

Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Marie-Laure Massot, Jean-Philippe Moreux et Vincent Ventresque

CAPHÉS, UMS 3610 – CNRS/ENS PSL

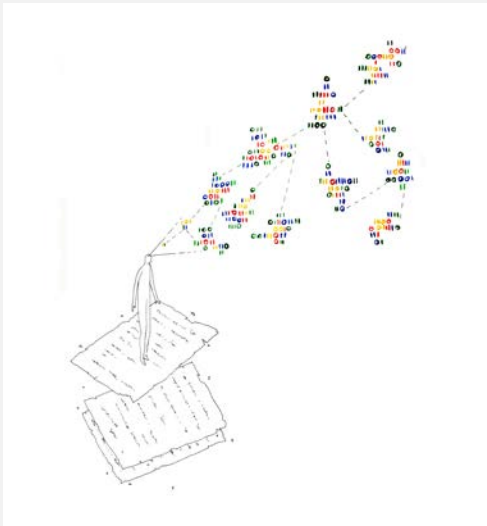
Service de la coopération numérique et de Gallica, BnF

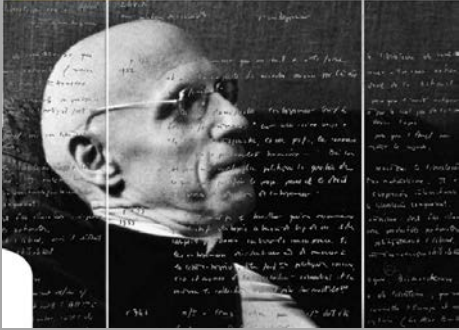
TRIANGLE, UMR 5206 – CNRS/ENS Lyon

Colloque de clôture du projet ANR Foucault Fiches de lecture

Seconde partie 26 septembre 2020 à la BnF

« Editer Michel Foucault (1994-2021) »





Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

- ❑ Usages de Transkribus pour les activités de recherche – L'ex. du projet ANR Ffl
- ❑ Pourquoi transcrire les fiches de lecture de Michel Foucault ?
- ❑ Transcrire automatiquement les fiches de lecture avec Transkribus
- ❑ La fonction de repérage de mots clés (Keyword Spotting)
- ❑ Résultats et perspectives
- ❑ Ressources
- ❑ Démonstration de l'outil (2 vidéos)
- ❑ Expérimenter le logiciel Transkribus sur un corpus test (scénario)



Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Présentation de Transkribus



Transkribus est une **plateforme complète** pour la reconnaissance, la transcription et la recherche automatisées de documents historiques manuscrits ou imprimés.

Elle facilite la lecture, l'enrichissement et l'exploration des textes. Le logiciel est développé par le [projet européen READ](#) (*Recognition and Enrichment of Archival Documents*). Ce projet souhaiterait révolutionner l'accès aux documents d'archives en développant des outils numériques innovants ([Transkribus](#), [ScanTent](#), [read@search](#)).

Le logiciel est gratuit. Il suffit de s'inscrire sur le site web et télécharger Transkribus qui fonctionne sous Windows, MacOS et Linux.

The screenshot shows the Transkribus website homepage. At the top, there's a navigation bar with 'Transkribus' and links for 'Register' and 'Login'. Below this is a large banner with the text 'Transcribe. Collaborate. Share...' and '...and benefit from cutting edge research in Handwritten Text Recognition!'. There are buttons for 'Download version 1.11' and 'Download version 1.11 for Mac', along with a link to a '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 user group and a 'View details' button. Below these columns is a section titled 'Humanities Scholars' which lists the features of Transkribus, such as registering, downloading expert tools, uploading documents, and segmenting images. To the right of this list is a 'Get started!' section with a numbered list: 1. Register, 2. Download Transkribus, 3. Try out some test documents with automatically produced full-text in the TranskribusCloud collection. At the bottom right, there is an illustration of St. Jerome by Albrecht Dürer.

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

Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Présentation de Transkribus



La plateforme est développée depuis 2015, à l'université d'Innsbruck (Autriche) et a été financée par le programme de recherche européen H2020.

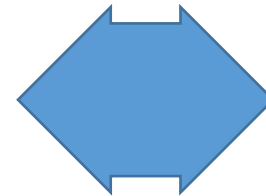
Fondé sur l'**intelligence artificielle**, le **moteur de reconnaissance de texte manuscrit** (Handwritten Text Recognition ou **HTR**) doit être préparé avec des **données d'apprentissage**, obtenues par la transcription d'une centaine de pages minimum, en établissant la correspondance ligne à ligne entre l'image du texte numérisé et sa transcription.

Cette plate-forme peut-être utilisée à différentes fins.

The screenshot shows the Transkribus website homepage. At the top, there's a navigation bar with 'Transkribus' and links for 'Register' and 'Login'. Below this is a large banner with the text 'Transcribe. Collaborate. Share...' and '...and benefit from cutting edge research in Handwritten Text Recognition!'. There are buttons for 'Download version 1.11' and 'Download version 1.11 for Mac', along with a link to a '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 user group and a 'View details' button. Below these columns is a section titled 'Humanities Scholars' which lists the features of the platform, such as registering, downloading expert tools, uploading documents, and segmenting images. To the right of this list is a 'Get started!' section with a numbered list of steps: 1. Register, 2. Download Transkribus, and 3. Try out some test documents. At the bottom right of the 'Humanities Scholars' section is an illustration of St. Jerome by Albrecht Dürer.

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

Usages de Transkribus pour les activités de recherche – ex. du projet Ffl



ANR-17-CE38-0001

triangle
UMR 5206

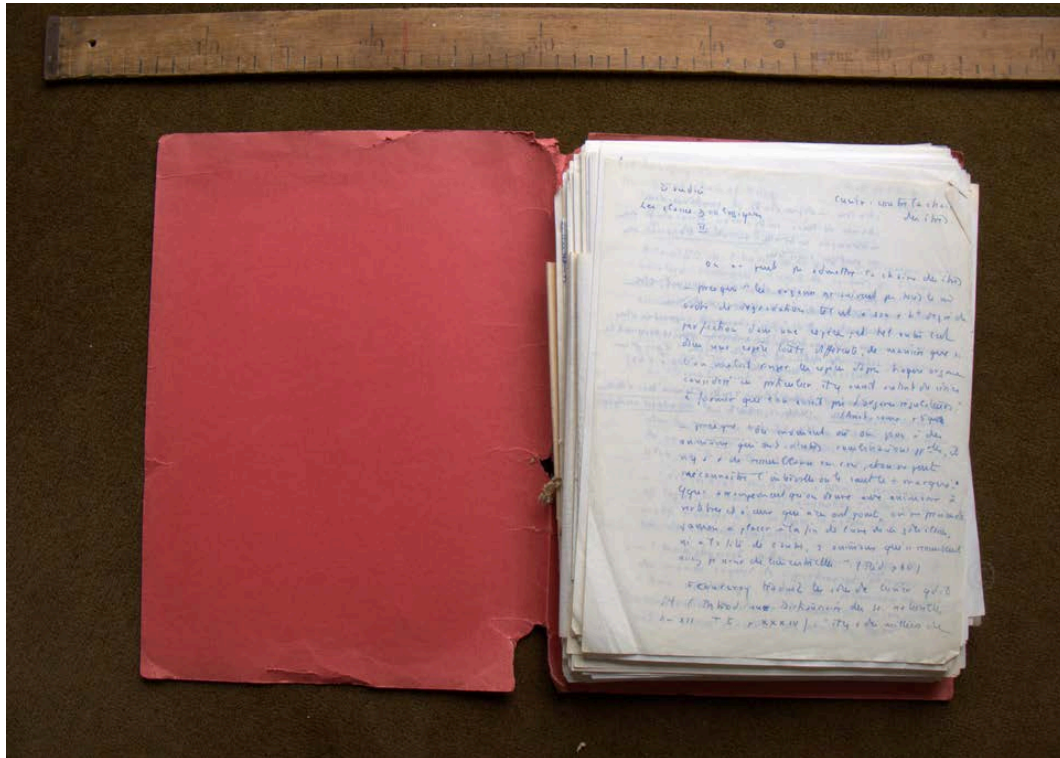
{ BnF



Une collaboration réussie entre un projet national et un projet européen.

Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Pourquoi transcrire les fiches de lecture de Michel Foucault ?



<p>Pièce n° : b001-f0001 Révoltes au moyen-âge. [couverture chemise]</p> <p>Revoltes au moyen-âge. [couverture chemise]</p>	<p>Pièce n° : b022-f0001 Noces spirituelles [rayé : Chair (Antiquité) Virginité] [couve]</p> <p>Noces spirituelles [rayé : Chair (Antiquité) Virginité] [couve]</p>	<p>Pièce n° : b038-f0107 L'éternité et l'amour</p> <p>L'éternité et l'amour</p>
<p>Pièce n° : b001-f0002 Duby. Eco rurale II. Exploitation des paysans au XIè-XI</p> <p>Duby. Eco rurale II. Exploitation des paysans au XIè-XI</p>	<p>Pièce n° : b022-f0002 Brouéhoux, p 86. Noces spirituelles chez Clément d'Alex</p> <p>Brouéhoux, p 86. Noces spirituelles chez Clément d'Alex</p>	<p>Pièce n° : b038-f0108 L'amour</p> <p>L'amour</p>
<p>Pièce n° : b001-f0003 Duby. Économie rurale II. Exploitation des paysans au</p> <p>Duby. Économie rurale II. Exploitation des paysans au</p>	<p>Pièce n° : b022-f0003 D. Perler. Ein Hymnus zur Ostervigil von Meliton. Freiburg</p> <p>D. Perler. Ein Hymnus zur Ostervigil von Meliton. Freiburg</p>	<p>Pièce n° : b038-f0109 Totalité et existence</p> <p>Totalité et existence</p>
<p>Pièce n° : b001-f0004 Siméon Luce. Histoire de la Jacquerie. Le droit à la ré</p> <p>Siméon Luce. Histoire de la Jacquerie. Le droit à la ré</p>	<p>Pièce n° : b022-f0004 Rouët de Journal. Textes ascétiques des pères de l'Église.</p> <p>Rouët de Journal. Textes ascétiques des pères de l'Église.</p>	<p>Pièce n° : b038-f0110 Psychologie et physiologie</p> <p>Psychologie et physiologie</p>
<p>Pièce n° : b001-f0005 A. Réville. Le soulèvement des travailleurs d'Angleterre</p> <p>A. Réville. Le soulèvement des travailleurs d'Angleterre</p>	<p>Pièce n° : b022-f0005 Virginité : mariage avec le Christ [?].</p> <p>Virginité : mariage avec le Christ [?].</p>	

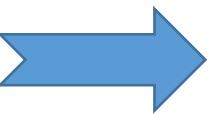
Rôle de la transcription au sein du projet FFL ?

Pourquoi transcrire en règle général ?

- **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)

Dans le cadre du projet :

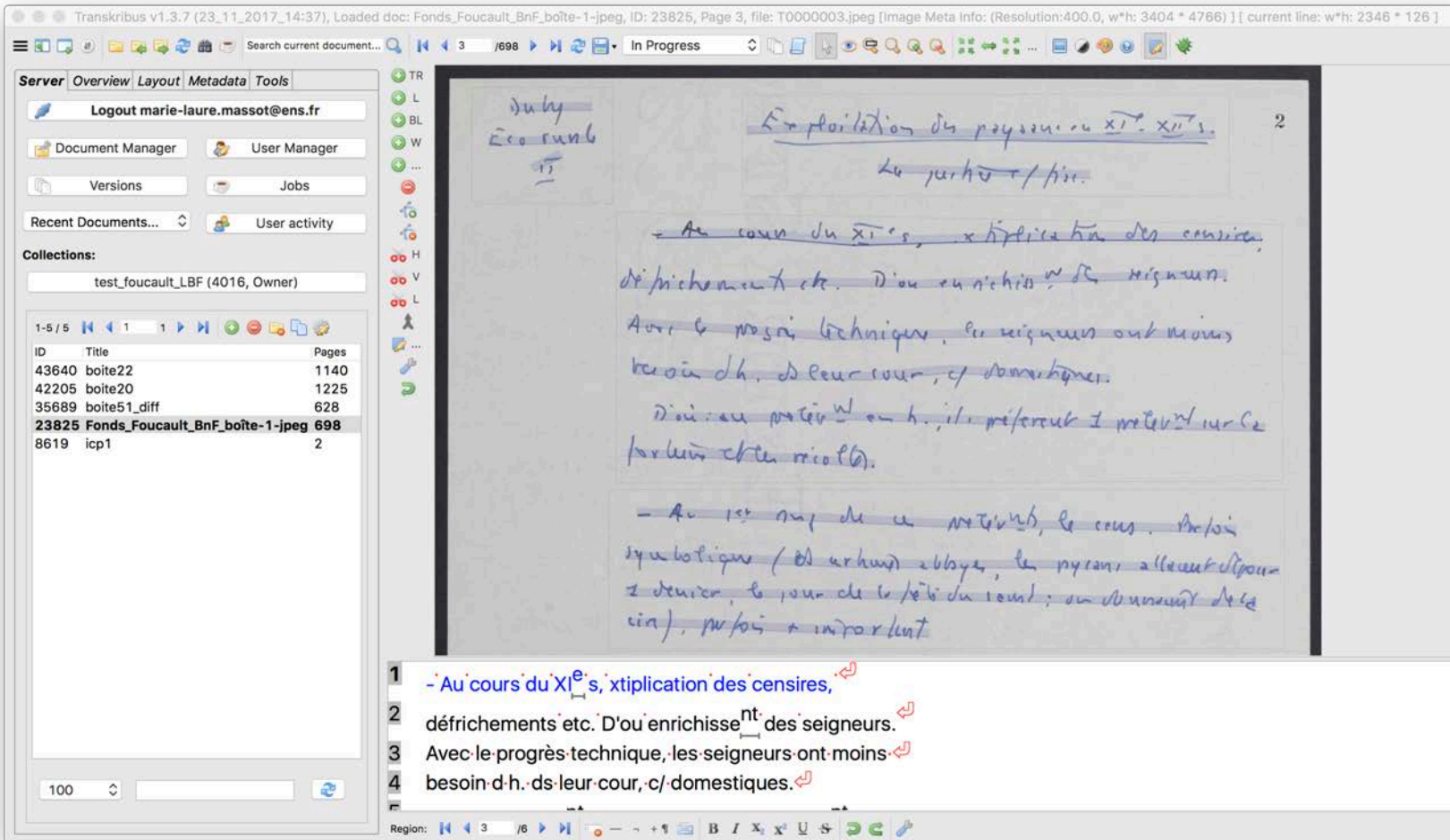
- La transcription avait été envisagée, puis retirée du projet (pertinence de transcrire des notes, des bouillons ; énormité du corpus ; écriture de Foucault peu accessible; etc.).
- Idée de faire un test de transcription automatique, en collaborant avec un projet européen en cours de développant d'un outil de transcription automatique prometteur et peu encore utilisé et connu de la communauté scientifique française.
- **Recherche d'occurrences** (auteurs, ouvrages, concepts, etc.) : Plein texte, recherche floue et KWS ; fouille de texte.
- **Indexation du corpus.**
- **Entraîner un Modèle HDR** pour l'écriture de M. Foucault qui pourrait ensuite être utilisé pour d'autres éditions du philosophe.



La transcription est la première phase de l'enrichissement d'un corpus.
Elle facilite la lecture et permet des recherches automatisées.

Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Transcrire automatiquement les manuscrits avec Transkribus



Transkribus v1.3.7 (23_11_2017_14:37), Loaded doc: Fonds_Foucault_BnF_boîte-1-jpeg, ID: 23825, Page 3, file: T0000003.jpeg [Image Meta Info: (Resolution:400.0, w*h: 3404 * 4766)] [current line: w*h: 2346 * 126]

Server Overview Layout Metadata Tools

Logout marie-laure.massot@ens.fr

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections:

test_foucault_LBF (4016, Owner)

ID	Title	Pages
43640	boite22	1140
42205	boite20	1225
35689	boite51_diff	628
23825	Fonds_Foucault_BnF_boîte-1-jpeg	698
8619	icp1	2

1-5 / 5

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

1001

1002

1003

1004

1005

1006

1007

1008

1009

1010

1011

1012

1013

1014

1015

1016

1017

1018

1019

1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063

1064

1065

1066

1067

1068

1069

1070

1071

1072

1073

1074

1075

1076

1077

1078

1079

1080

1081

1082

1083

1084

1085

1086

1087

1088

1089

1090

1091

1092

1093

1094

1095

1096

1097

1098

1099

1100

1101

1102

1103

1104

1105

1106

1107

1108

1109

1110

1111

1112

1113

1114

1115

1116

1117

1118

1119

1120

1121

1122

1123

1124

1125

1126

1127

1128

1129

1130

1131

1132

1133

1134

1135

1136

1137

1138

1139

1140

1141

1142

1143

1144

1145

1146

1147

1148

1149

1150

1151

1152

1153

1154

1155

1156

1157

1158

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

1172

1173

1174

1175

1176

1177

1178

1179

1180

1181

1182

1183

1184

1185

1186

1187

1188

1189

1190

1191

1192

1193

1194

1195

1196

1197

1198

1199

1200

1201

1202

1203

1204

1205

1206

1207

1208

1209

1210

1211

1212

1213

1214

1215

1216

1217

1218

1219

1220

1221

1222

1223

1224

1225

1226

1227

1228

1229

1230

1231

1232

1233

1234

1235

1236

1237

1238

1239

1240

1241

1242

1243

1244

1245

1246

1247

1248

1249

1250

1251

1252

1253

1254

1255

1256

1257

1258

1259

1260

1261

1262

1263

1264

1265

1266

1267

1268

1269

1270

1271

1272

1273

1274

1275

1276

1277

1278

1279

1280

1281

1282

1283

1284

1285

1286

1287

1288

1289

1290

1291

1292

1293

1294

1295

1296

1297

1298

1299

1300

1301

1302

1303

1304

1305

1306

1307

1308

1309

1310

1311

1312

1313

1314

1315

1316

1317

1318

1319

1320

1321

1322

1323

1324

1325

1326

1327

1328

1329

1330

1331

1332

1333

1334

1335

1336

1337

1338

1339

1340

1341

1342

1343

1344

1345

1346

1347

1348

1349

1350

1351

1352

1353

1354

1355

1356

1357

1358

1359

1360

1361

1362

1363

1364

1365

1366

1367

1368

1369

1370

1371

1372

1373

1374

1375

1376

1377

1378

1379

1380

1381

1382

1383

1384

1385

1386

1387

1388

1389

1390

1391

1392

1393

1394

1395

1396

1397

1398

1399

1400

1401

1402

1403

1404

1405

1406

1407

1408

1409

1410

1411

1412

1413

1414

1415

1416

1417

1418

1419

1420

1421

1422

1423

1424

1425

1426

1427

1428

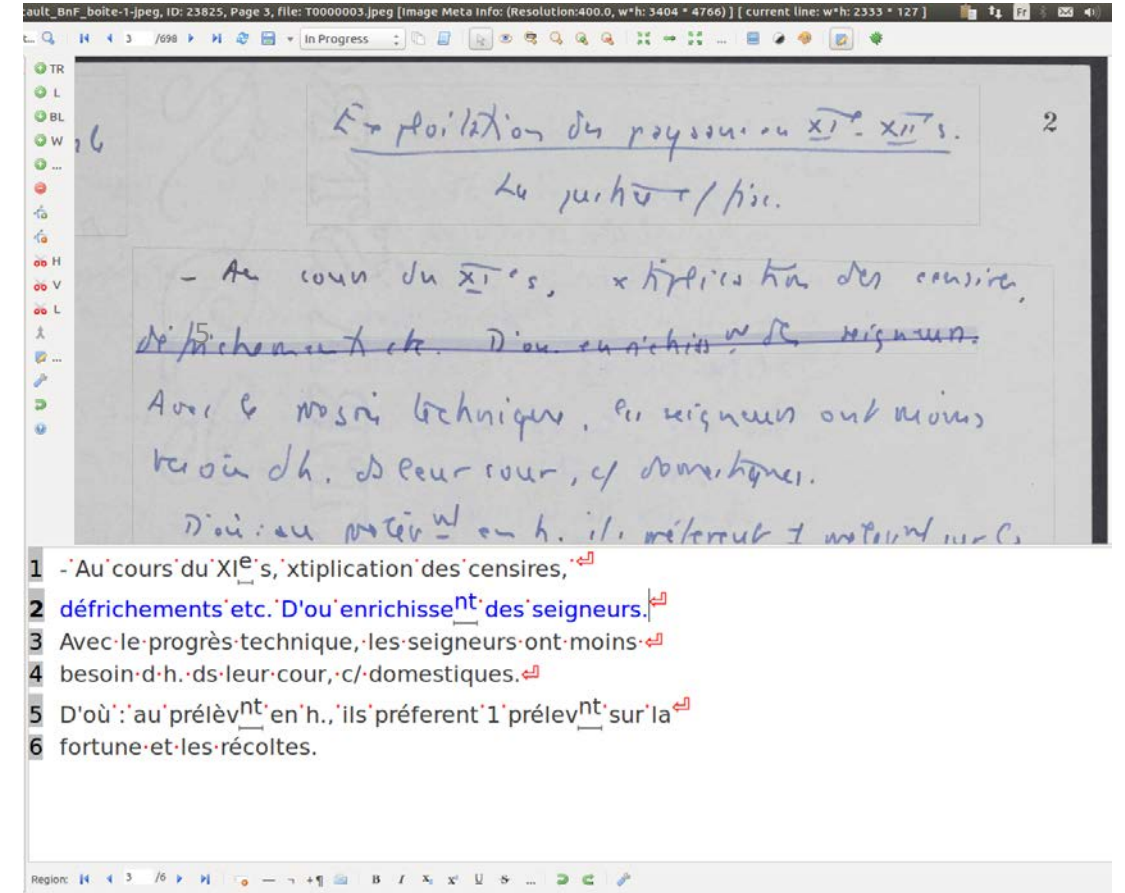
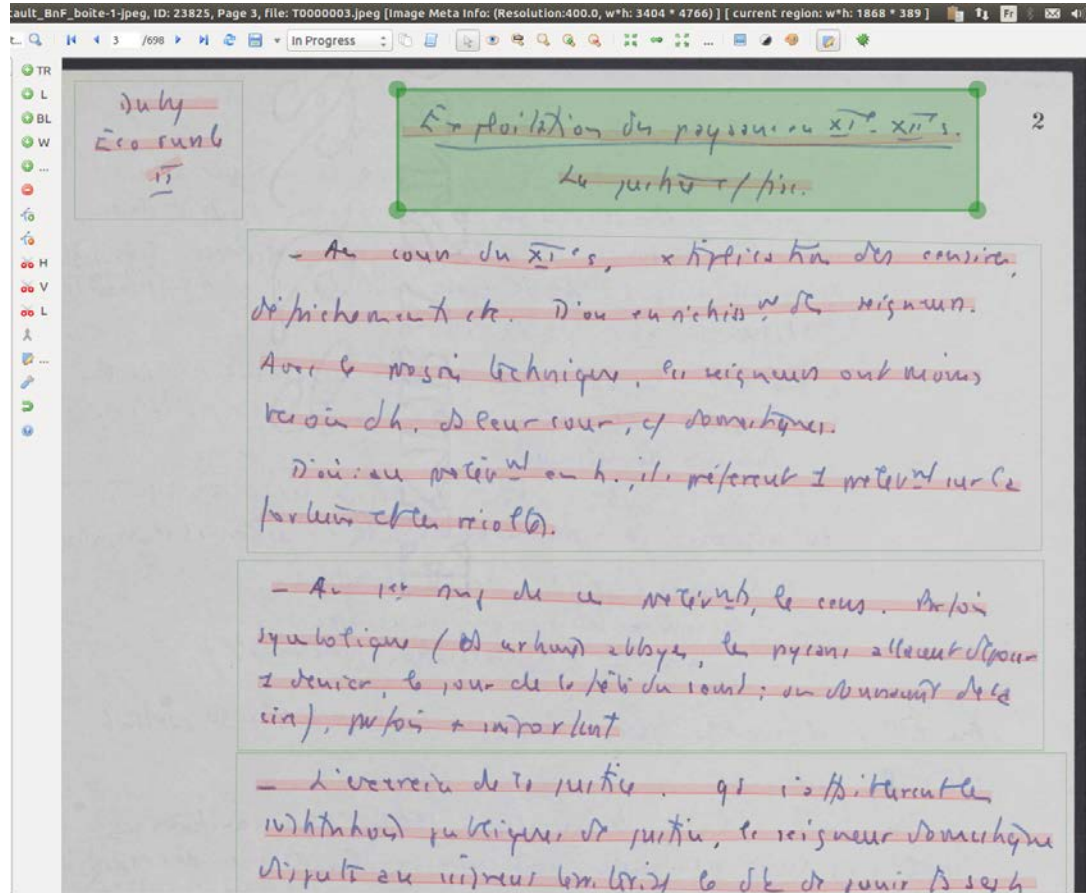
1429

1430

<

Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

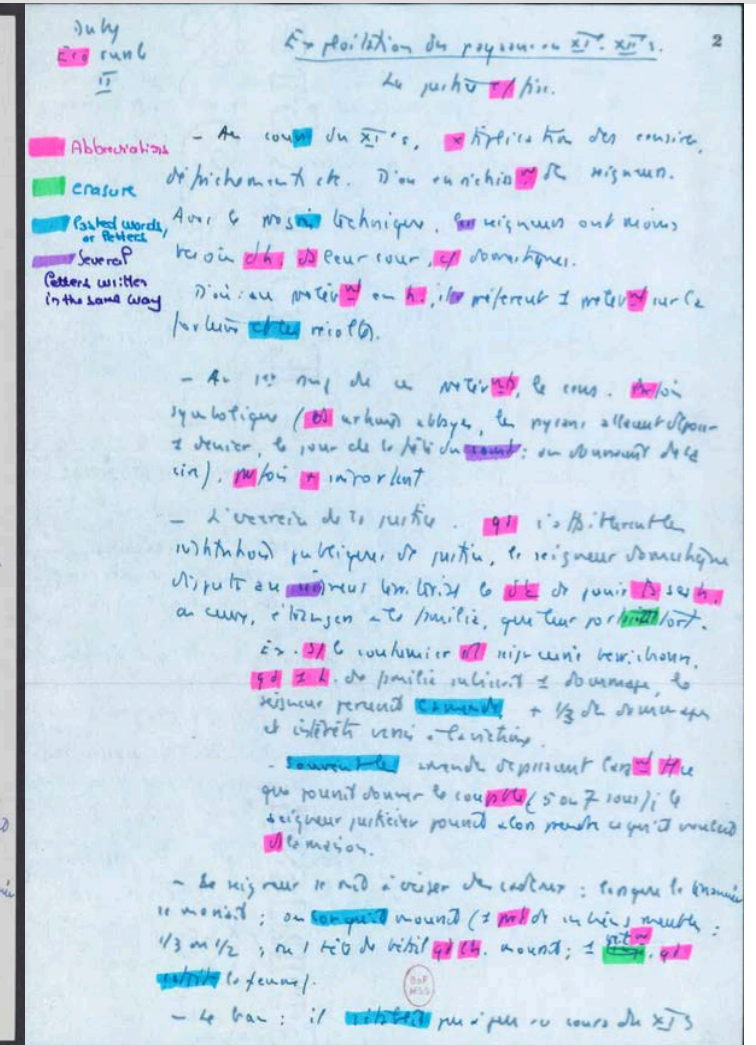
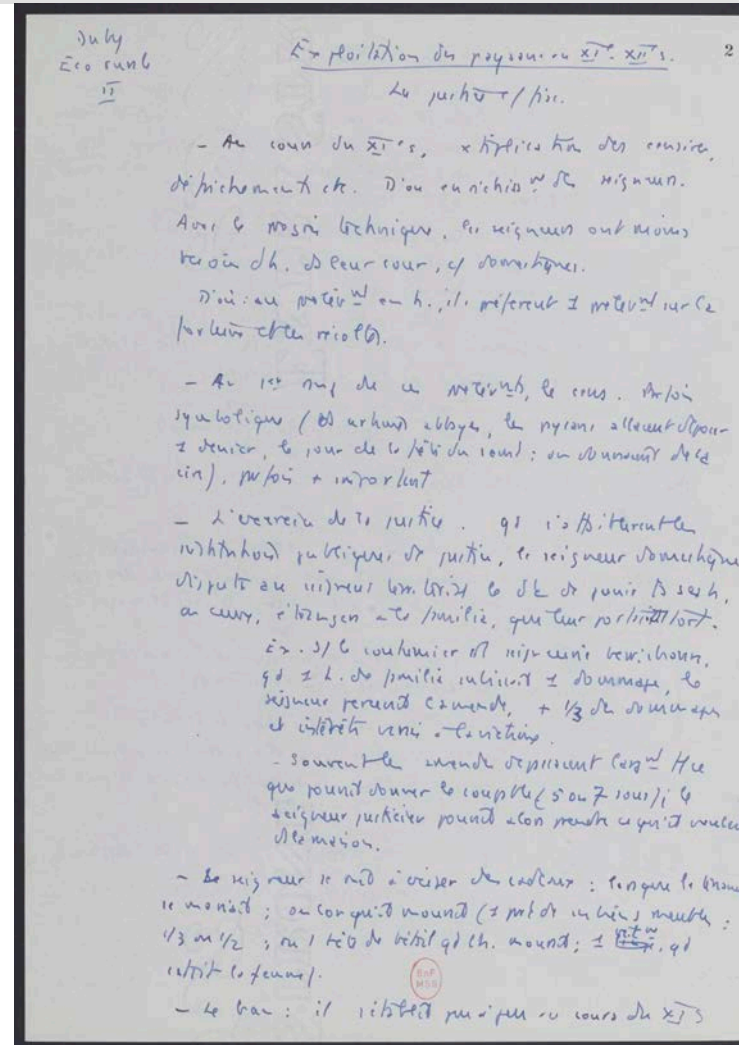
Créer des données d'apprentissage



Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Difficultés rencontrées

- L'écriture de Foucault : notes de lecture avec des **abréviations** nombreuses et ambiguës, **plusieurs lettres écrites de la même façon**, **rautes**, fautes d'orthographe, **mots collés**, etc.
- Les mots illisibles ont été identifiés avec un codage TEI
- Prise en main du logiciel et temps moyen par carte
- Ergonomie du logicielle
- Transcriptions à plusieurs mains



Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Transcriptions automatiques avec le modèle HTR

Loaded doc: boîte51_diff, ID: 35689, Page 51, file: T0000424.jpeg [Image Meta Info: (Resolution:400.0, w:h: 3496 * 4918)] [current line: w:h: 2394 * 252]

Search current document...

TR
L
BL
W
...

162

166

Faute des parents qui punissent les enfants

Pon qu'ils se sont blessés ds des jeux, si bien que

les enfants cochent leurs blessures, qui denenrent

+ grave.

167

167

"L'on ne doit presque jamais battre les

enfants, car, l'ans compter que c est les avilir et

le ravater au rang des malheureux, et qu on

leur inspire des sentiments bas, ainpans,

suvt le mensonge et pa. des vices encore+

gdts, c est qu'il est très ivident que cela est

5 des enfants

6 1762.

7 - Faute des parents qui punissent les enfants

8 166

9 lorsqu'ils se sont blessés ds des jeux, si bien que

10 les enfants cochent leurs blessures, qui denenrent

11 + grave.

12 167

13 - "L'on ne doit presque jamais battre les

14 enfants, car, l'ans compter que c est les avilir et

15 le ravater au rang des malheureux, et qu on

16 leur inspire des sentiments bas, ainpans,

17 suvt le mensonge et pa. des vices encore+

18 gdts, c est qu'il est très ivident que cela est

Loaded doc: boîte51_diff, ID: 35689, Page 45, file: T0000418.jpeg [Image Meta Info: (Resolution:400.0, w:h: 3524 * 4936)] [current line: w:h: 2299 * 224]

Search current document...

TR
L
BL
W
...

1776

122

Le meilleur alement pr l'enfant

de l'est de la mère

1776

122

11 L'aliment qui convient le mieue aux enfants

en ga jusqu a ce que leurs forces soient accrues

eu x de pouvoir digérer des aliments solides est

icontestibint le lait et surtt celui de la mère

si elle est saine Le lait des mammelles de la

13 mère, qui n est qu'une gradation du lait utérain

14 123

15 est la nourriture la + analogue la constitution

16 parhiuleure de l'infant. La nature, qui proportionne

17 ttrs les ressources au besom ... donne à chaque

1 anit

2 Essai sur la sant et

3 Le meilleur alement pr l'enfant

4 l'éducatim medat des filles

5 le lait de la mère

6 1776

7 122

8 11 L'aliment qui convient le mieue aux enfants

9 en ga jusqu a ce que leurs forces soient accrues

10 eu x de pouvoir digérer des aliments solides est

11 icontestibint le lait et surtt celui de la mère

12 si elle est saine Le lait des mammelles de la

13 mère, qui n est qu'une gradation du lait utérain

14 123

15 est la nourriture la + analogue la constitution

16 parhiuleure de l'infant. La nature, qui proportionne

17 ttrs les ressources au besom ... donne à chaque

Expérimenter **Transkribus** sur les fiches de lecture de Michel Foucault

Transcriptions automatiques avec le modèle HTR

Transkribus v1.11.0 (02_06_2020_17:00): loaded doc: Tests BNF, ID: 484199, Page 1, file: B1_T0000003.jpeg [Image Meta Info: (Resolution:400.0, w*h: 3404 * 4766)] [current line: w*h: 1785 * 238]
Search current document...

Server Overview Layout Metadata Tools

Logout marie-laure.massot@ens.fr

Document... Find

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections: Tests_BNF (76435, Owner) Col-ID

Documents HTR Model Data

ID	Title	Pages	Uploader
484...	Tests BNF	4	marie-laure

50 Filter

The image shows a handwritten document with the following text:

Duby
Eco rurale
II

Exploitation des paysans au XIe-XIIe s.

La justice c/ fisc.

- Au cours du XIe s, xtiplication des censires, défrichements etc. D'ou enrichissnt des seigneurs.

Avec le progrès technique, les seigneurs ont moins besoin d h. ds leur cour, c/ domestiques.

D'où: au prélèvt en h., ils préfèrent 1 prélevnt sur la fortune et les récoltes.

- Au 1er rang de ces prélèvnrs, le cens. Parfois

La fonction de repérage de mots-clés (Keyword Spotting)

Documents | Fulltext (Solr) | Tags | **KWS**

Search for...

Search in: Current document ☐ Partial Matches ☒ Case-sensitivity ☐ Expert Syntax

Confidence Threshold: 0.05

Queries

Keyword 1: justice

Keyword 2: seigneur

Keyword 3:

Keyword Spotting Results

Confidence	Page Nr.	Line transcription	Preview
0.8061	677	D où des conflits entre le juuticier et le seigneur	
0.8052	618	Cetle justierrofitais au seigneur : d où le fait qd il	
0.7784	610	à son seigneur se developpèrent lavent tout en	
0.7641	605	fait simplint appe lacu seigneur territoriasl pr les cs	
0.7583	638	de fiice le seigneur le + puissant retenant les cas qui li	
0.7528	457	. Le procureur du roi cs ou du ceigneur est tprs partie	
0.7478	671	arre du pouvoir central) : mais par le seigneur	
0.7476	439	justicier n est fonde dêtre seigneur féodae de tt ce qui	
0.7410	389	A bolition du "hériot" (dt pr les seigneus de semaparer	
0.7405	629	de clase. reste que te justice est possessiond1seigneur	
0.7404	683	le seigneur dt depdait le condamne	
0.7339	679	oes Normandié tte frande sur les dts du seigneur.	
0.7304	619	d1 seigneur ; mais de personnalités diversn. Aisi en 1074	

Search Results

Created Status

21.09.20 21:18:27 Com

03.04.19 13:33:08 Com

03.04.19 13:30:32 Com

Preview

- Nouvel outil puissant (calcul intensif) pour **rechercher des mots distincts** dans une collection de documents ;
- **Lancer un modèle HTR** pour produire une transcription, puis rechercher immédiatement dans le document ;
- Même si la transcription générée contient des erreurs, **KWS trouvera de manière fiable des mots, phrases, etc.** dans les documents ;
- Le programme indique le nombre d'occurrences trouvées, les pages et les images et évalue la fiabilité des résultats ;
- Dans notre ex., le programme retrouve 259 fois « justice » et 154 fois « seigneur ». On remarque que même si la transcription est fautive « Ceigneur » (p. 457), le KWS a bien identifié l'image du mot.

Résultats et perspectives

- Entraînement du modèle HTR sur **600 images** transcrites **manuellement**
- **92% des caractères reconnus** sans erreur
- Les transcriptions automatiques peuvent être **corrigées** et de **nouveaux modèles peuvent être générés** pour améliorer encore les résultats
- Aujourd'hui avec HTR + = Taux d'erreur de 5% (CER)
- Les **résultats des tests** du logiciel sont très **encourageants** : Certes, il reste des erreurs... mais avec un taux d'erreur de 5% au caractère, les fiches deviennent **lisibles** et **indexables** par un moteur d'indexation*.
- Produire des **transcriptions automatiques pour toutes les images** et **indexation** du corpus
- Utiliser la plateforme Ffl-Eman (Omeka) pour mettre en place un **système de correction collaborative** ou l'interface Web proposée par Transkribus
- **Keywords spotting (KWS)** cad recherche approximative sur l'image : un véritable + même dans le cas de transcription fautive (voir vidéo)
- **Flux de travail envisagé** : corpus numérisé /image -> transkribus (modèle HTR) / texte -> FFL prototype / annotation (enrichissement et espace de travail privé) -> Plateforme publique Ffl-Eman (Consultation des fiches, recherche de termes et correction des transcriptions)

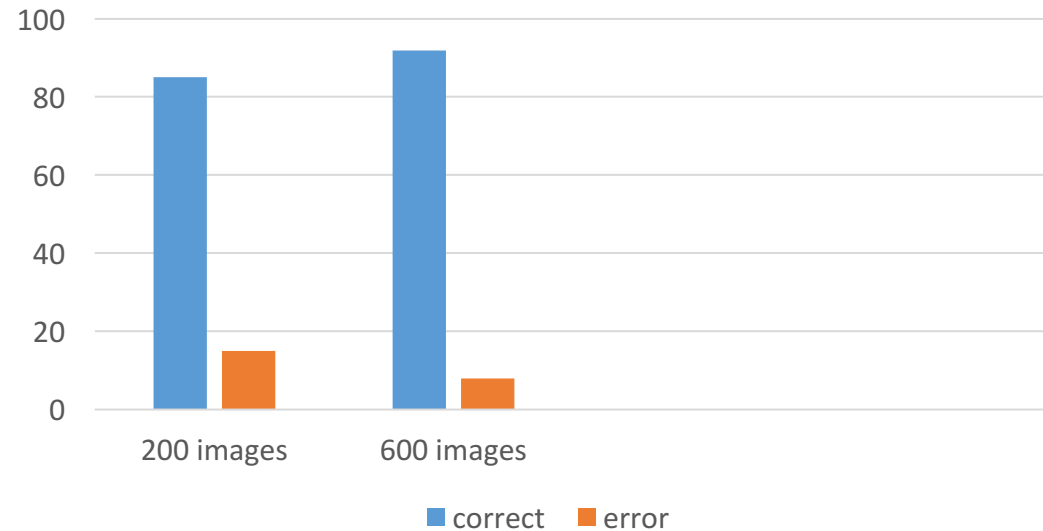
*L'**indexation automatique de documents** est un domaine de l'informatique et des sciences de l'information et des bibliothèques qui utilise des méthodes logicielles pour organiser un ensemble de documents et faciliter ultérieurement la recherche de contenu dans cette collection.



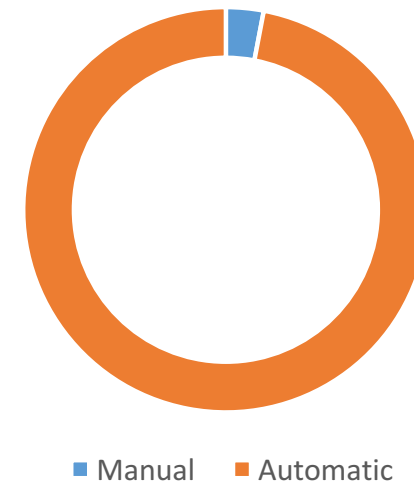
Une base de données textuelles qui peut être vérifiée, documentée, enrichie et explorée.

Conclusion

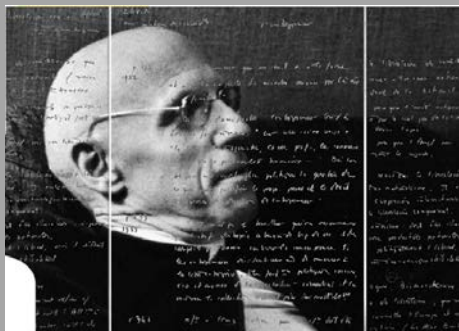
Average success rate of the tests



Transcriptions



La collaboration avec le projet européen Transkribus/READ a donné une nouvelle dimension au projet FFL !



Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Compte-rendu, article et communications dans le cadre du projet Ffl

Marie-Laure Massot, Arianna Sforzini, Vincent Ventresque.

[*Transcrire les fiches de lecture de Michel Foucault avec le logiciel Transkribus : compte rendu des tests.*](#) Juin 2018 : Compte-rendu des tests consultable dans HAL.

[*Transcribing Foucault's handwriting with Transkribus – Report*](#)

Nov. 2018 : Communication au deuxième congrès des utilisateurs de Transkribus (Vienna), Autriche.

[*Transcribing Foucault's handwriting with Transkribus. Jdmdh:5043 – Journal of Data Mining & Digital Humanities, March 5, 2019, Atelier Digit_Hum.*](#)

Mars 2019 : Article en anglais publié dans jdmdh.

[*Transcrire automatiquement les fiches de lecture de Michel Foucault avec Transkribus : retour d'expérience.*](#)

Avr. 2019 : Communication à l'atelier international de travail EMAN-latext (Las Palmas Canaries), Espagne.

« Lire les fiches de lecture avec Transkribus »

Sept. 2020 : Communication au colloque de clôture du projet ANR FFL – 1ère Partie Fiches et fichiers. L'archive Foucault à l'ère du numérique (IMEC), Abbaye d'Ardenne.



Ressources

Contacter l'équipe :

TRANSKRIBUS Team at University of Innsbruck

email@transkribus.eu

S'enregistrer sur la plateforme et télécharger le logiciel Transkribus

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

Guides d'utilisation et tutoriels sur Transkribus Wiki

https://transkribus.eu/wiki/index.php/Main_Page

https://transkribus.eu/wiki/images/7/77/How_to_use_TRANSKRIBUS_-_10_steps.pdf

https://transkribus.eu/wiki/images/1/1d/Comment_utiliser_Transkribus_%E2%80%93_10_%C3%A9tapes_ou_moins_with_Screenshots.pdf

Vous pouvez aussi consulter le [tutoriel en français établi par Emmanuelle Perrin](#) (Archéorient, Université de Lyon), réalisé en novembre 2019 : https://hal.archives-ouvertes.fr/hal-02472234/file/Tutoriel_Transkribus_BBM.pdf

The screenshot shows the Transkribus Wiki Main Page. The browser address bar displays https://transkribus.eu/wiki/index.php/Main_Page. The page features a sidebar with navigation links such as 'Main page', 'Recent changes', 'Random page', and 'Help'. The main content area includes a 'Main Page' section with a description of Transkribus as a platform for automated recognition, transcription, and searching of historical documents. It lists key features: an expert tool, a webinterface, and several cloud services. Below this, there is a section titled 'Get an overview of all the tools we are building in the Transkribus interfaces map!'. The page also mentions the main objective of Transkribus, which is to support users in transcribing printed or handwritten documents. A list of tools offered is provided, including Handwritten Text Recognition (HTR), Layout Analysis, Document Understanding, Writer Identification, and Optical Character Recognition (OCR) using ABBYY Finereader Engine 11. The page is funded by the European Commission as part of the READ project. A 'Registration' section is also visible, followed by 'Download and Installation', 'How to Guides and Videos', 'Handwritten Text Recognition Workflow', 'Handwritten Text Recognition - Example Documents', 'Transkribus Cloud Services', 'Questions and Answers', and 'Credits'. The credits section states that Transkribus is hosted by the Digitisation and Digital Preservation group (DEA) at the University of Innsbruck, funded by the European Commission as part of the H2020 Project READ (2016-2019). It also mentions previous funding by the FP7 Project tranScriptorium (2013-2015) and notes that Transkribus is still under development with most parts being Open Source. The page footer includes a note about the last modification date (11 December 2018) and a Creative Commons license.

Expérimenter **Transkribus** sur les fiches de lecture de Michel Foucault

Démonstration

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

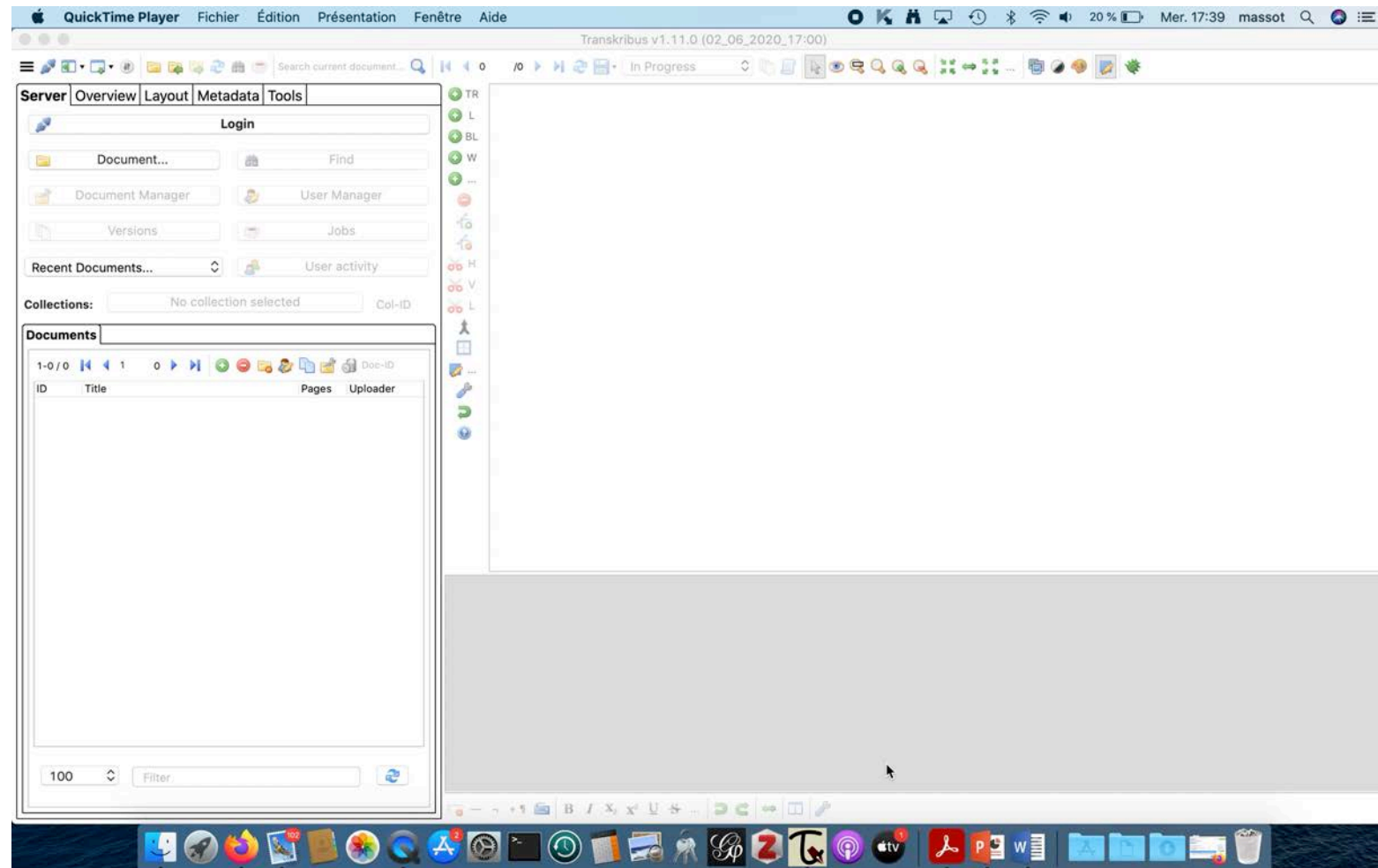
Are you a computer scientist and working in the fields of computer vision, document analysis, pattern recognition, natural language processing or a related field? You are seeking interesting documents from 1000 years of handwriting, printing and publishing? And you would really enjoy to get some reference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

[View details »](#)



Expérimenter **Transkribus** sur les fiches de lecture de Michel Foucault

Démonstration : transcription automatique avec le dernier modèle HTR



http://ffl-public.huma-num.fr/doc/Transkribus_automatique.mov

Expérimenter Transkribus sur les fiches de lecture de Michel Foucault

Démonstration : recherche approximative sur l'image (KWS)

The screenshot displays the Transkribus v1.11.0 web interface within a QuickTime Player window. The interface is divided into several sections:

- Server:** Includes a login field for 'marie-laure.massot@ens.fr' and buttons for 'Document...', 'Find', 'Document Manager', 'User Manager', 'Versions', 'Jobs', 'Recent Documents...', and 'User activity'.
- Collections:** Shows 'Foucault_collection_principale (4016, Ow)'.
- Documents:** A table listing documents with columns for ID, Title, Pages, and Uploader. The document '180... b38-test' is highlighted.
- Document View:** Displays a scanned image of a handwritten document. The text is: 'L'heure c/ication avec Satan. 10', 'La c/ication de P'h avec les anges est "dérivée de la c/auté de P'h que l'on remarque en ces 2 natures : Au contraire " l'association de P'h cherie du Paradis avec l'Arge banni du ciel est fondée en la condition de leur c/n accidents'.
- Transcription:** A list of transcribed text segments with line numbers (1-10) and a search bar.

The bottom of the screen shows a macOS dock with various application icons.

http://ffl-public.huma-num.fr/doc/Transkribus_KWS.mov

ANR

59

Ffl

605

FOUCAULT

FICHES DE

LECTURE

FOUCAULT'S

READING

NOTES

manière et m'indiquent d'une façon

de l'être, un secret m'égale

l'innocence d'un être qui se montre "les fautes"

rompre m'indiquent d'une façon

Accueil

Présentation du projet

Activités

Partenaires

Recherche

Éditer Michel Foucault (1994-2021)

Publié le 14/09/2020 par Marie-Laure Massot

Colloque de clôture du projet ANR Foucault Fiches de Lecture – Seconde partie à la BnF, site François-Mitterrand, Paris, 26 septembre 2020


Entre 1994 et 2021, les textes du philosophe Michel Foucault ont fait l'objet d'une édition posthume. Un colloque dont la seconde partie a lieu à la BnF revient sur les questions qui se sont posées lors de cette entreprise de longue haleine.

Avec Philippe Chevallier, Laurent Dartigues, Daniel Defert, Claude-Olivier Doron, François Ewald, Henri-Paul Fruchaud, Frédéric Gros, Laurence Le Bras, Isabelle Le Masne de Chermont, Marie-Laure Massot, Jean-Philippe Moreux, Judith Revel, Michel Senellart, Vincent Ventresque, Richard Walter.

Retrouvez le colloque sur le site de la BnF : <https://www.bnf.fr/fr/agenda/editer-michel-foucault-1994-2021>

Entrée gratuite – Réservation recommandée via l'application Affluences ou sur [affluences.com](https://www.bnf.fr/fr/agenda/editer-michel-foucault-1994-2021) (rubrique Bibliothèques) – Réservation obligatoire avant le 18 septembre pour la matinée.

Il est recommandé de se présenter en avance (jusqu'à 20 minutes avant la manifestation).



ÉDITER MICHEL FOUCAULT (1994-2021)

26 sep. 2020

14h - 18h

François Mitterrand

Grand auditorium

Publié dans Billets | Marqué avec ANR, Archives, BNF, Colloques, Édition numérique, Humanités numériques, Michel Foucault

Modifier

Recherche

PRÉSENTATION

Carnet de recherche dédié au projet Foucault Fiches de Lecture

Research notebook dedicated to the Foucault project Reading Notes

CATÉGORIES

Billets (17)

ARCHIVES

septembre 2020 (1)

juin 2020 (1)

février 2020 (1)

novembre 2019 (1)

mai 2019 (1)

avril 2019 (1)

mars 2019 (1)

janvier 2019 (1)

décembre 2018 (3)

octobre 2018 (1)

août 2018 (1)

juin 2018 (1)

mai 2018 (2)

avril 2018 (1)

ÉTIQUETTES

ANR Archives Atelier

Auteurs BNF Colloques

Corpus Développements

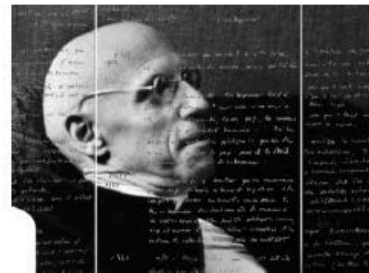
Édition numérique EMAN

Carnet de recherche Ffl << <https://ffl.hypotheses.org/>>>



EXPERIMENTER TRANSKRIBUS

SUR LES FICHES DE LECTURE DE MICHEL FOUCAULT



Transkribus est une plateforme complète pour la reconnaissance, la transcription et la recherche automatisées de documents historiques manuscrits ou imprimés. Elle facilite la lecture, l'enrichissement et l'exploration des textes. Le logiciel est développé par le [projet européen READ](#) qui souhaiterait révolutionner l'accès aux documents d'archives en développant des outils numériques innovants ([Transkribus](#), [ScanTent](#), [read@search](#)).

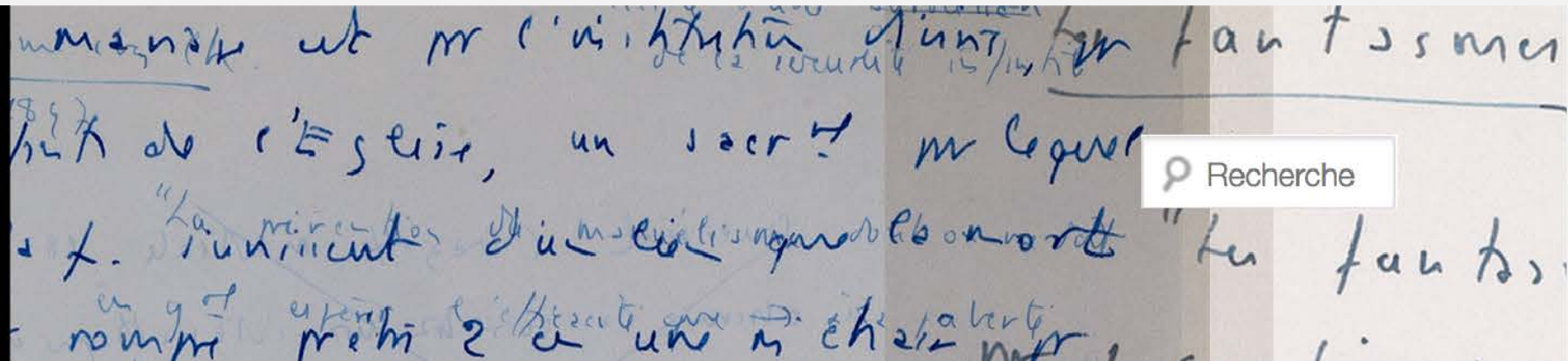
Expérimentez Transkribus en téléchargeant le scénario test :

https://f-origin.hypotheses.org/wp-content/blogs.dir/4503/files/2020/09/tester_transkribus-1.pdf

<https://ffl.hypotheses.org/>



FOUCAULT
FICHES DE
LECTURE
FOUCAULT'S
READING
NOTES



MERCI POUR VOTRE ATTENTION ET N'HÉSITEZ PAS À EXPÉRIMENTER TRANSKRIBUS SUR NOTRE CORPUS TEST !