

# *Apprendre à vivre avec Transkribus : premiers pas*

Version préliminaire : 14 août 2020



**Préparé par :**  
Catherine Broué  
Justin Dubé  
Maxime Gohier



*Nouvelle France*  
**NUMÉRIQUE**

## **Table des matières**

<b>Introduction</b>	<b>3</b>
<b>1) Obtenir Transkribus</b>	<b>4</b>
<b>2) Entrer dans Transkribus</b>	<b>5</b>
<b>3) Comprendre l'organisation de l'interface Transkribus</b>	<b>6</b>
<b>4) Accéder aux collections de documents</b>	<b>7</b>
<b>5) Téléverser un document</b>	<b>7</b>
<b>6) Lancer une reconnaissance de texte</b>	<b>8</b>
<b>7) Ajuster les zones de texte</b>	<b>11</b>
<b>8) Déchiffrer le document</b>	<b>14</b>
<b>9) Autres fonctions</b>	<b>15</b>

## Introduction

Le présent document constitue un abrégé des principales étapes nécessaires pour se familiariser avec les fonctionnalités du logiciel Transkribus. Il est bien sûr possible de consulter directement le guide introductif (en français) élaboré par le consortium READ qui a mis au point ce logiciel : [https://transkribus.eu/wiki/images/1/1d/Comment\\_utiliser\\_Transkribus\\_%E2%80%93\\_en\\_10\\_%C3%A9tapes\\_ou\\_moins\\_with\\_Screenshots.pdf](https://transkribus.eu/wiki/images/1/1d/Comment_utiliser_Transkribus_%E2%80%93_en_10_%C3%A9tapes_ou_moins_with_Screenshots.pdf)

Le consortium READ fournit aussi, sur son site wiki ([https://transkribus.eu/wiki/index.php/How\\_to\\_Guides](https://transkribus.eu/wiki/index.php/How_to_Guides)), des guides spécifiques (en anglais, pour la plupart) à l'utilisation de certaines fonctionnalités du logiciel.

Enfin, une vidéo introductive (3:57 min.) est également disponible (en anglais) à l'adresse suivante : <https://www.youtube.com/watch?v=8Ei0a7WIITl>

## 1) Obtenir Transkribus

\*Pour plus d'informations sur le téléchargement et l'installation de Transkribus (en anglais): [https://transkribus.eu/wiki/index.php/Download\\_and\\_Installation](https://transkribus.eu/wiki/index.php/Download_and_Installation)

\*L'utilisation de Transkribus nécessite un logiciel Java approprié (version 8, contenant un module de développement JDK ou Java Development Kit). Ce logiciel peut être téléchargé directement sur le site web de Java, en choisissant la version adaptée à votre système d'exploitation : <https://java.com/de/download/manual.jsp#win>

Il est possible de vérifier la version de Java sur votre ordinateur en cliquant ici : [https://java.com/en/download/help/version\\_manual.xml](https://java.com/en/download/help/version_manual.xml)

Sur MacOS, plusieurs problèmes concernent l'utilisation du module de développement Java. Le cas échéant, consulter la foire aux questions de Transkribus ([https://transkribus.eu/wiki/index.php/Questions\\_and\\_Answers](https://transkribus.eu/wiki/index.php/Questions_and_Answers)), qui propose de multiples solutions, notamment sous la rubrique « **Wrong JAVA Version on Mac** ».

- a) Pour obtenir Transkribus, visitez le site web : <https://transkribus.eu/>  
L'image suivante devrait s'afficher :



- b) Cliquez sur « Register » pour créer un compte.

- c) Téléchargez Transkribus, en sélectionnant la case appropriée pour votre ordinateur.

\* Sur MacOS, il faudra d'abord glisser l'icône du logiciel dans le dossier « Applications ». Ensuite, il faut tenter de lancer le logiciel une première fois, puis se rendre dans les Préférences Système/Sécurité et confidentialité pour autoriser le lancement de l'application (malgré le fait qu'elle ait été téléchargée à partir d'un site inconnu).

\*Prenez soin d'effectuer les mises à jour proposées par Transkribus.

\*Il se peut que votre Transkribus ne soit plus fonctionnel après un ou deux ans. Vous devrez le télécharger de nouveau, à partir du site de Transkribus, pour obtenir une version plus récente. Pour ce faire vous devrez vous assurer que votre logiciel Java soit aussi à jour.

\*Pour les ordinateurs Macintosh datant d'avant 2014 et dont les systèmes d'exploitation ne peuvent plus être mis à jour, il faut utiliser une version plus ancienne de Transkribus. Toutes les anciennes versions sont disponibles pour téléchargement à partir du répertoire suivant : <https://transkribus.eu/TranskribusX/releases/mac/>.

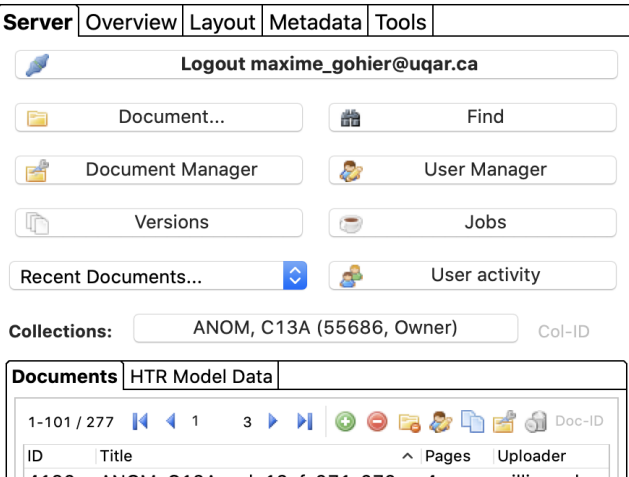
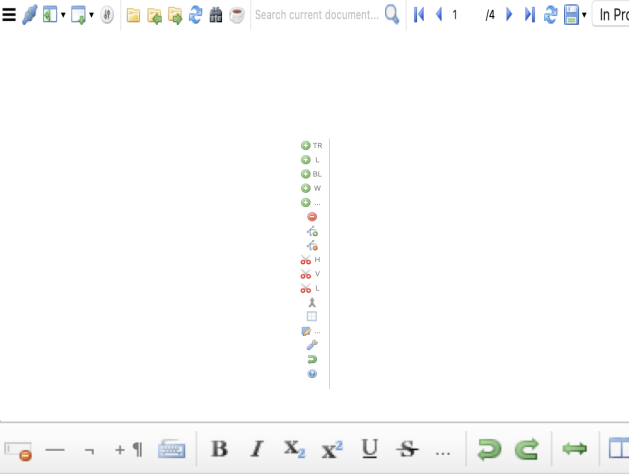
Pour obtenir une version plus ancienne pour Windows, consultez le répertoire suivant : <https://transkribus.eu/TranskribusX/releases/>.

## 2) Entrer dans Transkribus

Ouvrez Transkribus et cliquez sur « Login » pour accéder à votre compte.

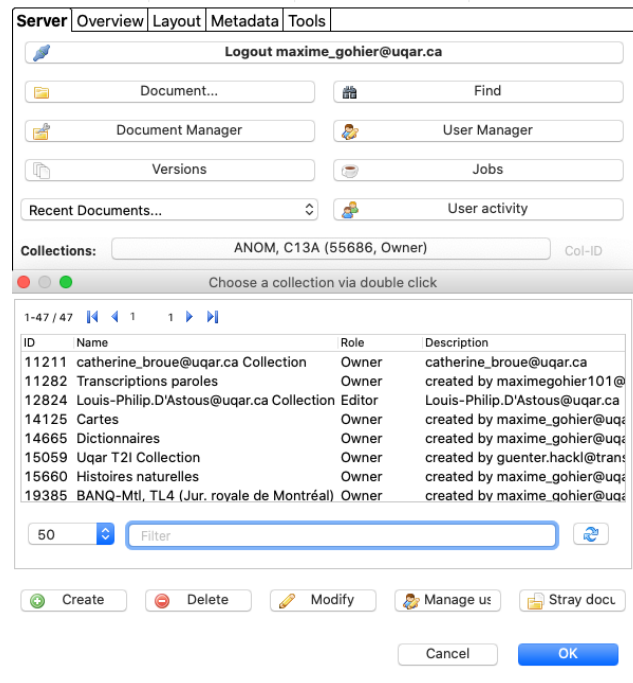


### 3) Comprendre l'organisation de l'interface Transkribus


a)	<p>Repérez le menu de travail, situé à gauche de la fenêtre. Ce menu comprend des onglets donnant accès aux outils de gestion et de traitement des documents.</p>	
b)	<p>Repérez les différentes barres d'outils :</p> <ul style="list-style-type: none"> <li>* la barre d'outils principale est située tout en haut de la fenêtre ;</li> <li>* située au centre de la fenêtre, la barre d'outils verticale donne accès aux outils de segmentation et de structuration du texte ;</li> <li>* située au bas de l'écran, la barre d'outils de mise en forme offre différents raccourcis pour accéder aux fonctions les plus courantes.</li> </ul>	

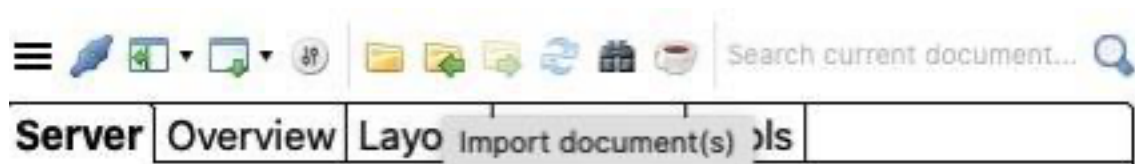
#### 4) Accéder aux collections de documents

Dans Transkribus, tous les documents sont associés à des collections. Pour accéder à une collection, cliquer dans le menu « Server » sur le nom de la collection situé immédiatement à droite de l'onglet « Collections ». Le menu « Choose a collection » s'ouvre alors, dans lequel s'affichent toutes les collections auxquelles vous avez accès, aussi bien celles que vous avez créées vous-mêmes que celles que d'autres utilisateurs ont partagées avec vous.




#### 5) Téléverser un document

- a) Pour télécharger un document sur Transkribus, il faut cliquer sur l'icône « Import document(s) », représentée par «  ».



- b) Vous arrivez sur la fenêtre « Document ingest/upload ».

	<ul style="list-style-type: none"><li>- Sélectionnez le format de document à téléverser. S'il s'agit de pages numérisées individuelles, sélectionnez « Upload single document ».</li><li>- Vous pouvez aussi choisir la collection dans laquelle mettre votre dossier. Pour un premier exercice, utilisez votre collection personnelle, attachée à votre compte.</li><li>- Cliquer sur «  » pour aller chercher votre document dans vos dossiers.</li></ul> <p>* Transkribus est en mesure de recevoir à peu près tous les types de fichiers images existants : .jpg, .jpg2000, .tiff, .png, etc.</p> <p>* Notez qu'il est préférable de limiter la résolution des images versées dans Transkribus. Ce logiciel, en effet, n'exige pas une très haute résolution pour obtenir des résultats satisfaisants. Au contraire, des images de trop haute résolution risquent de ralentir le traitement des données et le logiciel.</p>
c)	<p>Sélectionnez les pages du document à téléverser et cliquer sur « Upload ».</p> <p>* Pour téléverser un document composé de plusieurs images, il faut d'abord placer toutes les images dans un même dossier. Transkribus attribue automatiquement le titre de votre dossier au document.</p>
<b>6) Lancer une reconnaissance de texte</b>	
<p>Le tutoriel de Transkribus <i>Automatic segmentation</i> (en anglais) présente comment segmenter un texte en régions, en lignes et en lignes de base à l'aide des outils du logiciel : <a href="https://www.youtube.com/watch?v=b0MqNwSIKC0">https://www.youtube.com/watch?v=b0MqNwSIKC0</a></p>	
a)	<div><div>Dans la barre d'outils principale, sous l'onglet « Server », vous pouvez sélectionner votre collection et naviguer dans vos documents.</div><div><div>Server</div><div>Overview</div><div>Layout</div><div>Metadata</div><div>Tools</div><div></div></div></div>



Sous l'onglet « Overview », apparaissent toutes les pages de votre document et l'état de leur traitement.

Au-dessus de votre document, vous pouvez changer de page avec les flèches et enregistrer votre document.

Server	Overview	Layout	Metadata	Tools
Document:	ANOM, C13A, vol. 13, f. 265-266v, ID: 418212			
Collection:	ANOM, C13A, ID: 55686			
Filename:	FRANOM34_C13_013_0541.jpg			
Image URL:	<a href="https://files.transkribus.eu/Get?id=ALCAYDYDCFDXW">https://files.transkribus.eu/Get?id=ALCAYDYDCFDXW</a>			
Transcript URL:	<a href="https://files.transkribus.eu/Get?id=XTHBDMTXBTUBB">https://files.transkribus.eu/Get?id=XTHBDMTXBTUBB</a>			
Page/Transcript ID:	16658551/32781534			
Thumbnail Overview:				
<input type="button" value="Show Document Manager"/>				
1: In Progress   2: In Progress   3: In Progress   4: In Progress				



- b) Pour effectuer une reconnaissance de texte, allez sur l'onglet « Tools ».

Vous pouvez appliquer l'outil à toutes les pages ou à une sélection de pages.

Activez d'abord l'outil de « Layout Analysis » en cliquant le premier bouton « Run » pour exécuter la segmentation des zones de textes.

Ensuite, cliquez sur le bouton « Run » de l'outil « Text Recognition » pour activer la reconnaissance textuelle. Cette fonction vous permet de sélectionner un modèle de

Server	Overview	Layout	Metadata	Tools
▼ <b>Layout Analysis</b>				
Method:		CITlab Advanced		<input type="button" value="Configure..."/>
<input type="radio"/> Current page		<input checked="" type="radio"/> Pages (7): 1-7 <input type="button" value="..."/>		
<input checked="" type="checkbox"/> Find Text Regions		<b>Only use unsegmented page:</b>		
<input checked="" type="checkbox"/> Find Lines in Text Regions				
<input type="button" value="Run"/>				
▼ <b>Text Recognition</b>				
Method:		HTR (CITlab)		
<input type="button" value="f"/>		Models...		
<input type="button" value="Run..."/>				

	<p>reconnaissance approprié à votre document.</p> <p>Le bouton « Models » vous permet d'analyser tous les modèles de reconnaissance et de segmentation textuels associés à votre collection et disponibles pour votre document.</p>	
c)	<p>Sous l'onglet « Server », le sous-onglet « Jobs » permet de consulter l'état d'avancement des commandes que vous avez lancées (« Running » ou « Finished »).</p> <p>Dites oui (« Yes ») lorsque Transkribus vous demandera de recharger la page courante (« reload the current page »).</p>	
d)	<p>Une fois la segmentation effectuée, vos pages seront parcourues de cases vertes et bleues.</p> <p>Les cases vertes sont des zones de texte. Elles identifient et circonscrivent des paragraphes.</p> <p>Les cases bleues identifient et circonscrivent les lignes de textes.</p> <p>Chaque ligne de texte contient aussi une « Baseline » (en mauve), qui indique le sens de lecture du texte d'une ligne.</p>	

100=

On aura de la peine à se persuader, une chose, qui ne  
 laisse cependant pas d'être véritable parmi les paysans  
 des bois, qui ont quelque fois une si grande bêtise de croire  
 qu'ils sont obligés d'abandonner leur part de la viande  
 des animaux, et que la faim leur fait trouver tout bon,  
 et même cette fausse vertu. Je connois fort particulie-  
 rement un francois qui se trouvant en France à quel-  
 ques années, a plusieurs fois vu ce que l'enfant prodigue  
 des bois, il auroit plongé sa tête dans une mangeaille  
 immonde, pour y attraper quelque tron de chon pour  
 se satisfaire la faim, et je puis dire par ma propre

## 7) Ajuster les zones de texte

Pour en connaître plus sur la segmentation des zones, consultez le tutoriel vidéo de Transkribus :  
<https://www.youtube.com/watch?v=b0MqNwSIKC0>

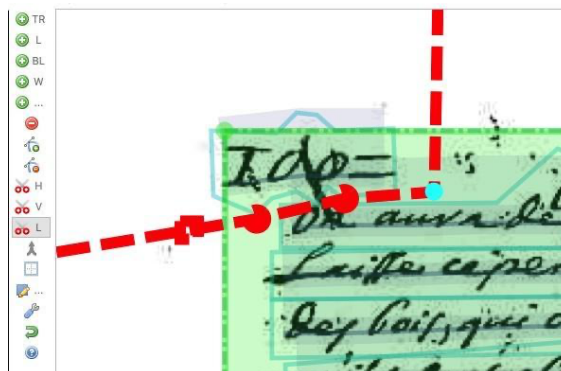
- |    |  |
|----|--|
| a) | <p>La première étape consiste à ajuster les zones de textes que le logiciel n'aura parfois délimité qu'imparfaitement.</p> <p>Par exemple, il est utile de séparer les zones de textes en fonction des blocs de textes réels (paragraphes, paratexte, numéros de pages, sceaux d'institutions d'archives, etc.). Ainsi, dans l'exemple précédent, il importe de séparer le numéro de folio (« 100 ») du reste du corps du texte.</p> |
|----|--|

Pour ce faire, sélectionnez la zone de texte contenant le texte à découper. Ensuite, utilisez les ciseaux pour exécuter le découpage.

Trois types de ciseaux peuvent être employés : le premier (« H ») coupe la région de manière horizontale, le second (« V »), de manière verticale, et le troisième (« L ») permet le découpage à angles.

Lorsque l'on utilise le ciseau angulé (« L »), on crée un angle chaque fois que l'on clique sur le bouton gauche de la souris, ce qui applique un point rouge permettant de continuer la sélection libre.

Pour terminer un découpage (avec les trois outils), double-cliquez sur le bouton gauche de la souris.



### Autres commandes

La barre horizontale propose plusieurs fonctions, dont nous détaillons les plus utiles ci-dessous :



+TR : Créer une zone de texte ou *Text Region* (verte)



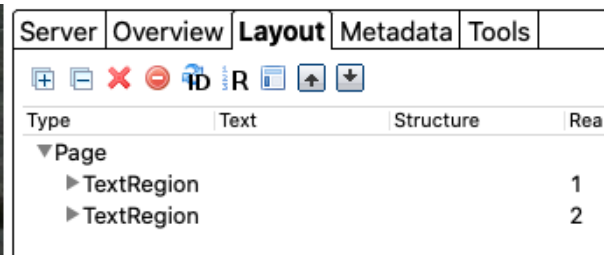
+L : Créer une ligne ou *Line* (bleue)

+BL : Créer une ligne de base ou *Baseline* (mauve)

+W : Créer une zone de mot ou *Word* (rouge)


+... : Créer tout autre type de zone

Notez que lorsque l'on crée une ligne de base (Baseline), une ligne lui est automatiquement associée par défaut.

		<p>Nous vous conseillons donc de vous borner à utiliser les fonctions « +TR » et « +BL », qui sont utiles pour ajouter/ajuster les zones de textes et les lignes.</p> <hr/> <p>La commande «  » sert à supprimer une zone.</p> <p>Les deux ciseaux (+) et (-) servent à ajouter ou supprimer des points délimitant les zones, ce qui permet de peaufiner la délimitation des zones.</p> <p>Les trois ciseaux (H, V et L) ont déjà été présentés.</p> <p>Les deux flèches noires convergentes servent à fusionner deux zones.</p> <p>Nous sautons quelques commandes moins utiles pour finir avec les deux suivantes :</p> <p>La flèche verte de retour permet d'annuler la dernière modification effectuée.</p> <p>Le « ? » permet de découvrir des raccourcis-clavier pour activer les différentes commandes.</p> <p>*La flèche verte située dans la barre verticale permet seulement d'annuler des modifications faites à la segmentation du texte. Pour annuler une modification faite à la transcription, il faut utiliser la même flèche située dans la barre de mise en forme, au bas de l'écran.</p>
b)	<p>Dans l'onglet « Layout » du menu de travail, vérifiez que les zones de texte et les lignes soient ordonnancées de façon cohérente, en respectant un ordre de lecture logique</p>	

Vous pouvez ajuster l'ordre des zones en usant des flèches ou manuellement.

Vous pouvez aussi ajuster l'ordre des zones de lignes de texte (bleues).

Notez que le découpage de zones de texte génère parfois des lignes résiduelles vides dans l'une des deux zones. Supprimez ces lignes vides au moyen de l'icône «  ».

ServerOverviewLayoutMetadataTools

Type	Text	Structure	Reading	ID
▼ Page				
▼ TextRegion			1	TextReg...
▶ Line	At		1	r1i1
▼ TextRegion			2	TextReg...
▶ Line	ou auà ée ne...		1	r1i2
▶ Line	Laie ce pe da...		2	r1i3
▶ Line	ey oiri que on...		3	r1i4
▶ Line	qu' rout aotge...		4	r1i5
▶ Line	ley anemauxj...		5	r1i6

8) Déchiffrer le document

a)

Une fois que les zones de texte sont ajustées et organisées correctement, la transcription peut commencer.

À moins que vous décidiez d'entraîner votre propre modèle de reconnaissance en effectuant la transcription vous-même (si vous êtes un paléographe chevronné et que votre corpus présente une écriture très particulière), il sera plus productif, pour commencer, d'utiliser un modèle de reconnaissance déjà existant. Selon l'efficacité de ce modèle, vous aurez des ajustements plus ou moins grands importants à effectuer : un modèle est toujours perfectible.

Dans le cas d'un modèle en développement, il est encore plus important de faire un bon travail de déchiffrement.

EXPLIQUER LA DÉMARCHE POUR CHOISIR UN MODÈLE DE RECONNAISSANCE.

1-1 At

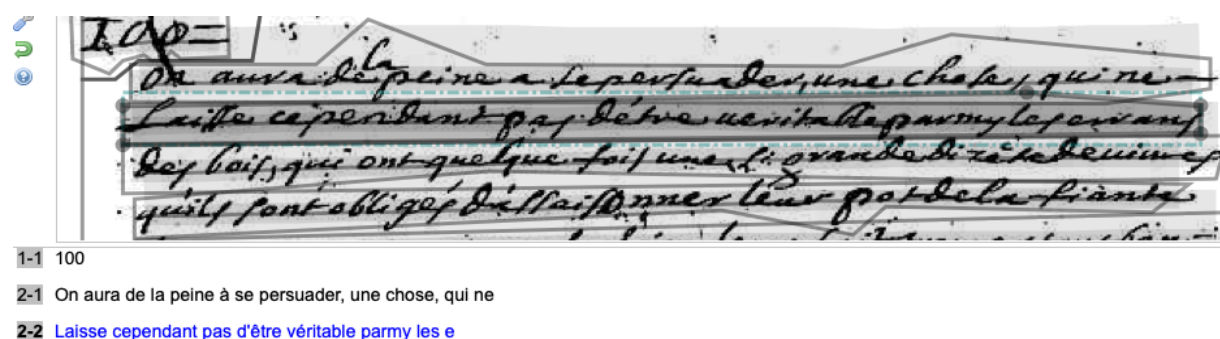
2-1 ou auà ée ne à se neiderune ctas, qu ne

2-2 Laie ce pe dant a dése uet e parny lereras

2-3 ey oiri que ont queque fisures cuxe se secou

- b) Dans le présent exemple, il est manifeste que le modèle de reconnaissance gagnerait à être peaufiné! Il faut donc user de ses compétences en paléographie pour déchiffrer le texte pour ce faire.

Nous vous invitons à consulter le « Protocole de transcription » disponible sur le site web de *Nouvelle-France numérique* pour connaître les normes les plus adéquates pour un usage optimal des transcriptions à des fins de recherche ou d'édition.



1-1 100

2-1 On aura de la peine à se persuader, une chose, qui ne

2-2 Laisse cependant pas d'être véritable parmi les e


## 9) Autres fonctions

- a) L'onglet « Metadata » permet d'appliquer des métadonnées à des passages du texte à l'aide de *balises* (« tags ») signalant des particularités de structure textuelle (« Structural ») ou du texte lui-même (« Textual »).

Ces balises permettent, par exemple, d'identifier les numéros de page, les abréviations, les mots incomplets, les lectures incertaines en raison d'un défaut du papier, d'écritures superposées ou de graphies difficiles à déchiffrer, les passages ajoutés en cours d'écriture ou lors d'une étape de révision ultérieure, les passages supprimés, les notes marginales, les mentions de lieux, de personnes, les dates, etc.

Cet onglet permet également d'ajouter des commentaires au texte.

Les usages des différentes balises et la façon de les appliquer sont décrites de façon détaillée dans le document *Protocole d'encodage TEI pour Transkribus*, disponible lui aussi sur le site *Nouvelle-France numérique*.

b)	Sachez que Transkribus effectue des sauvegardes incrémentielles de toutes les modifications apportées à un document. Ainsi, vos anciennes versions demeurent accessibles en tout temps sous l'onglet « Server », dans l'outil « Versions ».
c)	<p>Une fois le document transcrit, rechercher un mot ou une expression devient facile grâce à la loupe bleue (« Search current document ») située dans la barre d'outils principale :</p>  <p>* Des fonctions de recherches avancées sont accessibles grâce à l'outil « Search for documents, keywords, etc. » (🏠) situé lui aussi dans la barre d'outils principale.</p>
	<p>Pour mettre en pratique les informations contenues dans ce guide, il est possible de s'exercer sur le document intitulé « Test Document » généré automatiquement par Transkribus et qui est associé par défaut à votre collection originale.</p> <p>Mais le mieux est de vous lancer en téléversant un premier document.</p>

Pour des conseils et instructions complémentaires, voir le guide de Transkribus :

<https://transkribus.eu/Transkribus/docs/How%20to%20use%20TRANSKRIBUS-0.1.6.pdf>