichier (il faut revenir à votre dossier de travail contenant vos corpus, puis sélectionner le corpus que vous souhaitez éditer).

L'interface de JAXE est ensuite conçue pour être relativement intuitive. La fenêtre principale comprend deux volets, un à gauche et l'autre à droite (fig. 1).

Le volet de droite contient le texte en cours d'édition : les éléments sont notés par des balises de début et de fin apparaissant en couleur (jaune si l'élément est complet, rouge s'il est incomplet).

Le volet de gauche affiche soit la liste des éléments que l'on peut insérer dans le contexte en cours (onglet “Insertion”, cf. fig. 1), soit l'arborescence du document XML (onglet “Arbre”, cf. fig. 2.a), soit les attributs de l'élément en cours (onglet “Attributs”, cf. fig. 2.b).

4.1 Les opérations d'édition de corpus de contacts de langues

La description complète de la structure d'un document conforme au schéma XML *CORPUS-Contacts* est donnée dans le document “corpus_xsd.pdf”¹³. On ne donne donc ici que quelques indications techniques relevant plutôt de l'utilisation de JAXE pour éditer des documents de ce type.

4.1.1 Structure d'un corpus de contacts de langues

Pour mémoire : un corpus de contacts de langues (un document XML “*CORPUS-Contacts*”) est constitué d'un en-tête, suivi d'un ou de plusieurs textes constituant le corpus.

L'en-tête contient des informations générales sur le corpus (le titre général, la date de compilation, le nom du ou des éditeurs, une catégorisation des situations de contacts de langues — mises en évidence selon des typologies décrites dans la littérature —, ainsi qu'une éventuelle

12. Parfois, l'extension “.bat” est masquée : on reconnaît alors le fichier, qui apparaît simplement sous le nom “jaxe”, au fait qu'il est noté comme étant du type « fichier de commandes MS-DOS » (si son type est visible, ce qui est le cas par exemple si l'on laisse le curseur de la souris pointé sur lui pendant plus d'une seconde), ou qu'il apparaît sous la forme d'une icône représentant une petite roue dentée (si l'on est en mode affichage des icônes).

13. Disponible à l'adresse : http://clapoty.vjf.cnrs.fr/contacts/doc/corpus_xsd.pdf.

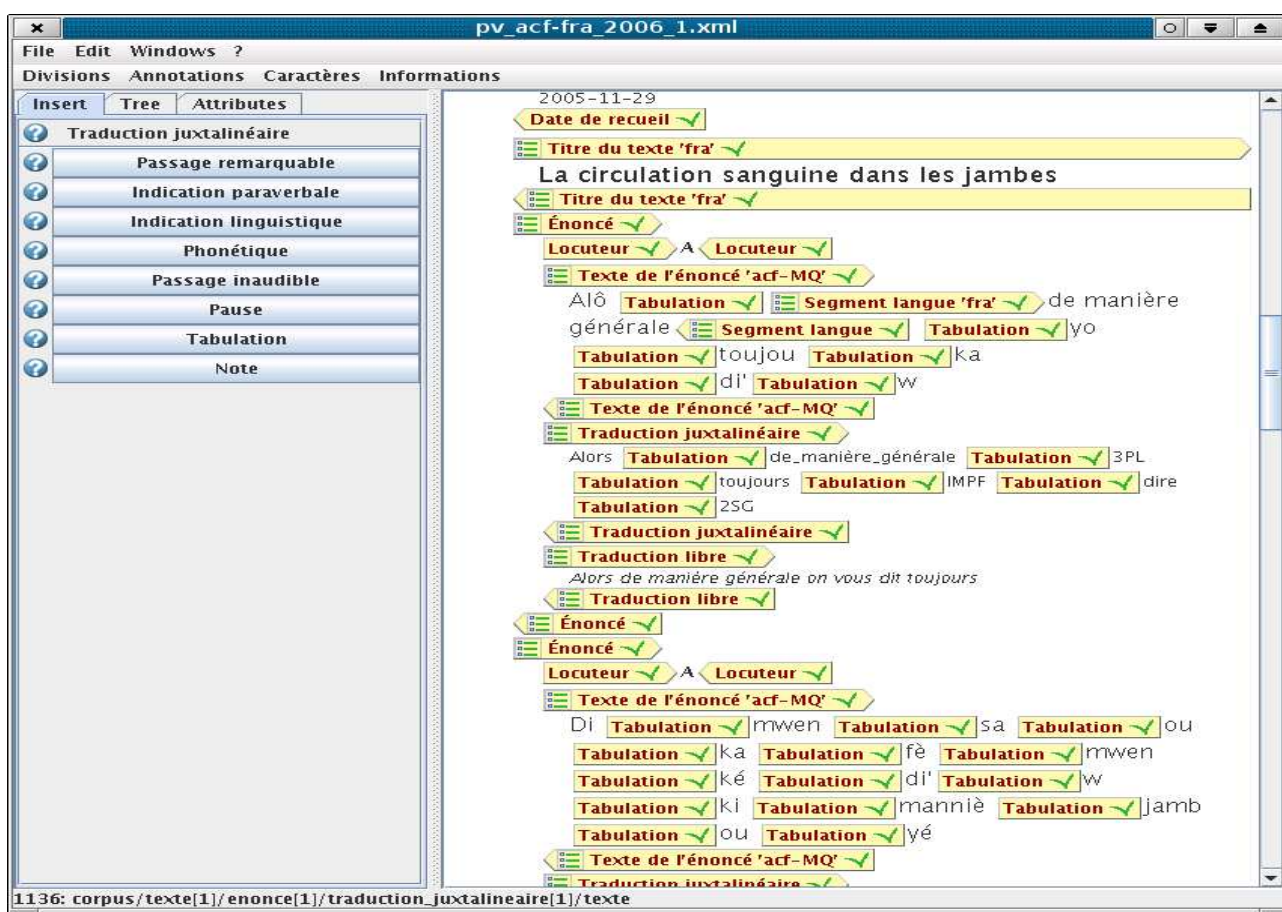


FIG. 1 – Fenêtre principale de JAXE, avec onglet d'insertion à gauche.

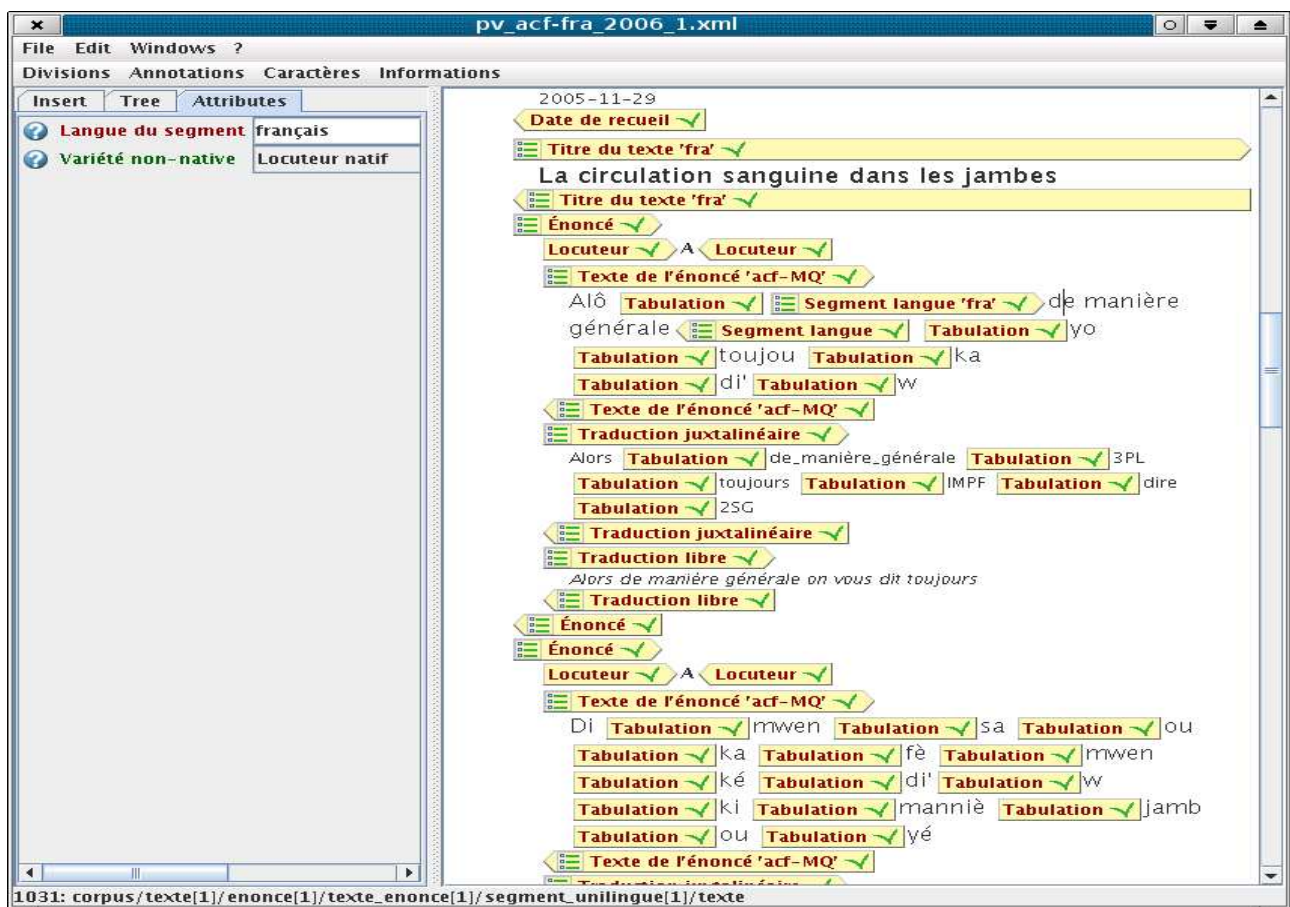
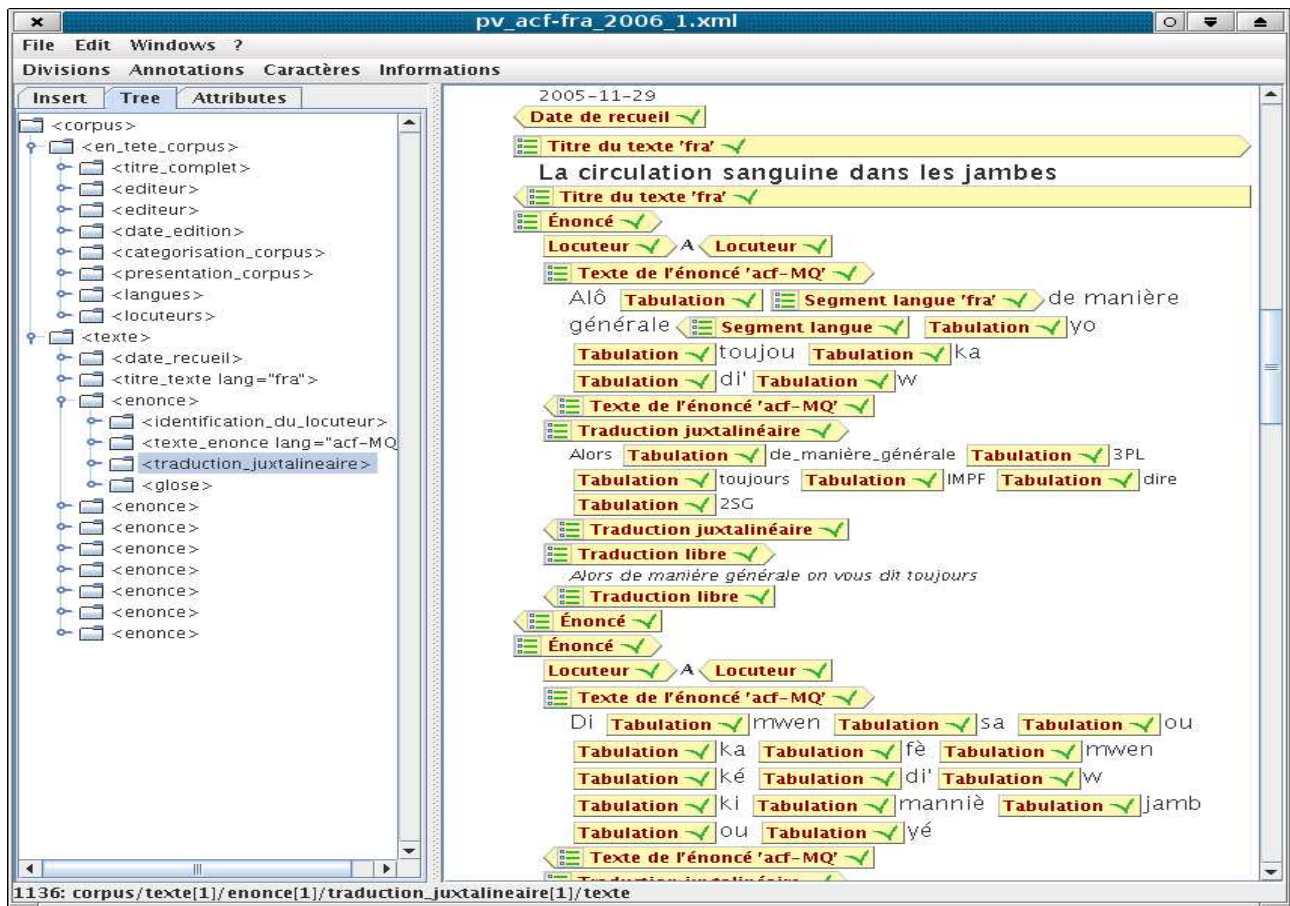


FIG. 2 – Fenêtre principale de JAXE, avec onglet d'arborescence (a, en haut) et onglet d'attributs (b, en bas).

description supplémentaire en texte libre) ; un inventaire complet des langues utilisées dans le corpus ; et un inventaire complet des locuteurs intervenant dans le corpus.

Chaque texte, pour sa part, contient un titre, une date de recueil, une éventuelle description particulière en texte libre, et — surtout — une série d'énoncés, qui constituent son contenu.

Chaque énoncé contient l'indication d'un locuteur et un texte original transcrit ; et, facultativement : une traduction juxtalinéaire (morphème par morphème) ; une ligne contenant la liste des catégories grammaticales de chaque morphème (« POS-tags », i.e. *Part-of-Speech tags* : étiquettes de parties du discours) ; et une traduction libre.

4.1.2 Insertion d'éléments de description linguistique

Dans les textes des énoncés, on peut rajouter des notes de bas de page, des indications concernant les changements de langue, les passages remarquables, indications non-verbales, etc. : voir les éléments disponibles selon le contexte dans le menu de gauche (onglet "Insertion").

À chaque fois que l'on veut insérer une balise spéciale de ce genre, on a le choix entre les deux possibilités suivantes :

- soit on clique d'abord sur le bouton dans le menu à gauche (mettons par exemple sur "Passage remarquable"), ce qui insère à la fois la balise de début et la balise de fin, et positionne le curseur à l'intérieur de la zone (après, pour en sortir, il suffit d'avancer avec la flèche droite jusqu'après la balise de fin) ;
- soit on tape d'abord le texte à annoter, puis on le sélectionne à la souris, et ensuite seulement on clique sur le bouton du menu (dans ce cas il insère automatiquement la balise de début et la balise de fin de part et d'autre de la zone sélectionnée).

4.1.3 Traduction juxtalinéaire

Alignement. Indication particulière concernant la traduction juxtalinéaire morphème par morphème : on doit pouvoir l'aligner avec des points de repère correspondants dans l'énoncé ; il faut donc pour cela insérer un certain nombre de points de tabulation dans l'énoncé d'une part, et dans la traduction juxtalinéaire d'autre part. (Convention : espace avant un point de tabulation, pas d'espace après). Exemple : fig. 3.

L'insertion d'un point de tabulation se fait en utilisant simplement la touche "Tabulation" du clavier.

N.B. Si l'on utilise la ligne contenant la liste des catégories grammaticales des morphèmes (liste des étiquettes de parties du discours), il faut impérativement y positionner aussi le même nombre de points de tabulation, aux positions correspondantes.

Abréviations. La traduction juxtalinéaire utilise fréquemment des abréviations en petites capitales pour transcrire les sens et fonctions des morphèmes de la langue d'origine. Afin de faciliter le choix et l'insertion de ces abréviations, un menu « Abréviations » a été inclus dans la configuration de JAXE pour l'édition de corpus.

Ce menu est composé de sous-menus, qui regroupent tout d'abord, à un premier niveau, des grandes catégories d'abréviations (indices, relations, constructions, parties du discours, symboles) ; puis, au second niveau, différentes valeurs d'un même attribut (p.ex. *genre* regroupe *masculin*, *féminin*, *neutre*).

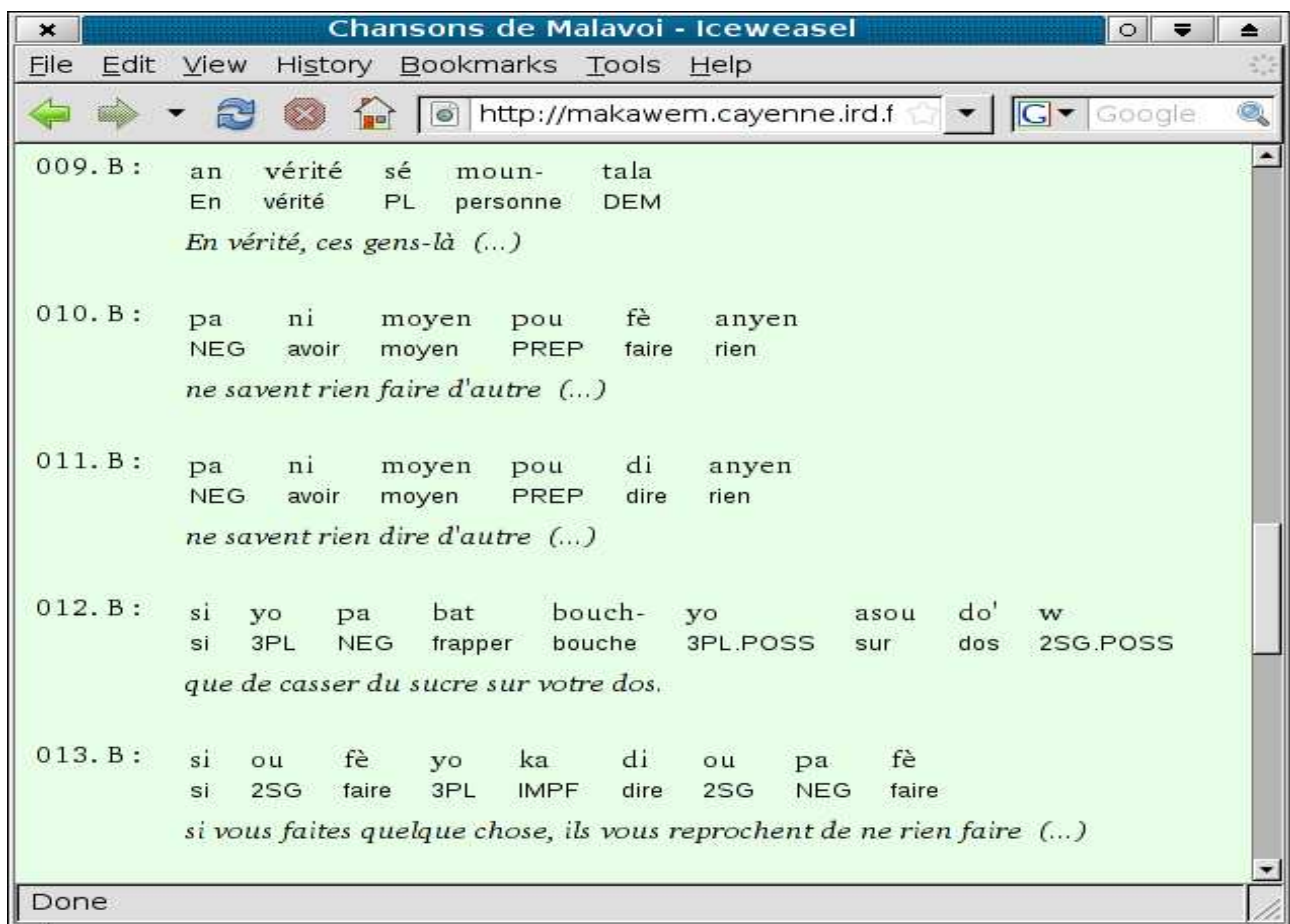
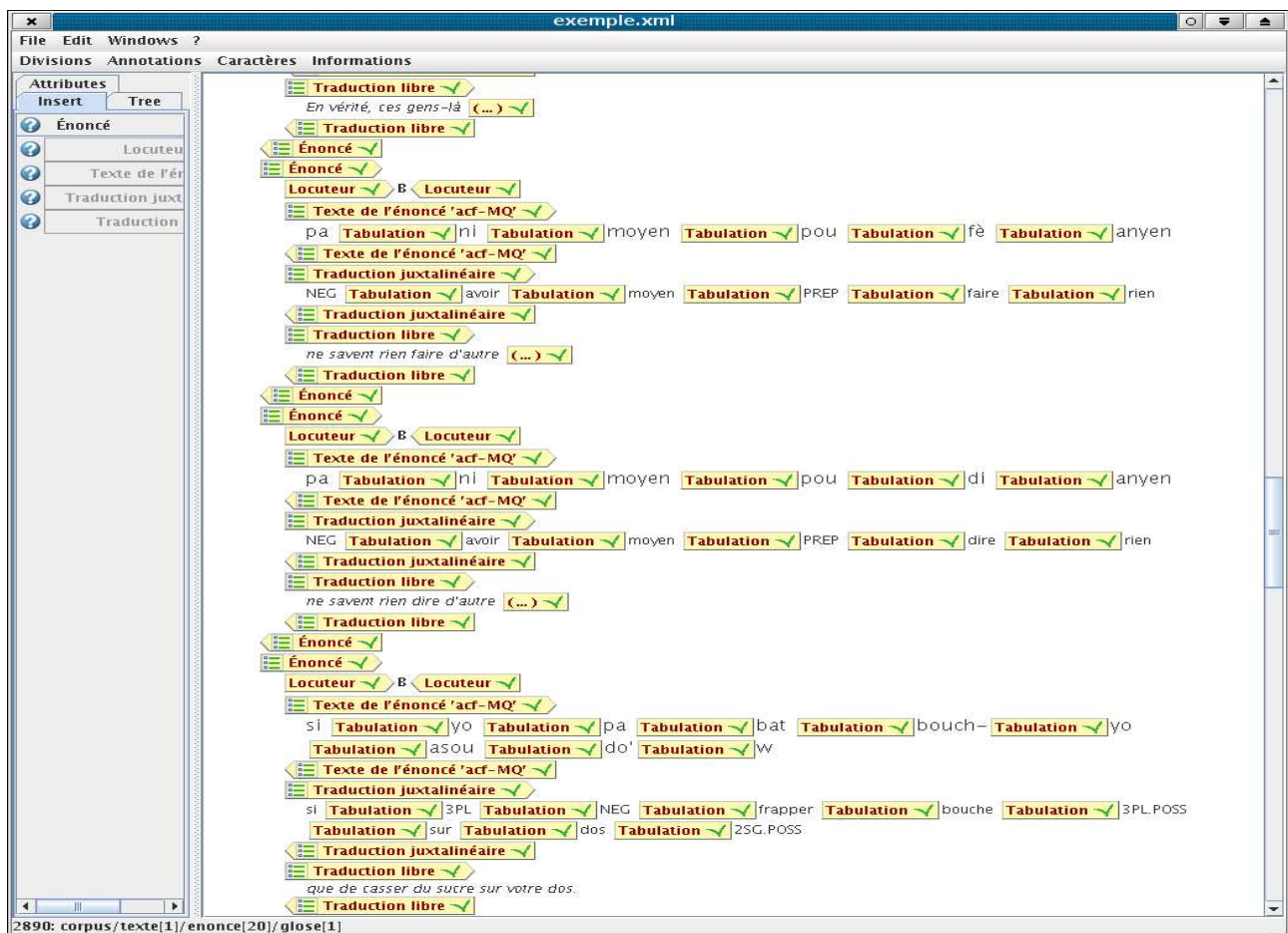


FIG. 3 – (a) en haut : exemple d'insertion de points de tabulation dans JAXE, pour aligner texte de l'énoncé et traduction interlinéaire; (b) en bas : visualisation à travers la feuille de transformation XSLT.

La liste des abréviations disponibles à la date d'édition de ce document est indiquée dans les tableaux 1 à 12. Cette liste a été conçue au départ sur la base de diverses sources récapitulant les usages les plus répandus dans les travaux de typologie linguistique (notamment la « liste de Leipzig » — cf. plus bas pour la liste complète des sources), augmentée d'abréviations qui ont semblé utile pour la représentation des langues manifestées dans les corpus du projet CLAPOTY.

Pour être précis, voici quelles sont les sources de la liste d'abréviations établie :

1. La petite liste de référence incluse dans le guide synthétique édité par les linguistes de Leipzig — Comrie, Haspelmath et Bickel (*The Leipzig Glossing Rules. Conventions for interlinear morpheme-by-morpheme glosses*, collectif, Max-Planck-Gesellschaft et Université de Leipzig¹⁴, dernière édition 2008, p. 8-10).
2. Des annotations courantes relevées (ou *adaptées*) au fil de divers articles du numéro d'Amerindia sur les langues de Guyane. (*Langues de Guyane*, Amerindia 26/27, CNRS, 2e trimestre 2002¹⁵).
3. La liste donnée par Christian Lehmann dans son article “Interlinear Morphemic Gloss” (in G. Booij, C. Lehmann, J. Mugdan & S. Skopeteas (eds.), *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung. 2. Halbband*. Berlin, Walter de Gruyter [Handbücher der Sprach- und Kommunikationswissenschaft, 17.2]¹⁶, p. 15-23).
4. La liste incluse dans l'ouvrage de William Croft (*Typology and Universals*, 2nd edition, [Cambridge Textbooks in Linguistics], Cambridge University Press, 2003, p. xix-xxiii).
5. La liste donnée par Croft sur son site internet, qui unifie celle qu'il donne dans *Typology and Universals* avec celle de Leipzig¹⁷.
6. Les indications élaborées dans le cadre du projet EuroTyp (1990s)¹⁸.

[1] a servi de source principale, [2] (et quelques autres ouvrages, par exemple sur l'arawak ou le ndyuka) de source d'inspiration pour quelques ajouts utiles pour les langues de Guyane ; et les autres (surtout [4]) de sources de vérification ou de recherche de concordance. Dans le fichier texte où est stockée la liste d'abréviations, les sources utilisées sont consignées.

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories indicielles			
<i>Catégories indicielles (principalement nominales)</i>			
nombre	singulier	<i>singular</i>	SG
nombre	duel	<i>dual</i>	DU
nombre	pluriel	<i>plural</i>	PL
nombre	collectif	<i>collective</i>	COLL
genre	masculin	<i>masculine</i>	M
genre	féminin	<i>feminine</i>	F
genre	neutre	<i>neuter</i>	N

TAB. 1 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : indices de genre et nombre

Il est important de noter que cette liste a vocation à évoluer. Le menu « Abréviations » de l'interface JAXE n'est pas « codé en dur » en XML dans le fichier de configuration, mais il est

14. URL : <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>

15. URL : http://celia.cnrs.fr/Fr/Am_Num.htm#26-27

16. Version révisée 2004 disponible à l'URL : http://www.christianlehmann.eu/CL_Publ/IMG.PDF

17. URL : <http://www.unm.edu/~wcroft/Papers/TypAbbrev.pdf>

18. URL : <http://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaire/eurotyp-guidelines.php>

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories indicielles (suite)			
<i>Catégories indicielles (principalement nominales)</i>			
personne	première personne	<i>first person</i>	1
personne	deuxième personne	<i>second person</i>	2
personne	troisième personne	<i>third person</i>	3
personne	inclusif	<i>inclusive</i>	INCL
personne	exclusif	<i>exclusive</i>	EXCL
personne	nous inclusif duel	<i>dual inclusive we</i>	12
personne	nous exclusif	<i>exclusive we</i>	13
personne	nous inclusif général	<i>plural inclusive we</i>	123
personne	2ème personne pluriel	<i>2nd person plural</i>	22
personne	3ème personne pluriel	<i>3rd person plural</i>	33
personne	1ère personne singulier	<i>1st person singular</i>	1SG
personne	2ème personne singulier	<i>2nd person singular</i>	2SG
personne	3ème personne singulier	<i>3rd person singular</i>	3SG
personne	1ère personne pluriel	<i>1st person plural</i>	1PL
personne	2ème personne pluriel	<i>2nd person plural</i>	2PL
personne	3ème personne pluriel	<i>3rd person plural</i>	3PL
personne	nous inclusif	<i>inclusive we</i>	1PL;INCL
personne	nous exclusif	<i>exclusive we</i>	1PL;EXCL

TAB. 2 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : indices de personne (simple)

ré-engendré automatiquement par un programme, à partir d'un fichier contenant la liste des abréviations en format texte¹⁹.

Pour insérer une abréviation de catégorie morphosyntaxique, l'usage des abréviations incorporées dans le menu qui vient d'être décrit n'est pas obligatoire. Chaque utilisateur peut toujours préférer taper à la main, s'il le souhaite, et s'il trouve cela plus rapide, les caractères composant l'abréviation souhaitée. Ces menus d'insertion n'ont donc pas pour but de *contraindre* de manière stricte l'ensemble des abréviations morphologiques. Leur justification, malgré tout, est double :

- Ils ont pour but premier de faciliter l'accès aux abréviations qui ne sont pas suffisamment fréquentes pour être connues « par cœur », et d'éviter ainsi à l'utilisateur d'aller se référer à une source comme celles citées plus haut ([1] à [6]).
- Mais ils ont également pour but de contribuer à *normaliser* les abréviations utilisées, tant en interne, au sein de l'équipe CLAPOTY ; qu'en externe, dans une optique de réutilisation éventuelle et de partage de données au sein d'un groupe d'utilisateurs plus large. Sans contraindre (l'utilisateur est toujours libre de choisir l'abréviation qu'il préfère), le système de suggestion d'abréviations promeut l'emploi d'abréviations standard.

Un exemple : l'article de Marie-France Patte, dans le numéro d'Amerindia consacré aux *Langues de Guyane*, utilise le terme de *localisateur* (abrégié LOC) pour désigner le morphème arawak qui, suffixé à un nom absolu (aliénable), indique qu'il fonctionne comme le terme possédé d'une relation de possession (le contraire du génitif, en quelque sorte, puisqu'il indique la possession, mais s'applique au terme possédé et non au terme possédant). Dans le même numéro d'Amerindia, l'article d'Odile Lescure emploie le terme de *relateur* (abrégié REL) pour désigner un morphème kali'na de même fonction. Toujours dans le même numéro, Eliane Camargo utilise pour un morphème de même rôle en Apalai l'abréviation ALIEN (pour : suffixe

19. URL : <http://clapoty.vjf.cnrs.fr/contacts/ABREVIATIONS.txt>.

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories indicielles (suite)			
<i>Catégories indicielles (principalement nominales)</i>			
personne	1ère personne agent	<i>1st person agent</i>	1A
personne	1ère personne patient	<i>1st person patient</i>	1P
personne	2ème personne agent	<i>2nd person agent</i>	2A
personne	2ème personne patient	<i>2nd person patient</i>	2P
personne	3ème personne agent	<i>2nd person agent</i>	3A
personne	3ème personne patient	<i>2nd person patient</i>	3P
personne	nous inclusif duel agent	<i>dual inclusive we agent</i>	12A
personne	nous inclusif duel patient	<i>dual inclusive we patient</i>	12P
personne	nous exclusif agent	<i>exclusive we agent</i>	13A
personne	nous exclusif patient	<i>exclusive we patient</i>	13P
personne	nous inclusif général agent	<i>plural inclusive we agent</i>	123A
personne	nous inclusif général patient	<i>plural inclusive we patient</i>	123P
personne	2ème personne pluriel agent	<i>2nd person plural agent</i>	22A
personne	2ème personne pluriel patient	<i>2nd person plural patient</i>	22P
personne	3ème personne pluriel agent	<i>3rd person plural agent</i>	33A
personne	3ème personne pluriel patient	<i>3rd person plural patient</i>	33P
personne	1A et 2P	<i>1A and 2P</i>	1>2
personne	1A et 3P	<i>1A and 3P</i>	1>3
personne	1A et 22P	<i>1A and 22P</i>	1>22
personne	1A et 33P	<i>1A and 33P</i>	1>33
personne	2A et 1P	<i>2A and 1P</i>	2>1
personne	2A et 3P	<i>2A and 3P</i>	2>3
personne	2A et 13P	<i>2A and 13P</i>	2>13
personne	2A et 33P	<i>2A and 33P</i>	2>33
personne	3A et 1P	<i>3A and 1P</i>	3>1
personne	3A et 2P	<i>3A and 2P</i>	3>2
personne	3A et 3P	<i>3A and 3P</i>	3>3
personne	3A et 12P	<i>3A and 12P</i>	3>12
personne	3A et 13P	<i>3A and 13P</i>	3>13
personne	3A et 123P	<i>3A and 123P</i>	3>123
personne	3A et 22P	<i>3A and 22P</i>	3>22
personne	3A et 33P	<i>3A and 33P</i>	3>33
personne	12A et 3P	<i>12A and 3P</i>	12>3
personne	12A et 33P	<i>12A and 33P</i>	12>33
personne	13A et 2P	<i>13A and 2P</i>	13>2
personne	13A et 3P	<i>13A and 3P</i>	13>3
personne	13A et 22P	<i>13A and 22P</i>	13>22
personne	13A et 33P	<i>13A and 33P</i>	13>33
personne	123A et 3P	<i>123A and 3P</i>	123>3
personne	123A et 33P	<i>123A and 33P</i>	123>33
personne	22A et 1P	<i>22A and 1P</i>	22>1
personne	22A et 3P	<i>22A and 3P</i>	22>3
personne	22A et 13P	<i>22A and 13P</i>	22>13
personne	22A et 33P	<i>22A and 33P</i>	22>33
personne	33A et 1P	<i>33A and 1P</i>	33>1
personne	33A et 2P	<i>33A and 2P</i>	33>2
personne	33A et 3P	<i>33A and 3P</i>	33>3
personne	33A et 12P	<i>33A and 12P</i>	33>12
personne	33A et 13P	<i>33A and 13P</i>	33>13
personne	33A et 123P	<i>33A and 123P</i>	33>123
personne	33A et 22P	<i>33A and 22P</i>	33>22
personne	33A et 33P	<i>33A and 33P</i>	33>33

TAB. 3 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : indices combinant les catégories de personne et de rôle actanciel

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories indicielles (suite)			
<i>Catégories indicielles (principalement nominales)</i>			
détermination	défini	<i>definite</i>	DEF
détermination	indéfini	<i>indefinite</i>	INDF
détermination	spécifique	<i>specific</i>	SPE
détermination	générique	<i>generic</i>	GNR
détermination	démonstratif	<i>demonstrative</i>	DEM
détermination	déictique	<i>deictic</i>	DEIC
détermination	proximal	<i>proximal</i>	PROX
détermination	non-proximal	<i>non proximal</i>	NPROX
détermination	distal	<i>distal</i>	DIST
détermination	déictique de 1ère personne	<i>1st person deictic</i>	D1
détermination	déictique de 2ème personne	<i>2nd person deictic</i>	D2
détermination	déictique de 3ème personne	<i>3rd person deictic</i>	D3
détermination	déictique de nous inclusif dual	<i>dual inclusive we deictic</i>	D12
classe sémantique	animé	<i>animate</i>	AN
classe sémantique	inanimé	<i>non animate</i>	NAN
classe sémantique	humain	<i>human</i>	HUM
classe sémantique	non-humain	<i>non human</i>	NHUM
classe sémantique	classe nominale	<i>noun class</i>	CL

TAB. 4 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : indices de détermination et de classe

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories indicielles (suite)			
<i>Catégories indicielles (adjectivales : degrés adjectivaux)</i>			
degré	comparatif	<i>comparative</i>	CMPR
degré	superlatif	<i>superlative</i>	SUP

TAB. 5 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : degrés adjectivaux

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories indicielles (suite)			
<i>Catégories indicielles (principalement verbales)</i>			
voix	passif	<i>passive</i>	PASS
voix	antipassif	<i>antipassive</i>	ANTIP
voix	applicatif	<i>applicative</i>	APPL
voix	causatif	<i>causative</i>	CAUS
voix	récioproque	<i>reciprocal</i>	RECP
voix	réflexif	<i>reflexive</i>	REFL
aspect	complétif	<i>completive</i>	COMPL
aspect	duratif	<i>durative</i>	DUR
aspect	imperfectif	<i>imperfective</i>	IPFV
aspect	perfectif	<i>perfective</i>	PFV
aspect	progressif	<i>progressive</i>	PROG
aspect	résultatif	<i>resultative</i>	RES
aspect	dispersion	<i>scattering</i>	SCAT
aspect	simultané	<i>simultaneous</i>	SIM
aspect	totalisation	<i>gathering</i>	GATH
aspect	imminent	<i>imminent</i>	IMMIN
aspect	prospectif	<i>prospective</i>	PROSP
aspect	accompli	<i>perfect</i>	PRF
aspect	inaccompli	<i>non perfect</i>	NPRF
aspect	habituel	<i>habitual</i>	HAB
aspect	itératif	<i>iterative</i>	ITER
aspect	statif	<i>stative</i>	STAT
aspect	processif	<i>processive</i>	PROC
temps	présent	<i>present</i>	PRS
temps	passé	<i>past</i>	PST
temps	futur	<i>future</i>	FUT

TAB. 6 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : indices de voix, d’aspect, de temps

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories indicielles (suite)			
<i>Catégories indicielles (principalement verbales)</i>			
mode	conditionnel	<i>conditional</i>	COND
mode	co-verbe	<i>converb</i>	CVB
mode	déclaratif	<i>declarative</i>	DECL
mode	impératif	<i>imperative</i>	IMP
mode	indicatif	<i>indicative</i>	IND
mode	infinitif	<i>infinitive</i>	INF
mode	irréel	<i>irrealis</i>	IRR
mode	prohibitif	<i>prohibitive</i>	PROH
mode	participe	<i>participle</i>	PTCP
mode	intentionnel	<i>purposive</i>	PURP
mode	citation	<i>quotative</i>	QUOT
mode	subjonctif	<i>subjunctive</i>	SBJV
mode	exhortatif	<i>(ex)hortative</i>	HORT
mode	interrogatif	<i>interrogative</i>	INT
mode	suppositif	<i>presumptive</i>	PRSM
mode	tendanciel	<i>tendential</i>	TEND
mode	volitif	<i>volitional</i>	VOL
mode	déontique	<i>deontic</i>	DEO
mode	assertif	<i>assertive</i>	ASRT
mode	intensif	<i>intensive</i>	INTS

TAB. 7 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : indices de mode / modalité

possessif appliqué aux noms aliénables). Par ailleurs, une recherche dans les tables des abréviations les plus courantes proposées dans des tentatives de normalisation *de facto* (comme celles mentionnées plus haut sous les entrées [1], [3], [5], et [6]) montre que l’abréviation généralement suggérée, pour ce type de morphème de cas nominal (qui a pour but de marquer la possession en s’affixant au terme possédé), est CONST (pour *forme construite*, terme qui vient à l’origine de la grammaire hébreue et arabe).

Dans la mesure où l’utilisation de termes plus standards n’écrase pas des nuances indispensables (ce qui reste, bien sûr, à l’appréciation du linguiste utilisateur de l’interface), il semble qu’il est préférable, à portée descriptive égale, de choisir le terme le plus normalisé ; cela tend à rendre la transcription plus facilement déchiffrable, en réduisant la fréquence des recours à des tables d’abréviations propres à chaque auteur. Le menu d’insertion d’abréviations morphosyntaxiques est proposé dans cet esprit.

En revanche, dès qu’un utilisateur détecte une lacune dans cette liste d’abréviations normalisées, il peut le signaler immédiatement au gestionnaire de l’interface informatique²⁰, afin que cette abréviation utile soit incorporée dans une version ultérieure du fichier de configuration. Dans l’intervalle, il peut taper l’abréviation dont il a besoin à partir du clavier.

20. À la date d’édition de ce document : Pascal Vaillant <vaillant@vjf.cnrs.fr>.

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories relationnelles			
actant verbal	actant unique	<i>single</i>	S
actant verbal	agent	<i>agent</i>	A
actant verbal	patient	<i>patient</i>	P
actant verbal	thème	<i>theme</i>	T
actant verbal	sujet	<i>subject</i>	SBJ
actant verbal	objet	<i>object</i>	OBJ
actant verbal	oblique	<i>oblique</i>	OBL
actant verbal	focus actant	<i>actant focus</i>	AF
actant verbal	focus patient	<i>patient focus</i>	PF
actant verbal	focus locatif	<i>locative focus</i>	LF
cas nominal	ablatif	<i>ablative</i>	ABL
cas nominal	absolutif	<i>absolutive</i>	ABS
cas nominal	accusatif	<i>accusative</i>	ACC
cas nominal	allatif	<i>allative</i>	ALL
cas nominal	bénéfactif	<i>benefactive</i>	BEN
cas nominal	comitatif	<i>comitative</i>	COM
cas nominal	datif	<i>dative</i>	DAT
cas nominal	distributif	<i>distributive</i>	DISTR
cas nominal	ergatif	<i>ergative</i>	ERG
cas nominal	génitif	<i>genitive</i>	GEN
cas nominal	instrumental	<i>instrumental</i>	INS
cas nominal	locatif	<i>locative</i>	LOC
cas nominal	nominatif	<i>nominative</i>	NOM
cas nominal	possessif	<i>possessive</i>	POSS
cas nominal	prédicatif	<i>predicative</i>	PRED
cas nominal	vocatif	<i>vocative</i>	VOC
cas nominal	latif	<i>lative</i>	LAT
cas nominal	associatif	<i>associative</i>	ASSOC
cas nominal	privatif	<i>privative</i>	PRV
cas nominal	construit	<i>construct</i>	CONST
cas nominal	médiatif	<i>mediative</i>	MEDT
cas nominal	mouvement	<i>movement</i>	MOV
cas nominal	directionnel	<i>directional</i>	DIR
cas nominal	essif	<i>essive</i>	ESS
place dans la structure informationnelle	focus	<i>focus</i>	FOC
place dans la structure informationnelle	topic	<i>topic</i>	TOP

TAB. 8 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : indices relationnels (actant, cas, rôle informationnel)

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Catégories constructionnelles			
cadre actanciel du verbe	intransitif	<i>intransitive</i>	INTR
cadre actanciel du verbe	transitif	<i>transitive</i>	TR
cadre actanciel du verbe	ditransitif	<i>ditransitive</i>	DITR
cadre actanciel du nom	aliénable	<i>alienable</i>	AL
cadre actanciel du nom	inaliénable	<i>inalienable</i>	INAL
type prédicat / compatibilités aspectuelles	statif	<i>stative</i>	STAT
type prédicat / compatibilités aspectuelles	processif	<i>processive</i>	PROC

TAB. 9 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : indices constructionnels (déterminant la structure du syntagme) : cadres actanciels, types inhérents

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Partie du discours et morphèmes divers			
partie du discours	nom	<i>noun</i>	N
partie du discours	nom propre	<i>proper noun</i>	N;PROPR
partie du discours	verbe	<i>verb</i>	V
partie du discours	adjectif	<i>adjective</i>	ADJ
partie du discours	adverbe	<i>adverb</i>	ADV
partie du discours	article	<i>article</i>	ART
partie du discours	auxiliaire	<i>auxiliary</i>	AUX
partie du discours	copule	<i>copula</i>	COP
partie du discours	déictique	<i>deictic</i>	DEIC
partie du discours	déterminant	<i>determiner</i>	DET
partie du discours	déterminant défini	<i>definite determiner</i>	DET;DEF
partie du discours	déterminant indéfini	<i>indefinite determiner</i>	DET;INDF
partie du discours	déterminant spécifique	<i>specific determiner</i>	DET;SPEC
partie du discours	déterminant générique	<i>generic determiner</i>	DET;GENR
partie du discours	pronom	<i>pronoun</i>	PRN
partie du discours	pronom réciproque	<i>reciprocal pronoun</i>	RECP
partie du discours	pronom réflexif	<i>reflexive pronoun</i>	REFL
partie du discours	pronom relatif	<i>relative pronoun</i>	REL
partie du discours	interrogatif (qu-)	<i>interrogative (wh-)</i>	INT
partie du discours	exclamation	<i>exclamation</i>	EXCL
partie du discours	interjection	<i>interjection</i>	INTJ
partie du discours	adposition	<i>adposition</i>	ADP
partie du discours	preposition	<i>preposition</i>	PREP
partie du discours	postposition	<i>postposition</i>	POSTP

TAB. 10 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : parties du discours, morphèmes indépendants

Attribut	Valeur (fr)	Valeur (en)	Abréviation
Autres morphèmes indépendants			
morphème indépendant	classificateur	<i>classifier</i>	CLF
morphème indépendant	quantificateur	<i>quantifier</i>	QUANT
morphème indépendant	nombre cardinal	<i>cardinal number</i>	CARD
morphème indépendant	nombre ordinal	<i>ordinal number</i>	ORD
morphème indépendant	complémenteur	<i>complementizer</i>	COMP
morphème indépendant	négation	<i>negation</i>	NEG
morphème indépendant	question particule	<i>marqueur interrogatif</i>	Q
morphème indépendant	subordonnant	<i>subordinator</i>	SUBR
morphème indépendant	joncteur	<i>linker</i>	LNK
morphème indépendant	coordonateur	<i>co-ordinator</i>	CO
morphème indépendant	particule	<i>particle</i>	PRT
morphème indépendant	particule de temps	<i>tense particle</i>	PRT.TE
morphème indépendant	particule de aspect	<i>aspect particle</i>	PRT.AS
morphème indépendant	particule de mode	<i>mood particle</i>	PRT.MO
morphème indépendant	particule énonciative	<i>enunciation particle</i>	PRT.EN
morphème flexionnel	accord	<i>agreement</i>	AGR
morphème dérivationnel	nominalisateur	<i>nominalizer</i>	NR
morphème dérivationnel	verbalisateur	<i>verbalizer</i>	VR
morphème dérivationnel	augmentative	<i>augmentatif</i>	AUG
morphème dérivationnel	diminutif	<i>diminutive</i>	DIM

TAB. 11 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : parties du discours, morphèmes indépendants

Attribut				Valeur (fr)	Valeur (en)	Abréviation
Partie du discours et morphèmes divers						
Symboles conventionnels						
symbole unique				signifiant vide	<i>null expression element</i>	Ø
symbole unique séparant deux éléments				frontière de morpheme	<i>morpheme boundary</i>	-
symbole unique séparant deux éléments				frontière de clitique	<i>clitic boundary</i>	=
symbole unique séparant deux éléments				combinaison d'éléments du métalangage	<i>metalanguage elements combination</i>	.
symbole unique séparant deux éléments				combinaison d'éléments du métalangage pour un morphème non-segmentable dans la langue d'origine	<i>metalanguage elements combination for a non-segmentable morpheme of the original language</i>	—
symbole unique séparant deux éléments				combinaison d'éléments du métalangage pour des éléments sémantiques ou fonctionnels amalgamés dans la langue d'origine	<i>metalanguage elements combination for semantic or functional elements which are fused in the original language</i>	;
symbole unique séparant deux éléments				combinaison d'éléments du métalangage pour des éléments sémantiques ou fonctionnels dont la frontière n'est pas montrée dans la langue d'origine	<i>metalanguage elements combination for semantic or functional elements which are not separately displayed in the original language</i>	:
symbole unique séparant deux éléments				combinaison d'éléments du métalangage se manifestant par une altération morphophonologique dans la langue d'origine	<i>metalanguage elements combination expressed by a morphophonological alteration in the original language</i>	\
symbole unique séparant deux éléments				réduplication	<i>reduplication</i>	~
paire de symboles encadrant un élément				indication de morphème(s) non-manifesté(s) dans la langue d'origine	<i>indication of a non-overt element in the original language</i>	
paire de symboles encadrant un élément				indication de catégorie inhérente	<i>indication of an inherent category</i>	()
paire de symboles encadrant un élément				infixe	<i>infix</i>	<>
paire de symboles encadrant un élément				circonfixe	<i>circumfix</i>	><

TAB. 12 – Abréviations disponibles dans les menus pour la traduction juxtalinéaire : symboles conventionnels

4.1.4 Caractères spéciaux

Le menu “Caractères”, en haut de la fenêtre d’interface principale de JAXE, contient quelques commandes d’insertion de caractères spéciaux difficiles à atteindre directement au clavier, sur un clavier français (AZERTY) ²¹.

Ceci concerne notamment l’espace insécable (également disponible avec le raccourci clavier “Ctrl-Espace”).

Pour la voyelle « i » étirée fermée d’arrière du kali’na, il existe un élément de menu qui insère le caractère « i-barré ». Celui-ci est également accessible avec le raccourci clavier “Ctrl-i”.

Les autres fonctions de ce menu sont faites pour insérer les caractères accentués difficiles d’accès sur le clavier français et utiles pour la transcription du nenge (ex. a accent aigü), du portugais (a et o til), ou d’autres langues.

Pour les autres linguistes, et pour rendre le programme plus générique, il est prévu dans une version ultérieure l’ajout d’un menu qui permette d’insérer n’importe quel caractère IPA.

4.2 Validation de la conformité d’un corpus

JAXE propose une fonction “Validation” (dans le menu “Fenêtres”) qui permet de vérifier qu’un document de type CORPUS-Contacts est bien valide (c’est-à-dire : conforme à la structure idéale décrite par le schéma de documents XML ²²).

Cette fonction permet notamment de s’assurer que les références croisées (référence d’un énoncé à un autre énoncé en cas de chevauchement, référence à une langue ou à un locuteur devant être catalogué dans l’en-tête du corpus ...) sont bien conformes : c’est-à-dire, d’une part, qu’aucune référence ne pointe « dans le vide » (vers un identifiant qui n’existe pas) ; d’autre part, qu’il n’y a pas deux énoncés qui partagent le même identifiant.

Cette fonction peut être utilisée pour ne pas avoir à se préoccuper à l’avance d’avoir un inventaire exhaustif de toutes les langues et de tous les locuteurs du corpus : il suffit de taper son corpus « au kilomètre », puis de vérifier à la fin (ou à intervalles réguliers) que tous les éléments sont bien référencés en utilisant la fonction “Validation” (s’il manque une référence, celle-ci dit en effet exactement de quelle référence il s’agit, et où elle doit se situer). On peut ainsi constituer *a posteriori* un inventaire des langues et/ou des locuteurs du corpus.

21. Cette modification fait partie des quelques « bricolages » que j’ai dû faire sur le code source de JAXE pour l’adapter à certains besoins particuliers.

22. Schéma de documents *CORPUS-Contacts* : http://clapoty.vjf.cnrs.fr/contacts/doc/corpus_xsd.pdf.