
Journaux intimes franco-phones en Russie au XIX^e siècle : une approche textométrique de l'interaction entre langues et cultures

Alexei Lavrentiev

alexei.lavrentev@ens-lyon.fr

IRHIM, CNRS & ENS de Lyon, France

Michèle Debrenne

micheledebrenne@gmail.com

Université d'État de Novossibirsk, Russie

Dmitry Dolgushin

d_dolgushin@mail.ru

Université d'État de Novossibirsk, Russie

Nina Panina

pa.nina@mail.ru

Université d'État de Novossibirsk, Russie

Andrey Borodiknin

borodichin@spsl.nsc.ru

Branche Sibérienne de l'Académie des sciences, Russe

Serge Heiden

slh@ens-lyon.fr

IRHIM, CNRS & ENS de Lyon, France

La rédaction de journaux intimes en français était un phénomène assez répandu parmi les élites russes au XIX^e siècle. Les aristocrates de l'époque apprenaient souvent le français dès le plus jeune âge et continuaient à pratiquer cette langue tout au long de leur vie. Le français n'était cependant pas leur langue maternelle, et leurs écrits offrent des données précieuses sur l'interaction entre les deux langues et cultures. Par ailleurs, certains journaux présentent de grandes qualités littéraires et artistiques (grâce aux dessins d'auteur qui les accompagnent) et méritent d'être plus largement connus par la communauté académique et le grand public.

Le projet présenté dans ce poster est porté par une équipe franco-russe de l'Université d'état de Novossibirsk et de l'École normale supérieure de Lyon. Il vise à analyser et à publier numériquement plusieurs corpus, dont les journaux d'Alexandre Tchitchérine et d'Olga Orlova-Davydova.

A. Tchitchérine, lieutenant au régiment Semionovski de l'armée russe, a tenu son journal de septembre 1812 à août 1813, son manuscrit contient 83 illustrations, dont la plupart en couleurs. Son texte a été traduit en russe et publié (Chicherin 1966), mais cette traduction dont les qualités littéraires sont indéniables, est parfois inexacte et présente des lacunes. L'original français de ce journal est inédit.

Les journaux d'O. Orlova-Davydova (1814-1876) sont le fruit d'un vaste projet autobiographique que l'auteure a mené tout au long de sa vie. Ces journaux n'ont jamais fait l'objet d'une recherche scientifique. Dans le cadre du présent projet les passages franco-phones de ses journaux (1830-1847) servent de corpus pour étudier le bilinguisme de l'auteure et pour déceler les voies de formation de personnalités bi- ou

plurilingues dans la Russie du XIX^e siècle (Debrenne 2016a et 2016b). Le noyau du corpus est constitué par une copie des journaux d'Orlova-Davydova réalisée vraisemblablement à sa demande et corrigée par sa propre main. Ce document est conservé à Novossibirsk, à la Bibliothèque scientifique et technique de la Branche Sibérienne de l'Académie des sciences russe (BST). Il fait partie du fonds patrimonial Tikhomirov (http://www.spsl.nsc.ru/rbook/ogl_tix.html).

Les journaux du corpus sont transcrits et « encodés par stylage » dans un logiciel de traitement de texte (Microsoft Word) de façon à permettre une conversion quasi-automatique dans un format XML-TEI enrichi d'un certain nombre d'annotations. Il s'agit de repérer et de classer les différents types d'erreurs par rapport au français standard de l'époque (orthographiques, grammaticales ou lexicales), de recenser les différents types d'entités nommées présents (index hiérarchique), ainsi que les différents passages rédigés en russe. La conversion se fait en deux étapes : d'abord à l'aide de l'outil de conversion en ligne Odette (Glorieux 2015), puis avec une feuille de transformation XSLT spécifique au projet. Après validation formelle, la forme finale des sources est importée dans la plateforme TXM (Heiden et al. 2010). TXM permet d'analyser et de publier des corpus complexes encodés en XML-TEI en combinant des outils de traitement automatique de la langue (TAL :tokenisation paramétrable, étiquetage morphosyntaxique avec TreeTagger), de philologie numérique (éditions présentant de façon

synoptique le fac-simile de la source et plusieurs niveaux de transcription), et d'analyse qualitative (recherche de motifs lexicaux avec le moteur plein texte CQP) et quantitative (analyse de spécificités, analyse factorielle de correspondances, etc.). Le projet utilisera les outils de TXM pour lire de façon intégrée l'édition du texte avec les illustrations correspondantes. Il permettra également de synthétiser le vocabulaire utilisé juste devant les passages en russe et étudier les emprunts français propres à chaque auteur par rapport à la langue russe. Les illustrations ci-dessous présentent un exemple d'encodage XML-TEI dans les sources d'un dessin situé au début d'une page du journal de Tchitchérine (Fig. 1) suivi de la copie d'écran d'une visualisation synoptique correspondante dans TXM de la transcription comprenant le dessin et le fac-similé du manuscrit (Fig. 2).

de la BST (bibliothèque numérique du fonds Tikhomirov). Des exemples de réalisations similaires peuvent être consultés sur le [portail de démonstration](#).

```
<p> [...] ¶
...<pb:n="20"/>...
...<figure>...
...<graphic:url="img/tchitcherine/Tchitcherine_p20_dessin1.jpg"/>...
...</figure>...
instants me semblent précieux. Peut-être macusera-ton d'inconstance mon journal
est moins rempli que l'autre moins détaillé...</p>¶
```

Fig. 1. Encodage XML-TEI d'une illustration insérée dans le corps du texte

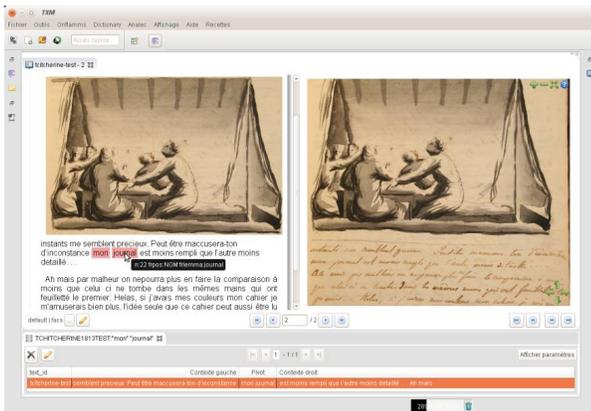


Fig. 2. Affichage synoptique d'une transcription et d'un fac-similé d'un texte doté d'une illustration sous TXM en mode « retour au texte » depuis une concordance (de l'expression « mon journal » dont l'occurrence est mise en évidence dans la transcription)

Remerciements

Ce travail a bénéficié du soutien la Fondation scientifique de Russie pour les sciences humaines (RGNF) [16-24-08001] et de la Fondation Maisons des Sciences de l'Homme (FMSH).

Note

Tous les corpus édités dans le cadre du projet seront publiés en ligne via un portail TXM et sur le site