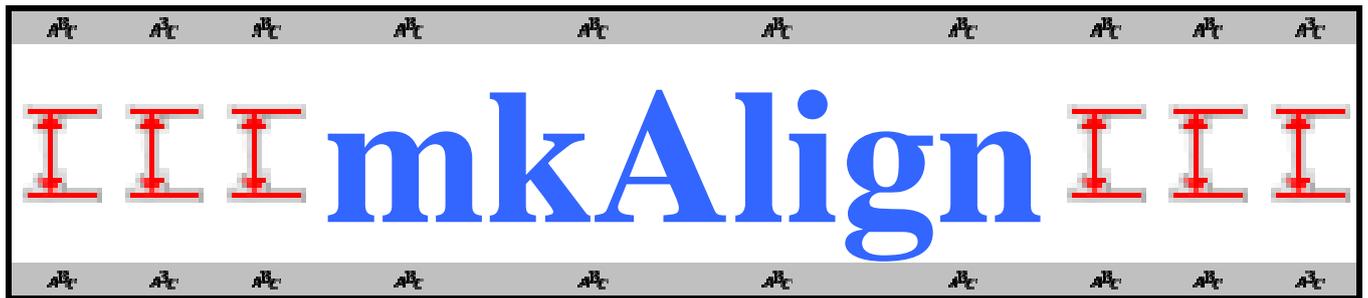


Centre de textométrie - SYLED / CLA²T
[U. Paris 3 Sorbonne nouvelle]
<http://syled.univ-paris3.fr/cla2t.html>



Serge Fleury

SYLED/CLA²T

Université Sorbonne Nouvelle Paris 3

URL : <http://www.tal.univ-paris3.fr/mkAlign/>
Téléchargement : <http://www.tal.univ-paris3.fr/mkAlign/>
Contact : serge.fleury@univ-paris3.fr

Versions exécutables sous Windows disponibles en ligne
Documentation en ligne sur la page du projet

Manuel d'utilisation

mise à jour : juin 2012
version 2.0 (b144)

Sommaire

1	Figures	6
2	Préambule.....	9
3	Présentation générale.....	10
4	Interface de mkAlign.....	11
5	Descriptif des icônes	13
6	Chargement des fichiers	14
6.1	Mode général.....	15
6.1.1	Chargement des textes à aligner.....	15
6.1.2	Choix du segmenteur.....	17
6.1.3	Prétraitement des textes à aligner.....	17
6.1.4	Découpage des textes en parties.....	18
6.1.5	Paramètres lexicométriques des textes chargés.....	18
6.2	Mode « alignement par recherche de cognats ». Alignement automatique de fichiers.....	22
6.3	Mode «Import d'alignement au format TMX »	24
7	Principales fonctionnalités de mkAlign pour l'édition de l'alignement	26
7.1	Recherche de chaînes	26
7.2	Exporter sous-corpus contenant un motif.....	27
7.3	Figement de cellule (lecture/écriture)	27
7.4	Fractionnement de cellule	27
7.5	Fusion de cellule.....	27
7.6	Etat de l'alignement	28
7.7	Remarques sur le segmenteur.....	28
7.8	Les modes.....	29
7.9	Fonctionnalité complémentaire pour le MODE SPLIT : lecture LR/RL	29
8	Représentation cartographique de l'alignement.....	30
8.1	Construction de la carte de l'alignement.....	30
8.2	Recherche de motifs dans la carte de l'alignement	31
8.3	Importation/exportation du vecteur des positions d'une sélection.....	33
8.4	Affichage d'une partition dans la carte de l'alignement	34
8.5	Navigation vers l'alignement	34
8.6	Le vocabulaire spécifique d'une section de la carte.....	35
8.7	Le vocabulaire spécifique d'une sélection de sections de la carte	36
8.8	Fonctionnalités complémentaires sur la carte des sections	37
8.9	Le vocabulaire spécifique des sections de la carte contenant une forme ou un motif : recherche de cooccurrents	38
8.10	Retour aux contextes à partir du vocabulaire spécifique.....	39
8.11	Filtrage de la « zone miroir » au cours d'une recherche (« au-delà du miroir »).....	40
9	Concordances	44
10	Importation de listes	45
11	Segments répétés	46
12	Cooccurrences – Poly-Cooccurrences.....	48
13	Graphiques : ventilation / accroissement de vocabulaire	50
13.1	Ventilation de formes graphiques	50
13.2	Courbe d'accroissement de vocabulaire.....	52
14	Exportation des traitements	53
14.1	Export vers Lexico3	53
14.2	Export au format XML.....	54

14.3	Export au format XML pour sauvegarder une session de travail.....	55
14.4	Export complet au format HTML	55
14.5	Export partiel au format HTML	56
14.6	Export de bi-textes au format TXT	58
14.7	Export de l'alignement au format TMX.....	59
15	Import d'alignement	61
16	Mise au jour de la variation dans les textes.....	63
16.1	Repérage de la variation	63
16.2	Projections lexicométriques	65
17	Le rapport : enregistrement des résultats produits	66
17.1	Ajouter un élément au rapport.....	66
17.2	Ajouter un élément externe au rapport.....	68
17.3	Enregistrer le rapport.....	69
18	Références	70
19	Annexes.....	71
19.1	Les expressions régulières.....	71
19.2	Fonctionnalités de l'éditeur d'alignement.....	72

1 Figures

Figure 1 : <i>mkAlign</i> Fenêtre principale	11
Figure 2 : <i>mkAlign</i> Onglet PARAM	12
Figure 3 : <i>mkAlign</i> Onglet ALIGN, chargement des fichiers	14
Figure 4 : <i>mkAlign</i> Onglet ALIGN, fichiers chargés	16
Figure 5 : <i>mkAlign</i> Onglet PARAM, paramétrage de l'encodage	16
Figure 6 : <i>mkAlign</i> Onglet ALIGN, fichiers chargés après sélection d'encodage	17
Figure 7 : <i>mkAlign</i> choix du <i>segmenteur</i> onglet ALIGN ou onglet PARAM	17
Figure 8 : <i>mkAlign</i> pré-formatage de la SOURCE et de la CIBLE	18
Figure 9(1) : <i>mkAlign</i> Onglet DIC, chargement	19
Figure 9(2) : <i>mkAlign</i> Onglet DIC, projection sur la carte d'une sélection de formes	19
Figure 9(3) : <i>mkAlign</i> Onglet DIC, concordance d'une sélection de formes	20
Figure 9(4) : <i>mkAlign</i> Onglet DIC, recherche de formes	20
Figure 10 : <i>mkAlign</i> Onglet RAPPORT	21
Figure 11 (1) : <i>mkAlign</i> Alignement par recherche de cognats	22
Figure 11 (2) : <i>mkAlign</i> Alignement par recherche de cognats	23
Figure 12 (1) : <i>mkAlign</i> Alignement par importation d'un fichier au format TMX (choix du fichier)	24
Figure 12 (2) : <i>mkAlign</i> Alignement par importation d'un fichier au format TMX (choix de 2 langues)	25
Figure 13 : <i>mkAlign</i> Onglet ALIGN, recherche de chaînes	26
Figure 14 : <i>mkAlign</i> Alignement en cours	28
Figure 15 : <i>mkAlign</i> Onglet MAP	30
Figure 16 : <i>mkAlign</i> Recherche dans la carte des sections	31
Figure 17 : <i>mkAlign</i> Recherche et Navigation dans la carte des sections	32
Figure 18 : <i>mkAlign</i> Carte / Sections / Parties	34
Figure 19 : <i>mkAlign</i> Carte / Sections / Spécificités	35
Figure 20 : <i>mkAlign</i> Carte / Sections / Sélection	36
Figure 21 (1) : <i>mkAlign</i> Carte / Sections / Spécificité sections	36
Figure 21 (2) : <i>mkAlign</i> Carte / Sections / Spécificité sections	37
Figure 22 (1) : <i>mkAlign</i> Carte / Sections / Spécificités / Cooccurents	38
Figure 22 (2) : <i>mkAlign</i> Carte / Sections / Spécificités / Cooccurents	39
Figure 23 (1) : <i>mkAlign</i> « Au-delà du miroir »	40
Figure 23 (2) : <i>mkAlign</i> « Au-delà du miroir »	41
Figure 23 (3) : <i>mkAlign</i> « Au-delà du miroir »	42
Figure 23 (4) : <i>mkAlign</i> « Au-delà du miroir »	42
Figure 24 (1) : <i>mkAlign</i> Onglet CONCORDANCE, édition d'une concordance	44
Figure 24 (2) : <i>mkAlign</i> exportation bi-concordance	44
Figure 25 : <i>mkAlign</i> Onglet LISTES, importation de listes	45
Figure 26 : <i>mkAlign</i> Onglet LISTES, édition de listes importées	45
Figure 27 (1) : <i>mkAlign</i> Onglet SEGMENTS	46
Figure 27 (2) : <i>mkAlign</i> Onglet PARAM, paramétrage du calcul des segments répétés	46
Figure 27 (3) : <i>mkAlign</i> Onglet SEGMENTS, sélection de segments	47
Figure 27 (4) : <i>mkAlign</i> Onglet SEGMENTS, sélection de segments	47
Figure 28 : <i>mkAlign</i> Onglet COOCS, calcul de cooccurents et de poly-cooccurents	49
Figure 29 (1) : <i>mkAlign</i> Onglet DIC, ventilation de formes	50
Figure 29 (2) : <i>mkAlign</i> Onglet DIC, sélection de formes pour ventilation	51
Figure 29 (3) : <i>mkAlign</i> Onglet GRAPHE, ventilation des formes sélectionnées	51
Figure 30 : <i>mkAlign</i> Onglet GRAPHE, courbe d'accroissement du vocabulaire	52

Figure 31 : <i>mkAlign</i> Onglet EXPORT-L3.....	53
Figure 32 : <i>mkAlign</i> Onglet EXPORT-XML	54
Figure 33 : <i>mkAlign</i> Schéma Export XML.....	54
Figure 34 : <i>mkAlign</i> Export HTML.....	55
Figure 35 : <i>mkAlign</i> Export sélectif (sélection d'une forme).....	56
Figure 36 : <i>mkAlign</i> Résultat d'export sélectif - Fichier complet : example-export.html	57
Figure 37 : <i>mkAlign</i> , Export de bi-textes.....	58
Figure 38 : <i>mkAlign</i> , Export de l'alignement au format TMX.....	59
Figure 39 : <i>mkAlign</i> , Ouverture de l'alignement au format TMX dans Word avec feuille de styles.....	60
Figure 40 : <i>mkAlign</i> Onglet EXPORT.....	61
Figure 41 : <i>mkAlign</i> Schéma Import	62
Figure 42 (1) : <i>mkAlign</i> Onglet Variation (paramètres et outils)	63
Figure 42 (2) : <i>mkAlign</i> Repérage de la variation (alignement initial).....	63
Figure 42 (3) : <i>mkAlign</i> Mise au jour de la variation	64
Figure 43 : <i>mkAlign</i> Ajouter un élément au rapport.....	66
Figure 44 : <i>mkAlign</i> Onglet EXPORT : Rapport en cours	67
Figure 45 : <i>mkAlign</i> Insertion d'image externe dans le rapport	68
Figure 46 : Index du rapport enregistré	69

2 Préambule

La notion de *corpus parallèle*, qui émerge actuellement dans les travaux de différents chercheurs comme *corpus comportant plusieurs volets qui correspondent chacun à une version d'un même texte dans deux ou plusieurs langues différentes*, renvoie à des situations connues de coexistence de textes présentant des liens forts dans leur structuration.

Le traitement de corpus parallèles suppose une phase préalable d'*alignement*¹, c'est-à-dire de mise en correspondance dans chacun des volets de différents types d'unités textuelles [Zimina, 2004].

Aligner des corpus de textes originaux et de leurs traductions c'est mettre en relation des unités textuelles qui se correspondent. On peut établir des correspondances entre des unités de différents niveaux : mots, syntagmes, phrases, paragraphes, sections, etc.

¹ L'alignement (multilingue) [Véronis, 2000] part de deux textes qui sont en rapport de traduction. Il consiste à établir des correspondances de plus en plus fines (entre les parties du texte : alignement structurel ; entre les phrases : alignement phrastique ; entre les mots : alignement lexical)

3 Présentation générale

Le programme **mkAlign** permet de construire, corriger et visualiser un alignement de deux textes *via* un éditeur à double entrée. Il permet d'afficher simultanément les textes *source* et *cible* pour y rajouter ou corriger des segments équivalents.

Ce programme n'est pas (seulement) un aligneur automatique.

Il est conçu pour aider l'utilisateur dans la création, l'alignement, la correction et la validation de textes traduits. L'utilisateur garde la maîtrise sur l'ensemble des processus, depuis la mise en correspondance initiale des segments équivalents jusqu'à l'export final du bi-texte produit.

Il est aussi conçu pour mener des *calculs lexicométriques*² sur les contenus textuels chargés.

Il appartient à l'utilisateur de construire l'alignement et de définir son degré de précision (*résolution*). Cette résolution peut varier pour mettre en évidence les correspondances entre les segments textuels des différents niveaux.

La notion de sauvegarde de session de travail (création de fichiers d'*export/import* de bi-textes au format *xml* et *html*) permet de commencer le travail sur un corpus à deux volets textuels, l'exporter au format désiré, puis le réimporter plus tard pour y apporter des modifications.

La visualisation de l'alignement dans une *représentation cartographique (bi-text map)* offre plusieurs possibilités de gestion de corpus qui partagent des similitudes au plan traductionnel³.

² <http://tal.univ-paris3.fr/wakka/wakka.php?wiki=Glossaire>

³ Par exemple, la recherche de chaînes de caractères présentées sous forme d'*expressions régulières* à l'aide de la carte met en valeur les cas de parallélisme (ou les ruptures de parallélisme) dans les distributions lexicales au sein du bi-texte [Zimina, 2005].

4 Interface de mkAlign

L'interface du programme est composée d'une fenêtre graphique disposant de différents onglets.

L'onglet **HOME** visible au chargement contient un mode d'emploi du programme.

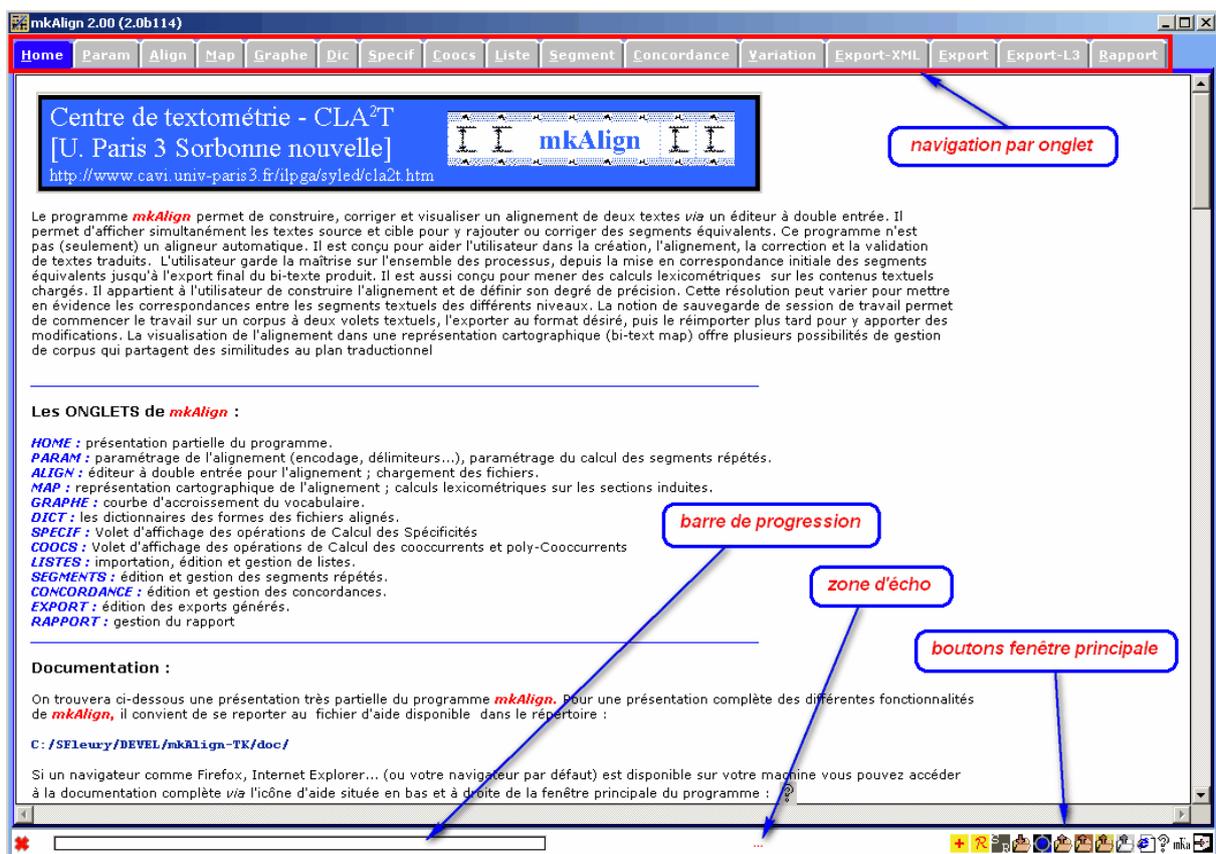


Figure 1 : **mkAlign** Fenêtre principale

L'onglet **PARAM** permet de modifier le paramétrage de certaines fonctionnalités du programme (taille des polices d'affichage des textes, couleurs de cellules d'édition, encodage des fichiers *source* et *cible* etc.).

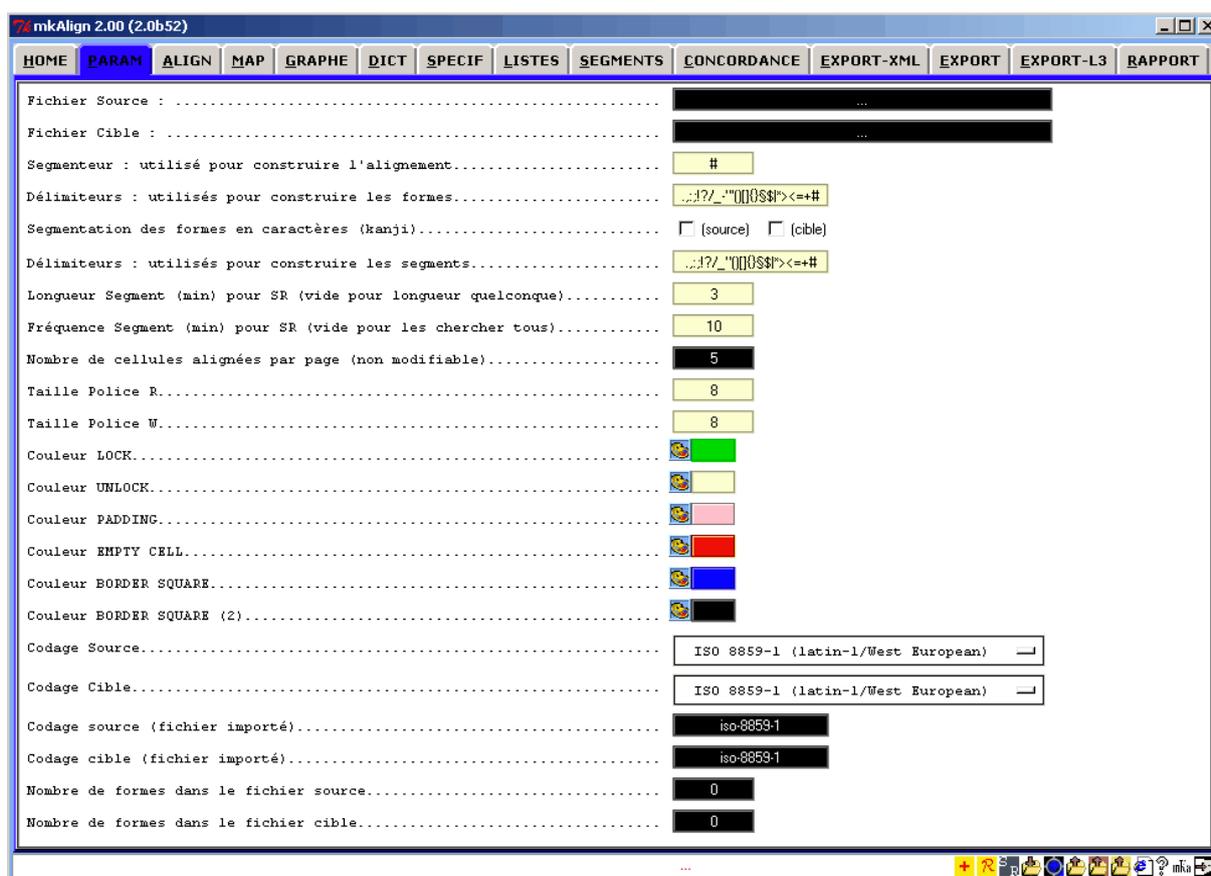


Figure 2 : *mkAlign* Onglet **PARAM**

Les autres onglets sont décrits *infra*.

5 Descriptif des icônes

Icône	Fonction	Localisation
	Chargement du fichier source	Onglet ALIGN
	Chargement du fichier cible	Onglets ALIGN
	Alignement du fichier cible et du fichier source	Onglets ALIGN
	Sauvegarde (fichier source, cible, export, carte)	Onglets ALIGN, EXPORT, MAP
	Rafraîchissement éditeur/carte	Onglets ALIGN, MAP
	MODE MERGE	Onglet ALIGN
	MODE SPLIT	Onglet ALIGN
	Dessin de la carte des sections	Onglet MAP
	Export au format HTML	Onglets EXPORT, CONCORDANCE
	Parser XML	Onglet ALIGN
	Import d'alignement	Fenêtre principale
	Export au format XML	Fenêtre principale
	Export de bitexte	Fenêtre principale
	Export au format Lexico3	Fenêtre principale
	Export au format HTML	Fenêtre principale
	Documentation du programme	Fenêtre principale
	Sortie du programme	Fenêtre principale
	Palette	Onglet PARAM
	Page suivante	Onglet ALIGN
	Dernière page	Onglet ALIGN
	Première page	Onglet ALIGN
	Page précédente	Onglet ALIGN
	Editeur annexe	Fenêtre principale
	Export "Recherche Source"	Onglet MAP
	Export "Recherche Cible"	Onglet MAP
	Export Intersection "Recherche Source/Cible"	Onglet MAP
	Export Vecteur Recherche Source	Onglet MAP
	Export Vecteur Recherche Cible	Onglet MAP
	Import Vecteur Recherche Source	Onglet MAP
	Import Vecteur Recherche Cible	Onglet MAP
	Import de liste (Source)	Onglet MAP
	Import de liste (Cible)	Onglet MAP
	Calcul des segments répétés dans les fichiers SOURCE et CIBLE	Fenêtre principale
	Recherche des cooccurrents d'une forme dans la source	Onglet MAP
	Recherche des cooccurrents d'une forme dans la cible	Onglet MAP
	Calcul des spécificités dans une sélection de sections dans la source	Onglet MAP
	Calcul des spécificités dans une sélection de sections dans la cible	Onglet MAP
	Sélection du type de recherche dans la <i>zone miroir</i>	Onglet MAP
	Ventilation Source/Cible : présence/absence/intersection	Onglet MAP
	Ajout au rapport / Enregistrement du rapport	Fenêtre principale
	Calcul des cooccurrents et des poly-cooccurrents	Onglet COOCS

6 Chargement des fichiers

Trois modes de chargement des textes de travail sont disponibles :

- **Mode général** : Sélection d'un *segmenteur*. Chargement d'un fichier **SOURCE** puis d'un fichier **CIBLE**.
- **Mode « alignement par recherche de cognats**». Alignement automatique du fichier **SOURCE** et du fichier **CIBLE** basé sur la recherche de cognats⁴.
- **Mode « Import d'alignement au format TMX »**. Chargement d'un alignement enregistré au format TMX⁵.

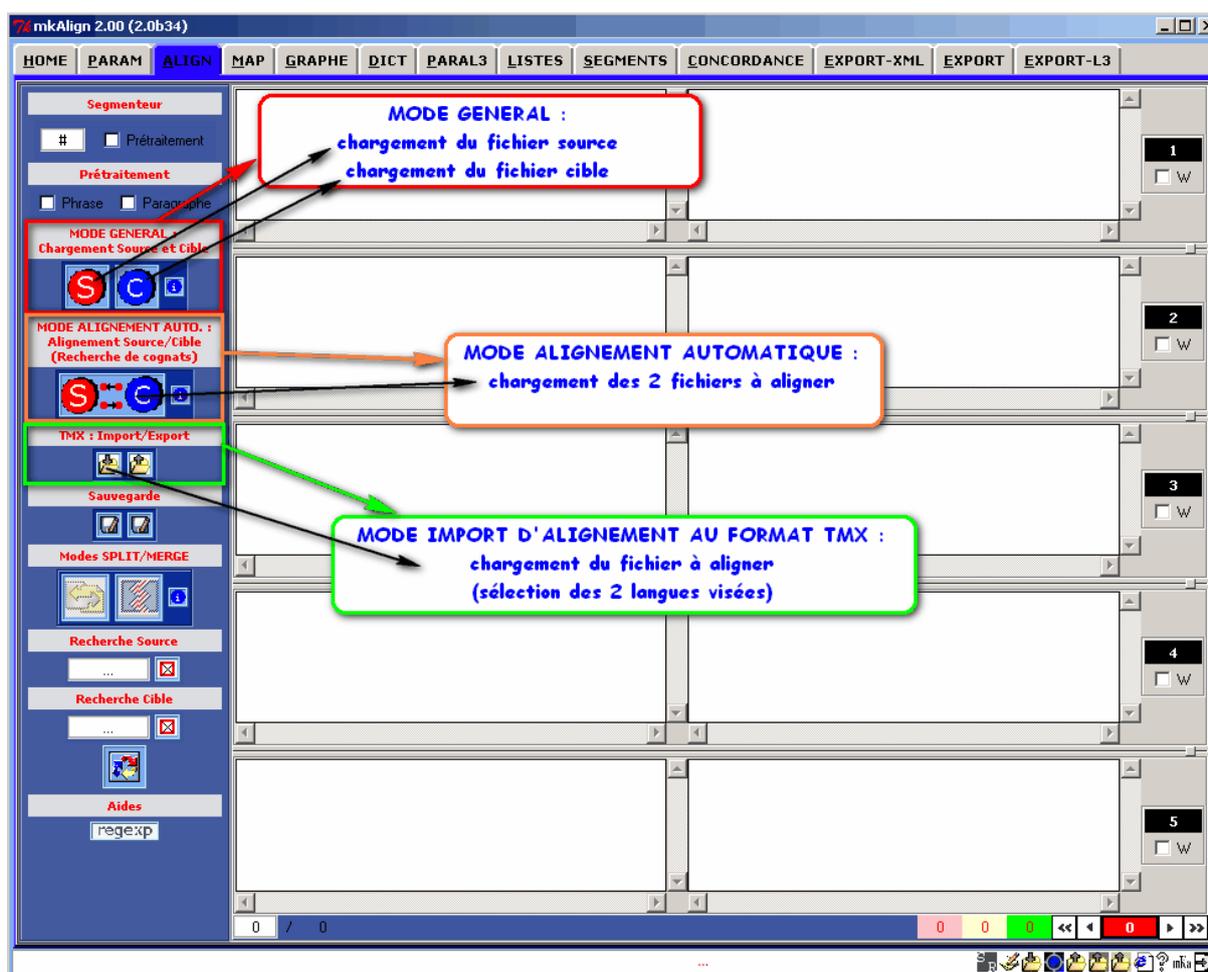


Figure 3 : *mkAlign* Onglet **ALIGN**, chargement des fichiers

⁴ Cognat est l'anglicisme couramment employé pour parler des congénères : à savoir des mots qui dérivent d'une même souche et présentent par conséquent des affinités formelles (graphiques, morphologiques, phonétiques et éventuellement sémantiques)

⁵ <http://www.lisa.org/Translation-Memory-e.34.0.html>

6.1 Mode général

6.1.1 Chargement des textes à aligner

L'onglet **ALIGN** contient les fenêtres d'édition pour l'alignement et les points d'entrée pour les différentes fonctionnalités associées.

La construction d'un alignement utilise en entrée deux volets d'un même corpus (**SOURCE** et **CIBLE**) : le texte **SOURCE** sera chargé dans la partie gauche de **mkAlign** et le texte **CIBLE** dans la partie droite.

- Pour charger le texte **SOURCE**: activez le bouton  (onglet **ALIGN**)
- Pour charger le texte **CIBLE**: activez le bouton  (onglet **ALIGN**)

Le chargement peut être réalisé en tenant compte d'un caractère délimiteur (appelé le *segmenteur*, par défaut le caractère #) qui sera utilisé pour aligner les 2 volets du corpus. Si le *segmenteur* n'est pas présent dans les 2 volets initiaux, les 2 volets seront alignés globalement affichés intégralement dans les premières cellules **SOURCE** et **CIBLE**.

Par exemple, si le caractère *segmenteur* choisi est #, les fichiers en entrée peuvent avoir l'allure suivante :

Fichier Source :

```
ssssssssssssssss #
ssssssssssssssss #
etc.
```

Fichier Cible :

```
cccccccccccccccc #
cccccccccccccccc #
etc.
```

Dans cet exemple, le fichier **SOURCE** sera chargé dans la partie gauche de **mkAlign** avec au moins 2 cellules correspondant au découpage sur la base du *segmenteur* sélectionné. Idem pour le fichier **CIBLE**, dans la partie droite. Les 2 volets du corpus aligné sont présentés par page de 5 blocs alignés, on peut passer d'une page à l'autre de l'alignement *via* les boutons présents au bas de l'onglet **ALIGN**, ou en sélectionnant une page donnée (puis touche Entrée).

Après chargement de 2 fichiers, l'onglet **ALIGN** a l'allure suivante :

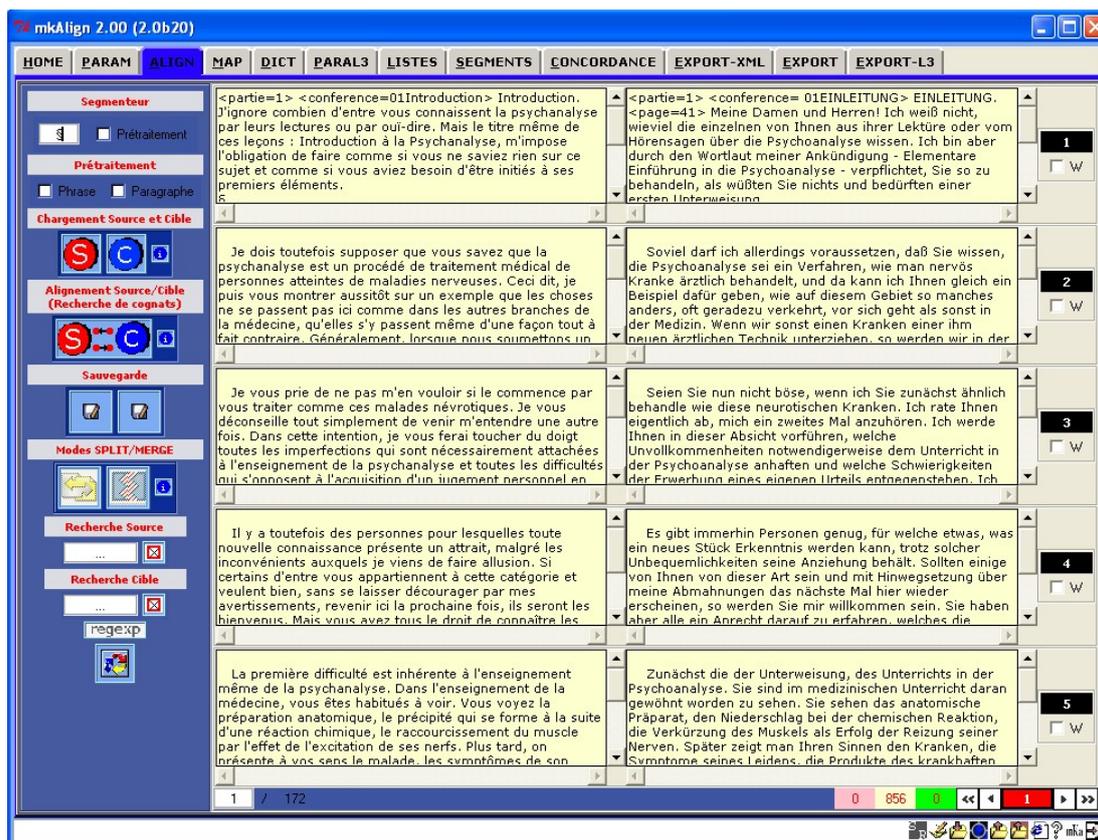


Figure 4 : *mkAlign* Onglet **ALIGN**, fichiers chargés

Dans les figures qui suivent, on présente le chargement des fichiers après sélection des paramètres d'encodage : iso-8859-1 pour le texte **SOURCE** et utf-8 pour le texte **CIBLE**.

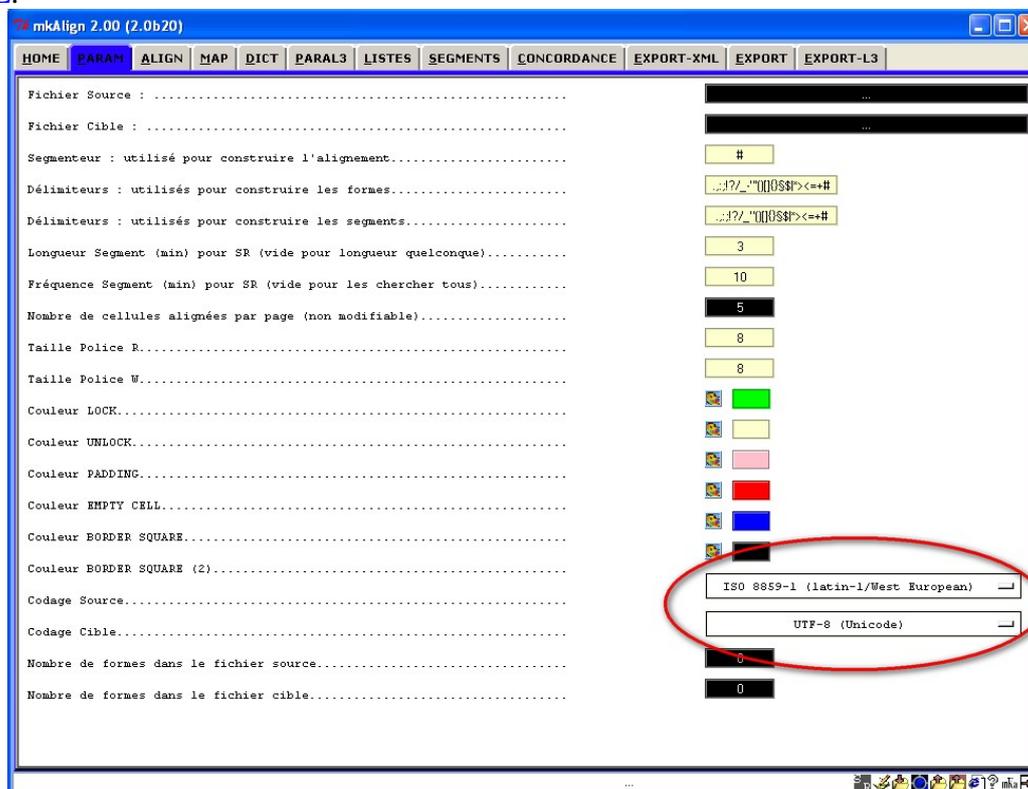


Figure 5 : *mkAlign* Onglet **PARAM**, paramétrage de l'encodage

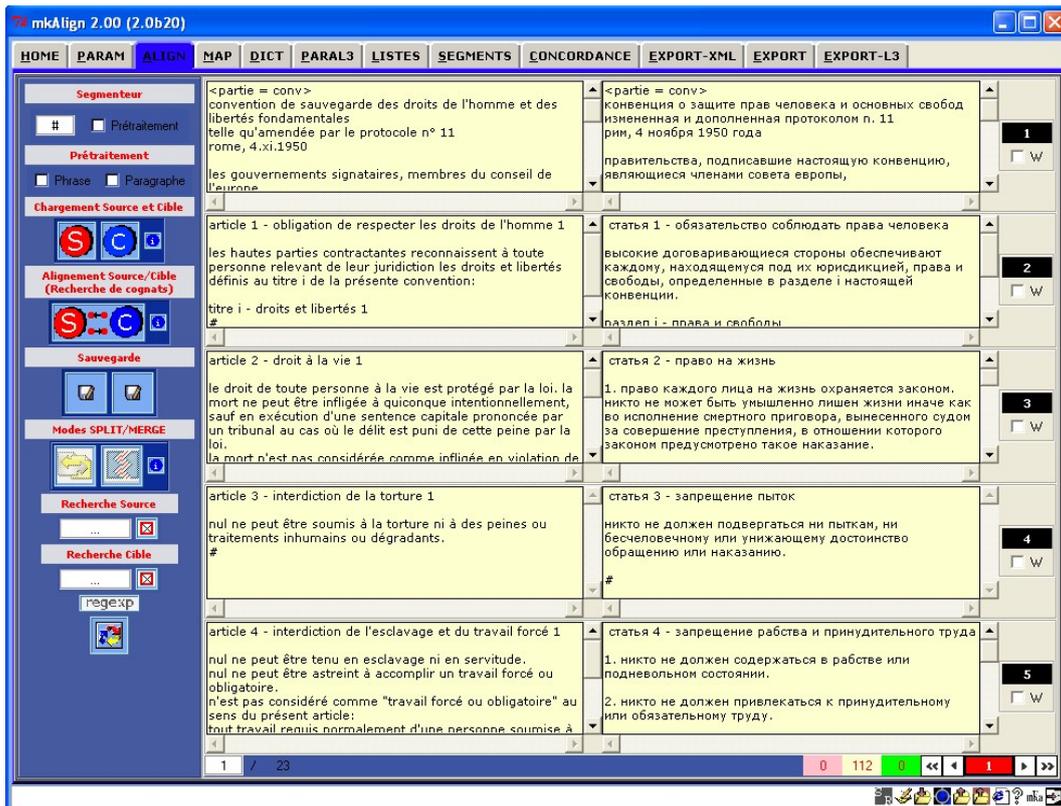


Figure 6 : *mkAlign* Onglet **ALIGN**, fichiers chargés après sélection d'encodage

6.1.2 Choix du segmenteur

Il est possible de paramétrer le *segmenteur* avant de charger le corpus : il suffit pour cela de changer la valeur dans la zone de saisie idoine (onglet **ALIGN** ou onglet **PARAM**)

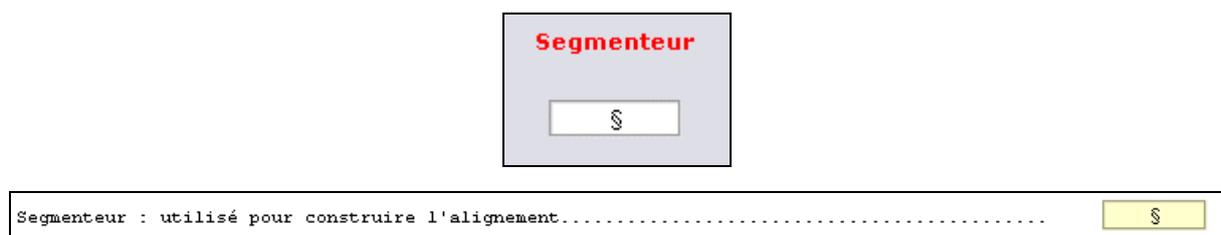


Figure 7 : *mkAlign* choix du *segmenteur* onglet **ALIGN** ou onglet **PARAM**

Si la zone de saisie permettant de définir le *segmenteur* est vide, c'est le caractère *retour chariot* qui sera défini comme étant le *segmenteur*.

Remarque : il est aussi possible de forcer la segmentation des textes à aligner en caractères en cochant dans l'onglet **PARAM** la case idoine associée au fichier visé.

6.1.3 Prétraitement des textes à aligner

Par défaut, le chargement du texte **SOURCE** et du texte **CIBLE** se réalise en ne tenant compte que du *segmenteur* choisi et n'opère aucune modification sur les 2 volets à charger.

Il est aussi possible de prétraiter les 2 textes à charger et d'attribuer ainsi une valeur prédéfinie au *segmenteur*. Dans l'onglet **ALIGN**, l'activation de la case à cocher « *Prétraitement* » permet de pré-formater les 2 textes à charger en phrases ou en paragraphes.

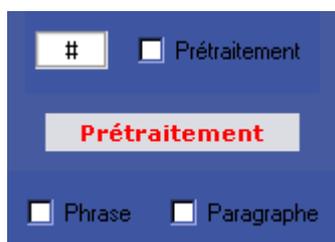


Figure 8 : *mkAlign* pré-formatage de la SOURCE et de la CIBLE

Le découpage en paragraphes / phrases a pour objectif la création automatique d'un premier alignement très approximatif. Ce « brouillon » du bi-texte aligné devrait ensuite être retravaillé par l'utilisateur à l'aide des fonctionnalités de *mkAlign* (voir la section §6 « [Principales fonctionnalités de mkAlign](#) »).

Le découpage automatique en phrases suit les règles typographiques élémentaires :

- Une phrase est définie grossièrement comment une chaîne de caractères se terminant par les caractères suivants : point (.), 3 points (...), point d'interrogation (?) et point d'exclamation (!).
- Un paragraphe étant défini grossièrement aussi comme une suite de phrase terminée par un retour à la ligne.

Si l'option « *Prétraitement* » est choisie, il reste à déterminer le type de pré-formatage à réaliser. Par défaut, l'activation du prétraitement active le pré-formatage en paragraphes ; pour activer le pré-formatage en phrases, il suffit de cocher la case idoine.

6.1.4 Découpage des textes en parties

Le chargement des 2 fichiers alignés tient compte d'un marquage dans les textes d'une éventuelle **partition**, ce marquage est détecté s'il est réalisé de la manière suivante :

```
<nomdelapartie="valeur">
```

Ce marquage signifie que l'utilisateur a prédéfini un marquage de partie *via* une série de balises dans laquelle la partie est désignée par un type, ici par "nomdelapartie", et qu'elle est associée à chaque fois à une valeur donnée. Un fichier en entrée pourra donc avoir l'allure suivante (2 parties définies ici) :

```
<para="xxx">ssssssssssssss #
<part="yyy">ssssssssssssss #
<para="zzz">ssssssssssssss #
<part="uuu">ssssssssssssss #
etc.
```

6.1.5 Paramètres lexicométriques des textes chargés

Le chargement des fichiers **SOURCE** et **CIBLE** déclenche automatiquement un module de segmentation générant un dictionnaire des formes graphiques du fichier chargé. Ce programme de segmentation est paramétrable dans l'onglet **PARAM** : une liste des délimiteurs utilisés pour cette segmentation est donnée par défaut, l'utilisateur peut la modifier :

Délimiteurs : utilisés pour construire les formes..... :!?'_'"000\$!*"><=+

Le résultat de cette segmentation est visible dans l'onglet **DIC** et dans l'onglet **RAPPORT** :

- le dictionnaire sera visible dans le premier

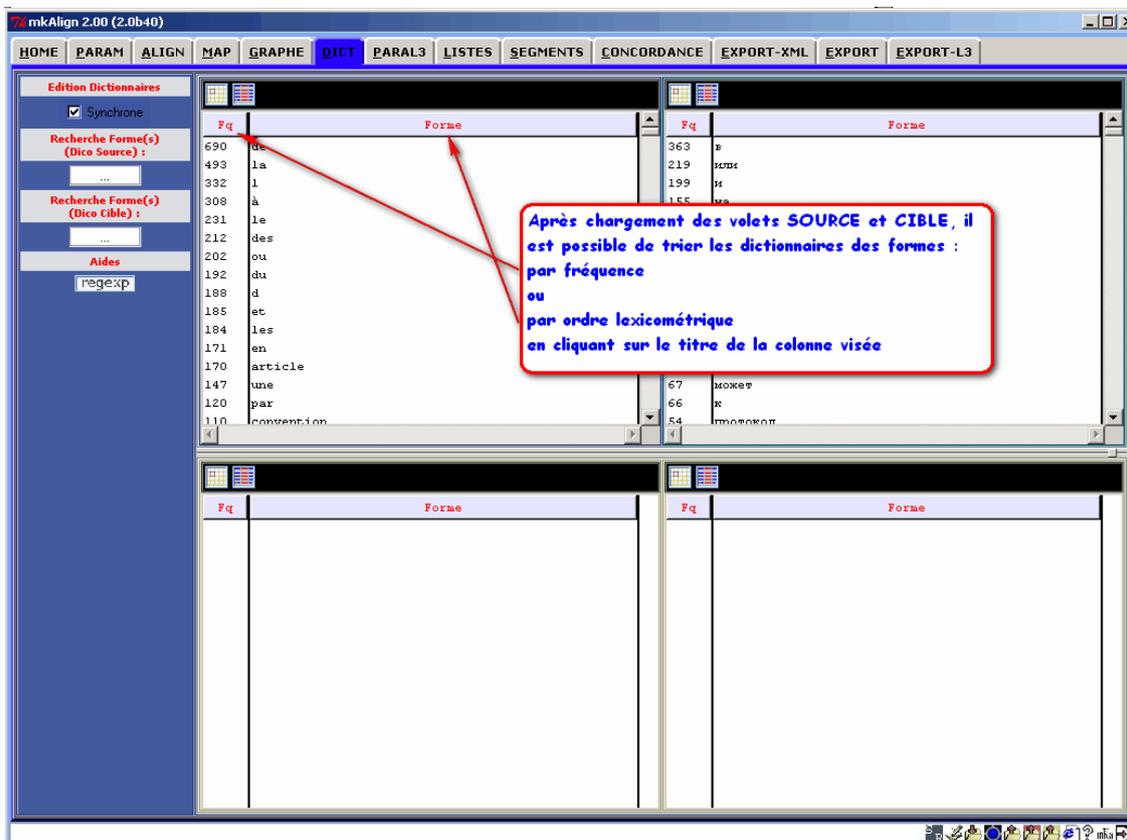


Figure 9(1) : *mkAlign* Onglet **DIC**, chargement

Après sélection d'une ou de plusieurs formes du dictionnaire, il est possible de déclencher :

- la recherche de cette forme (ou de ces formes) dans la représentation cartographique de l'alignement présentée *infra* (cf onglet **MAP**).

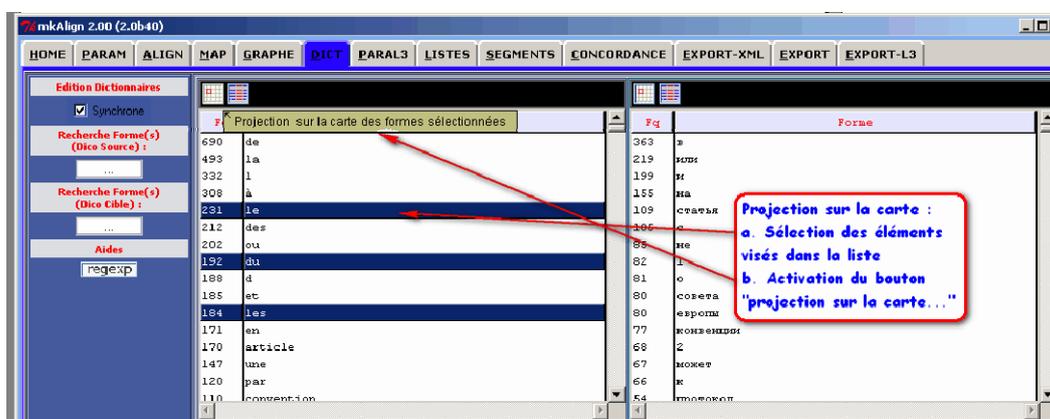


Figure 9(2) : *mkAlign* Onglet **DIC**, projection sur la carte d'une sélection de formes

- la construction d'une concordance de cette forme (ou de ces formes) et l'affichage du résultat de cette concordance dans l'onglet **CONCORDANCE**.

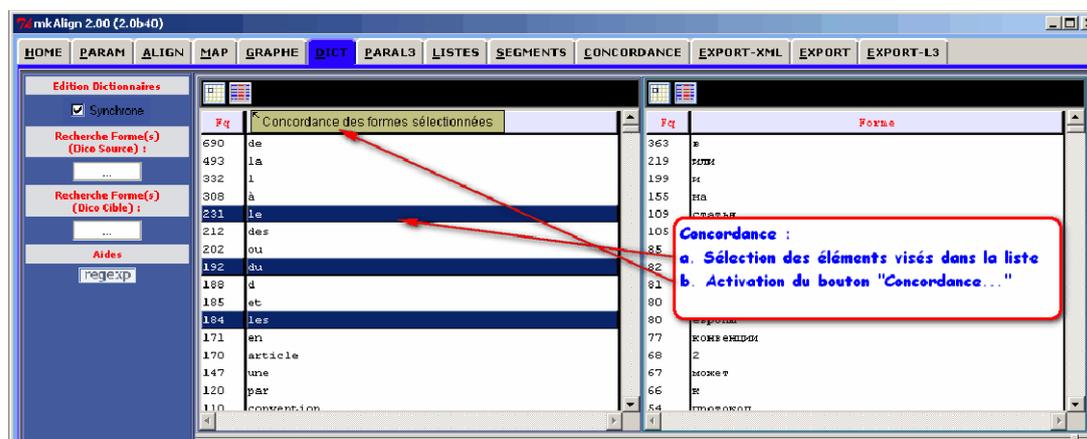


Figure 9(3) : *mkAlign* Onglet **DICT**, concordance d'une sélection de formes

On peut rechercher des formes dans le dictionnaire en définissant un motif de recherche dans les zones de saisie « Recherche Forme(s) ». Les formes trouvées seront affichées dans les zones d'édition situées au bas de l'onglet **DICT**. On peut réaliser la même opération en sélectionnant un item dans la liste des formes puis en activant le raccourci *clic-droit* sur la forme visée.

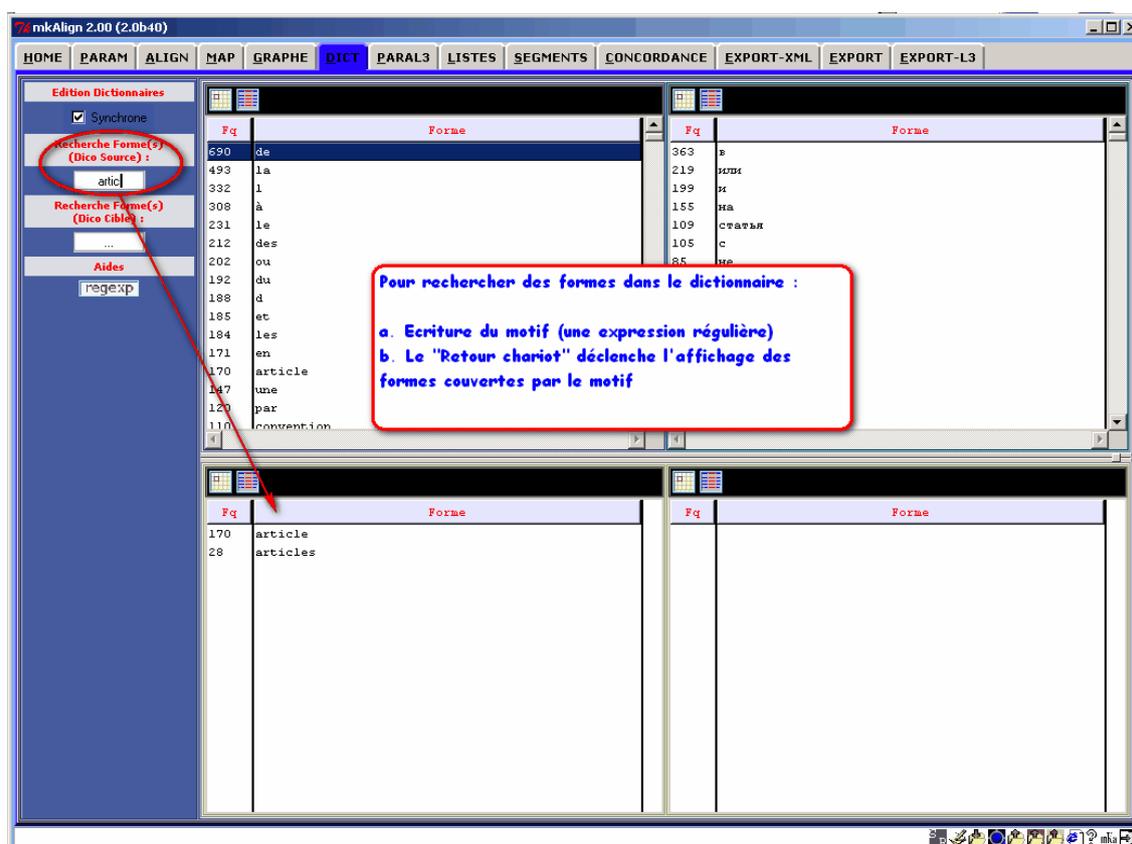


Figure 9(4) : *mkAlign* Onglet **DICT**, recherche de formes

Là encore, on peut sélectionner une forme (ou plusieurs) dans la liste construite puis déclencher (1) la recherche de cette forme dans la représentation cartographique de

l'alignement présentée *infra* (cf onglet **MAP**), (2) la construction d'une concordance sur cette forme et l'affichage du résultat de cette concordance dans l'onglet **CONCORDANCE**.

- les paramètres de la segmentation sont disponibles dans le l'onglet **RAPPORT**

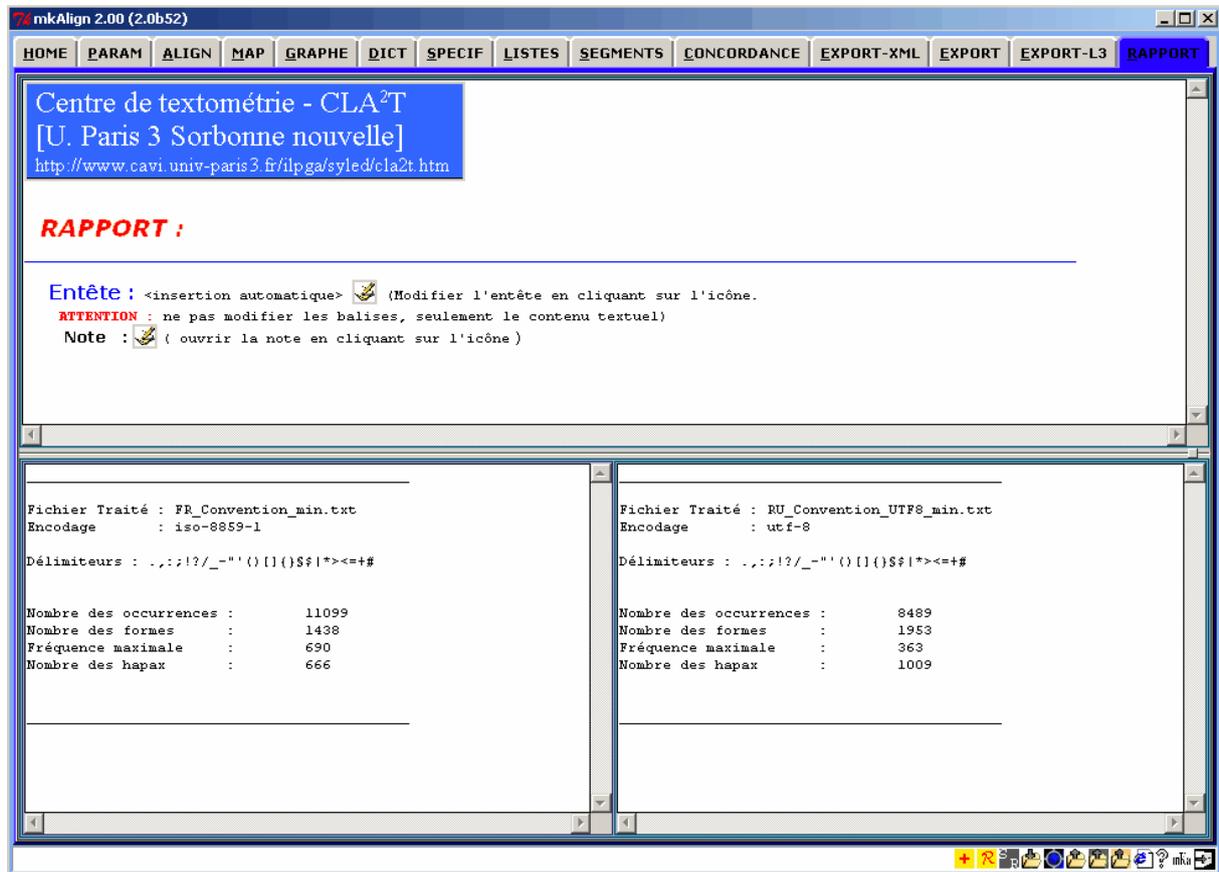


Figure 10 : *mkAlign* Onglet **RAPPORT**

6.2 Mode « alignement par recherche de cognats ». Alignement automatique de fichiers

Ce mode de chargement s'inspire de travaux réalisés dans le domaine de l'alignement automatique des phrases utilisant la recherche de points d'ancrage lexicaux pour mettre en correspondance des segments textuels plus longs. Cette méthode permet, pour des langues apparentées, de construire un alignement en recherchant tout d'abord des équivalents traductionnels sous forme de mots apparentés (ou cognats)⁶, les points d'ancrage obtenus déterminent ensuite l'alignement des zones en correspondance. Sur ce type de méthodes, on consultera en particulier [Kraif, 1999].

Pour lancer ce mode de chargement *via* **mkAlign**, il faut utiliser le bouton disponible sous l'item « **Alignement Source/Cible (Recherche de cognats)** ». Celui-ci déclenche (1) l'ouverture d'une fenêtre demandant à l'utilisateur de préciser les paramètres pour construire l'alignement :

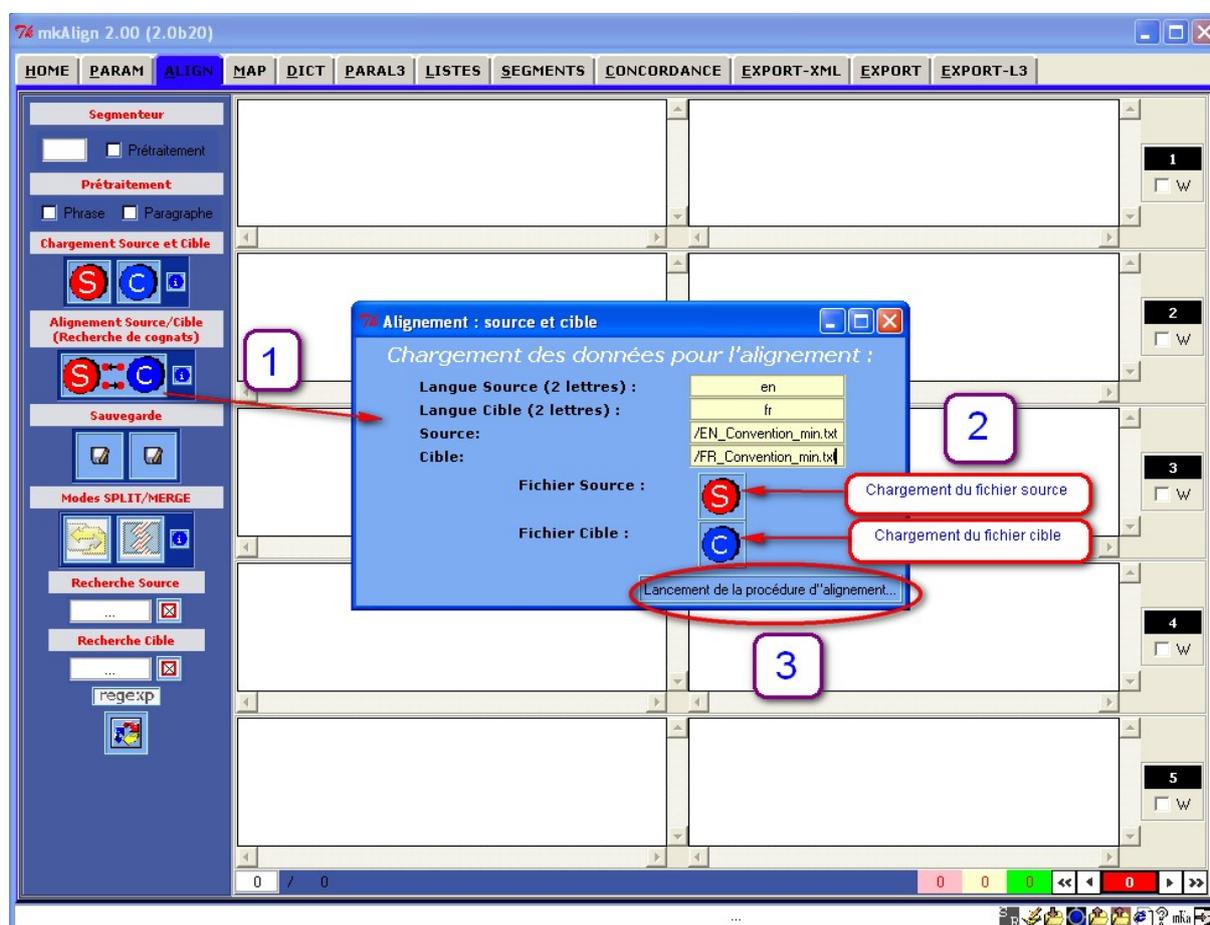


Figure 11 (1) : **mkAlign** Alignement par recherche de cognats

Les 4 paramètres étant définis (2), on peut lancer le processus d'alignement (3).

⁶ De manière informelle, les mots apparentés sont des mots qui partagent des propriétés phonologiques, orthographiques et sémantiques facilement repérables. Ces mots représentent souvent des traductions mutuelles. Parmi les exemples types pour l'anglais et le français, des mots comme *comprehension/compréhension*, *text/texte*, etc. Logiquement, les noms propres sont souvent dans cette catégorie (*Paris/Paris*, *London/Londres*, *Russia/Russie*), ainsi que les expressions numériques, et même parfois des signes de ponctuation (le point d'interrogation, les parenthèses, etc.).

Après alignement des 2 fichiers, **mkAlign** a l'allure suivante :

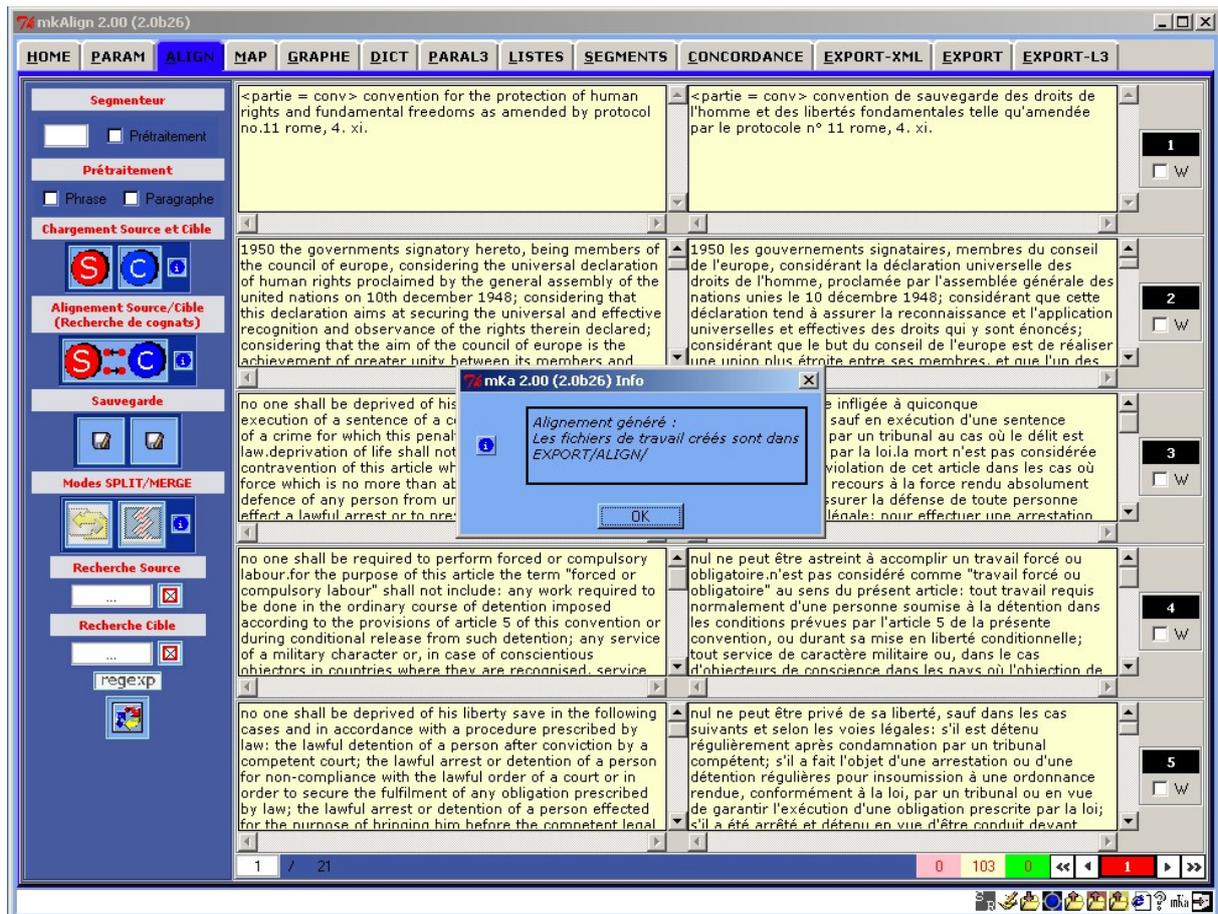


Figure 11 (2) : **mkAlign** Alignement par recherche de cognats

Avec **mkAlign**, les résultats de l'alignement par recherche de cognats peuvent toujours être corrigés ou affinés par l'utilisateur à l'aide des fonctionnalités de création / gestion du bi-texte (voir la section §6 « [Principales fonctionnalités de mkAlign](#) »).

6.3 Mode «Import d'alignement au format TMX »

Ce mode de chargement permet de charger un alignement construit sur deux ou plusieurs fichiers et enregistré en utilisant le format TMX⁷. Le format TMX (*Translation Memories eXchange*) est un format d'échange entre mémoires de traduction développé par Lisa⁸. Dans ce format, un lien implicite relie les segments équivalents dans une même unité de traduction :

```
- <tuv>
- <tuv xml:lang="fr">
  <seg>article 1 - obligation de respecter les droits de l'homme 1 les hautes parties contractantes reconnaissent à toute personne relevant de leur juridiction les droits et libertés définis au
  titre I de la présente convention: titre I - droits et libertés 1</seg>
</tuv>
- <tuv xml:lang="ru">
  <seg>статья 1 - обязательство соблюдать права человека высокие договаривающиеся стороны обеспечивают каждому, находящемуся под их юрисдикцией, права и свободы,
  определенные в разделе I настоящей конвенции. раздел I - права и свободы</seg>
</tuv>
</tuv>
```

Les figures suivantes illustrent le chargement d'un fichier de ce type. Après sélection du fichier à importer, il suffit de sélectionner les 2 langues associées aux deux volets visés :

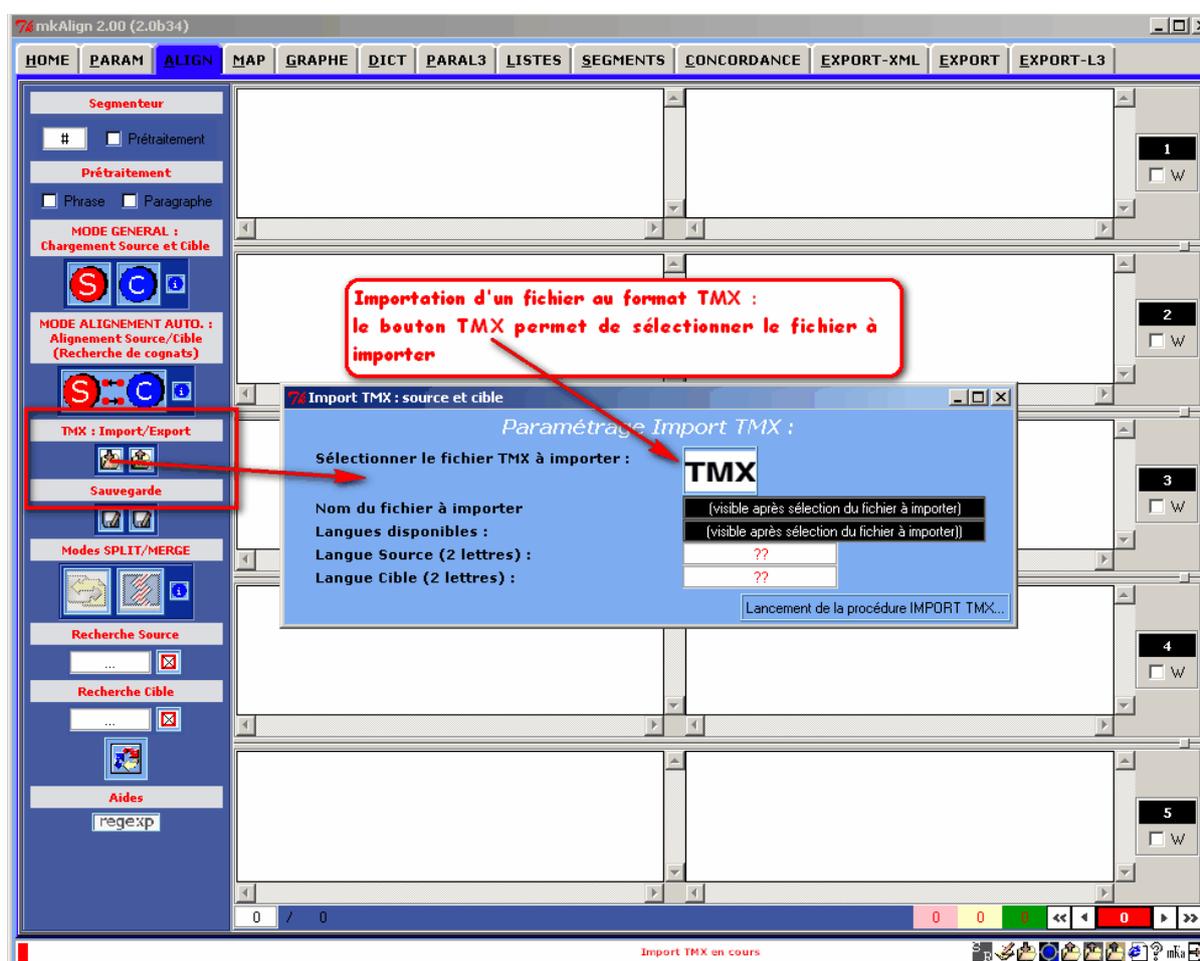


Figure 12 (1) : *mkAlign* Alignement par importation d'un fichier au format TMX (choix du fichier)

⁷ <http://www.lisa.org/fileadmin/standards/tmx1.4/tmx.htm>

⁸ <http://www.lisa.org/Translation-Memory-e.34.0.html>

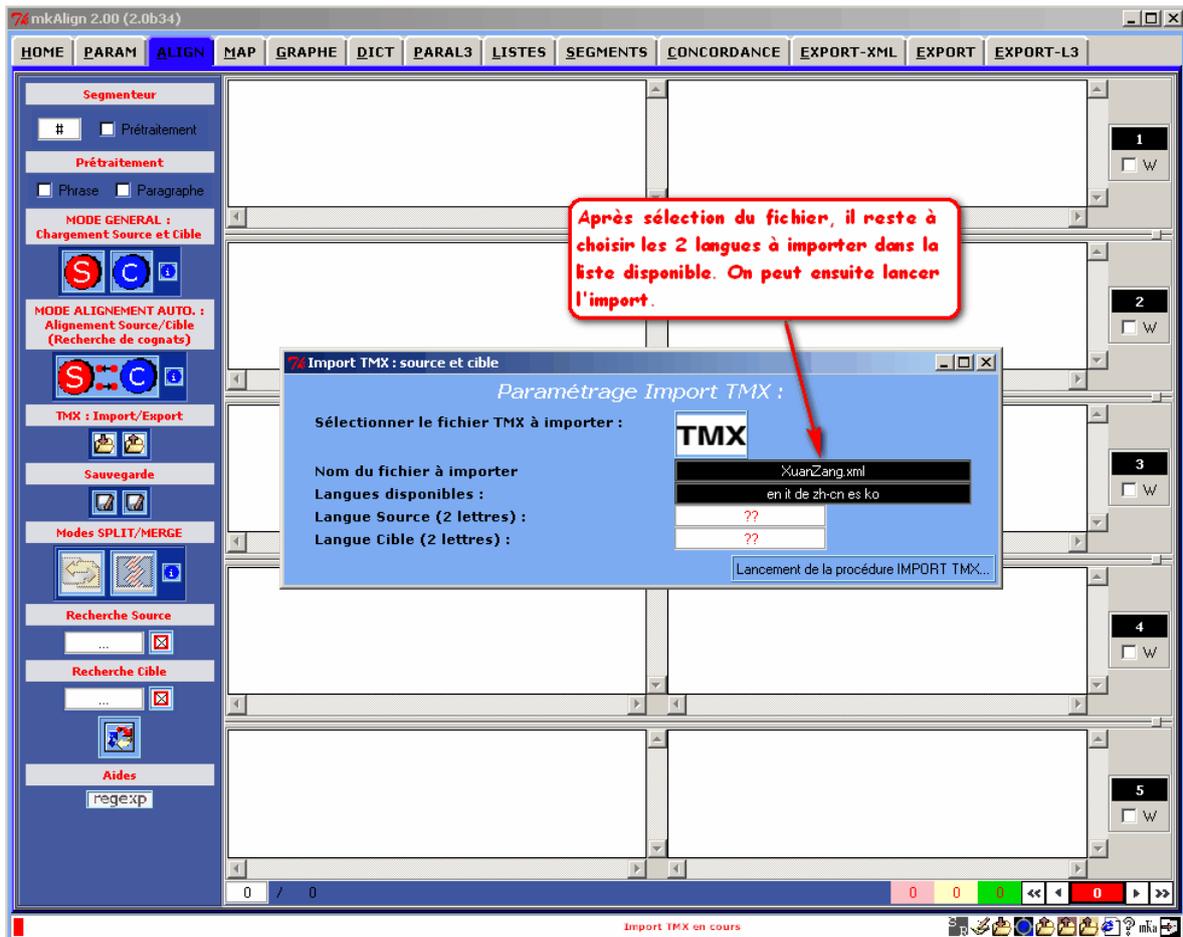


Figure 12 (2) : *mkAlign* Alignement par importation d'un fichier au format TMX (choix de 2 langues)

Le dossier [EXPORT/TMX](#) contient 2 exemples de ce type de fichier qu'il est possible d'importer dans *mkAlign*.

7 Principales fonctionnalités de **mkAlign** pour l'édition de l'alignement

7.1 Recherche de chaînes

La figure 12 illustre une des fonctionnalités classique d'éditeurs de texte que l'on retrouve dans **mkAlign** : la recherche de chaîne de caractères.

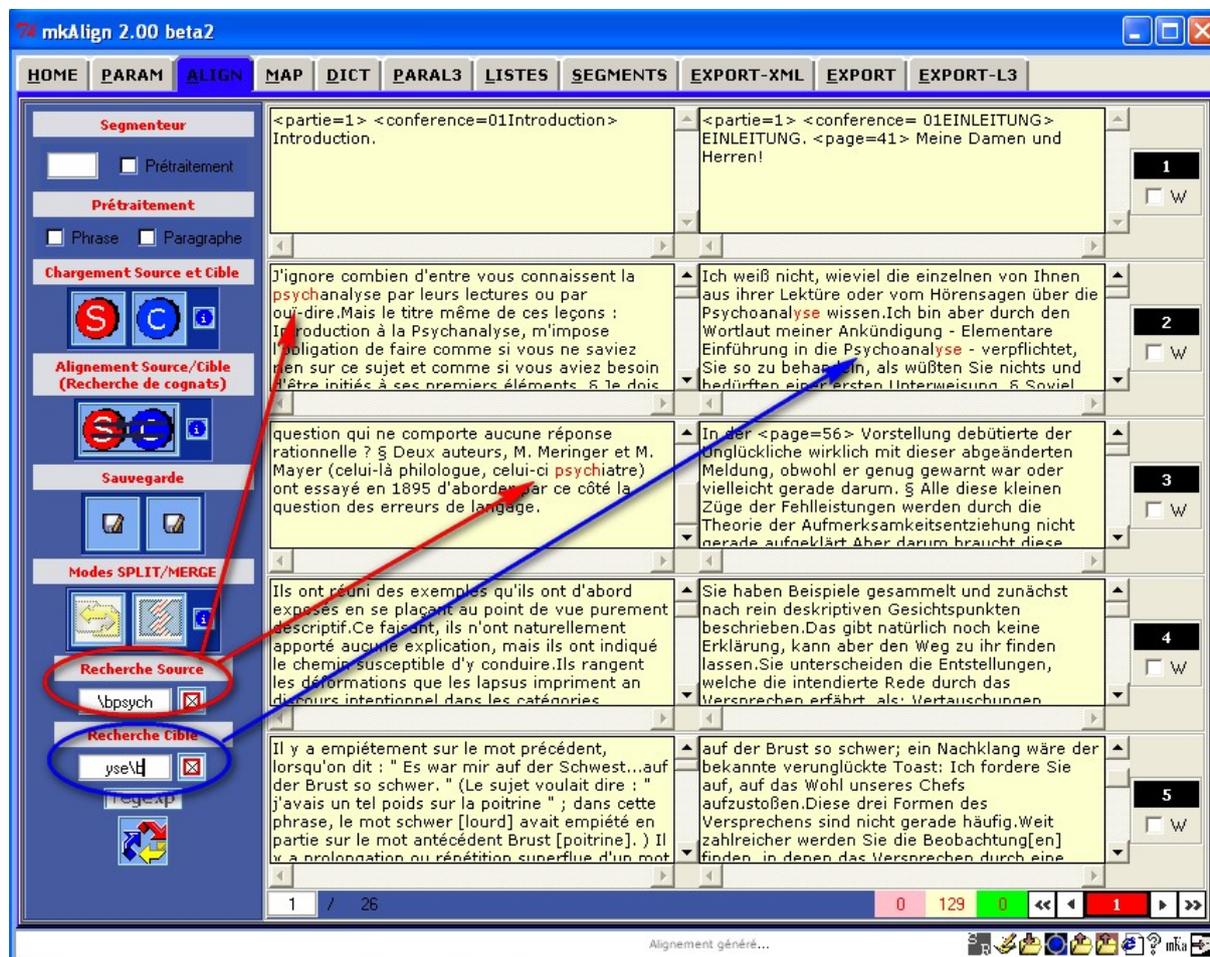


Figure 13 : **mkAlign** Onglet **ALIGN**, recherche de chaînes

On dispose dans l'onglet **ALIGN** de 2 zones de saisie permettant de lancer des recherches de chaînes : "*Recherche Source*" et "*Recherche Cible*". Les requêtes de recherche peuvent s'exprimer sous la forme d'expressions régulières [Fourmond, 2005].

Une expression régulière a pour fonction de définir un "modèle" de chaîne de caractères. Seules les formes lexicales contenant une chaîne de caractères conforme à ce modèle seront sélectionnées. Une des formes d'expression régulière simple est, par exemple, **ique**. Seules les formes qui contiennent **ique** seront sélectionnées.

Dans la figure précédente par exemple, les 2 requêtes exprimées sont :

Recherche Source : `\bpsych`

Recherche Cible : `yse\b`

Pour la requête **SOURCE**, on recherche en début de mot (`\b`) la présence des caractères `psych` : on cherche, par exemple, tous les mots qui commencent par `psych`. Comme le montre la figure 12, les séquences trouvées dans la figure sont colorées en rouge (`psychanalyse`).

Pour la requête **CIBLE**, on recherche la présence des caractères `yse` en fin de mot (`\b`): on cherche, par exemple, tous les mots qui se terminent par `yse`. Sur la figure 12, les séquences trouvées dans le texte **CIBLE** sont colorées en rouge (`Psychoanalyse`).

On présente en annexe les différents opérateurs d'expression régulière disponibles avec **mkAlign** (opérateurs classiques).

7.2 Exporter un sous-corpus contenant un motif

On trouve à côté de chaque zone de saisie de recherche un bouton permettant de sauvegarder l'ensemble des cellules contenant le motif défini dans la zone de saisie : on exporte ainsi un sous-corpus contenant toutes les sections contenant le motif sélectionné. Le fichier généré par l'exportation peut-être réimporté dans **mkAlign** (voir la procédure décrite *supra* : « [Chargement des textes à aligner](#) »).



7.3 Figement de cellule (lecture/écriture)

- il est possible de figer 2 cellules alignées (case à cocher en regard de chaque couple de cellules alignées), dans ce cas les cellules sont protégées en écriture (*état R*), les cellules figées sont colorées en vert
- si une cellule n'est pas protégée, elle est modifiable en écriture (*état W*) (couleur blanche)

7.4 Fractionnement de cellule⁹

- l'insertion du caractère *segmenteur* provoque automatiquement le découpage de la cellule concernée,
- si un couple de cellules est figée à une position inférieure dans **mkAlign**, l'insertion du *segmenteur* (dans une cellule **SOURCE** ou **CIBLE**) provoque aussi l'insertion d'une nouvelle cellule vide (dans la colonne **CIBLE** ou **SOURCE**) juste avant la cellule figée, cette cellule insérée est colorée en rose.

7.5 Fusion de cellule

- la suppression du caractère *segmenteur* dans une cellule provoque automatiquement la **fusion** de la cellule concernée et de la cellule juste au dessous

⁹ Cf mode SPLIT *infra* : lecture LR ou RL.

- si un couple de cellules est figé à une position inférieure dans **mkAlign**, la fusion provoque aussi l'insertion d'une nouvelle cellule vide dans la même colonne juste avant la cellule figée, cette cellule insérée est colorée en rose.

7.6 Etat de l'alignement

- on trouvera au bas de l'onglet **ALIGN** 3 cellules de couleur contenant chacune un compteur des différents états (de couleur) des cellules, par défaut
 - o le compteur vert indique le nombre de lignes protégées,
 - o le compteur rose indique le nombre de lignes contenant une cellule rose provenant d'une insertion ou d'une fusion,
 - o et le compteur blanc le nombre de ligne qui ne sont pas les précédentes

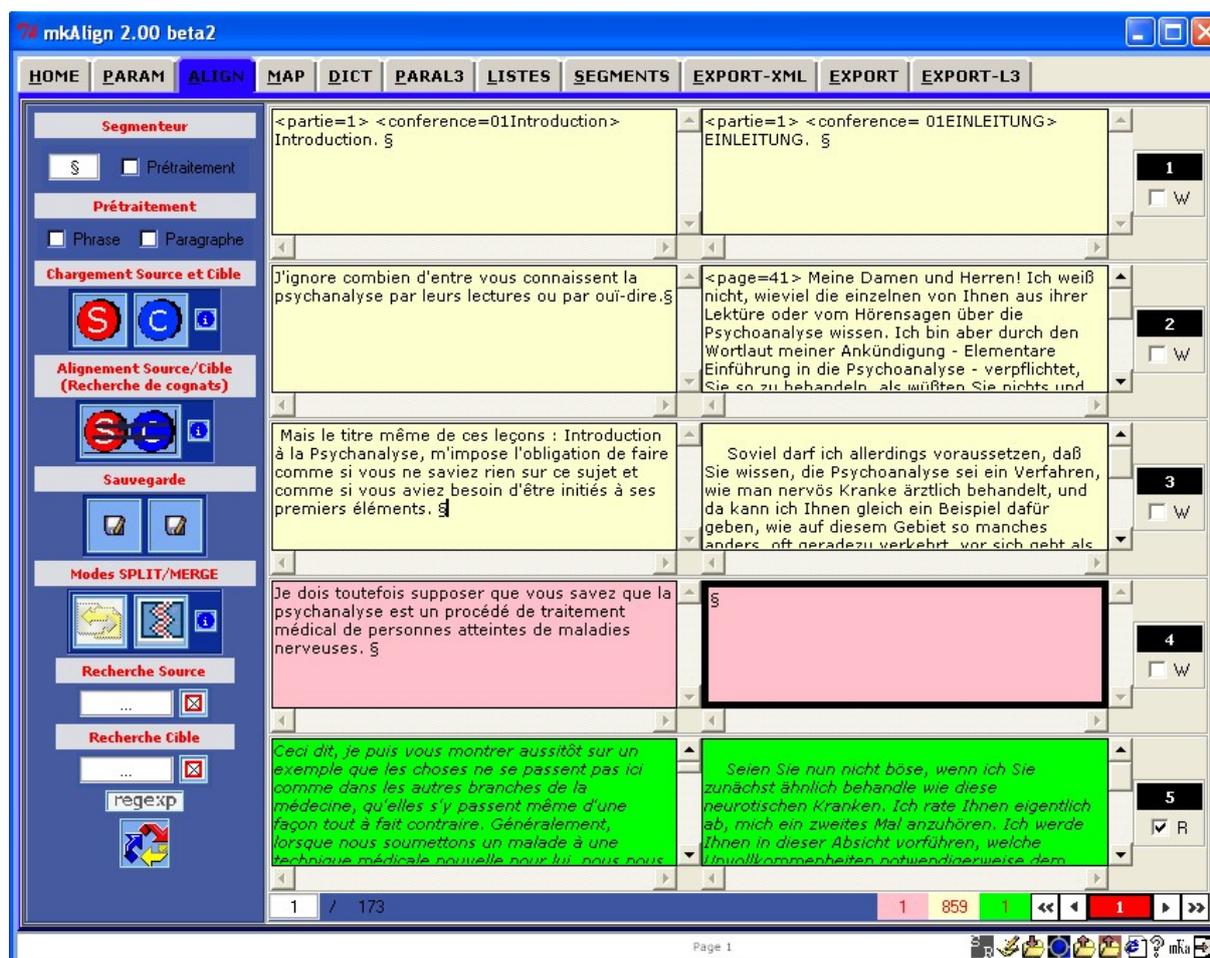


Figure 14 : **mkAlign** Alignement en cours

7.7 Remarques sur le segmenteur

- si la zone de saisie permettant de définir le **segmenteur** est vide, le **segmenteur** utilisé par le programme est le *retour chariot*, dans ce cas le fait de taper un retour chariot dans une cellule provoque le découpage de la cellule concernée
- si la zone de saisie permettant de définir le **segmenteur** est vide, le **segmenteur** utilisé par le programme est le *retour chariot*, dans ce cas, pour réaliser une fusion de cellule, il faut utiliser le **MODE MERGE** présenté *infra*.

7.8 Les modes

Dans les fonctionnalités précédentes, l'insertion ou la suppression du caractère *segmenteur* est faite au clavier en insérant/supprimant dans la zone d'édition choisie le caractère *segmenteur*.

Il existe aussi des *modes* particuliers permettant de réaliser ces opérations de fractionnement ou de fusion de cellules :



le **MODE SPLIT**



le **MODE MERGE**.

Ces deux modes sont disponibles après avoir chargé les fichiers de travail et après activation de l'un ou l'autre de ces modes (*via* les boutons SPLIT/MERGE dans l'onglet **ALIGN**). Ces deux modes sont désactivés par défaut.

Une fois activé (*via* un *clic droit* sur le bouton correspondant au mode choisi), ce mode permet, *via* le *clic gauche* de la souris :

- de scinder une cellule (équivalent à l'insertion du caractère *segmenteur*) à l'endroit où le *clic* a été réalisé
- de fusionner la cellule dans laquelle le clic a été réalisé avec la cellule suivante (équivalent à la suppression dans la cellule d'édition visée du caractère *segmenteur*)

7.9 Fonctionnalité complémentaire pour le **MODE SPLIT** : lecture LR/RL

Par défaut, l'activation du **MODE SPLIT** scinde la cellule visée en maintenant le *contenu à gauche* du *clic gauche* (ou du *segmenteur* inséré) dans cette cellule et insère le *contenu à droite* dans une nouvelle cellule en dessous (tout en décalant toutes les autres vers le bas). Ce mode de scission par défaut (*lecture LR*) peut être modifié pour permettre de réaliser l'opération inverse (*lecture RL*) : maintien de contexte droit en place et descente du contexte gauche. Ce paramétrage est disponible dans l'onglet **PARAM**.

Exemple :

Cellules initiales :

<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,	<LIGNE="VERS1"> C' est un trou de verdure où chante une rivière,
--	--

Cellules finales : SPLIT avant le « où ». *SPLIT LR* (source) et *SPLIT RL* (cible) ;

<LIGNE="VERS1"> C' est un trou de verdure	où chante une rivière,
où chante une rivière,	<LIGNE="VERS1"> C' est un trou de verdure

8 Représentation cartographique de l'alignement

8.1 Construction de la carte de l'alignement

Après avoir chargé les 2 volets d'un alignement il est possible de visualiser l'alignement en cours dans une représentation cartographique (identique à celle produite par **Lexico3**¹⁰ via l'outil "*Carte des Sections*"¹¹). On trouvera dans l'onglet **MAP**, un bouton  permettant de construire cette représentation cartographique dans laquelle on disposera d'une carte des sections pour le volet **SOURCE** et d'une autre pour le volet **CIBLE**, les carrés construits portent la couleur de leur état dans l'aligneur (par défaut blanc, rose ou vert) ; au bas des 2 cartes on dispose aussi d'une zone d'édition pour éditer (via un *clic gauche*) le contenu d'un carré donné : le contour du carré sélectionné devient bleu et le contour du carré aligné devient noir.

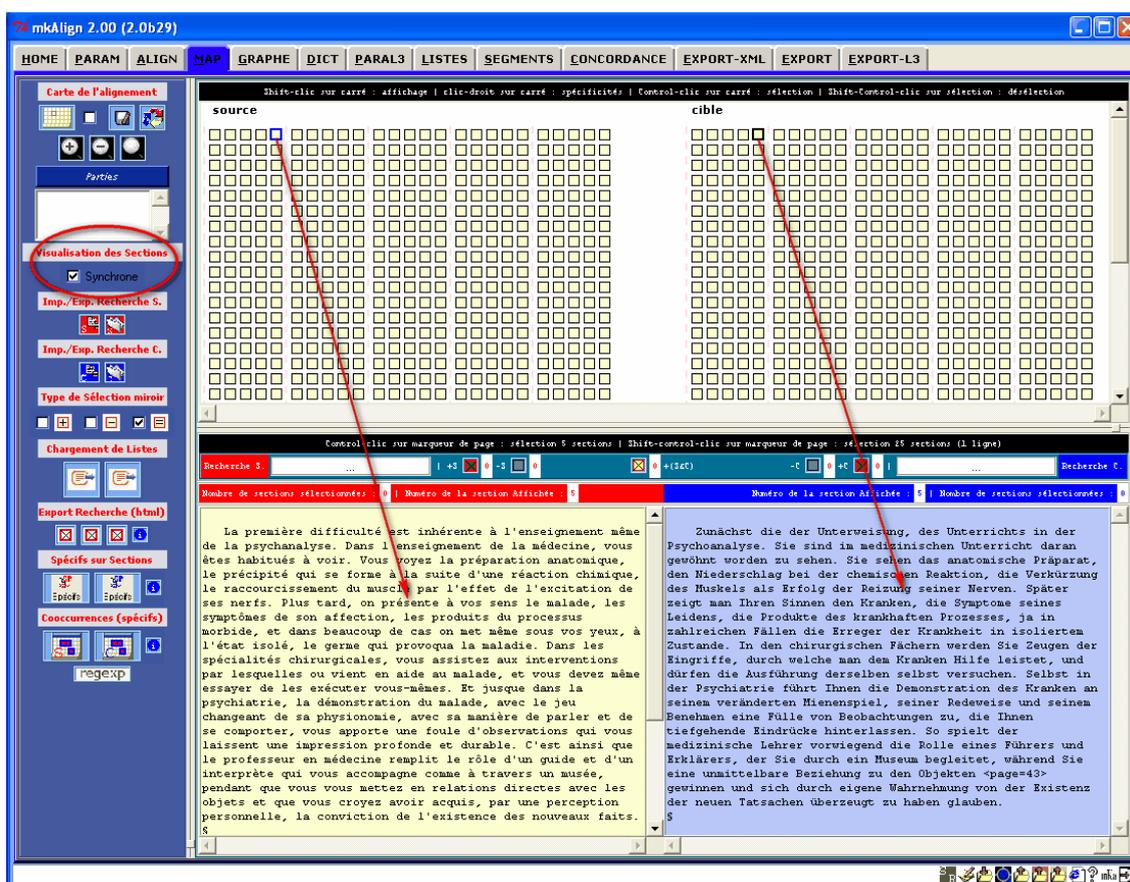


Figure 15 : *mkAlign* Onglet MAP

Paramétrage de l'affichage de la carte des sections



Le bouton  active le chargement de la carte. Par défaut, cet affichage insère linéairement les sections, sans alignement à gauche après insertion d'un marqueur de partie (*cf infra*). De

¹⁰ <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/>

¹¹ http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/manuelsL3/manuel-3.41.htm#_Toc34045613

plus, un séparateur de blocs de sections est inséré après affichage de 5 sections : ce séparateur est utilisable ensuite pour sélectionner les sections (*cf infra*).

- Alignement à gauche 

Ce bouton est inhibé par défaut, son activation (*clic-droit*) permet de représenter la carte en alignant les sections à gauche après chaque marquage de partie. Désactivation de cette fonction par un *clic* sur le bouton

- Insertion d'un marqueur de bloc de sections 

Ce bouton est actif par défaut, son inhibition permet de ne pas afficher le marqueur de blocs de sections. Pour inhiber cette fonction : *clic-droit*. Pour l'activer : *clic*

8.2 Recherche de motifs dans la carte de l'alignement

Il est possible de rechercher des chaînes de caractères, écrites sous la forme d'expressions régulières (*cf annexes*), dans le volet **SOURCE** et/ou dans le volet **CIBLE** : on dispose dans l'onglet **MAP** de 2 zones de saisie permettant d'y définir un motif. Si la chaîne est trouvée, le carré correspondant est surligné en rouge (dans le volet **SOURCE** ou **CIBLE**) et les diagonales du carré sont dessinées. Les sections correspondantes (respectivement dans le volet **SOURCE** ou **CIBLE**) seront-elles aussi matérialisées par un surlignage rouge (trait fin) : la *zone miroir*.

Remarque : Le surlignage rouge sera inhibé après sélection (par un *clic gauche*) d'un carré donné et la croix sera maintenue *i.e* les résultats de la précédente requête resteront ainsi accessibles. Par contre le surlignage rouge restera actif après activation du raccourci clavier *shift-clic* sur le carré visé (affichage de la section associée).

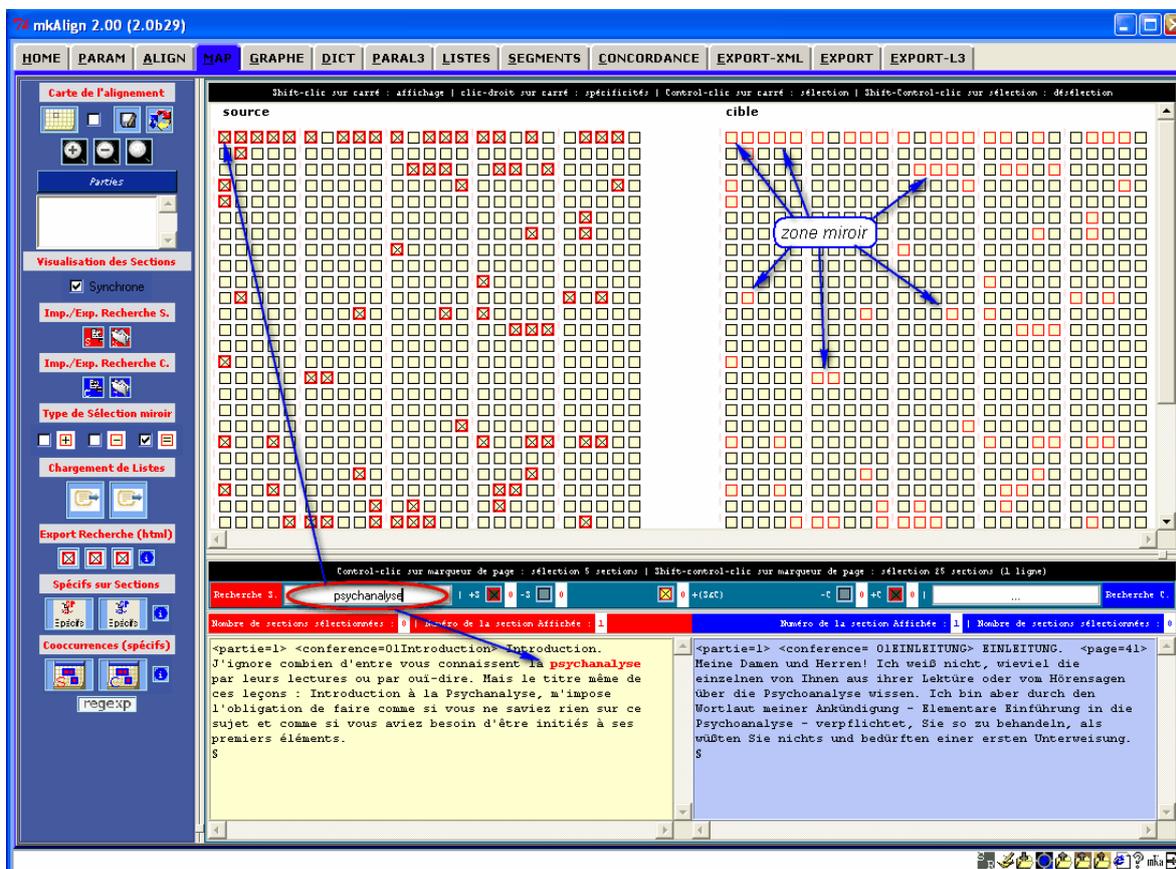


Figure 16 : *mkAlign* Recherche dans la carte des sections

A l'issue d'une recherche de motif dans la carte, la zone d'édition obtenue par un *clic gauche* (ou via *shift-clic*) sur un carré donné prend en compte le motif défini dans la zone de saisie associée : la sélection ou l'affichage d'une section de la carte déclenche une recherche du motif dans l'éditeur et produit le cas échéant l'affichage coloré de ce motif.

La recherche de 2 motifs respectivement dans chacun des volets de l'alignement permet de donner une représentation contrastée de la ventilation de ces 2 motifs :

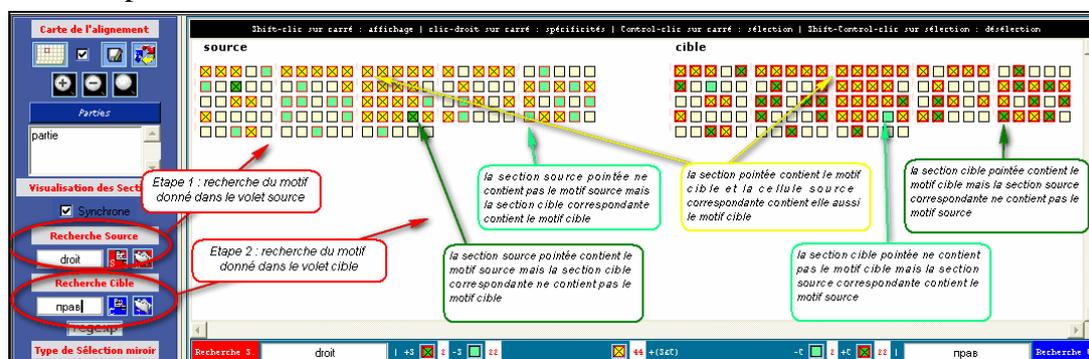


Figure 17 : *mkAlign* Recherche et Navigation dans la carte des sections

Dans la carte précédente, on cherche à donner une représentation de la ventilation d'un motif **SOURCE** ET d'un motif **CIBLE**. Le résultat final permet de distinguer 3 types de sections dans chacun des volets :

Volet **SOURCE**

- ☒ le motif **SOURCE** est présent dans la section **SOURCE** visée **ET** le motif **CIBLE** est présent dans la section **CIBLE** correspondante
- ☒ le motif **SOURCE** est présent dans la section **SOURCE** visée **MAIS** le motif **CIBLE** n'est pas présent dans la section **CIBLE** correspondante
- ☑ le motif **SOURCE** n'est pas présent dans la section **SOURCE** visée **MAIS** le motif **CIBLE** est présent dans la section **CIBLE** correspondante

Volet **CIBLE**

- ☒ le motif **CIBLE** est présent dans la section **CIBLE** visée **ET** le motif **SOURCE** est présent dans la section **SOURCE** correspondante
- ☒ le motif **CIBLE** est présent dans la section **CIBLE** visée **MAIS** le motif **SOURCE** n'est pas présent dans la section **SOURCE** correspondante
- ☑ le motif **CIBLE** n'est pas présent dans la section **CIBLE** visée **MAIS** le motif **SOURCE** est présent dans la section **SOURCE** correspondante

Remarque : ce type de visualisation n'est possible que sous « le mode  » (mode par défaut *cf infra*).

Un *clic gauche* sur chacun des boutons    permet aussi de sauvegarder un état de l'alignement au format HTML prenant en compte la ventilation des motifs visés :

- ☑ état de l'alignement, noté (S-, C+), avec le motif **SOURCE** absent dans une section du volet **SOURCE** et le motif **CIBLE** présent dans la section correspondante du volet **CIBLE** (et respectivement (S+, C-))

8.4 Affichage d'une partition dans la carte de l'alignement

Il est aussi possible de visualiser une éventuelle partition présente dans le codage du texte *via* des balises (*cf supra*). Après avoir activé le bouton **Parties** puis sélectionné une ou plusieurs clés de cette partition (*via* la liste de choix), la construction de la représentation cartographique de l'alignement intègrera ce marquage des parties sélectionnées.

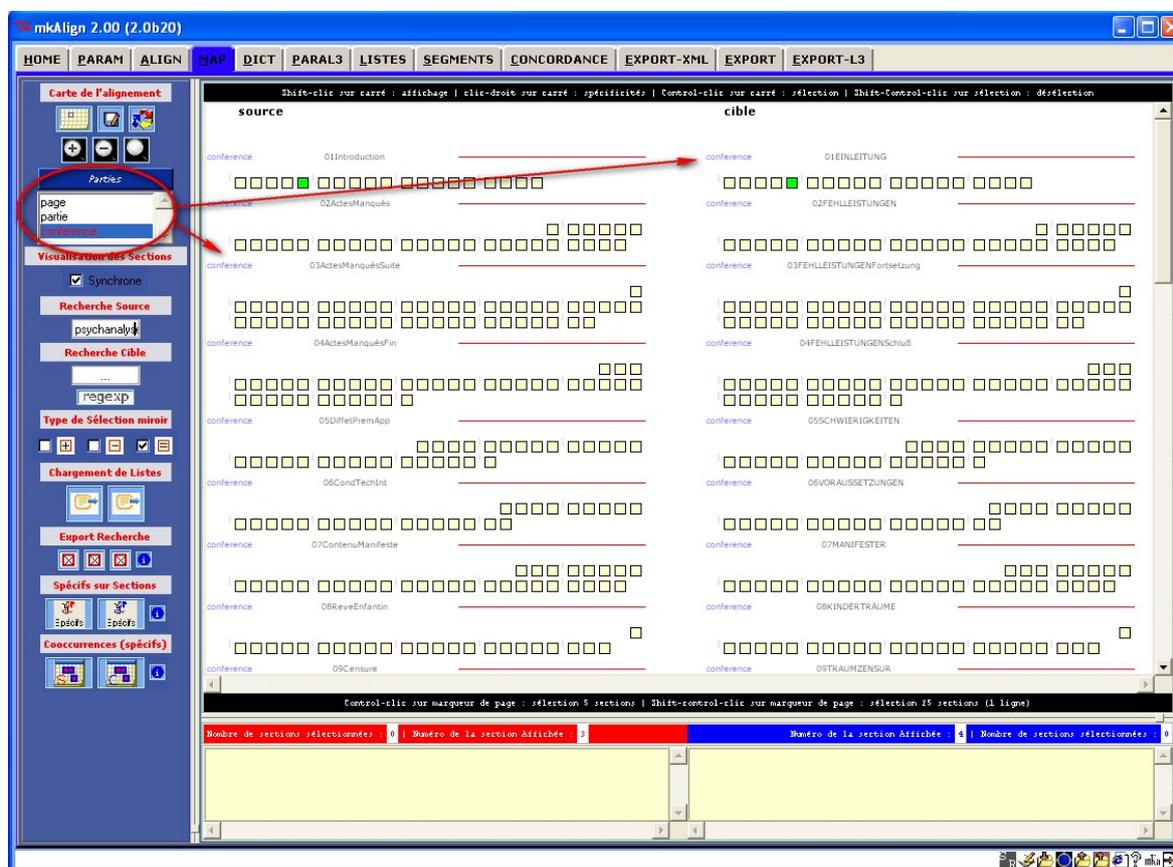


Figure 18 : *mkAlign* Carte / Sections / Parties

8.5 Navigation vers l'alignement

La carte contient aussi un *marqueur des pages d'édition* dans l'aligneur :



Dans la figure précédente, le marqueur | indique le passage d'une page à l'autre dans l'édition de l'alignement visible dans l'onglet **ALIGN**. Un *clic gauche* sur ce marqueur permet de retourner dans l'onglet **ALIGN** à la page visée par le marqueur.

8.6 Le vocabulaire spécifique d'une section de la carte

La carte dispose d'une fonctionnalité supplémentaire pour agir sur les contenus textuels des sections. On a vu qu'un *clic* sur un carré déclenchait l'affichage du contenu textuel associé à ce carré dans la zone d'édition prévue à cet effet. Un *clic droit* sur un carré déclenche le calcul du vocabulaire spécifique de la section considérée.

L'analyse des spécificités permet de porter un jugement sur la fréquence de chacune des unités textuelles dans chacune des sections du corpus¹².

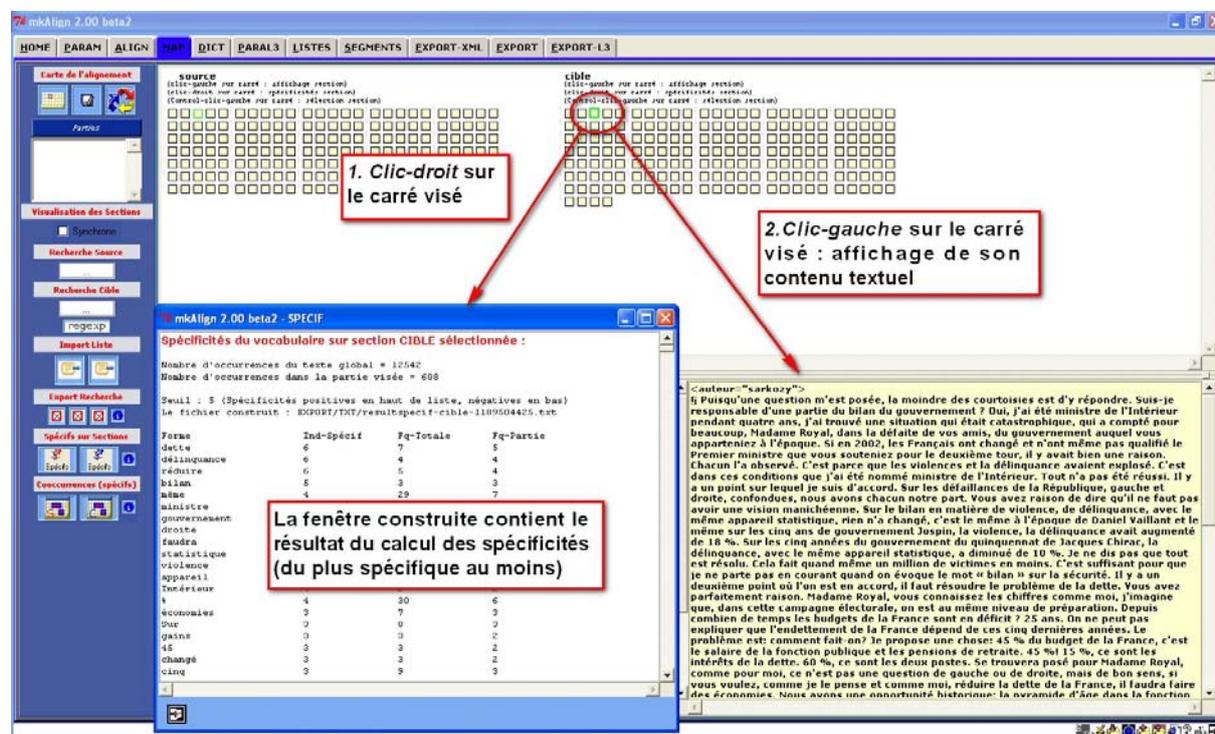
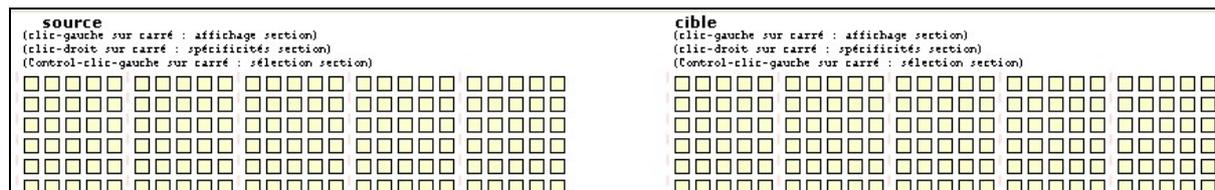


Figure 19 : mkAlign Carte / Sections / Spécificités

¹² Sur la méthode des spécificités on consultera par exemple : (Lafon, 1984) ou en ligne : <http://weblex.ens-lsh.fr/doc/weblex/contraste.html#specif>

8.7 Le vocabulaire spécifique d'une sélection de sections de la carte

Il est possible de sélectionner une série de sections dans la carte de l'alignement (**SOURCE** ou **CIBLE**) et de calculer le vocabulaire spécifique de cette sélection. Pour sélectionner une section, il convient d'activer la combinaison de touches : *Control+clic-gauche*. Après avoir été sélectionné, le contour du carré associé à la sélection devient vert gras, la zone cible associée a elle aussi un contour vert mais fin.



Figure 20 : *mkAlign* Carte / Sections / Sélection

Après avoir sélectionné une série de sections, on peut activer le calcul du vocabulaire spécifique de cette sélection *via* les 2 boutons disponibles dans l'onglet **MAP**.

	recherche du vocabulaire spécifique d'une sélection de section dans la SOURCE
	recherche du vocabulaire spécifique d'une sélection de section dans la CIBLE

Le résultat produit donne à voir le vocabulaire spécifique des sections sélectionnées (dans la **SOURCE** ou dans la **CIBLE**) et des sections associées (respectivement dans la **CIBLE** ou dans la **SOURCE**).

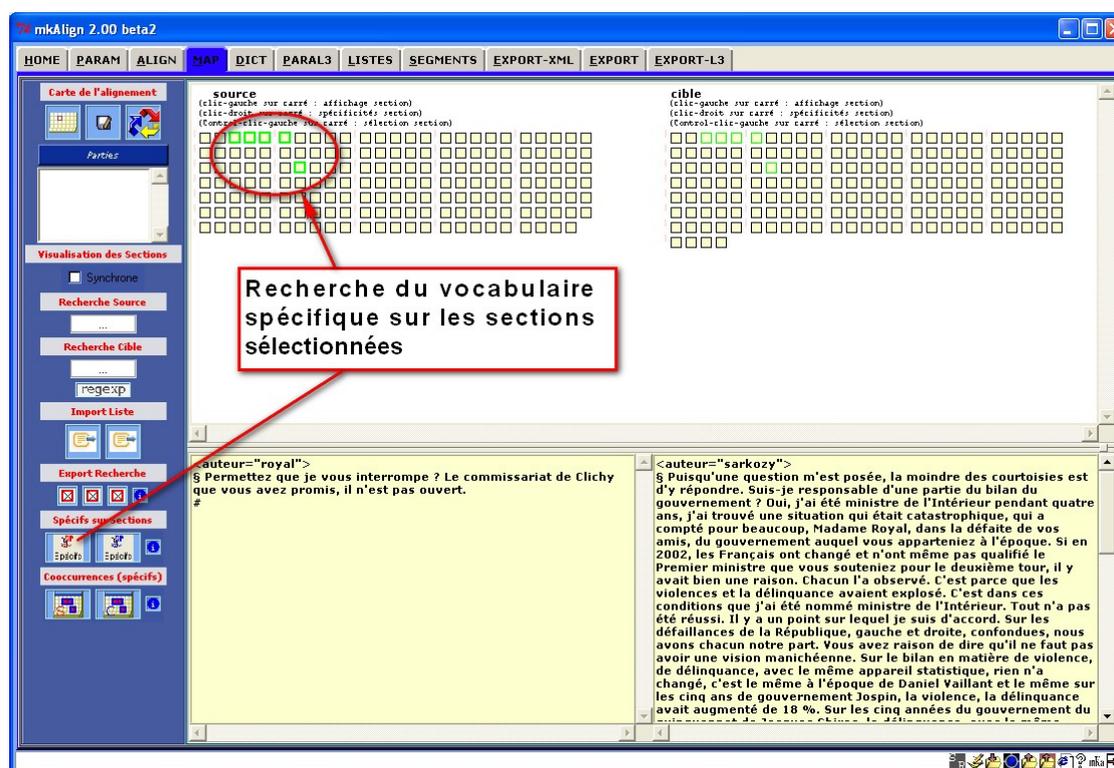


Figure 21 (1) : *mkAlign* Carte / Sections / Spécificité sections

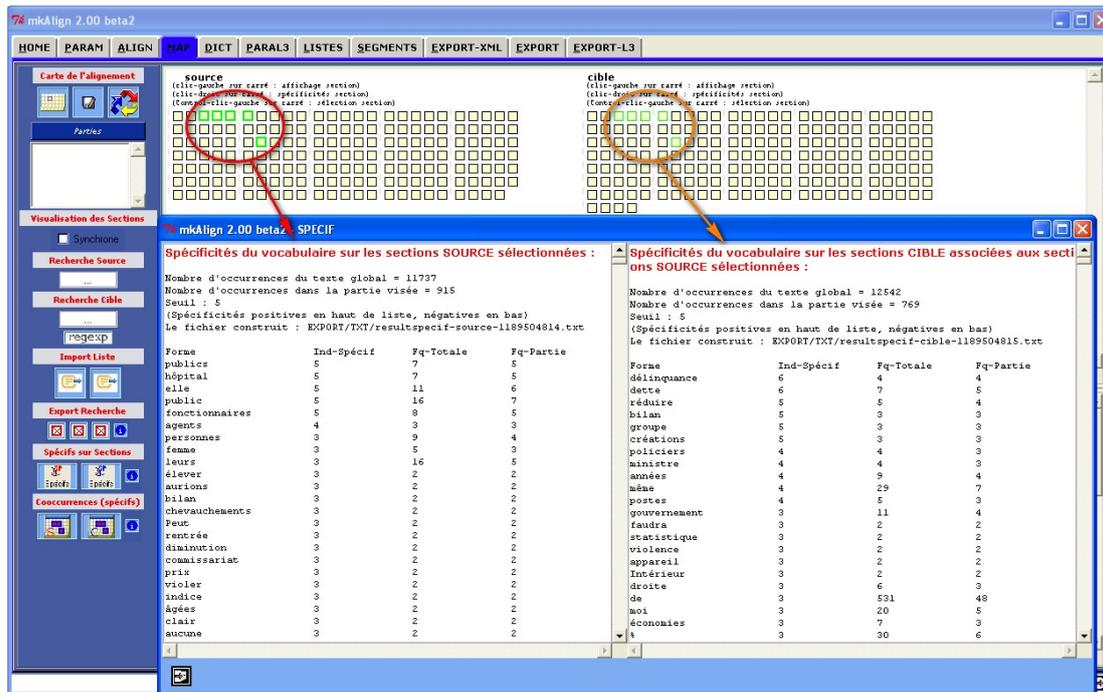


Figure 21 (2) : mkAlign Carte / Sections / Spécificité sections

Les résultats produits par ce calcul de vocabulaire spécifique seront insérés dans l'onglet **SPECIF**.

8.8 Fonctionnalités complémentaires sur la carte des sections

Control-Clic sur un marqueur des pages d'édition : sélection des 5 sections à sa droite



Shift-Control-Clic sur un marqueur des pages d'édition : sélection des 25 sections à sa droite (une ligne dans la carte des sections)



Shift-Control-Clic sur un carré dans la carte des sections : dé-sélection de la section visée.

8.9 Le vocabulaire spécifique des sections de la carte contenant une forme ou un motif : recherche de cooccurents

L'onglet **MAP** dispose de deux boutons supplémentaires permettant de calculer les cooccurents d'une forme ou d'un motif donné.

	recherche des cooccurents d'une forme dans la source
	recherche des cooccurents d'une forme dans la cible

A partir de la carte de l'alignement, on peut rechercher (sur la **SOURCE** ou sur la **CIBLE**) une forme graphique ou un motif. On obtient une série de carré indiquant la présence du mot ou du motif. On peut ensuite lancer un calcul du vocabulaire spécifique de l'ensemble de ces sections. On obtient les mots spécifiques contenus dans l'ensemble des sections contenant le mot ou le motif initial. On obtient une liste de mots qui portent soit un indice de spécificité positif soit un indice de spécificité négatif, dans le premier cas, on aboutit en gros à une liste des cooccurents du mot ou du motif cherchés, dans le second cas on obtient des mots qui n'apparaissent pas avec le mot ou le motif cherchés.

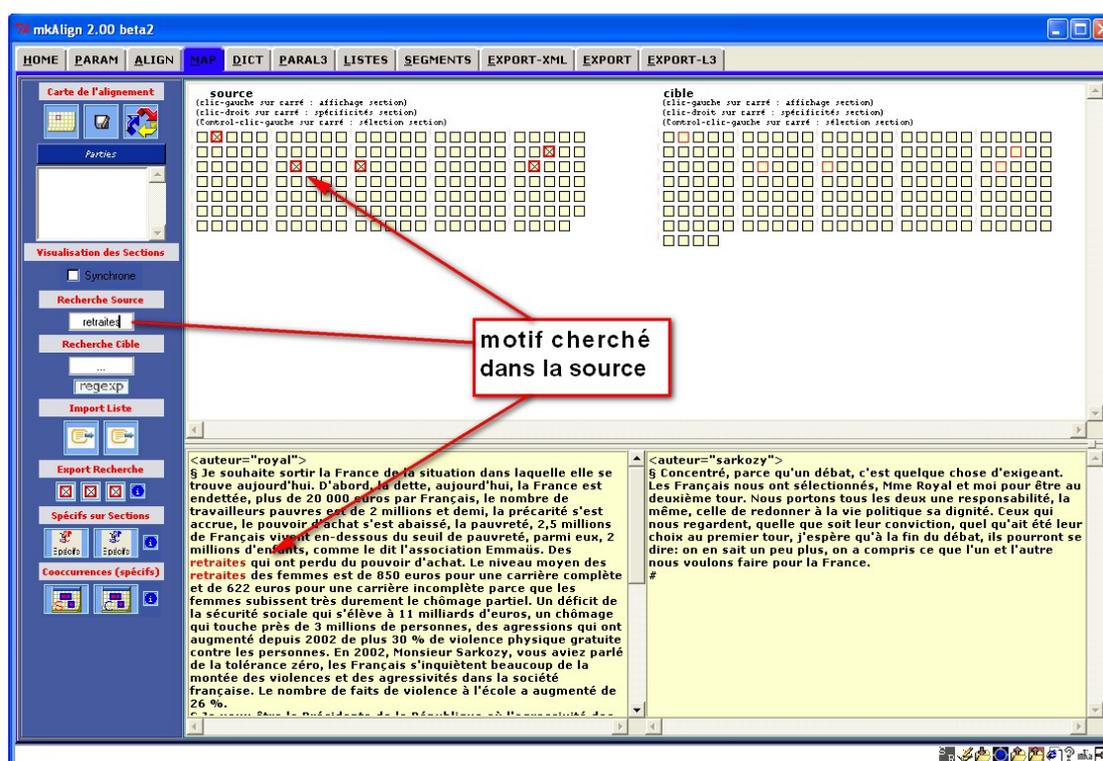


Figure 22 (1) : *mkAlign* Carte / Sections / Spécificités / Cooccurents

Le résultat produit donne à voir le vocabulaire spécifique des sections contenant le motif (dans la **SOURCE** ou dans la **CIBLE**) et des sections associées (respectivement dans la **CIBLE** ou dans la **SOURCE**).

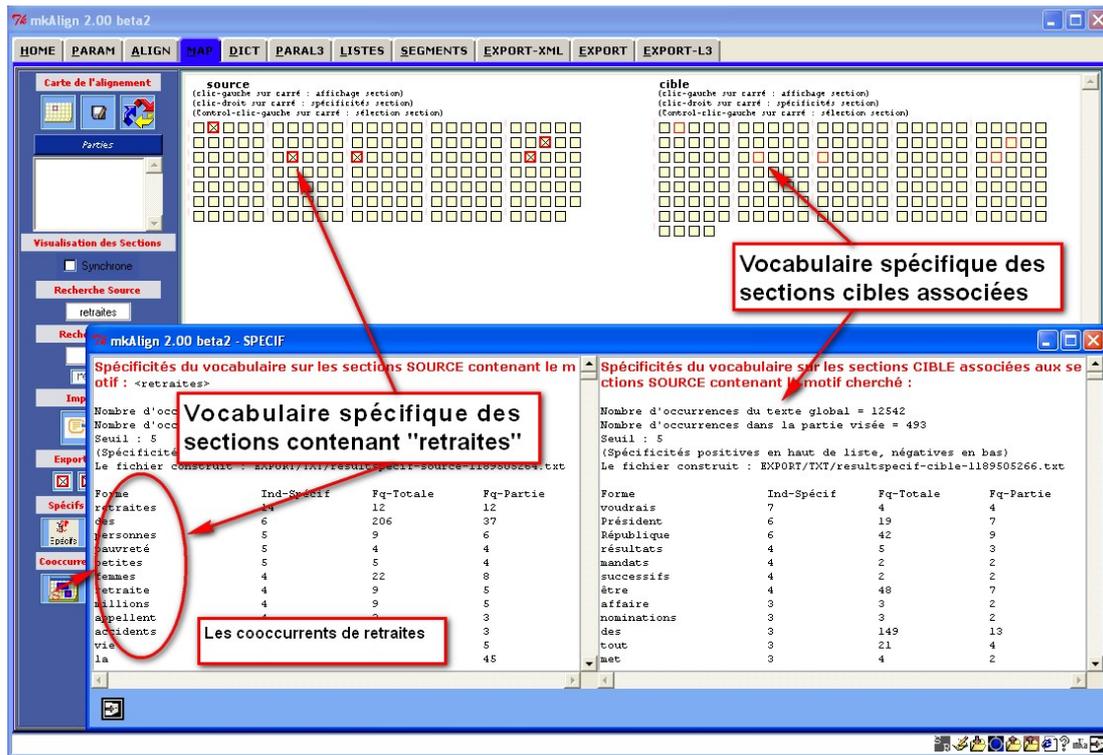


Figure 22 (2) : mkAlign Carte / Sections / Spécificités / Cooccurrents

8.10 Retour aux contextes à partir du vocabulaire spécifique

Les résultats produits par le calcul de vocabulaire spécifique sont insérés dans l'onglet **SPECIF**. Ces résultats sont associés à des processus de calcul déjà disponibles dans d'autres onglets :

- *Concordance* : calcul d'une concordance sur toutes les positions des items ayant la forme visée
- *Ventilation-section* : ventilation sur la carte des sections des formes sélectionnées
- *Ventilation*: ventilation sur la partition section des formes sélectionnées

Ces différents processus sont activables par les raccourcis clavier visibles dans l'onglet **SPECIF** :

Clic-droit sur forme : concordance | shift-Clic-droit sur forme : ventilation-section | Control-Clic sur forme : ventilation

8.11 Filtrage de la « zone miroir » au cours d'une recherche (« au-delà du miroir »)

Nous allons exposer dans cette section une fonctionnalité avancée de **mkAlign** permettant de raffiner un processus de recherche dans un contexte de mise à jour progressive d'équivalences traductionnelles dans des fichiers alignés. Cette fonctionnalité illustre la méthode présentée par [Zimina, 2004]. L'expérience à réaliser peut être résumée de la manière suivante :

1. On charge 2 fichiers alignés (par exemple dans le volet **SOURCE** un texte en français, et dans le volet **CIBLE** un texte en anglais)
2. On construit la carte de l'alignement
3. On cherche sur un volet une forme (ou un ensemble de formes) : par exemple la forme « fonctionnaire » sur le volet **SOURCE** en français (cf [Zimina, 2004]). Dans le volet **CIBLE**, les zones associées aux zones **SOURCE** sélectionnées (marquées en *rouge gras*) sont repérables par un marquage *rouge fin* : ce marquage fin dans le volet **CIBLE** est appelée la *zone miroir de la recherche* dans le volet **SOURCE**.
4. On construit les cooccurents de « fonctionnaire » (via la fonctionnalité présentée *supra*). On obtient 2 listes de formes :
 1. une pour le volet **SOURCE** (les cooccurents de « fonctionnaire »)
 2. une pour le volet **CIBLE** qui contient (au sommet de la liste) les candidats traductionnels et ses cooccurents ;
5. On veut ensuite s'assurer ou vérifier que ce candidat existe (ou pas) ailleurs dans la carte de l'alignement que dans les zones « sélectionnées » précédemment (*rouge fin*) ce qui conduirait éventuellement à considérer ce candidat comme pouvant avoir d'autres traductions dans le volet **SOURCE** (d'autres configurations sont aussi envisageables).

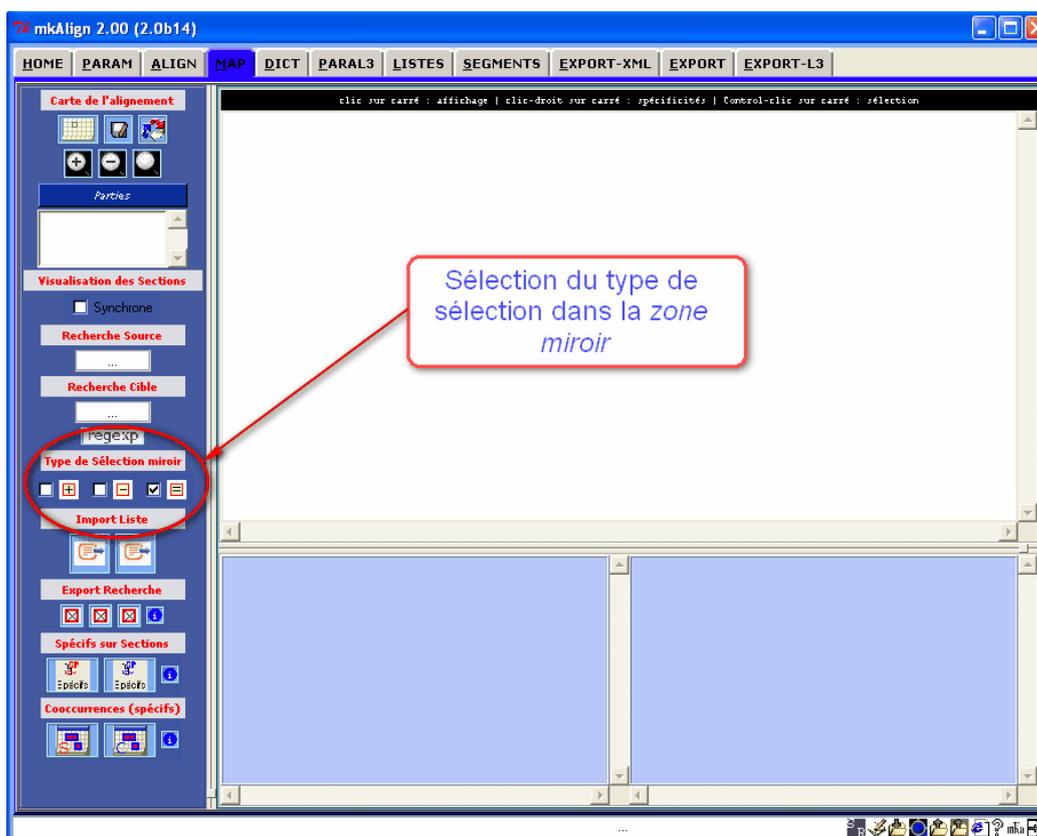


Figure 23 (1) : **mkAlign** « Au-delà du miroir »

Cette expérience est possible via **mkAlign**. Un mode (*expert*) de recherche dans la *zone miroir* de la carte de l'alignement est disponible (onglet **MAP**). L'item « *Type de sélection miroir* » permet de sélectionner 3 types de choix.

A. « Le mode »

C'est en fait le « **mode normal** » déjà présenté : si on lance une recherche dans le volet **SOURCE**, on obtient dans la carte de l'alignement :

- l'affichage des sections dans le volet **SOURCE** contenant la forme visée (les carrés sont marqués d'une croix et « cerclés » d'un trait **rouge gras**)
- et l'affichage de la *zone miroir* dans le volet **CIBLE** (les carrés sont « cerclés » d'un trait **rouge fin**)

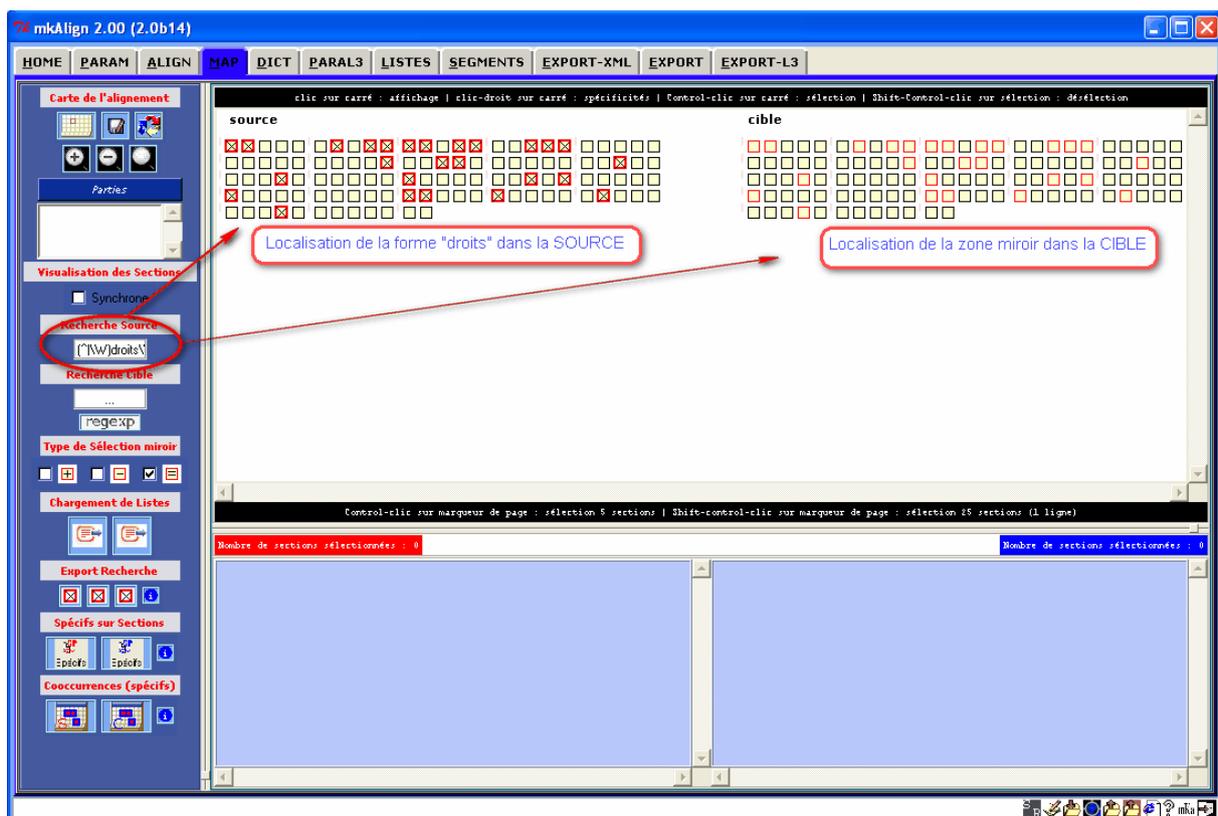


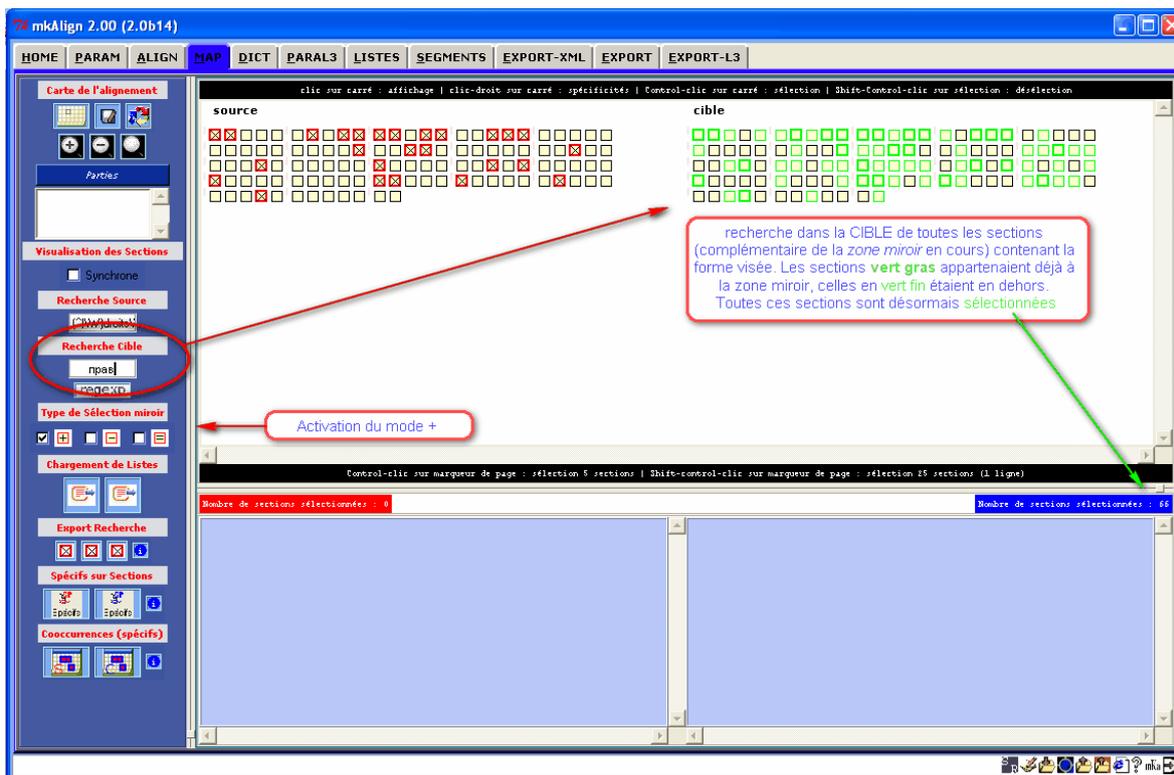
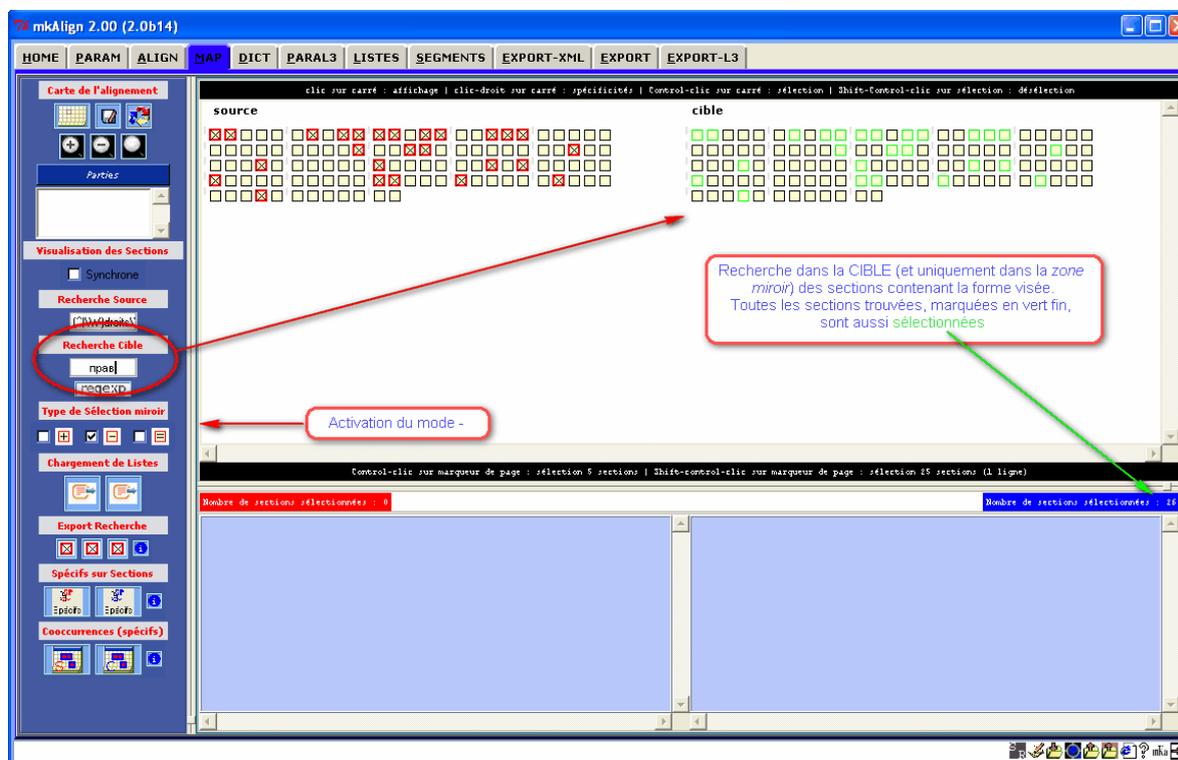
Figure 23 (2) : **mkAlign** « Au-delà du miroir »

Si on lance ensuite une recherche dans le volet **CIBLE**, le même phénomène se produit sur le volet **CIBLE** (*Remarque* : les croix initiales dans le volet **SOURCE** restent visibles).

B. « Le mode¹³ »

Si ce mode est activé (après avoir fait une recherche dans le volet **SOURCE**), on peut rechercher une forme dans le volet **CIBLE**, le résultat affiché donnera à voir les sections du volet **CIBLE** correspondant à la *zone miroir* déjà visible (même cerclage **rouge fin** pour les sections ne contenant pas la forme visée et cerclage **vert gras** pour celles la contenant) en y ajoutant toutes les sections **CIBLE** contenant la forme visée (les carrés sont cette fois de couleur **verte** et il n'y a pas de marquage de la présence de la forme par une croix). Toutes les sections colorées de vert sont aussi sélectionnées (on peut y lancer un calcul de spécificités par exemple).

¹³ A priori, il convient d'activer ce mode après avoir fait une recherche en « mode = »

Figure 23 (3) : *mkAlign* « Au-delà du miroir »C. « Le mode¹⁴  »Figure 23 (4) : *mkAlign* « Au-delà du miroir »¹⁴ Cf note précédente.

Si ce mode est activé (après avoir fait un recherche dans le volet **SOURCE**), on peut rechercher une forme dans le volet **CIBLE**, le résultat affiché donnera à voir les seules sections du volet **CIBLE** appartenant à la *zone miroir* initiale et contenant la forme visée (ce qui peut donc correspondre à une sélection de sections moins importante que la *zone miroir* initialement présente), les carrés sont cette fois-ci de couleur **vert fin** et il n'y a pas de marquage de la présence par une croix. Toutes les sections colorées de vert sont aussi sélectionnées (on peut y lancer un calcul de spécificités par exemple).

9 Concordances

L'onglet **CONCORDANCE** permet de construire des concordances sur le volet **SOURCE** et sur le volet **CIBLE** de l'alignement en cours. On peut aussi exporter une bi-concordance au format HTML.

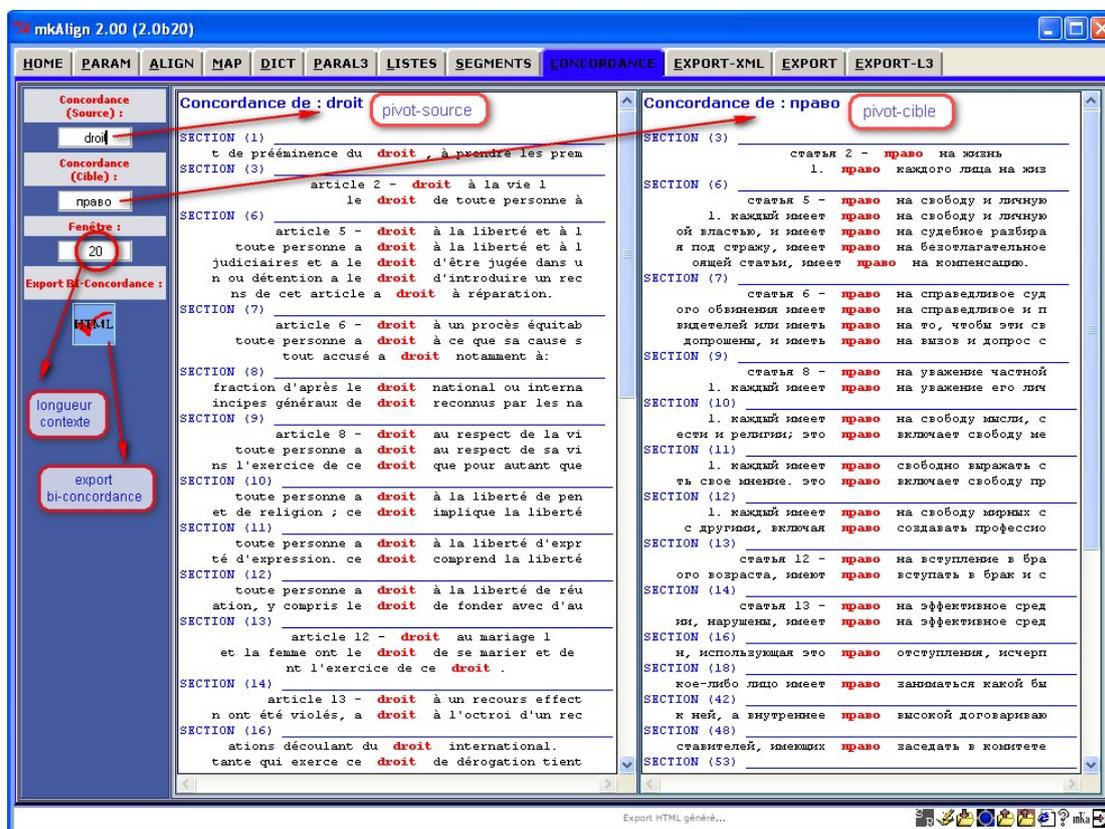


Figure 24 (1) : *mkAlign* Onglet **CONCORDANCE**, édition d'une concordance

Section	Contexte Source	Contexte Cible
1	t de prééminence du droit , à prendre les prem	
3	article 2 - droit à la vie l le droit de toute personne à	статья 2 - право на жизнь 1. право каждого лица на жиз
6	article 5 - droit à la liberté et à l toute personne a droit à la liberté et à l judiciaires et a le droit d'être jugée dans u n ou détention a le droit d'introduire un rec ns de cet article a droit à réparation.	статья 5 - право на свободу и личную 1. каждый имеет право на свободу и личную ой властью, и имеет право на судебное разбира я под стражу, имеет право на безотлагательное ошей статьи, имеет право на компенсацию.
7	article 6 - droit à un procès équitab toute personne a droit à ce que sa cause s tout accusé a droit notamment à:	статья 6 - право на справедливое суд ого обвинения имеет право на справедливое и п видетелей или иметь право на то, чтобы эти св допрошены, и иметь право на вызов и допрос с
8	fraction d'après le droit national ou interna incipes généraux de droit reconnus par les na	
9	article 8 - droit au respect de la vi toute personne a droit au respect de sa vi ns l'exercice de ce droit que pour autant que	статья 8 - право на уважение частной 1. каждый имеет право на уважение его лич
10	toute personne a droit à la liberté de pen et de religion ; ce droit implique la liberté	1. каждый имеет право на свободу мысли, с ести и религии; это право включает свободу ме
11	toute personne a droit à la liberté d'expr té d'expression. ce droit comprend la liberté	1. каждый имеет право свободно выражать с ть свое мнение. это право включает свободу пр
12	toute personne a droit à la liberté de réu ation, y compris le droit de fonder avec d'au	1. каждый имеет право на свободу мирных с с другими, включая право создавать профессио
13	article 12 - droit au mariage l et la femme ont le droit de se marier et de nt l'exercice de ce droit .	статья 12 - право на вступление в бра ого возраста, имеют право вступать в брак и с
14	article 13 - droit à un recours effect n ont été violés, a droit à l'octroi d'un rec	статья 13 - право на эффективное сред и, нарушены, имеет право на эффективное сред
16	ations découlant du droit international. tante qui exerce ce droit de dérogation tient	н, используя эту право отступления, исчерп

Figure 24 (2) : *mkAlign* exportation bi-concordance

10 Importation de listes

L'onglet **LISTES** dispose de 2 boutons  permettant d'importer des listes de formes graphiques (mots ou segments). Cette liste peut ensuite être utilisée pour la recherche dans le corpus *par le biais* de sa représentation cartographique.

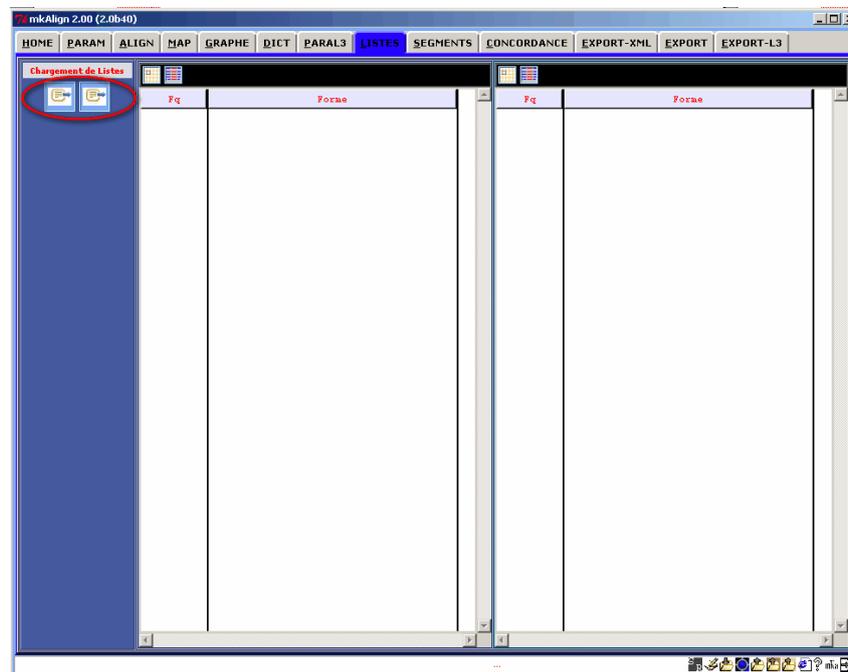


Figure 25 : *mkAlign* Onglet **LISTES**, importation de listes

L'importation d'une liste *source* ou *cible* déclenche l'importation de cette liste dans l'onglet **LISTES**.

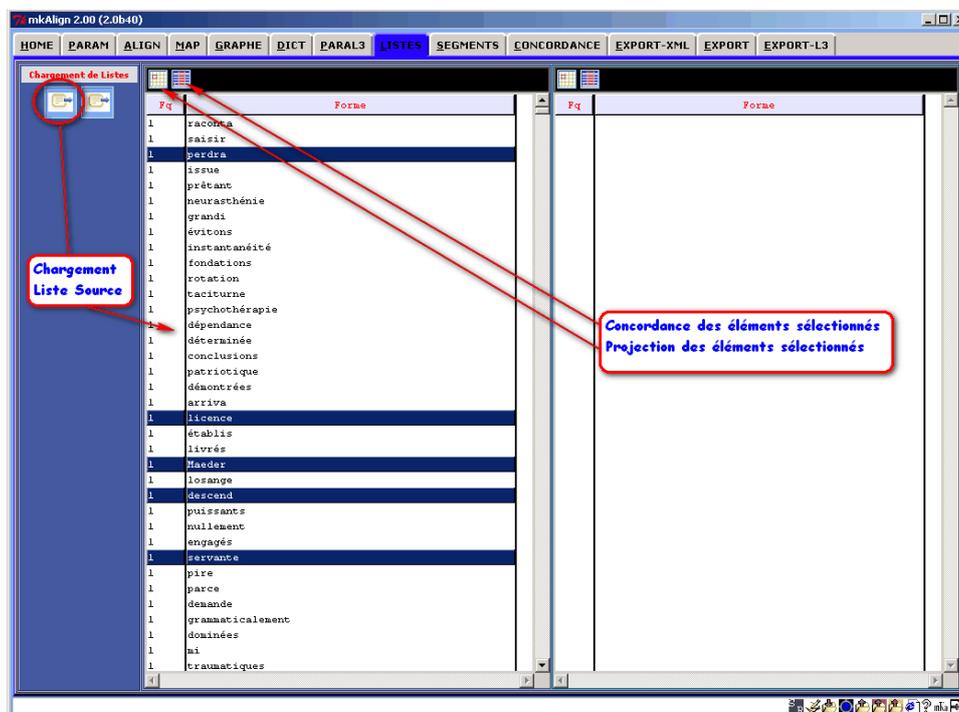


Figure 26 : *mkAlign* Onglet **LISTES**, édition de listes importées

11 Segments répétés

La fenêtre principale dispose d'un bouton  permettant de calculer les segments répétés des fichiers **SOURCE** et **CIBLE** puis d'afficher les 2 listes dans l'onglet **SEGMENTS**.

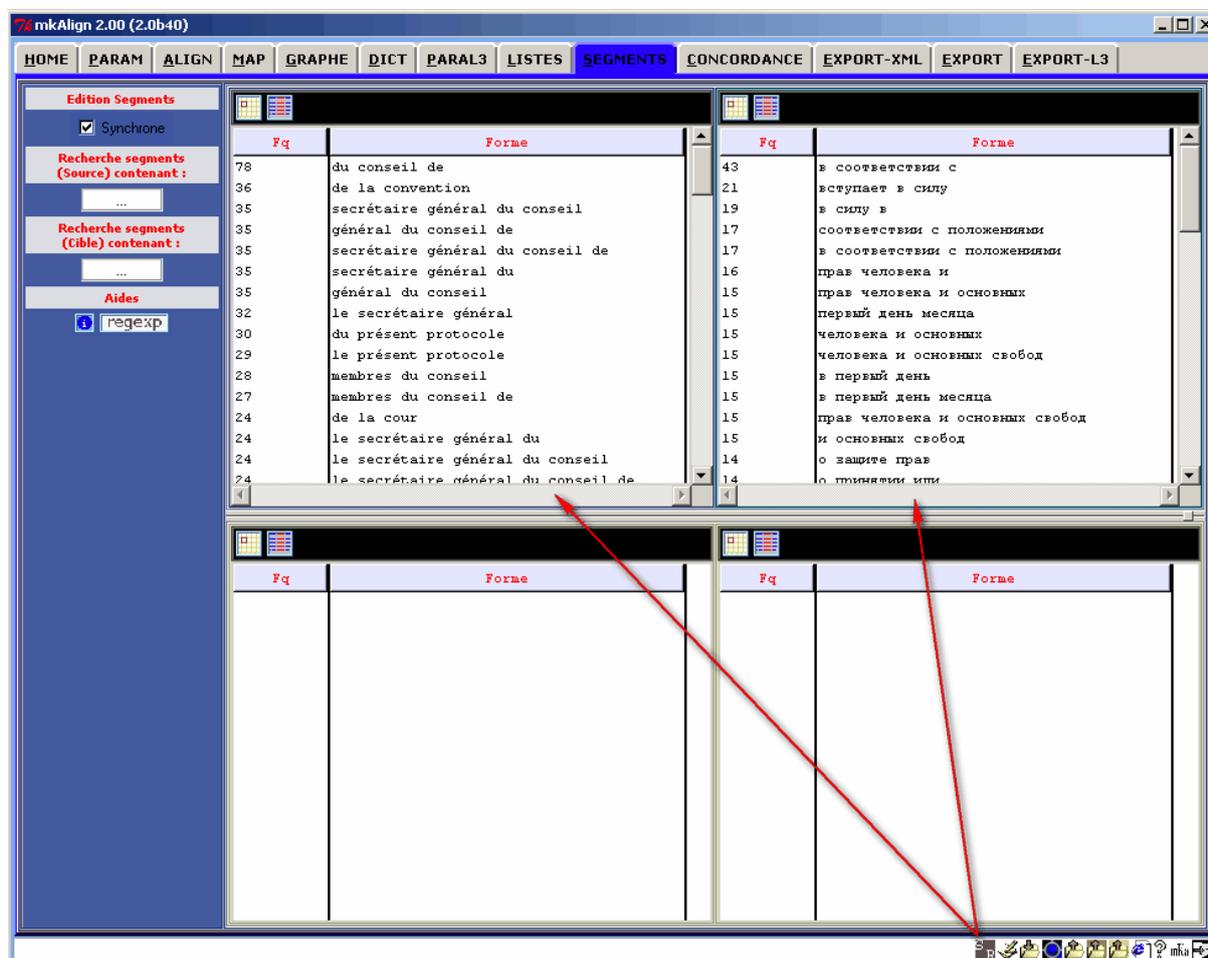


Figure 27 (1) : *mkAlign* Onglet **SEGMENTS**

Remarque : Les segments apparaissent dans les 2 volets supérieurs de l'onglet **SEGMENTS**.

Le calcul des segments est paramétrable (cf onglet **PARAM**) : on peut choisir les délimiteurs à utiliser, la longueur minimale des segments à rechercher ou leur fréquence minimale. Si ces 2 derniers paramètres ne sont pas renseignés tous les segments de longueur supérieure ou égale à 2 sont recherchés.

Longueur Segment (min) pour SR (vide pour longueur quelconque).....	<input type="text"/>
Fréquence Segment (min) pour SR (vide pour les chercher tous).....	<input type="text"/>

Figure 27 (2) : *mkAlign* Onglet **PARAM**, paramétrage du calcul des segments répétés

On peut ensuite sélectionner un (ou plusieurs) segments et déclencher (1) l'affichage du segment dans la représentation cartographique de l'alignement (cf onglet **MAP**), (2) l'affichage d'une concordance de ce segment :

Un *clic gauche* sur un des mots d'un segment donné déclenche la recherche de tous les segments contenant ce mot. Cette sélection de segments est affichée dans la partie inférieure de l'onglet **SEGMENTS**. Un *clic droit* sur un des segments déclenche le même processus que ci-dessus.

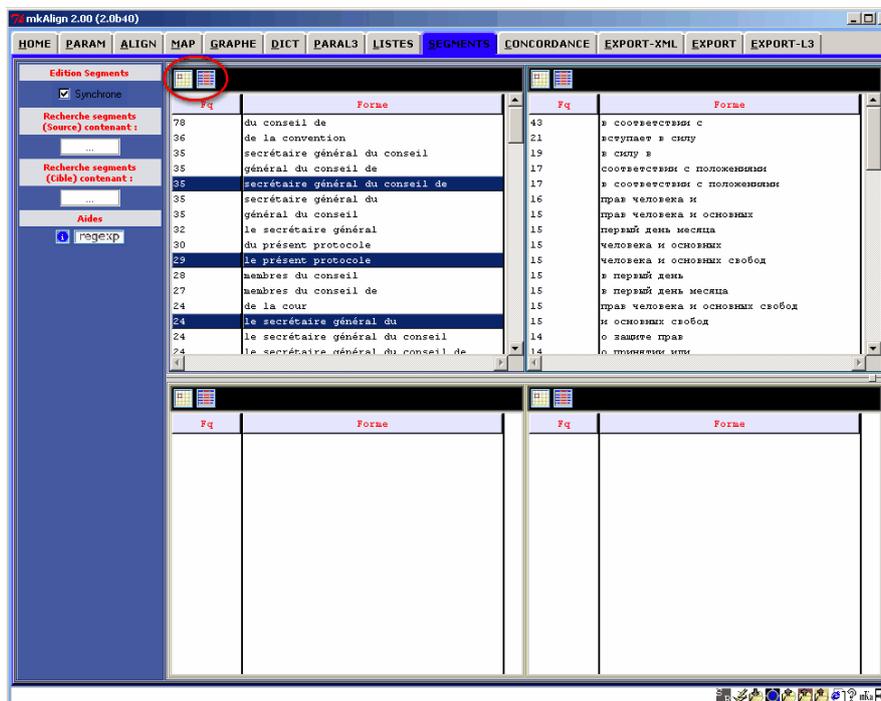


Figure 27 (3) : *mkAlign* Onglet **SEGMENTS**, sélection de segments

On peut aussi rechercher des segments contenant une forme donnée en utilisant les zones de saisie disponibles sur la gauche de l'onglet **SEGMENTS**.

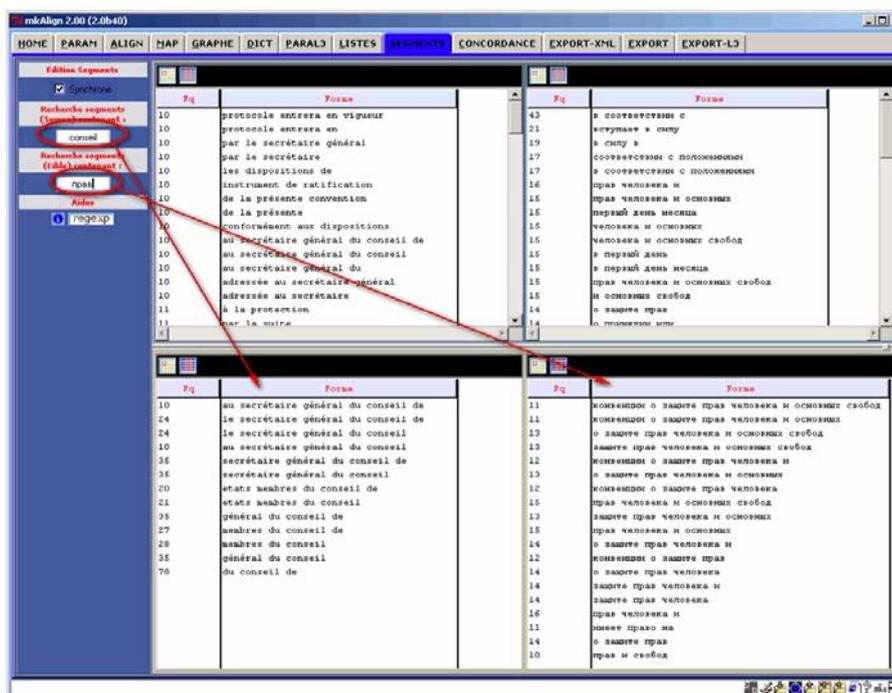


Figure 27 (4) : *mkAlign* Onglet **SEGMENTS**, sélection de segments

12 Cooccurrences – Poly-Cooccurrences

L'onglet **COOCS** permet de lancer des modules de calculs de cooccurrences et de poly-cooccurrences sur le volet **SOURCE** et sur le volet **CIBLE** de l'alignement en cours.

Les modules disponibles s'inscrivent dans la démarche mise en œuvre dans le travail de William Martinez (2002, 2003, 2006).

- Une *cooccurrence* désigne l'apparition de deux mots en même temps et dans le même contexte.

Le module de cooccurrences mis en œuvre prend appui sur l'alignement en cours, les contextes dans lesquels on examine la co-présence sont donc ceux qui coïncident aux différentes cellules dans l'éditeur d'alignement (ou aux sections dans la carte des sections)

- Le terme *poly-cooccurrence* désigne les attractions lexicales au-delà de la cooccurrence binaire.

Le module de poly-cooccurrences intégré reprend l'algorithme décrit dans [Martinez, 2006] :

- On calcule pour le pôle A les cooccurrents spécifiques B, C et D
- Dans leurs contextes communs, on calcule pour les pôles A+B les cooccurrents spécifiques E et F
- Les pôles A+B+E ont pour cooccurrent spécifique H
- Les pôles A+B+E+H n'ont pas de cooccurrent spécifique et l'exploration s'interrompt pour ce chemin
- Les pôles A+B+F ont pour cooccurrents spécifiques I, etc.
- Durant l'exploration, différents filtres conditionnent l'épuisement des explorations contextuelles et réduisent le bruit dans les résultats pour privilégier l'information la plus spécifique : seuils maximaux de fréquence et de spécificité du cooccurrent.

Par défaut, le calcul se fait en travaillant en parallèle sur les sections d'un volet (source ou cible) contenant le pôle visé et sans tenir compte de ce qui se passe dans le volet d'en face.

On peut aussi ne retenir que les sections pour lesquelles les pôles source et cible sont présents en même temps dans les sections alignées. Pour lancer cette option, il suffit de cocher l'option disponible dans l'onglet COOCS (*Pôles alignés uniquement*).

Paramètres :	
Pôle Source	droit
Pôle Cible	right
Fq-Cooc source	2
Fq-Cooc cible	2
Seuil source	5
Seuil cible	5
<input checked="" type="checkbox"/> Pôles alignés uniquement	
	

La figure qui suit illustre la démarche mise en œuvre pour construire des listes de cooccurrences puis des graphes donnant à voir les poly-cooccurents :

paramétrage des calculs :

1. Choix de ou des pôles
2. Choix des paramètres Fq-Cooc et Seuil pour les pôles visés

1. Lancement du module de cooccurrences

2. Lancement du module de poly-cooccurrences

résultat : un graphe

résultat : la (ou les) liste des cooccurents

les formes obtenues sont réutilisables dans les modules déjà disponibles :

- ventilation sur la carte des sections
- concordance
- ventilation sur le graphe de partition induite par les sections

Forme	Fq	co-freq	specif
et	185	94	5.2
protection	15	15	7.5
liberté	21	19	7.3
a	52	34	5.6
y	18	15	5.1
loi	26	23	6.1
personne	28	26	10.1

Forme	Fq	co-freq	specif
law	37	22	5.8
detection	13	11	5.8
or	217	85	6.1
person	20	15	6.3
his	33	25	9.6
public	21	17	7.7
lawful	12	12	8.1
interests	12	10	5.3
exercice	12	10	5.3
freedom	14	13	7.6
everyone	19	18	10.5

Figure 28 : mkAlign Onglet COOCS, calcul de cooccurents et de poly-cooccurents

Les résultats produits par ce calcul sont insérés dans l'onglet **COOCS** et sont associés à des processus de calcul déjà disponibles dans d'autres onglets :

- le raccourci *clic-droit* sur un nœud du graphe déclenche le calcul de la concordance de la forme associée au nœud visé
- le raccourci *control-clic* sur un nœud permet de le sélectionner

Des nœuds du graphe étant sélectionnés, on peut ensuite les projeter sur la carte des sections :



Si l'option « Global » est activée, les sections mises au jour seront celles contenant l'ensemble des formes visées (ET logique) : on peut ainsi visualiser en contexte les chemins de polycooccurrences calculés précédemment.

13 Graphiques : ventilation / accroissement de vocabulaire

13.1 Ventilation de formes graphiques

On peut sélectionner des formes dans les dictionnaires (onglet **DIC**) et construire la ventilation de la ou des unités textuelles choisies dans la partition du corpus induite par les sections de l'alignement.

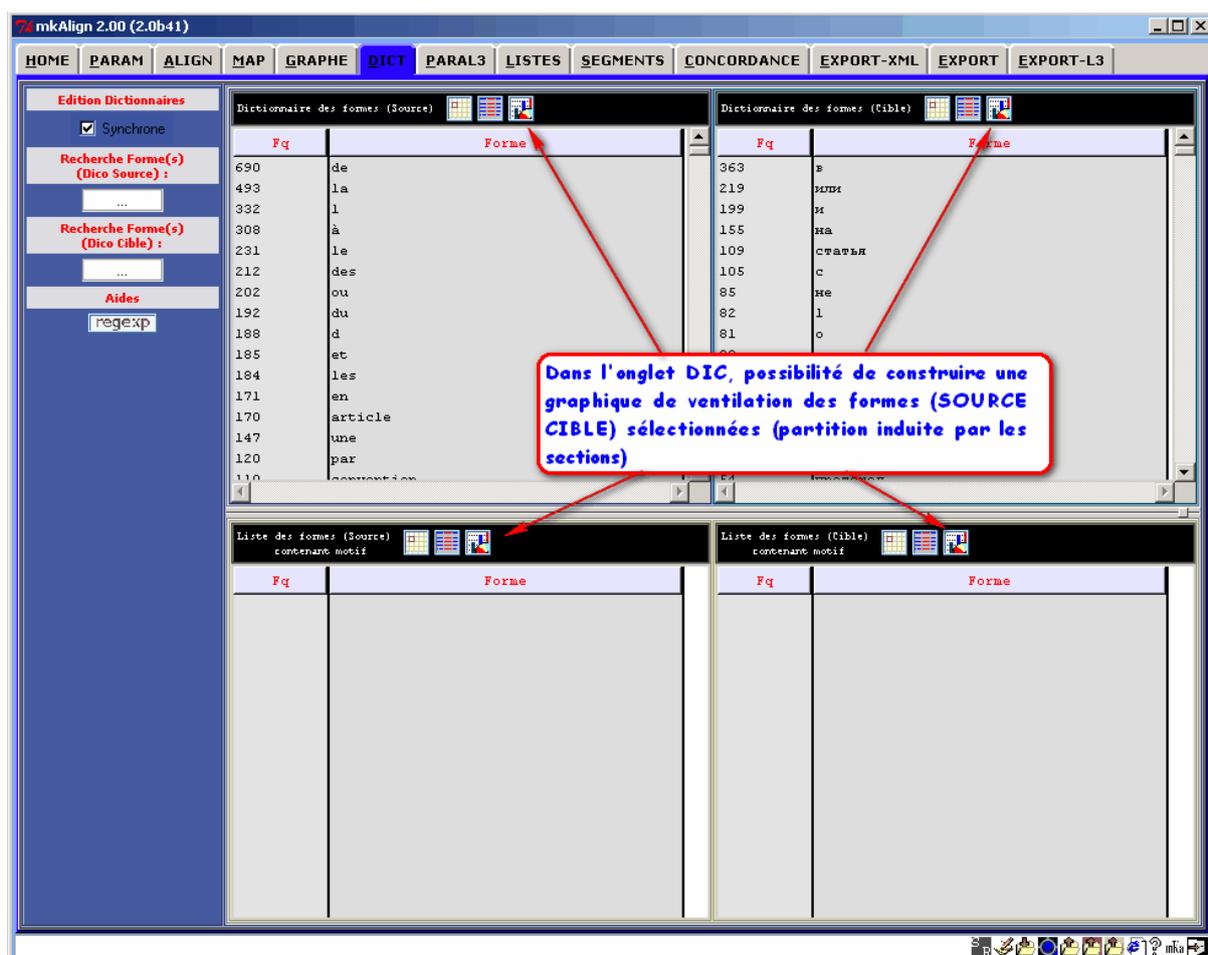


Figure 29 (1) : *mkAlign* Onglet **DIC**, ventilation de formes

On peut sélectionner une ou plusieurs formes dans le volet **SOURCE** (supérieur) et une ou plusieurs formes dans le volet **CIBLE** (supérieur) : la ventilation donnera à voir en parallèle la ventilation des formes choisies dans les 2 volets.

Idem pour les volets **SOURCE CIBLE** inférieurs.

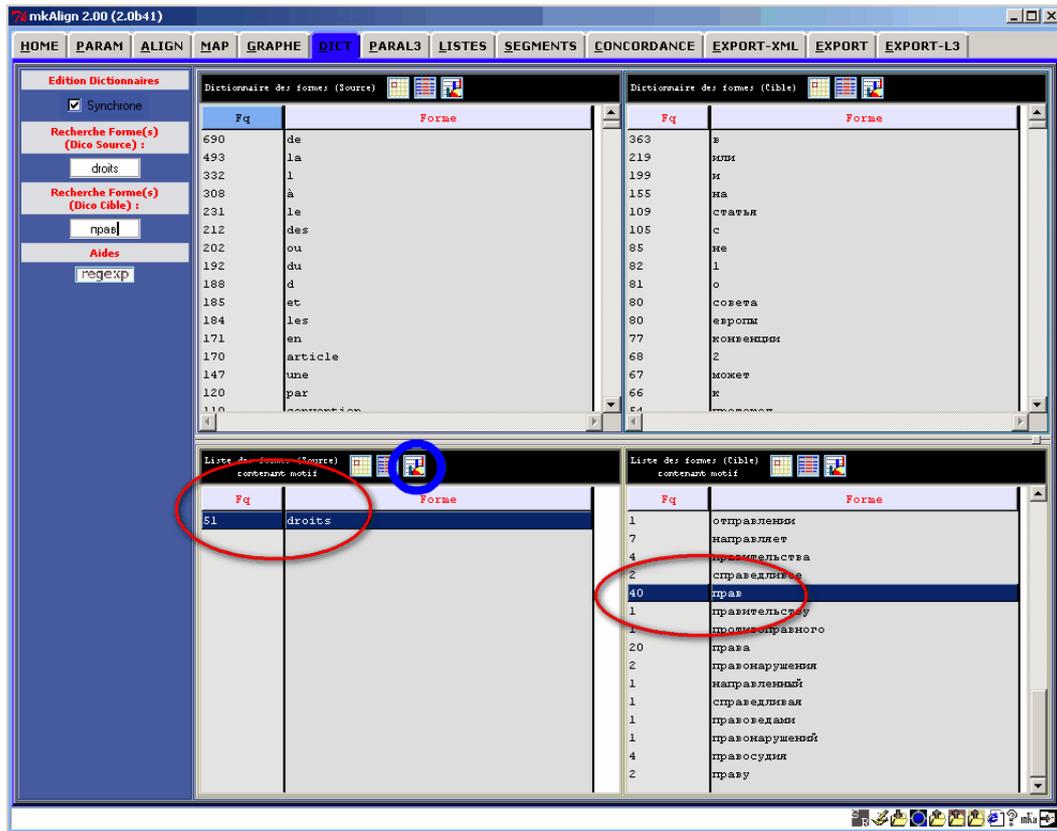


Figure 29 (2) : mkAlign Onglet DIC, sélection de formes pour ventilation

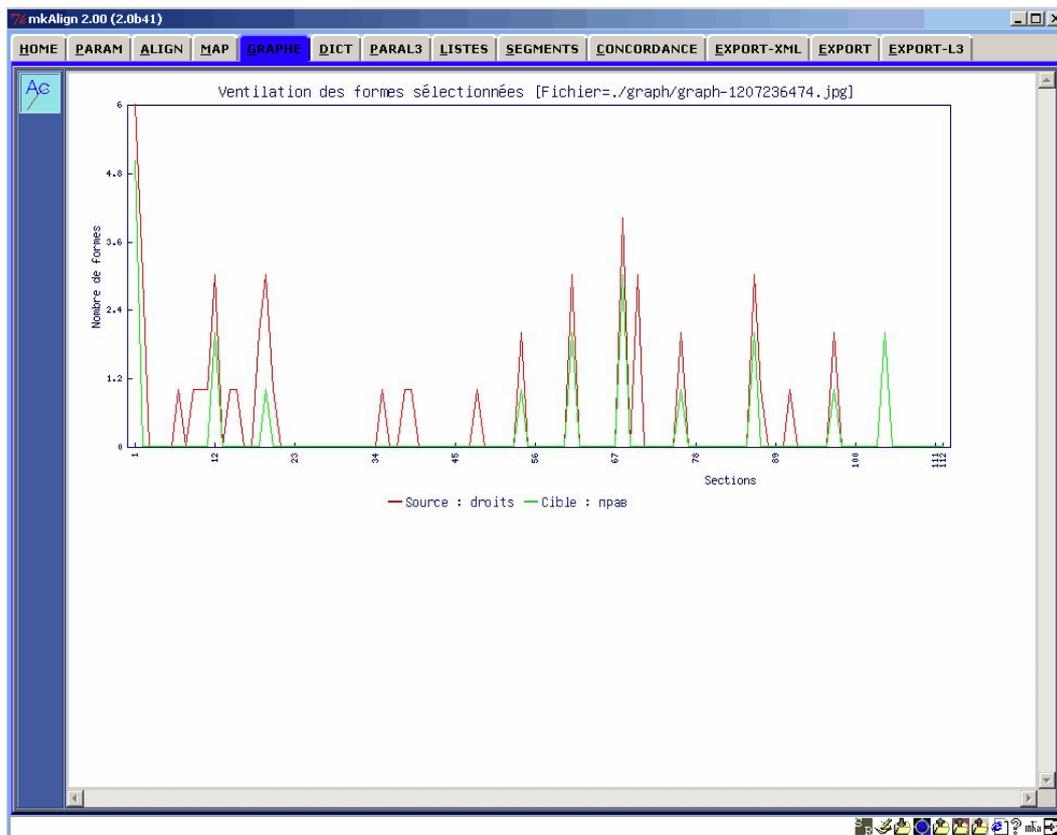


Figure 29 (3) : mkAlign Onglet GRAPHE, ventilation des formes sélectionnées

13.2 Courbe d'accroissement de vocabulaire

On dispose dans l'onglet **GRAPHE** d'un bouton permettant de construire la courbe d'accroissement du vocabulaire sur les volets **SOURCE** et **CIBLE**.



Figure 30 : *mkAlign* Onglet **GRAPHE**, courbe d'accroissement du vocabulaire

14 Exportation des traitements

14.1 Export vers Lexico3

Il est possible d'exporter l'alignement au format **Lexico3** via le bouton **EXPL3** , l'onglet **EXPORT-L3** contiendra le résultat de cette exportation. Le fichier contenant le résultat de cette exportation se trouve dans le sous-répertoire EXPORT/TXT du répertoire de travail du programme et porte le nom MK-ALIGN-FOR-LEX3-date.txt.



Figure 31 : mkAlign Onglet EXPORT-L3

14.2 Export au format XML

Il est possible d'exporter l'alignement au format XML *via* le bouton **EXPXML** , l'onglet **EXPORT-XML** contiendra la version XML produite. Le fichier contenant le résultat de cette exportation se trouve dans le sous-répertoire EXPORT/XML du répertoire de travail du programme et porte le nom MK-BUILT-ALIGN-*date*.txt.

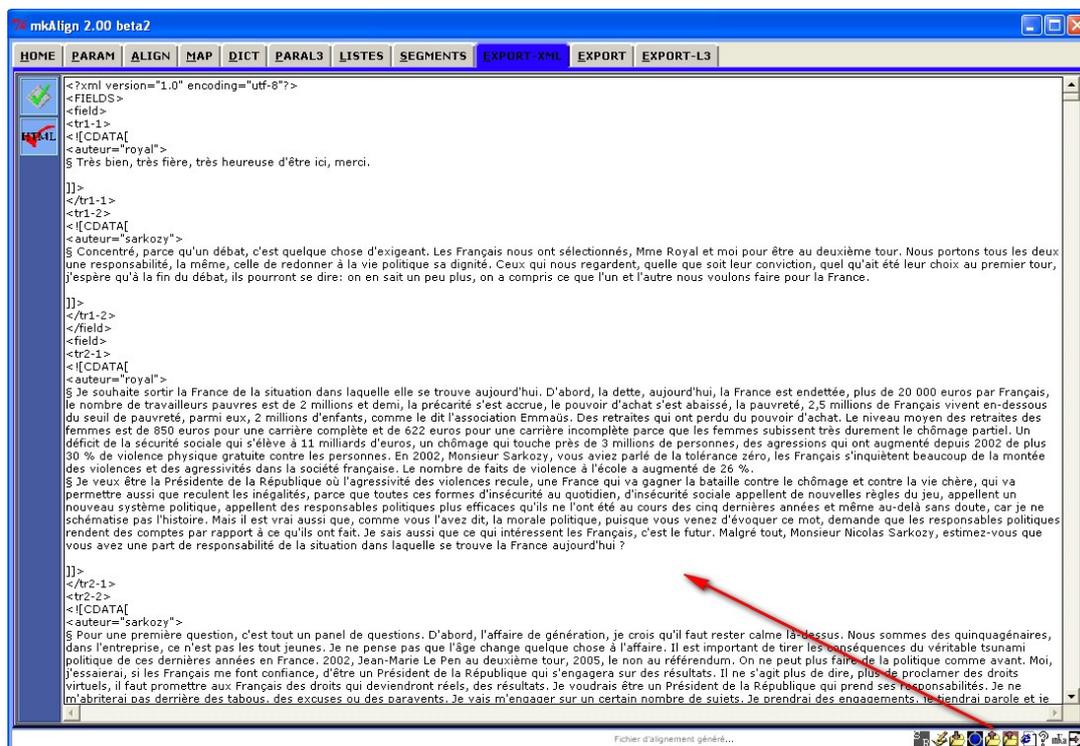


Figure 32 : *mkAlign* Onglet EXPORT-XML

Le schéma du fichier d'export XML est décrit dans la figure suivante :

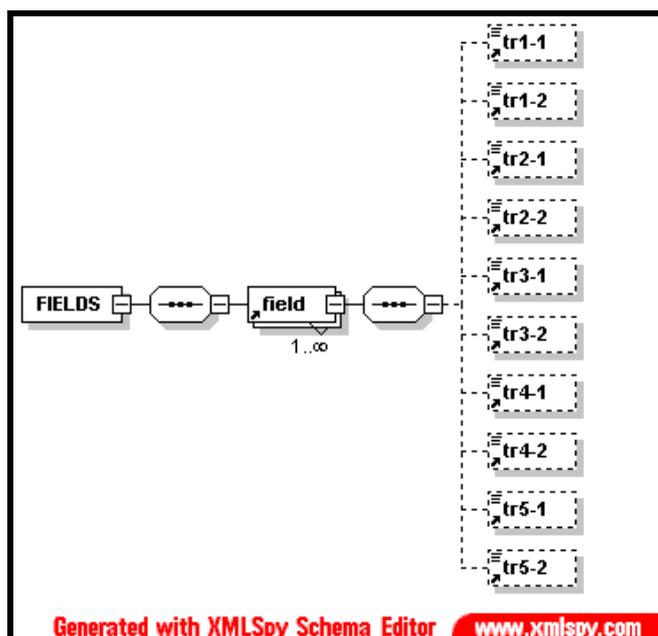


Figure 33 : *mkAlign* Schéma Export XML

14.3 Export au format XML pour sauvegarder une session de travail

Au moment de l'export XML, un autre fichier d'export est construit, le nom de ce fichier est du type : EXPORT-MKALIGN-date.xml (dans le dossier EXPORT/XML). Ce fichier enregistre l'état complet de l'alignement et pourra être réimporté dans **mkAlign** après redémarrage du programme (*cf infra*).

14.4 Export complet au format HTML

On dispose aussi d'un bouton  permettant de visualiser l'état de l'alignement au format HTML *via* le navigateur internet défini sur la machine de travail.

CLA²T [U. DE PARIS 3, Sorbonne nouvelle]
mkAlign

<p>§ <mkALIGN="SOURCE-0000000001"> <partie=1> <conference=01Introduction> Introduction</p> <p>§ <mkALIGN="SOURCE-0000000002"> <p=001> J&apos;ignore combien d&apos;entre vous connaissent la psychanalyse par leurs lectures ou par oui-dire. Mais le titre même de ces leçons : Introduction à la Psychanalyse, m&apos;impose l&apos;obligation de faire comme si vous ne saviez rien sur ce sujet et comme si vous aviez besoin d&apos;être initiés à ses premiers éléments.</p> <p>§ <mkALIGN="SOURCE-0000000003"> <p=002> Je dois toutefois supposer que vous savez que la psychanalyse est un procédé de traitement médical de personnes atteintes de maladies nerveuses. Ceci dit, je puis vous montrer aussitôt sur un exemple que les choses ne se passent pas ici comme dans les autres branches de la médecine, qu&apos;elles s&apos;y passent même d&apos;une façon tout à fait contraire. Généralement, lorsque nous soumettons un malade à une technique médicale nouvelle pour lui, nous nous appliquons à en diminuer à ses yeux les inconvénients et à lui donner toutes les assurances possibles quant au succès du traitement. Je crois que nous avons raison de le faire, car en procédant ainsi nous augmentons effectivement les chances de succès. Mais on procède tout autrement, lorsqu&apos;on soumet un névrotique au traitement psychanalytique. Nous le mettons alors au courant des difficultés de la méthode, de sa durée, des efforts et des sacrifices qu&apos;elle exige ; et quant au résultat, nous lui disons que nous ne pouvons rien promettre, qu&apos;il dépendra de la manière dont se comportera le malade lui-même, de son intelligence, de son obéissance, de sa patience. Il va sans dire que de bonnes raisons, dont vous saisirez peut-être l&apos;importance plus tard, nous dicent cette conduite inaccoutumée.</p> <p>§ <mkALIGN="SOURCE-0000000004"> <p=003> Je vous prie de ne pas m&apos;en vouloir si le commence par vous traiter comme ces malades névrotiques. Je vous déconseille tout simplement de</p>	<p>§ <mkALIGN="CIBLE-0000000001"> <partie=1> <conference= 01EINLEITUNG> EINLEITUNG</p> <p>§ <mkALIGN="CIBLE-0000000002"> <p=001> <page=41> Meine Damen und Herren! Ich weiß nicht, wieviel die einzelnen von Ihnen aus ihrer Lektüre oder vom Hörensagen über die Psychoanalyse wissen. Ich bin aber durch den Wortlaut meiner Ankündigung - Elementare Einführung in die Psychoanalyse - verpflichtet, Sie so zu behandeln, als wülten Sie nichts und bedürften einer ersten Unterweisung.</p> <p>§ <mkALIGN="CIBLE-0000000003"> <p=002> Soviel darf ich allerdings voraussetzen, daß Sie wissen, die Psychoanalyse sei ein Verfahren, wie man nervös Kranke ärztlich behandelt, und da kann ich Ihnen gleich ein Beispiel dafür geben, wie auf diesem Gebiet so manches anders, oft geradezu verkehrt, vor sich geht als sonst in der Medizin. Wenn wir sonst einen Kranken einer ihm neuen ärztlichen Technik unterziehen, so werden wir in der Regel die Beschwerden derselben vor ihm herabsetzen und ihm zuversichtliche Versprechungen wegen des Erfolges der Behandlung geben. Ich meine, wir sind berechtigt dazu, denn wir steigern durch solches Benehmen die Wahrscheinlichkeit des Erfolges. Wenn wir aber einen Neurotiker in psychoanalytische Behandlung nehmen, so verfahren wir anders. Wir halten ihm die Schwierigkeiten der Methode vor, ihre Zeitdauer, die Anstrengungen und die Opfer, die sie kostet, und was den Erfolg anbelangt, so sagen wir, wir können ihn nicht sicher versprechen, er hänge von seinem Benehmen ab, von sein! em Verständnis, seiner Gefügigkeit, seiner Ausdauer. Wir haben natürlich gute Motive für ein anscheinend so verkehrtes Benehmen, in welche Sie vielleicht später einmal Einsicht gewinnen werden.</p>
--	--

Figure 34 : **mkAlign** Export HTML

SYLED/CLA²T

<http://syled.univ-paris3.fr>

14.5 Export partiel au format HTML

On dispose aussi dans l'onglet **MAP** de trois boutons  permettant d'exporter le résultat d'une requête de recherche de formes graphiques (exprimée sous la forme d'une expression régulière).

Dans la figure qui suit, l'onglet présente le résultat d'une requête de recherche de la forme "psychanalyse" dans la carte :

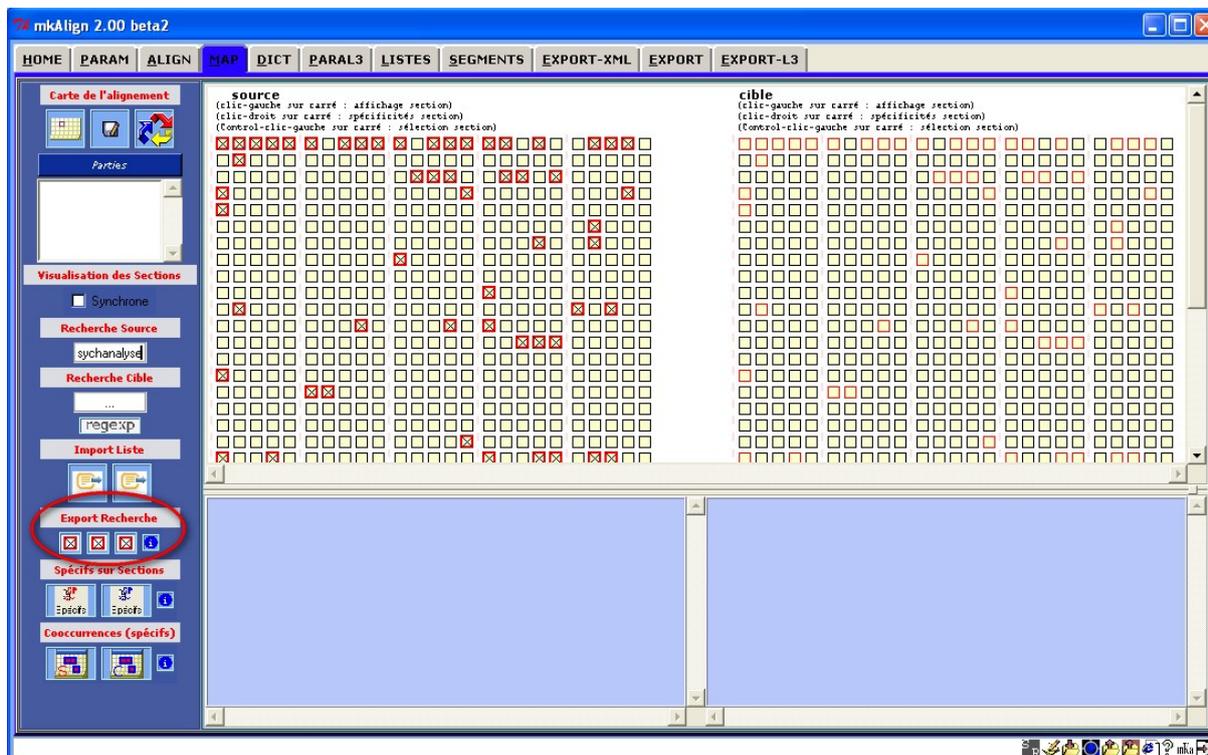


Figure 35 : *mkAlign* Export sélectif (sélection d'une forme)

1. L'activation du bouton d'export de la "Recherche Source" (bouton gauche de la zone cerclée ci-dessus) provoque la génération d'un fichier aligné (au format XHTML) regroupant uniquement les cellules du fichier *source* contenant la forme cherchée.
2. Une fonctionnalité similaire est disponible pour le fichier **CIBLE** (bouton du centre de la zone cerclée ci-dessus).

Le résultat de cet export est visible dans l'onglet **EXPORT-XML** et *via* le navigateur internet défini sur la machine de travail sous la forme suivante :

Dans cette sortie, la forme utilisée pour générer l'export est colorée en rouge (en suivant notre exemple précédent, c'est la forme "psychanalyse" qui est colorée ci-dessus).

Le nom de ce fichier est du type : MK-EXPORT-SEARCH-SOURCE-*date*.html (dans le répertoire EXPORT/HTML).

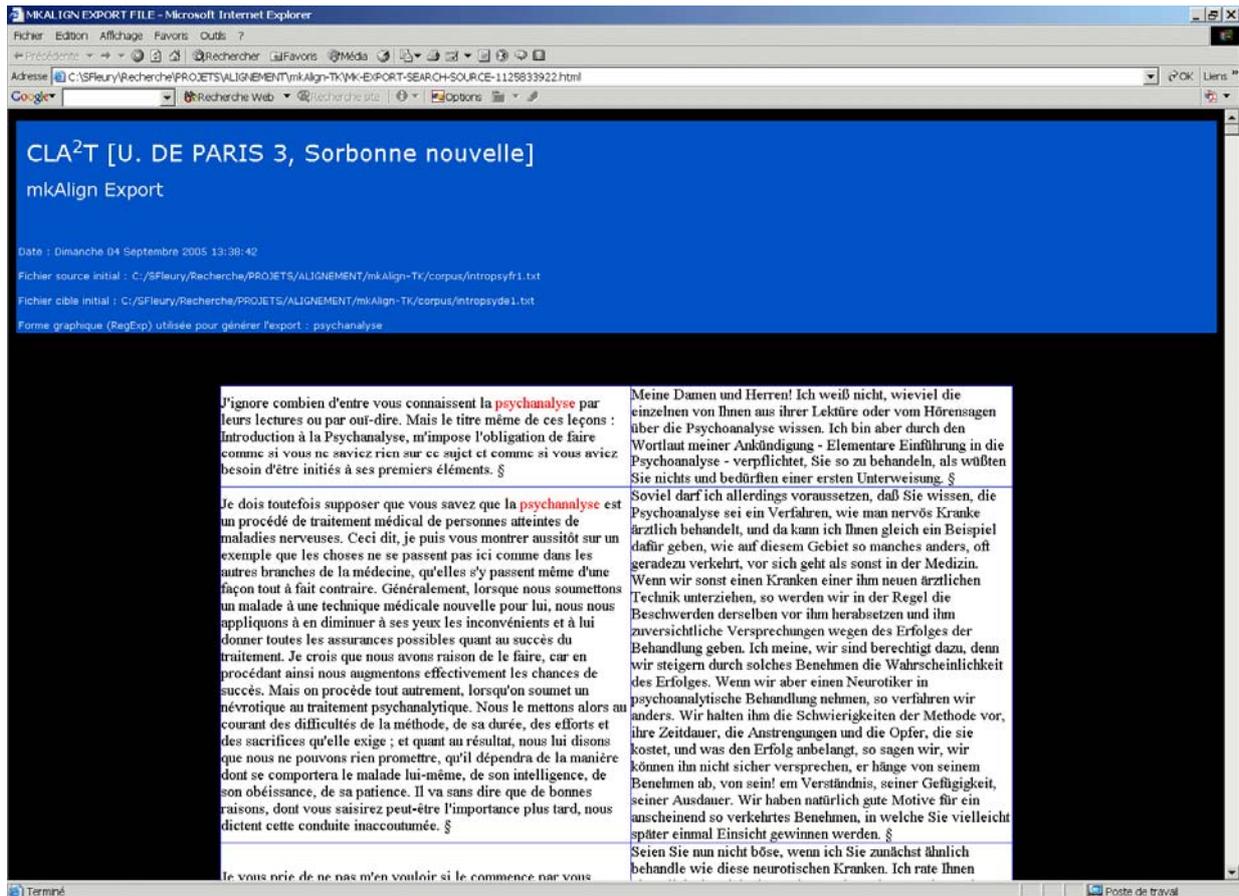


Figure 36 : *mkAlign* Résultat d'export sélectif - Fichier complet : [example-export.html](#)

3. Le troisième bouton d'export partiel permet :

1. d'exporter les seules zones bi-textuelles correspondant à la zone d'intersection entre les distributions d'une forme (source) cherchée et d'une forme (cible) cherchée
2. d'exporter les seules zones bi-textuelles dans lesquelles la forme (source) cherchée est présente dans une zone source et la forme (cible) cherchée n'est pas présente dans la zone cible associé
3. d'exporter les seules zones bi-textuelles dans lesquelles la forme (source) cherchée n'est pas présente dans une zone source et la forme (cible) cherchée est présente dans la zone cible associée

Ces trois rapports sont accessibles via un menu construit au moment de cet export (document HTML avec liens hypertextes donnant accès aux trois types d'export).

14.6 Export de bi-textes au format TXT

On dispose aussi d'une procédure permettant d'exporter des bi-textes correspondant d'une part aux cellules du fichier *source* contenant une forme cherchée sur ce volet et d'autre part aux cellules du fichier *cible* contenant une autre forme recherchée sur cet autre volet. Les 2 textes constituant ce bi-texte sont sauvegardés séparément. On peut ensuite les recharger dans **mkAlign**. La figure suivante illustre cette procédure d'export de bi-textes. Les cellules sélectionnées sur le fichier *source* contiennent la forme graphique « ich », les cellules sélectionnées sur le fichier *cible* contiennent la forme « je ».

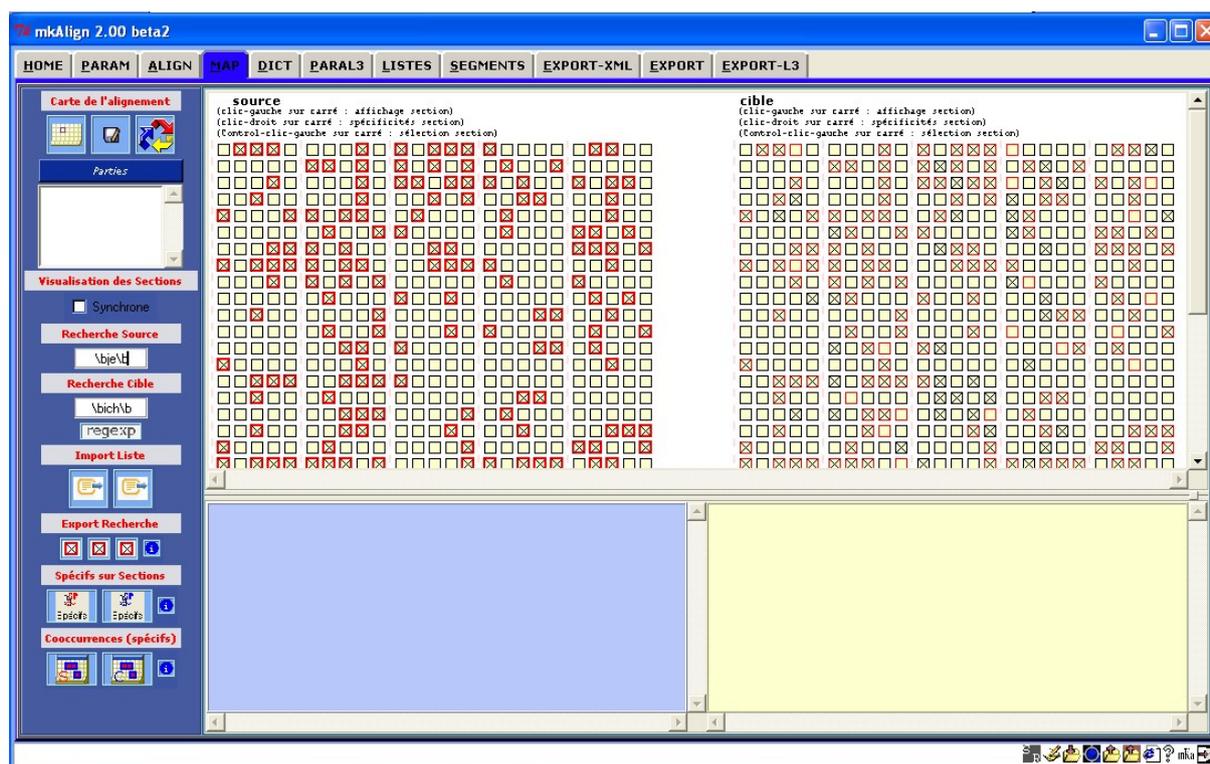


Figure 37 : **mkAlign**, Export de bi-textes

L'activation du bouton  déclenche la génération des 2 fichiers attendus : le fichier construit à partir du fichier *source* et ne contenant que les cellules (sur la gauche) marquées ci-dessus (*i.e.* les cellules contenant la forme graphique « ich ») et le fichier construit à partir du fichier *cible* et ne contenant que les cellules (sur la droite) marquées ci-dessus (*i.e.* les cellules contenant la forme graphique « je »).

14.7 Export de l'alignement au format TMX

On dispose dans l'onglet **ALIGN** d'une procédure permettant d'exporter un alignement construit avec **mkAlign** au format TMX. Le fichier produit est réimportable par la suite (cf chargement de fichiers). La figure suivante illustre ce processus :

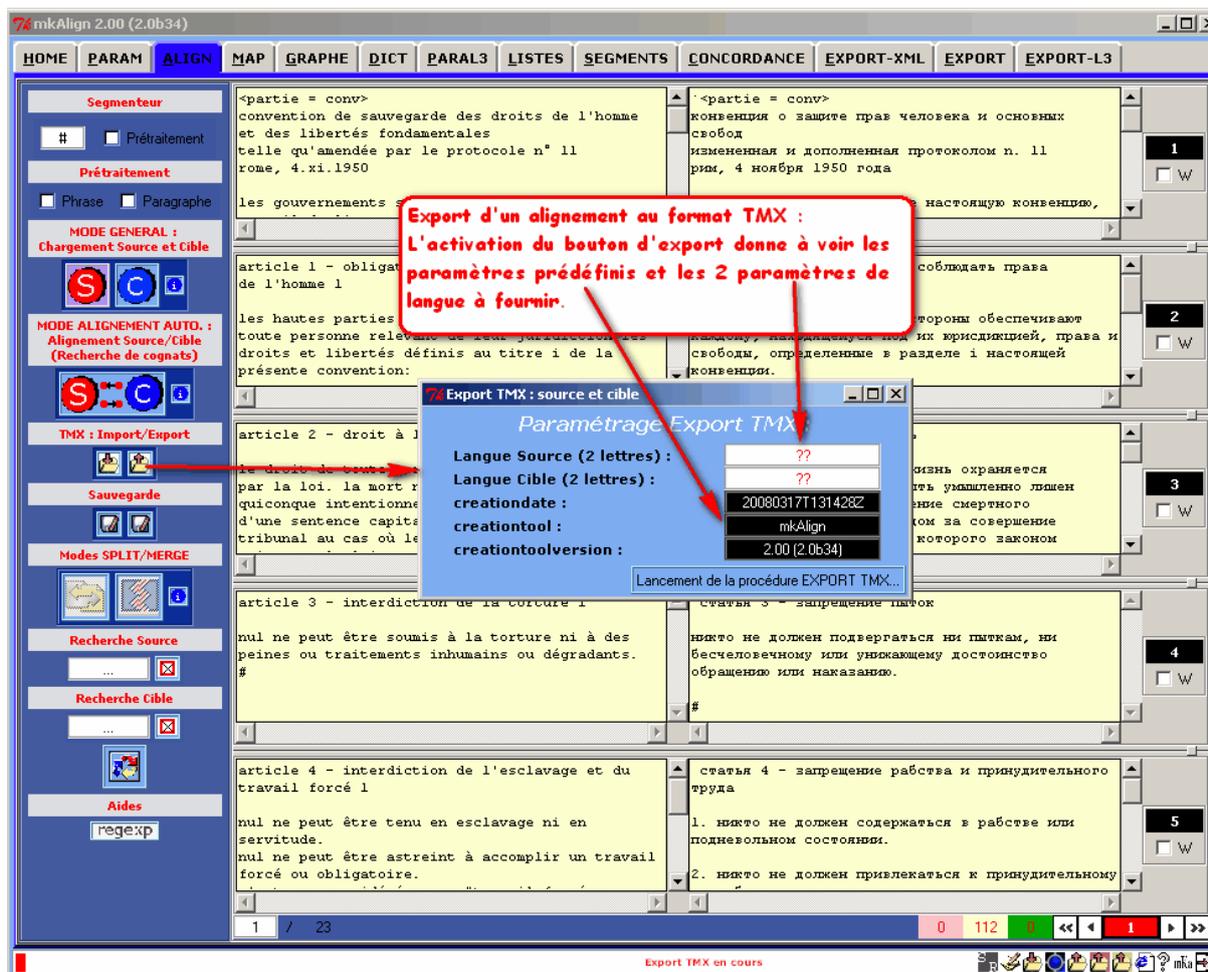


Figure 38 : **mkAlign**, Export de l'alignement au format TMX

Ce processus d'export au format TMX crée tout d'abord le fichier TMX (sous la forme : MK-EXPORT-TMX-date.tmx) dans le dossier **EXPORT/TMX** puis crée un fichier au format HTML (sous la forme : MK-EXPORT-TMX-date.html) résultat de la transformation du précédent par la feuille de styles XSLT fournie dans ce même dossier (styles-tmx.xsl).

On peut ensuite rééditer le fichier TMX dans un éditeur du type MSOFFICE-WORD en choisissant d'afficher ce document avec la même feuille de styles (ou une autre).

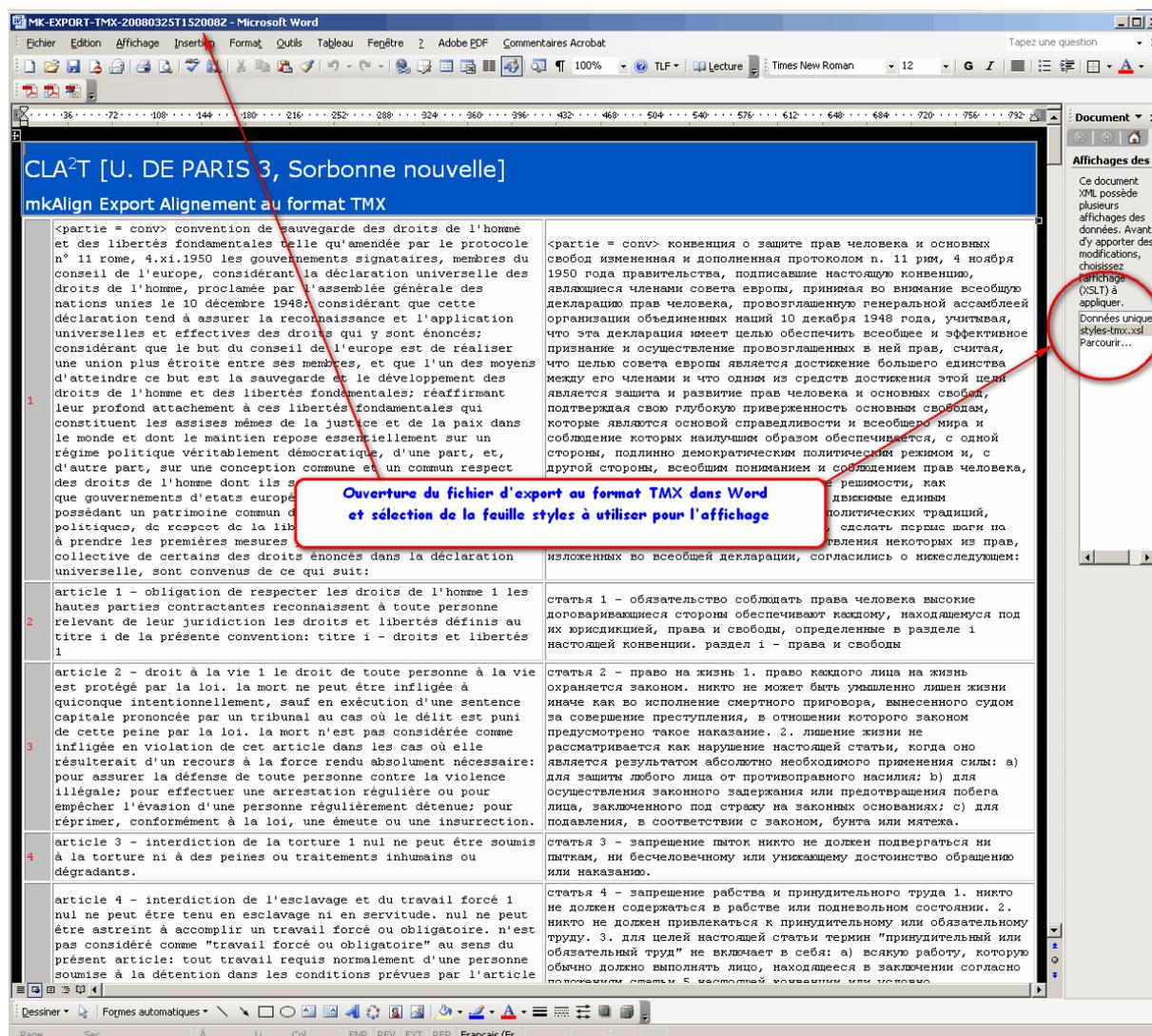


Figure 39 : **mkAlign**, Ouverture de l'alignement au format TMX dans Word avec feuille de styles

15 Import d'alignement

Il est possible d'importer  un alignement préalablement construit par **mkAlign**. Au moment de l'export XML présenté ci-dessus, un fichier d'export interne au programme est construit (cet export est chargé dans l'onglet **EXPORT**), le nom de ce fichier est du type : EXPORT-MKALIGN-*date*.xml .



Figure 40 : **mkAlign** Onglet **EXPORT**

L'importation de ce fichier reconstruit l'état complet de l'alignement préalablement construit dans **mkAlign**.

Le schéma du fichier d'export produit par **mkAlign** est décrit dans la figure ci-dessous :

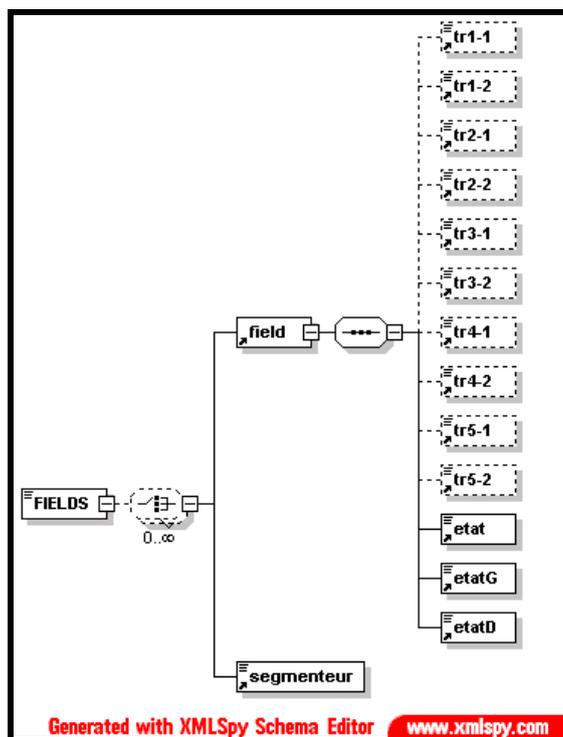


Figure 41 : **mkAlign** Schéma Import

Chacune des cellules alignées est décrite dans un élément `field`, cette description contient le contenu de chaque cellule et ses états d'édition (`lock`, `unlock`...)

16 Mise au jour de la variation dans les textes

16.1 Repérage de la variation

Ce processus s'appuie sur l'implémentation de la commande diff¹⁵ dans la bibliothèque Tk::DiffText¹⁶ (*composite widget for colored diffs*)



Figure 42 (1) : *mkAlign* Onglet Variation (paramètres et outils)

On illustre ici une expérience utilisant ce processus de mise au jour de la variation entre 2 textes.

Etape n°1 : alignement des 2 volets d'un même texte

Il s'agit de 2 versions du discours de Ségolène Royal du 11/02/2007 (pour en savoir plus sur ces 2 versions : [billet de J. Véronis sur son blog](#))

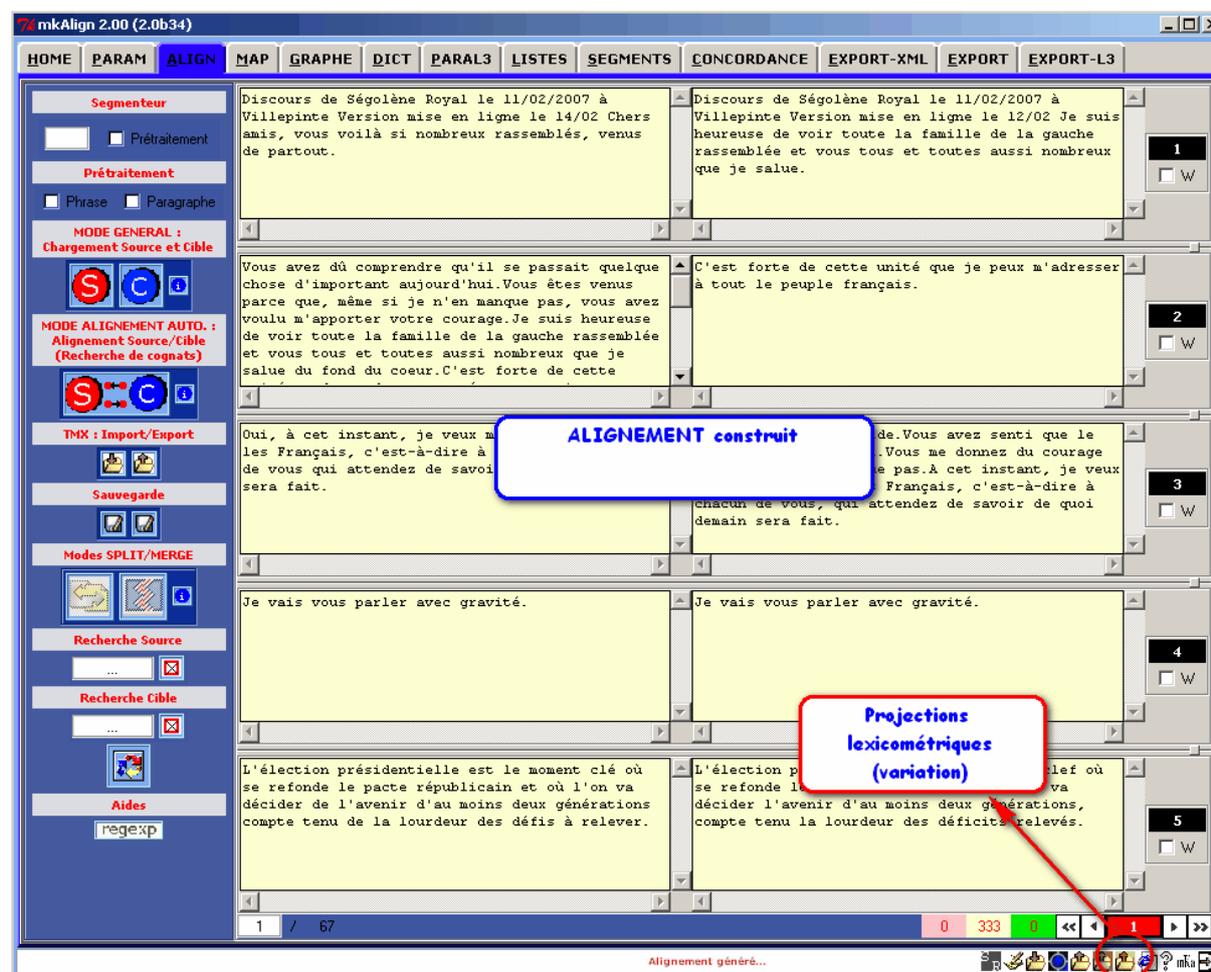


Figure 42 (2) : *mkAlign* Repérage de la variation (alignement initial)

¹⁵ compare deux fichiers et affiche les différences (cf <http://fr.wikipedia.org/wiki/Diff>)

¹⁶ <http://search.cpan.org/~mjarman/Tk-DiffText-0.17/lib/Tk/DiffText.pm>

Etape n°2 : repérage de la variation par coloration des ajouts, suppressions, modifications dans l'onglet VARIATION

Après avoir choisi le grain de la variation (mot, ligne, caractère), on lance la visualisation de la variation en activant le bouton idoine :

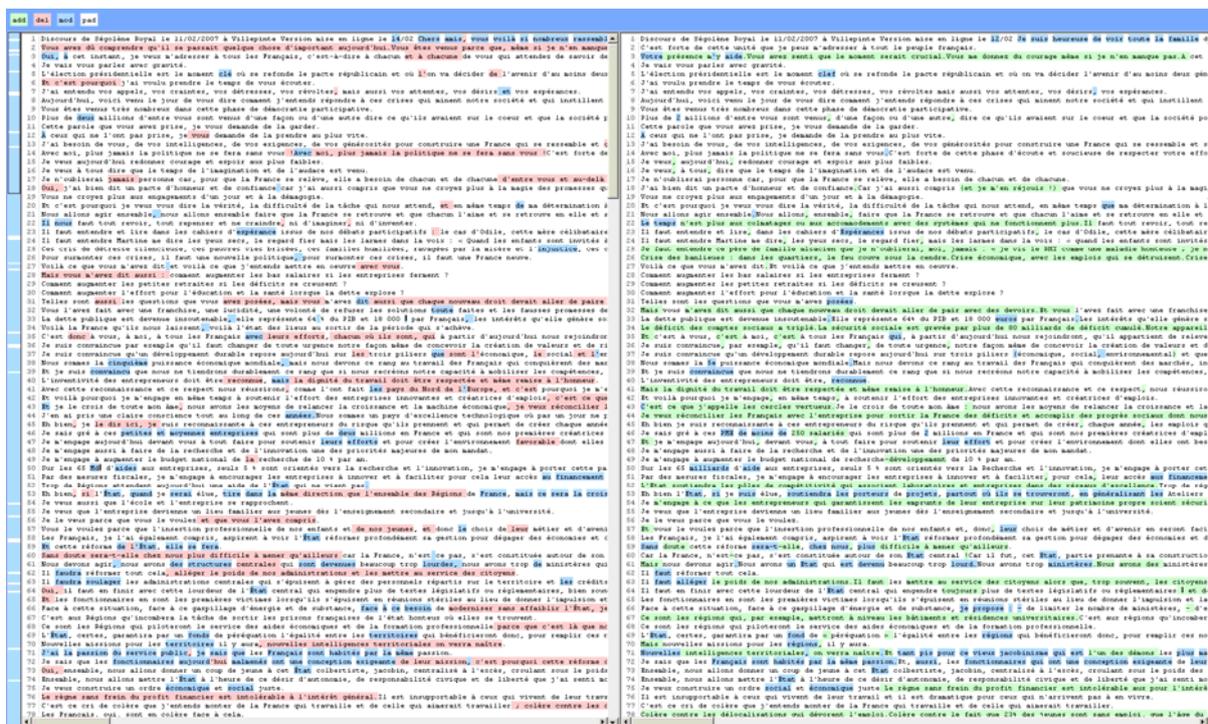


Figure 42 (3) : mkAlign Mise au jour de la variation

Ce résultat est exportable au format HTML ; on trouve en ligne plusieurs illustrations de ces exports :

- Deux traductions du discours d'investiture de B. Obama :
 - [export comparaison](#)
 - [graphique de comptage de la variation](#)
- Deux discours de Ségolène Royal (campagne 2007) :
 - [export comparaison](#) (après alignement automatique)
- Deux discours de Nicolas Sarkozy (conférence de presse 2008) :
 - [export comparaison](#) (après alignement automatique)

On peut aussi calculer des indicateurs de la variation (fond commun, mots ajoutés, supprimés, modifiés...) : illustration en ligne *supra*.

16.2 Projections lexicométriques

(module expérimental)

Objectif : Repérage de la variation dans 2 versions d'un même texte par projection lexicométrique.

On trouvera en ligne des exemples de couples de textes disponibles dans 2 états "proches".

- **Exemple n°1** : 2 versions du texte de la conférence de presse de Nicolas Sarkozy le 8 janvier 2009 (voeux à la presse) ; pour en savoir plus sur ces 2 versions : [billet de Philippe Gambette sur son blog](#).
- **Exemple n°2** : [2 discours de Ségolène Royal](#) au cours de la campagne 2007. On trouvera derrière ce lien une première phase du processus de *projection lexicométrique* réalisé ici. Le processus intégré à mkAlign intègre un mécanisme d'alignement automatique réalisé avant la projection.

Méthode :

1. Alignement automatique par recherche de cognats des 2 volets traités
2. Projection (*via* mkAlign) sur les 2 volets d'un texte (ce lien donne à voir la concaténation des fichiers traités ici) :
 - des segments maximaux déterminés à partir du calcul des segments répétés de l'ensemble des 2 volets
 - des sous-segments maximaux propres à chaque volet,
 - des hapax de l'ensemble des 2 volets,
 - des hapax associés à chacun des 2 volets du texte,
 - des formes propres à chaque volet.

Des exemples de résultats de ce type projection sont disponibles en ligne :

Sur l'exemple n°1 :

<http://www.tal.univ-paris3.fr/mkAlign/mkalign-variation/discours-sarko/projection-1200418948.html>

Sur l'exemple n°2 :

<http://www.tal.univ-paris3.fr/mkAlign/mkalign-variation/discours-sego/projection-1200419509.html>

17 Le rapport : enregistrement des résultats produits

Les résultats qui intéressent l'utilisateur pour une exploitation ultérieure peuvent être rassemblés dans un dossier construit dans le répertoire `rapport`. Ce dossier aisément manipulable à l'aide d'un navigateur web (*Firefox*, *Internet Explorer*, *Safari*, etc.) contient un fichier d'index qui permet la navigation parmi les résultats sélectionnés. Le rapport peut être consulté dès que l'utilisateur l'a enregistré.

17.1 Ajouter un élément au rapport

Pour ajouter un document au rapport, il suffit de se positionner sur un onglet (ou une fenêtre) contenant un résultat produit par le programme puis de cliquer sur l'icône *Ajouter au rapport* présente dans la fenêtre principale ou dans la fenêtre de résultats (c'est le cas pour les spécificités).

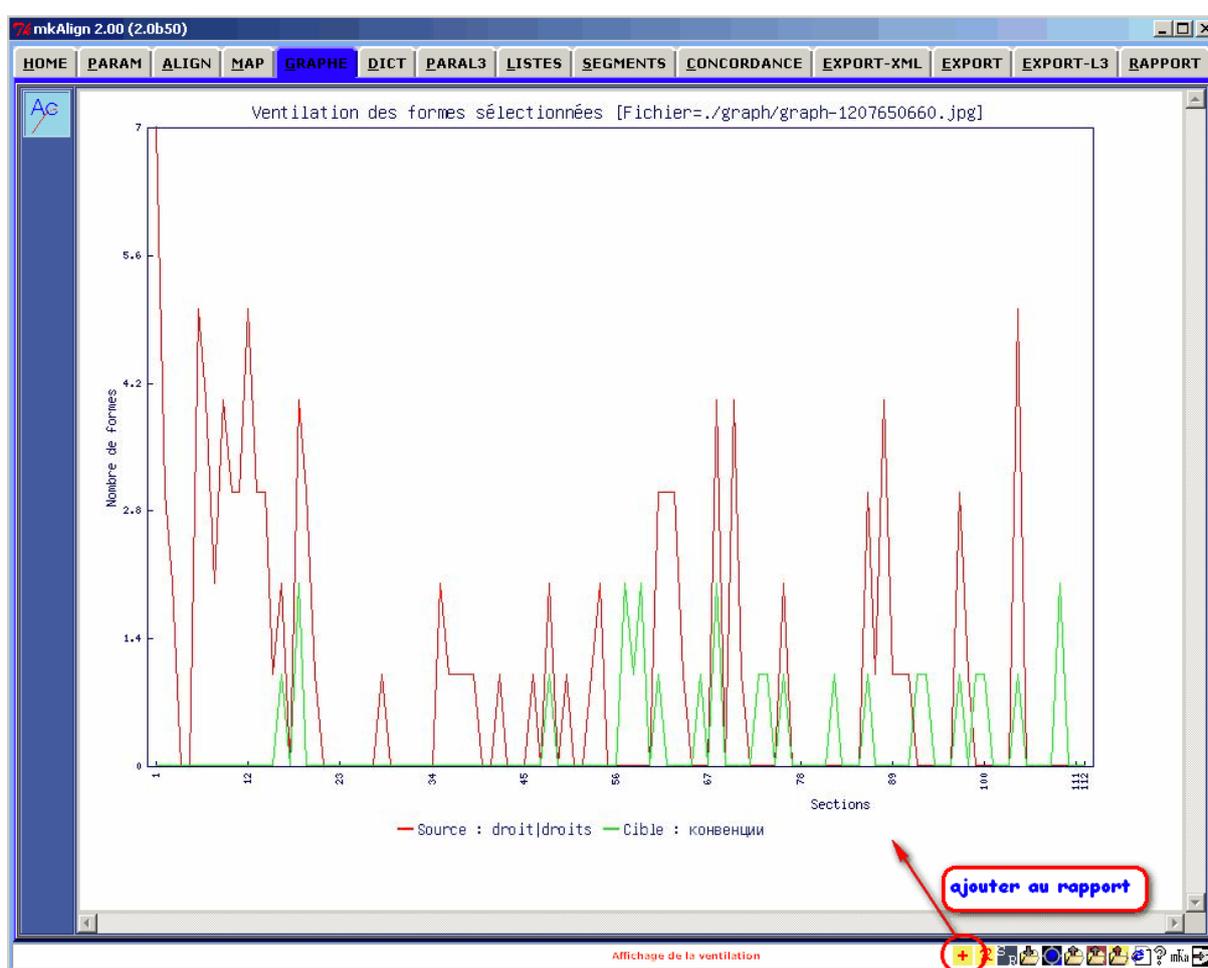


Figure 43 : *mkAlign* Ajouter un élément au rapport

Une trace des éléments ajoutés au rapport est visible de manière synthétique dans l'onglet **RAPPORT**.

Le rapport est constitué par un cartouche et la liste des éléments ajoutés par l'utilisateur. Chacun des éléments du rapport dispose d'une zone de texte « libre » dans laquelle l'utilisateur peut y noter ses observations. Pour éditer ou modifier cette note, il suffit d'activer l'icône de l'éditeur présent près de chaque élément du rapport dans l'onglet **RAPPORT**.

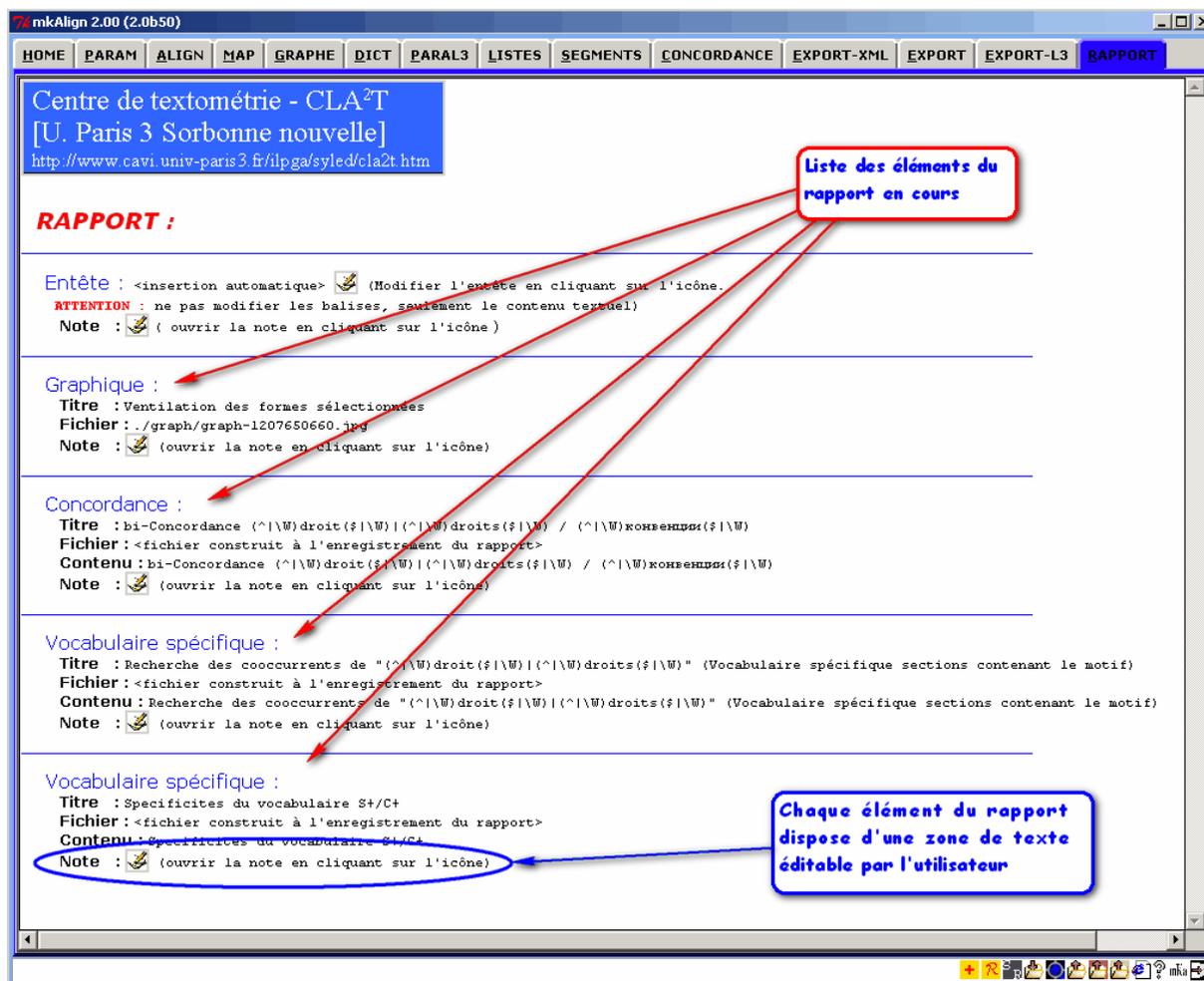


Figure 44 : *mkAlign* Onglet **EXPORT** : Rapport en cours

17.2 Ajouter un élément externe au rapport

Cette fonctionnalité permet à l'utilisateur d'insérer dans le rapport des images externes (non produites directement par **mkAlign**). Par exemple, si l'on souhaite insérer une image donnant à voir la carte des sections, on peut le faire en procédant de la manière suivante :

- créer une copie d'écran de la carte des sections (par exemple avec un outil comme *FastStone Capture* <http://www.faststone.org/>) et sauvegarder la copie d'écran au format jpeg ou gif
- insérer l'image dans le rapport en utilisant le menu idoine dans l'onglet rapport

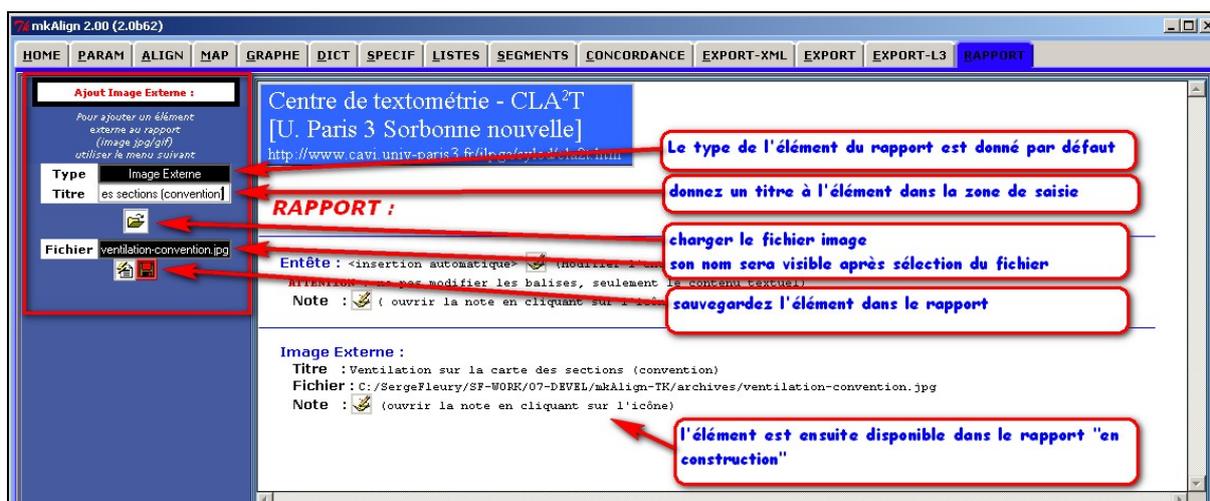


Figure 45 : **mkAlign** Insertion d'image externe dans le rapport

17.3 Enregistrer le rapport

On peut à tout moment visualiser le rapport en cliquant sur le bouton *Enregistrer le rapport* . Une fois activé, ce bouton déclenche la génération d'un « dossier rapport ». Ce dossier contient, dans la distribution initiale de **mkAlign**, des feuilles de styles XSL (pour chaque type de résultat à afficher) et une feuille de styles CSS (pour gérer l'affichage général des toutes les pages du rapport). Les fichiers construits pour un rapport sont au format XML et ils sont rassemblés dans un sous-dossier horodaté du dossier `rapport`. Après enregistrement, le rapport apparaît dans le navigateur web paramétré par défaut sur la machine utilisée (ici *Firefox*) :



Figure 46 : Index du rapport enregistré

Chaque page du rapport contient la description de l'élément visé. Cette description d'un élément du rapport (sur la droite) est constituée :

- d'un *corps* (le contenu de l'*élément du rapport* pour les fichiers résultats ou une liste des *éléments du rapport* pour le fichier d'index)
- d'une *note* : zone de texte libre renseignée par l'utilisateur

On trouvera sur le site de **mkAlign** des exemples complets de rapports produits sur des corpus bilingues.

18 Références

- [Fourmond, 2005] Fourmond, V. Les expressions régulières par l'exemple. H&K, Technique & Pratique, Paris.
- [Kraif, 1999] Kraif, Olivier. Identification des cognats et alignement bi-textuel : une étude empirique, in *Actes TALN'99*, Cargèse, 12-17 juillet 1999, pp.205-214
- [Lafon, 1984] Lafon Pierre. *Dépouillements et statistiques en lexicométrie*. Genève-Paris, Slatkine-Champion.
- [Lamalle, 2001] Lamalle C., Martinez W, Fleury S., Salem A., Kuncova A., Maisondieu A., "Dix premiers pas avec Lexico3", Manuel d'utilisation abrégé ([PDF](#)¹⁷), ([HTML](#)¹⁸) (sur le [site de Lexico](#)¹⁹).
- [Lebart, 1994] Lebart L. et Salem A. *Statistique textuelle*. Dunod, Paris ([livre disponible en ligne](#)²⁰).
- [Martinez, 2002] William Martinez, Zimina Maria. "Utilisation de la méthode des cooccurrences pour l'alignement des mots de textes bilingues" ([PDF](#)), in Actes [JADT'2002](#)²¹, Journées Internationales d'Analyse Statistiques des Données Textuelles, St Malo.
- [Martinez, 2003] William Martinez. *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*. Thèse de Doctorat en Sciences du Langage, Université de la Sorbonne nouvelle - Paris 3, sous la direction d'André Salem, Paris.
- [Martinez, 2006] William Martinez, Jean-Marc Leblanc. "L'analyse contrastive des réseaux de cooccurrence Le monde dans les discours des présidents de la Cinquième République" ([PDF](#)), in Actes [JADT'2006](#), Journées Internationales d'Analyse Statistiques des Données Textuelles, Besançon.
- [Véronis, 2000] Véronis, Jean. *Alignement de corpus multilingues* ([PDF](#)²²), in Pierrel, J.-M., éditeur, *Ingénierie des langues*, Informatique et systèmes d'information, chapitre 6, pages 151–172. Hermès Science, Paris.
- [Zimina, 2004a] Zimina Maria. *Approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles* ([slides](#)²³). Présentation à la soutenance de thèse, Université de la Sorbonne nouvelle - Paris 3.
- [Zimina, 2004b] Zimina Maria. *L'alignement textométrique des unités lexicales à correspondances multiples dans les corpus parallèles*. ([PDF](#)²⁴), in Actes [JADT'2004](#), Louvain-la-Neuve (Belgique).
- [Zimina, 2005] Zimina Maria, *Topographie bi-textuelle et approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles* ([PDF](#)), in Actes des 7es Journées scientifiques du Réseau de chercheurs²⁵ "Lexicologie, Terminologie, Traduction", Institut supérieur de traducteurs et interprètes (ISTI), Bruxelles, 8-10 septembre 2005.
- [Zimina, 2007a] Fleury Serge, Maria Zimina, "Exploring Translation Corpora with MkAlign²⁶", in *Translation Journal*, Volume 11, n°1.
- [Zimina, 2007b], Zimina Maria, *Corpus multilingues : exploration textométrique dans l'espace intertextuel*, in Ballard M., Pineira-Tresmontant C. (éd) *Les corpus en linguistique et en traductologie*" (p. 107-121), Artois Presses Université.

¹⁷ <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/lex3-10pas/Lexico3-10premierspas.pdf>

¹⁸ <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/lex3-10pas/index.htm>

¹⁹ <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/index.htm>

²⁰ <http://www.cavi.univ-paris3.fr/lexicométrica/livre/st94/st94-tdm.html> ou <http://egsh.enst.fr/lebart/ST.html>

²¹ <http://www.jadt.org/>

²² <http://www.up.univ-mrs.fr/~veronis/pdf/2000hermes6.pdf>

²³ http://www.cavi.univ-paris3.fr/ilpga/ED/student/stmz/ED268-PagePersoMZ_fichiers/stmz/page6_fichiers/26novembre_MZ.zip

²⁴ http://www.cavi.univ-paris3.fr/lexicométrica/jadt/jadt2004/pdf/JADT_118.pdf

²⁵ <http://www.ltt.auf.org/Journees-scientifiques-LTT-2005>

²⁶ <http://accurapid.com/journal/39mk.htm>

19 Annexes

19.1 Les expressions régulières

Une expression régulière peut comporter les éléments suivants :

- **^** désigne le début de champ²⁷.
 - Exemple : **^part** sélectionne les champs commençant par **part**
- **\$** désigne la fin de champ.
 - Exemple : **isme\$** sélectionne les champs terminés par **isme**
- **.** (le caractère "point") désigne un caractère quelconque.
 - Exemple : **o.me** sélectionne les champs contenant **omme**, **oume**, **orme**, **osme** etc.
- **[...]** désigne un caractère quelconque qui est un des caractères entre les crochets.
 - Exemple : **o[mu]me** sélectionne les champs contenant **omme**, **oume**, mais pas **orme**.
- **[^...]** désigne un caractère quelconque qui n'est pas un des caractères entre les crochets.
 - Exemple : **o[^mu]me** sélectionne les champs contenant **orme**, **osme**, mais pas **omme**, **oume**.
- **(choix₁|choix₂|...|choix_n)** désigne une des possibilités **choix₁** ou **choix₂** ou ... **choix_n**.
 - Exemple : **(voul|pers)** sélectionne les champs contenant **voul** ou **pers**.
- **Caractères de modification**
 - Placés derrière un caractère quelconque, ou devant un caractère "point" ou devant une expression entre crochets ou devant une expression parenthésée, les caractères "?", "+", "*" ont le rôle suivant :
 - "?" : rend optionnel le caractère ou l'expression derrière lequel/laquelle il est placé.
 - Exemples :
 - **ismes?\$** sélectionne les champs se terminant par **isme** ou **ismes**.
 - **c.?oupe** sélectionne les champs contenant **coupe**, mais aussi par exemple **croupe**.
 - "*" : signifie que le caractère ou l'expression derrière lequel/laquelle il est placé peuvent se répéter un nombre quelconque de fois, ou être absent.
 - Exemples :
 - **vr[aeiouypqrs]*** sélectionne les champs contenant **vr** suivi d'un nombre quelconque de caractères choisis dans l'ensemble [aeiouypqrs].
 - **^c.*oupe\$** sélectionne les champs commençant par **c** et se terminant par **oupe** (un champ contenant **coupe** est sélectionné).
 - "+" : a presque la même signification que le caractère *, à ceci près que le caractère ou l'expression derrière lequel/laquelle il est placé doit être présent et peut se répéter un nombre quelconque de fois.
 - Exemple : **^c.+oupe\$** sélectionne les champs commençant par **c** et se terminant par **oupe** (un champ contenant **coupe** n'est pas sélectionné).

²⁷ Le terme champ désigne soit la forme graphique d'un mot soit un segment (plusieurs mots consécutifs)

19.2 Fonctionnalités de l'éditeur d'alignement

- [1] Clicking mouse button 1 positions the insertion cursor just before the character underneath the mouse cursor, sets the input focus to this widget, and clears any selection in the widget. Dragging with mouse button 1 strokes out a selection between the insertion cursor and the character under the mouse.
- [2] Double-clicking with mouse button 1 selects the word under the mouse and positions the insertion cursor at the beginning of the word. Dragging after a double click will stroke out a selection consisting of whole words.
- [3] Triple-clicking with mouse button 1 selects the line under the mouse and positions the insertion cursor at the beginning of the line. Dragging after a triple click will stroke out a selection consisting of whole lines.
- [4] The ends of the selection can be adjusted by dragging with mouse button 1 while the Shift key is down; this will adjust the end of the selection that was nearest to the mouse cursor when button 1 was pressed. If the button is double-clicked before dragging then the selection will be adjusted in units of whole words; if it is triple-clicked then the selection will be adjusted in units of whole lines.
- [5] Clicking mouse button 1 with the Control key down will reposition the insertion cursor without affecting the selection.
- [6] If any normal printing characters are typed, they are inserted at the point of the insertion cursor.
- [7] The view in the widget can be adjusted by dragging with mouse button 2. If mouse button 2 is clicked without moving the mouse, the selection is copied into the text at the position of the mouse cursor. The Insert key also inserts the selection, but at the position of the insertion cursor.
- [8] If the mouse is dragged out of the widget while button 1 is pressed, the entry will automatically scroll to make more text visible (if there is more text off-screen on the side where the mouse left the window).
- [9] The Left and Right keys move the insertion cursor one character to the left or right; they also clear any selection in the text. If Left or Right is typed with the Shift key down, then the insertion cursor moves and the selection is extended to include the new character. Control-Left and Control-Right move the insertion cursor by words, and Control-Shift-Left and Control-Shift-Right move the insertion cursor by words and also extend the selection. Control-b and Control-f behave the same as Left and Right, respectively. Meta-b and Meta-f behave the same as Control-Left and Control-Right, respectively.
- [10] The Up and Down keys move the insertion cursor one line up or down and clear any selection in the text. If Up or Right is typed with the Shift key down, then the insertion cursor moves and the selection is extended to include the new character. Control-Up and Control-Down move the insertion cursor by paragraphs (groups of lines separated by blank lines), and Control-Shift-Up and Control-Shift-Down move the insertion cursor by paragraphs and also extend the selection. Control-p and Control-n behave the same as Up and Down, respectively.
- [11] The Next and Prior keys move the insertion cursor forward or backwards by one screenful and clear any selection in the text. If the Shift key is held down while Next or Prior is typed, then the selection is extended to include the new character. Control-v moves the view down one screenful without moving the insertion cursor or adjusting the selection.
- [12] Control-Next and Control-Prior scroll the view right or left by one page without moving the insertion cursor or affecting the selection.
- [13] Home and Control-a move the insertion cursor to the beginning of its line and clear any selection in the widget. Shift-Home moves the insertion cursor to the beginning of the line and also extends the selection to that point.
- [14]

- End and Control-e move the insertion cursor to the end of the line and clear any selection in the widget. Shift-End moves the cursor to the end of the line and extends the selection to that point.
- [15] Control-Home and Meta-< move the insertion cursor to the beginning of the text and clear any selection in the widget. Control-Shift-Home moves the insertion cursor to the beginning of the text and also extends the selection to that point.
- [16] Control-End and Meta-> move the insertion cursor to the end of the text and clear any selection in the widget. Control-Shift-End moves the cursor to the end of the text and extends the selection to that point.
- [17] The Select key and Control-Space set the selection anchor to the position of the insertion cursor. They don't affect the current selection. Shift-Select and Control-Shift-Space adjust the selection to the current position of the insertion cursor, selecting from the anchor to the insertion cursor if there was not any selection previously.
- [18] Control-/ selects the entire contents of the widget.
- [19] Control-\ clears any selection in the widget.
- [20] The F16 key (labelled Copy on many Sun workstations) or Meta-w copies the selection in the widget to the clipboard, if there is a selection.
- [21] The F20 key (labelled Cut on many Sun workstations) or Control-w copies the selection in the widget to the clipboard and deletes the selection. If there is no selection in the widget then these keys have no effect.
- [22] The F18 key (labelled Paste on many Sun workstations) or Control-y inserts the contents of the clipboard at the position of the insertion cursor.
- [23] The Delete key deletes the selection, if there is one in the widget. If there is no selection, it deletes the character to the right of the insertion cursor.
- [24] Backspace and Control-h delete the selection, if there is one in the widget. If there is no selection, they delete the character to the left of the insertion cursor.
- [25] Control-d deletes the character to the right of the insertion cursor.
- [26] Meta-d deletes the word to the right of the insertion cursor.
- [27] Control-k deletes from the insertion cursor to the end of its line; if the insertion cursor is already at the end of a line, then Control-k deletes the newline character.
- [28] Control-o opens a new line by inserting a newline character in front of the insertion cursor without moving the insertion cursor.
- [29] Meta-backspace and Meta-Delete delete the word to the left of the insertion cursor.
- [30] Control-x deletes whatever is selected in the text widget.
- [31] Control-t reverses the order of the two characters to the right of the insertion cursor.