



HAL
open science

Tutoriel Transkribus

Emmanuelle Perrin

► **To cite this version:**

| Emmanuelle Perrin. Tutoriel Transkribus. 2019. hal-02472234

HAL Id: hal-02472234

<https://hal.science/hal-02472234>

Preprint submitted on 10 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



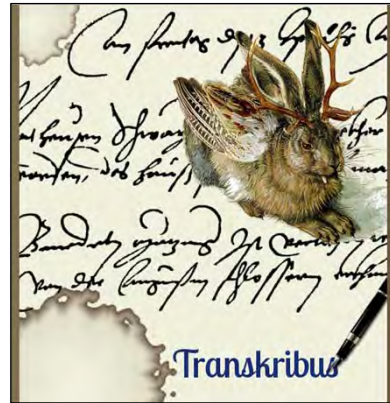
Distributed under a Creative Commons Attribution 4.0 International License

Bulliot, Bibracte et moi

Transcription collaborative des archives archéologiques du site de Bibracte

Tutoriel Transkribus

version 2



Avec le soutien financier du ministère de la Culture



| un projet porté par



| en partenariat avec



Transkribus est une plateforme de transcription automatique de textes manuscrits, développée depuis 2015, à l'université d'Innsbruck (Autriche), dans le cadre du projet européen [READ](#) (Recognition and Enrichment of Archival Documents), financé par le programme de recherche européen H2020. Fondé sur l'intelligence artificielle, le **moteur de reconnaissance de texte manuscrit** (*Handwritten Text Recognition* ou HTR) doit être préparé avec des données d'apprentissage, obtenues par la transcription d'une centaine de pages minimum, en établissant la correspondance ligne à ligne entre l'image du texte numérisé et sa transcription.

Pour télécharger Transkribus et s'enregistrer comme utilisateur :

<https://transkribus.eu/Transkribus/>

Tutoriels, wiki et guide :

[Comment utiliser Transkribus en 10 étapes \(voire moins\)](#)

[How To Transcribe Documents with Transkribus — Advanced Mode](#)

[Transkribus Wiki](#)

[Guide de l'utilisateur](#)

Créer un compte

1) <https://transkribus.eu/Transkribus/>

The screenshot shows the Transkribus website with a 'Create account' modal form open. The form contains the following fields: Email*, Password*, Repeat password*, Given name*, Family name*, ORCID, and Gender (with a dropdown menu set to 'Other'). Below the fields are checkboxes for 'I accept the terms of use*' and 'Je ne suis pas un robot' (with a reCAPTCHA logo). A 'Sign up' button is at the bottom of the form. The background shows the website header with 'Transkribus' and 'Register' and 'Login' buttons. The page content includes sections for 'Scholars' and 'Humanities Scholars'.

2) Cliquer sur « Register »

3) Remplir le formulaire

4) Valider

5) Vous recevez ensuite un e-mail « Transkribus registration » avec un lien pour activer votre compte

S'identifier

The image shows a browser window at <https://transkribus.eu/Transkribus/>. The page features a dark header with the 'Transkribus' logo and 'Register' and 'Login' buttons. A central 'Login' modal is open, containing an 'Email' input field, a 'Password' input field, a green 'Login' button, a red 'Login with Google' button, and a link for 'Or sign up. Forgot password?'. Three red arrows point to the 'Login' button in the header, the 'Login' button in the modal, and the 'Login' button in the modal. Three red dashed boxes with white text provide instructions: '1) Cliquer sur Login' (pointing to the header button), '2) Remplir le formulaire' (pointing to the modal form), and '3) Valider' (pointing to the modal 'Login' button). Below the modal, the page content is divided into four columns: 'Scholars', 'Archives', 'Volunteers', and 'Scientists', each with a brief description of their role.

1) Cliquer sur Login

2) Remplir le formulaire

3) Valider

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

Archives
Are you responsible for large collections of handwritten and printed documents? Do you believe that digitisation paves the way to realise new opportunities to access, enrich and explore archival material? And are you open to involve humanifies scholars and volunteers so that they can work with these documents

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

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

Télécharger

Transkribus

Il faut être identifié pour télécharger Transkribus

Register

Login

Transcribe. Collaborate. Share...

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

Download version 1.9

Download version 1.9 for Mac

Wiki » How-to guide (pdf) »

Téléchargement

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 digitisation 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

research?

Volunteers

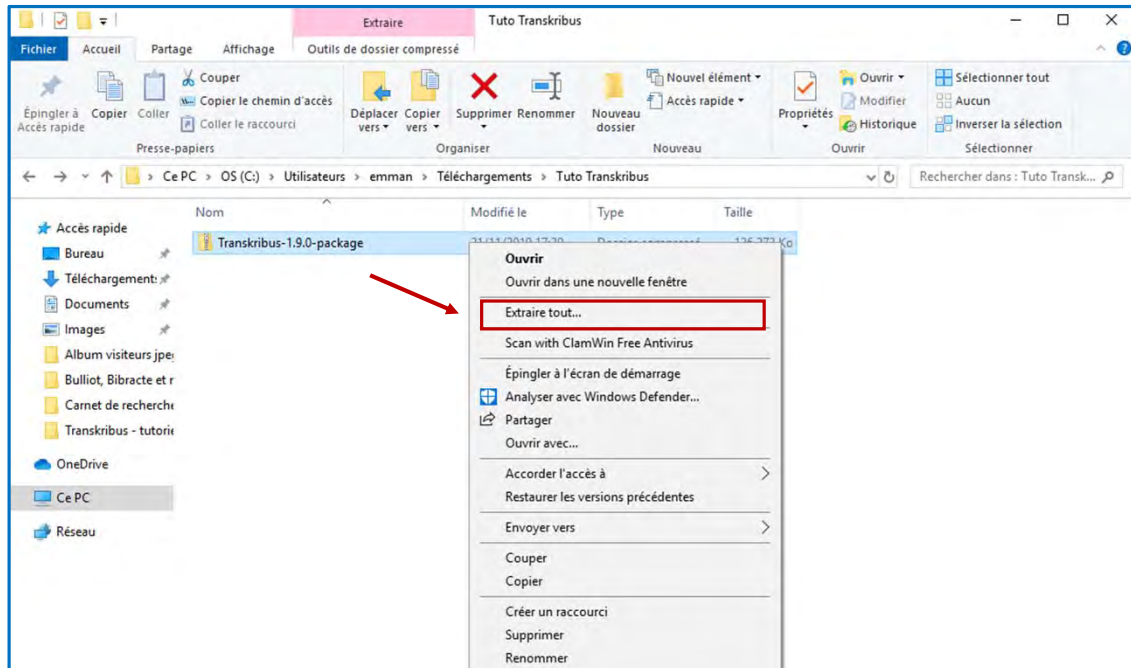
Are historical letters, postcards, manuscripts or medieval documents fascinating for you? Do you enjoy deciphering handwriting – this wonderful feeling when you can read something which may be hidden to most other people? And do you believe that everyone can make a valuable contribution to scholarship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

[View details »](#)

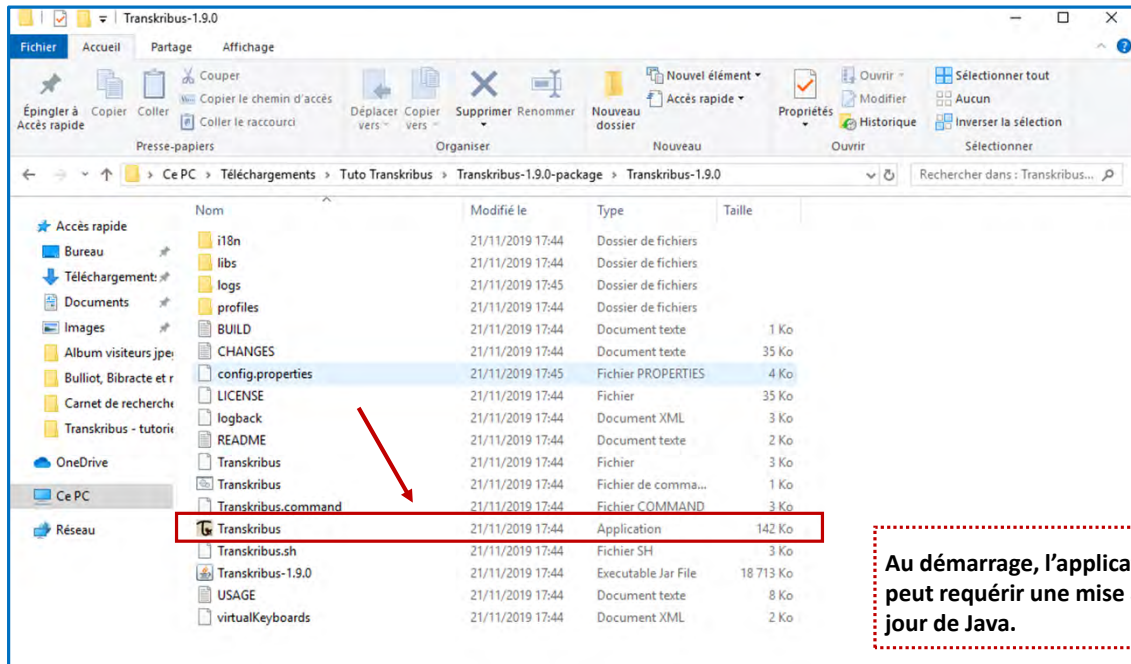
Scientists

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?

Décompresser



Démarrer Transkribus

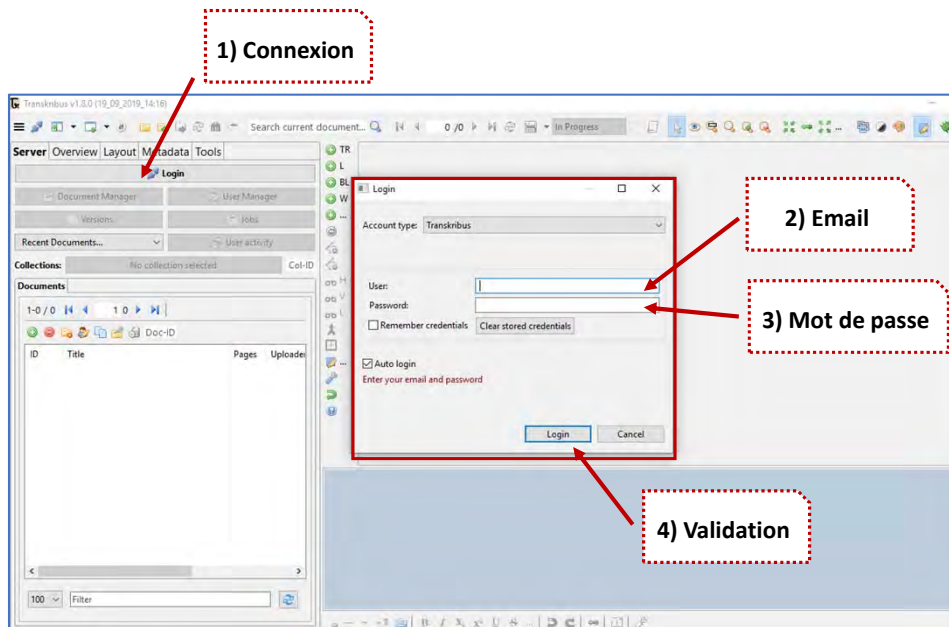


The screenshot shows a Windows File Explorer window titled 'Transkribus-1.9.0'. The address bar indicates the path: 'Ce PC > Téléchargements > Tuto Transkribus > Transkribus-1.9.0-package > Transkribus-1.9.0'. The main pane displays a list of files and folders with columns for 'Nom', 'Modifié le', 'Type', and 'Taille'. A red box highlights the 'Transkribus' application file (142 Ko). A red arrow points from the 'config.properties' file to the 'Transkribus' file. A text box on the right contains the note: 'Au démarrage, l'application peut requérir une mise à jour de Java.'

Nom	Modifié le	Type	Taille
i18n	21/11/2019 17:44	Dossier de fichiers	
libs	21/11/2019 17:44	Dossier de fichiers	
logs	21/11/2019 17:45	Dossier de fichiers	
profiles	21/11/2019 17:44	Dossier de fichiers	
BUILD	21/11/2019 17:44	Document texte	1 Ko
CHANGES	21/11/2019 17:44	Document texte	35 Ko
config.properties	21/11/2019 17:45	Fichier PROPERTIES	4 Ko
LICENSE	21/11/2019 17:44	Fichier	35 Ko
logback	21/11/2019 17:44	Document XML	3 Ko
README	21/11/2019 17:44	Document texte	2 Ko
Transkribus	21/11/2019 17:44	Fichier	3 Ko
Transkribus	21/11/2019 17:44	Fichier de comman...	1 Ko
Transkribus.command	21/11/2019 17:44	Fichier COMMAND	3 Ko
Transkribus	21/11/2019 17:44	Application	142 Ko
Transkribus.sh	21/11/2019 17:44	Fichier SH	3 Ko
Transkribus-1.9.0	21/11/2019 17:44	Executable Jar File	18 713 Ko
USAGE	21/11/2019 17:44	Document texte	8 Ko
virtualKeyboards	21/11/2019 17:44	Document XML	2 Ko

Au démarrage, l'application peut requérir une mise à jour de Java.

Se connecter pour accéder aux documents à transcrire



Sélectionner une collection

1) Cliquer dans le champ « Collections »

2) Dans la fenêtre qui apparaît, double-cliquer sur la collection à votre nom

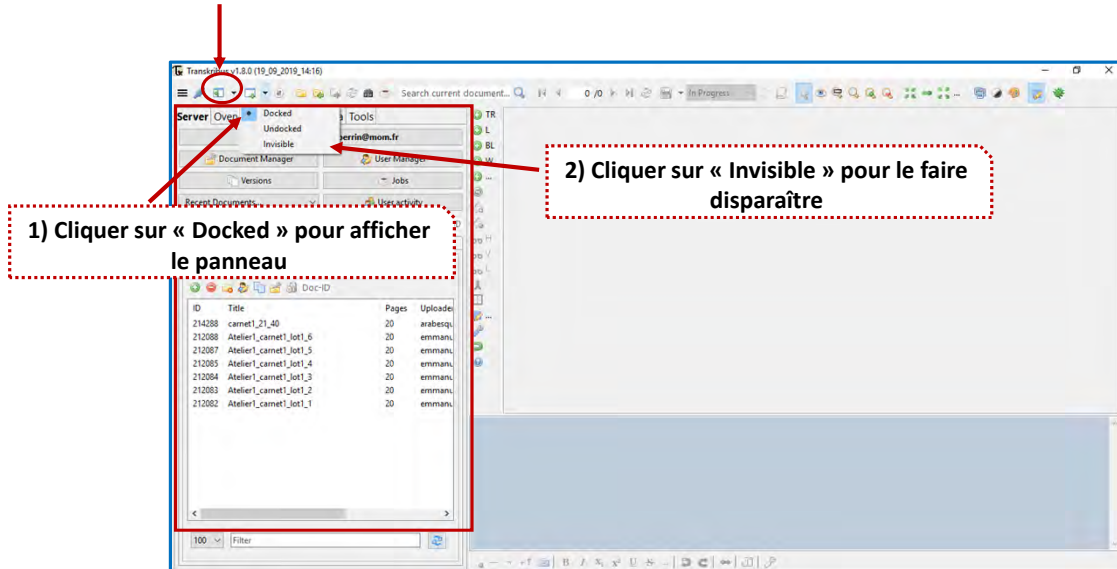
3) Double-cliquer sur le lot à transcrire

The screenshot shows a document management system interface. The main window displays a list of documents under the heading 'Documents'. A secondary window titled 'Choose a collection via double click' is open, showing a list of collections. Red arrows point from the text instructions to the corresponding UI elements.

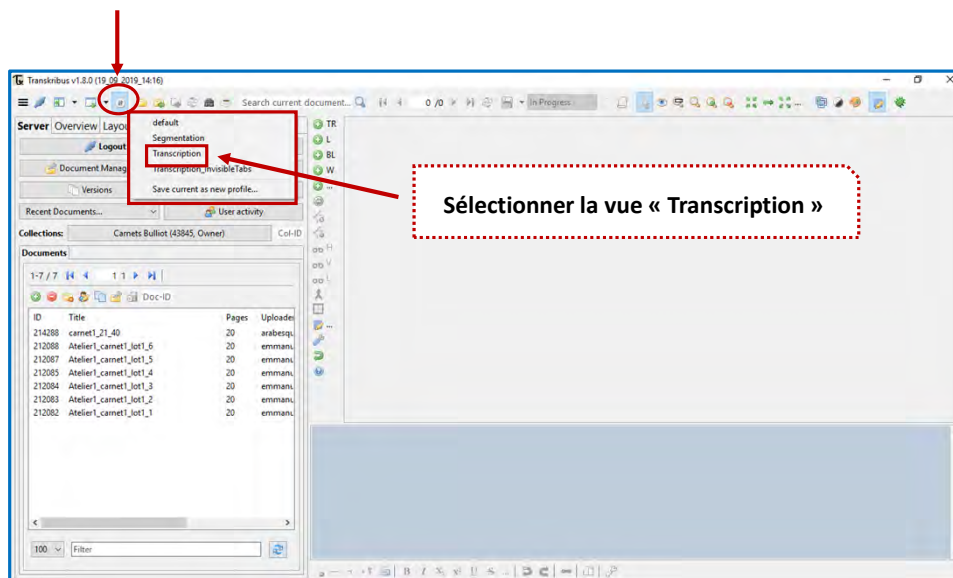
ID	Name	Role	Description
13052	Correspondance Rey	Owner	created by emmanuelle.perrin.touche@univ-st-etienne.fr
43845	Carnets Bulletin	Owner	created by emmanuelle.perrin.touche@univ-st-etienne.fr
49122	emmanuelle.perrin@mom.fr Collection	Owner	emmanuelle.perrin@mom.fr

ID	Title	Pages	Uploader
214288	carnet1_21_40	20	arabesqu
212088	Atelier1_carnet1_lot1_6	20	emmanu
212087	Atelier1_carnet1_lot1_5	20	emmanu
212085	Atelier1_carnet1_lot1_4	20	emmanu
212084	Atelier1_carnet1_lot1_3	20	emmanu
212083	Atelier1_carnet1_lot1_2	20	emmanu
212082	Atelier1_carnet1_lot1_1	20	emmanu

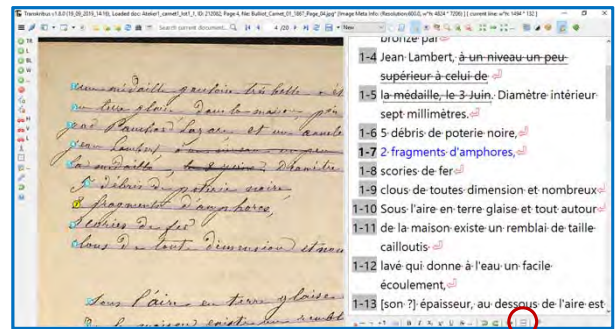
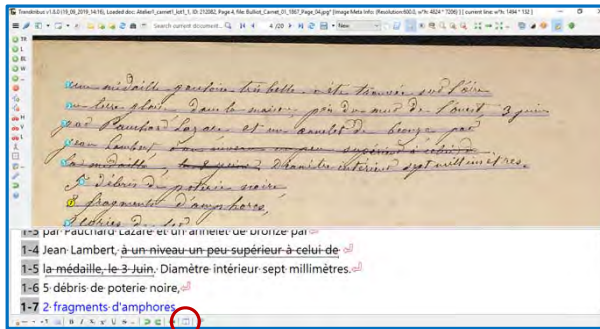
Affichage du panneau de gauche



Vue Transcription

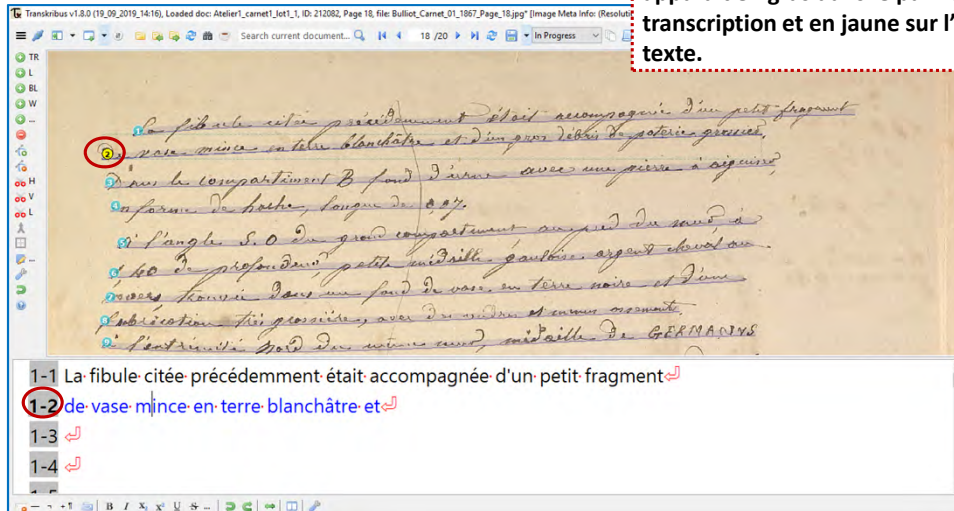


Diviser l'écran en blocs horizontaux ou verticaux



Transcrire

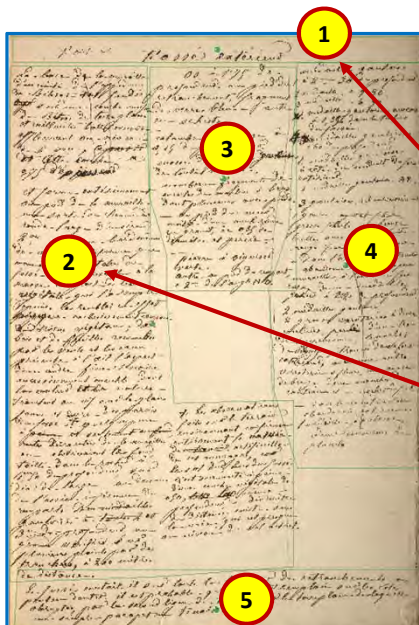
La transcription s'effectue ligne à ligne. Le numéro de la ligne sélectionnée apparaît en gras dans le panneau de transcription et en jaune sur l'image du texte.



The screenshot displays the Transkribus v1.8.0 interface. The main window shows a handwritten document with several lines of text. A red circle highlights the first line of the document. Below the document, a transcription panel is visible, showing a list of lines with their corresponding transcription. The first line is highlighted in yellow, and the number '1-2' is bolded in the list.

1-1 La fibule citée précédemment était accompagnée d'un petit fragment
1-2 de vase mince en terre blanchâtre et d'un gros débris de poterie grossière,
1-3 dans le compartiment B fond d'ivoire avec une pierre à siffler,
1-4 en forme de balle, longue de 0,5 cm.
Si l'angle S.O. du grand compartiment au sud du sud, à
60° de profondeur, petite médaille d'argent devant au
dessus trouvée dans un fond de vase en terre noire et d'une
fabrication très grossière, avec des motifs et enroulements
de l'extrémité sud. Des mêmes années, médaille de GERMANUS

Les zones de texte (text regions)



Pour faciliter leur transcription, les textes ont parfois été divisés en plusieurs zones. C'est le cas des textes en colonnes ou des textes illustrés de plans.

Le numéro de la zone de texte est indiqué avant celui de la ligne.

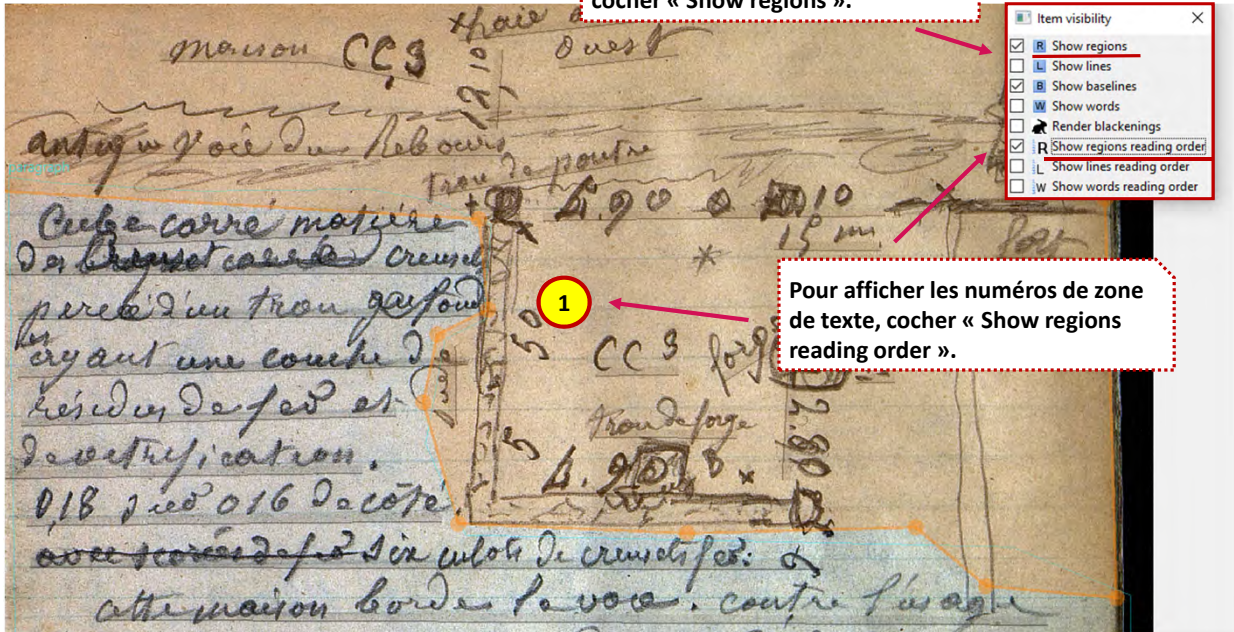
Zone 1

- 1-1 Fort est
- 1-2 Fossé extérieur
- 2-1 La base de la muraille
- 2-2 d'enceinte de l'oppidum

Zone 2

Les zones de texte (text regions)

Pour afficher les zones de texte, cocher « Show regions ».



Pour afficher les numéros de zone de texte, cocher « Show regions reading order ».

Indiquer les ratures, les mots soulignés, les exposants

à dix mètres au sud de la crête de
retranchement supérieur au revers nord de
la butte du Champlain vaste construction avec
assises en pierres brutes reliées par la terre glaise, ~~aire~~ travail rudimentaire
pour solide ~~1 m² de base~~ ~~1 m² de base~~ elle a en largeur 7 m
du nord au sud et 4 m de l'est à l'ouest.
L'aire, enfouie à 1,30 de profondeur est en terre
glaise

1-6 a dix mètres
1-7 retranchem
1-8 la butte du Champlain, vaste construction avec
1-9 assises en pierres brutes reliées par la terre glaise, ~~aire~~ travail rudimentaire
1-10 informe à 1,30 de profondeur.

Utiliser les différents styles de texte pour rendre
les ratures, les mots soulignés et les exposants

Les caractères spéciaux

1-1 Feuilles du Beauvais

1-3 Champlain

1-4 Tranchée autour de la butte

Virtual keyboards

Greek/Coptic Hebrew Latin-1 Extended-A Latin-1 Extended-B Latin-1 Extended-D Latin-1 Suppl. »

« ¼ ½ ¾ ζ À Á Â Ã Ä Å Æ Ç È É Ê Ë Ì Í Î Ï Ð Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß à á â ã ä å æ ç è é ê ë ì í î ï ð ñ ò ó ô õ ö ÷ ø ù ú û ü ý þ ÿ

1) Cliquer sur le clavier virtuel

2) Cliquer sur le caractère

Affichage des lignes et des numéros de ligne

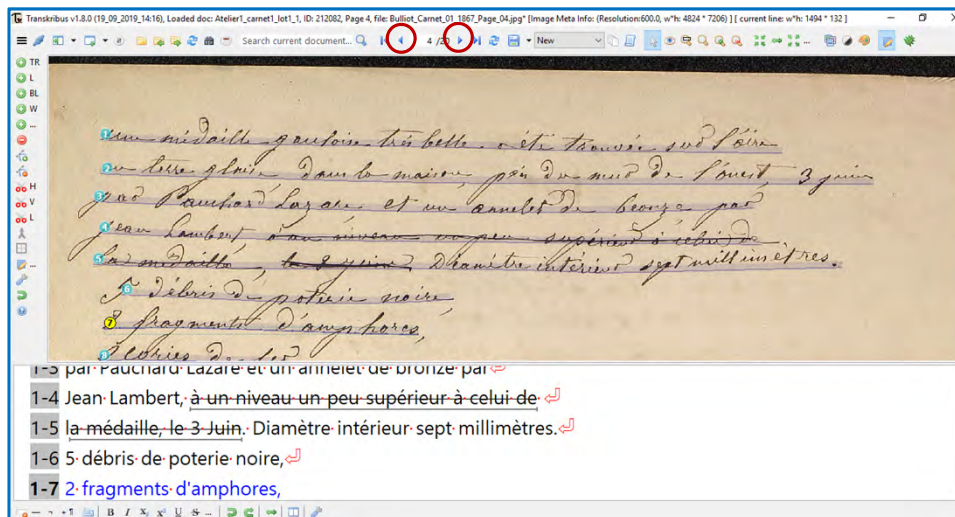
1) Pendant la saisie du texte, décocher « Show region »

2) Cocher ou décocher « Show baselines » pour souligner ou non les lignes

3) Cocher ou décocher « Show lines reading order » pour afficher ou non les numéros de ligne

1-1 une médaille gauloise très belle a été trouvée sur l'aire
1-2 en terre glaise dans la maison, près du mur de l'ouest 3 juin
1-3 par Pauchard Lazare et un anneau de bronze par
1-4 Jean Lambert, à un niveau un peu supérieur à celui de

Changer de pages



Modifier un numéro de ligne

1) Cliquer sur le numéro de la ligne

Change Reading Order

Please enter new reading order value:

1

Do it for all following

OK Cancel

2) Inscrire le nouveau numéro

3) Valider

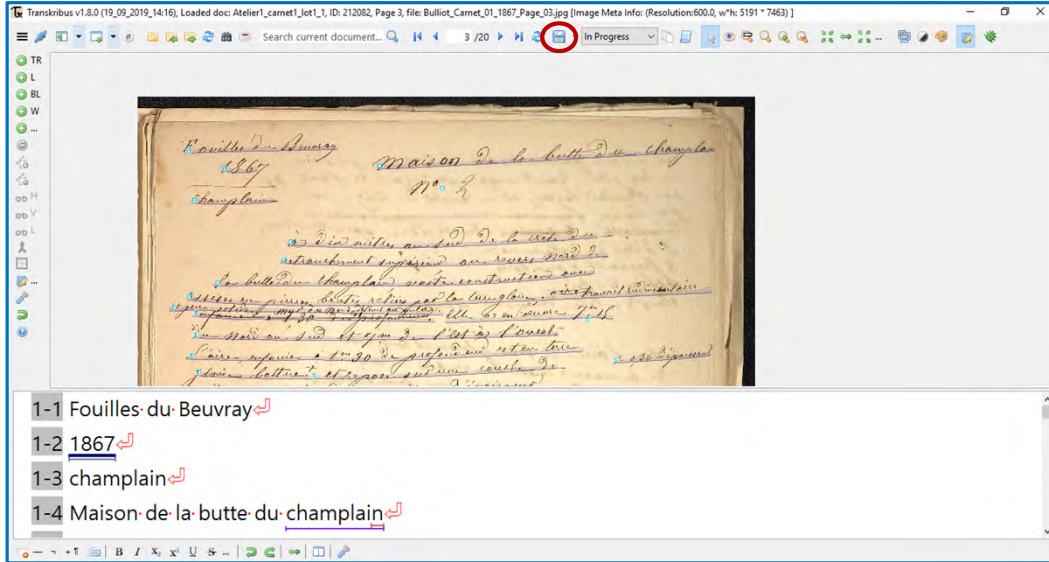
1-1 une médaille gauloise très belle a été trouvée sur l'aire

1-2 en terre glaise dans la maison, près du mur de l'ouest 3 juin

1-3 par Pauchard Lazare et un anneau de bronze par

1-4 Jean Lambert, à un niveau un peu supérieur à celui de

Sauvegarder



Indiquer une incertitude de lecture (*unclear*)

1) Cliquer sur l'onglet « Metadata » puis sur l'onglet « Textual »

2) Dans l'onglet « Tags », cocher « Show all »

Tag	Value	Text
1	textStyle	à un nh Jean Lambert, 30
2	textStyle	a méda supérieur à celui
3	unclear	son facile écoulement
4	textStyle	antéris débris de poterie
5	textStyle	à hauteur, un vase fait au tour,
6	textStyle	au-des: la bande unie au-dessus de 05 de la

3) Sélectionner le mot

4) Cliquer sur + dans la ligne « unclear »

5) Dans « reason », indiquer la cause de l'incertitude (rature, tache, etc.) et dans « alternative », une autre possibilité de lecture si besoin est

Indiquer un mot illisible (gap)

1) Cliquer sur l'onglet « Metadata » puis sur l'onglet « Textual »

2) Dans l'onglet « Tags », cocher « Show all »

3) Dans la phrase, cliquer à l'emplacement du mot illisible

4) Cliquer sur + dans la ligne « gap » (lacune)

Tag	Value	Text
1	gap	extérieurement. du peu de consista

Tags of current Transcript

Tag specification	Color	Shortcut
div		
gap		
romanisation		

Props for tag: no tag selected

1-1 Un enduit noir la recouvre extérieurement. du peu ↵
1-2 de consistance de la pâte; il s'est levé de nombreuses écailles. ↵
fragments d'un vase analogue; mais d'une terre ↵
ême fouille; Les ornements étaient les ↵

Restituer un mot (*supplied*)

1) Cliquer sur l'onglet « Metadata » puis sur l'onglet « Textual »

2) Dans l'onglet « Tags », cocher « Show all »

3) Sélectionner la partie du mot à restituer

4) Cliquer sur + dans la ligne « supplied » (restitué)

Tag	Value	Text
1	date	1867
2	textStyle	Fouilles du Beuvray 1867 champlair
3	place	champl de la butte du champlair n° 2 à dix
4	supplied	n butte du champlair n° 2 à dix mètr
5	comment	2 du champl
6	place	champl de la butte
7	textStyle	aire terre glaise
8	textStyle	enfoiie travail rudi
9	unclear	l' milieu. Elle
10	textStyle	m a en l'oeuvre 705 du nord au sud
11	textStyle	nord au sud m de l'est à l'ouest
12	textStyle	m l'aire enfonc à 1m30 de profondes

All Tags Show all Customize...

Tag specification	Color	Shortcut
speech		
supplied		
unclear		

Props for tag: 'supplied' - value: 'n'
reason bord de la page

1-1 Fouilles du Beuvray
1-2 1867
1-3 champlair
1-4 Maison de la butte du champlair

Les abréviations

1) Cliquer sur l'onglet « Metadata » puis sur l'onglet « Textual »

2) Dans l'onglet « Tags », cocher « Show all »

3) Sélectionner l'abréviation

4) Cliquer sur + dans la ligne « abbrev »

5) Développer l'abréviation

Tag	Value	Text
1	textStyle	à un ni Jean Lambert, à un niveau un peu
2	textStyle	a méda supérieur à celui de la médaille, le
3	textStyle	antérie débris de poterie, antérieurement rc
4	abbrev	N.O environ de l'angle N.O il a été possi
5	textStyle	à hauteur, un vase à fait au tour,
6	textStyle	au-dess la bande unie
7	textStyle	Le vase de 05 de large,

Tags of current Transcript

All Tags

Show all

Tag specification	Color	Shortcut
abbrev		
add		
addrec		

Props for tag: 'abbrev' - value: 'N.O'

expansion nord-ouest

1-19 l'intérieur, à un mètre environ de l'angle N.O

reconstituer en partie

à hauteur, un vase à fait au tour,

Indiquer une date

1) Cliquer sur l'onglet « Metadata » puis sur l'onglet « Textuel »

The screenshot shows a digital manuscript viewer with a sidebar on the left and a main document area on the right. The sidebar has tabs for 'Server Overview', 'Layout', 'Metadata', and 'Tools'. The 'Metadata' tab is active, and within it, the 'Textuel' sub-tab is selected. A table titled 'Tags of current Transcript' is visible, listing tags like 'place', 'date', and 'champl' with their corresponding values and text. A red box highlights the 'date' tag in this table. Below the table, there is a 'Tags' panel with a 'Show all' checkbox checked. A red box highlights this checkbox. In the 'All Tags' section, the 'date' tag is highlighted with a red box. A red box also highlights a '+' icon next to the 'date' tag. The main document area shows a handwritten manuscript page with a red box around the word '1867'. Below the manuscript, a list of tagged text segments is shown, with '1-2 1867' highlighted by a red box. A red arrow points from the '1867' in the list to the '1867' in the manuscript image. The list includes: '1-1 Fouilles du', '1-2 1867', '1-3 champlain', and '1-4 Tranchée autour de la butte'.

2) Dans l'onglet « Tags », cocher « Show all »

3) Sélectionner la date

4) Cliquer sur + dans la ligne « date »

Indiquer un nom de lieu

1) Cliquer sur l'onglet « Metadata » puis sur l'onglet « Textual »

2) Dans l'onglet « Tags », cocher « Show all »

3) Sélectionner le nom de lieu

4) Cliquer sur + dans la ligne « place »

Tag	Value	Text
1	place	Beuvray
2	date	1867
3	place	champl
4	place	champl
5	place	champl
6	place	champl
7	place	champl

Tag specification	Color	Shortcut
gap		
organization		
person		
place		
sic		

Props for tag: place

- country
- placeName
- type

1-1 Fouilles du Beuvray

1-2 1867

1-3 **champlain**

1-4 Tranchée autour de la butte

Indiquer un nom de personne

1) Cliquer sur l'onglet « Metadata » puis sur l'onglet « Textual »

Tag	Value	Text
1 person	Paucha	l'ouest, 3 juin par Pauchard Lazare,
2 person	Jean La	de bronze par Jean Lambert, à un n
3 textStyle	à un ni	Jean Lambert, à un niveau un peu
4 textStyle	a méda	supérieur à celui de la médaille, le
5	Son	facile écoulement. Son épaisseur, a
6 textStyle	antérieu	débris de poterie, antérieurement r
7 abbrev	N.O	environ de l'angle N.O il a été poss

2) Dans l'onglet « Tags », cocher « Show all »

Tag specification	Color	Shortcut
gap		
organization		
person		
place		
sic		

3) Sélectionner le nom de personne

4) Cliquer sur + dans la ligne « person »

1-1 une médaille gauloise
1-2 en terre glaise dans la
1-3 par Pauchard Lazare et un anneau de bronze par
1-4 Jean Lambert, à un niveau un peu supérieur à celui de

Faire un commentaire

1) Cliquer sur l'onglet
« Metadata » puis sur
l'onglet « Comments »

The screenshot shows the Transkribus v1.8.0 interface. The top menu bar includes 'Server', 'Overview', 'Layout', 'Metadata', and 'Tools'. The 'Metadata' tab is active, and the 'Comments' sub-tab is selected. On the left, there is a 'Comments' panel with an 'Add' button and a 'Comment' input field. The main area displays a handwritten document page with the text 'Maison de la butte du Champlain' and 'n° 2' highlighted. A red box highlights the 'n° 2' in the original image. Below the document, a list of comments is shown, with '1-5 n° 2' highlighted. A red arrow points from the 'Add' button to the '1-5 n° 2' comment.

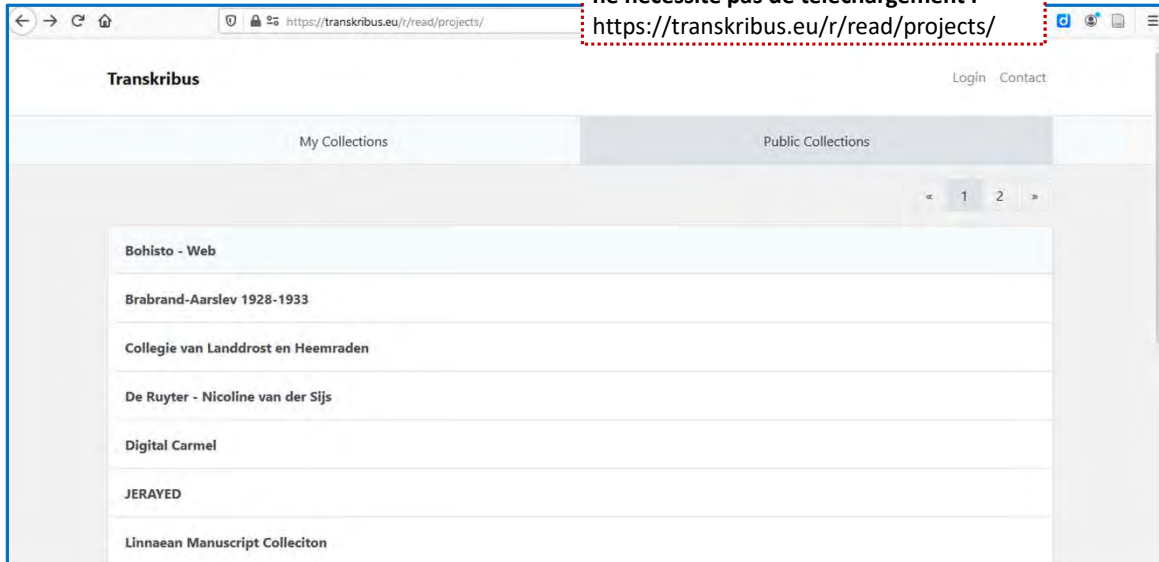
3) Écrire le
commentaire puis
cliquer sur « Add »

2) Sélectionner le mot ou le
passage à commenter

1-3 champlain ↵
1-4 Maison de la butte du Champlain ↵
1-5 n° 2 ↵
1-6 à dix mètres au sud de la crête du ↵

L'interface Web

Il existe également une interface web qui ne nécessite pas de téléchargement :
<https://transkribus.eu/r/read/projects/>



Se connecter à l'interface Web

Ce sont les mêmes identifiants que ceux utilisés pour l'enregistrement sur la plateforme principale

<https://transkribus.eu/Transkribus/>

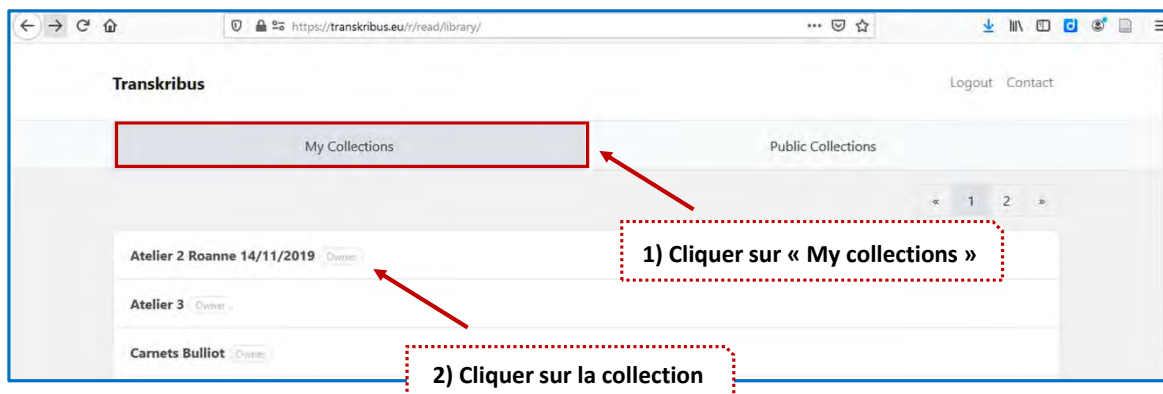
The screenshot shows the Transkribus login page. The browser address bar displays <https://transkribus.eu/r/read/login>. The page header includes the Transkribus logo and navigation links for 'Login' and 'Contact'. The main content area is titled 'Login' and contains a form with the following elements:

- A 'Username' input field.
- A 'Password' input field.
- A 'Login' button.
- Links for 'Lost password?' and 'Create Account'.

Three numbered instructions are overlaid on the page:

- 1) Cliquer sur Login (indicated by an arrow pointing to the 'Login' button in the top right navigation bar).
- 2) Remplir le formulaire (indicated by an arrow pointing to the Username and Password input fields).
- 3) Valider (indicated by an arrow pointing to the 'Login' button).

Accéder à ses documents depuis l'interface Web



Accéder à ses document depuis l'interface Web

Atelier 2 Roanne 14/11/2019

Document
carnet01_61_67 (252071)

Status
--

Page 1
In Progress
Changed by philippe.chassignet
Plaintext Annotation

Page 2
In Progress
Plaintext Annotation

Page 3
In Progress
Changed by philippe.chassignet
Plaintext Annotation

Page 4
In Progress
Changed by emmanuelle.perrin
Plaintext Annotation

Cliquer sur « Plaintext »

Transcrire ses documents depuis l'interface Web

The screenshot displays the Transkribus web interface for transcribing a handwritten document. The document image is at the top, showing cursive text. Below it, the transcription is visible in a text editor. The interface includes a toolbar with text styling options (bold, italic, underline, strikethrough) and a dropdown menu for annotations (COMMENT, DATE, ORGANIZATION, PERSON, PLACE, UNCLEAR). A 'Save Changes' button is also visible.

Annotation

Styles de texte

Sauvegarde

In Progress

Save Changes

Tutoriel établi par Emmanuelle Perrin (Archéorient, université de Lyon), dans le cadre du projet *Bulliot, Bibracte et moi*, à partir des remarques des participants aux ateliers de transcription, qui se sont déroulés le 21 septembre 2019, au musée de Bibracte et le 14 novembre 2019 au musée Déchelette à Roanne.

Version 2 du 22 novembre 2019

