HAL
open science

IntraTexTT (Intralingual Textual Transformation Tool) Development of a Digital Text Comparison Tool for the Analysis of Rewriting and Intralingual Translation Processes in Pre-Modern Chinese Texts

Rainier Lanselle

To cite this version:

Rainier Lanselle. IntraTexTT (Intralingual Textual Transformation Tool) Development of a Digital Text Comparison Tool for the Analysis of Rewriting and Intralingual Translation Processes in Pre-Modern Chinese Texts. 2018. hal-01863346

HAL Id: hal-01863346

https://hal.science/hal-01863346v1

Preprint submitted on 28 Aug 2018

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.

IntraTexTT

Intralingual Textual Transformation Tool

Elaboration d'un outil numérique de comparaison textuelle pour l'analyse des procédés de réécriture et de traduction intralinguale dans les textes de la Chine prémoderne
(Document de travail)

Development of a Digital Text Comparison Tool for the Analysis of Rewriting and Intralingual Translation Processes in Pre-Modern Chinese Texts
(Working Paper)

Rainier Lanselle

EPHE, PSL, Section des sciences historiques et philologiques, Paris, France

CRCAO Centre de recherches sur les civilisations de l'Asie orientale (UMR 8155, CNRS)

Paris, France

Table des matières — Table of Contents

FRANÇAIS *3*

Résumé 3

Principes généraux 3

Projet de construction d'un outil d'analyse des transformations textuelles dans le contexte intralingual 5

Typologie des transformations textuelles — Micro-stratégies et caractéristiques stylistiques 6

Indications techniques pour l'élaboration de l'outil numérique 10

Conclusion 10

ENGLISH *11*

Abstract 11

General Outline 11

The Current Project: Building a Digital Tool for the Analysis of Intralingual Textual Transformation 13

Typology of textual transformations - Micro-strategies and stylistic characteristics 14

Technical information for the development of the digital tool 17

Conclusion 18

Références bibliographiques citées Bibliographical References *19*

Tableau des catégories et sous-catégories Chart of Categories and Subcategories *21*

FRANÇAIS

Résumé

Ce document de travail (juillet-août 2018) présente le projet en cours d'élaboration d'un outil d'humanités numériques nommé ***IntraTexTT*** *(pour « **Intralingual Textual Transformation Tool** »). Cet outil est destiné à l'analyse des transformations textuelles entre textes sources et textes cibles dans le contexte de la production de récits littéraires en Chine prémoderne (ca. XIIIeme-XVIIIeme s.). Cette production recourait de façon régulière à des processus complexe d'intertextualité, avec recours à des opérations de réécriture et de traduction intralinguale, motivées en particulier en raison du contexte de diglossie de la langue et de l'écriture chinoises, entre langue classique et langue vernaculaire. L'outil* ***IntraTexTT*** *sera destiné à l'analyse, au relevé statistique, à la mise en ligne et au partage des données relatives à ces opérations de production textuelle par transformation intertextuelle. Le présent document de travail expose les motivations et le contexte du besoin d'élaboration d'un tel outil. Il en donne les principes généraux, et en expose de façon détaillée les typologies ; il décrit les stratégies et micro-stratégies textuelles, avec leurs dénominations précises, à prendre en compte pour la réalisation d'un outil numérique pertinent et performant. Il indique que cet outil est actuellement développé dans le cadre du CRCAO, UMR 8155, entre l'auteur et un ingénieur d'études CNRS chargé de l'édition de corpus numériques, rattaché au laboratoire CRCAO. Le résultat devrait être opérationnel au cours de l'année 2019.*

Principes généraux

Mes recherches et mon enseignement sur la littérature prémoderne en Chine (époques Yuan à première moitié des Qing, XIIIeme-XVIIIeme siècles) m'amènent régulièrement à décrire et à analyser des cas de réécriture et de traduction intralinguale. Dans la littérature de récit de cette époque, sous toutes ses formes, la stratégie de création littéraire des auteurs recourt massivement à des emprunts, reprises et retraitements de thèmes antérieurs qu'ils recyclent, réécrivent, amplifient, et très souvent réorientent idéologiquement en fonction de leurs objectifs.

Je traite principalement ici de la littérature de récit en langue vernaculaire. Dans ce vaste domaine, qui recouvre principalement les formes du conte, du roman et du théâtre, mais aussi des formes de poésies narratives ou chantées, les auteurs de la période considérée ont constamment tenu compte des caractéristiques du paysage linguistique au sein duquel ils évoluaient : à savoir un environnement marqué par la diglossie, ou plus généralement par des formes de « schizoglossie [1] » entre langue(s) classique(s) *(wenyan* 文言*)* et langue(s)

[1] Pour reprendre en la modifiant une appellation suggérée par King 2015, p. 13, n. 23, parlant de « schizoglossographia ». Cf. également les propos de Ross King dans son intervention du 14 mai 2018, "Ditching 'Diglossia': Describing Ecologies of the Spoken and Inscribed in Pre-modern Korea", au séminaire mensuel

vernaculaire(s) *(tongsu* 通俗*)* — cette dernière catégorie pouvant comprendre aussi bien des aspects standards que des aspects dialectaux. On assiste alors à des phénomènes massifs de *réécriture intralinguale*, voire de *traduction intralinguale*[2], qui vont presque toujours dans le même sens, soit de la langue littéraire classique vers la langue vernaculaire. Les auteurs qui pratiquent ce genre de réécriture ou traduction intralinguales montrent toujours une claire conscience de la situation schizoglossique de la langue chinoise de leur époque dans ses formes écrites. Ils ne cessent, comme cela a été souvent décrit dans la littérature savante dès les premières décennies du siècles dernier (Lu Xun 魯迅, Sun Kaidi 孫楷第, Tan Zhengbi 譚正璧…) de faire référence à l'art des conteurs, qu'ils imitent sciemment dans leurs œuvres, où ils font intervenir de façon habituelle des éléments d'oralité lorsqu'ils expliquent, paraphrasent et re-racontent des récits antérieure sous une forme nouvelle. Le fait de rendre un thème accessible au plus grand nombre constitue un argument constant cité comme justifiant leur action de réécriture/ traduction intralinguale. Les auteurs s'expliquent souvent de façon explicite à ce sujet. Les textes critiques, dès l'époque des Song, et jusqu'aux Ming-Qing, en particulier les préfaces et avant-propos des œuvres, font fréquemment référence à la situation schizoglossique de leur époque. En particulier ils reviennent régulièrement sur la difficulté de la langue classique et sur les avantages prêtés à la réécriture en langue vulgaire : cette dernière est vue comme possédant une plus grande capacité de pénétration vis-à-vis du grand public, et donc des effets en termes d'éducation ou de transmission des connaissances et des valeurs (en particulier morales)[3].

Ces phénomènes font aujourd'hui l'objet d'un intérêt renouvelé, notamment en lien avec une volonté d'approfondir notre compréhension de l'émergence des expressions vernaculaires, non seulement dans la Chine prémoderne, mais dans l'ensemble de l'Asie orientale de l'époque prémoderne/moderne (XIIIeme-XIXeme siècles). C'est dans ce cadre que je conduis aujourd'hui un projet intitulé « Intralingual Translation, Language Shifting, and the Rise of Vernaculars in East Asian Classical and Premodern Cultures », avec Barbara Bisetto, maître de conférences à l'Universita Degli Studi di Verona[4], avec qui je dirige un programme « Diglossie, traduction intralinguale, interprétation » dans le cadre du CRCAO (Centre de recherches sur les civilisations de l'Asie orientale, UMR 8155, CNRS), pour le contrat pluriannuel 2019-2023 de l'UMR.

Dans ce cadre, nous prévoyons des collaborations avec des spécialistes des textes chinois considérés du point de vue de leurs rapports aux vernaculaires, qu'il s'agisse des vernaculaires de Chine ou des pays environnants. Pour ma part, et cela vaut pour tout ce qui est traité ci-après, je m'intéresse exclusivement aux phénomènes intralinguaux, c'est-à-dire à ceux qui ont

---

(organisé par Andréas Stauder et moi-même) EPHE Section des Sciences historiques et philologiques & SCRIPTA-PSL History and Practices of Writing : *Diglossie, traduction intralinguale, réécriture, commentaire — Diglossia, intralingual translation, rewriting, and commentary*. La notion de « diglossie », telle que définie d'abord par Ferguson 1959, est aujourd'hui contestée parmi les spécialistes de l'Asie orientale, qui jugent sa distribution des rôles entre langues « hautes » et « basses » (H/L) peu pertinente, voire très problématique, dans le cas des langues et pratiques d'écriture de la région.

[2] Au sens donné au phénomène, par exemple, par Korning Zethsen et Hill-Madsen 2016, Zethsen 2009, Zethsen 2007, et nombre d'auteurs du domaine des *Translation Studies*.

[3] Voir par exemple *passim* dans : Dalian tushuguan cankaobu (ed.) 1983, Zeng et al. (eds.) 1982, Dars et al. 2001, Rolston et Lin (eds.) 1990.

[4] Barbara Bisetto conduit actuellement des recherches sur une forme de commentaire à dimension traductive, le *yanyi* 演義 (« élaboration du sens »). Voir Bisetto 2018.

lieu à l'intérieur du périmètre des seules langue et écriture chinoises, sous les formes diverses qu'elles comportent, formes dites « classiques » et formes dites « vernaculaires »[5].

Projet de construction d'un outil d'analyse des transformations textuelles dans le contexte intralingual

Les phénomènes de réécriture/traduction intralinguale jouent donc un rôle majeur dans les mécanismes de production des textes littéraires, dans le contexte historique et culturel esquissé ci-dessus. Ils ont été depuis longtemps décrits dans la littérature sinologique. On a établi, pour un très grand nombre de cas, comment tels textes ont été à l'origine de tels autres textes, qui ont été, en totalité ou partiellement, le produit de réécritures, de traductions ou d'amplifications. Cela s'applique en particulier dans le domaine des contes en langue vernaculaire couramment appelés *huaben* 話本, dans la lignée des œuvres éditées ou créées, à la fin des Ming, par Feng Menglong 馮夢龍(1574-1646) et Ling Mengchu 凌濛初 (1580-1644), dont les collections totalisant approximativement 200 contes ont inspiré nombre d'autres auteurs et nombre d'autres collections. Ces relations intertextuelles ont été massivement identifiées, et ont fait l'objet d'études et d'inventaires plus ou moins complets, depuis les travaux pionniers de Hanan, de Lévy & al., de Tan Zhengbi, etc[6].

Cependant le *comment* de ces processus de réécriture, le détail des mécanismes de transfert intertextuel, les stratégies et micro-stratégies développées par les auteurs lorsqu'ils élaboraient un texte sur la base d'un autre texte plus ancien, tous ces phénomènes ont été rarement décrits ou analysés de façon détaillée. Tout un travail de philologie, au plus près des textes et des versions, paragraphe par paragraphe, souvent phrase par phrase, et quelquefois mot à mot, reste à faire. Cette entreprise paraît intéressante, voire fondamentale, car lorsqu'elle est menée, elle est susceptible de révéler un nombre considérable d'informations sur la manière dont les textes ont été produits, et donc sur la manière dont les auteurs conduisaient leur travail d'écriture, utilisaient leurs références dans l'environnement hautement intertextuel qui était le leur, et donnaient sens à leur démarche.

En Occident, Vibeke Børdahl a mené un travail pionnier sur les différents types de réécriture du thème de « Wu Song combattant le tigre » 武松打虎, constituant l'un des cycles du grand roman du XVeme siècle *Au Bord de l'eau* (*Shuihu zhuan* 水滸傳)[7] lorsqu'elle en a livré, de façon extensive, des exemples comparaisons textuelles, en citant *verbatim* des extraits, avec marquages des différentiations entre versions par des moyens typographiques. Alors que la plupart des travaux antérieurs susmentionnés se limitaient à des évocations généralistes des transferts intertextuels, se bornant à des constats souvent impressionnistes exprimés en

[5] Il m'est difficile, à cet égard, de souscrire à l'habitude qui s'est répandue récemment dans la littérature anglophone, de parler du chinois classique en termes de « literary Sinitic [ou : Chinese] » plutôt que « classical Chinese » ; la langue vernaculaire n'est en effet pas moins « littéraire » que la langue classique, et il serait trompeur de considérer l'expression vernaculaire comme l'équivalent d'une expression « populaire ». L'expression « chinois classique » / « classical Chinese » est sans aucun doute critiquable, mais employée faute d'un meilleur substitut.

[6] Etudes fondamentales : Hanan 1973, Hanan 1981, Lévy 1981. Travail d'identification et d'inventaire des sources des *huaben*, recueils de contes en langue vernaculaire du XVIIeme siècle : Tan 1980, Lévy et al. (eds.) 1978–2006.

[7] Børdahl 2013.

termes de rapports de « sources » et d'« influences » des textes les uns sur les autres, elle a pu démontrer comment une description non généraliste, faisant état du détail, au plus près du texte, des processus de transformation à l'œuvre pouvaient être riches d'enseignements. Elle l'a fait en accordant en particulier une grande attention aux choix lexicaux ou stylistiques opérés entre une version et une autre. Cette méthodologie lui a permis de révéler des pans entiers de la réalité des réseaux sémiotiques tissés entre ces textes par différents auteurs / traducteurs intralinguaux, avec ce que cela comportait de stratégies de sens, d'intentions, de recontextualisation et de généalogies textuelles.

Cependant Vibeke Børdahl s'est plutôt attachée à situer les évolutions du thème entre les différents types de textes (roman, théâtre, textes de conteurs ou d'arts chantés) auquel ce thème s'est prêté, sans pour autant entrer dans le détail de la description des stratégies micro-textuelles employées pour ces opérations de réécriture. De plus, son entreprise se limitait à des moyens « analogiques », *i.e.*, visibles sur la page, sans aucune retombée numérisée capable de produire des outils d'évaluation, de description des typologies, d'analyse des verbatim, et de statistique des transformations à l'œuvre.

Il faut donc aller plus loin, et surtout apporter des innovations qualitatives permettant de répondre aux objectifs d'analyse exhaustives des transformations intertextuelles observées. Pour cela, nous avons besoin :

1. d'une part de définir une typologie précise des stratégies et micro-stratégies textuelles utilisées par les auteurs dans leurs opérations de production textuelle ;
2. d'autre part de disposer d'un outil numériques capables de constituer les résultats obtenus en base de données systématisée, interprétable et exploitable.

C'est dans cet objectif qu'est actuellement élaboré et mis au point un outil d'humanités numériques capable de rendre compte de ces objectifs. Cet outil, nommé **IntraTexTT** (pour « **Intralingual Textual Transformation Tool** ») est actuellement (2018) développé par moi-même en collaboration avec Philippe Pons, ingénieur d'études CNRS chargé de l'édition de corpus numériques, dans le cadre du laboratoire CRCAO, UMR 8155.

Typologie des transformations textuelles — Micro-stratégies et caractéristiques stylistiques

Au cours de mon enseignement à l'Ecole pratique des hautes études, notamment autour du thème des récits de suicide et de leurs réécritures[8], j'ai pu analyser les différents types de micro-stratégies développées par les auteurs de la période considérée dans leurs opérations de réécriture/traduction intralinguale de textes sources dans leurs objectifs de production de nouveaux textes. Je l'ai fait principalement autour de la production des contes en vernaculaire, de type *huaben* 話本, susmentionnés, mais aussi en examinant la production de textes théâtraux.

[8] https://hal.archives-ouvertes.fr/hal-01518163 ; https://hal.archives-ouvertes.fr/hal-01654675.

Le résultat de cette analyse m'a amené à définir les opérations de transformation textuelle à l'œuvre entre **texte source (TS)** et **texte cible (TC)** comme comportant obligatoirement la combinaison concomitante de deux catégories :

> **Catégorie I : Micro-stratégies textuelles** (*i.e.*, types de procédés de réécriture / traduction intralinguale)
>
> **Catégorie II : Caractéristiques stylistiques / linguistiques** (*i.e.*, qualités de la langue et du point de vue narratif utilisés)

La catégorie I relève de l'énoncé, de ce que le texte dit, autrement dit du niveau dénotatif. La catégorie II relève de l'énonciation, des caractéristiques de la langue utilisée, autrement dit du niveau indexical.

En accord avec ces deux axes, nous définissions les typologies des opérations de transformation textuelles comme suit[9] :

Catégorie I (Micro-stratégies textuelles [types de procédés de traduction / réécriture])

Citation verbatim

Un segment de TS est repris sans changement dans TC. Dans la grande majorité des cas, le segment en question est en langue classique, puisque c'est la langue dans laquelle sont écrits la plupart des TS.

Quasi citation

Un segment de TS est repris presque à l'identique dans TC. Des modifications diverses peuvent être introduites : ordre des mots, remplacement d'un syntagme de la langue classique par son équivalent vernaculaire, passage à une langue mixte… Le but est souvent de fluidifier l'insertion de la citation au sein de son environnement dans TC. Parfois le quasi citation est l'effet d'une volonté de citation *verbatim*, mais entachée d'erreurs (de mémoire, de copie…).

Traduction / remplacement par (quasi) synonyme

Il s'agit de la traduction intralinguale à proprement parler. Il n'y a pas d'ajout sémantique exogène. Un énoncé en langue classique donne lieu à une translation vers un énoncé équivalent en langue vulgaire ou mixte (mais parfois aussi en restant dans la langue classique). Il y a recherche de synonymes ou quasi synonymes. Cette recherche peut s'accompagner de distorsions sémantiques limitées (si elles sont importantes, on tombe dans le cas de la traduction amplifiée, ou explicitante).

Amplification

Il s'agit d'un procédé de réécriture, classiquement de la langue classique vers la langue vernaculaire, avec ajout libre de détails ou adjuvants, adjonction d'éléments narratifs *ad libitum*, mais liés au contexte et non séparés du noyau narratif de TS. C'est un

[9] Voir en fin de document le tableau synthétique des catégories et sous-catégories.

procédé très commun, souvent très imaginatif, et l'une des principales ressources pour donner plus d'épaisseur narrative et de qualités littéraires à TC, en conférant de l'ampleur à l'énoncé de TS. La différence spatiale (nombre de caractères, quantité de texte) entre TS et TC peut être importante, voire très importante. Après amplification, TC peut représenter plusieurs fois la longueur de TS.

Traduction amplifiée

La traduction intralinguale ne s'en tient pas à une fidélité ou à une recherche de synonymies entre TS et TC, mais s'accompagne de l'addition d'éléments sémantiques non présents dans le champ sémantique de TS. Cette adjonction reste limitée, sans quoi on tombe, soit dans le cas de l'amplification, soit dans le cas de la traduction explicitante / reformulation.

Traduction explicitante / reformulation

La traduction intralinguale entre TS et TC s'accompagne d'une volonté explicative. Il s'agit fondamentalement d'une paraphrase. Il n'y a pas d'ajout d'éléments narratifs ou discursifs supplémentaires lors du passage d'un texte à l'autre, comme dans le cas de la traduction amplifiée. La réécriture concerne prioritairement le champ linguistique, mais affecte le champ sémantique. Si les éléments modifiés ou ajoutés ne sont pas directement présents dans TS, ils restent cependant dans son champ sémantique. Le procédé est classiquement utilisé dans le cas d'une narration supposant une relation d'expert à profane, où le sens de TS a besoin d'être expliqué, avec ajouts de détails et visée pédagogique, volonté de clarifier le TS ou d'y apporter une forme de commentaire.

Ajout

Ajout d'éléments narratifs *ad libitum* et non spécifiquement liés au contexte. Typiquement, des éléments d'une biographie ou d'un récit sont inventés de toute pièce et insérés au sein du continuum de TC. Ce procédé se distingue de l'amplification, car dans le cas de l'ajout, les éléments additionnels ne sont pas liés à un noyau narratif de TS, mais viennent s'adjoindre de façon interpolée.

Omission

Des éléments narratifs de TS sont supprimés lors du passage à TC.

Rappelons qu'un segment donné d'un TC ne peut être le fruit que d'une et une seule des micro-stratégies textuelles désignées ci-dessus. Celles-ci sont exclusives les unes des autres. Il ne peut y avoir deux micro-stratégies concomitantes.

NB : Il n'y a pas de définition *a priori* de ce qui constitue un « segment ». Un segment peut consister en un ou plusieurs paragraphes, en une phrase, en une partie de phrase, voire en un seul mot. Si plusieurs micro-stratégies textuelles sont constatées dans une même phrase, celle-ci doit être divisée en autant de segments qu'il y a de micro-stratégies. La notion de segment est fondamentale pour la construction de l'outil numérique ; elle permettra de mettre en regard de façon systématique les éléments issus de TS vis-à-vis de leur

transformation dans TC, car il est important d'être capable à tout moment d'identifier quelle partie de TS est transformée en quelle autre partie de TC.

Catégorie II : Caractéristiques stylistiques / linguistiques

Dans cette catégorie, deux sous catégories sont obligatoirement combinées de manière concomitante :

A. Caractéristiques linguistiques
B. Point de vue narratif

S'ajoute une catégorie optionnelle : le déplacement textuel

Les caractéristiques linguistiques sont :

a. langue classique
b. langue vulgaire standard (≈ *guanhua*)
c. langue vulgaire dialectale
d. langue mixte (combinaison d'éléments de la langue classique et de la langue vernaculaire)

Le point de vue narratif concerne :

a. la narration à la 3° personne
b. la narration à la 1° personne
c. l'énoncé dialogique

Déplacement textuel

Aux deux sous-catégories ci-dessus s'en ajoute une troisième, optionnelle : la possibilité d'indiquer les segments qui, dans l'ordre de la narration, auraient été déplacés, soit par antéposition, soit par postposition, dans le TC, par rapport à leur position d'origine dans TS.

Comme on le voit, les micro-stratégies textuelles sont donc inséparables de leurs caractéristiques linguistiques et stylistiques. Par exemple, dans TS, un passage ou segment narratif hétérodiégétique, à la troisième personne et en langue classique, peut faire l'objet, dans TC, d'une traduction amplifiée ou d'une amplification en langue vulgaire standard, intradiégétique, avec passage à la première personne ; pour des raisons de stratégie narrative, des éléments du texte peuvent de plus être déplacés, par exemple en postposition afin de maintenir le suspens ou l'intérêt du lecteur.

L'expérience montre que les auteurs de TC utilisent généralement, par segments, la gamme plus ou moins complète des options présentées ci-dessus à chaque fois qu'ils produisent un texte narratif par recyclage / réécriture d'un texte (de textes) antérieur(s).

Indications techniques pour l'élaboration de l'outil numérique

Concernant l'approche technique employée pour l'élaboration de l'outil numérique visé, Philippe Pons, ingénieur d'études CNRS chargé de l'édition de corpus numériques, dans le cadre du laboratoire CRCAO, UMR 8155, chargé du projet, indique :

> « Les textes chinois du corpus sont d'abord transcrits et stylés dans le traitement de texte Microsoft Word. Ensuite, un script XQuery permet de les encoder automatiquement en XML-TEI (Text Encoding Initiative)[10].
> Cette base de données textuelles est alors enregistrée dans un logiciel de base de données XML, BaseX[11], qui assure *via* des scripts en XQuery (XML Query) la mise en ligne du corpus.
> Enfin, l'affichage des textes et leur alignement est assurée par le logiciel Versioning Machine[12],entièrement traduit en XQuery pour les besoins de ce projet, et par du code complémentaire en Javascript. »

Tous les types de transformation et de caractéristiques stylistiques / linguistiques, relevant des catégories I et II, indiqués ci-dessus, possèdent une dénomination complète et une dénomination abrégée, en français et en anglais. La dénomination abrégée est destinée à faciliter leur usage dans la construction de l'outil numérique, notamment au niveau du traitement stylé du texte sous Word. Voir à ce propos le tableau synthétique en fin de document.

Conclusion

Le projet en cours de création consiste donc à créer un outil d'humanités numériques susceptible d'effectuer des relevés de tous les aspects textuels susmentionnés à la fois, de les mettre en évidence de façon visible, de les rendre exploitable et interprétables, notamment par la production de statistiques.

NB : dans un nombre significatif de cas, le TC est le fruit de la réécriture / traduction intralinguale non pas d'un TS unique mais de plusieurs TS combinés. L'outil doit pouvoir rendre compte sans difficulté de cette multiplicité de sources, quel que soit le nombre de ces sources.

Grâce à la mise en ligne, les résultats des opérations de comparaison textuelle pourront être stockés et leurs résultats partagés.

L'outil numérique élaboré sera lui-même destiné à être mis gratuitement au service de la communauté des chercheurs. L'outil lui-même sera adaptable dans toute langue, mais restera exclusivement pensé pour et destiné à des comparaisons textuelles entre des textes chinois.

IntraTexTT doit être finalisé au cours de l'année 2019.

---

[10] http://www.tei-c.org.
[11] http://basex.org.
[12] http://v-machine.org.

ENGLISH

Abstract

*This working paper (July-August 2018) introduces the current project to develop a digital humanities tool called **IntraTexTT** (for "**Intralingual Textual Transformation Tool**"). This tool is intended for the analysis of textual transformations between source texts and target texts in the production of literary narratives in pre-modern China (ca. 13th-18th centuries). As is well known, this production involved complex intertextuality processes, entailing rewriting and intralingual translation, particularly in the context of the diglossic situation of Chinese between classical and vernacular languages. The **IntraTexTT** tool will be used for the analysis, statistical reporting, online publishing and sharing of data related to said situations, in which textual production was operated via intertextual transformations. This working paper sets out the rationale and context for the need to develop such a tool. It gives the general principles, details the typologies, and describes the textual strategies and micro-strategies, with their precise names, to be considered for the realization of a relevant and efficient digital tool. It states that this tool is currently being developed with the help of the CRCAO, UMR 8155, between the author and a CNRS research engineer in charge of digital corpus editing, member of the CRCAO laboratory. The result should be operative during the year 2019.*

General Outline

My research and teaching on pre-modern literature in China (Yuan to mid-Qing, 13th-18th centuries) lead me to describe and analyze cases of rewriting and intralingual translation on a regular basis. In the narrative literature of the period, whatever the genres, the strategy of literary creation massively draws on borrowing and reprocessing of previous themes. These themes are recycled, rewritten, amplified and quite often reoriented ideologically according to the authors' objectives.

In the undergoing project, I deal mainly with vernacular narrative literature. In this vast field, which encompasses different genres—short stories, novels, theater plays, but also forms of narrative or sung poetry—, the authors always take careful account of the linguistic environment of their time, *i.e.*, an environment characterized by diglossia, or more generally by forms of "schizoglossia"[13] between classical language(s) *(wenyan* 文言*)* and vernacular

[13] To use a name suggested by King 2015, 13, n. 23, speaking of "schizoglossographia". See also Ross King in his lecture of 14 May 2018, "Ditching'Diglossia': Describing Ecologies of the Spoken and Inscribed in Pre-modern Korea", at the monthly seminar (organized by Andréas Stauder and myself) EPHE Section des Sciences historiques et philologiques & SCRIPTA-PSL History and Practices of Writing : *Diglossie, traduction intralinguale, reécriture, commentaire - Diglossia, intralingual translation, rewriting, and commentary*. The notion of "diglossia", as first defined by Ferguson 1959, is contested today among East Asian scholars, who find its distinction of so-called "high" and "low" (H/L) languages levels quite problematic and irrelevant in the case of languages and writing practices in the area.

language(s) *(tongsu* 通俗*)*.[14] In this context we witness massive phenomena of intralingual rewriting, even intralingual translation,[15] typically from the classical literary language to the vernacular language. Authors who practice this kind of intralingual rewriting or translation always show a keen awareness of the schizoglossic situation of the Chinese language of their time in its written forms. As has often been described in the scholarly literature since the first decades of the last century (Lu Xun 魯迅, Sun Kaidi 孫楷第, Tan Zhengbi 譚正璧...), premodern authors of fiction constantly hint at the art of storytellers, whom they pretend to imitate in their works. They did so when they used elements of orality in explaining, paraphrasing, and re-telling previous stories under a new form. Making a theme accessible to as many people as possible, bridging the gap from expert to lay, are constant arguments when it comes to justifying their intralingual rewriting/translation endeavor. They often explain themselves explicitly on this subject. The literati of the time, from the Song up to the Ming-Qing, refer massively to the prevailing schizoglossic situation, in particular in the prefaces and forewords that they add to the published novels or short story collections. In those prefaces, they typically refer to the intrinsic difficulty of the classical language and to the advantages of rewriting in a language "accessible to the vulgar" *(tongsu* 通俗*)*. They consider the latter having a greater capacity for communicating *(tong* 通*)* with the general public, and being more influent in terms of education, of capacity to transmit knowledge and (especially moral) values.[16]

These phenomena are currently the subject of an expanding interest, particularly in connection with a renewed attention regarding the understanding and evaluation of the emergence of vernaculars in pre-modern/modern China and East Asia (13th-19th centuries). In this context I am currently conducting a project entitled "Intralingual Translation, Language Shifting, and the Rise of Vernaculars in East Asian Classical and Premodern Cultures", co-directed by Barbara Bisetto, lecturer at the Universita Degli Studi di Verona.[17] Bisetto and I are co-organizers of a research program entitled "Diglossia, intralingual translation, interpretation" that is part of the pluriannual research scheme of the CRCAO (Centre de recherches sur les civilisations de l'Asie orientale, UMR 8155, CNRS, Centre National de la Recherche Scientifique, France), for the 2019-2023 period (CNRS UMR's [Unité mixte de recherche] five-year contract).

In this framework, we intend to collaborate with specialists studying Chinese texts in relation with vernaculars, either in China or in surrounding countries. For my part, and this applies to everything that is discussed below, I am interested in intralingual phenomena exclusively, *i.e.*, taking place within the sole perimeter of Chinese language and writing. In these I include their various, "classical" or "vernacular" forms.[18]

---

[14] The *tongsu* category can include both standard and dialectal aspects.

[15] In the sense given to the phenomenon, for example, by Korning Zethsen and Hill-Madsen 2016, Zethsen 2009, Zethsen 2007, and numerous authors working in the field of Translation Studies.

[16] See for example: Dalian tushuguan cankaobu 1983, Zeng & a. 1982, Dars & a. 2001, Rolston & Lin 1990.

[17] Barbara Bisetto is currently exploring a form of commentary with a translatory dimension, the *yanyi* 演義 ("elaboration of the meaning"). See Bisetto 2018.

[18] In this respect, it can hardly subscribe to the habit which has recently become widespread in Anglophone literature, to refer to classical Chinese in terms of "literary Sinitic [or: Chinese]". The vernacular language is indeed no less "literary" than its classical counterpart, and furthermore it would be misleading to consider vernacular expression as the equivalent of a "popular" expression. The expression "classical Chinese" /"chinois classique" is undoubtedly open to criticism, but I still use it for lack of a more convincing substitute.

The Current Project: Building a Digital Tool for the Analysis of Intralingual Textual Transformation

The case of intralingual rewriting/translation thus plays a major role in the mechanisms of literary text production, in the historical and cultural background outlined above. Such textual production has long been described in the sinological literature. How some texts originate in some other, earlier texts, and are, in whole or in part, the product of rewritings, translations or amplifications, is a well-established fact. This applies in particular to vernacular tales commonly called *huaben* 話本, in the tradition made popular in the late Ming by authors/editors like Feng Menglong 馮夢龍 (1574-1646) and Ling Mengchu 凌濛初 (1580-1644), whose short story collections, totalizing approximately 200 tales, inspired many imitators, who, like them, composed their narratives largely through rewriting. These intertextual relations have been identified in good part, and have been the subject of more or less complete studies and inventories since the pioneering work of Tan Zhengbi, Hanan, Lévy & al., etc.[19].

However, *how* exactly did the authors of the time carry out these rewriting processes, what were the particulars of the intertextual transfer mechanisms that they implemented, what kind of strategies and micro-strategies did they develop when putting together a text on the basis of another (other), older text(s)? These processes *per se* have rarely been described or investigated in full detail. A careful philological endeavor thus remains to be carried out, that would explore these textual transformations between versions as closely as possible; this should be done paragraph by paragraph, often sentence by sentence, sometimes word by word. This exploration appears as fundamental, as it is likely to reveal a considerable amount of information on the way in which the texts were produced. This could help us put into light how authors conducted their (re)writing task, how they managed their references in the highly intertextual network which was their intellectual environment, and to better understand what they intended by working the way they did.

In the West, Vibeke Børdahl has done a pioneering work on the different kinds of rewriting of the theme "Wu Song Fights the Tiger" 武松打虎, that constitute one of the cycles of the great 15th century novel *Water Margin (Shuihu zhuan* 水滸傳*)*[20]. She gave extensive examples of textual comparisons on the theme, citing verbatim excerpts, publishing her results in publications that would show differentiations between versions by means of typographical devices. While most of the previous works mentioned above were limited to generalist evocations of intertextual transfers, confining themselves to sometimes impressionistic observations expressed in rather vague terms of "source" and "influence" of such text on such other text, she was able to demonstrate to what extent a non-generalist description could be informative, when the textual comparison was conducted at a very detailed level. In paying particular attention to the lexical or stylistic choices made between one version and another, her efforts proved fruitful. This methodology allowed her to disclose whole sections of the semiotic networks woven between different texts by different authors / intralingual

[19] Basic references : Hanan 1973, Hanan 1981, Lévy 1981. Identification and inventory of *huaben* sources, collection of stories in vernacular language of the 17th century: Tan 1980, Lévy & a. 1978–2006.
[20] Børdahl 2013.

translators, with what it entailed in terms of writers' intents, semiotic strategies, recontextualization and textual genealogies.

However, although she has convincingly shown how to situate the evolution of the theme between the different types of texts to which this theme lent itself (novels, plays, storytellers' texts or sung arts), Børdahl did not go as far as describing and analyzing down to the minutest details all the textual micro-strategies involved in the production of rewritten versions. Besides, her endeavor was limited to "analogical" means, *i.e.*, devices typographically visible on the page, without any digitized spin-off capable of producing tools for evaluation, description of typologies, analysis of verbatim, and statistics of the transformations at work.

It is therefore necessary to go further, in particular to be able to provide qualitative improvements in order to meet the objectives of an exhaustive analysis of the observed intertextual transformations. For that, we need:

1. to define a precise typology of the textual strategies and micro-strategies used by authors in their textual production;
2. to benefit from a digital tool capable of constituting the results obtained into a systematized, interpretable and exploitable database.

It is with this objective in mind that a tool of digital humanities is currently (2018) being developed by myself in collaboration with Philippe Pons, CNRS design engineer in charge of the edition of digital corpus, as part of the CRCAO laboratory, UMR 8155, CNRS Centre National de la Recherche Scientifique, France. This tool is named **IntraTexTT (for "Intralingual Textual Transformation Tool")**.

Typology of textual transformations - Micro-strategies and stylistic characteristics

While teaching at the Ecole Pratique des Hautes Etudes, in particular when exploring transformative textual operations focused on the theme of suicide stories,[21] I have had the opportunity to analyze the different types of micro-strategies developed by the authors of the period in question in their operations of intralingual translation and rewriting from source texts to new, target texts. I did this in large part by investigating the production of the above-mentioned *huaben* 話本 stories in the vernacular, but also the production of theatrical texts.

The result of this survey has led me to define the textual transformation operations at work between source text (ST) and target text (TT) as necessarily involving the concomitant combination of two categories :

Category I : Textual micro-strategies (i.e. types of rewriting processes / intralingual translation)

Category II : stylistic / linguistic characteristics

[21] https://hal.archives-ouvertes.fr/hal-01518163 ; https://hal.archives-ouvertes.fr/hal-01654675.

Category I is the statement, what the text says, in other words it concerns the denotative level. Category II is a matter of enunciation, of the characteristics of the language used; in other words that concerns the indexical level.

In accordance with these two directions, we defined the typologies of textual transformation operations as follows [22] :

Category I (Textual micro-strategies [types of translating / rewriting procedures])

Verbatim citation

An ST segment is included in TT unchanged. In most cases, the segment is in classical Chinese, since this is the language in which most ST are written.

Quasi-citation

An ST segment is almost identical in TT. Various modifications can be introduced: word order, replacement of a syntagm in the classical language by its vernacular equivalent, switching to mixed language... The aim is often to fluidify the insertion of the quotation within its environment in TT. Sometimes the quasi-quotation is motivated by an intent at *verbatim* quotation, but with some errors (in copying, or memorizing…).

Translation / replacement by (quasi-)synonym

This is intralingual translation strictly speaking. There is no exogenous semantic addition. A statement in classical is simply translated in its equivalent in vernacular (standard or dialectal) or mixed language (though it sometimes remains in classical Chinese). The textual transformation is aimed at finding synonyms or quasi-synonyms. This research can be accompanied by limited semantic distortions (if the distortions are important, the transformation pertains to amplified, or explanatory translation).

Amplification

This is a rewriting process, typically from classical to vernacular language, with free addition of details or adjuvants, addition of narrative elements *ad libitum*, but which are linked to the context and not separated from the narrative core of ST. This is a very common process, often very imaginative, and one of the main resources to give more narrative depth and literary qualities to TT, giving extra amplitude to the ST statements. The spatial difference (number of characters, quantity of text) between ST and TT can be large or very large. TT's length can end up amounting to several times the length of ST.

Amplified translation

Intralingual translation does not limit itself to be close to ST or to be searching for synonymies between ST and TT, but is accompanied by the addition of semantic elements which are not present in the semantic field of ST. This addition remains limited,

[22] See at the end of the document the summary table of the categories and subcategories.

though, otherwise the transformations pertains to the cases of either in the amplification, or explanatory translation / reformulation.

Explanatory translation / reformulating

The intralingual translation between ST and TT is accompanied by a clear explicative, explanatory intent. It is basically a paraphrase. There are no additional narrative or discursive elements added during the transfer from one text to another, as in the case of amplified translation. The rewriting concerns primarily the linguistic field, but affects the semantic field. If the modified or added elements are not directly present in ST, they remain however in its semantic field. The process is typically used in the case of a narrative involving an expert-to-lay relationship, where the meaning of ST needs to be explained, with added statements and pedagogical intent, and a desire to clarify or add some kind of comment on the ST.

Addition

Addition of narrative elements *ad libitum* and not specifically related to context. Typically, elements of a biography or story are invented from scratch and inserted into the TT continuum. This process is different from amplification, because in the case of addition, the additional elements are not linked to a narrative nucleus of ST, but are added as something entirely interpolated, not present even as a hint in any part of ST.

Omission

Narrative elements of ST are deleted when transferring to TT.

We must remember that a given segment of a TT can only be the effect of *one and only one* of the textual micro-strategies described above. Two micro-strategies cannot be concomitant.

NB: There is no *a priori* definition of what constitutes a "segment". A segment may consist of one or more paragraphs, a sentence, a part of a sentence or even a single word. If several text micro-strategies are found in the same sentence, it must be divided into as many segments as there are micro-strategies. The notion of segment is fundamental for the construction of the digital tool; it will make it possible to compare systematically the elements resulting from ST with their transformation in TT, because it is important to be able at any time to identify which part of ST is transformed into which corresponding part of TT.

Category II : Stylistic / linguistic characteristics

In this category, two subcategories must be combined at the same time:

A. Linguistic characteristics

B. Narrative point of view

An optional category is added: the textual displacement

Linguistic characteristics are:

a. classical (literary) Chinese
b. vernacular Chinese: "standard" *(guanhua)*
c. vernacular Chinese: dialectal
d. mixed language (combination of classical and vernacular Chinese)

The narrative point of view concerns:

a. the 3rd person narration
b. the 1st person narration
c. dialogue

Textual displacement

In addition to the two subcategories above, there is an optional third category: the possibility of indicating the segments which, in the order of narration, may have been moved, either by anteposition or by postposition, in the TT, with respect to their original position in ST.

As we can see, textual micro-strategies are therefore inseparable from both their linguistic and stylistic characteristics. For example, in ST, a heterodiegetic narrative passage or segment, in the third person and in the classical language, can result, when transposed into TT, in an amplified translation, or an amplification, written in standard, intradiegetic, vernacular language, expressed at the first person; for reasons of narrative strategy, elements of the original text can also be displaced within the textual continuum, for example in postposition in order to maintain the suspense or the interest of the reader.

Experience shows that TT authors generally use, in successive segments, more or less the full range of options presented above every time they produce a new text on the basis of a previous text.

Technical information for the development of the digital tool

Regarding the technical approach used to develop the digital tool currently under construction, Philippe Pons (CNRS design engineer in charge of the edition of digital corpus, member of the CRCAO, UMR 8155), in charge of the project, indicates:

> "The Chinese texts in the corpus are first transcribed and styled within MSWord word processor. Then, an XQuery script automatically encodes them in XML-TEI (Text Encoding Initiative).[23]

[23] http://www.tei-c.org.

This textual database is then registered in an XML database software, BaseX,[24] which ensures via scripts in XQuery (XML Query) the putting on line of the corpus.
Finally, text display and alignment is ensured by the software Versioning Machine,[25] fully translated into XQuery for the needs of this project, and by additional code in Javascript."

All types of transformation and stylistic/linguistic characteristics, falling under categories I and II, indicated above, have a full name and an abbreviated name, both in French and in English. The abbreviated names are intended to facilitate their use in the construction of the numerical tool, in particular at the level of styled text processing in MSWord. See the summary table at the end of the present document.

Conclusion

The design of the project currently being developed is therefore to create a tool in digital humanities capable of recording all the textual aspects mentioned above, highlighting them in a visible way, making them exploitable and interpretable, in particular through the output of statistical data.

NB: In many cases, TT is the result of rewriting / intralingual translation not of a single ST but of several combined STs. The tool must be able to account for this multiplicity of sources without difficulty, whatever the number of these sources.

The results of the textual comparison operations can be stored and their results shared online.

The digital tool developed will itself be intended to be made available free of charge to the community or scholars and researchers. The tool itself will be adaptable in any language, but will remain exclusively designed and intended for textual comparisons between Chinese texts.

IntraTexTT will be completed in the course of 2019.

[24] http://basex.org.
[25] http://v-machine.org.

Références bibliographiques citées

Bibliographical References

Bisetto, Barbara. 2018. “Commentary and Translation: Exploring the *Du Lü Yanyi* 杜律演義.” *Acta Universitatis Carolinae Philologica (Orientalia Pragensia)* 2017 (4): 97–108.

Børdahl, Vibeke. 2013. *Wu Song Fights the Tiger: The Interaction of Oral and Written Traditions in the Chinese Novel, Drama and Storytelling*. Nordic Institute of Asian Studies Monograph Series 122. Copenhagen: NIAS Press.

Dars, Jacques, Hingho Chan, Pierre-Henri Durand, Jean Levi, and Rémi Mathieu. 2001. *Comment lire un roman chinois : anthologie de préfaces et commentaires aux anciennes oeuvres de fiction*. Paris: Editions Philippe Picquier.

Ferguson, Charles A. 1959. “Diglossia.” *Word* (15): 325–340.

Hanan, Patrick. 1973. *The Chinese Short Story. Studies in Dating, Authorship, and Composition*. Harvard-Yenching Institute. Monograph Series 21. Cambridge, Mass: Harvard University Press.

———. 1981. *The Chinese Vernacular Story*. Harvard East Asian Series 94. Cambridge, Mass: Harvard University Press.

King, Ross. 2015. “Ditching ‘Diglossia’: Describing Ecologies of the Spoken and Inscribed in Pre-Modern Korea.” *Sungkyun Journal of East Asian Studies* 15 (1): 1–19.

Korning Zethsen, Karen, and Aage Hill-Madsen. 2016. “Intralingual Translation and Its Place within Translation Studies – A Theoretical Discussion.” *Meta : Journal Des Traducteurs / Meta : Translators’ Journal* 61 (3): 692–708.

Lévy, André. 1981. *Le Conte en langue vulgaire du XVIIe siècle*. Bibliothèque de l’Institut des Hautes études chinoises. Paris: Institut des Hautes études chinoises, PUF.

Lévy, André, Michel Cartier, Rainier Lanselle, Hingho Chan, and al., ed. 1978–2006. *Inventaire Analytique et Critique Du Conte Chinois En Langue Vulgaire*. 5 vols. Paris: Collège de France, Institut des Hautes Etudes Chinoises.

Rolston, David L., and Shuen-fu Lin, ed. 1990. *How to Read the Chinese Novel*. Princeton Library of Asian Translations. Princeton, N.J: Princeton University Press.

Zethsen, Karen Korning. 2007. “Beyond Translation Proper—Extending the Field of Translation Studies.” *TTR: Traduction, terminologie, rédaction* 20 (1): 281–308.

———. 2009. “Intralingual Translation: An Attempt at Description.” *Meta: Journal Des Traducteurs* 54 (4): 795–812.

Dalian tushuguan cankaobu 大连图书馆参考部, ed. 1983. *Ming Qing xiaoshuo xuba xuan* 明清小说序跋选. Shenyang: Chunfeng wenyi chubanshe.

Zeng, Zuyin 曾祖荫, Xianpei Wang 王先霈, Qingquan Huang 黄清泉, and Weimin Zhou 周伟民, ed. 1982. *Zhongguo lidai xiaoshuo xuba xuanzhu* 中国历代小说序跋选注. Xianning: Changjiang wenyi chubanshe.

Tan, Zhengbi 譚正璧. 1980. *Sanyan Liangpai ziliao* 三言兩拍資料. 2 vols. Shanghai: Shanghai Guji chubanshe.

Tableau des catégories et sous-catégories

Chart of Categories and Subcategories

IntraTexTT

Tableau des catégories et sous-catégories ## Chart of Categories and Subcategories

La transformation intertextuelle implique obligatoirement une combinaison de Catégorie I + Catégorie II

Intertextual transformation necessarily implies a combination of Category I + Category II

Catégories / Categories

Catégorie I / Category I				Catégorie II / Category II					
Français		*English*		*Français*			*English*		
Micro-stratégies textuelles (types de procédés de traduction / réécriture)		**Textual micro-strategies (types of translating / rewriting procedures)**		**Caractéristiques stylistiques / linguistiques**			**Stylistic / linguistic characteristics**		
➜ Niveau dénotatif (ce que le texte dit) (énoncé)		***➜ Denotational level (what the text says) (utterance)***		***➜ Niveau indexical (qualité de la langue utilisée) (énonciation)***			***➜ Indexical level (quality of the language used) (enunciation)***		
Obligatoirement un seul au choix		***Mandatory only one choice***		***Obligatoire : 1 seul parmi A et 1 seul parmi B***			***Mandatory: 1 only among A and 1 only among B***		
				Optionnel : ajout de C			***Optional: addition of C***		
Forme complète	*Forme abrégée*	*Complete form*	*Abbreviated form*			*Forme complète*	*Forme abrégée*		*Complete form*	*Abbreviated form*
citation verbatim	verbatim	verbatim citation	verbatim	A	Caractéristiques linguistiques	langue classique	l_class	A	Linguistic characteristics	classical Chinese	l_class
quasi citation	quasi_cit	quasi-citation	quasi_cit	A	Caractéristiques linguistiques	langue vulgaire : guanhua « standard »	l_standard	A	Linguistic characteristics	vernacular Chinese: “standard” guanhua	l_standard
traduction / remplacement par (quasi) synonyme	traduction	translation / replacement by (quasi-)synonym	translation	A	Caractéristiques linguistiques	langue vulgaire : dialecte	l_dialect	A	Linguistic characteristics	vernacular Chinese: dialectal	l_dialect
amplification	amplification	amplification	amplification	A	Caractéristiques linguistiques	langue mixte	l_mixt	A	Linguistic characteristics	mixed language	l_mixt
traduction amplifiée	trad_amplif	amplified translation	amplif_transl	B	Point de vue narratif	narration à la 3° personne	pers3	B	Narrative point of view	3rd person narration	pers3
traduction explicitante / reformulation	reform	explanatory translation / reformulating	reform	B	Point de vue narratif	narration à la 1° personne	pers1	B	Narrative point of view	1st person narration	pers1
ajout	ajout	addition	addition	B	Point de vue narratif	dialogue	dialog	B	Narrative point of view	dialogue	dialog
omission	omission	omission	omission	C		déplacement textuel	déplac	C		textual displacement	displacement

Sous-catégories / Subcategories