

Centre de textométrie - CLA<sup>2</sup>T  
[U. Paris 3 Sorbonne nouvelle]  
<http://www.cavi.univ-paris3.fr/ilpga/syled/cla2t.htm>



*(mise à jour : septembre 2007, 1.64)*

# Manuel d'utilisation

Serge Fleury

Maria Zimina

URL : <http://tal.univ-paris3.fr/mkAlign/>

Téléchargement : <http://tal.univ-paris3.fr/mkAlign/>

Contact : [serge.fleury@univ-paris3.fr](mailto:serge.fleury@univ-paris3.fr)

Version exécutable sous Windows disponible en ligne

Version script Perl/Tk disponible sur demande

(documentation en ligne sur la page du projet, ce document mis à jour)

## Sommaire

1	Figures .....	3
2	Préambule.....	4
3	Présentation générale.....	5
2	Interface de mkAlign.....	6
4	Descriptif des icônes .....	8
5	Mode d'emploi .....	9
5.1	Mode général.....	9
5.1.1	Chargement des textes à aligner .....	9
5.1.2	Choix du segmenteur.....	12
5.1.3	Prétraitement des textes à aligner .....	12
5.1.4	Découpage des textes en parties.....	13
5.1.5	Paramètres lexicométriques des textes chargés.....	13
5.2	Mode « alignement par recherche de cognats » .....	15
6	Principales fonctionnalités de mkAlign .....	17
6.1	Recherche de chaînes .....	17
6.2	Exporter sous-corpus contenant un motif.....	18
6.3	Figement de cellule (lecture/écriture) .....	18
6.4	Fractionnement de cellule .....	18
6.5	Fusion de cellule.....	18
6.6	Etat de l'alignement .....	18
6.7	Remarques sur le segmenteur.....	19
6.8	Les modes.....	19
7	Représentation cartographique de l'alignement.....	21
7.1	Construction de la carte de l'alignement .....	21
7.2	Recherche de motifs dans la carte de l'alignement .....	21
7.3	Affichage d'une partition dans la carte de l'alignement .....	23
7.4	Navigation vers l'alignement .....	24
7.5	Le vocabulaire spécifique d'une section de la carte.....	24
7.6	Le vocabulaire spécifique d'une sélection de sections de la carte .....	25
7.7	Le vocabulaire spécifique des sections de la carte contenant une forme ou un motif : recherche de cooccurents .....	26
8	Importation de listes .....	28
9	Segments répétés .....	30
10	Export des traitements .....	32
	Export vers Lexico3 .....	32
	Export au format XML.....	33
	Export au format XML pour sauvegarder une session de travail.....	34
	Export complet au format HTML .....	35
	Export partiel au format HTML .....	36
	Export de bi-textes au format TXT .....	38
11	Import d'alignement.....	39
12	Références .....	41
13	Annexes.....	43
13.1	Les expressions régulières.....	43
13.2	Fonctionnalités de l'éditeur d'alignement.....	44

# 1 Figures

Figure 1 : <i>mkAlign</i> Fenêtre principale .....	6
Figure 2 : <i>mkAlign</i> Onglet PARAM .....	7
Figure 3 : <i>mkAlign</i> Onglet ALIGN .....	9
Figure 4 : <i>mkAlign</i> Onglet ALIGN, fichiers chargés .....	10
Figure 5 : <i>mkAlign</i> Onglet PARAM, paramétrage de l'encodage .....	11
Figure 6 : <i>mkAlign</i> Onglet ALIGN, fichiers chargés après sélection d'encodage .....	12
Figure 7 : <i>mkAlign</i> choix du <i>segmenteur</i> onglet ALIGN ou onglet PARAM .....	12
Figure 8 : <i>mkAlign</i> pré-formatage de la SOURCE et de la CIBLE .....	13
Figure 9 : <i>mkAlign</i> Onglet DICT .....	14
Figure 10 : <i>mkAlign</i> Onglet PARAL3 .....	14
Figure 11 (1) : <i>mkAlign</i> Alignement par recherche de cognats .....	15
Figure 11 (2) : <i>mkAlign</i> Alignement par recherche de cognats .....	16
Figure 12 : <i>mkAlign</i> Onglet ALIGN, recherche de chaînes .....	17
Figure 13 : <i>mkAlign</i> Alignement en cours .....	19
Figure 14 : <i>mkAlign</i> Onglet MAP .....	21
Figure 15 : <i>mkAlign</i> Recherche et Navigation dans la carte des sections (1) .....	22
Figure 16 : <i>mkAlign</i> Recherche et Navigation dans la carte des sections (2) .....	23
Figure 17 : <i>mkAlign</i> Carte / Sections / Parties .....	23
Figure 18 : <i>mkAlign</i> Carte / Sections / Spécificités .....	24
Figure 19 : <i>mkAlign</i> Carte / Sections / Sélection .....	25
Figure 20 (1) : <i>mkAlign</i> Carte / Sections / Spécificité sections .....	25
Figure 20 (2) : <i>mkAlign</i> Carte / Sections / Spécificité sections .....	26
Figure 21 (1) : <i>mkAlign</i> Carte / Sections / Spécificités / Cooccurents .....	27
Figure 21 (2) : <i>mkAlign</i> Carte / Sections / Spécificités / Cooccurents .....	27
Figure 22 : <i>mkAlign</i> Import de listes .....	28
Figure 23 : <i>mkAlign</i> Onglet LISTES .....	28
Figure 24 (1) : <i>mkAlign</i> Onglet SEGMENTS .....	30
Figure 24 (2) : <i>mkAlign</i> Onglet PARAM, paramétrage du calcul des segments répétés .....	30
Figure 24 (3) : <i>mkAlign</i> Onglet SEGMENTS – sélection de segments .....	31
Figure 24 (4) : <i>mkAlign</i> Onglet SEGMENTS – sélection de segments .....	31
Figure 25 : <i>mkAlign</i> Onglet EXPORT-L3 .....	32
Figure 26 : <i>mkAlign</i> Onglet EXPORT-XML .....	33
Figure 27 : <i>mkAlign</i> Schéma Export .....	33
Figure 28 : <i>mkAlign</i> Export HTML .....	35
Figure 29 : <i>mkAlign</i> Export sélectif (sélection d'une forme) .....	36
Figure 30 : <i>mkAlign</i> Résultat d'export sélectif - Fichier complet : example-export.html .....	37
Figure 31 : <i>mkAlign</i> , Export de bi-textes .....	38
Figure 32 : <i>mkAlign</i> Onglet EXPORT .....	39
Figure 33 : <i>mkAlign</i> Schéma Import .....	40

## 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*<sup>1</sup>, 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.

---

<sup>1</sup> 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 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 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 traductionnel<sup>2</sup>.

---

<sup>2</sup> 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].

## 2 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.

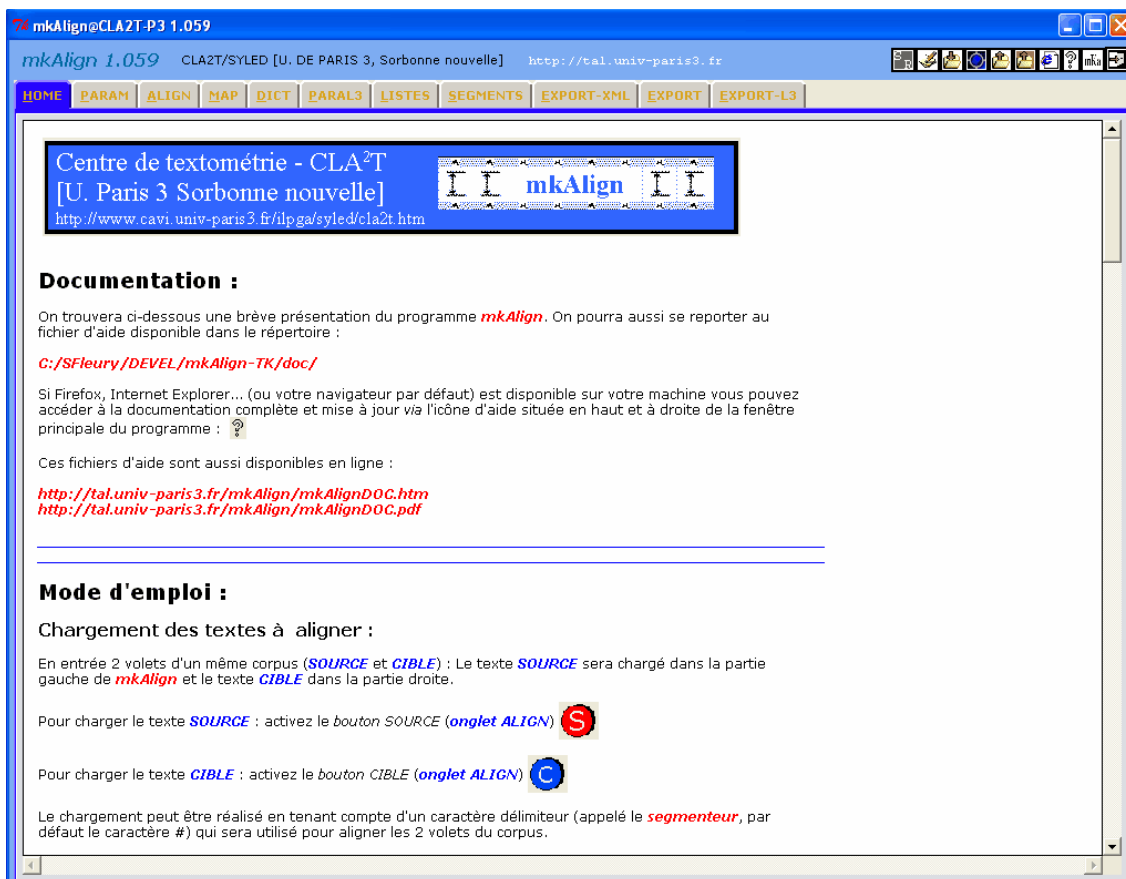


Figure 1 : **mkAlign** Fenêtre principale

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.).

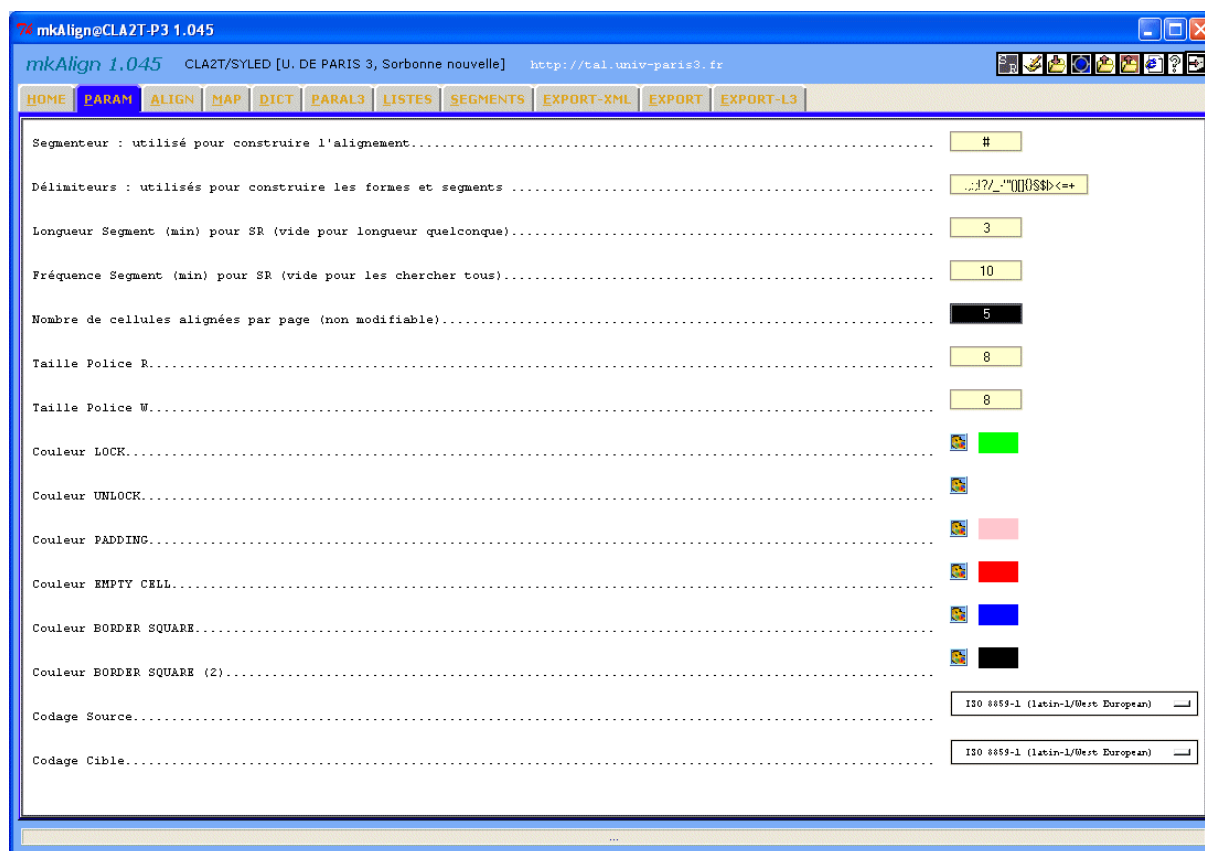



































Figure 2 : **mkAlign** Onglet **PARAM**

Les autres onglets sont décrits *infra*.

## 4 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	Onglet EXPORT
	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
	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 cooccurents d'une forme dans la source	Onglet MAP
	Recherche des cooccurents 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

## 5 Mode d'emploi

Deux 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 cognats »**. Alignement automatique du fichier **SOURCE** et du fichier **CIBLE** basé sur la recherche de cognats.

### 5.1 Mode général

#### 5.1.1 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.

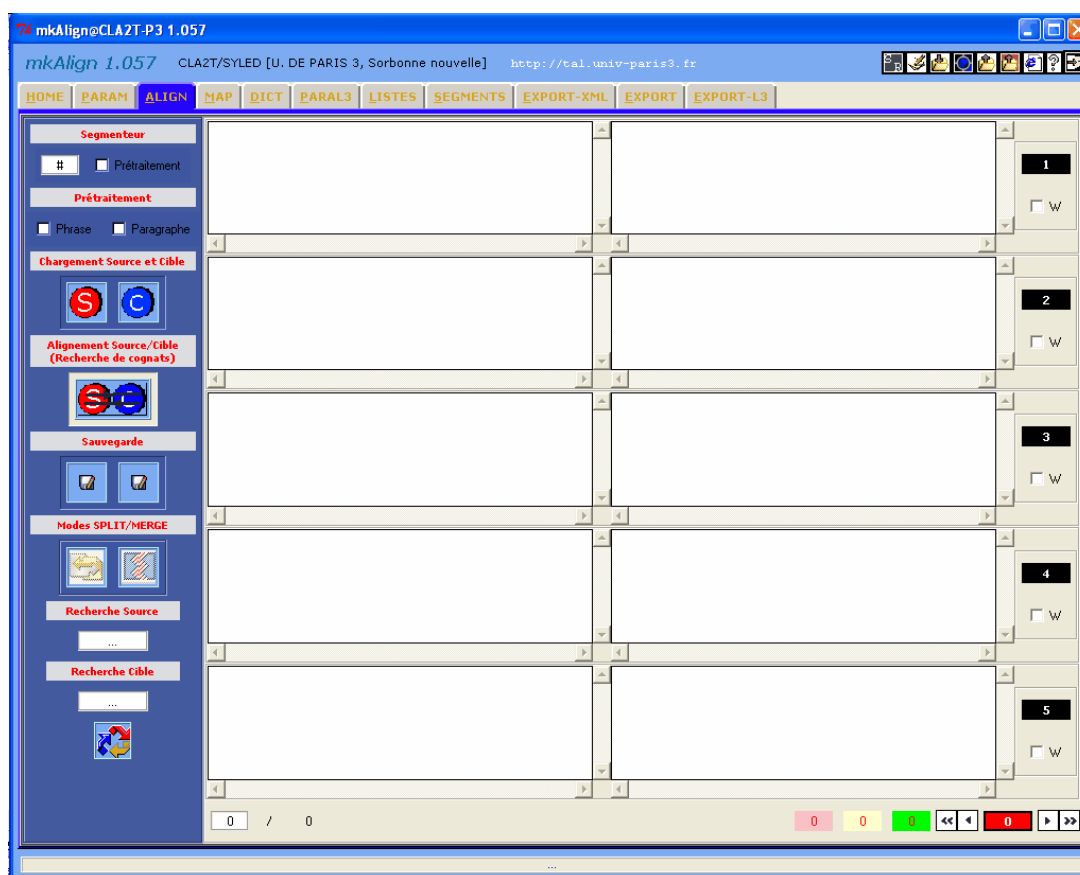




Figure 3 : **mkAlign** Onglet **ALIGN**

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 source  (onglet **ALIGN**)
- Pour charger le texte *cible* : activez le bouton cible  (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.

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

**Fichier Source :**

```

SSSSSSSSSSSSSSSSSS #
SSSSSSSSSSSSSSSSSS #
etc.

```

**Fichier Cible :**

```

CCCCCCCCCCCCCCCCC #
CCCCCCCCCCCCCCCCC #
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*, sur la partie droite.

Les 2 volets du corpus alignés 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 :

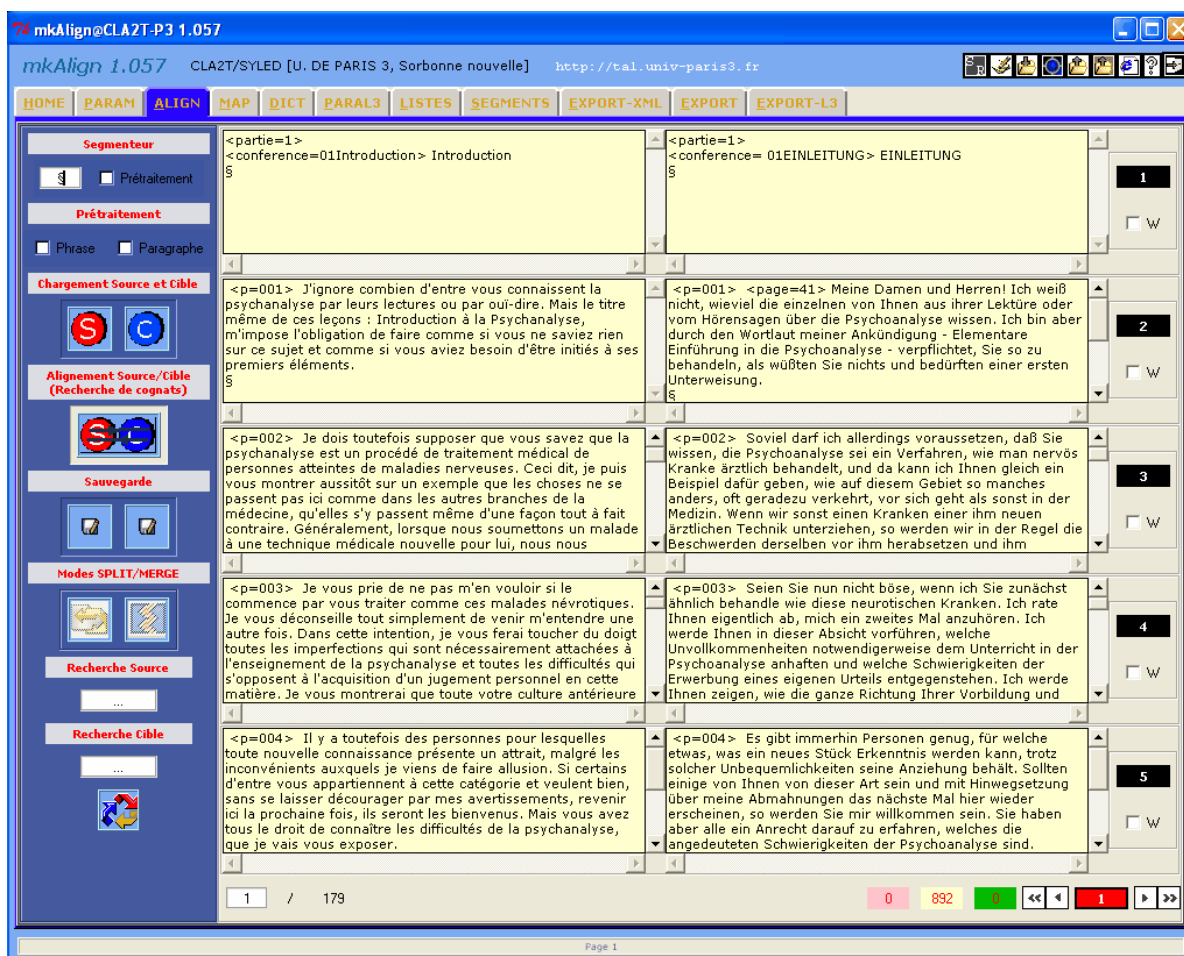


Figure 4 : **mkAlign** Onglet **ALIGN**, fichiers chargés

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

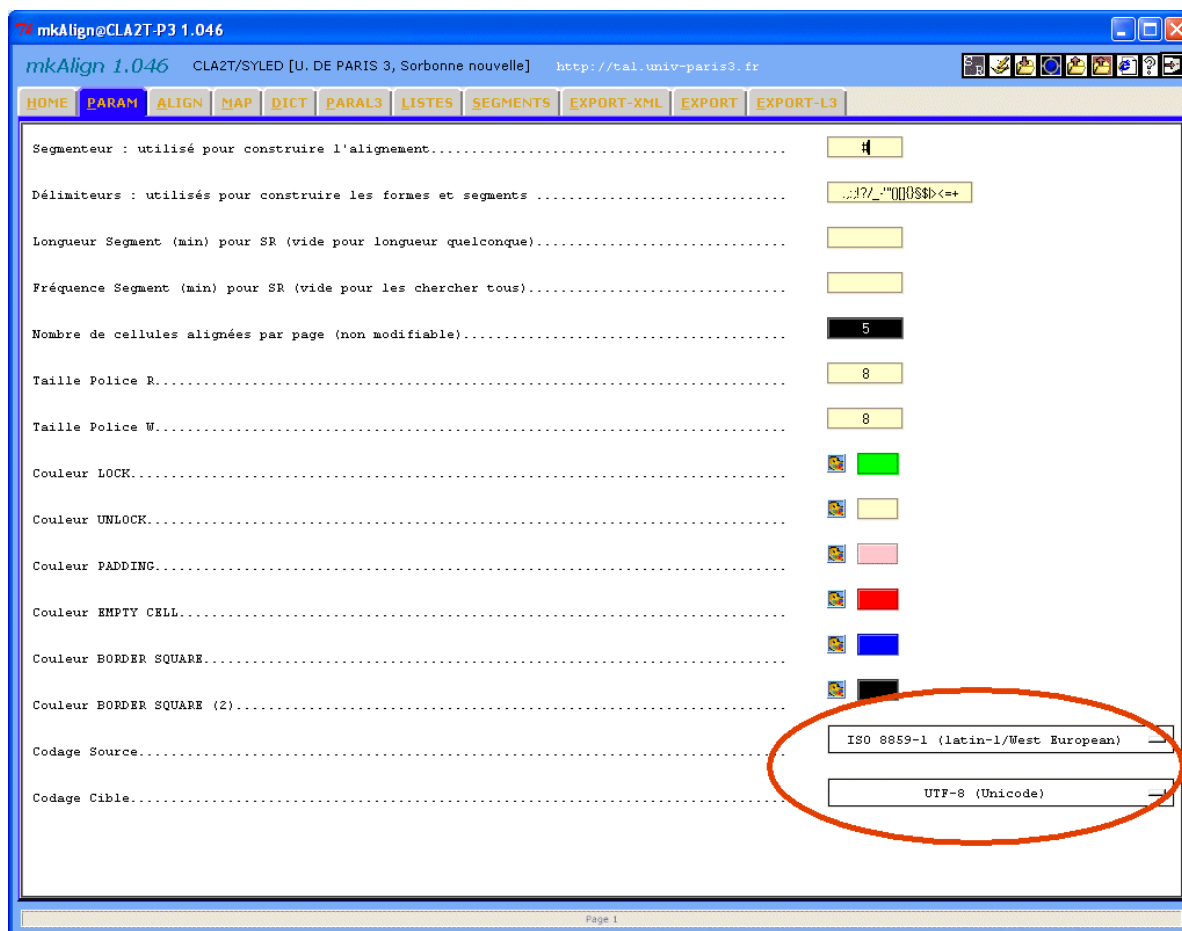


Figure 5 : **mkAlign** Onglet **PARAM**, paramétrage de l'encodage

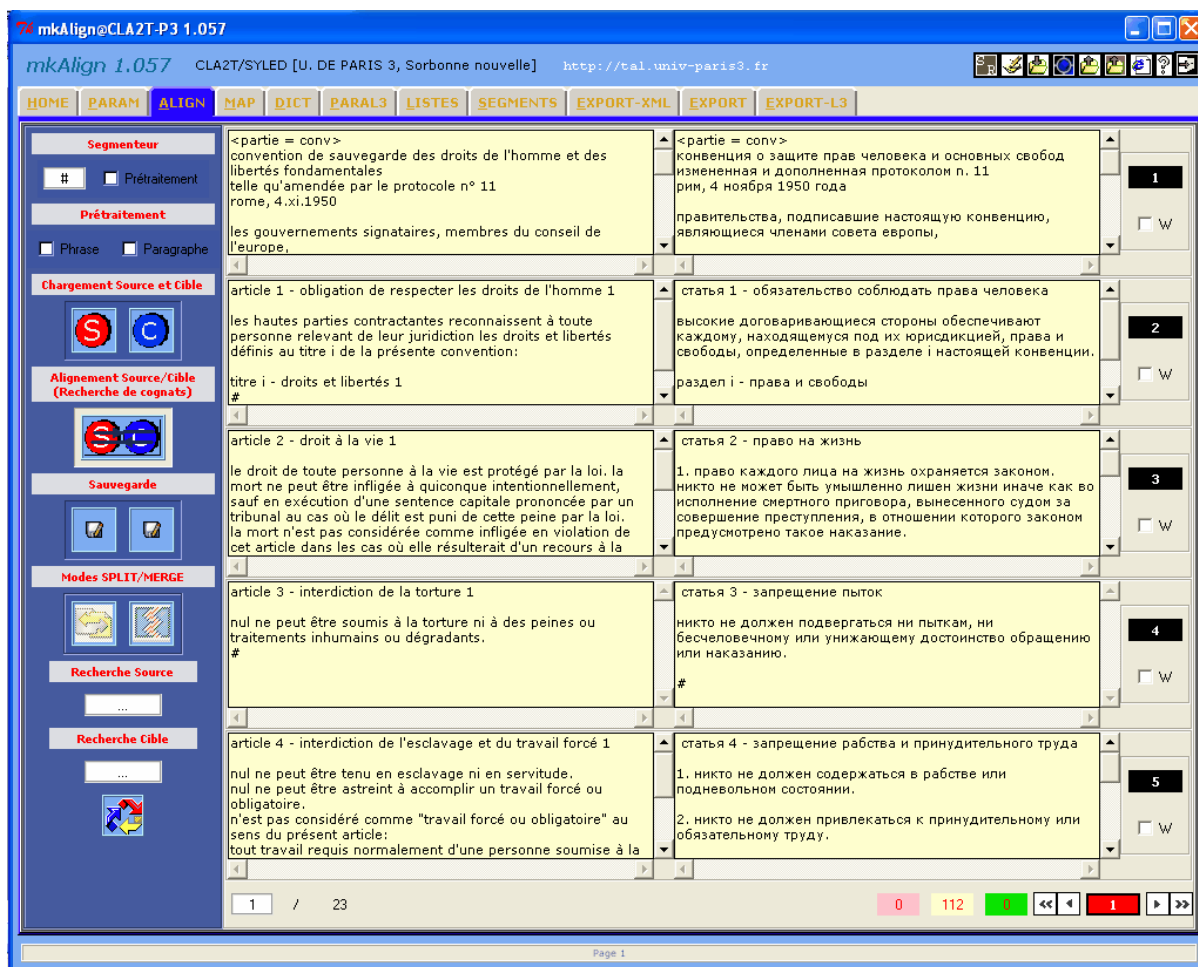
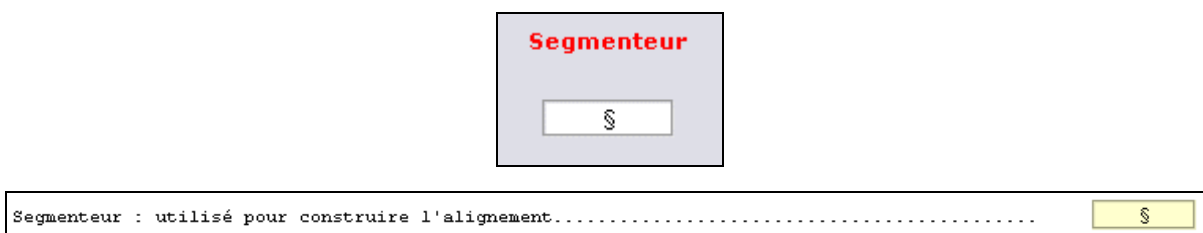


Figure 6 : mkAlign Onglet ALIGN, fichiers chargés après sélection d'encodage

### 5.1.2 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**.

### 5.1.3 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 modifications sur les 2 volets à charger.

Il est aussi possible de prétraiter les 2 textes à charger et d'attribuer 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

Une phrase est définie grossièrement comment une chaîne de caractères se terminant par les caractères : 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 choisi, 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.

#### 5.1.4 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.
```

#### 5.1.5 Paramètres lexicométriques des textes chargés

Le chargement des fichiers *source* et *cible* déclenche automatiquement un module de segmentation (issu du programme **Lexico3** [Lamalle & al., 2001]) 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 et segments .....	.,!/?_'"008\$%><=+
---------------------------------------------------------------------	--------------------

Ces délimiteurs sont utilisés aussi par le programme calculant les segments répétés présenté *infra*.

Le résultat de cette segmentation est visible dans l'onglet **DICT** et dans l'onglet **PARAL3** :

- le dictionnaire sera visible dans le premier

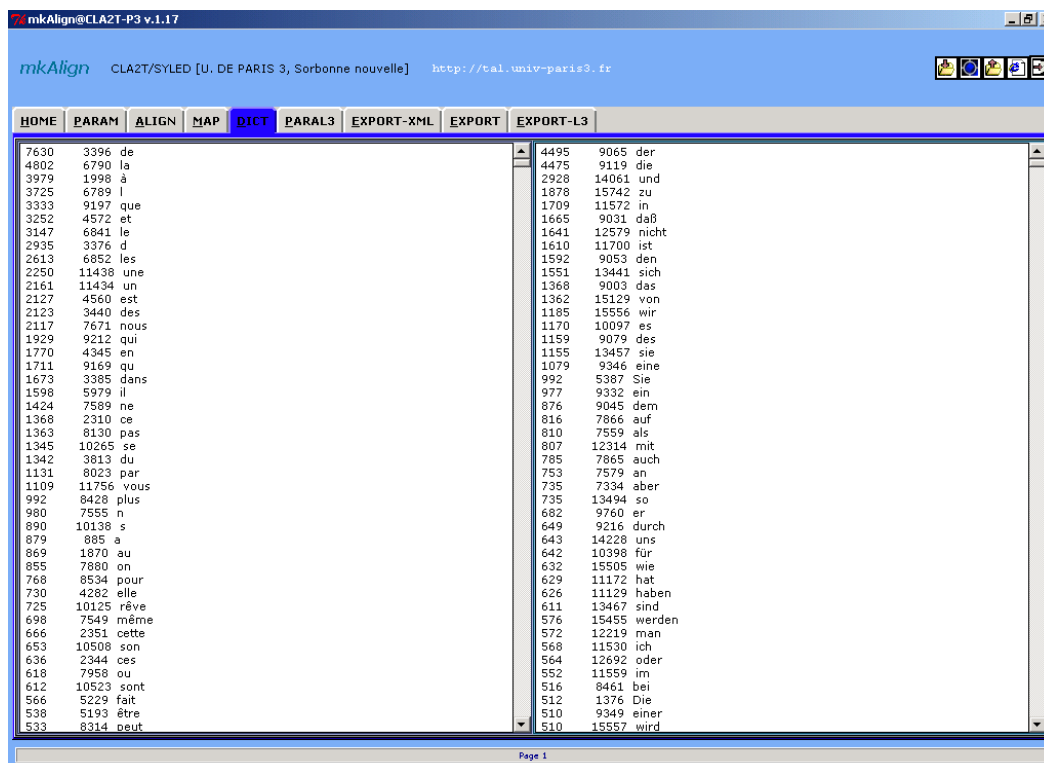


Figure 9 : mkAlign Onglet DICT

- et les paramètres de la segmentation dans le second

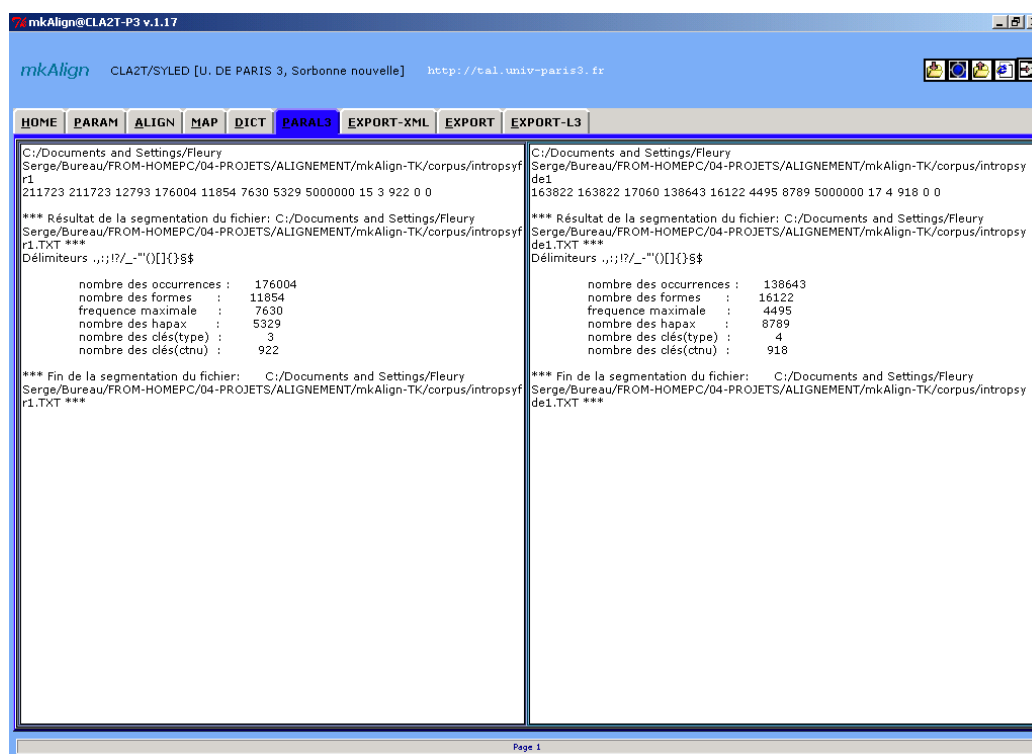


Figure 10 : mkAlign Onglet PARAL3

Un clic droit sur un des mots du dictionnaire présent dans l'onglet **DICT** déclenchera la recherche de ce mot et l'affichage du résultat de cette recherche dans la représentation cartographique de l'alignement présentée *infra* (cf onglet **MAP**).

## 5.2 Mode « alignement par recherche de cognats »

Ce mode de chargement s'inspire de travaux réalisés dans le domaine de l'alignement multilingue utilisant l'extraction de points d'ancrage pour aligner des textes. Cette méthode permet, sur des langues apparentées, de construire un alignement en recherchant tout d'abord des équivalents traductionnels sous la forme de mots apparentés (ou cognats), les points d'ancrage obtenus dessinent des zones textuelles préalignées éventuellement à affiner. On regardera en particulier [Kraif, 1999].

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

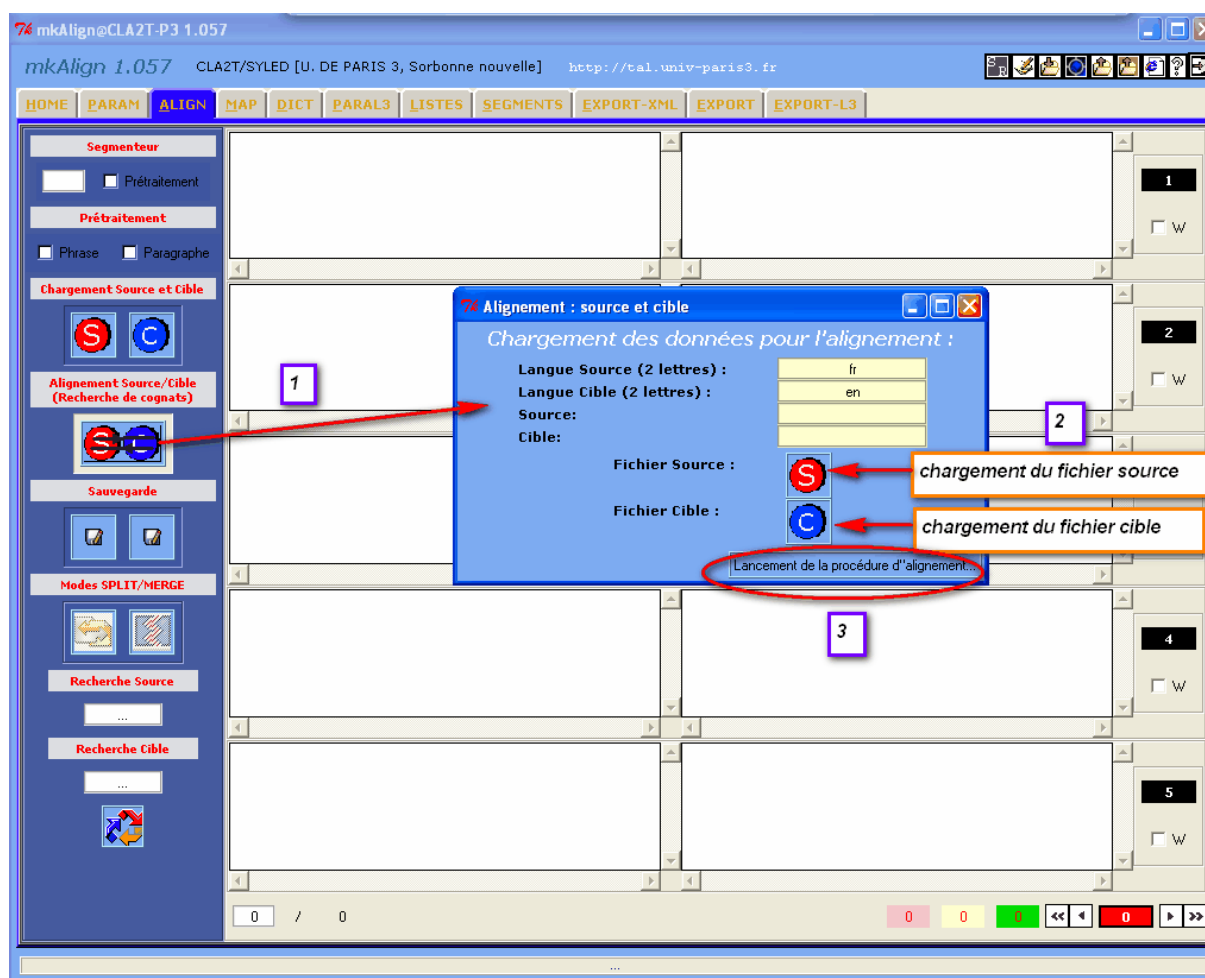


Figure 11 (1) : **mkAlign** Alignement par recherche de cognats

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

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

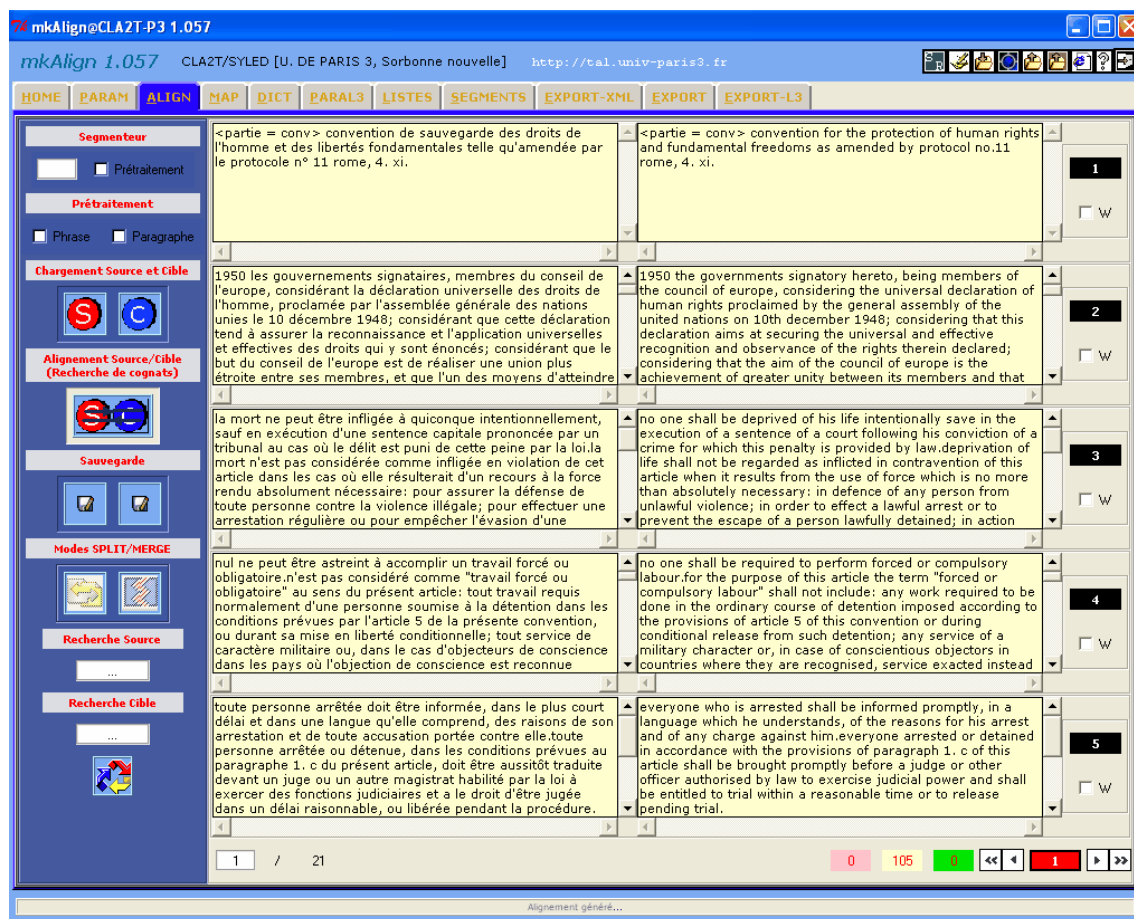


Figure 11 (2) : mkAlign Aligement par recherche de cognats

## 6 Principales fonctionnalités de mkAlign

### 6.1 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.

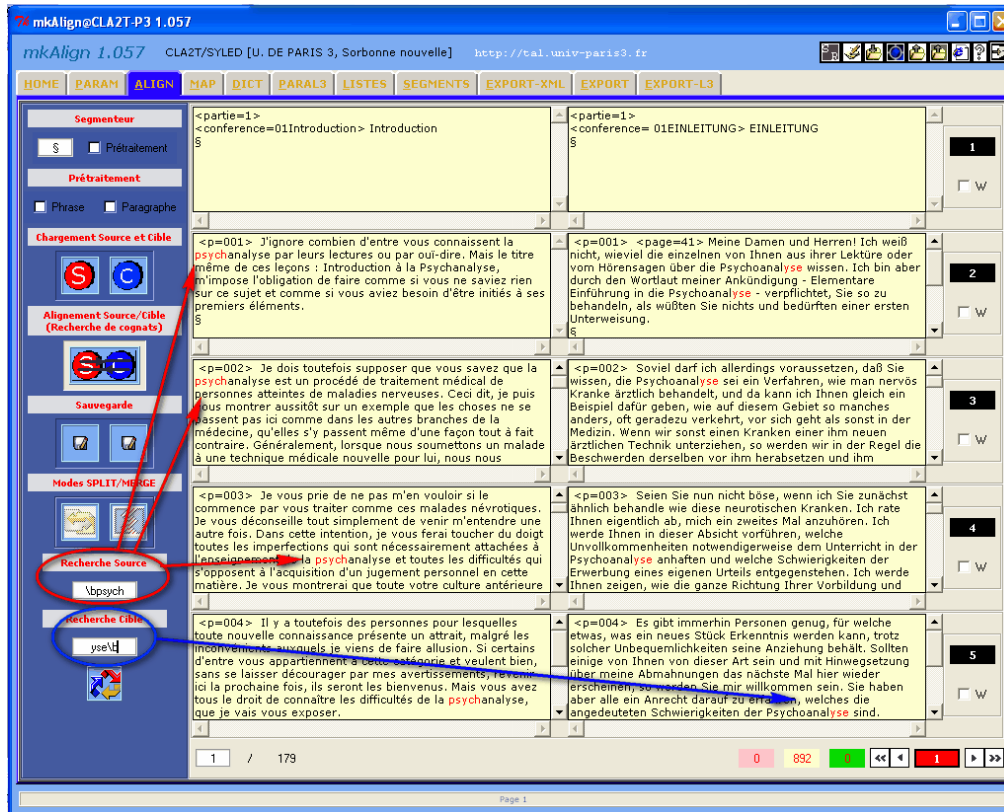


Figure 12 : **mkAlign** Onglet **ALIGN**, recherche de chaînes

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**. Les mots trouvés dans la figure sont colorés en rouge (**psych**analyse).

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*. Les mots trouvés dans la figure sont colorés en rouge (Psychoanal*yse*).

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

## 6.2 Exporter sous-corpus contenant un motif

A partir de la version 1.62 de **mkAlign**, on trouve à côté de chaque zone de saisie de recherche un bouton permettant de sauvegarder l'ensemble des cellules contenant le motif inséré dans la zone de saisie : on exporte ainsi un sous-corpus contenant toutes les sections contenant le motif sélectionné, ce texte est ensuite « réimportable » dans **mkAlign**.



## 6.3 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)

## 6.4 Fractionnement de cellule

- 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.

## 6.5 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
- 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.

## 6.6 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,

- et le compteur blanc le nombre de ligne qui ne sont pas les précédentes

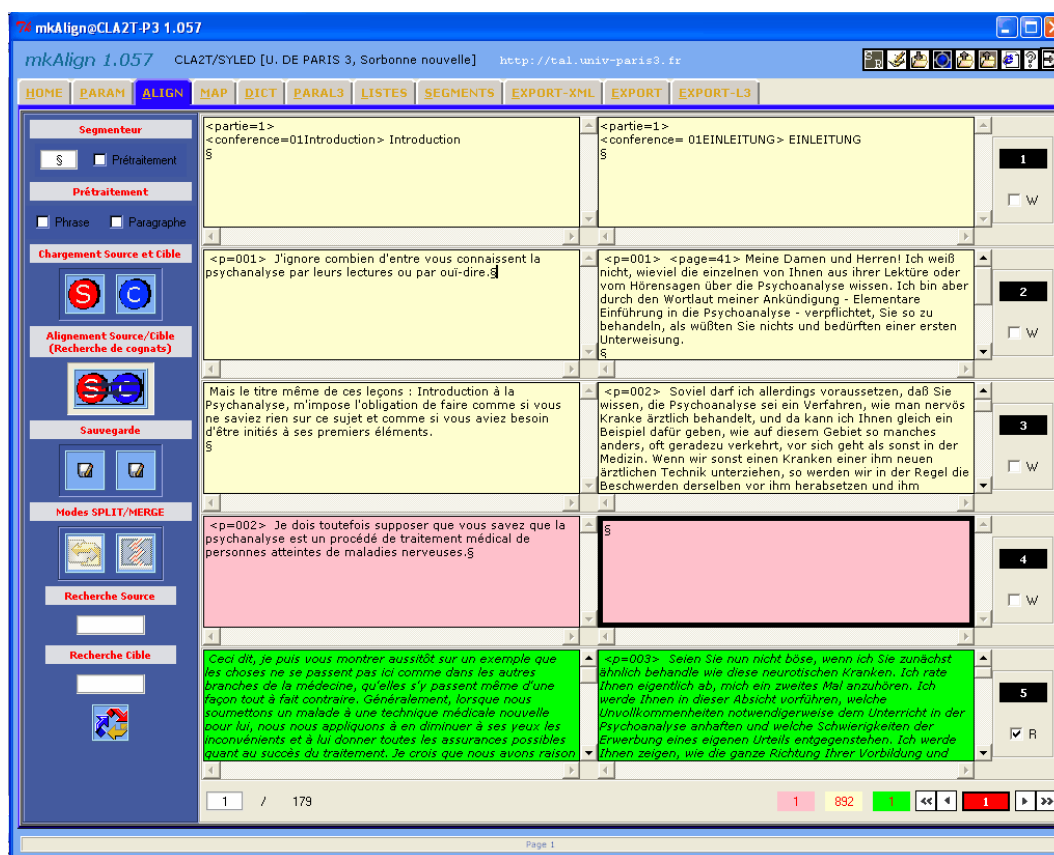


Figure 13 : mkAlign Alignement en cours

## 6.7 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*.

## 6.8 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 Représentation cartographique de l'alignement

### 7.1 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**<sup>3</sup> via l'outil "Carte des Sections"<sup>4</sup>). On trouvera dans l'onglet **MAP**, un bouton  permettant de construire cette représentation cartographique dans laquelle on disposera d'une carte des sections pour la *source* et d'une autre pour la *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 afficher (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.

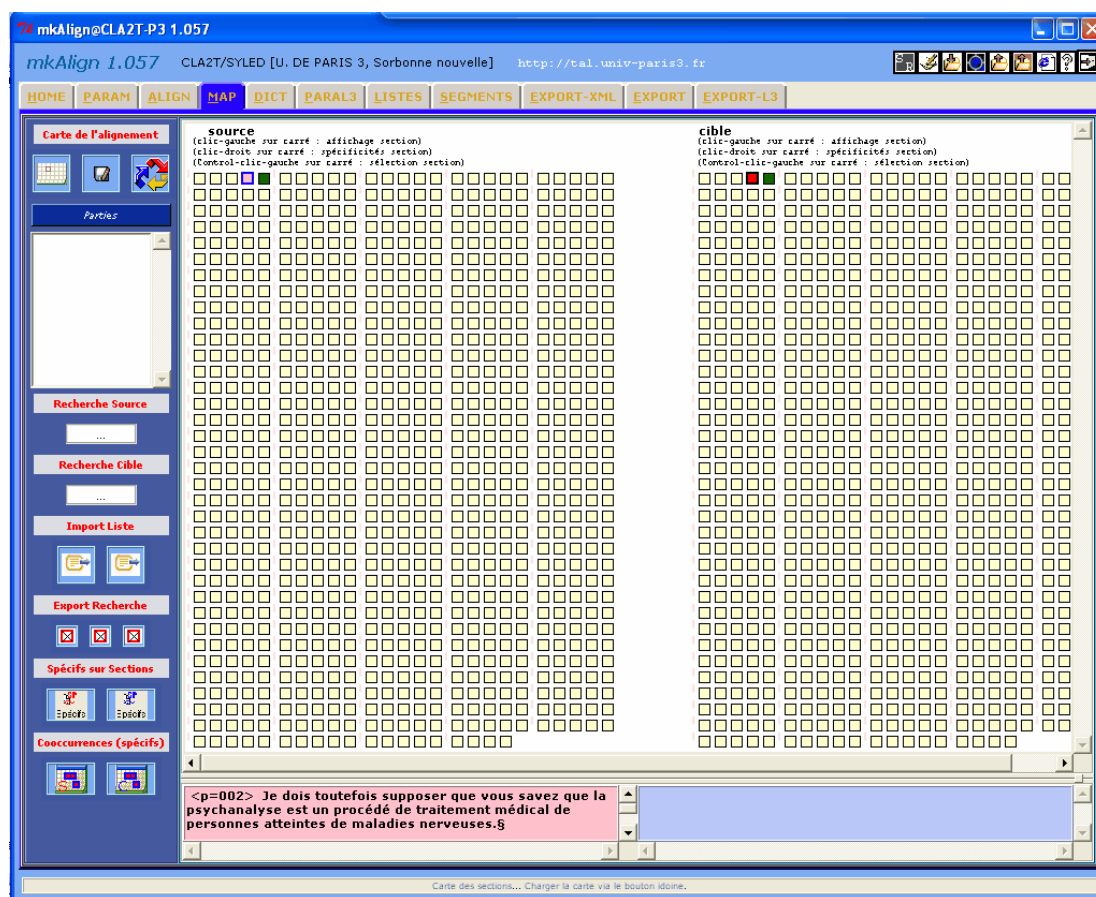


Figure 14 : **mkAlign** Onglet **MAP**

### 7.2 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 la carte *source* et dans la *cible*. On trouvera dans l'onglet **MAP** deux zones de saisie similaires à celles déjà présentées dans l'onglet **ALIGN**. La recherche peut être lancée dans le volet *source* ou *cible*.

<sup>3</sup> <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/>

<sup>4</sup> [http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/manuelsL3/manuel-3.41.htm#\\_Toc34045613](http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/manuelsL3/manuel-3.41.htm#_Toc34045613)

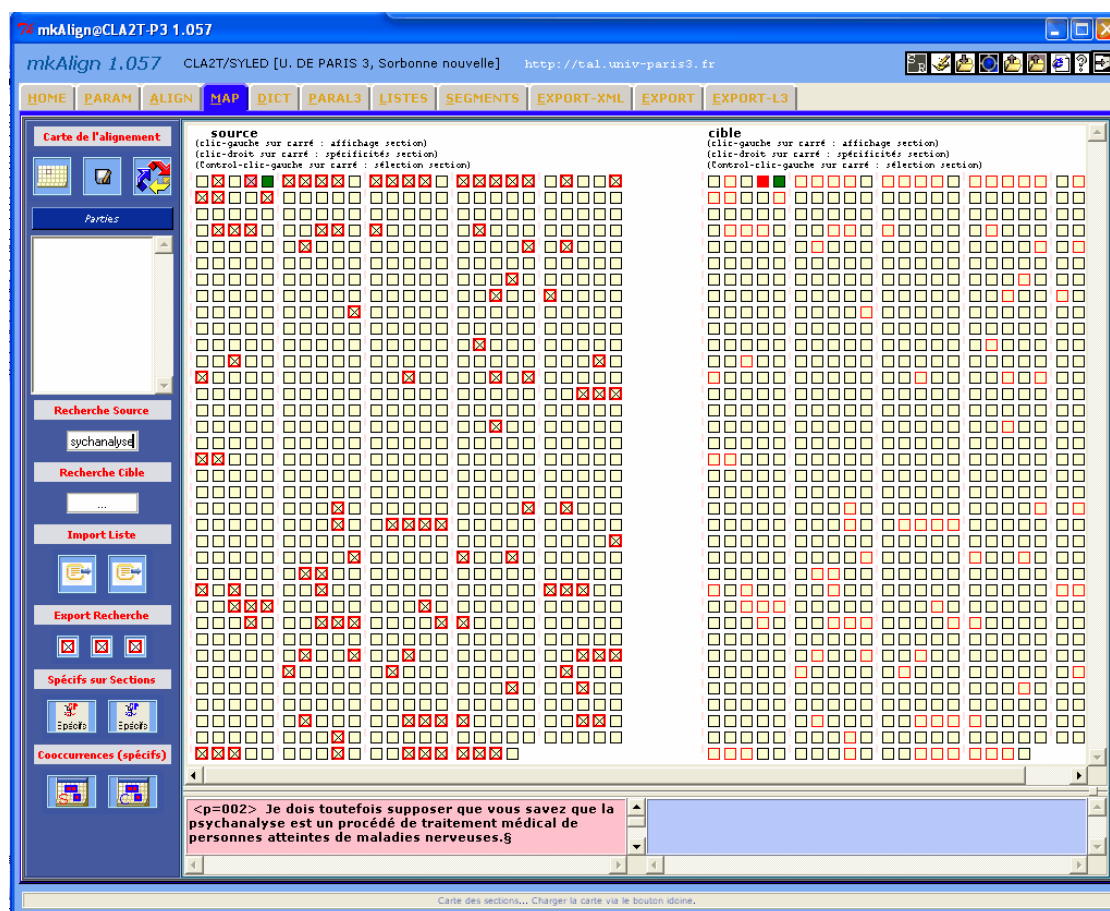
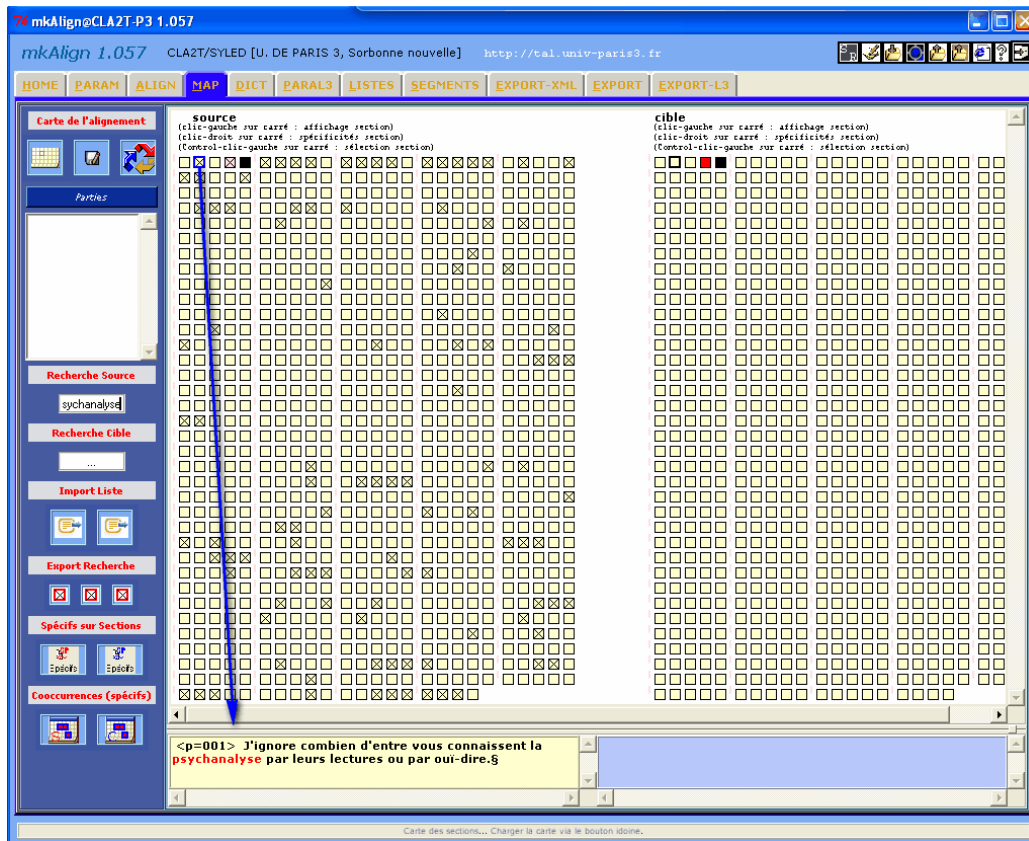


Figure 15 : **mkAlign** Recherche et Navigation dans la carte des sections (1)

Si la chaîne est trouvée, le carré correspondant est surligné en rouge (dans la carte SOURCE ou CIBLE) et les diagonales du carré sont dessinées. Les sections correspondantes (respectivement dans la carte SOURCE ou CIBLE) seront-elles aussi matérialisées par un surlignage rouge (trait fin). Le surlignage rouge sera inhibé après sélection d'un carré donné, par contre la croix sera maintenue et les résultats de la précédente requête resteront ainsi accessibles.

A l'issue d'une recherche de chaîne dans la carte, la zone d'édition obtenue par un clic gauche sur un carré donné prend en compte la chaîne de caractères présente dans la zone de saisie associée : la sélection d'un carré de la carte déclenche une recherche de la chaîne cherchée dans l'éditeur et produit le cas échéant l'affichage colorée de celle-ci.

Figure 16 : **mkAlign** Recherche et Navigation dans la carte des sections (2)

### 7.3 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 17 : **mkAlign** Carte / Sections / Parties

## 7.4 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.

## 7.5 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 corpus<sup>5</sup>.

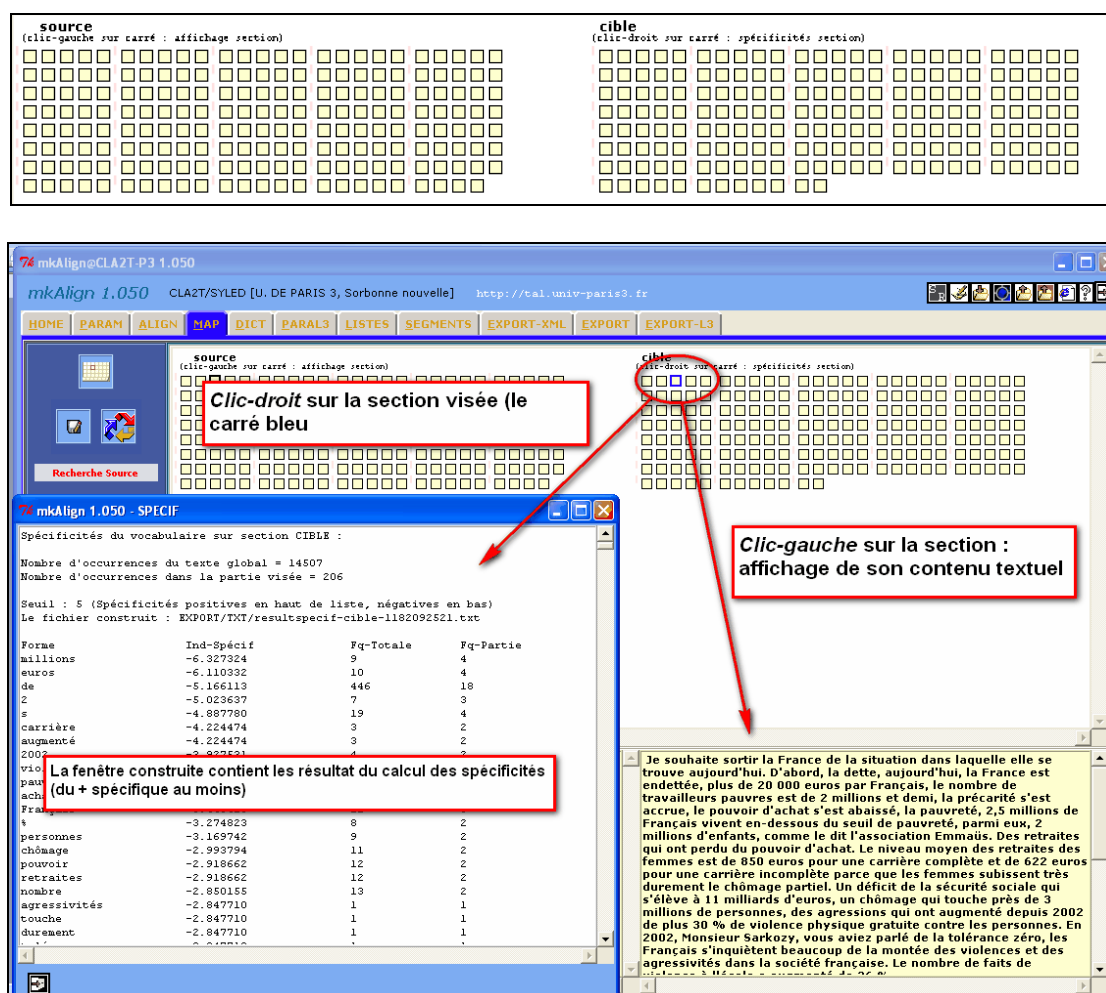


Figure 18 : mkAlign Carte / Sections / Spécificités

<sup>5</sup> Sur la méthode des spécificités on consultera par exemple : (Lafon, 1984) ou en ligne : <http://weblex.ens-lsh.fr/doc/weblex/contraste.html#specif>

## 7.6 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.

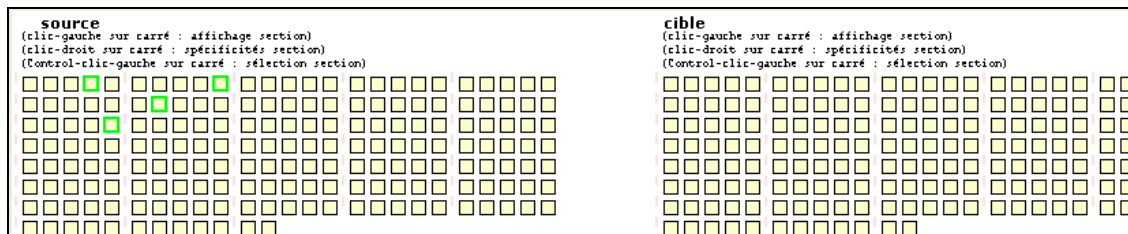




Figure 19 : **mkAlign** Carte / Sections / Sélection

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).

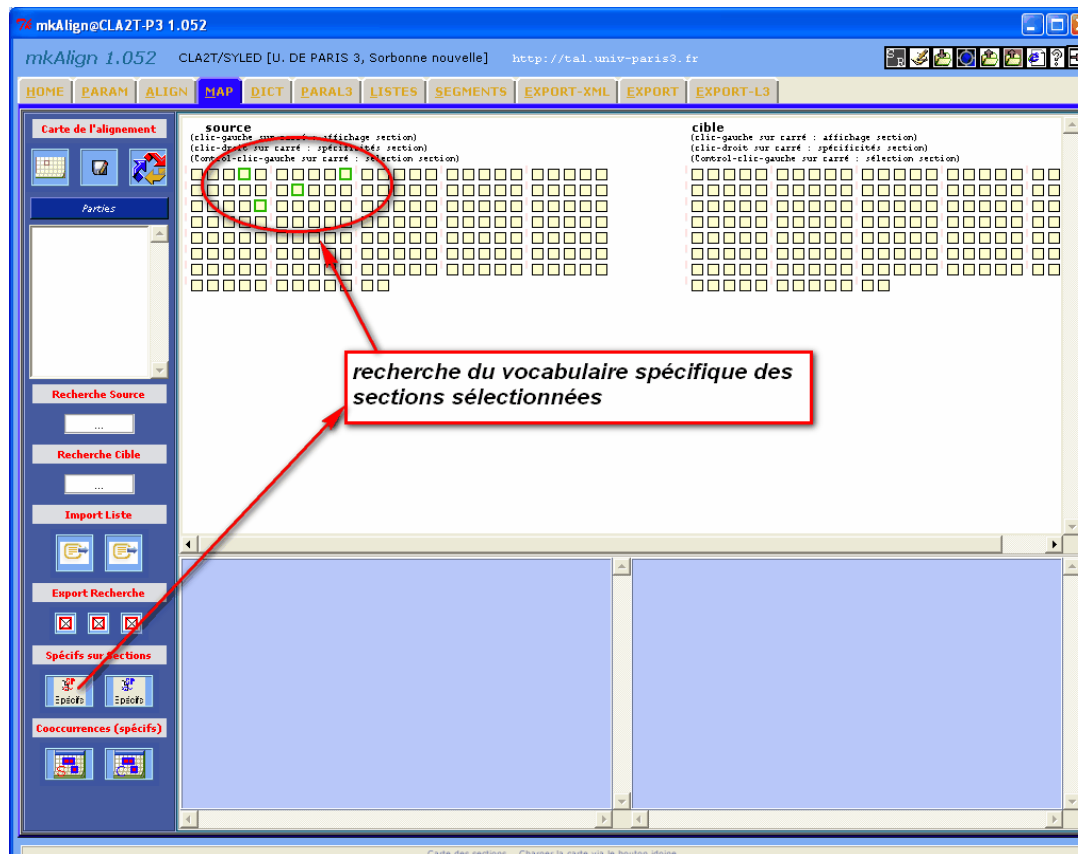


Figure 20 (1) : **mkAlign** Carte / Sections / Spécificité sections

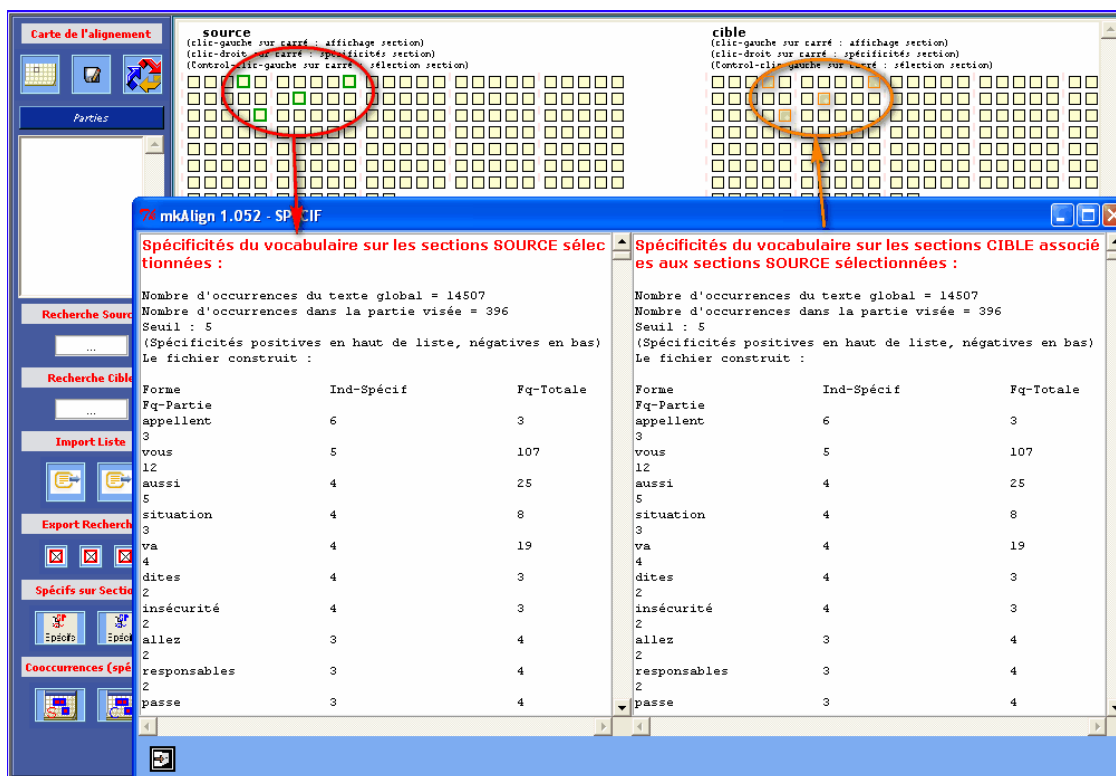





Figure 20 (2) : mkAlign Carte / Sections / Spécificité sections

### 7.7 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.

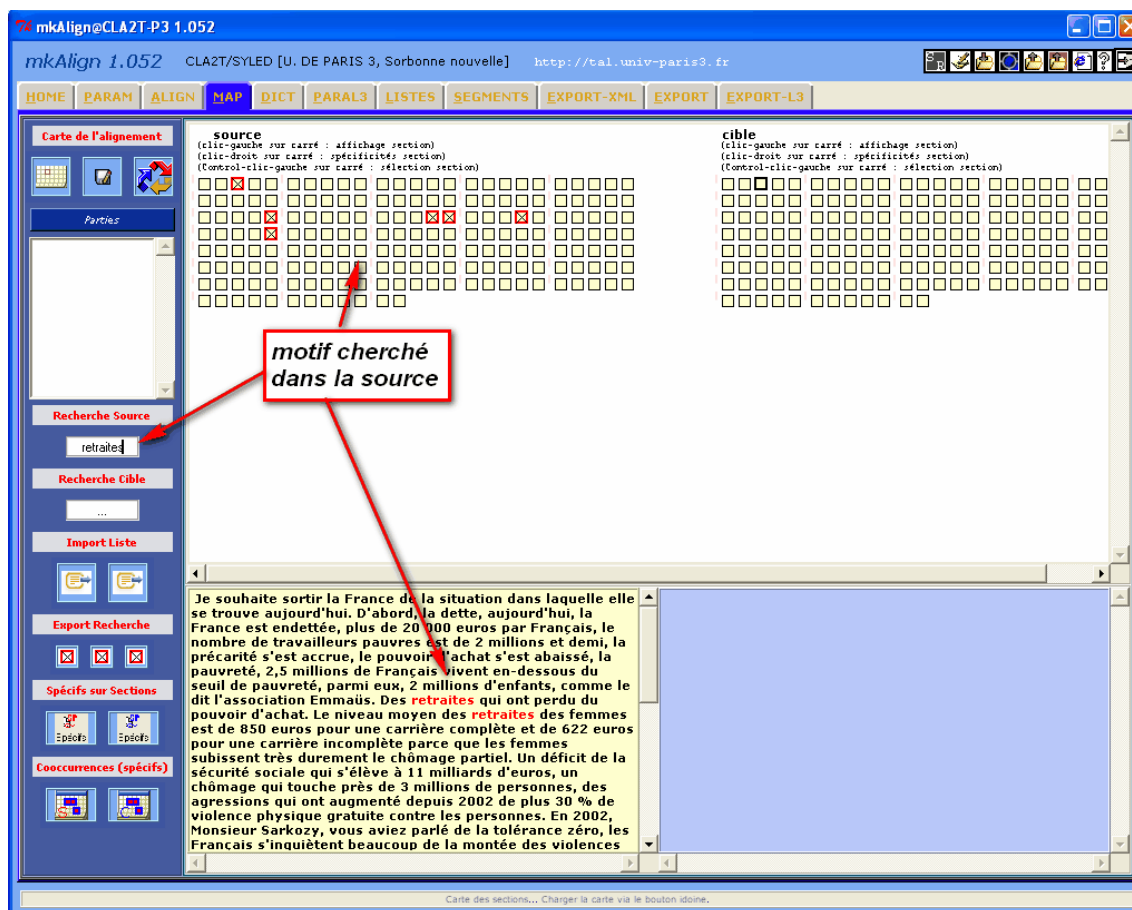


Figure 21 (1) : mkAlign Carte / Sections / Spécificités / Cooccurents

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).

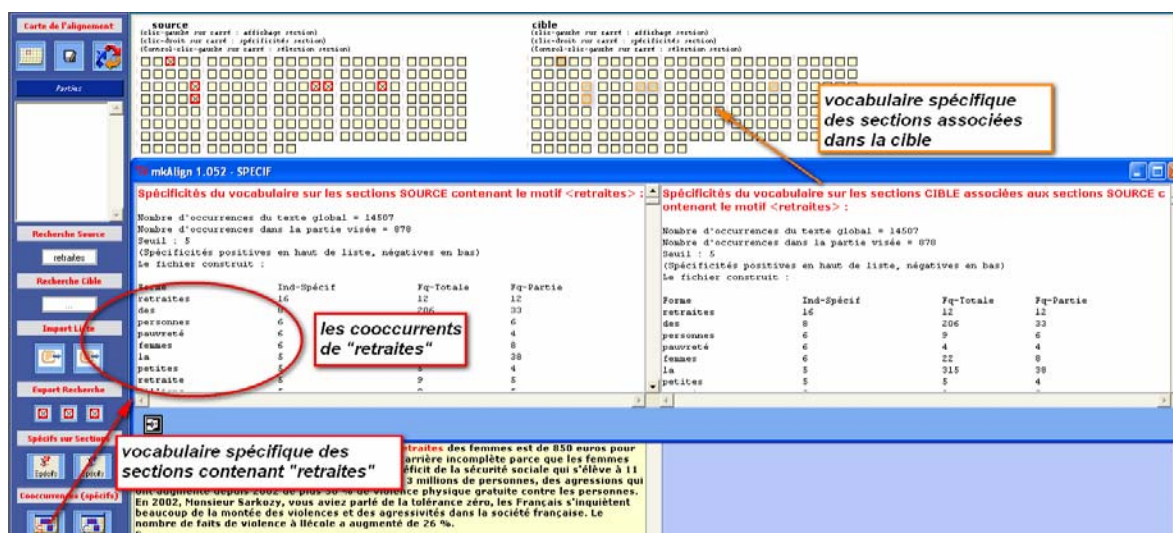



Figure 21 (2) : mkAlign Carte / Sections / Spécificités / Cooccurents

## 8 Importation de listes

L'onglet **MAP** 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.

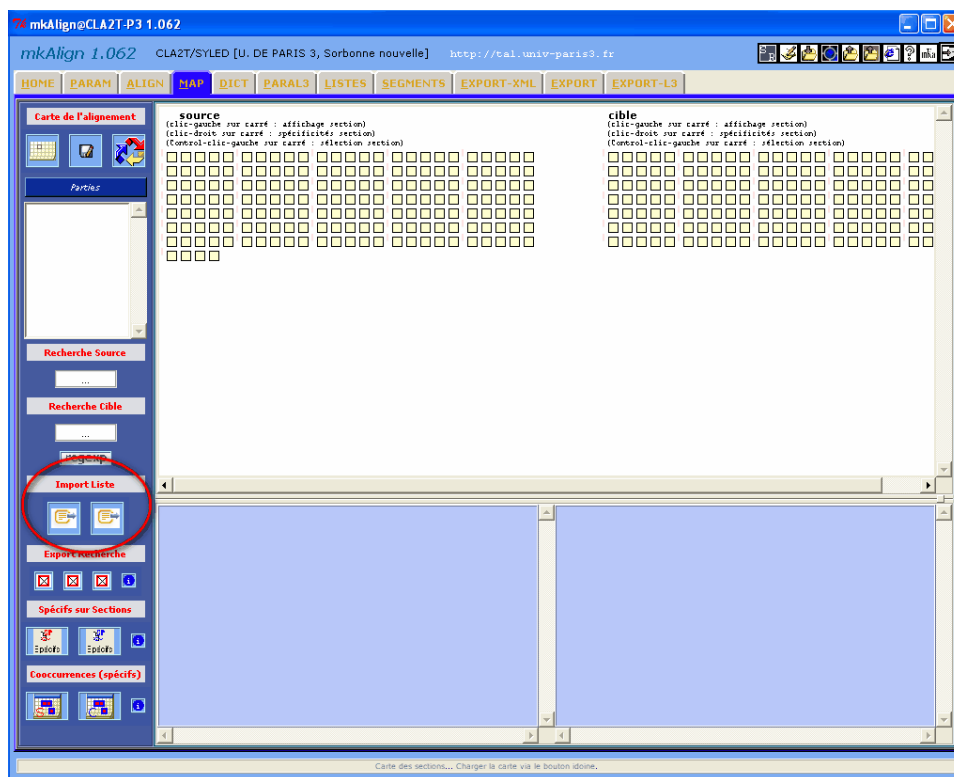


Figure 22 : **mkAlign** Import de listes

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

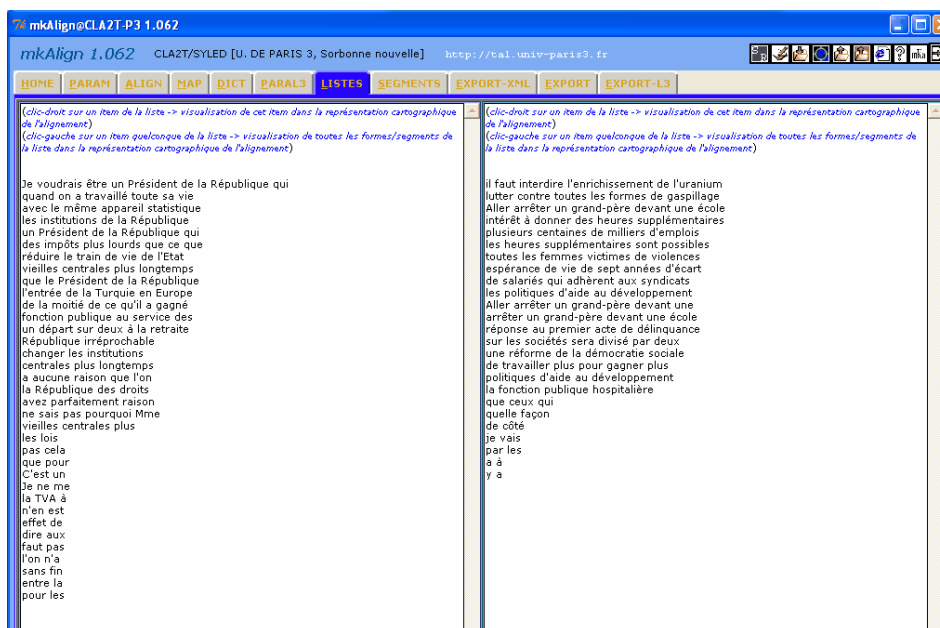


Figure 23 : **mkAlign** Onglet **LISTES**

Un clic droit sur un des items de la liste importée déclenche la recherche de cet item (forme ou segment) et l'affichage du résultat de cette recherche dans la représentation cartographique de l'alignement (*cf* onglet **MAP**).

Un clic gauche sur n'importe quel item de la liste déclenche la recherche de l'ensemble des items de la liste et l'affichage du résultat de cette recherche dans la représentation cartographique de l'alignement (*cf* onglet **MAP**).

## 9 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**.

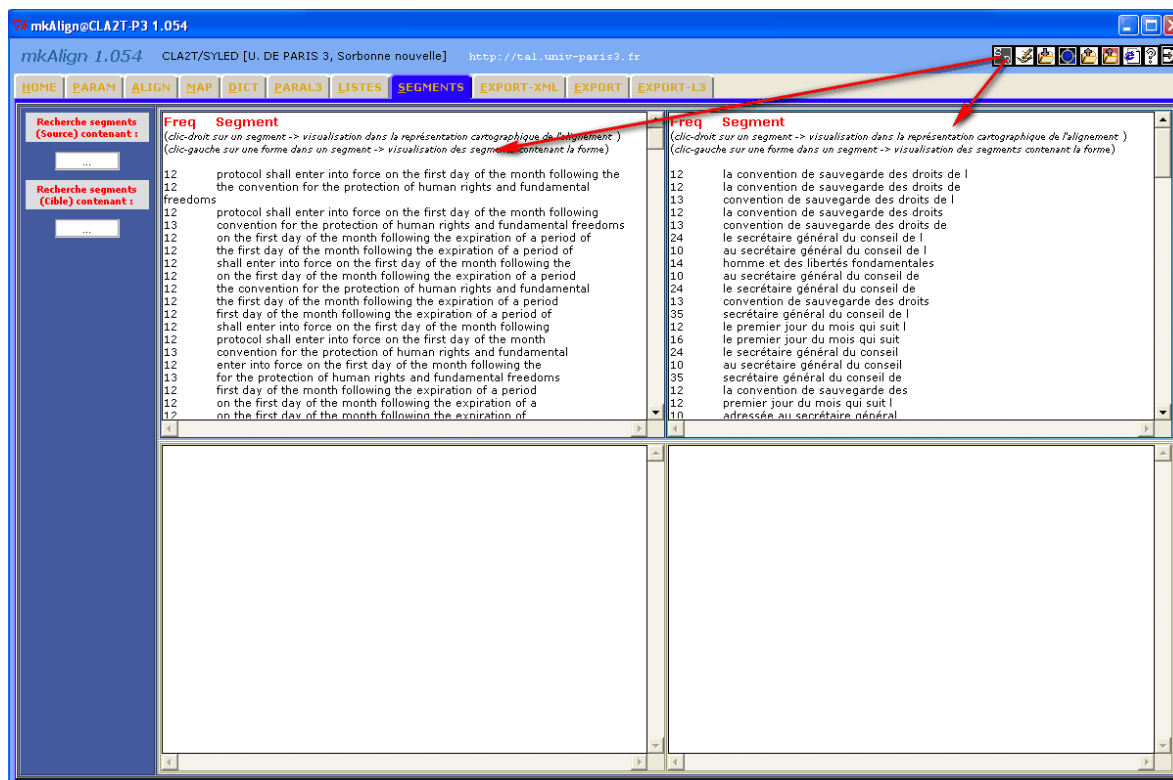


Figure 24 (1) : **mkAlign** Onglet **SEGMENTS**

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

Un *clic droit* sur un des segments de la liste déclenche l'affichage de la localisation du segment dans la représentation cartographique de l'alignement (cf onglet **MAP**).

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.

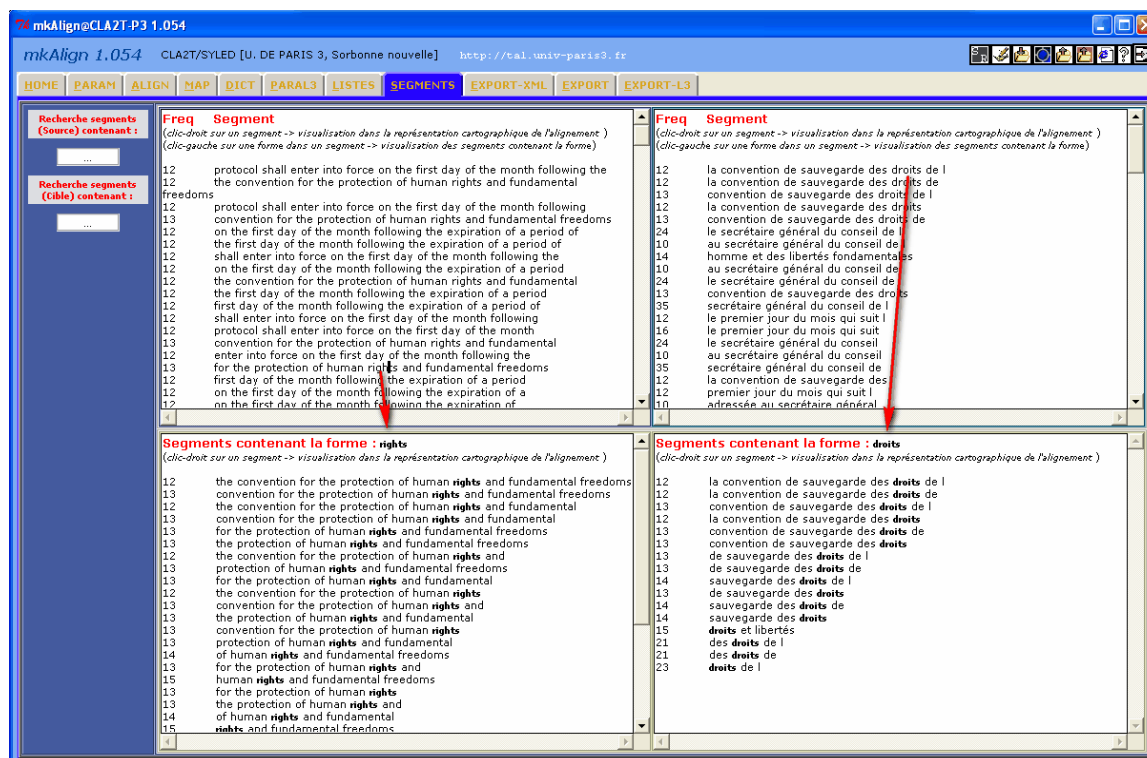


Figure 24 (3) : mkAlign Onglet SEGMENTS – sélection de segments

On peut aussi rechercher des segments contenant une forme donnée en utilisant les zones de saisie disponibles sur la gauche de l'onglet **SEGMENTS**. Un *clic droit* sur un des segments déclenche le même processus que ci-dessus.

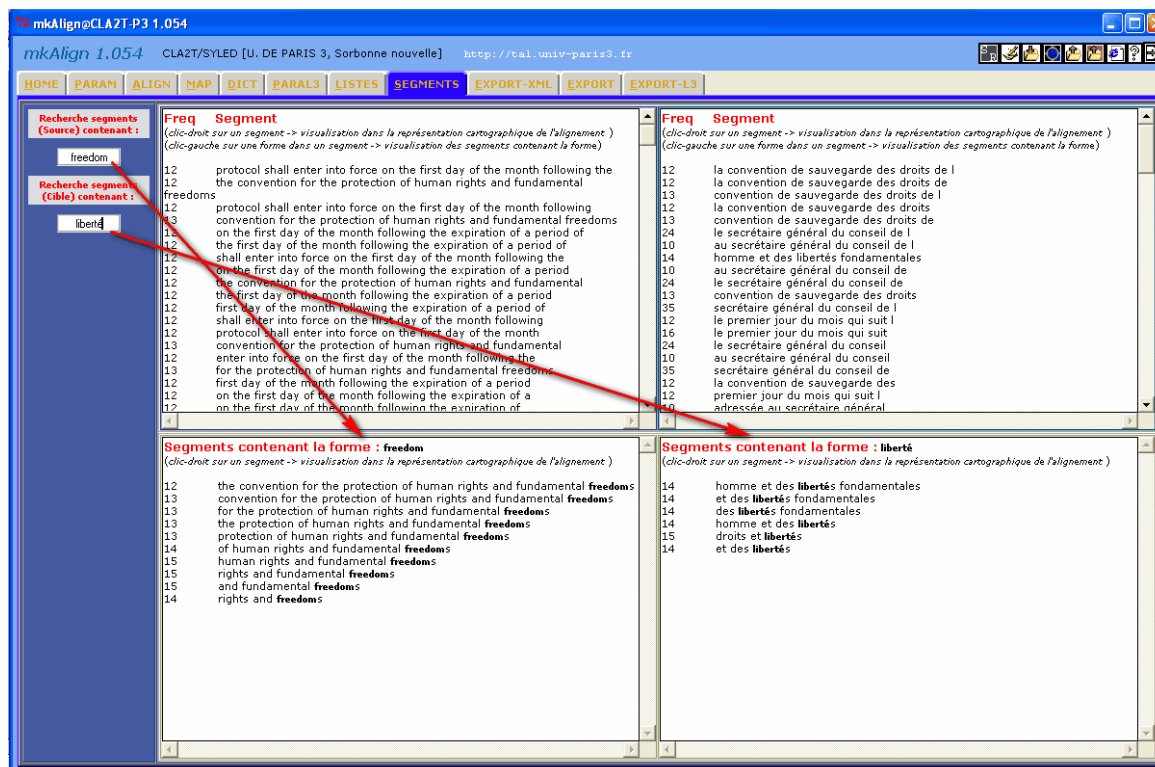



Figure 24 (4) : mkAlign Onglet SEGMENTS – sélection de segments

## 10 Export des traitements

### 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 cet export se trouve dans le répertoire de travail du programme et porte le nom MK-ALIGN-FOR-LEX3-date.txt (dans le répertoire EXPORT).

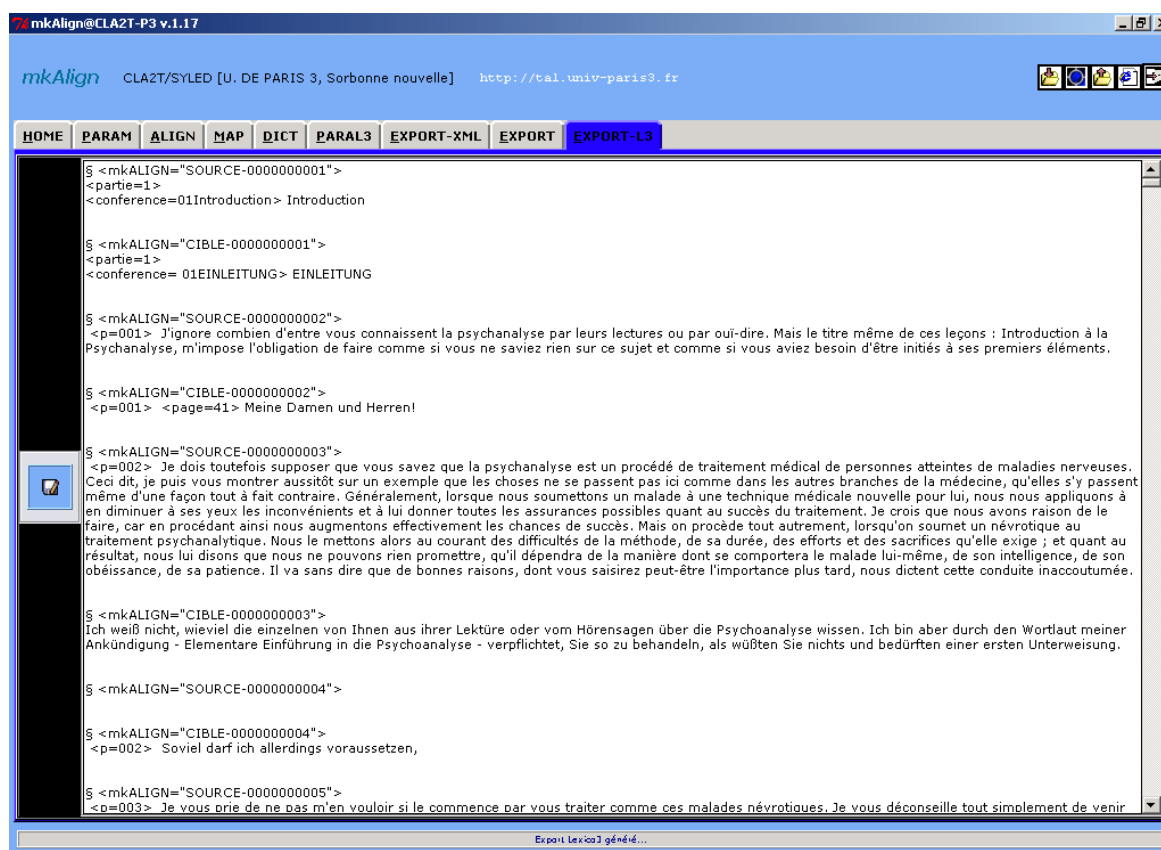


Figure 25 : *mkAlign* Onglet **EXPORT-L3**

## 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 cet export se trouve dans le répertoire de travail du programme et porte le nom MK-BUILT-ALIGN-date.txt. (dans le répertoire EXPORT/XML).



Figure 26 : **mkAlign** Onglet **EXPORT-XML**

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

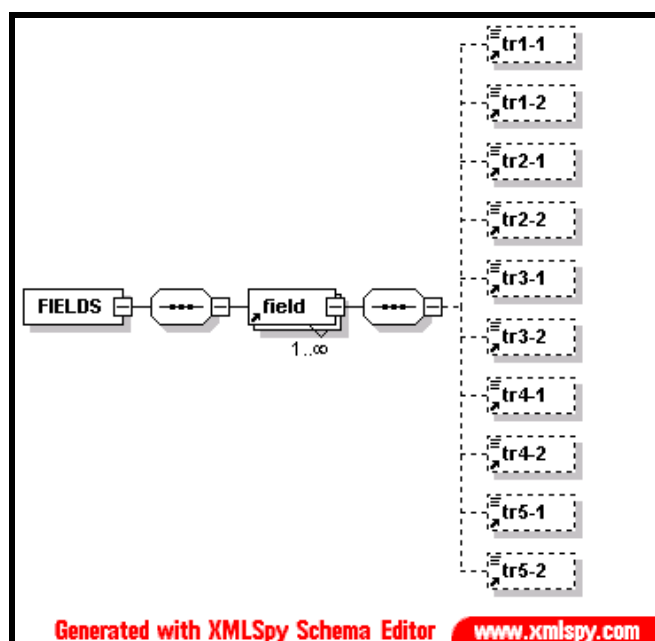



Figure 27 : **mkAlign** Schéma Export

***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*).

**Important** : il est nécessaire de renommer ce fichier d'export d'une session de travail à l'issue de sa création car ils seront supprimés au prochain lancement du programme.


## Export complet au format HTML

On dispose aussi dans l'onglet **EXPORT-XML** d'un bouton  permettant de visualiser l'état de alignement au format HTML via Internet Explorer (si celui-ci est disponible).

CLA <sup>2</sup> T [U. DE PARIS 3, Sorbonne nouvelle] mkAlign	
<p>§ &lt;mkALIGN="SOURCE-0000000001"&gt; &lt;partie=1&gt; &lt;conference=01Introduction&gt; Introduction</p> <p>§ &lt;mkALIGN="SOURCE-0000000002"&gt; &lt;p=001&gt; J&amp;apostrophe;ignore combien d&amp;apostrophe;entre vous connaissent la psychanalyse par leurs lectures ou par oui-dire. Mais le titre même de ces leçons : Introduction à la Psychanalyse, m&amp;apostrophe;impose l&amp;apostrophe;obligation de faire comme si vous ne saviez rien sur ce sujet et comme si vous aviez besoin d&amp;apostrophe;être initiés à ses premiers éléments.</p> <p>§ &lt;mkALIGN="SOURCE-0000000003"&gt; &lt;p=002&gt; 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&amp;apostrophe;elles s&amp;apostrophe;y passent même d&amp;apostrophe;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&amp;apostrophe;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&amp;apostrophe;elle exige ; et quant au résultat, nous lui disons que nous ne pouvons rien promettre, qu&amp;apostrophe;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 saisissez peut-être l&amp;apostrophe;importance plus tard, nous dictent cette conduite inaccoutumée.</p> <p>§ &lt;mkALIGN="SOURCE-0000000004"&gt; &lt;p=003&gt; Je vous prie de ne pas m&amp;apostrophe;en vouloir si le commence par vous traiter comme ces malades névrotiques. Je vous déconseille tout simplement de</p>	<p>§ &lt;mkALIGN="CIBLE-0000000001"&gt; &lt;partie=1&gt; &lt;conference= 01EINLEITUNG&gt; EINLEITUNG</p> <p>§ &lt;mkALIGN="CIBLE-0000000002"&gt; &lt;p=001&gt; &lt;page=41&gt; 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 wüßten Sie nichts und bedürften einer ersten Unterweisung.</p> <p>§ &lt;mkALIGN="CIBLE-0000000003"&gt; &lt;p=002&gt; 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 28 : mkAlign Export HTML

### 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 :

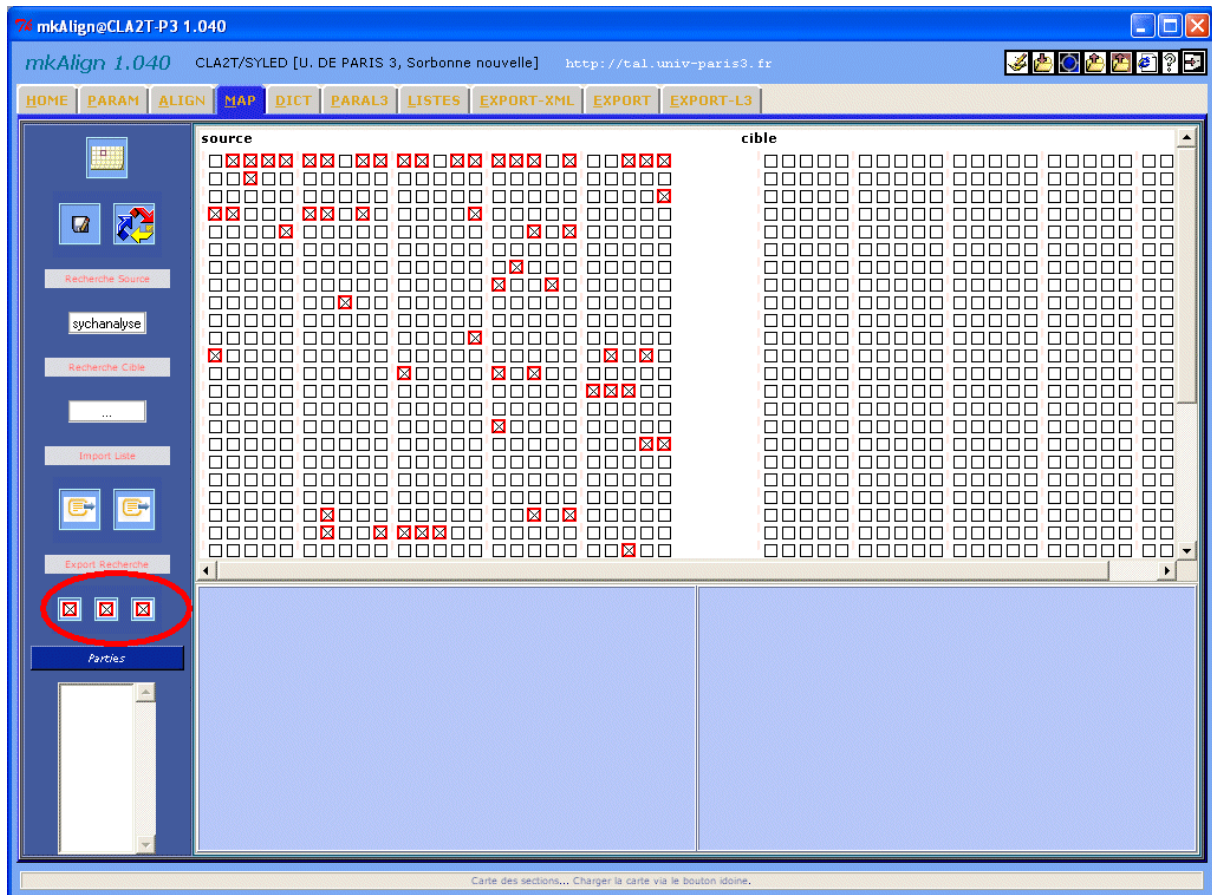


Figure 29 : **mkAlign** Export sélectif (sélection d'une forme)

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* Internet Explorer (si celui-ci est disponible) sous la forme suivante :

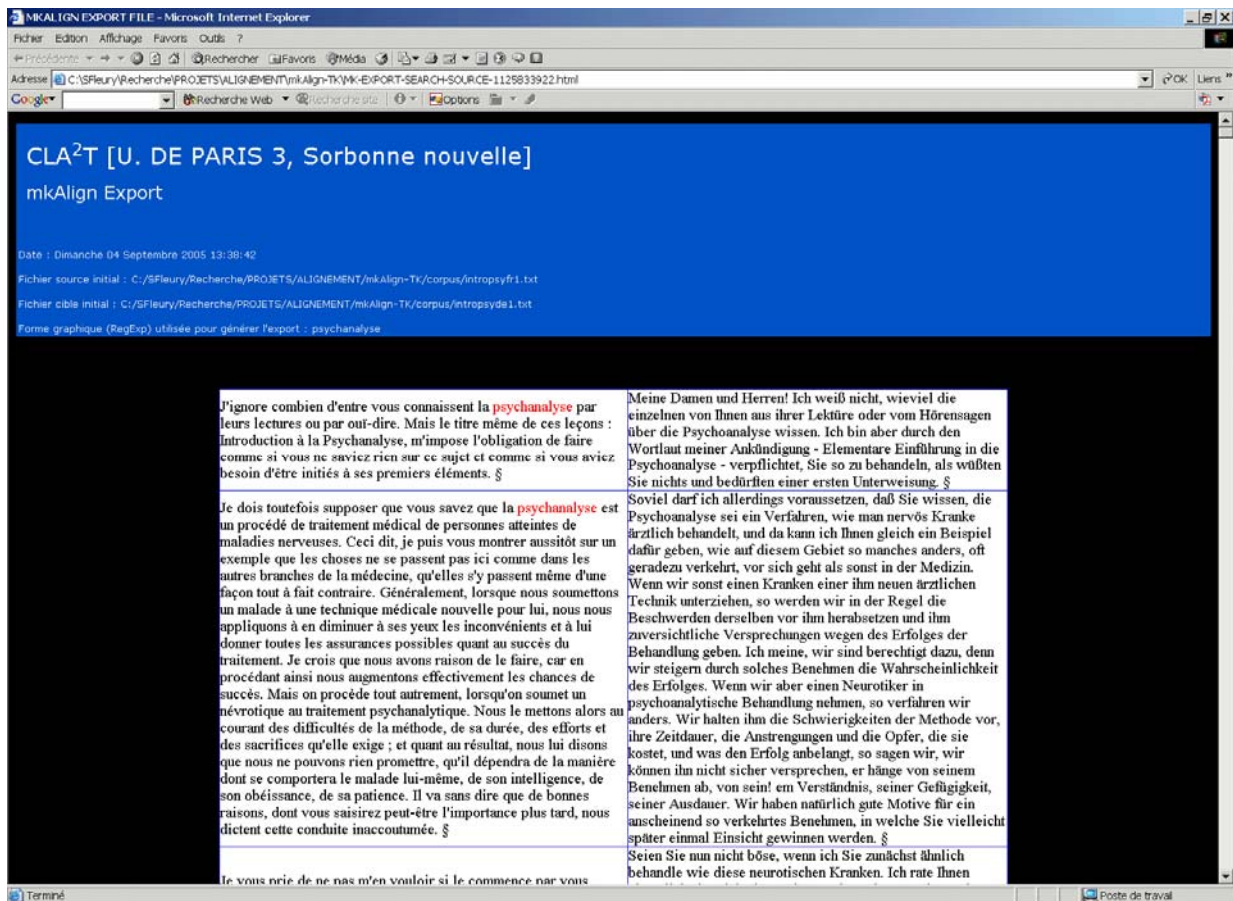


Figure 30 : **mkAlign** Résultat d'export sélectif - Fichier complet : [example-export.html](#)

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).

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).

### 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 ».

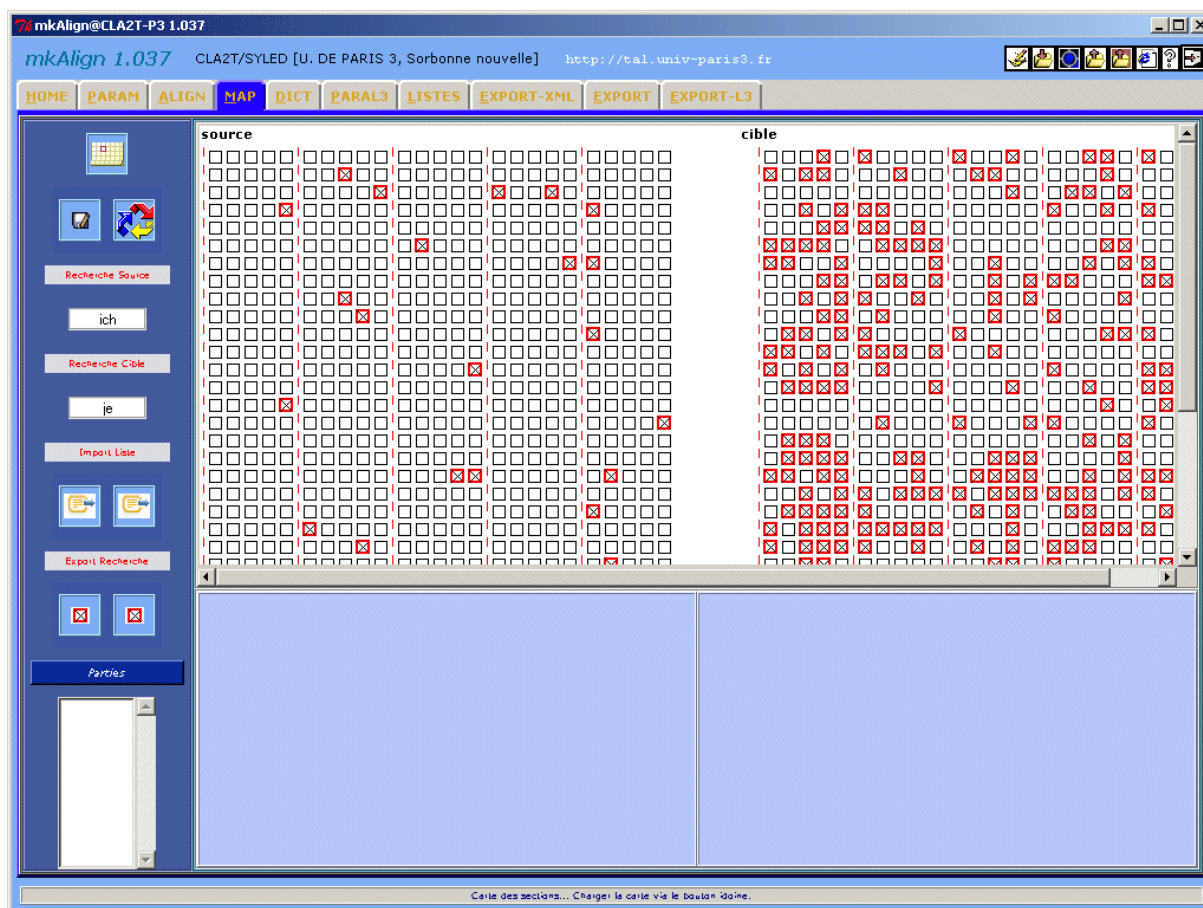




Figure 31 : **mkAlign**, Export de bi-textes

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 »).

## 11 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 .

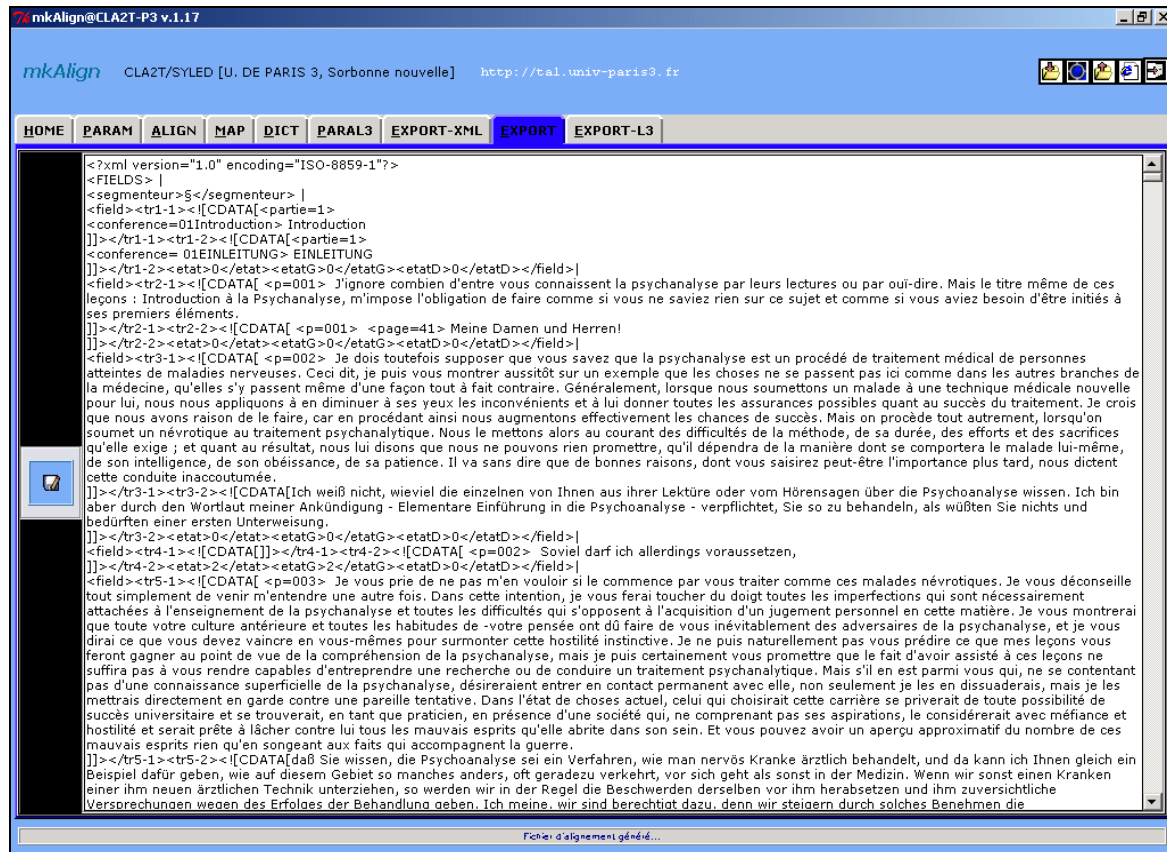
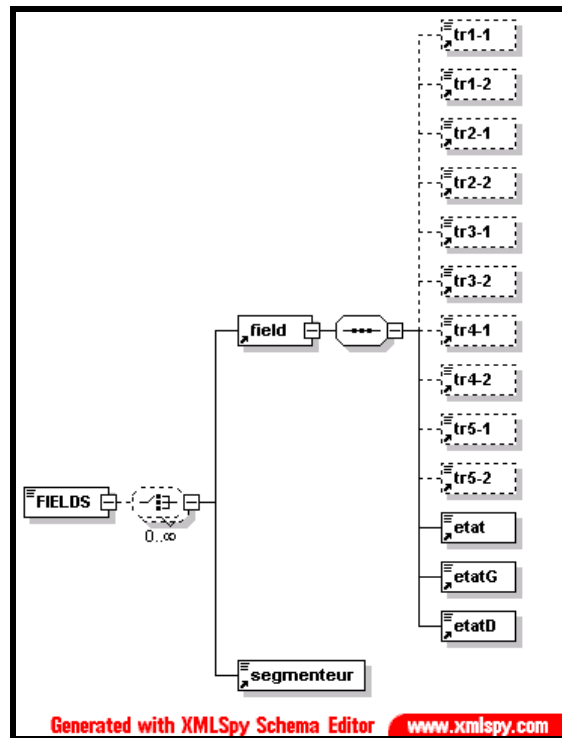


Figure 32 : **mkAlign** Onglet **EXPORT**

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 33 : *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`...)

## 12 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, *Actes de la 6ème conférence annuelle sur le Traitement Automatique des Langues Naturelles, 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 ([Version PDF](#)<sup>6</sup>), ([Version HTML](#)<sup>7</sup>) (sur le [site de Lexico](#)<sup>8</sup>).

[Lebart, 1994], Lebart L. et Salem A. (1994). *Statistique textuelle* ([livre en ligne](#)<sup>9</sup>) Paris, Dunod.

[Véronis, 2000] Véronis, Jean. Alignement de corpus multilingues ([Version PDF](#)<sup>10</sup>). 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](#)<sup>11</sup>) Présentation à la soutenance de thèse, Université de la Sorbonne nouvelle - Paris 3, le 26 novembre 2004.

[Zimina, 2004b] Zimina Maria, *L'alignement textométrique des unités lexicales à correspondances multiples dans les corpus parallèles*. ([Version PDF](#)<sup>12</sup>) Conférence aux 7es Journées internationales d'Analyse statistique des Données Textuelles [JADT'2004](#)<sup>13</sup>, 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*. ([Version PDF](#)) *Actes des 7es Journées scientifiques du Réseau de chercheurs*<sup>14</sup> "Lexicologie, Terminologie, Traduction", Institut supérieur de traducteurs et interprètes (ISTI), Bruxelles, 8-10 septembre 2005.

<sup>6</sup> <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/lex3-10pas/Lexico3-10premierspas.pdf>

<sup>7</sup> <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/lex3-10pas/index.htm>

<sup>8</sup> <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/index.htm>

<sup>9</sup> <http://www.cavi.univ-paris3.fr/lexicometrica/livre/st94/st94-tdm.html> ou <http://egsh.enst.fr/lebart/ST.html>

<sup>10</sup> <http://www.up.univ-mrs.fr/~veronis/pdf/2000hermes6.pdf>

<sup>11</sup> [http://www.cavi.univ-paris3.fr/ilpga/ED/student/stmz/ED268-](http://www.cavi.univ-paris3.fr/ilpga/ED/student/stmz/ED268-PagePersoMZ_fichiers/stmz/page6_fichiers/26novembre_MZ.zip)

[PagePersoMZ\\_fichiers/stmz/page6\\_fichiers/26novembre\\_MZ.zip](http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_118.pdf)  
<sup>12</sup> [http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT\\_118.pdf](http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_118.pdf)

<sup>13</sup> <http://www.jadt.org/>

<sup>14</sup> <http://www.ltt.auf.org/Journees-scientifiques-LTT-2005>

[Zimina, 2007] Fleury Serge, Maria Zimina, "*Exploring Translation Corpora with MkAlign*<sup>15</sup>", in *Translation Journal*, Volume 11, n°1 January 2007.

---

<sup>15</sup> <http://accurapid.com/journal/39mk.htm>

## 13 Annexes

### 13.1 Les expressions régulières

Une expression régulière peut comporter les éléments suivants :

- **^** désigne le début de champ<sup>16</sup>.
  - 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**.
- **(choix<sub>1</sub>|choix<sub>2</sub>|...|choix<sub>n</sub>)** désigne une des possibilités **choix<sub>1</sub>** ou **choix<sub>2</sub>** ou ... **choix<sub>n</sub>**.
  - 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).

<sup>16</sup> Le terme champ désigne soit la forme graphique d'un mot soit un segment (plusieurs mots consécutifs)

### 13.2 *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.