



HAL
open science

Transcription, Parcours Humanités numériques de l'École normale supérieure, programme gradué Translitterae ENS/PSL, 2018-2019

Marie-Laure Massot

► To cite this version:

Marie-Laure Massot. Transcription, Parcours Humanités numériques de l'École normale supérieure, programme gradué Translitterae ENS/PSL, 2018-2019. Licence. France. 2018. <hal-02398087>

HAL Id: hal-02398087

<https://hal.science/hal-02398087v1>

Submitted on 6 Dec 2019

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.



HAL Authorization

Parcours Humanités numériques de l'ENS Transcription

Marie-Laure Massot

Ingénieure d'études SHS

CAPHÉS, UMS 3610 – CNRS/ENS Paris

Programme gradué Translitterae 2018-2019

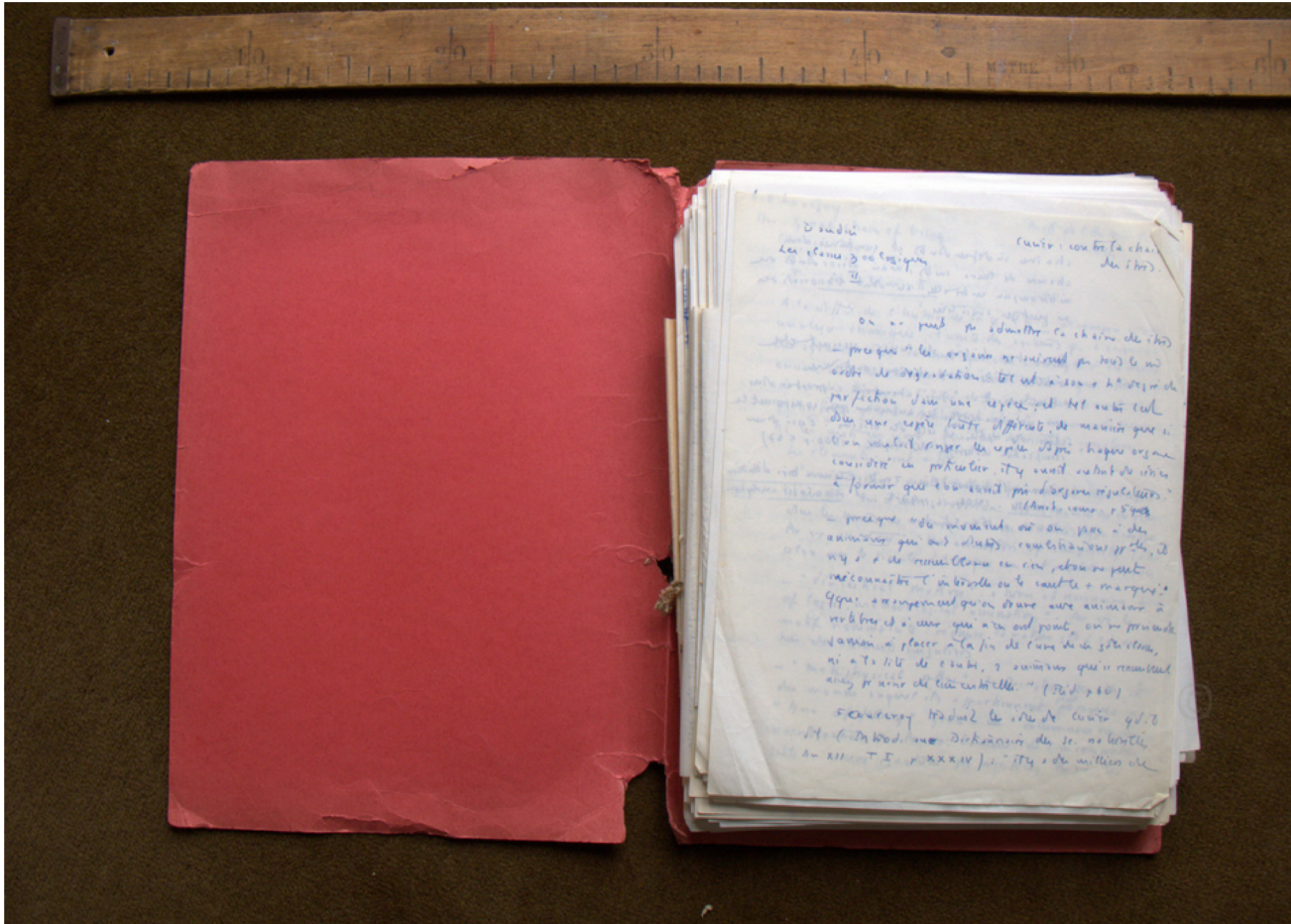
Les bases de la gestion de données numériques pour la recherche

Parcours Humanités numériques de l'ENS

Transcription

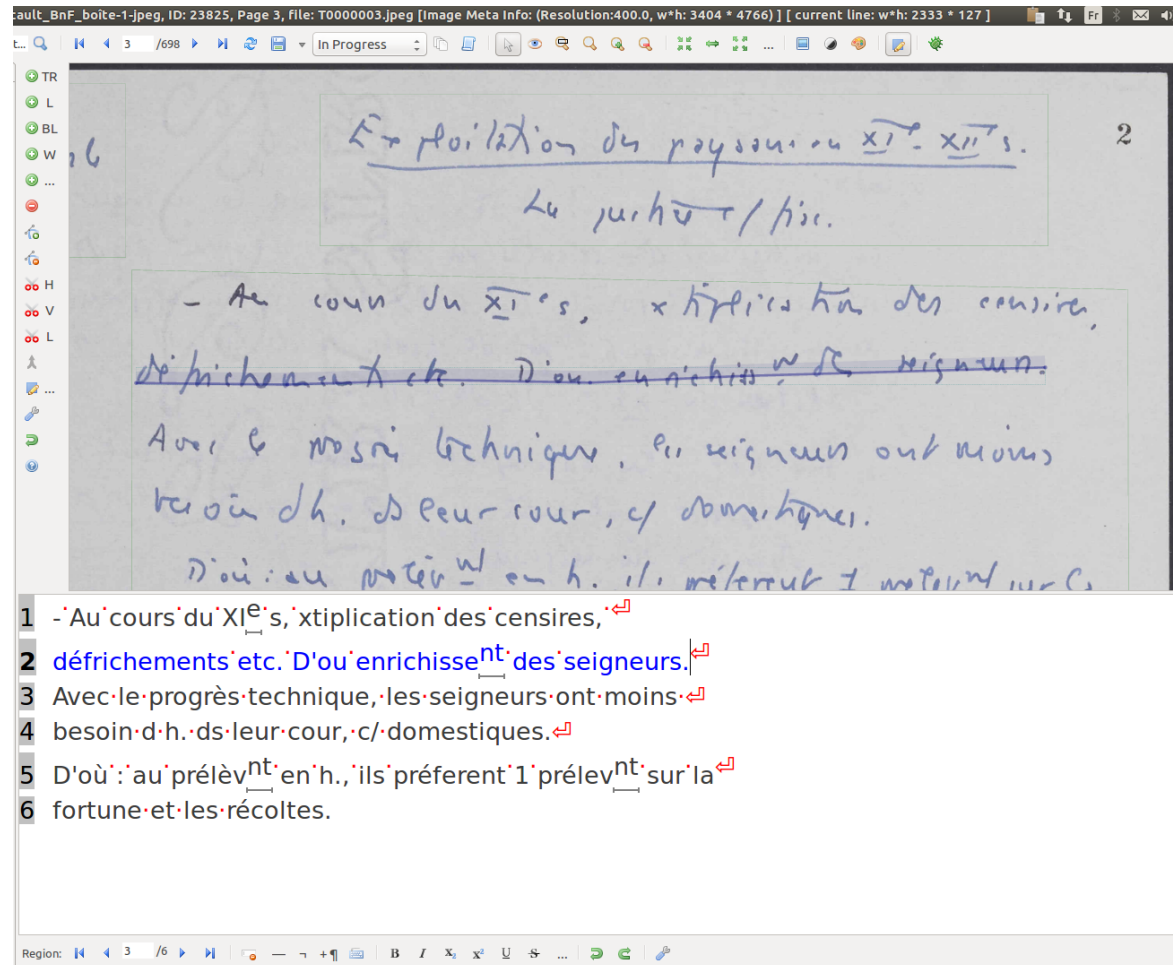
- Qu'est-ce que transcrire ?
- Pourquoi transcrire ?
- Transcriptions collaborative et automatique : *Transcrire* et *Transkribus*
- Avantages et limites de la transcription
- Ressources

Qu'est-ce que transcrire ?



- En **édition numérique**, transcrire c'est **reproduire sous forme textuelle un texte qui a été numérisé en mode image**.
- 20 000 pages manuscrites
- Conservation
- Mise à disposition
- Exploration du corpus
- Mieux comprendre le processus d'écriture du philosophe

Qu'est-ce que transcrire ?



The screenshot shows a digital interface for transcribing handwritten text. The top part displays the original handwriting on a page titled "Exploitation des paysans au XI^e - XII^e s." with a page number "2". The text is written in a medieval script. Below the original text, a transcription is shown with numbered markers (1-6) and red arrows indicating corrections or insertions. The transcription reads: "1 - Au cours du XI^e s, xtiplication des censires, 2 défrichements etc. D'ou enrichisse^{nt} des seigneurs. 3 Avec le progrès technique, les seigneurs ont moins 4 besoin d'h. ds leur cour, c/ domestiques. 5 D'où : au prélè^{nt} en h., ils préfèrent 1 prélev^{nt} sur la 6 fortune et les récoltes."

1 - Au cours du XI^e s, xtiplication des censires,
2 défrichements etc. D'ou enrichisse^{nt} des seigneurs.
3 Avec le progrès technique, les seigneurs ont moins
4 besoin d'h. ds leur cour, c/ domestiques.
5 D'où : au prélè^{nt} en h., ils préfèrent 1 prélev^{nt} sur la
6 fortune et les récoltes.

Pourquoi transcrire ?

- **Enregistrer** un texte oral ou écrit en vue de **conservation**
- **Mettre à disposition** du public
- **Valoriser** des corpus et les rendre plus accessibles
- **Faciliter** la lecture, ou le déchiffrage
- **Rechercher** plus facilement dans le texte
- Permettre des **analyses**, la **Fouille de texte**
- **Interpréter** : pas un simple miroir, mais une représentation des **choix du transcripteur** (Pas de transcription neutre ou objective, choix à faire en fonction de l'analyse visée)

Définition

The screenshot shows the CNRTL website interface. At the top, there's a search bar with the text 'www.cnrtl.fr/definition/transcrire'. Below it, the CNRTL logo and 'Centre National de Ressources Textuelles et Lexicales' are visible. A navigation menu includes 'Accueil', 'Portail lexical', 'Corpus', 'Lexiques', 'Dictionnaires', 'Métalexicographie', 'Outils', and 'Contact'. The main content area is titled 'Lexicographie' and shows the search results for 'transcrire'. The definition is presented in a structured format with sections A, B, and 3, each containing detailed linguistic information and references.

www.cnrtl.fr/definition/transcrire

Centre National de Ressources Textuelles et Lexicales

Ortolang
Outils et Ressources pour un Traitement Optimisé de la LANGue

Accueil Portail lexical Corpus Lexiques Dictionnaires Métalexicographie Outils Contact

Morphologie **Lexicographie** Etymologie Synonymie Antonymie Proxémie Concordance Aide

Entrez une forme transcrire Chercher

options d'affichage catégorie : toutes

TRANSCRIRE, verbe trans.

A. –

1. Reproduire très exactement, par l'écriture, ce qui a déjà été écrit. Synon. copier, enregistrer, recopier. *Transcrire des inscriptions, une lettre, un manuscrit, des notes, des passages, des phrases, des scholies, un texte, des versets; transcrire littéralement, textuellement, mot à mot. Ce qu'il copiait lui semblait si absurde, qu'il en vint à transcrire ligne par ligne, sans songer au sens* (STENDHAL, *Rouge et Noir*, 1830, p. 407). *Je me suis demandé cent fois, en transcrivant ces pages, si Justin avait réellement tenu le rôle de « meneur de jeu » qu'il semble s'attribuer dans toute la fin de l'ouvrage* (DUHAMEL, *Nuit St-Jean*, 1935, p. 10).
♦ [Avec un compl. prép. sur] *Je me reproche de n'avoir pas, au jour le jour, transcrit sur un carnet spécial les phrases glanées au cours de mes lectures* (GIDE, *Journal*, 1942, p. 157).
– **Transcrire qqc. à qqn.** *Transcrivez-moi ce cahier* (Ac.). *Je viens de lire une bonne phrase dans l'Émancipation (...). Je te la transcris* (HUGO, *Corresp.*, 1852, p. 72).

2. **DR.** „Reproduire par copie manuscrite la teneur d'un acte sur un autre document, généralement un registre“ (ROLAND-BOYER 1983). Synon. enregistrer. *Transcrire un acte, des jugements. Transcrire un contrat sur le registre des hypothèques* (Ac. 1835-1935). *L'acte fut minuté par le maître clerc, revu par les hommes d'affaires des deux fiancés et transcrit définitivement sur un beau cahier de papier timbré où il ne manquait plus que les signatures* (ABOUT, *Nez notaire*, 1862, p. 189). V. conservateur ex. du Code civil.

B. – Reproduire exactement par écrit à l'aide d'autres signes, d'un système de notation différent, d'un autre code.
Transcrire du grec en caractères latins; transcrire en clair un message chiffré; transcrire un papyrus.

1. [Le compl. désigne un texte écrit ou prononcé] *Les organes vocaux sont aussi extérieurs à la langue que les appareils électriques qui servent à transcrire l'alphabet morse sont étrangers à cet alphabet* (SAUSS. 1916, p. 36). *Tout cela sans écrire une seule formule, mais dans un langage si transparent et si précis qu'on peut immédiatement le transcrire en formules comme il vient d'être fait* (BOURBAKI, *Hist. math.*, 1960, p. 207).
– **LING.** [Le compl. désigne des paroles prononcées ou un texte enregistré sur un support sonore] „Faire correspondre terme à terme les unités discrètes de la langue parlée et les unités graphiques“ (Ling. 1972, s.v. transcription).
Transcrire en alphabet phonétique. Un polytechnicien (...) qui se fait parfaitement comprendre et qui séduit son auditoire par son éloquence; une fois transcrit, son discours donne l'impression du décousu (Cl. BLANCHE-BENVENISTE, C. JEANJEAN, *Le Fr. parlé*, 1987, p. 152):
• Si le français était une langue de sauvages, non fixée par l'écriture, un voyageur-linguiste, recueillant sur les lèvres des indigènes le présent du verbe *aimer*, le **transcrirait** ainsi: *Jèm, tuèm, ilèm, nouzèmon, vouzémé, ilzèm*. BALLY, *Lang. et vie*, 1952, p. 25.

2. **INFORMAT.** **Recopier des données, avec ou sans changement de code, et les transférer** (d'apr. LE GARFF 1975).

3. [Le compl. désigne un élément non ling., paysage ou mouvement] *Comme Chardin, comme Vermeer, il [Poussin] transcrit la nature, mais il est loin de se soumettre à elle* (MALRAUX, *Voix sil.*, 1951, p. 296). *Pour transcrire exactement la position des diverses articulations et les trajets accomplis, on exigea du chorégraphe: 1. la perception, rapide, précise et fidèle des formes; 2. certaines connaissances anatomiques* (BOURGAT, *Techn. danse*, 1959, p. 37).

Si transcrire s'est reproduire très exactement, terme à terme, recopier des données avec ou sans changement de code, c'est aussi choisir, mettre en place des pratiques, conventions de transcription en fonction du type d'édition scientifique visée.

Définitions issues du glossaire

Edition fac-similé : **Reproduction** du corpus de manuscrits (copie, procédé photographique)

Edition critique : **Transcription** et **analyse** des manuscrits (enrichissement, documentation, annotation, visualisation, fouille de texte, etc.)

Edition génétique : Dossiers génétiques des œuvres. L'ambition d'une édition génétique est de publier l'œuvre d'un auteur de manière à représenter et à **rendre intelligible la genèse de ses projets d'écriture ou de création artistique**. Voir Paolo D'Iorio, « Qu'est-ce qu'une édition génétique numérique ? », *Genesis* [En ligne], 30 | 2010, mis en ligne le 30 mai 2012, consulté le 17 septembre 2018. URL : <http://journals.openedition.org/genesis/116> ; DOI : 10.4000/genesis.116

Transcription : En **paléographie**, la transcription consiste à **reproduire un texte manuscrit**, en notant les particularités du texte et rétablissant (ou non) les erreurs ou les abréviations qu'il peut contenir ; on parle de **transcription diplomatique** quand tous **les phénomènes visibles du texte sont reproduits** (comme la reproduction des retours à la ligne). En **édition numérique**, il s'agit de **reproduire sous forme textuelle un texte qui a été numérisé en mode image**. La transcription en contexte numérique obéit aux mêmes principes et méthodologies que la transcription « classique », les problèmes étant les mêmes.

Parcours Humanités numériques de l'ENS Transcription

Origami

dalembert.obspm.fr/Intranet/Correspondance_ORIGAMI/navhorizontale.php?book=5620&version=manuscrit1

Confrontation des folios et de leur transcription

Options | Mode de transcription

LETTRE 56.20 | 26 novembre 1756
De D'Alembert (Paris) à Formey (Berlin)

Présentation de la lettre | Description matérielle | Citer la lettre | Cachet de la lettre

Version : Manuscrit autographe Folio : 1r

Monsieur

Je reponds sur le champ à la lettre que vous me faites l'honneur de m'ecrire en date du 13, & que je viens de recevoir. M. Euler me rendroit bien peu de justice, s'il croyoit que mon intention fût le moins du monde de le desobliger, ou de paroître lui chercher querelle. Je vous avoie cependant que si j'ay jamais cru avoir raison sur quelque chose, c'est sur les differens points qui font l'objet du memoire en question, & si quelque chose pouvoit m'en faire desirer l'impression, c'est de voir la reponse que M^f. Euler y a jointe, car plus j'examine la chose, moins je vois ce qu'il peut repliquer à mes raisons. Ce n'est point, comme il le croit, dans des momens de vivacité que j'ai escrit ce memoire, mais a tête bien reposée, & je crois y avoir traité mes deux adversaires avec toute la politesse possible. Si la note que j'ay mise à la page 3 deplait à M^f. Euler, je le prie de se mettre un moment à ma place, et de juger si je n'ai pas raison de me plaindre. J'espere qu'il y fera reflexion, & qu'il voudra bien sur les deux points dont il s'agit dans cette note me rendre la justice qui m'est dûe. J'ose l'assurer que j'ai eu bonne part, et peut être la principale à'

Parcours Humanités numériques de l'ENS Transcription

Origami

dalembert.obspm.fr/Intranet/Correspondance_ORIGAMI/navhorizontale_couleur.php?taille=2&ocf=0&book=5620&version=man

Google

CAPHES GESTION http://www.doh... caes.cnrs.fr — ca... uncorked - Engl... D'Alembert : édi... SYRTE - Observ... ENCYCLOPEDIE Recherches sur ...

Confrontation des folios et de leur transcription

Options | Mode de transcription

LETTRE 56.20 | 26 novembre 1756
De D'Alembert (Paris) à Formey (Berlin)

Description matérielle | Description bibliographique | Cachet de la lettre

Version : **Manuscrit autographe** Folio : 2r

M^f. Euler voudra bien me faire parvenir **en même tems et dans le même paquet** pour ma propre instruction, un extrait de sa reponse, ou sa reponse même. Pour l'y engager, je lui donne ma parole d'honneur de me retracter publiquement dans vos memoires, si je trouve qu'il ait raison. Si M^f. Euler me refuse cette grace, en ce cas je vous prie d'imprimer mon memoire, tel qu'il est, et sans y changer une syllabe. Je vous prie instamment de me rendre reponse sur le champ, parce que si mon memoire est imprimé, je vous enverrois **tout de suite** une addition assez curieuse, qui ne regarde nullement M. Euler ni M. Bernoulli. M. Euler peut etre assuré que je ne garderai pas longtems son memoire, et que je le lui renverrai, par quelque occasion.

M. de Maupertuis est à S^t. Malo ; je n'en ai point de nouvelles directes, je le crois pourtant en bonne santé, mais les circonstances presentes rendent sa position embarrassante et fâcheuse.

J'ai recu le tome X des memoires, & je vous en suis infiniment obligé. Je suis aussi très sensible au parti que vous

à Paris le 26 Nov. 1756. *Maupertuis* D'Alembert

Transcriptions collaborative et automatique

Transcrire : un exemple de transcription collaborative

The screenshot shows the homepage of the Transcrire website. At the top, there is a navigation bar with the logo 'TRANSCRIRE' and links for 'TRANSCRIRE PAR COLLECTION', 'SE CONNECTER', and 'A PROPOS'. Below the navigation bar, a green banner contains the text: 'Participez à la transcription des archives et matériaux de terrains de la recherche en sciences humaines et sociales'. The main content area features a large image of a handwritten letter with a green overlay on the right side. The overlay contains the text: 'Correspondances, père Camille De la Croix', 'Camille de la Croix (1831-1911), prêtre jésuite et archéologue, a entretenu une correspondance abondante avec de nombreux savants de son époque, comme Ernest Babelon, Emile Espérandieu, Jean-Camille Formigé, Antoine Héron de Villefosse, Camille Jullian, Robert de Lasteyrie ou Salomon Reinach.' Below the image, there are three columns of text: 'Contribuez', 'Rejoignez-nous!', and 'Suivez-nous sur Twitter'. The 'Contribuez' section encourages users to discover archives and transcribe documents. The 'Rejoignez-nous!' section lists ways to contribute: 'Ouvrir un compte', 'Se connecter', and 'Contact'. The 'Suivez-nous sur Twitter' section shows a tweet from Transcrire (@transcrire) about a course on handwritten text recognition.

Transkribus : un exemple de transcription automatique

The screenshot shows the homepage of the Transkribus website. At the top, there is a navigation bar with the logo 'Transkribus' and links for 'Register' and 'Login'. Below the navigation bar, a large banner features the text: 'Transcribe. Collaborate. Share...' and '...and benefit from cutting edge research in Handwritten Text Recognition!'. Below the banner, there are two buttons: 'Download version 1.4' and 'Download version 1.4 for Mac', along with a link for 'Wiki • How-to guide (pdf)'. The main content area is divided into four columns: 'Scholars', 'Archives', 'Volunteers', and 'Scientists'. Each column contains a brief description of the target audience and a 'View details »' button. The 'Scholars' section asks if users transcribe historical documents and if they appreciate automated tools. The 'Archives' section asks if users are responsible for large collections and if they believe digitisation paves the way for new opportunities. The 'Volunteers' section asks if users find historical letters fascinating and if they believe everyone can contribute. The 'Scientists' section asks if users are computer scientists working in fields like computer vision and natural language processing.

La plateforme Transcrire

- <http://transcrire.huma-num.fr/> :
Un projet de sciences participatives en SHS
- Une plateforme libre de transcription collaborative de matériaux de terrains (collections patrimoniales de bibliothèques, carnets de terrains, correspondances, etc.)

The screenshot shows the homepage of the Transcrire platform. At the top, there is a dark navigation bar with the logo 'TRANSCRIRE' in green and white, and links for 'TRANSCRIRE PAR COLLECTION', 'SE CONNECTER', and 'A PROPOS'. Below the navigation bar, a green banner reads 'Participez à la transcription des archives et matériaux de terrains de la recherche en sciences humaines et sociales'. The main content area features a large image of a handwritten letter with a green overlay on the right side containing the text: 'Correspondances, père Camille De la Croix. Camille de la Croix (1831-1911), prêtre jésuite et archéologue, a entretenu une correspondance abondante avec de nombreux savants de son époque, comme Ernest Babelon, Emile Espérandieu, Jean-Camille Formigé, Antoine Héron de Villefosse, Camille Jullian, Robert de Lasteyrie ou Salomon Reinach.' Below the image, there are three columns of text: 'Contribuez' (Venez découvrir nos archives...), 'Rejoignez-nous !' (Aidez les bibliothèques à rendre ces documents plus utiles à la recherche...), and 'Suivez-nous sur Twitter' (Tweets de @transcrire). A 'Débutez la transcription' button is visible on the image.

La plateforme Transcrire

- Un espace en ligne pour **transcrire des documents manuscrits numérisés** en mode image.
- **Suivi, vérification, correction et validation** des textes transcrits.
- Les **corpus de données textuelles**, ainsi obtenus, peuvent être mis à disposition des chercheurs pour la **fouille, l'analyse et l'édition numérique de sources**.
- Les **archives** déposées sur *Transcrire* proviennent principalement de carnets de terrains de **chercheurs** mais aussi de **collections patrimoniales de bibliothèques** de recherches en sciences humaines et sociales (ethnologie, archéologie, géographie, histoire).

Parcours Humanités numériques de l'ENS Transcription

TRANSCRIRE TRANSCRIRE PAR COLLECTION | SE CONNECTER | A PROPOS

Carnets d'Annie-Hélène Dufour

Annie-Hélène Dufour (1947-2002), ethnologue. Ses matériaux de terrain portent sur la pêche, les espaces sauvages, la forêt, les jeux en Provence, l'ethnobotanique et plus largement sur les pratiques et les représentations de l'espace ou l'expression des identités collectives. Le fonds d'archives est déposé à la [Phonothèque de la Maison méditerranéenne des sciences de l'homme](#). L'inventaire est consultable sur [Calames](#).

 <p>Recherches sur les méthodes de pêche (Hyères)</p> <p>6% débuté</p>	 <p>Recherches sur la pêche et les pêcheurs (Var), 1984-1985</p> <p>8% débuté</p>	 <p>Notes et entretiens sur le mode de vie dans la région de Prali, 1981-1982</p> <p>65% débuté</p>	 <p>Notes et entretiens (Brusc, Six-Four), 1980</p> <p>À réviser</p>	 <p>Les jeux d'enfants (Saint Michel de l'Observatoire, Mane), 1981-1982</p> <p>À réviser</p>
---	--	--	---	---

TRANSCRIRE TRANSCRIRE PAR COLLECTION | SE CONNECTER | A PROPOS

Carnets d'Annie-Hélène Dufour

Les jeux d'enfants (Saint Michel de l'Observatoire, Mane), 1981-1982

Page 1

Transcrire

Ecole maternelle de Mane le 28.1.81

La marelle (garçons peut être) tout le tps de l'école primaire
Les billes
La toupie - garçons et filles

Toupies de bois + ficelle de lieuse prise à la maison pour attacher les gerbes de blé
Cadeau de Noël (père aîné)
Son père a fait un cheval de bois pour le Noël d'un père - sculpté lui-même - Il y avait des chevaux de labour chez les gds parents.

Grimper aux arbres - Cabanes dans les arbres, près de la maison et une près de la Laye vers Montagne de Lure.

Poupées, quelquefois achetées. Très laide mais habitait n'importe quoi = des courges ou des morceaux de bois choisis à cause de leur forme
Courges poussaient chez le gd père. Potage, gratin. Après qu'on les ait ramassées. Stockées dans un coin. On les prenait, on les reposait après avoir joué.

S'inscrire pour transcrire

◀ préc. suiv. ▶

Le logiciel Transkribus

- <https://read.transkribus.eu/transkribus/>
- Un outil pour transcrire automatiquement des manuscrits
- Un **logiciel de reconnaissance automatique** de l'écriture manuscrite, accompagné d'une **plateforme de transcription** d'images numérisées de manuscrits.

Transkribus Register Login

Transcribe. Collaborate. Share...

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

[Download version 1.4](#) [Download version 1.4 for Mac](#) [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 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 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

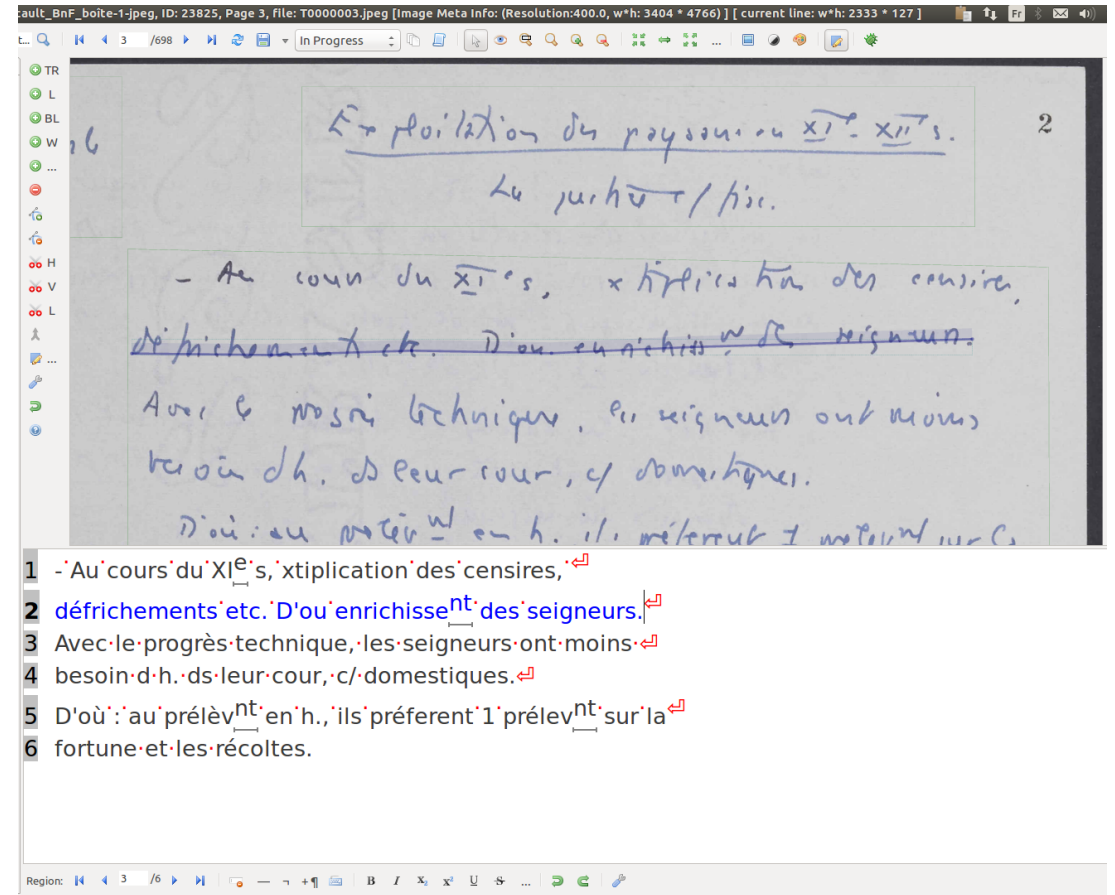
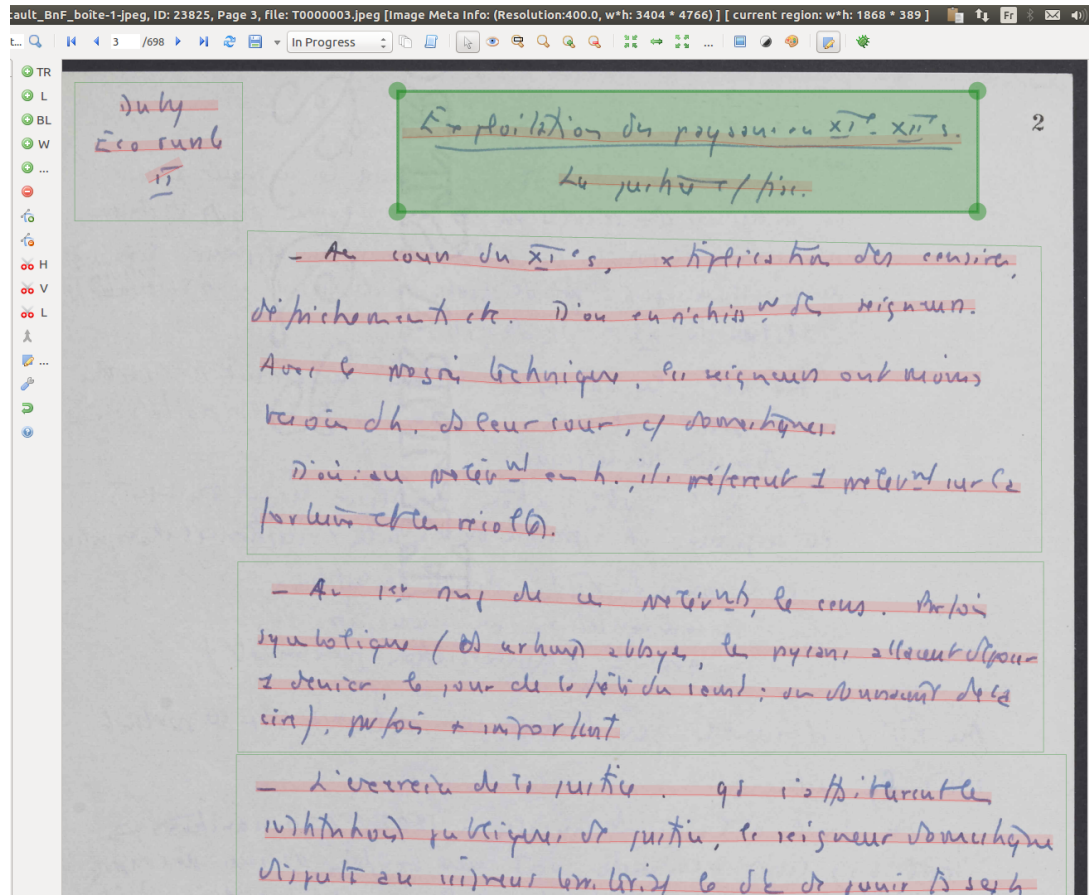
Are you a computer scientist and working in the fields of computer vision, document analysis, pattern recognition, natural language processing or a related field? You are seeking interesting documents from 1000 years of handwriting, printing and publishing? And you would really enjoy to get some reference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

[View details >](#)

Principaux usages de Transkribus


- Entraîner le **moteur de reconnaissance d'écriture manuscrite** (HTR : *Handwritten text recognition*) puis l'utiliser pour transcrire automatiquement les images fournies ;
- Transcrire des documents pour une **édition scientifique** (interface de saisie *Wysiwyg* avec encodage TEI et création de balises personnalisées) ;
- Faire des **recherches de termes** dans les documents manuscrits, y compris sur des termes proches (recherche "floue", *keyword spotting*).

Le logiciel Transkribus



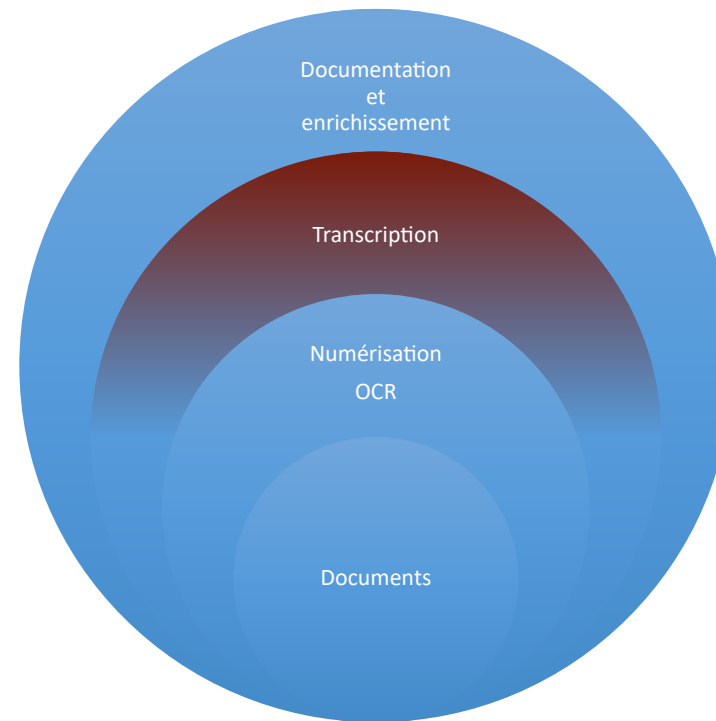
Avantages et limites de la transcription « massive » (collaborative ou automatique)

- Transcription facilitée
- Transcription accélérée
- Transcription de plus vaste corpus
- Valorisations de travaux scientifiques
- Edition et réutilisation des textes pour l'exploration et l'analyse informatique des données (standardisation, balises, TEI, etc.)
- Problème de qualité et d'uniformité des transcriptions obtenues
- Nécessité de mise en place de règles de transcriptions
- Révision, validation nécessaire par des spécialistes du corpus



Une base de données textuelles qui pourra être documentée, enrichie et explorée.
Création d'un savoir nouveau.

Gestion et manipulation de documents numériques



Pour conclure, dans l'**édition numérique**, transcrire s'est **reproduire sous forme textuelle un texte qui a été numérisé en mode image**, mais c'est aussi préparer les **analyses** envisagées en introduisant des **balises** pour ajouter des informations complémentaires au texte et préparer l'encodage TEI par exemple. Préparer le texte pour les explorations futures.

Ressources

D'Alembert en toutes lettres : <http://dalembert.academie-sciences.fr/Correspondance/>

Edition Numérique Collaborative et Critique de l'Encyclopédie (ENCCRE) :

<http://enccre.academie-sciences.fr/encyclopedie/>

Edition des manuscrits de Madame Bovary : <http://www.bovary.fr/>

Edition en ligne des Pensées de Montesquieu (Montedite) :

<https://www.unicaen.fr/services/puc/sources/Montesquieu/index.php?page=projet>

Eman Notes de cours de l'ENS : <http://www.eman-archives.org/coursENS/index.php>

Foucault fiches de lectures : <https://ffl.hypotheses.org/presentation-du-projet>

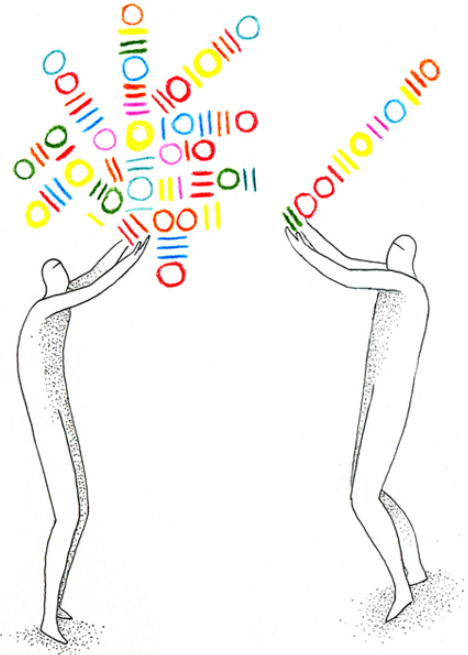
Letters of 1916 : <http://letters1916.maynoothuniversity.ie/>

Transcrire : <http://transcrire.huma-num.fr/a-propos-du-projet>

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

Transcription collaborative des Testaments de Poilus :

<http://www.archives-nationales.culture.gouv.fr/web/guest/transcription-collaborative-des-testaments-de-poilus>



les humanités numériques

