VariaLog : how to locate words in Early Modern Stages of French and English

Marie-Hélène Lay
marie-helene.lay@univ-poitiers.fr
Jean-Louis Duchet
jlduchet@univ-poitiers.fr

† *Linguistics Department, University of Poitiers, France*
EA 3816 FORELL, Université de Poitiers
et Centre d'Études Supérieures de la Renaissance
UMR 7323 CNRS, Université François Rabelais de Tours

Abstract: *The efficiency of search engines is based on the principle that the information searched can be retrieved by "looking for words" conveying the information. Which amounts to taking for granted that words are always written in the same way. This view, which is well adapted to texts produced in contemporary periods of language history, is not suited to texts produced during the earlier stages of the language, be it the French of the Renaissance, or the Early Modern English of the same period.*

After due consideration of the strategies based on text annotation, we will focus on a tool using query expansion, in order to locate forms, while being insensitive to spelling variation.

Another purpose of the paper is to show the relevance of resorting to linguistic expertise in order to generate the forms to be searched in texts.

This paper describes VariaLog, a tool first elaborated for the French Renaissance context, and the adaptation we have made to Early Modern English.

Keywords: *French Renaissance, Early Modern English, Spelling variation, Search engine, Query expansion, Virtual Library .*

I. INTRODUCTION : SEARCHING WORD FORMS IN A PRE-ORTHOGRAPHIC CONTEXT

The efficiency of search engines is based on the principle that the information searched can be retrieved by "looking for words" conveying the information and that these words can be identified thanks to the string of characters they are comprised of. This view takes for granted that the words are always spelt in the same way and that they comply with orthographic rules.

This view may well be suitable for texts which are produced in contemporary stages of language history, and which correspond to the vast majority of texts available in digitized form. Such is not the situation which prevails for the texts produced during the French Renaissance period and the Early Modern English period. Therefore the availability of older texts for purposes of archiving and disseminating the cultural heritage tradition raises a particular problem.

This context induces us to retrieve the exact quotation in an original version of a work : in order to do this, the need is to bridge the gap between the different forms of spelling: e. g. Montaigne's well known question *Que sais-je*, spelt *Que sçay ie.*

Texts edited in French before the 18th century are characterized by an irregularity in spelling which raises obstacles to an efficient use of search engines: spellings are not consistent, as "proper" spelling has not been "invented" yet. One and the same word may therefore be spelt in a variety of forms. This is not only a variation in time, as would be expected from the evolution of the language between the 15th and the 17th century, but in one and the same book many different spellings may be identified for the one and the same word. For example, one may find *un* or *ung*, and for the word *côté*, either *coté, cotté, cote, costé,* or *couste*, etc. The verb *savoir* may be spelt either *scavoir* or *sçavoir*, « *je sais* » may be spelt « *ie sçay* », and its past participle « *su* » may appear as « *sceu* ».

It is therefore necessary to adapt search engines based on word form identification if they are to render the service expected : our aim is to spot all the written forms which could correspond to a query. Several strategies can be envisaged and the purpose of this paper is to focus on those which resort to linguistic expertise, either injected into the documents themselves (by annotation) or into the search engine (by query expansion). The solutions considered are produced in the particular context of the Virtual Humanistic Library Project and its evolution. The part of the project we describe here, VariaLog, is supported by a Google Digital Humanities Research Award.

II. THE VIRTUAL HUMANISTIC LIBRARY CONTEXT

A. Books and tools

The Virtual Humanistic Library project (VHL) is run by a research unit at the Centre d'Etudes Supérieures de la Renaissance (Center for Renaissance Studies) at the University of Tours, France. This virtual environment aims to disseminate the cultural heritage of the French

Renaissance period.

The digitization project, which was started in 2003, (http://www.bvh.univ-tours.fr; Demonet & Lay, 2008) is the continuation of an editorial project which produced in 1995 a first version of the Epistemon database and an "electronic edition" of Rabelais' s *Works*, (Demonet, 1995), in which the text was already included in a lexicometric environment called Hyperbase (Demonet, 1996), using new tools, available in an electronic environment, to improve the reading and interpretation of literature.

The BVH/VHL project offers two types of digital representations: the image of a copy (its "facsimile"), and its corresponding transcription; 483 books or manuscripts of the Renaissance period (out of a total of 700 digitized works), and 31 transcribed texts are currently online.

Several tools are available on the BVH/VHL website, such as the XTF search engine (http://xtf.cdlib.org/documentation/programming-guide), a software package developed by the University of California. The reader can access the text as he would in a real library, by finding the book and read it. He can also enjoy the possibility of accessing the content of the book, viz. linguistic information (thanks to PhiloLogic, a search engine developed by the University of Chicago, www.artamene.org/philologic.php), or graphic information (thanks to IconClass, http://www.iconclass.nl/).

The database currently gives access to 21000 pictures extracted from the books, whether they be portraits (300), dropped initial capitals (14000), tailpieces, headpieces or illustrations (3000).

In order to achieve its goals, the BVH/VHL project also develops its own tools:

(1) The AGORA project, a graphic analyzer for rare books, which performs the analysis of the page structure and the extraction of all the connected components of each page (Ramel, 2006);

(2) The RETRO project, (REconnaissance et TRanscription par Ordinateur) dealing with Optical Character Recognition for old printed texts, the critical point being that a given font may have been used just for one book (Ramel, 2008);

(3) The "Renaissance TEI protocol" (2009) for the XML encoding of texts (http://www.c-tei.org, http://www.bvh.univ-tours.fr/XML-TEI/index.asp.

(4) The DissimiLog project, a prototype of I/J and U/V normalization tool already prototyped in Tours and Poitiers, with a set of rules and specific dictionaries, dealing with the old usage of i/j and u/v alternation: *vne iambe* morphing to *une jambe*, *viure enyure* morphing to *vivre enivré*;

(5) The Analog project, (Lay, 2010), a tool for lemmatization and morphosyntactic tagging. This tool, developped in Java by François Raynaut, is freely available. AnaLog also provides a concordancer well suited for literary purposes: it helps formulate complex requests with annotated texts and to perform comparisons of several texts.

B. The normalization of spelling variation

The BVH/VHL context considered here is that of a highly expert environment, of a relatively moderate size, aiming at a complete editorial treatment and the dissemination of annotated and validated resources.

Within this context, two solutions have been designed, which are being put to use in the current work.

(1) One possibility is to enrich texts with linguistic information gained from lemmatization and morphosyntactic tagging. The forms retrieved, whatever their spelling, are lemmatized under a canonic form which then becomes the pivot of further requests: for example the lemma for *nuit* groups together forms like *nuit* (which is "regular French") or *nuyctz* (old spelling).

A first solution, Humanistica (Lay, 2000) was based on the adaptation of a probabilistic tagger/lemmatizer. The results achieved were rather satisfactory but the adaptation of the analyzer had to be started all over again to take into account the specific features of almost every text and it was soon found out that the size of the corpus could not allow the automatic treatment of learning procedures. The tuning could only take place by the creation of rules and the adaptation of lexical resources. Moreover, the editorial rules of the web site required careful and consistent proofreading. The mistakes detected in the analyzer's output could hardly be edited by human readers expert at Renaissance studies : they are not necessarily aware of computer procedures.

Another solution was therefore necessary, viz. the development of an environment helpful at all stages of corpus observation, lexical resource creation and text annotation. This tool, called AnaLog (Lay, 2010a, 2010b), is currently being used in the framework of a new "enriched" virtual library.

But as is well known the enrichment of text through linguistic annotation is a slow and costly process. Though this solution is very useful to go on producing slowly but safely a reference environment, it is nonetheless desirable to avail oneself of efficient research tools on corpora of texts already available but not annotated yet.

(2) One other possibility is to use "Query expansion", without requiring the lemmatization process. The aim is not to produce exactly the right forms (like in EEBO -VosPos-, Impact, ToTrTaLe, LGeRM, or for Old Czech, or Old German projects). We will do so, in order to help in an editorial process (DISSIMILOG), but here, we just want to spot all the written forms which could correspond to a query, being insensitive to variation, without requiring the lemmatization process.

C. Identifying variation through query extension

There are more and more texts which are "ready", if one is content with a rather low level of enrichment, and there are more and more digitized documents which have hardly been edited and which are nonetheless relevant for consultation and dissemination purposes.

The setting up of virtual libraries cannot be reduced to a mere transposition of real libraries: the amateur and the expert alike have new requests in terms of access to the information content. In a real library the book may be consulted by turning its pages. The scanned image of a book provides the same facility. But in a virtual environment the reader expects to be able to use the tools which are usually available to access the content of the document through functions based on the retrieval of character strings, as is the case with the XTF platform, for example. But this is of very limited use for Renaissance texts as shown above. The character strings aimed at are in fact all those that correspond to the intention of the query : this is what happens when the form searched has only one spelling.

To solve this problem one has to go back to observational evidence, viz. the texts which are the targets of searches. The variability they exhibit must be precisely measured. Two directions may be taken in this respect: either observe the texts or observe the variants attested for a given form.

(1) Concerning text observation, the aim is to evaluate for a given text the number of forms which do not correspond to the norm. Moreover one must take into consideration the extent to which the texts can be compared. We intend to illustrate this with two short extracts from Montaigne and Rabelais, two authors of paramount significance.

(2) Concerning the observation of variants attested for one word, the idea is to formulate the rules which govern the "production" of abnormal forms. That will be shown on a small sample of forms. We will then put forward a strategy to produce rules to expand the query, turning the search of a word into the search of all the forms assumed by this word, and match the result with forms in texts

III. CHARACTERISATION OF EQUIVALENT FORMS

A. Observational evidence

The observation of texts with spelling variants dating back to a time before the stabilization or generalization of prescriptive spelling helps perceive how frequent the phenomenon was.

Here is an extract from Montaigne, in which the "unexpected" forms (for a French-speaking contemporary reader) are printed in bold:

« De la **coustume** & de ne changer **aisement vne loy receüe**.

Celuy me semble **auoir tres**-bien **conceu** la force de la **coustume**, qui premier forgea ce conte, **qu'vne** femme de village ayant **apris** de caresser & porter entre ses bras **vn** veau **des** l'heure de sa naissance, & continuant **tousiours a** ce faire, **gaigna** cela par **l'accoustumance** que tout grand **beuf** qu'il **estoit**, elle le **portoit** encore. Car c'est a la **verité vne** violente & **traistresse maistresse d'escole**, que la **coustume**. Elle **establit** en nous peu a peu a la **desrobée** le pied de son **authorité**: mais par ce doux & humble commencement l'ayant rassis & planté **auec l'ayde** du temps, elle nous **decouure tantost vn furieux** & **tirannique** visage, contre lequel nous **n'auons** plus la liberté de **haulsser** seulement les yeux. »

For the sake of comparison, a short extract from Rabelais's *Quart Livre* is quoted below. Montaigne's text is easily deciphered by the Francophone contemporary reader with no special expertise whereas Rabelais's text is much more difficult to understand, even in the dissimilated form here adopted (i.e. with the <i> and <j> and the <u> and <v> one expects in contemporary French : *iamais* appears as *jamais*, *vn* as *un* etc., thus creating islands of confidence on which the mind may rest). For each of the words in bold the difference with the contemporary form is much larger than in the text above. In fact, except the "dissimilation point", the forms extracted from Montaigne's text could be mistaken for spelling errors as found in schoolchildren's papers, which is not the case with Rabelais's text:

« Vous **estez deuement adverty**, Prince **tresillustre**, de quants grands **personaiges j'ay esté**, et suis **journellement** stipulé, requis, & importuné pour la continuation des mythologies Pantagruelicques: **alleguans** que plusieurs gens **languoureux**, malades, ou autrement **faschez** & **desolez avoient** a la lecture **d'icelles** trompé leurs **ennuictz**, temps joyeusement passé, & **repceu alaigresse** & consolation nouvelle. »

The conclusion reached after the close study of a substantial set of texts is that the texts themselves comprise a highly unstable environment, hardly compatible with knowledge acquisition strategies based on statistical regularities. The human reader, however, eventually adapts herself. The forms remain floating for a time in front of the reader, letters are combined and recombined and finally the possible "parolles suspendues" eventually reach their final expected form.

There are undoubtedly structures which help to interpret the text and which are based on a "regularization" of data and their alignment with contemporary familiar forms. But one cannot see them just by reading texts. What is needed is a given set of the forms imagined among which the correct form has been identified.

The idea is therefore to observe the lexical items and to detect all the forms they may take, in order to make regular patterns more visible. Here are some of them:

.

vices →	vyces, visces
témoignage →	Tesmognage, tesmoignage, tesmoignaige, tesmoinaige, tesmoingnage, tesmoingnaige, tesmongnage
souverain →	Souverein, souuerain, souverain, soulverain, soulverein, souverayn, soverain, soverein, souvrain, sovrain

Table 1. Equivalent forms attested

If one is to observe the word *souverain* for example, it is possible to identify among the variants encountered a certain number of regular patterns concerning the alternations observed:

- between *i* and *y (vice/vyce), u* and *ul (autre/aultre)*

- or also *a* and *e* in certain contexts. It is impossible to test all the forms supposing that *a* and *e* are interchangeable but it is possible to identify contexts such as *{a,e}|{i,y,n,m}* for example,
  - *ay* may be the same as *ey*, (but also as *ois; : seray/serois/seroye)*
  - *en* and *am* may be equivalent;
  - one may also find this phenomenon with certain prefixes such as *a/ad (avis/advis),* etc.

The observation of long lists of examples gives the impression of a dense jungle of possible targets, the combination of possible substitutions skyrocketing as the words get longer... One may have the feeling that almost everything is possible: not only *y* for *i* or *a* for *e* as seen before. Let us consider the *c:*

accord /acord: colère/cholere;
avec/avecques; échec/eschecqt;
carré/quarre; défense/defance;
face/fasse, enrichicent; donc/doncq

It becomes difficult to keep in mind that each element cannot result in anything. The identification of regular patterns is strongly needed: otherwise the text could not be read.

B. Spotting a word in a given text

Let us keep in mind our initial objective: formulate requests which would provide all the forms corresponding to the requested word. The next stage is therefore to compare a list of "words to be searched" with their actual occurrences in the text: this is a form of contextualization. Let us start with an extract from Montaigne's text, which will be "searched" (shown in bold in the following extract) :

" De la coustume & de ne changer aisement vne loy **receüe**.

Celuy me semble auoir tres-bien conceu la force de la coustume, qui premier forgea ce conte, qu'vne femme de village ayant **apris** de caresser & porter entre ses bras vn veau des l'heure de sa naissance, & continuant **tousiours** a ce faire, gaigna cela par l'accoustumance que tout grand beuf qu'il estoit, elle le portoit encore. Car c'est a la verité vne violente & traistresse **maistresse** d'escole, que la coustume. Elle **establit** en nous peu a peu a la desrobée le **pied** de son **authorité**: mais par ce doux & humble commencement l'ayant rassis & planté auec l'ayde du temps, elle nous decouure **tantost** vn furieux & tirannique visage, contre lequel nous n'auons plus la liberté de hausser seulement les yeux. Nous luy voyons forcer tous les coups les reigles de nature: i'en croy les medecins, qui quitent si souuent a son **authorité** les **raisons** de leur art: […] I'en vi vn autre **estant enfant** qui manioit vne **espée** a deux mains, vne hallebarde du pli du col a faute de mains, les iettoit en l'air & les reprenoit, lançoit vne dague & faisoit craqueter vn foët aussi bien que charretier de France. […] la c'est office de pieté de tuer son pere en certain aage: **alleurs** les peres ordonnent des **enfans** encore au ventre des meres, ceux qu'ils veulent estre nourris & conseruez, & ceux qu'ils veulent estre abandonnés & tués: **ailleurs** les vieux maris prestent leurs femmes a la ieunesse pour s'en seruir: & ailleurs elles sont communes sans peché: voire en tel **pais** portent pour merque d'honneur autant de belles houpes frangées au bord de leurs robes, qu'elles ont acointé de **masles**. N'a elle pas **faict** encore vne chose publique de femmes a part? leur a elle pas mis les armes a la main? **faict** dresser des armées, & **liurer** des batailles? […] Car nous **sçauons** des nations entieres, ou non seulement l'horreur de la mort estoit mesprisée, mais l'heure de sa venue a l'endroit des plus cheres personnes qu'on eut, festoiée auec grande alegresse. […] il n'est rien qu'elle ne **face**, ou qu'elle ne puisse: & auec **raison** l'appelle Pindarus, a ce qu'on m'a dict, la **royne** & **Emperiere** du monde. Mais le principal effect de sa puissance c'est de nous saisir & **ampieter** de telle sorte qu'a peine soit il en nous de nous r'auoir de sa **prinse**."

The aim is to spot the following words:

"pied, nature, raisons, reçue, appris, celui, toujours, tantôt, autorité, épée, enivrés, maîtresse, école, établi, allégresse, loi, âge, enfant, pays, mâles, fait, livrer, savons, fouettés, mets, fasse, reine, impératrice, empiéter, prise, vue"

Comparing the searched forms and their spelling in text, a typology of the situations occurring may be offered. The form being searched is sometimes that which does occur in the text (*raison/raison*); in some cases the link seems to be very weak (*impératrice/empériere*), and between these two types a whole gradation of situations can be organised on a linguistic basis :

- relations between sounds and spellings in their different forms (sometimes still occurring in modern French) : *c=ss; n=nn; r=rr; s=z; t=th; ai=ei,ai,ey,ay,oi,oy; [uv]=u,v; u=eu,*
- inflectional history (*serais/seray/serois/seroye*)
- morphological history (*hôpital/hospitalier; forêt/forestier; advis/avis*).

Linguistic knowledge helps recognize regular replacement patterns, which can be turned into rules, and also helps recognize "hopeless" situations such as *caietz/cahiers; ajoutée/adjouxtee; échec/eschecqtz.*

IV. CONCEPTION OF RULES

A. Presentation of the typology

Obvious findings:

1. tokens similar to the types :

pied > pied	*nature > nature*	*raisons > raisons*

2. well known phenomena

(a) Some of these are "transparent" because morphological or derivational traces of them may be found in contemporary French:

(a-1) Based on such derivational sets

forêt/forestier	*hôpital/hospitalier*

little wonder that many occurrences of the circumflex ("^") should be equated to an occurrence of *s* :

maîtresse > maistresse	*tantôt > tantost*
mâles > masles	*traîtresse > traistresse*

(a-2) Similarly, there are many occurrences of *é* which appear as *es* : the French of today provides the following:

étude, étudiant, estudiantin, studieux

Hence our rule:

école > escole	*épée > espée*	*établi > estably*

(a-3) It could also be a *c* occurring between a *i* and a *t* : *fruict, faict, dict, effect, nuyct*. In contemporary French, the *c* may occur in a word from the same "family" due to the learned derivatives created at the Renaissance period precisely: *fruit/fructueux; fait / facture; dit / dicton; nuit / nocturne; effet/effectif.* It is therefore rather easy to "guess" that the *c* can be dropped.

(b) Some of these are "transparent", based on homophony :
(b-1) Occurrences of *i* are also interchangeable with those of *y* and vice versa:

celui > celuy	*loi > loy*	*pays > pais*

(b-2) Different way to write [ɛ] are still possible in contemporary French, *règle, fait, fouet, prête, est, allégresse, neige…* and they may alternate, in the spontaneous spelling of native speakers,... or in Renaissance texts : *reigle, faict, foët, preste, alaigresse*
(b-3) Same thig for [ã] : *ampieter, empiéter, melancolye, mellencolie, mélancolie, semble, samble, etc.*

(b-4) Variation also affects consonants and double consonants, with several spellings for the same consonant sound (resulting in what is still today common spelling errors):

appris > apris	*empiéter > ampietter*
quittent > quitent	*accointé > acointé*
face > fasse // fasse > face	*autorité > authorité*

3. A regular but obsolete way of spelling

Some regular alternations do not occur any longer. Some cases are more difficult because they are further apart from contemporary production in some way or other, even though they remain legible.
(a) There is the problem already mentioned of alternations now normalized in modern editions, those between *u* and *v* and between *i* and *j*:

livrer > liurer	*enivrés > enyures*

(b) Multiple alternations in spelling
Inflections, especially verb inflections, also provide examples of variation: *portais, portay, portois*, etc. are well-known examples but there are other examples such as the past participles written *eu* instead of *u :*

reçue > receüe	*vue > veüe*	*lu > leu*

(c) Very frequent verbs
Some verbs, frequently used in discourse are really puzzling, even if one can "explain" everything:

savons > sçauons	*prise > prinse*	*né > nai*

These verbs are also irregular in contemporary French. And so is a fringe of the most frequent vocabulary such as the verbs *avoir, être, aller, faire*. For such cases it may be preferable to provide access to a small lexicon of extension to generate the most common forms on the basis of the base form (infinitive).

4. cases for which little can be done

In a certain number of cases already mentioned, the variation is on the borderline between spelling variation and morphological variation. Once the borderline is clearly traced, it may seem legitimate to abandon the hope of identifying two forms: thus *icelle/ycelle* will not be identified as a variant of *celle*. The same may be said about forms such as *royne* (the feminine of *roi*, nowadays *reine*), *Emperière*, (the feminine of *Empereur*, now *Impératrice*, or even *tousiours* for *toujours*. Providing rules to identify these forms would generate such noise that one would too often end up with results which have little to do with the expected forms. For similar reasons one may drop the idea of identifying all the forms for *dé* ("dice") or *né* ("born") because they share their unaccented variant with the words *de* and *ne* and the vastly higher frequency of occurrence of *de* and *ne* would not provide a reasonable interpretation of the data.

B. Formalizing substitution rules

The situation seems to be "intellectually" rather simple : due to the structural instability of this linguistic data, equivalences between character strings are difficult to track statistically, and no model-based approach can be developed.

Sometimes the choice made by an author must not be memorized. An author, for instance, may realise that the verb *avoir* originates in the Latin *habere*, and that other European languages "preserve the initial h" (*have, haben, haber*), and may therefore decide to reintroduce an "etymological h" for the verb *avoir* in French. Such a phenomenon is of course not to be generalized.

The next point which needs to be taken into account is the relevance of the substitution rules formulated : do they really help find all the forms concerned (with a low silence and good recall), and do they avoid generating too much noise (good precision)?

The noise may have two different aspects: on the one hand, one may generate forms which correspond to several other forms actually existing, but need to be distinguished (*écouler, écolier*), and on the other hand, one may generate a large number of candidate words which are possible in terms of calculation, but totally unacceptable from a linguistic point of view (*autorité <> àüttolrrythêz*).
To test these results, let us look at a small corpus of 7 words (*vices, une, face, fesse, lu, vu, souverain*) transformed by 8 substitution rules : *c = ss ; n = nn ; r = rr ; s = z ; s = c ; ai = ei,ai,ey,ay,oi,oy ; [uv] = u, v ; u = eu*
The results reached do contain all the relevant forms, but for each base form here is the number of forms generated:

vices 648	*fesse* 93	*lu* 7	*vu* 19
une 84	*face* 90	*souverain* 117504	

Thus 7 words have been extended to 118445 forms. There is obviously some correlation between the length of the word and the number of generated forms, due to the combinatory process.
The solution chosen to fix that problem is to describe, for each rule, the context in which the substitution is allowed. This aims at constraining strongly their application, and limiting their productivity. This contextualisation is based on a good knowledge of the linguistic process involved. In the example given, the 8 simple rules are transformed into 9 more complex rules. Most of the time, one simple rule will be derived into 5 to 15 contextualised rules.

(?<\=[bdflmnprstv])u = eu (?<\=[aeiouy])c(?\=[eiy]) = ss (?<\=[aeiouy])ss(?\=[eiy]) = c (?!^.+)n(?<!.+$) = nn (?!^.+)r(?<!.+$) = rr ain = ein,ain,eyn,ayn	^s(?\=[eiy]) = c (?!^.+)v = u ^u = v s$ = z

The results reached are satisfactory: the rules produce all the linguistically permissible variants, and the number of variants is much lower: the 7 words generate 37 forms

vices 4	une 4	lu 2	vu 2
face 2	fesse 2	souverain 21	

These results meet our expectation and the contextualisation of rules is the solution chosen to improve the process. The final set of rules includes slightly less than 200 rules and one hundred lexical substitutions such as *prins* for *pris* or inflected radical forms of *sçauoir : sçai,sçau* : it was not useful to create a rule to replace s with sç, even limited to the beginning of words. The need to test precisely the recall and precision of each rule does increase with the complexity of the specification, and the output files generated to check the results at the various steps of the process are very useful.

V. DESCRIPTION OF THE TOOL

The tool itself is thought to be really user-friendly especially for the tuning of rules and the evaluation of their consequences (efficiency and non regression tests). It is a free available JAVA program which first transforms a list of words into an extended list of forms. This being done, the need is to localise the different forms attested in the old spelling in a text, according to the requested form. Therefore the programme needs three files to be used: the request list, the set of rules and the text to be searched.

```
java -jar VariaLog.jar
Path\rules-filename.txt
Path\words-filename.txt
Path\text-filename.txt
```

Hence one can identify two “phases” : query expansion and form spotting.

A. Generating the expanded form of the request

In this phase, the tool needs two input files :

- a file in .txt format containing a list of words written in the modern spelling : the requested words
- another .txt file containing the set of rules. The formalism, as shown above, is quite simple: character strings that must be replaced by others are described as regular expressions. For example, `(?<\=[aeiouy])c(?\=[eiy]) = ss` express that a “*c*” can be changed to “*ss*” when it appears between two vowels and the second one is a palatal vowel. So that the word *face* can be written *fasse*

At this stage, the program generates 3 files :

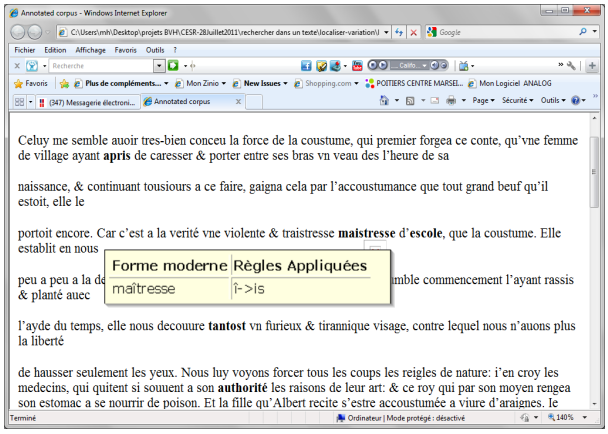
- Two of them are dedicated to a synthetic report about the process (using the rules) and the result (generated forms). For the rules, the point is to know how many times they have been used : by so many and sometimes such close rules; without such a report, it is difficult to know exactly which one is never used, and which one is too often used. For the generated forms list, the need is to see how many forms have been generated for each single form morphed into a set of other forms. It is otherwise not possible to identify the rules generating a huge amount of variants. For example, the word *autorité* has been morphed into 407 different words.
- The third one is a file containing, for each word, the list of the forms generated as well as the rules used in the process. This information is really useful to tune the set of rules : to detect errors in the formulation of rules, conflicts between rules, redundancies, noisy rules and so on. The following example shows how this is expressed.

Example of generated forms for **autorité**

```
    autoricté(it->ict)
    authorithe([eé]->e,t->th)
    authorrythes([iîïyÿ]->y,(?!^.+)r(?<!.+$)->rr,é-
>es,t->th)
    autoryté([iîïyÿ]->y)
    autoricte(it->ict,[eé]->e)
    äuthorrïte([iîïyÿ]->ï,(?!^.+)r(?<!.+$)->rr,[eé]-
>e,[aàä]->ä,t->th)
```

B. Finding the right form within a text

When the extended request is calculated, the ultimate test is to identify all the variants really attested in the text. This the second phase of our program. The third input file needed to run the program is the text being searched. The output file of this last part of the process is an html file with graphic highlighting (or bold characters) of the identified variant. Moreover, each form is connected to a bubble showing the rules used to derive the variant. A table containing the summary of the rules used for the text is also avalaible. So, the human validation process is rather user-friendly.



Tests on Montaigne's text yield a 100% recall, and our expectation was fully reached. The set of rules just had to be improved.
In order to do this, 1254 words have been collected from three reference texts :

- Champfleury, *art et science de la deue et vraye proportion des lettres attiques*,
- Des Perriers, *les nouvelles recreations et joyeux devis*
- Ronsard, *discovrs des miseres de ce temps*

This words were chosen to test the set of rules : phenomena occurring really often are tested once (sometimes with a few words, to be sure that all aspects are covered), hapax forms are tested in the same way. Therefore, this test list is a demanding but hopefully reliable one.

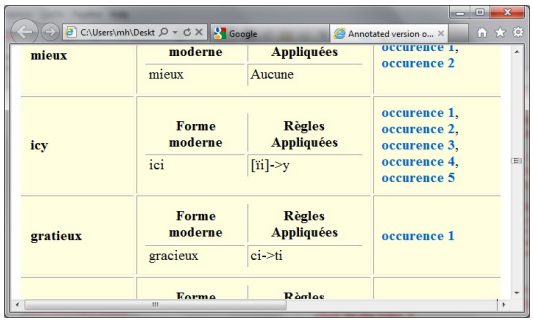
In this context, the results we have are surprisingly good, exhibiting a 99,5 recall rate.

To test the precision, we had to use the same list of words with a full text; the aim was to detect if the words identified were linked to a wrong occurrence. This cannot be totally avoided, but the aim is to restrict it to linguistic ambiguities: the rules may sometimes generate ambiguity. If "o" becomes "ou", (*profiter → prouffiter*) then, *école* may become *écoule*, which is not an acceptable variation, but *volant* becomes *voulant*, which is an acceptable variation; as a result, *volant* will correspond to *vouloir* ("want"), thus increasing the ambiguity of this form which already means "flying", "robbing", "steering-wheel", "flounce", and "shuttlecock". The ambiguity generated is no different from standard ambiguities occurring even in a regular orthographic environment.

Here is another attested example : the form *loy* is linked to the modern *lois,* and *Louis,*but when considering legal documents there is no reason to expect the occurrence of "Louis" instead of "law", but if it does occur, it should not be seen as an error : it was our surprise to discover both occurrences in the same paragraph of Ronsard's *Discours*

> Ainsi que nous voyons instruire nostre Roy Qui par vostre vertu n'a point changé de **loys (law)**. [...] Ha que diront là bas soubs les tombes poudreuses De tant de vaillans Roys les ames genereuses! Que dira Pharamond! Clodion, et Clouis! Nos Pepins! nos Martels! nos Charles, nos **Loys (Louis)**!

To allow a fast and easy appreciation of the matching process for a text, a table summarizes the word pairs matched by the rules .



The surprising thing is that the precision rate is as good as the recall rate: almost all matched forms could really be related to the query form. One point remains to be improved in the program: the rules could be more constrained and be more « elegant » and sophisticated : monosyllabic words do not have the same behaviour as polysyllabic ones and that should be examined closely.

VI. Application to Early Modern English

Already used to search several French dialects, VariaLog can be used to process any form of spelling variation, in any language. One just needs to adjust one's own specific spelling rules or dictionary.

This is what we have tried for Early English, assuming that the rules would work on the same principles, even if the variation appears on very different sequences : equivalences between *th/s ; ei/ea/ee/i/e* ; *o,u,ou,oo* are of course not relevant for French.

Fifty three years ago, Norman Davis in his contribution to the *Mélanges de linguistique et de philologie : Fernand Mossé in memoriam,* dealt with the problem of spelling variation in the English of the XVth c., concentrating on the corpus of the *Paston Letters*, of which he was to produce a few years later a modern and authoritative edition. In so doing he was aiming at the linguistic information which could be provided by the spontaneous language written by three of the letter writers of the corpus considered between 1460 and 1475.

Davis concentrated on the language written before the standardization brought about by the generalization of printing in the years which followed. Our aim is to concentrate on the variation still observable *after* the generalization of printing which is also to be observed in the French literature of the same period.

Our aim is rather limited although somewhat more ambitious since we hope to be able to generate rules to predict spelling variation in texts of the Renaissance period.

Our corpus is still rather limited: we have chosen one long text, George Puttenham's *The Arte of English Poesie,* (95,749 words without punctuation), and three short ones: Sir Philip Sidney's *The Psalms of David* (3,190 words), Edmund Spenser's *Ruins of Rome* (3,777 words) and *Letters to Gabriel Harvey* (3,528 words).

We followed the same approach as for French : we first built the index of the texts : 11,407 entries for Puttenham, 1,190 for Sidney *Psalms*, 1,713 for Spenser's *Letters*, and 1,468 for the *Ruins of Rome*. We then reduced these lists of words: eliminating the words which are written no differently today; identifying a typology of facts which are regular may be very productive : the *th* digraph at the end of word indicating the verbal inflection (*hath, expresseth, carieth, geueth*), the *-e* at the end of words (*againe, feete, minde, farre)* and sometimes very specific features, like the alternation between *ow* and *ew* in *show*, which lasted until the beginning of the XXth century.

The index was thus reduced to a list of 4723 words, which we used as a basis to formulate our rules.

To build the index, compile the concordances of occurrences and search letter combinations, we have used the tool called AnaLog, as mentioned above.

Some rules have been fairly easy to formulate, as the spelling showed a regular pattern in the 16th c. corresponding to a no less regular although different pattern nowadays (*fairly/fairely, cheerful/cheereful, entertainment/entertainement*).

The extensions of our rules to other texts will enable us to validate or refine them.

But our purpose is here to consider the rate of success of the tranfer of our method to Early Modern English texts, rather than immediately achieve a 99% matching or a finalized tool.

A. Presentation of the typology

Obvious findings:

1. tokens similar to the types :

with> with	*reason > reason*	*excellent > excellent*

2. well known phenomena affecting the spelling of both English and French

(a) As in French, the dissimilation between u/v is not made, or not standardized:

haue	*geueth*	*neuer*
until,vntil	*vpon,upon*	*live,liue*
vniversal		

(b) Phenomena regulary occurring at the end of words

(b-1) inflectional endings

has>hath,	*express >expresseth*
gives>geueth	*caries>carieth*

(b-2) the massive presence of -e at the end of words, irrespective of its role as marker of the length of the previous vowel:

foot>foote	*againe>againe*

mind>minde	*far>farre*

(c) Double consonants

(c -1) Double consonants and finale mute -e :

The most frequent case is that of the equivalence in monosyllables especially between a final simple consonant and and a double consonant followed by a final -e. The doubling of consonants before a final mute e is a well-known feature of EME, a trace of which is, after the loss of -e, the contemporary digraph -ck after a short vowel. Hence such spellings as:

toppe, hoppe, shippe, gappe, hatte, matte, witte, backe, blacke, flocke, lucke, musicke, penne, ginne, sinne, runne, tenne, winne, swimme, stemme, plumme, farre, starre, stirre, preferre, referre, differre, etc.

(c-2) Double or single consonants in medial position

Just as in French, one can find a word with a simple or a double consonant :

British, brittish		*Litle, little*	*title, tittle*
Elisabeth, elissabeth		*cometh, commeth*	*forein, forrein*
sory	*mary*	*leter*	*prety*

Our corpus yields occurrences of the last four words with only one consonant, which shows that the current standardisation had not been achieved in the status that double consonants have now acquired in the English spelling.

The doubling of m in *becommeth* is an interesting consequence of the syllabic status of the inflectional ending *-eth*, which makes it necessary for the stressed syllable to have its short vowel marked by a double *m*.

(c-3) Other alternations connected with assibilation processes

As in French, the double value of *c* and *t* depending on the environment has created variations in spelling such as:

Councellor, counsellor	lyrickes
Pronunciation, pronuntiation	scituation

3. Word formation (lexical and inflectional affxation)

(a) presence of an -e before the suffix

The frequent presence of a final -e which the modern spelling has now deleted often gives the impression that this vowel, more a diacritic than a vowel, is much more present in Early Modern English. This is overwhelming true at the end of words as we have seen above but it also regularly occurred before suffixes:

cheereful, cheereless,
empeachement, entertainement, commaundement,
fairely, *cleanely, clearely, truely, dearely,* and even *easely, curiousely,* and such similar cases.

(b) absence of -e now required after a prefixed element

Conversely the absence of the letter -e is also consistent in compounds the first element of which is *some, where* or *there*:

wherto, wherupon, wherby, therin, therwith
somwhere, somwhat, somtimes
concurrently with *sometimes, somewhat, somewhere*

(c) absence of -e now present before an inflectional suffix:

The spelling of final syllabic consonants did not require the presence of a prefinal -e- as it now does: the past tense form of *remember* was *remembred*, whereas the modern spelling leaves the verb form *remember* unaltered before adding the inflectional suffix *-ed.*

remembered>remembred	*encountered>encountred*
happened> hapned	*offered>offred*

(d) absence of -e now present before a derivational suffix:

The same holds true for derivational suffixes as well:

encounter → encountrer	*wander → wandrer*

4. Word formation : Romance and Germanic origins

(a) endings in -ull, -all, -ill, -ell

Another regular phenomenon is the presence of -ll in suffixes at the end of adjectives, e.g. in *-full* and *-al/-ial/-ical*:

beawtifull, behooffull, thankefull
artificiall, tragicall, comicall, astronomicall, partiall, proverbiall, memorial, cathedrall, morall

Final *-ll* is also frequent with unstressed endings other than suffixes:

Scandall, seuerall
angell, modell, kernell, counsell
euill, perill, pensill

and in adjectives ending in *-il* borrowed from the Latin or from the French:

facill, subtill, ciuill

The contemporary spelling of word-final syllabic consonants in *-le* had not been generalized yet as shown be-

low, whereas words of Old English origin already had the modern spelling:

mantell, castell, hostell, battell
title, little, gentle.

(b) suffixes *er/ar* and *or/our*

The occasional ending -ar (rather than -er) in *souldiar*, shows that the contemporary spelling in *liar* is not a single occurrence.

Words in -our/-or such as *coulour/coullour, fauour, doctour, errour, oratour, senatour,* all have the two spellings but end in -*our* rather more often than in -*or* in spite of a pronunciation which was certainly not that of a diphthong. The occurrences of final -or (*ancestor, actor, author, ambassador, conqueror, coulor, dolor, humor, inferior, inventor,* and even *neighbor*), although less frequent, do exist.

5. The treatment of vowels other than -e

(a) the case of <ou/u>

The same -ou- digraph mentioned in connection with the or/our ending seems to represent also a short vowel (although not unstressed) in *bloud* for *blood*, *floud* for *flood*, *toung* for *tongue,* and also *aboundantly* for *abundantly*. This <ou> → [ʌ] correspondance can still be traced in contemporary spelling in a limited number of words such as *double, couple, trouble, cousin.*

(b) Distribution of <ie> and <y>

The distribution between final -y and -ie is largely in favour of -ie, a spelling now restricted to the plural form when it is followed by -s: in Puttenham's *Arte* the spelling *city* occurs twice while *citie* has twelve occurrences and Spenser's *Ruins of Rome* have only six occurrences of *citie* and none of *city*.

But the complementary distribution between *y* and *ie* we are now familiar with was not established yet: the vowel *y* is not necessarily in final position and it is to be found also followed by -e: *faithfullye, presentlye, ordinarilye, partelye,* and in monosyllabic verbs: *flye, lye*. We have had to take this variation into consideration for the formulation of our rules:

y → ie, ie → ye, ys → ies (to account for such spellings as *alwaies*), *etc.*

agayne, mayne, vayne, entertaynes

(c) Vowels before nasals

empeachement/impeachement	employed/imployed
embassadour, ambassadour	nomber, number
pronouncing, pronuncing	somptuous, sumptuous
remembraunce, remembrance	aunswer, answer

(d) Other vowel digraphs

The variation in digraphs has resulted in a substantial number of rules, all of them justified by the examples provided below, but so far (perhaps due to the limted size of our corpus) the number of rules has been definitely smaller than in French:

e → ea : extreme→extreame

ee → ea, e: cheer→cheare

ea → ee: peace→peece, flea→flee, dear/deare→deere

ei → ea : receive→receaue

ei → ai : reign→raigne, sovereign→soueraign

discret / discreet (both are to be found).

CONCLUSION

The highly "instable" situation of spelling at the time of the French Renaissance or Early Modern English period is not conducive to the acquisition of rules by automatic procedures: on the one hand, the evolution caused by the passage of time is significant as the period extends over two centuries which have been marked by mutations in each language and their standardization; on the other hand, variation is observable on one and the same page, in one and the same book, and may change radically from one book to another or even from one edition to another, and each author has a different attitude in this respect. And the typology we made shows thet variation in Early Modern English is as frequently the case as in French in one and the same text: hence the need of a tool insensitive to spelling variation.

Such a tool performs what the speaker of a language does to make the necessary adjustments in situations of communication even though he may do this somewhat erratically. The fact is that speakers are used to dealing with "approximation". Errors in the transcription of oral sequences often are unexpectedly funny. P. Cappeau (2010) explores these "auditive illusions" that one can find in transcriptions: (on the telephone), *je suis rue Béranger* ("I am at rue Béranger") has been interpreted as *je ne veux pas déranger* ("I don't want to disturb you"). The listener has redesigned an utterance which he can interpret with the elements he has perceived. With a little training, the reader is led to elaborate a sort of method combining the different pronunciations possible for a given string of letters, substituting the ones for the others on the basis of substitutions otherwise possible in contemporary French, resorting to his knowledge of lexical structures and of derivations in word formation, exploring his mental lexicon and its organisation, and relying heavily on this "engine of mental approximation" which leads speakers to identify an existing form, occasionally running the risk of making a mistake.

After reaching acceptable results on a French corpus, our aim is to test the transferability of our method to other corpora in other languages, especially English.

The context has favorable elements: the two languages have numerous points of contact and the diachronic phenomena concerned have a lot in common in terms of

phonetic change and of its more or less delayed transposition in spelling, in terms of word formation or in terms of spelling standardisation to help differentiate possible homographs even though they may be homophones (French *vers/vert/verre/vair*; Engl. *to/too, for/four, corpses* with a p which is not recorded in the spelling of our corpus which has *corses* for *corpses*).

The analysis of the three texts of our corpus enables us to confirm our hypothesis and to produce rules adapted to the spelling variations of our English corpus. They are largely different from the rules for French and yet apply the same logic and their efficiency is equally acceptable: we have reached a recall rate which is above 99% and a precision rate which is even better than for French. This must be qualified by the fact that our test has been conducted on a limited corpus which has helped elaborating the method followed. Yet our previous experience with French texts entertain our hope to succeed. The challenge now is to improve the tool by using it in a wider project, which would mobilize the energy required to validate it.

There is an area to which we would like to attract the reader's attention, both for the improvement of the tool and for the study of texts and languages. In both English and French the rules applicable to monosyllables are significantly different from those applicable to longer words. And this is a field of study which is relevant for several other fields, such as the study of rhythm, language change or the identification of areas of text in different languages, etc.

Some facts left out of the batch treatment made possible by the rules help provide synthetic and documented evidence which is a welcome spinoff of the use of the tool developed and applied to the documents of earlier periods.

This tool plays the rôle of a magnifying glass for these facts which do require a lexical treatment specific for each word. But in terms of language history and grapho-phonemic correspondences the tiny problems thus revealed are worth consideration as they allow observations which might otherwise have been overlooked.

Some historical conclusions may be drawn from the exhaustive analysis of the spelling. For example, the phonemic merger of vowels before r has generated variants in spellings which could only be predicted if the context of the consonant r could be specified. Many phonetic respellings are to be observed and in such cases as *curteous,* or *aduirtisement* the whole context would have made the rules especially tricky to formulate.

Some words require special treatment. Such is the case of *bewtie/beautie* and its derivatives: there are 29 occurrences of *beaut-* and 39 occurrences of *bewt-* in Puttenham's *Arte*. The phonetic equivalence of *bewt-* and *beaut-* can hardly be generalised and therefore a lexical treatment of the variant was the only reasonable solution.

In *perswade, perswasiue, perswasion, perswader(s),* the presence of <w> is interesting phonetically since it shows that what follows the s was and is a semi-vowel (phonetically [w]) and not a vowel or the first element of a dipthong as in *evacuate, insinuate, casual, etc.*

The preference later given to the u spelling in *persuade* has obliterated this feature to comply with the faithfulness to the Latin etymology. But in the verb *swear* spelt *sware* in EME, a verb of Old English origin, the *w* has been retained.

The variation between *enmitie* and *ennimitie* provides evidence for the priority one must give diachronically to the placement of stress on the deriving form ['en (i) mi +it i] with a stress-neutral derivation of the noun in *-itie/-ity*, rather than a placement of stress governed by the presence of the stress-imposing ending *-ity*, which would have prevented the syllable syncopation of *-ni-*.

The variant spelling recorded in *moniment* is a consistent representation of a variant pronunciation, which happens to be echoed two centuries later by Walker's *Critical Pronouncing Dictionary* (1791), a prescription of which, on the basis of the modern spelling which in the meantime has prevailed (*monument*), stigmatizes the pronunciation with a reduced vowel for which *moniment* would have been an acceptable spelling.

The observation and the identification of variants has an impact on several possible applications:

1. in the digital document environment
    - it helps improve search engines
    - it helps identify quotations in certain contexts
    - it helps produce stndardized editions of modernized texts
2. in the field of language study
    - history of spelling
    - lexical evolution
    - historical phonetics

This is the experience which has been ours by observing facts from different viewpoints. We assume that other viewpoints will give rise to other such discoveries.

REFERENCES

Baron, A. and Rayson, P. (2009). Automatic standardization of texts containing spelling variation, how much training data do you need? In M. Mahlberg, V. González-Díaz and C. Smith (eds.) Proceedings of the Corpus Linguistics Conference, CL2009, University of Liverpool, UK, 20-23 July 2009

Burnard, L. (1995). 'Text Encoding for Information Interchange – An Introduction to the Text Encoding Initiative'. *Proceedings of the Second Language Engineering Conference*, 1995.

Bonnin, E., Dallo, A., (2003) Hyperbase et Lexico 3, outils lexicométriques pour l'historien, in *Histoire & Mesure*, XVIII, n°3/4

Cappeau, Paul. (2011). « La transcription et ses entours ». In Jean Chuquet (éd). Le langage et ses niveaux d'analyse - Cognition, production de formes, production du sens. Rennes. PUR.

Craig, H., Whipp, R. (2010).'Old spellings, new methods: automated procedures for indeterminate linguistic data'. *Literary and Linguistic Computing,* 25-1, 2010, pp. 37-52.

Davis, N. (1959), Scribal Variation in late Fifteenth-Century English, in *Mélanges de linguistique et de philologie : Fernand Mossé in memoriam,* Paris : Ed. Didier, 1959, pp. 95-103.

Demonet, M.-L. (1995), Pour une édition hypertextuelle de *La briefve declaration de Rabelais*. Wooldridge ed, CA

Demonet, M.-L. (1996), *Pronostiquer avec Hyperbase, Mots chiffrés et déchiffrés*, Slatkine, pp. 455-471.

Demonet, M.-L. (2006), Les Bibliothèques Virtuelles Humanistes (BVH) au Centre de la Renaissance de Tours : numériser en région pour l'Europe, *10e journée des pôles associés BNF*.

Demonet, M.-L., Lay M.-H. (2008), Digitizing European Renaissance prints: a 3-year experiment on image-and-text retrieval, Kolkata, *International Workshop on Digital Preservation of Heritage* (IWDPH07)

Erjavec, T. (2011). 'Automatic linguistic annotation of historical language: ToTrTaLe and XIX century Slovene'. *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 2011, Portland, pp. 33-38.

Hana, J., Feldman, A., Aharodnik, K. (2011). 'A Low-budget Tagger for Old Czech'. *Ibid*, pp. 10-18.

Habert B. & Zweigenbaum (2002), Régler les règles, *TAL*, 43-3, pp. 83-105.

Habert B. (2005), *Instruments et ressources électroniques pour le français,* Paris : Ophrys.

Kraif O., Chen B. (2004), Combining clues for lexical level aligning using the Null hypothesis approach, in *Proceedings of Coling 2004*, Geneva, August 2004, pp. 1261-1264.

Kraif O. (2011) Les concordances pour l'observation des corpus : utilité, outillage, utilisabilité, in *Actes des 10 ans de la MSHS de Poiters*, Rennes: PUR.

Lay-Antoni, M.-H., al. (2000), 'Adaptation d'un lemmatiseur au corpus rabelaisien : naissance d'Humanistica'. *JADT 2000,* Lausanne.

Lay, M.-H. (2011) "AnaLog, un outil pour l'étude et la comparaison de corpus annotés" In Jean Chuquet (éd). Le langage et ses niveaux d'analyse - Cognition, production de formes, production du sens. Rennes. PUR.

Lay, M.-H., al. (2010). 'Pour une exploration humaniste des textes: AnaLog'. *JADT 2010,* Rome.

Ramel, J.-Y., Busson, S., Demonet, M.-L. (2006) AGORA: the interactive document image analysis tool of the BVH project, *DIAL, Digital Image Analysis for Library,* Lyon.

Sánchez Marco, C., Boleda, G., Padró, L. (2011). 'Extending the tool, or how to annotate historical language varieties'. *ACL-HLT Workshop*, 2011, Portland, pp. 1-9.

Scheible, S., Whitt, R. J., Durrell, M., Bennett, B. (2011). 'Evaluating an 'off-the-shelf' POS-tagger on Early Modern German text'. *Ibid*, pp. 10-18.

Souvay, G., Pierrel, J.-M. (2009). 'LgeRM: Lemmatisation des mots en moyen français'. *TAL*, 50-2, 2009, pp. 149-172.

Thaisen, J. (2011). 'Probabilistic Analysis of Middle English Orthography: the Auchinleck Manuscript'. *Digital Humanities Conference Abstracts, 2011,* Stanford.

Vandendorpe, C., (1999) *Du papyrus à l'hypertexte. Essai sur les mutations du texte et de la lecture*. Éditions de la Découverte.

Walker, J., (1791) *A Critical Pronouncing Dictionary and Expositor of the English Language*

www.cesr.univ-tours.fr/Epistemon/

www.bvh.univ-tours.fr/

http://www.c-tei.org

http://www.bvh.univ-tours.fr/XML-TEI/index.asp. (Manuel d'encodage TEI Renaissance, 2009

www.artamene.org/philologic.php

http://xtf.cdlib.org/documentation/programming-guide/

http://www.monkproject.org/

http://eebo.chadwyck.com/help/whatis_wh.htm

http://impactocr.wordpress.com/

http://panini.northwestern.edu/mmueller/vospos.pdf

Corpus From the BVH : www.bvh.univ-tours.fr/

Champfleury, *art et science de la deue et vraye proportion des lettres attiques*, (qu'on dit autrement lettres antiques, et vulgairement lettres romaines proportionnees selon le corps et visage humain).

Des Perriers, L*es nouvelles recreations et joyeux devis,* Lyon, 1563

Montaigne, Les E*ssais,* Bordeaux, 1598

Rabelais, P*antagruel,* Lyon, 1532; 1542)

Ronsard, D*iscovrs des miseres de ce temps,* Lyon, 1563

Corpus From the http://www.luminarium.org/renlit/

Puttenham, *the Arte of English Poesie*,

Spenser, *Letters to Harvey*, London, 1882

Spenser, *Ruins of Rom*, Oxford, 1910

Sidney, *Psalms of David*,1877