

# Expérience d'entraînement de TreeTagger et d'intégration à l'interface Web de SATO

Fernande Dupuis, Robert Kapitan, François Daoust

Université du Québec à Montréal – Montréal – Canada

## Résumé

Le TreeTagger est un outil paramétrable qui permet d'attribuer des catégories grammaticales dans diverses langues suite à des entraînements appropriés. Ce texte a deux objectifs. D'abord, il s'agit de faire état d'une adaptation de TreeTagger modifiant les fichiers élaborés par Achim Stein pour l'ancien français en fonction d'un nouveau jeu de catégories plus élaboré pour les verbes et les noms. Le défi de la transformation est surtout relié à l'identification adéquate des positions des catégories dans le fichier d'entraînement. L'autre défi est relié à la très grande variation dans l'orthographe de l'ancien français. Étant donnée la taille des fichiers d'entraînement, il a fallu recourir à un processus d'automatisation par l'emploi de nombreuses expressions rationnelles modifiant le contenu des fichiers. L'outil a été testé sur des textes médiévaux. Ensuite, il s'agit de montrer comment l'appel au TreeTagger a été intégré, selon diverses langues, à l'interface Web du logiciel SATO permettant ainsi d'annoter tout texte en faisant appel au TreeTagger sur les parties du texte admissibles à la catégorisation automatique.

## Abstract

The TreeTagger is a configurable tool that allows one to assign grammatical categories in various languages as a result of proper training. This article pursues two goals. First, it describes how we adapted the TreeTagger by modifying the files developed by Achim Stein for Old French in order to make it compatible with a new set of more elaborate categories for verbs and nouns. The first challenge of the transformation is mainly related to the proper identification of the position of the categories in the training file. The other challenge is related to the great variation in the spelling of Old French. Given the size of the training files, we had to create an automation process using many regular expressions modifying the content of the files. The tool was tested on medieval texts. Then it is shown how the TreeTagger was integrated to the web interface of the software SATO in order to provide an automatic categorization on any eligible text.

**Keywords:** TreeTagger, SATO, grammatical tags, automatic categorization, medieval French

## 1. Introduction

L'essor des recherches en linguistique historique au cours des trois dernières décennies a donné lieu à la constitution d'importants corpus informatisés. La collaboration entre des équipes internationales reconnues pour leurs travaux sur l'évolution du français nous permet dorénavant d'accéder à des corpus ayant comme empan toute la période médiévale.

Nos recherches sur le changement linguistique dans l'évolution du français nous imposent toutefois de lourdes et coûteuses tâches de codage et d'indexation. Pour alléger ces tâches nous avons élaboré une chaîne de traitement qui fait appel à des analyseurs de différents types. Le but de cette recherche est de faire évoluer le jeu des étiquettes du TreeTagger (dorénavant TT),

un analyseur qui associe une étiquette morpho-syntaxique aux mots du texte, afin de le rendre compatible avec un jeu plus élaboré d'étiquettes. L'article est organisé comme suit. Après avoir présenté les caractéristiques des analyseurs que nous avons utilisés, nous expliquons la procédure qui nous a permis d'enrichir l'outil de *taggage* et nous soulignons les problèmes suscités par notre démarche. Nous présentons la plateforme informatique utilisée et son intégration à l'interface Web de SATO. Suivent, une évaluation de la performance du *taggage* enrichi sur deux textes du français médiéval et quelques remarques conclusives.

## 2. Problématique

Dans le cadre du projet *Modéliser le changement : les voies du français*<sup>1</sup>, nous avons développé PREMHis : *Plateforme de Recherche et d'Étiquetage en Morphologie Historique* (Kapitan et al., 2007) un outil informatique pour l'analyse morphologique du français médiéval. Cette plateforme établit une passerelle entre le logiciel d'analyse de texte SATO (Daoust, 2009) et des analyseurs externes : TT (Stein and Schmid, 1995) et Flemm (Namer, 2000). TT est un outil permettant d'assigner automatiquement des annotations morpho-syntaxiques. Flemm est un ensemble de modules Perl permettant d'analyser la flexion du français moderne dans des corpus préalablement étiquetés par le TT. Nous avons utilisé une version du TT entraîné par Achim Stein (2007) pour catégoriser des textes en français médiéval. Toutefois, le système de catégories grammaticales utilisé par cette version du TT était peu compatible avec le jeu d'étiquettes *pdd* (Partie du discours) développé dans le cadre du projet *Modéliser le changement*.

Afin d'améliorer les performances de cette première version de la plateforme, nous avons entrepris de modifier le TT pour y introduire le jeu d'étiquettes *pdd*. En fait, nous nous sommes limités pour l'essentiel aux verbes et aux noms.

L'éclatement de l'étiquette *Verbe* du TT pour la redistribuer en quatre sous-catégories dans *pdd* a nécessité un traitement assez complexe. Nous avons utilisé la version de 2007 du fichier de paramètres du TT : *français.par*. Trois fichiers d'entraînement correspondent à cette version : un lexique original du TT<sup>2</sup>, un dictionnaire d'entraînement original<sup>3</sup> et *openclass* qui sert à définir le jeu d'étiquettes. Afin de modifier les fichiers d'entraînement du TT, nous avons également utilisé les fichiers de sortie de SATO préalablement catégorisés avec les étiquettes *pdd*. Le logiciel SATO permet d'enrichir les corpus grâce à l'application PREMHis (Daoust et al., 2007) qui procède au *taggage* intégral des textes médiévaux.

L'entraînement du TT à l'aide des étiquettes *pdd* a nécessité une procédure en trois étapes. Nous avons d'abord remplacé les étiquettes originales du TT par celles de *pdd*. Ensuite, nous avons entraîné le TT avec le nouveau jeu d'étiquettes pour finalement affiner la catégorisation en fonction des résultats obtenus. Ainsi, dans un premier temps, il s'est agi de remplacer uniquement l'étiquetage affectant successivement les verbes (*VER* est devenu *VJ*, *VPP*, *VG*, *VX*)<sup>4</sup> et les noms (*NOM* est devenu *NP* et *NC*). Pour effectuer les remplacements, il a fallu modifier les fichiers du lexique et d'entraînement. Dans un deuxième temps, nous avons redéfini l'outil d'entraînement pour qu'il soit capable de prendre en charge les nouvelles étiquettes. À titre de test, nous avons ciblé les verbes infinitifs (*pdd* : *VX*) étant donné qu'ils sont morphologiquement

<sup>1</sup> Projet dirigé par France Martineau de l'Université d'Ottawa.

<sup>2</sup> Le fichier *af:flex* comprend 235.197 lignes en tout.

<sup>3</sup> Le dictionnaire d'entraînement *amslit* contient 2.446.440 lignes.

<sup>4</sup> Respectivement : verbes conjugué, participe passé, gérondif, infinitif. En fait les étiquettes du jeu de *pdd* sont plus élaborées puisqu'elles distinguent aussi les auxiliaires et les modaux (dans leur formes conjuguées, infinitivales et gérondives). Nous ignorons ces étiquettes dans ce travail.

plus simples à repérer<sup>5</sup>. Une fois le nouvel étiquetage défini, nous avons, dans un troisième temps, testé le taggage sur de nouveaux textes et procédé aux remplacements des étiquettes.

### 3. Difficultés rencontrées

Les principaux problèmes qu'il nous a fallu résoudre au cours de la procédure résultent surtout de la nature des jeux d'étiquettes, de la variation orthographique, de l'architecture du TT et de l'ambiguïté catégorielle des mots. Voici une description plus détaillée des problèmes rencontrés :

- Les jeux d'étiquettes sont incompatibles parce que celles de TT sont moins spécifiques que celles de *pdd*;
- Le remplacement des verbes et des noms doit respecter leurs positions d'origine dans les fichiers d'entraînement;
- Le repérage des erreurs exige une stratégie de gestion des effets secondaires du raffinement automatique;
- La présence de plusieurs catégories associées à une seule unité provoque de l'ambiguïté lexicale;
- Les variations orthographiques doivent être prises en compte dans la gestion de l'ambiguïté.

#### 3.1. Nature des étiquettes

Le jeu de catégories du TT est limité aux catégories lexicales tandis que le jeu d'étiquettes *pdd* contient majoritairement des catégories grammaticales. Notre stratégie a consisté à réutiliser le module de filtrage utilisé dans PREMHis composé d'une suite d'expressions rationnelles pour corriger la sortie du TT afin de la préciser au plan morphologique. Il s'agissait alors de compléter ce travail pour y introduire les catégories de *pdd*, toujours pour les unités non ambiguës. Le module de filtrage ainsi complété a été utilisé pour modifier les fichiers d'entraînement du TT.

#### 3.2. Question d'architecture

Dans un premier temps, nous avons tenté de remplacer des verbes extraits des fichiers d'entraînement. Nous avons constaté que la position de l'unité dans les fichiers jouait un rôle important ; en fait, cette position est nécessaire pour effectuer des calculs statistiques dans le dictionnaire d'entraînement. Nous avons donc décidé de remplacer les unités verbales et nominales directement dans leurs positions d'origine, c'est-à-dire dans l'ordre initial qui existait dans les fichiers d'entraînement. Avant de procéder aux substitutions, il fallait nous assurer que les étiquettes qui n'ont pas été remplacées soient faciles à identifier.

Une fois le remplacement global terminé, un test a été effectué sur *La chanson de Roland*. Sur les 756 verbes du fichier test, 260 ont été correctement catégorisés avec les nouvelles étiquettes issues de *pdd*, les autres verbes ayant conservé l'étiquette *VER*. Le remplacement dans les fichiers originaux nous a permis d'atteindre environ 35% de précision.

Il fallait ensuite passer à l'étape de raffinement afin de corriger les erreurs et augmenter la précision de l'étiquetage. Nous avons automatisé le processus par l'emploi de nombreuses expressions rationnelles modifiant automatiquement le contenu des fichiers. Nous avons utilisé plusieurs textes médiévaux afin d'enrichir l'outil avec comme objectif un seuil de précision

---

<sup>5</sup> Les désinences de l'infinitif ont un nombre réduit de formes régulières, c'est-à-dire qu'elles présentent peu de variantes graphiques.

de 80%. Nous estimons avoir atteint la limite de l'approche automatisée par expressions rationnelles.

### 3.3. Variation orthographique

Des problèmes reliés à la variation orthographique de l'époque ont émergé et nous n'avons pas pu continuer à utiliser des expressions rationnelles dans les dictionnaires d'apprentissage. Par exemple, les désinences de la deuxième personne du pluriel (*gardez, trovez*) chevauchent souvent celles des participes passés (*ont gardez, a trovez*) et il est impossible de tout remplacer à l'aide d'une seule expression  $ez\|tVJ$  qui remplacerait  $ez\|tVER$ . Le remplacement doit se faire en mode cas par cas afin de nous assurer que le verbe auxiliaire n'est pas présent dans la phrase.

Par contre, les noms ne présentent pas le même problème et les substitutions ont pu être effectuées assez rapidement. Cependant, nous avons dû ajouter plusieurs noms propres repérés dans les textes médiévaux directement dans le fichier de lexique. Cependant, nous avons constaté qu'il fallait également nous assurer que ces unités soient présentes dans un contexte bien défini dans le dictionnaire d'entraînement. Dans certains cas, nous avons dû enrichir le contexte c'est-à-dire ajouter la phrase complète.

### 3.4. Ambiguïté lexicale

Nous avons consacré beaucoup de temps à gérer les positions des unités ambiguës. Que ce soit des verbes ou des noms, il existe des mots qui peuvent appartenir à différentes catégories à la fois, par exemple : *sains* étant un adjectif, un nom commun, un nom propre et une préposition, *fait* étant un verbe au présent, un participe passé et un nom commun. Dans ce cas, il a fallu créer plusieurs entrées et effectuer des calculs manuels dans le dictionnaire afin de privilégier l'entrée la plus répandue.

## 5. La plateforme informatique

L'utilisation du TT et la validation des résultats s'effectuent depuis le logiciel SATO (Daoust, 2009) et son interface Web. TT effectue ses calculs sur une liste d'occurrences (*token* en anglais) à raison d'un mot par ligne. Il est aussi possible d'ajouter une catégorie au mot, catégorie TT qui sera prise en compte par l'analyseur.

Dans le cas où le point de départ de l'analyse est un simple document en format texte, un programme en Perl est fourni par les développeurs du TT pour produire un découpage en mots selon le format décrit. Dans notre cas toutefois, le découpage est déjà réalisé par SATO qui représente le texte comme une suite d'occurrences de formes lexicales. Dans cette suite, chaque occurrence possède un numéro séquentiel, valeur de la propriété *NoOcc* de l'occurrence. Pour l'utilisation de TT, il suffira d'exporter le texte dans un fichier en format tabulaire et de garder un fichier de correspondance entre le numéro de la ligne dans le fichier TT et le numéro du mot dans la représentation interne de SATO. Au besoin, on pourra aussi choisir de n'exporter qu'une partie du corpus. Il suffit ensuite d'appeler un programme agissant à titre de passerelle entre SATO et TT.

Cette passerelle, en plus d'appeler le TT, ajustera les formats de données avant et après l'appel. Il suffira d'affecter les résultats de l'analyse du TT sur chacune des occurrences numérotées dans SATO.

Suite à l'application de cette procédure, on procède au protocole de validation permettant de comparer les catégories importées du TT, dans la propriété SATO *tt*, avec les valeurs existantes

de la propriété *pdd* qui ont fait l'objet d'une validation manuelle minutieuse. Pour ce faire, on définira une troisième propriété, appelée *conforme*, avec des valeurs *oui* et *non* indiquant que la catégorie dans *tt* est conforme ou non avec celle de *pdd*. Voici le contenu du scénario de commandes SATO.

---

```

PROPRIETE DÉFINIR Conforme symbolique pour Texte VALEUR oui non
PROPRIETE ATTRIBUER Conforme = VALEUR nil POUR $
PROPRIETE ATTRIBUER Conforme = VALEUR oui POUR $*pdd=(VJ,EJ,AJ,MDJ,LJ)*tt=VJ
PROPRIETE ATTRIBUER Conforme = VALEUR non POUR $*pdd=(VJ,EJ,AJ,MDJ,LJ)*tt~VJ
PROPRIETE ATTRIBUER Conforme = VALEUR oui POUR $*pdd=VG*tt=VG
PROPRIETE ATTRIBUER Conforme = VALEUR non POUR $*pdd=VG*tt~VG
PROPRIETE ATTRIBUER Conforme = VALEUR oui POUR $*pdd=VPP*tt=VPP
PROPRIETE ATTRIBUER Conforme = VALEUR non POUR $*pdd=VPP*tt~VPP
PROPRIETE ATTRIBUER Conforme = VALEUR oui POUR $*pdd=(VX,AX,EX,LX,MDX)*tt=VX
PROPRIETE ATTRIBUER Conforme = VALEUR non POUR $*pdd=(VX,AX,EX,LX,MDX)*tt~VX
PROPRIETE ATTRIBUER Conforme = VALEUR non POUR
$*pdd~(VJ,VG,VX,VPP,EJ,AJ,EX,AX,LX,MDX,MDJ,LJ)*tt=V$

```

---

*Scénario SATO de vérification de la conformité entre l'étiquetage du TT et pdd*

Le critère de conformité entre les deux propriétés ne correspond pas à une égalité stricte. Certains traits syntaxiques et morphologiques fins, marqués dans les catégories *pdd* sont neutralisés dans les catégories *tt*, calculées par le TT. Les divers filtres utilisés permettent de rendre compte des diverses possibilités. Par exemple, la première affectation à *oui* s'applique au filtre  $*pdd=(VJ,EJ, AJ, MDJ, LJ)*tt=VJ$ , qui s'interprète comme l'ensemble des mots (\$) dont la valeur *pdd* est comprise dans l'ensemble  $(VJ,EJ, AJ, MDJ, LJ)$  et pour lesquels la propriété *tt* contient la valeur *VJ*. Dans le fichier de commandes, l'opérateur  $\sim$  correspond à l'opérateur d'inégalité. La diversité des étiquettes référées dans le scénario illustre bien le nombre élevé d'étiquettes de *pdd* par rapport à *TT*. Par exemple, les étiquettes  $pdd=EJ,AJ,MDJ,LJ$  signifient respectivement *auxiliaire être*, *auxiliaire avoir*, *modaux* et *semi-auxiliaires aller* conjugués.

Muni de ces trois propriétés, *pdd*, *tt* et *conforme*, il devient facile de tracer un portrait de la performance du TT et des erreurs commises. Par exemple, de simples commandes de statistique descriptive permettront de produire les résultats présentés dans ce texte. Ainsi, la commande SATO *PROPRIÉTÉ DÉCRIRE pdd POUR \$\*conforme=non\*tt=VJ* nous fournit le nombre et le pourcentage de toutes les catégories de *pdd* pour pour lesquelles le TT a faussement repéré un verbe conjugué (VJ).

## 6. Validation sur corpus

La validation sur corpus a été réalisée sur deux textes du 13<sup>e</sup> siècle étiquetés dans le projet Modéliser le changement, les voies du français : *La conquête de Constantinople*, une œuvre en prose du XIII<sup>e</sup> siècle de 38.355 mots (ci-après TEXTE 1) et *Aucassin et Nicolette*, une chantefable et vers et prose de 11.747 mots (TEXTE 2). Ces textes ont été étiquetés selon une procédure semi-automatique dans le jeu *pdd*.

### 6.1. Résultats du TEXTE 1

Dans le jeu *pdd* sur les 4775 mots qui ont été étiquetés VJ, 4.481 ont reçu la même étiquette du TT. Le pourcentage de conformité de l'opération de catégorisation par le TT modifié est le

suivant : 94% pour les verbes conjugués (VJ), 95% pour les verbes infinitifs (VX), 71% pour les participes passés (VPP) et 89% pour les participes présents (VG).

Il reste maintenant à examiner les cas pour lesquels le TT a attribué incorrectement une étiquette en commençant par la catégorie VPP (359 occurrences). Les résultats se vérifient dans Tab. 1 où l'on voit que la confusion se produit surtout entre le VJ et le VPP soit dans presque 64% des cas. On peut aussi remarquer que le nombre de catégories impliquées dans la confusion est extrêmement limité, 98% des attributions erronées touchent le verbe, le nom ou l'adverbe.

<i>Occ.</i>	<i>% Occ.</i>	<i>pdd</i>
230	63.81%	VJ
83	22.31%	NCS (Nom commun singulier)
32	8.60%	NCPL (Nom commun pluriel)
14	3.75%	ADV (Adverbe)

Tableau 1 : Niveau de non conformité suite à l'étiquetage du TT pour VPP dans le TEXTE 1

De façon symétrique, l'erreur d'étiquetage des VJ (Tab. 2) (168 occurrences) est dû à une confusion avec VPP quoique dans une proportion moindre. Le nombre de catégories impliquées de façon significative est cependant identique, quatre catégories cumulant 94% des occurrences.

<i>Occ.</i>	<i>% Occ.</i>	<i>pdd</i>
63	37.50%	VPP
43	25.59%	(NCS, NCPL)
34	20.24%	P (Préposition)
18	10.71%	VX

Tableau 2 : Niveau de non conformité suite à l'étiquetage du TT pour VJ dans le TEXTE 1

La totalité des cas d'erreurs d'étiquetage de VX se retrouvent majoritairement sur les noms (17 cas soit 68%) et sur les verbes conjugués (8 cas soit 32%). Pour terminer, 6 occurrences ont été catégorisées VG alors qu'elles ont l'étiquette ADJ dans *pdd*. Dans la conclusion, nous présenterons quelques hypothèses pour éclairer ces résultats.

## 6.2. Résultats du TEXTE 2

Le pourcentage de conformité de l'opération de catégorisation par le TT modifié pour le deuxième texte est le suivant : 78% pour les verbes conjugués (VJ), 73% pour les participes passés (VPP) et 97% pour les verbes infinitifs (VX). En comparant les résultats obtenus sur le premier texte à ceux du deuxième, on constate donc que le taux de conformité est moins fort pour les VJ mais légèrement supérieur pour les VPP et les VX. Nous expliquerons un peu plus loin l'absence de VG.

Lorsque VPP est attribuée de manière erronée (197 cas) chez *Aucassin et Nicolette*, c'est de loin la catégorie VJ qui est en cause.

<i>Occ.</i>	<i>% Occ.</i>	<i>pdd</i>
150	76.14%	VJ
13	6.60%	(NS, NCPL)
9	4.57%	ADV
5	2.54%	ADJ

Tableau 3 : Niveau de non conformité suite à l'étiquetage du TT pour VPP dans le TEXTE 2

Tab. 4 résume la non conformité de l'étiquetage de VJ. Si l'on retranche des 76 cas fautifs les 27 dus à un problème dans l'attribution de la propriété non conforme elle-même, on observe sans surprise la même confusion entre VPP et VJ d'abord et NC ensuite, plus faiblement.

<i>Occ.</i>	<i>% Occ.</i>	<i>pdd</i>
33	43.43%	VPP
7	9.21%	(NCS, NCPL)
4	5.26%	P
3	3.95%	ADJ

Tableau 4 : Niveau de non conformité suite à l'étiquetage du TT pour VJ dans le TEXTE 2

Enfin, les cas où VX (9) et VG (42) sont distribuées à tort, sont aussi dus à VJ en très large part. Pour VX, 6 cas, soit 67% et, pour VG, 55%. Le reste des erreurs se répartit entre ADJ et NCS pour VX. La non conformité dans l'attribution de VG s'explique du fait que dans le texte 2, une nouvelle étiquette, VP, est utilisée dans l'étiquetage en *pdd* ce dont la procédure de conformité n'a pas tenu compte. Cela provoque une distorsion dans le calcul.

## 7. Conclusion

Malgré la complexité de la procédure d'enrichissement du TT et les problèmes qu'engendrent les différences dans la nature des catégories des deux jeux d'étiquettes, la gestion des effets de bord lié au raffinement automatique comme le repérage des erreurs, la variation orthographique, etc., les résultats que nous avons obtenus sur les deux textes médiévaux nous permettent d'affirmer que notre recherche mérite d'être poursuivie.

En comparant les deux textes de notre corpus test, on voit que le niveau de conformité de l'étiquetage du TT est très satisfaisant pour les étiquettes VJ et VX, un peu moins pour VPP. Les cas de non conformité sont assez révélateurs de la performance de la procédure. En effet, la distribution des erreurs d'attribution ne résulte pas du *hasard*, un petit nombre de catégories, sensiblement le même sont touchées, principalement les participes passés et les noms. Il est clair que certaines terminaisons ambiguës en ancien français, par exemples – *é* ou –*és* jouent un rôle parce qu'elles peuvent être à la fois des désinences de noms, de verbes, de participes et d'adjectifs. Dans le texte 1, 319 cas de terminaison en –*é*, –*és* sont non conformes dont 284 proviennent de la confusion dans l'identification du VPP.

126 cas sont des formes verbales conjugués comme *voliés*, *vaurrés*, *volés*, *veuillés*, *pregniés*, respectivement des formes de « vouloir » et de « prendre », ou encore, dans plus de 120 cas, des noms comme *croisié(s)*, *congié*, *abés* ou des adverbes comme *assés*.

Ce dispositif expérimental permet donc de reprendre facilement la procédure sur tout corpus pour lequel on dispose d'une catégorisation *pdd* validée pour un état de langue comparable. En fait, une fois établie la fiabilité du TT adapté à un état de la langue et du système catégoriel, on peut s'en servir comme outil diagnostique permettant d'évaluer les différences entre corpus représentants divers états de la langue. Ainsi entraîné, le TT peut s'avérer un outil précieux pour le chercheur en linguistique historique.

## Références

- Daoust F. (2009). Système d'analyse de texte par ordinateur, SATO, Manuel de référence, version 4.3. Centre d'analyse de texte par ordinateur, UQAM, 2007; modifié en 2009. <http://www.ling.uqam.ca/sato/satoman-fr.html>.
- Kapitan R., Daoust F. and Dupuis F. (2007). Prémhis, une plateforme pour l'analyse morphologique du français médiéval. Communication présentée à CILH-18, Montréal, août 2007.
- Namer F. (2000). FLEMM : Un analyseur flexionnel du français à base de règles. *TAL*, 41 (2) : 523-547.
- PREMHis (2007). <http://www.ling.uqam.ca/forum/premhis/>.
- Stein A. (2007). *Part of speech tagging and lemmatisation of Old French*. <http://www.uni-stuttgart.de/lingrom/stein/forschung/resource.html>.
- Stein A. and Schmid H. (1995). Etiquetage morphologique de textes français avec un arbre de décision. *TAL*, 36 (1-2) : 23–36.
- TreeTagger. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>.

### Liste des textes médiévaux utilisés

- Anonyme (1977). *La Chanson de Roland* [Jonin, P., editor]. Paris : Gallimard.
- Anonyme (1925). *Aucassin et Nicolette*, chantefable du XIIIe siècle [Roques, M., editor]. Paris : Champion.
- Benedeit (1979). *Voyage de saint Brendan*. I. Short, B. Merrilees. Manchester University Press.
- Clari, Robert de (2004). *La Conquête de Constantinople*. Paris : Champion.
- Commynes, Philippe de (1925). *Mémoires, tomes I et II* [Clamette, J., editor]. Paris : Champion
- Marie de France (2003). *Lais* [Rycher, J., editor]. Paris : Champion.