

READ

Recognition and Enrichment
of Archival Documents



Comment utiliser Transkribus en 10 étapes (voire moins)

Ce document est une introduction de base à Transkribus. Il offre un flux de travail standard pour travailler avec la plate-forme¹.

Pour plus d'informations, consultez les documents et sites Web suivants :

- Comment transcrire des documents avec Transkribus - Mode Simple
- Comment transcrire des documents avec Transkribus - Mode Avancé
- Comment préparer des projets de test avec Transkribus - pour les Archives et Bibliothèques

Téléchargez Transkribus Expert Client ou assurez-vous d'utiliser la dernière version:

- <https://transkribus.eu/>

Consultez le Wiki Transkribus pour plus d'informations et un guide de l'utilisateur:

- <https://transkribus.eu/wiki/>
- https://transkribus.eu/wiki/index.php/Users_guide

Transkribus et la technologie qui l'accompagne sont mis à disposition via les projets et sites suivants:

- <https://read.transkribus.eu/>
- <https://transcriptorium.eu/>
- <https://github.com/transkribus/>

Contact

- L'équipe Transkribus: email@transkribus.eu

Le projet READ a reçu un financement du programme de recherche et d'innovation Horizon 2020 de l'Union européenne en vertu de la convention de subvention n ° 674943.

¹ Ce document est une traduction de « How to use Transkribus - in 10 steps (or less) » réalisée par Régis Schlagdenhauffen (EHESS – IRIS).

1. Introduction

- a. Transkribus peut être utilisé à plusieurs fins. Les plus importantes sont :
 - i. La création de données d'apprentissage pour alimenter le moteur de reconnaissance de texte manuscrit (HTR) afin qu'il puisse apprendre à déchiffrer une écriture.
 - ii. Exécuter HTR sur vos documents et obtenir des transcriptions générées automatiquement.
 - iii. Transcrire des documents pour une édition scientifique.
- b. Transkribus est proposé comme infrastructure de recherche par le **projet H2020 READ** (*Reconnaissance et enrichissement des documents d'archives*).
- c. Sachez qu'il faudra du temps pour explorer Transkribus et se familiariser avec son fonctionnement. Bien sûr, nous sommes heureux de vous aider avec ceci - n'hésitez pas à **nous contacter** par email (email@transkribus.eu) ou utilisez le bouton "Envoyer un rapport de bogue / demande de fonctionnalité" (icône de bogue verte) dans Transkribus.

2. Inscrivez-vous sur le site web

- a. <http://transkribus.eu/>
- b. Lisez notre **contrat d'utilisateur**:
- c. https://transkribus.eu/Transkribus/docs/TranskribusTermsOfUse_v04-2016.pdf/
- d. Tous les documents dans Transkribus sont "privés", ce qui signifie que personne d'autre que vous n'a accès à eux.
- e. Nous respectons votre vie privée et n'utilisons les données que pour améliorer nos services et soutenir la recherche en sciences humaines et en informatique !

3. Télécharger Transkribus depuis le site Web

- a. Notre **programme expert** vous permet de travailler avec vos documents d'une façon professionnelle.
- b. Téléchargez l'outil - il fonctionnera sous Windows, MacOS et Linux.
- c. Décompressez le fichier - vous ne pouvez pas démarrer l'outil à partir du fichier compressé !
- d. Démarrez l'outil et utilisez le bouton "Connexion" du menu principal pour accéder aux documents du serveur.
- e. Chaque utilisateur a accès à une «collection maison » composée de deux pages de matériel manuscrit et d'une page de texte imprimé. N'hésitez pas à expérimenter Transkribus avec ces pages!
- f. Vous pouvez télécharger notre **package d'exemple** pour tester le processus de téléchargement, de segmentation et de transcription de documents : https://transkribus.eu/wiki/images/d/d6/Example_Package.zip

4. Travailler avec ses propres documents

- a. Transkribus vous permet de travailler avec vos propres documents, localement ou de les télécharger sur nos serveurs.
- b. Remarque : Les processus automatisés ne peuvent être exécutés que si les documents sont téléchargés sur la plateforme Transkribus. La plate-forme peut traiter les fichiers PDF, JPEG, PNG, TIFF et JP2.
- c. Vous pouvez télécharger des documents que vous avez scannés vous-même.

- d. **Vous pouvez également télécharger des documents** (PDF) d'Internet et les télécharger sur Transkribus. De nombreuses bibliothèques et archives suivent des « politiques Open Access » et encouragent par conséquent l'utilisation de leurs collections - il suffit de demander directement aux archives et bibliothèques si vous pouvez télécharger des images de leurs documents vers Transkribus!

5. Téléchargez vos documents

- a. Il est simple de télécharger des images de vos documents sur Transkribus.
- b. Avec le téléchargement "http", **vous pouvez télécharger un total de 500 Mo en même temps**, donc vous devrez peut-être diviser vos documents en lots.
- c. Dans le menu principal, cliquez sur "Importer document (s)" (icône de dossier avec la flèche pointant vers la gauche) pour transférer les images de votre ordinateur vers la plate-forme. Remarque: les images doivent se trouver dans un dossier distinct sur votre ordinateur avant de les télécharger sur Transkribus !
- d. Lorsque vous téléchargez un document, vous devez créer votre propre collection en utilisant les options au bas de la case " Télécharger un document".
- e. Pour accéder à vos documents, cliquez sur votre collection dans le menu déroulant de l'onglet "Serveur", puis double-cliquez sur l'un des documents dans la case en dessous.
- f. **Remarque: tous les documents téléchargés sur Transkribus sont privés par défaut.** Vous pouvez ensuite donner aux autres utilisateurs l'autorisation de voir vos documents si vous le souhaitez. Utilisez le bouton "Gérer les collections ..." dans l'onglet "Serveur" pour ajouter des utilisateurs (qui doivent être enregistrés avec Transkribus) à votre collection.

6. Créer des données d'entraînement pour le moteur HTR

- a. Afin de pouvoir alimenter le moteur HTR avec des données d'apprentissage, les documents doivent être segmentés.
- b. Transkribus vous demande donc de faire **un lien direct** entre les images de votre document et le texte transcrit correspondant.
- c. Ceci est réalisé en divisant chaque page en différentes sections qui sont connues sous le nom de régions de texte, lignes et lignes de base. Le processus de division d'une page de cette manière est appelé segmentation. Ne vous inquiétez pas, cela peut être fait semi-automatiquement et ne devrait prendre que quelques minutes.
- d. **La meilleure pratique consiste à définir manuellement des régions de texte sur plusieurs pages, puis à détecter automatiquement les lignes et les lignes de base à l'aide du bouton "Détecter les lignes et les lignes de base" dans l'onglet "Outils".**
- e. Cliquez sur le bouton "Visualiser les profils" dans le menu principal et sélectionnez vue "Segmentation" Cela vous permettra de voir plus facilement les régions de texte et les lignes de base.
- f. Pour définir une zone de texte, cliquez sur le bouton "+ TR" dans le menu Canevas. Cliquez une fois sur le coin supérieur gauche d'une section de texte, puis cliquez à nouveau dans le coin inférieur droit. Remarque: il n'y a pas de problème si les régions de texte se chevauchent!
- g. Pour détecter automatiquement les régions de texte et/ou les lignes et lignes de base, allez dans l'onglet "Outils" et cliquez sur "Détecter les lignes et les lignes de

base". Vous pouvez traiter plusieurs pages en même temps en cliquant sur le bouton "Traitement par lots ..." dans le même onglet.

- h. **Les lignes de base sont les plus importantes dans Transkribus** et peuvent donc avoir besoin d'être corrigées. Vous pouvez corriger les lignes de base en déplaçant ou en faisant glisser (des parties de) la ligne.

7. Commencez votre transcription

- a. Une fois les lignes de base visibles sur votre image, vous pouvez écrire du texte dans le champ de l'éditeur de texte.
- b. Cliquez sur le bouton "Visualiser les profils" dans le menu principal et sélectionnez Vue "Transcription" Cela facilitera la lecture et la transcription du texte dans votre document.
- c. **Pour chaque ligne de base, il y aura une ligne correspondante dans l'éditeur de texte.** Cela devrait faciliter la transcription de la page. Vous devez transcrire le texte ligne par ligne, exactement comme il apparaît dans l'image.
- d. Les caractères spéciaux peuvent être trouvés dans le bouton "Claviers virtuels" dans l'éditeur de texte **barre d'outils**.

8. Sauvegardez et exportez votre transcription

- a. Appuyez sur le bouton "Enregistrer" dans le menu principal pour enregistrer le document dans Transkribus. Si vous cliquez sur le bouton "Versions de la page ..." dans l'onglet "Serveur", vous verrez qu'une **nouvelle version est générée**. Cela signifie que vous pouvez toujours accéder aux versions précédentes d'un document si vous en avez besoin.
- b. Vous pouvez également exporter le document entier à tout moment du processus en cliquant sur "Exporter le document" (icône du dossier avec la flèche pointant vers la droite) dans le menu principal. Vous pouvez exporter des documents dans plusieurs formats différents: le package complet comprenant les fichiers image d'origine, les fichiers XML et les métadonnées ajoutés au document, un PDF (avec le texte transcrit inclus), doc et TEI (*Text Encoding Initiative*). Tous les fichiers sont stockés dans un dossier.

9. Utilisez la reconnaissance automatique de texte (HTR) sur vos documents

- a. Il est simple d'avoir vos documents reconnus par l'ordinateur. **Vous devrez créer des données de formation pour environ 20 000 mots (100 pages)** - après l'ordinateur peut être formé. Le processus de formation doit être démarré par l'équipe de Transkribus - envoyez-nous simplement un e-mail une fois que vous aurez segmenté et transcrit un premier lot de pages (email@transkribus.eu). Nous formerons le moteur HTR et vous indiquerons quand il sera prêt à traiter vos documents.
- b. Une fois qu'un modèle HTR a été formé pour vos documents, il peut être appliqué via le module "Exécuter la reconnaissance de texte" dans l'onglet "Outils". Vous pouvez sélectionner une ou plusieurs pages de vos documents et démarrer le traitement HTR.
- c. Les résultats apparaîtront directement sur une nouvelle version de votre document dans Transkribus. Il est également possible d'évaluer la précision de la transcription automatique en utilisant le bouton "Comparer" dans l'onglet "Outils".

- d. Pour voir un exemple de transcription automatique générée par HTR, jetez un coup d'œil à la page 2 de votre "collection maison".

10. Utilisations avancées

- a. Marge structurelle : Vous pouvez marquer différentes parties de votre document, telles que des notes de bas de page, marges, texte souligné etc. Vous pouvez le faire en cliquant sur l'onglet "Métadonnées" puis sur l'onglet "Structurel".
- b. Marquage : Vous pouvez également ajouter des balises plus spécifiques à transcription telles que « personne », « date », « lieu », etc. Utilisez l'onglet « Métadonnées », puis l'onglet « Étiquetage » (*Tag*) pour ce faire. Vous pouvez appliquer des tags existants ou créer les vôtres.
- c. Autres Services : Vous pouvez accéder à Transkribus via des services Web. Dans ce cas, utilisez l'interface REST: <https://transkribus.eu/TrpServer/Swadi/wadl.html>

Captures d'écran

Site Internet

Transkribus [Register](#) [Login](#)

Transcribe. Collaborate. Share...
...and benefit from cutting edge research in Handwritten Text Recognition!

[Download Transkribus beta 0.10.1 »](#) [Wiki » How-to guide \(pdf\) »](#)

Scholars

Are you transcribing historical documents? Handwritten or printed, from the middle ages or from the 20th century? Would you like to do this in a highly standardized, flexible and reliable way? And do you appreciate to get support from automated tools such as Handwritten Text Recognition and Layout Analysis?

[View details »](#)

Archives

Are you responsible for large collections of handwritten and printed documents? Do you believe that digitalisation paves the way to realise new opportunities to access, enrich and explore archival material? And are you open to involve humanities scholars and volunteers so that they can work with these documents in an effective way – producing data which can also be integrated in your repository?

[View details »](#)

Volunteers

Are historical letters, postcards, manuscripts or medieval documents fascinating for you? Do you enjoy deciphering handwriting – this wonderful feeling when you can read something which may be hidden to most other people? And do you believe that everyone can make a valuable contribution to scholarship and science?

[View details »](#)

Scientists

Are you a computer scientist and working in the fields of computer vision, document analysis, pattern recognition, natural language processing or a related field? You are seeking interesting documents from 1000 years of handwriting, printing and publishing? And you would really enjoy to get some reference data in a well-acknowledged format, such as PAGE?

[View details »](#)

Enregistrement

Transkribus [Register](#) [Login](#)

Transcribe. Collaborate. Share...
...and benefit from cutting edge research in Handwritten Text Recognition!

[Download Transkribus beta 0.10.1 »](#) [Wiki » How-to guide \(pdf\) »](#)

Scholars

Are you transcribing historical documents? Handwritten or printed, from the middle ages or from the 20th century? Would you like to do this in a highly standardized, flexible and reliable way? And do you appreciate to get support from automated tools such as Handwritten Text Recognition and Layout Analysis?

[View details »](#)

Archives

Are you responsible for large collections of handwritten and printed documents? Do you believe that digitalisation paves the way to realise new opportunities to access, enrich and explore archival material? And are you open to involve humanities scholars and volunteers so that they can work with these documents in an effective way – producing data which can also be integrated in your repository?

[View details »](#)

Volunteers

Are historical letters, postcards, manuscripts or medieval documents fascinating for you? Do you enjoy deciphering handwriting – this wonderful feeling when you can read something which may be hidden to most other people? And do you believe that everyone can make a valuable contribution to scholarship and science?

[View details »](#)

Scientists

Are you a computer scientist and working in the fields of computer vision, document analysis, pattern recognition, natural language processing or a related field? You are seeking interesting documents from 1000 years of handwriting, printing and publishing? And you would really enjoy to get some reference data in a well-acknowledged format, such as PAGE?

[View details »](#)

Télécharger

Transkribus [Modify Account](#) [Logout](#)

Transcribe. Collaborate. Share...
...and benefit from cutting edge research in Handwritten Text Recognition!

[Download Transkribus beta 0.10.1 »](#) [Wiki » How-to guide \(pdf\) »](#)

Scholars

Are you transcribing historical documents? Handwritten or printed, from the middle ages or from the 20th century? Would you like to do this in a highly standardized, flexible and reliable way? And do you appreciate to get support from automated tools such as Handwritten Text Recognition and Layout Analysis?

[View details »](#)

Archives

Are you responsible for large collections of handwritten and printed documents? Do you believe that digitalisation paves the way to realise new opportunities to access, enrich and explore archival material? And are you open to involve humanities scholars and volunteers so that they can work with these documents in an effective way – producing data which can also be integrated in your repository?

[View details »](#)

Volunteers

Are historical letters, postcards, manuscripts or medieval documents fascinating for you? Do you enjoy deciphering handwriting – this wonderful feeling when you can read something which may be hidden to most other people? And do you believe that everyone can make a valuable contribution to scholarship and science?

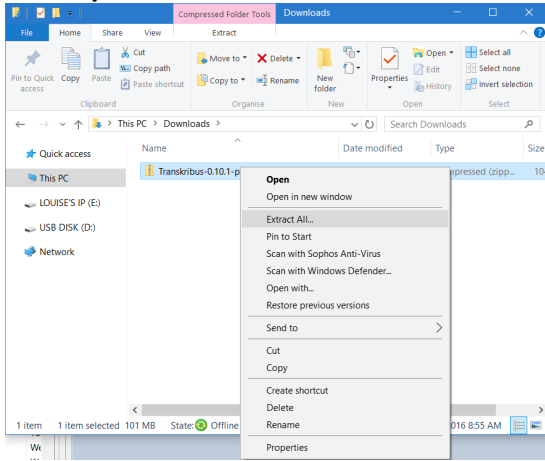
[View details »](#)

Scientists

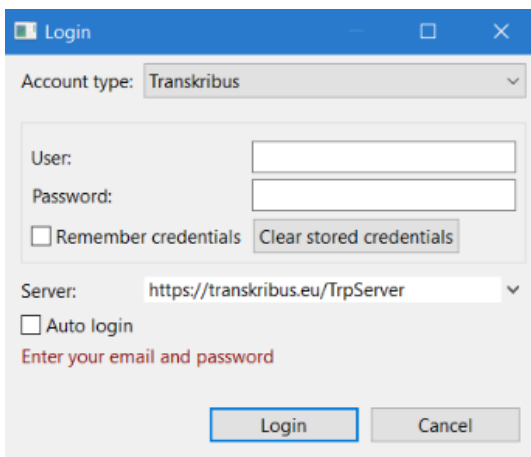
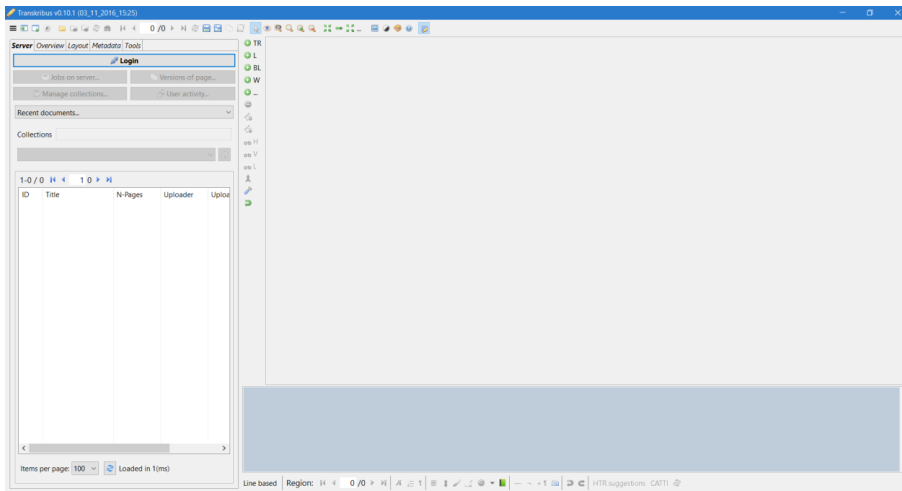
Are you a computer scientist and working in the fields of computer vision, document analysis, pattern recognition, natural language processing or a related field? You are seeking interesting documents from 1000 years of handwriting, printing and publishing? And you would really enjoy to get some reference data in a well-acknowledged format, such as PAGE?

[View details »](#)

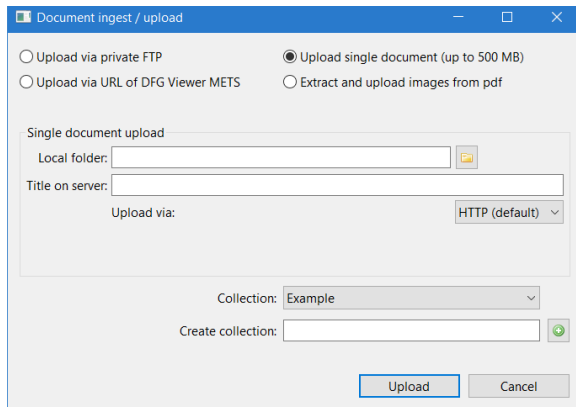
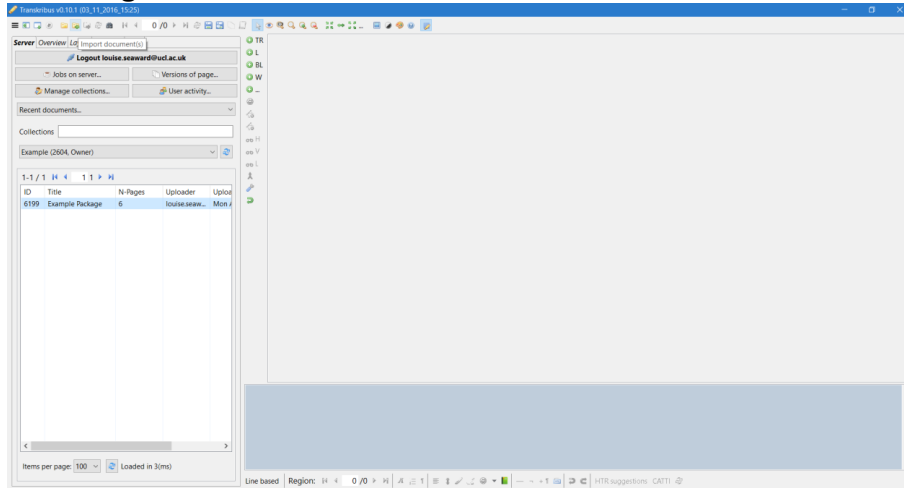
Décompresser



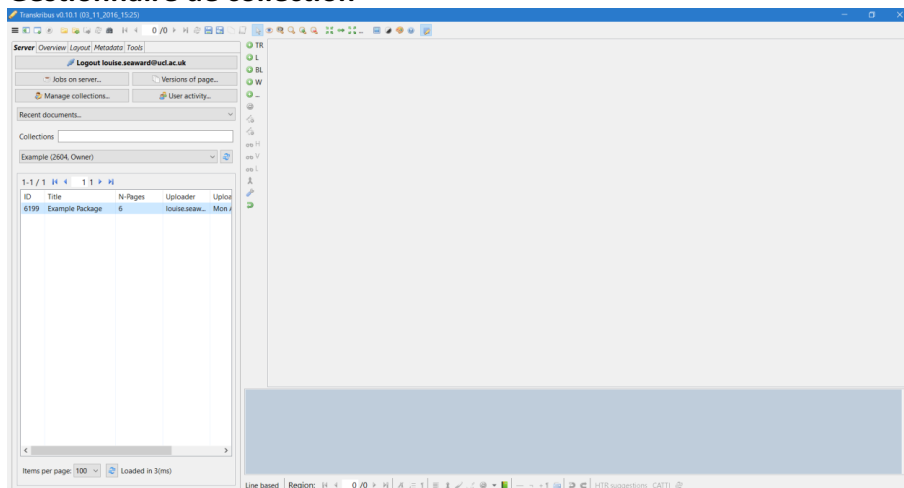
S'identifier



Télécharger un document



Gestionnaire de collection



The screenshot shows the 'Collection Manager' interface. It is divided into three main sections:

- Collections:** A table listing collections with columns for ID, Name, Role, and Description. The first row is '2604 Example Owner created by...'. Below the table are buttons for 'Create collection...', 'Delete collection', and 'Modify collection...'.
- Documents:** A table listing documents with columns for ID, Title, N-Pages, Uploader, Uploaded, and Collections. The first row is '6199 Example Package 6 louise.seaw... Mon Aug 0...'. Below the table are buttons for 'Add to collection...', 'Remove from collection', 'Delete document', and 'Duplicate document'.
- Users in collection:** A table listing users with columns for Username, Firstname, Lastname, and Role. The first row is 'louise... Louise Seaward Owner'. Below the table are buttons for 'Add user', 'Remove user', and 'Change Role: Transcriber'.

Document de segment - sélectionnez "Vue de segmentation"

The screenshot shows the 'Transcribe' interface. The left sidebar contains a 'Server Overview' menu with 'Segmentation' selected. The main area displays a document with handwritten text. The text is organized into sections: 'Segmentation', 'Transcription', and 'A meaningful line'. The 'Segmentation' section contains the following text:

This is an example page for
How to transcribe tab in Transcriber

Segmentation
It is really simple to transcribe
text and to generate training data
for the HTK engine.

First of all you will need to
segment the image. It's up
to you how many tab regions
you create.

Use the 'line/baseline segmenta-
tion' tool to create lines/baselines
automatically.

Transcription
When ~~transcribing~~ transcribing
keep to the alphabet of the
source document.

A meaningful line ~~is~~ bold, underlined, starts
with a tab char, and ends
with a tab char.

Delimiters are transcribed if
possible, otherwise use the
next 'tag'.

Baseline additions can be added
in the usual way as separate
lines.

Do not record other - just Es.
mark your use.

Ajouter une région de texte

The screenshot shows the 'Transcribe' interface. The left sidebar contains a 'Server Overview' menu with 'Logout louise.seaward@ucl.ac.uk' selected. The main area displays the same document as the previous screenshot. The text is organized into sections: 'Segmentation', 'Transcription', and 'A meaningful line'. The 'Segmentation' section contains the following text:

This is an example page for
How to transcribe tab in Transcriber

Segmentation
It is really simple to transcribe
text and to generate training data
for the HTK engine.

First of all you will need to
segment the image. It's up
to you how many tab regions
you create.

Use the 'line/baseline segmenta-
tion' tool to create lines/baselines
automatically.

Transcription
When ~~transcribing~~ transcribing
keep to the alphabet of the
source document.

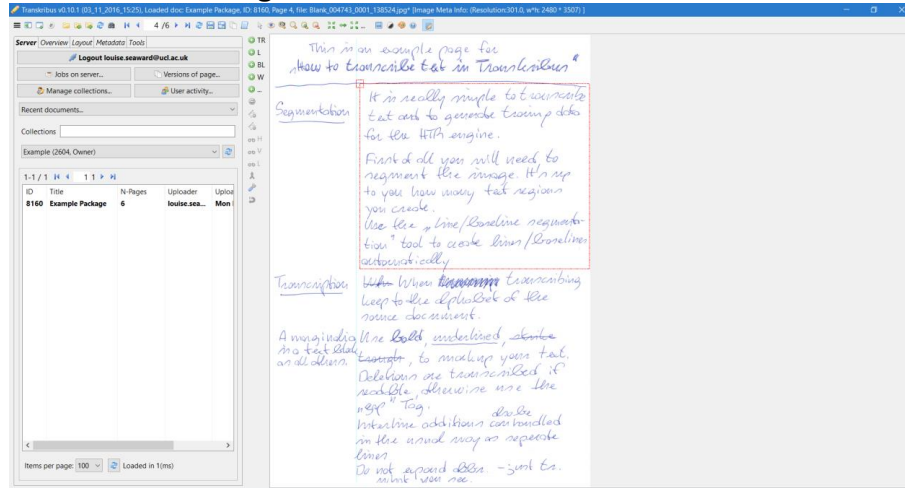
A meaningful line ~~is~~ bold, underlined, starts
with a tab char, and ends
with a tab char.

Delimiters are transcribed if
possible, otherwise use the
next 'tag'.

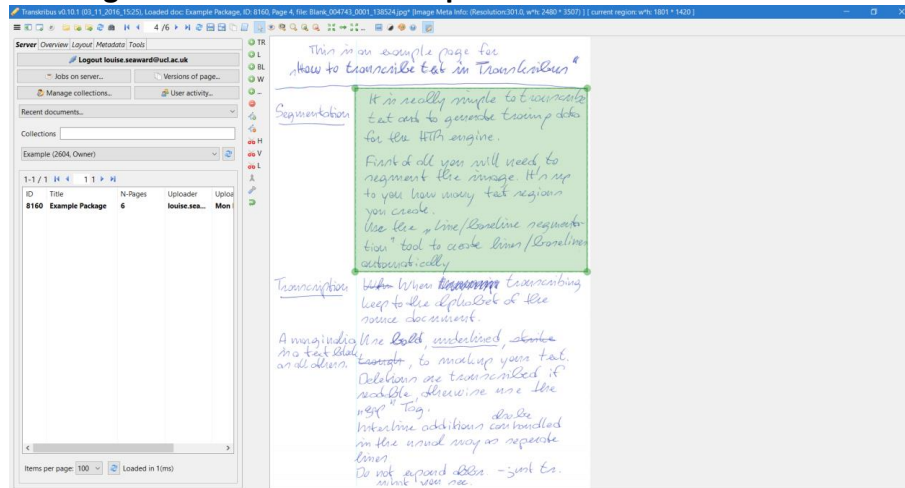
Baseline additions can be added
in the usual way as separate
lines.

Do not record other - just Es.
mark your use.

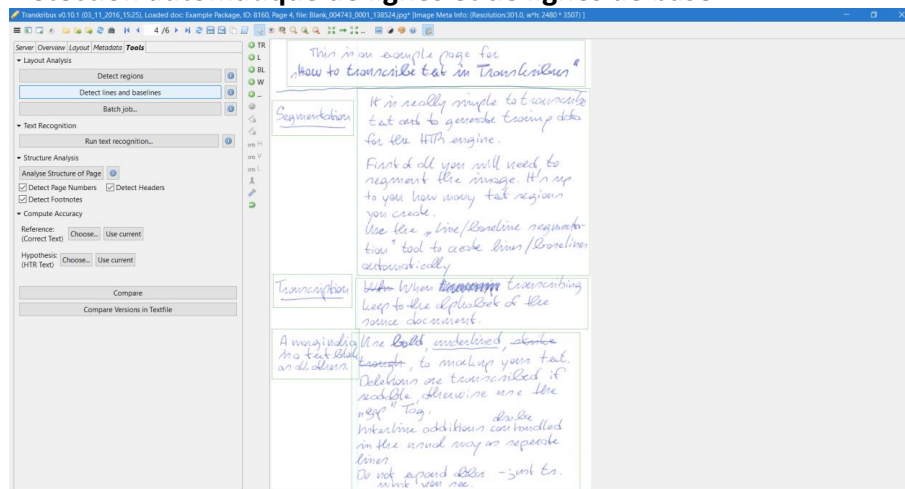
Dessiner un rectangle

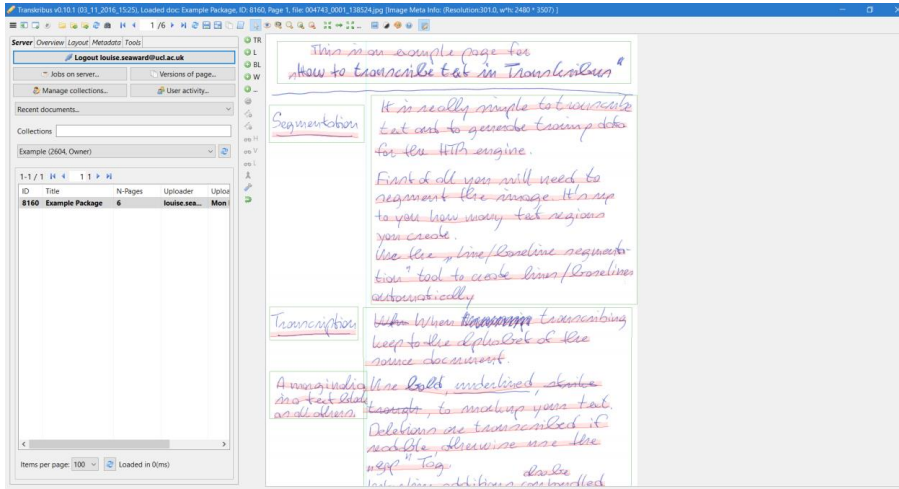


Affichage de la zone de texte - représenté ici en vert

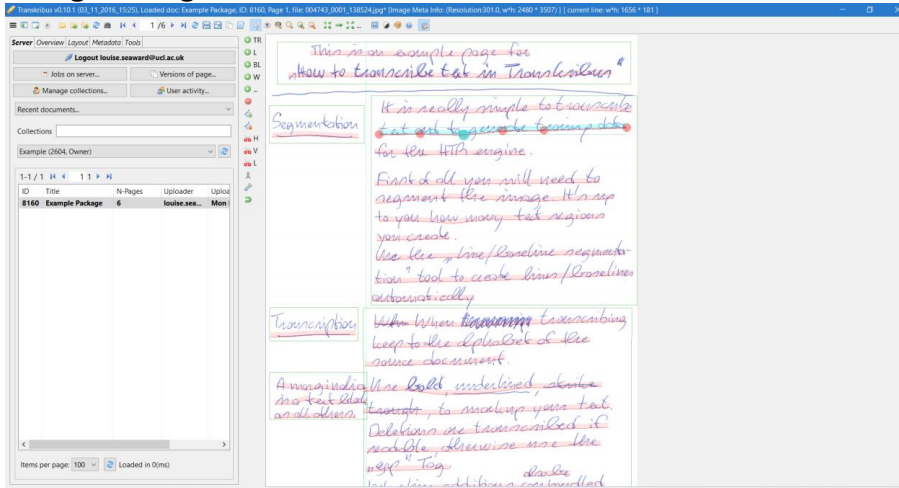


Détection automatique de lignes et de lignes de base

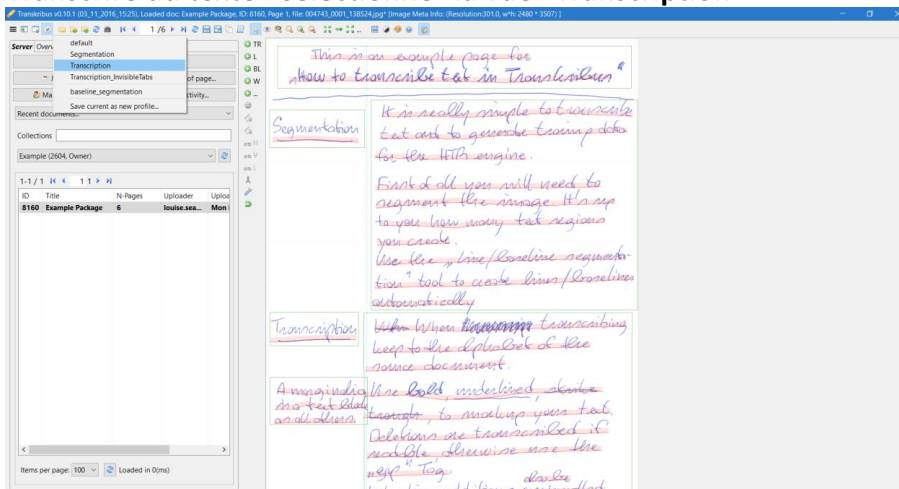




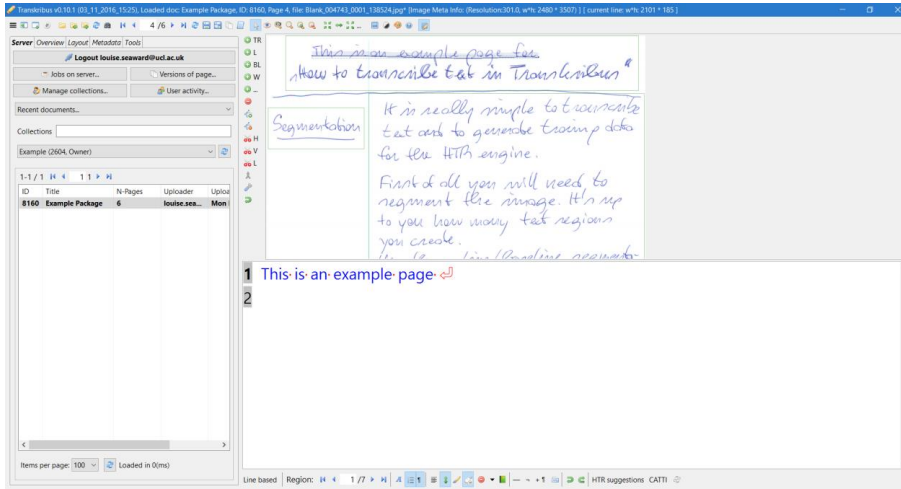
Corriger les lignes de base



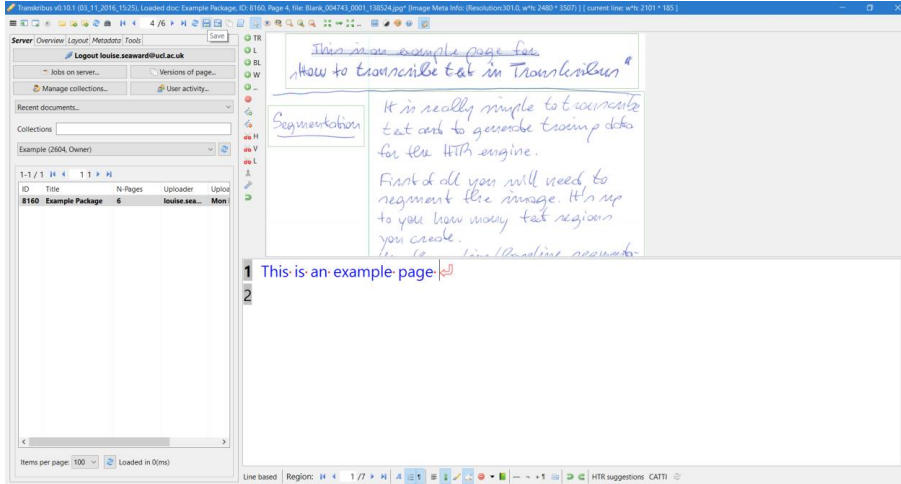
Transcrire du texte - sélectionnez la vue "Transcription"



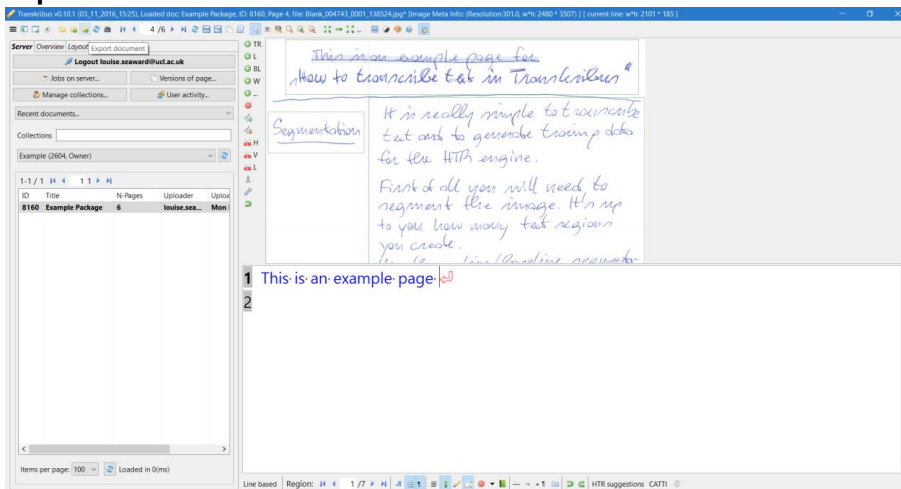
Transcrire du texte



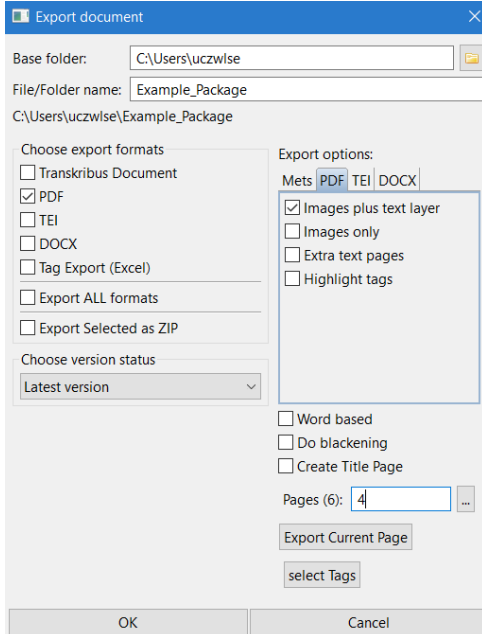
Sauvegardez votre transcription



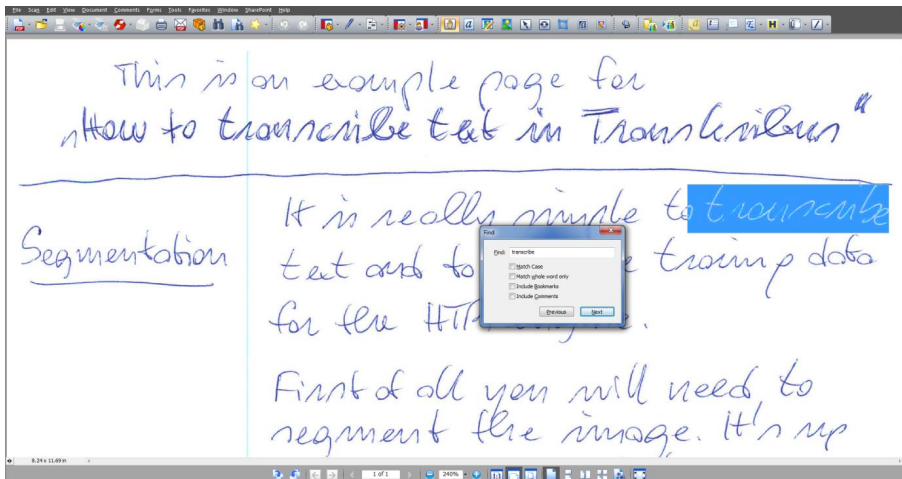
Exportez votre document



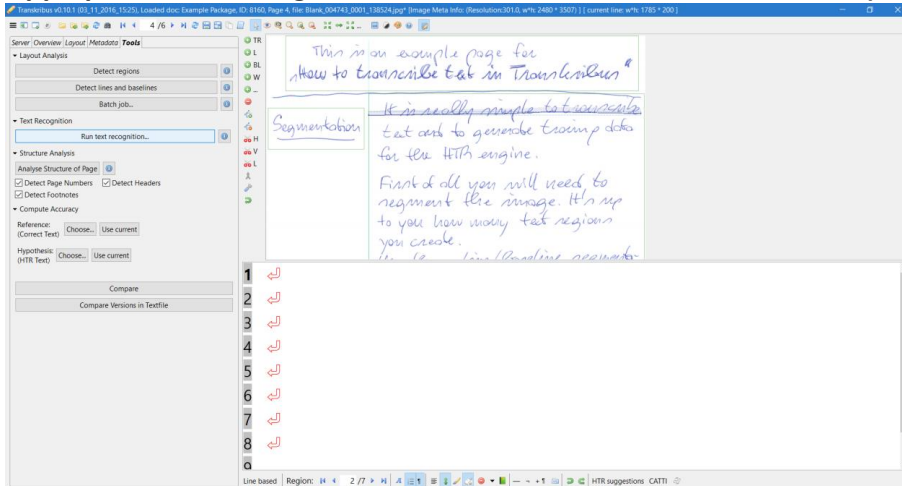
Exporter dans différents formats incluant TEI, RTF et PDF (montré ici)

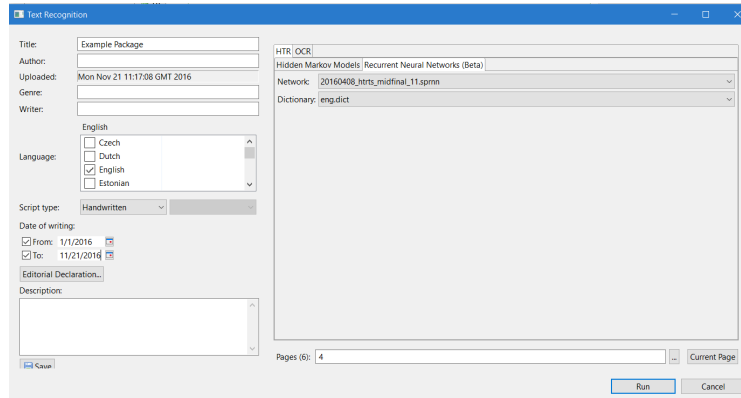


Rechercher votre texte transcrit dans le PDF

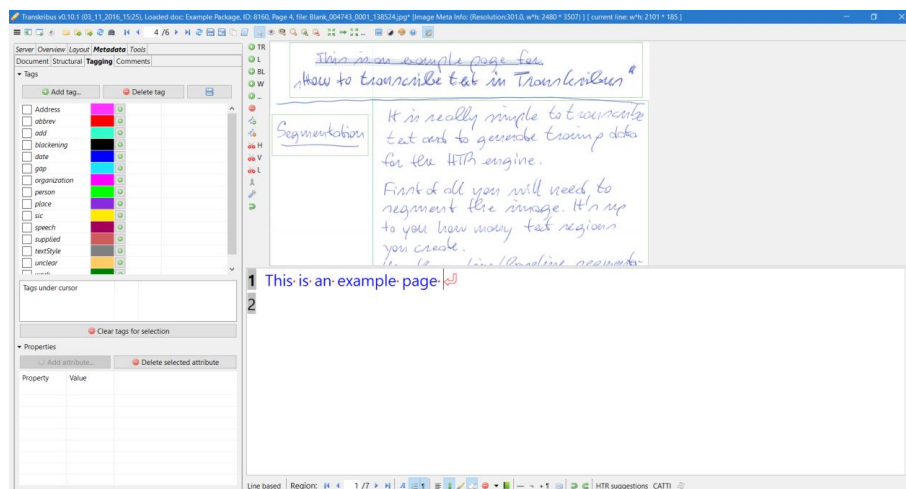
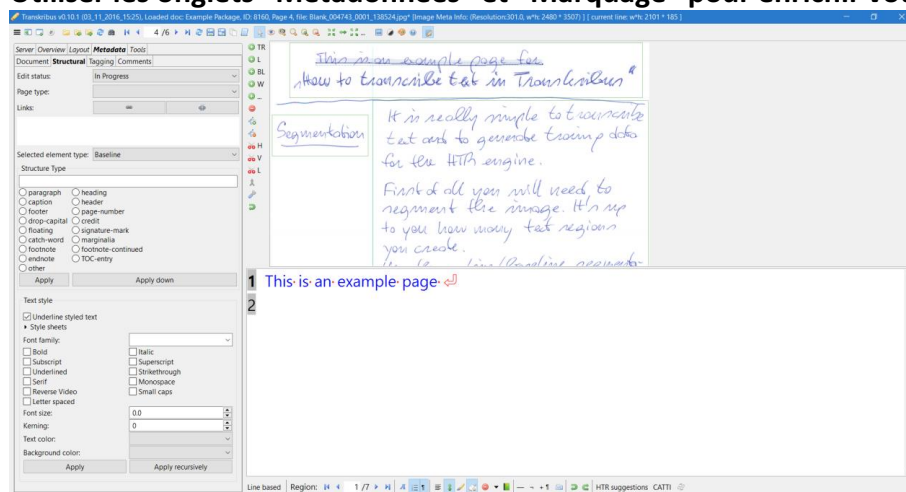


Appliquer la technologie de reconnaissance de texte manuscrite (HTR) à vos documents





Utiliser les onglets "Métadonnées" et "Marquage" pour enrichir votre transcription



Pour plus d'informations, veuillez consulter le wiki de **Transkribus** ou contactez l'équipe de Transkribus :

- https://transkribus.eu/wiki/index.php/Main_Page
- email@transkribus.eu