1. Introduction

Ancient French texts are increasingly numerous on the Web. This considerable growth of electronic databases allows the user to familiarize himself with various famous authors. However, for the researcher interested in literary or linguistics aspects, these texts are often being presented in a heterogeneous way, devoid of critical and inquisitive approaches. While this lack of theoretical and methodological insight does not directly affect an easy access to the Ancient French texts, it becomes a major impediment to the thorough study of corpora from a specific literary period.

In this article, I will present two research programs, the first on Old French, the second on vernacular French of the Classical period, which both have a common purpose: a modular study of texts.

II. The Laboratoire d’ancien français

The Laboratoire d’ancien français, directed by Pierre Kunstmann and myself, gathers researchers from
the University of Ottawa as well as researchers from other Canadian universities and international research centres. As stated in its mandate: “it aims at a large broadcasting of texts carried out in accordance to the French linguistics and philology standards and criteria.” Since the foundation of the Laboratoire in 1996, the LFA website has presented a broad spectrum of original texts (approximately twenty) (Old and Middle French manuscripts, original publications for the Renaissance and Classical French language, for instance, Couronnement de Louis, Chevalier au Lion, Miracles de Notre-Dame tirés du Rosarius, Le Diable boiteux) along with modules of interrogation. Two successive research projects, subsidized by the SSHRC, were conducted on Chrétien de Troyes’ Chevalier au Lion; this led to the development of modules of interrogation on this major Old French text.

III. Modular Research Program on Chevalier au Lion

III. A. Manuscripts, images from manuscripts, lemmatized indexes

Chevalier au Lion presents several manuscripts (8) and excerpts (6), some of which exhibit typical dialectal features. Most of these manuscripts were published on the LFA website along with a transcription; two of them, the H manuscript and the Princeton manuscript, are presented with images. The transcript aims to reproduce the original texts accurately and legibly; however, punctuation and word separation have been established according to modern written use. Lines have been numbered according to Guiot’s H manuscript, which we have used as a template. Pierre Sala’s manuscript is also part of this collection. Sala’s text dated a few centuries later (1522) is a rewriting of the original text.

From a linguistic and philological perspective, a possible comparison established between the different manuscripts and images is certainly useful to identify line by line all the variants. Such a comparative study would not have been possible using a standard publication such as Flammarion’s, which, in various aspects, resembles that of Roerster and attempts to “re-establish” Chevalier au Lion “authentic text” through the most well-known manuscripts.

Since the format and storyline remain quite identical in Old French manuscripts and Sala’s manuscript, the latter is very effective to compare, after an interval of a few centuries, the evolution of linguistic phenomena. Let’s consider the beginning of Chevalier au Lion and compare Guiot’s manuscript to Sala’s. Thanks to the manuscripts, we notice a major grammatical change in the first few lines. In Guiot’s, the verb enseigner ‘to teach’ is followed by a subjunctive, correct for this period’s grammatical rules, in spite of the pronoun co-indexation (nous enseigne que nous soyons ‘teaches us to be’). However, in Sala’s, the verb is followed by an infinitive (nous enseigne a estre). During the Renaissance period, the latter grammatical structure was directly competing with the preceding one.
Table 1  Comparison between the H and L manuscripts

<table>
<thead>
<tr>
<th>H Manuscript (Guiot)</th>
<th>L Manuscript (Sala)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Artus, li boëns rois de Brataigne,</td>
<td>Le bon roy Artus de Breaigne,</td>
</tr>
<tr>
<td>2. La cui prœseca nos enseigne</td>
<td>La qui prœseca nos enseigne</td>
</tr>
<tr>
<td>3. Que nos soiens preu et cortois,</td>
<td>Aestre hardi et courtois</td>
</tr>
<tr>
<td>4. Tint cort si riche cume rois</td>
<td>Depuis le chef jusqu'aux cortois,</td>
</tr>
<tr>
<td>5. A cele feste qui tant coste,</td>
<td>Tint court, a une Penteoste,</td>
</tr>
<tr>
<td>6. Qui'an doit clamer la Panteoste,</td>
<td>Qui fut de moult grand pris et coste,</td>
</tr>
<tr>
<td>7. Li rois fu a Cardœul et Gales;</td>
<td>C'estoit dedans Cardœul en Gales;</td>
</tr>
<tr>
<td>8. Après mangier, parmi ces salos,</td>
<td>La furont pleines les grans salos;</td>
</tr>
<tr>
<td>9. Ci chevalier s'atropelerent</td>
<td>De dames et de demoiselles,</td>
</tr>
<tr>
<td>10. La ou dames les apererent;</td>
<td>De chevaliers et de pucales.</td>
</tr>
<tr>
<td>11. Ou dameiselles ou pucales</td>
<td></td>
</tr>
</tbody>
</table>

(Martineau 1996, 1997). Our hypothesis which states that the latter structure is most common in Old French is further reinforced by the fact that all five other manuscripts already transcribed on our website present the same structure followed by the subjunctive for this specific line.

As shown in Table 1, manuscripts are presented line by line. In order to facilitate the reading and to allow for faster downloading, the text has been divided into 'chapters' which correspond to the adventures of Yvain, the protagonist. Clearly, each manuscript has been treated separately. Therefore, for the user interested in variants, this comparative approach could become a tedious process. We have consequently teamed up with Kaja Meyer from the University of Copenhagen who has identified all the different versions line by line, in accordance with the manuscripts. Table 2 displays the first line.

For the philologist, such comparative studies (K. Meyer's work in particular) are very valuable. However, for the linguist who is interested in morphosyntax and vocabulary, the comparative work remains quite tedious. A basic lemmatized index (with grammar indication [verb, noun, adjective, etc.] occurrence [forms and form frequency] such as the one created by P. Kunstmann for the H manuscript, is still not efficient enough for extensive grammar analysis; such a tool would not highlight the variations to be found in the different manuscripts of the Chevalier au Lion.

We therefore had to design a module which contained not only the qualities of a lemmatized index and those of K. Meyer's work, but that would also integrate more complex parameters. This module would have to include:

- a morphosyntactic analysis of essential verbs and their arguments,
- a word order study,
- a comparative study of the verbs in the different manuscripts of the Chevalier au Lion,
- an interlinking with other different analysis modules.
In order to achieve these goals, we needed to use different software programs; the first step with FileMaker Pro has been completed; the second, with SATO WEB, is still in progress.

III. B. Base d'analyse verbale du Chevalier au Lion

The Base d'analyse verbale du Chevalier au Lion was built using FileMaker Pro 5, a software program that is designed to create relational databases. In association with Direx enterprise, we adapted this software in order to create a user-friendly interface on the Web. Verbs were selected according to their semantic categories (for instance, volition verbs or affirmative verbs); many categories remain to be added to the database (auxiliary verbs être and avoir and factitive verbs faire and laisser). Each verb occurrence is linked to its own file displaying, in addition to written form and lemma, morphosyntactic features: tense, mood, person, voice, personal/impersonal/possessive construction. The line number of the occurrence on each file leads the user to the transcript line, sparing him a link between the verbal analysis database and the manuscript.

The user can either perform single queries:

Find the lemma ‘cuidier’
Find all imperfect verbs

or multi-variables queries:

Find ‘cuidier,’ third person, indicative mood
Find ‘cuidier,’ present tense, subjunctive mood

This latter type of query is especially helpful for verbal variant written forms analysis and lemmatization. FileMaker Pro enables the user to add a certain number of criteria on each file. We used this feature to integrate a comparative analysis of verbal behaviour in manuscripts. Since we used the H manuscript as our template, we added a link leading to it on every file.
Thus, it is possible to compare variants in each manuscript studied so far (H, V and P manuscripts). One can consider the lexical variants such as the alternation between the synonyms *cider*, *penser*, and *croire*, the morphosyntactic variants such as the alternation between the imperfect tense and past historic tense or the graphic variants. Hence, we note in the *Chevalier au Lion* that Guiot’s H manuscript has a much greater graphic form stability, whereas the P manuscript contains greater variation (read Pignatelli 2003 for an in-depth study of the subject, based on another Chrétien de Troyes text, *Chevalier à la Charette*, analyzed with the Tustep software).

We have also integrated certain syntactic elements in the file, including valency analysis (realized or maximal), the presence of other complements, the reduplication of the complement by a pronoun and word order.

Maximal Valency: 2
Realized Valency: 1
Other complements: no
Reduplication: no
Excerpt:
*Ne cuit qu’an plain ne an boschage
Puisse an garder beste sauvage,
N’en autre leu, por nule chose,*

The *Base d’analyse du Chevalier au Lion* was designed in such a way as to allow the user to establish links with other modules, either through searching by reference line or by lemma. Two other FileMaker Pro platform databases were created for other projects: the *Base d’analyse des formes verbales du premier miracle de Notre-Dame par personnages*, created by Lene Schoesler and the *Base d’analyse du discours rapporté dans le Chevalier au Lion*, created by Danielle Forget and France Martineau.

For example, let’s consider the reported speech analysis. It was shown in Forget and Martineau (2002) that the verb *dire* ‘to say’ is rather stable in reported speech occurrences since the morphosyntactic properties seldom vary and the verb is mainly conjugated to the present tense and to the third person. Moreover, the indirect object is often stated with this verb in reported speech occurrences, therefore emphasizing the interlocutor (*si li a dit* ‘and [she] said to him’).

*De l’ amor qui en luir s’est mise,*
*Le trova trespansé et vain;*
**Si li a dit:** «Messire Yain,*
*Quel siege avez vos puis eit?»
*– Tél, jet il, qui mout m’ a pleü.*
(vers 1548–1552)
To what extent, though, is this behaviour typical to the verb *dire*? How could one describe the use of this verb in such reported speech structure in relation to its general usage at this period? A basic verb behaviour comparison (through the intermediary of the lemma) using both the *Base d’analyse du discours rapporté* and the *Base d’analyse verbale du Chevalier au Lion* might be sufficient to state two facts: firstly, apart from when it is used in direct speech, the verb *dire* can be used with all tenses, moods, and persons, and, secondly, in cases in which three arguments could have been realized (subject, direct object, indirect object), the indirect object argument is always realized with *dire*, even to the detriment of the subject realization (see Table 3).

One could then consider whether the behaviour of the verb *dire*, which is frequently used with three arguments, is either characteristic of a certain period (i.e., the Old French) or of this particular work (the *Chevalier au Lion* manuscript). To answer this question, one would initially have to compare this verb in the *Base d’analyse verbale du Chevalier au Lion* to other manuscripts of the same period or even to compare the behaviour of this verb in Sala’s manuscript (a work written some centuries later) or in the *Base du premier miracle* dating from the Middle French. Thus, this latter comparison presented in Table 3 shows that the verb *dire*, contrary to what is found in the *Chevalier au Lion* (53.4%), seldom develops its maximal valency in this text (11.7%). In other words, the major change from Old French text *Chevalier au Lion* to Middle French text *Premier Miracle* is a less frequent development of maximal valency (3 arguments) to the advantage of a valency reduced to two arguments (direct object/indirect object and less frequent subject/direct object) and even one argument (direct object).

Then again, what about the element realized by this verb in a reduced valency? As Table 3 shows, the subject is often not realized in *Chevalier au Lion*, as it is expected since Old French is a V2 language (where the verb must be located in second position and the subject could be optional); thus in a two-term realized valency, the complements (direct and indirect) are more often realized than the subject. However, in *Premier Miracle*, the subject is more often realized in a two-term realized valency, following the language change to a subject-verb-object (SVO) pattern in which the subject is obligatory. One should also note the compulsory expression of the direct object, in Old French as well as in Middle French, a regular usage for this period (Schoesler 1999).

Lexical or morphosyntactic properties can be easily classified, and, as a matter of fact, lend themselves well to a query. However, an analysis of the nature and expression of the verb arguments (necessary to an analysis such as in Table 3) with FileMaker Pro is not so effective. How is it possible to determine which arguments are realized for a verb that has a maximal valency of three and carries out no more than two arguments? Contextual information such as the verb argument functions and their sequence in the sentence are poorly
Table 3  Comparison of the arguments development in a three-argument maximal valency with the verb dire (‘to say’) in Old and Middle French texts.

<table>
<thead>
<tr>
<th></th>
<th>1 (Direct object)</th>
<th>2 (Direct object/Indirect Object)</th>
<th>2 (Subject/Direct Object)</th>
<th>2 (Subject/Indirect Object)</th>
<th>3 (Subject/Direct Object/Indirect Object)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chevalier au Lion (Old French)</td>
<td>–</td>
<td>46.5% (40/86)</td>
<td>–</td>
<td>–</td>
<td>53.4% (46/66)</td>
</tr>
<tr>
<td>Premier Miracle (Middle French)</td>
<td>20.5% (7/34)</td>
<td>41.1% (14/34)</td>
<td>26.4% (9/34)</td>
<td>–</td>
<td>11.7% (4/34)</td>
</tr>
</tbody>
</table>

managed by FileMaker Pro. In order to recover this information somehow, we have incorporated an entry which specifies the word order of the major components and puts in parentheses arguments not realized, as shown below:

Word Order: Adverb (Subject) Verb – Direct Object – Indirect Object

This entry is treated the same way as text. Common queries such as “Find all verbs with a deleted subject followed by a direct object and an indirect object” provide good results, given that the linear sequence corresponds exactly to the query (without interfering elements such as an adverb, for example). Since it was impossible to create a file that would provide a more precise contextual query and would remain quite simple, we turn towards SATO WEB.

III. C. SATO WEB: For a contextual analysis of SATO software allows the user to submit contextual queries with more ease. We are currently working on the Base d’analyse verbale du Chevalier au Lion sur SATO, which will recover the morphosyntactic features defined in the Base d’analyse verbale FileMakerPro. The established lexicon will then be projected on the text in order to encode the verb’s arguments (their realization, their function, and their location in relation to the verb). For instance, the text’s occurrence of the verb cuer ‘to think’ is defined in the Lexicon by means of the information recovered in the Base d’analyse verbale FileMakerPro, as illustrated below:

**Text:**

*Ne cuiet qu’an plain ne an boschage*

*Puisse an garder beste sauvage,*

*N’en autre leu, por nule chose,*
**Lexicon:**

*Cuit*: lemma *cudier*, first person, present tense, indicative mood, active voice, personal construction.

The contextual information is then directly indexed over the word *cuit* in the text:

**Contextual information:**

Maximal Valency: 2  
Realized Valency: 1  
Subject: not realized  
Complement: realized, direct object, subjunctive clause, postverbal

The interaction between the three modules enables the user to submit a query on word order in the indexed text, starting either from the function or the grammatical nature of the word:

- Find the lemma *cudier* followed by a complement
- Find a verb without a realized subject
- Find the sequence Adverb-Verb

**IV. A Corpus of Vernacular French of the Classical Period**

**IV. A. Linguistic micro-variation and epistolarity in Nouvelle-France**

The research program that I am currently directing on *Evolution et variation en Français du Québec, du XVIIe s. au XIXe s.* is based on a collection of vernacular French corpus from North-West of France, Québec, and Acadia. Unlike the linguistic atlases, which examine the dialectal French of modern speakers and above all study pronunciation and vocabulary, our corpus had to be based on written texts of a relative length and informal nature, since we study grammar. It was imperative that the writers of these works would have no linguistic insight over their production. In other words, these texts had to be written spontaneously and devoid of all thought on the language itself.

Vernacular Ancient French corpora that are currently accessible to users are essentially literary ones (Ayres-Bennett 2000, Lodge). Given that the writer has a normalized perspective of his characters, the integration of most vernacular French elements is constantly diffracted: literary writers shall either caricaturize the familiar features or ultimately ignore these when they are not over-stigmatized. Based on such corpus, the overview of vernacular French is consequently fragmented and can never exemplify the interacting phenomena in a given system.
Since 1998, I have been building up a corpus of vernacular French epistolary texts from public and private archive collections. These texts were written by members of the same family or by soldiers with little education who lived in Québec, Acadia, and regions of North-West France during the seventeenth, eighteenth, and nineteenth centuries. The collection now consists of more than five hundred letters. To the best of my knowledge, this type of corpus is a first. It allows us to diachronically study variations in vernacular French and to compare dialectal variants of a given period.

Unlike literary texts which have designs on aesthetic quality and are intended for publishing, family and private letters imply a certain degree of informality. Epistolary genre’s style is also far less coded than any other writing genre, including administrative/commercial texts such as minutes reports or accounts books which all are addressed to general public.

Electronic data capture is almost complete. This step was rather complicated because of the material aspect of the letters which were often unreadable, given the calligraphy of less educated people and ink bleeding. We have now published four texts on our website which were authorized by the archive centres. The online version contains the images from the booklet, the original text’s transcript, and a modern orthography transcript.

Unlike the protocol followed by the Laboratoire de français ancien for Old and Middle French texts transcription, we have not standardized the word separation, punctuation (almost absent), and capital letters (frequently contained in the middle of the words). These elements are significant indications to help us comprehend these less-educated writers’ graphic system. Moreover, we wanted the transcript, including the lines disposition, to be very faithful to the booklet’s image.

IV. B. Lemmatization

The grammatical phenomenon analysis soon ran into problems related to unstable orthographic variants. Unlike Old and Middle French texts for which the written variants are well known and established, our corpus’s texts do not follow a known orthographical variation system. Two different research examples illustrate difficulties inherent in this research. In Martineau and Mougeon (2003), where we studied the deletion of the negative particle ne, simple location of this particle was rather complicated since it would often agglutinate with the next word when it started with a vowel (i.e., ncut as ‘did not have’, invité pas “n’invitai pas” as ‘did not invite’). We therefore had to search, line by line, through the text, due to the strong probability that the particle ne would agglutinate to any number of words starting with a vowel or even that the agglutinated particle ne would correspond instead to an en pronoun (i.e., nan né hu ocune ‘I have had no one’). In Déprez and Martineau (2003), we examined the occurrence of the indefinite adjective aucun ‘any’ competing with negative adverbs such as pas or plus (i.e., navais plus le droit de parler à aucun Matelau ‘We no longer
Table 4  Transcript of the original text and modern orthography transcript of an Acadian text (19th century).

<table>
<thead>
<tr>
<th>Transcript of the original text</th>
<th>Modern orthography transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cher père Cet a présent que ge</td>
<td>Cher père c'est à présent que je</td>
</tr>
<tr>
<td>pran le plesir de vous écrire</td>
<td>prends le plaisir de vous écrire</td>
</tr>
<tr>
<td>de mes nouvelles que ge sui an</td>
<td>de mes nouvelles que je suis en</td>
</tr>
<tr>
<td>bono santo dieu marci et g espêre</td>
<td>bonne santé dieu merci et j'espère</td>
</tr>
<tr>
<td>que vous lette aussi bien que moi</td>
<td>que vous êtes aussi bien que moi</td>
</tr>
<tr>
<td>Et tout leçipage le temps et bien</td>
<td>Et tout l'équipage le temps est bien</td>
</tr>
<tr>
<td>mauvait, il vante tout les deux</td>
<td>mauvais, il vante tous les deux</td>
</tr>
<tr>
<td>ou trois glour tré pour pecher</td>
<td>ou trois jours trop pour pêcher</td>
</tr>
<tr>
<td>quand nous on arives la morue</td>
<td>quand nous on arrive la morue</td>
</tr>
<tr>
<td>etait bien rare mais ge comman-</td>
<td>était bien rare mais je commençons</td>
</tr>
<tr>
<td>11. son a prendre courage pour</td>
<td>à prendre courage pour</td>
</tr>
</tbody>
</table>

had the right to talk to any sailors'). The numerous variants of a basic word such as *aucun* (i.e., *ocun, auqun, oqun, auqn*), in addition to erroneous word spacing (i.e., *ocun*), once again compelled us to read the text in full.

We have attempted to run an automatic semi-lemmatization software program (Tree Tagger) for Modern French, created by Achim Stein from Stuttgart University. The results of the lemmatization were disappointing: a success rate less than 32.4%. Due to the fact that this software also decodes the clause’s structure, there is a direct effect on the categorization of a specific word when the program does not accurately identify the grammatical category of the preceding word. For instance, since the computer software identified the *ge* ‘I’ pronoun as a verb, *pran* ‘take’ would then be interpreted as an adjective.

```
queCON:sub que
geVER:pl  <unknown>
puranADJ:masc:sg  <unknown>
```

We could certainly improve the efficiency of the software program by integrating several grammatical rules which would take into consideration some of the regularities found in texts which are written using orthography that is not standard. First of all, the complete absence of all French accents in this text presents a basic problem. Since French accents differentiate very few words (i.e., past historic and imperfect subjunctive third person) (Catach 1980) one would only have to indicate all possible variants to the software, with or without accents, for a given word. The word *était* ‘was,’ without an accent, would therefore be recognized in the lexis as the verb *être* ‘to be.’
Some well-known phonetic rules which cause identification problems can easily be integrated into the software program: the “a” in marci ‘thank you’ as well as in perdre ‘to lose’ refers to the open sound of the /e/ vowel followed by /r/. Moreover, problems originating from silent consonants and vowels can easily be solved: one would simply have to list all words containing silent vowels or consonants and then determine all possible variants for these words (i.e., troy ‘too much’, che ‘at’, moru ‘cook’).

It is possible to manage all of these orthographic variants using a lemmatization software program as long as one can formulate proper rules. In addition to this lemmatization purpose, we attempted to verify the ways in which rules’ identification would make it possible to reveal the spelling strategies of less-educated people.

IV.C. Spelling strategies and learning software program

Within the context of the interdisciplinary research program we are currently developing (Martineau, Desrochers and Morin), software programs will assist us in identifying all graphemes/archi-graphemes in written forms of less educated people, as well as in evaluating them in relation to French phonemes of various periods in language evolution insofar as it is possible to reconstruct them. Thus, we shall be able to attest the coherence of this reconstruction in linguistic history and verify the consistency of specific patterns which were ingrained in the writing skills of less-educated people when schooling was seldom provided to lower social classes during the seventeenth, eighteenth, and nineteenth centuries in French-speaking Canada (Corbeil 1976).

Several stages underlie this research (from locating forms to implementing rules). A first step aims at the creation of an equivalence grid between pronunciation, the word spelt by a less-educated writer, the standard/modern written form, and that of a particular period.

This manual lemmatization should be carried out parallel to the learning of certain rules and automatism by the software, which shall then be projected onto other texts, with special attention to grammatical orthographic learning scenarios.

As such, it will be possible to verify the use of the “s” to signal the plural form in on arrêtes ‘we arrive’ or even the imperfect inflection which is already invariable in singular or plural forms of certain texts, as shown in Table 6, in this Québec’ writer of the nineteenth century.

The drawn-out patterns allow us to evaluate the extent to which some of the processes associated with alphabetization vary depending on periods and dialects. In addition, these patterns allow us to eliminate certain ambiguities. For example, let’s consider the verb arrêtes in Table 4, line 9. Writers often use the e grapheme to indicate the schwa, the closed e, or the open e. In the first instance, it would indicate a present tense verb, third person singular (arrive ‘On arrive,’ which means ‘We arrive’). In the second, the e grapheme would indicate a historic past verb, first person singular (arrivai ‘I arrived’) and in the third case, it would be an im-
Table 5  Equivalence grid

<table>
<thead>
<tr>
<th>Pronunciation</th>
<th>Word as it was spelt</th>
<th>Standard spelling for that period</th>
<th>Standard Modern spelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>[prā]</td>
<td>pran</td>
<td>prends</td>
<td>prends</td>
</tr>
</tbody>
</table>

Table 6  Imperfect inflection, third person, singular and plural

<table>
<thead>
<tr>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>avaïs</td>
<td>Continuest</td>
</tr>
<tr>
<td>Esperais</td>
<td>anetais</td>
</tr>
<tr>
<td>Etaïs</td>
<td></td>
</tr>
</tbody>
</table>

perfect verb, third person singular (arrivait “On arrivait,” which means ‘We were arriving’). Since the second instance undoubtedly corresponds to a first-person singular conjugated verb, it must be excluded because the occurrence is third person as shown by the pronoun on ‘we’ (see note 12). However, can we determine whether the verb is imperfect or historical present? We might come to the conclusion that it is the latter alternative due to the fact that another historical present verb precedes this one (vante). Moreover, a more careful analysis of the spelling regularities clearly demonstrates that this Acadian writer follows a similar pattern to that of the Québécois writer shown in Table 6: imperfect and conditional, which are both pronounced with an open e are spelled with an a (était ‘was’ voudrais ‘would like’) while the present (which could also be pronounced with an open e when conjugated to the third person) is spelled with an e by this Québécois writer (et “est,” which means “is”). Thus, it is very likely that the verb arriver in the Acadian text (Table 4) is in the present tense, since e indicates a schwa as well as a present tense ending.

V. Conclusion

The unabridged version of a text and its most faithful transcript shall remain the key of literary and linguistic analysis. However, this could easily be supplemented with analysis modules in order to enlighten the user about the grammar, vocabulary, variations in written form, and themes of the text. Analysis modules operating on a same software platform as well as FileMaker Pro databases are indeed useful tools to perform grammatical categorization tasks. However, their efficiency is limited when operating with contextual analysis.
Lemmatization remains a necessary step to several of these analyses. While this step is more or less facilitated by Old and Middle French dictionaries, working with vernacular French remains very complex. It is therefore essential to elaborate new software programs that can expose certain regularities throughout the learning of non-standard rules.

References


Notes

1 This research received funding from two SSHRC grants (Martineau & Vinet; Kunstmann, Martineau & Forget).
Ancient French refers to Old French (9th–13th c.), Middle French (14th–15th c.), French of the 16th c. and Classical French (17th–18th c.), following the Laboratoire de français ancien.

Since the paper's presentation in 2002, some major developments occurred on both Web sites. P. Kunstmann, with Achim Stein, integrated Chevalier au Lion text in a new lemmatized database which could better handle the lemma with basic grammatical information. F. Martineau and her team were subsidized through a MCRI (SSHRC) grant Modéliser le changement: les voix du français (2005-2009) (www.voices.uottawa.ca). The purpose of this MCRI project is to develop a major corpus of texts, from Medieval French to 16th c. French to Canadian French (17th c. to 19th c.); this 5-million-word corpus will be completely lemmatized and parsed (i.e. with annotated syntactic structures). Contextual queries such as the ones exemplified in this paper with Chevalier au Lion are well handled in such a parsed corpus. We decided to integrate the Base d'analyse verbale du Chevalier au Lion in this parsed corpus. Written variant linked to vernacular letters are also handled in a program developed with 4D in the project Évolution et variation en français du Québec, du XVIIe au XIXe siècle (Martineau, Desrochers and Morin; SSHRC, 2004-2007).

Pierre Kunstmann did the transcriptions. The manuscripts with a transcription are Fr. 794 (H); Fr. 1433 (P); Fr. 12560 (G); Chantilly, Musée Condé 472 (A); Vatican, Christine 1725 (V) and Princeton, Garrett 125 R.

With Daniel Labonia, a research assistant.

G. Ernst (2002) published a CD-ROM consisting of a chronicle from the French town of Lille (17th century) and of three other texts from Paris (17th and 18th centuries).

On, which can be translated as 'we,' is morphologically third person singular.
In recent years, the application of computing technology to the arts and humanities has been a topic of increased focus in the post-secondary environment. With growing understanding of how these applications can serve the ongoing mission of humanities research, teaching, and training, technology is playing a larger role than ever before in these disciplines. Arising in part from a joint venture between the Consortium for Computers in the Humanities/Consortium pour ordinateurs en sciences humaines (COCH/COSH: now SDH/SEMI, the Society for Digital Humanities/Société pour l'étude des médias interactifs) and the Social Sciences and Humanities Research Council (SSHRC), Mind Technologies is the first volume to broadly document the internationally significant work of the Canadian academic community in the area of humanities computing. Edited by Raymond Siemens and David Moorman, Mind Technologies features contributions from those whose agencies provide research funding for such endeavours, from those whose institutions foster work in the area, and from innovative researchers whose work both reflects and has shaped inquiry into this rapidly growing field.

RAYMOND SIEMENS is Canada Research Chair in Humanities Computing and Associate Professor of English at the University of Victoria, as well as President (English) of the SDH/SEMI. Director of the Digital Humanities Summer Institute and founding editor of the electronic scholarly journal Early Modern Literary Studies, he is also author of a number of studies focusing on areas where literary studies and computational methods intersect, is editor of several Renaissance texts, and is co-editor of several book collections on humanities computing topics, among them the recent Blackwell Companion to Digital Humanities (with Susan Schreibman and John Unsworth).

DAVID MOORMAN is a Senior Policy Advisor with the Social Sciences and Humanities Research Council of Canada. His responsibilities include developing policies and support programs for research infrastructure. Dr. Moorman holds a PhD in History from the University of Ottawa.