# WARNING I have new documents that are not yet included to bring this to current state of the art.

GCW 30/01/2020

# Encoding Basnage

Although we are encoding the dictionary of Furetière, it is the 1701 Basnage de Beauval version, and as he brought in a large number of changes, it will be easier to simply refer to Basnage throughout.

It is necessary to accept that legacy dictionaries are, by their very nature, idiosyncratic. Many of the features that are now commonly accepted were purely experimental, and as such inconsistent across the dictionary. This means that markup will also be exploratory as the unexpected is to be expected and the markup must handle this diversity. We are currently working on the 1701 edition, but will move back to 1690 and forward to 1702, 1708 and 1725/27. These will mostly map, but other variations can be expected. As the aim is to move onto source dictionaries as those of Menage and Bayle, we need to cover a variety of possibilities. The essence of TEI is in being descriptive and not normative.

## Top-level entry structure

The dictionary was published in three volumes corresponding to the letters A-D, E-N and O-Z. In order to demonstrate this, the top-level division for each file is always for the volume with the volume, *tome* in French, being given as below for the A-D volume. The embedded division is the appropriate alphabetical section. The dictionary is currently being encoded by alphabetical section with one file per letter. Thus a file has been created for the letter ‘C’ with a second-level division within <body>.

<div type="tome\_A-D">  
 <div type="C">  
 <cb n="1"/>  
 <head>C. CA. CAA. CAB.</head>  
 <div type="C">

The entries in each alphabetical division are listed alphabetically based on the first one, two or three letters. Thus, there are sub-sections for each block and these are given their own division as below

<div type="C">

<div type="C">

<div type="CA">

<div type="CAA">

The entry for the letter itself is always given a separate division.

In our case, we use <entry for each separate words and its senses. An entry starts with a headword in capital letters and ends when the next entry starts. The entry has two parts: metalexicographical information about the headword and the sense structure. The information concerning the headword is contained within <form type=”lemmaGrp”>. The latter contains a series of other elements:

<form type=”lemma”> for the headword

<form type=”variant”> for second spellings if provided. Both contain the actual headword inside

<orth> The rendition attribute is linked the list of www attributes in the heard and is usually either #uc for Upper Case or #sc for Small Caps

<gramGrp> holds grammatical information if present - <pos>, <gen>, <subc> and <num>

<pron> contains information as to pronunciation

<per> holds a series of <seg> when a verb is conjugated. The latter are holders until full analytical markup is carried out.

<mentioned> is used by an external reference such as the *Académie française* is cited.

<usg> is used if usage information is provided.

<lbl> contains all link words and expressions such as <lbl>ou</lbl> or <ou plutot>

<cit> is temporarily used to contain <xr> for diminutives as in <xr><lbl> diminutif of <ref target=”cable”>cable</ref></xr>. As soon as the schema is updated, <cit> will go.

<note> is needed to include comments when they occur.

<dictScrap> is used when we are unsure how to markup.

Example 1

<form type="lemmaGrp"><form type="lemma"><orth rendition="#uc"  
 >battre</orth></form>  
 <gramGrp><pos expand="Verbe">v.</pos><subc expand="Actif"  
 >act.</subc></gramGrp><per><seg>J'ai battu.</seg><seg>Je batis </seg>  
 <seg>Je battrai.</seg>  
 <seg>Que je batti$$e</seg></per>  
 <note>Ce mot e$t de grand u$age, &amp; a plu$ieurs $ignifications, $elon les  
 cho$es à quoy on l'applique : &amp; premierement il $ignifie,  
 </note></form>

Example 2

<form type="lemmaGrp"><form type="lemma">orfelin </form><lbl>ou </lbl>  
 <form type="variant">orphelin,</form>  
 <form type="inflected">ine</form>  
 <gramGrp><pos>adj.</pos><lbl> &amp;</lbl><pos>subst.</pos></gramGrp></form>

This system of one file per letter obviously brings in problems when cross-references occur. This is overcome pragmatically by simply created the necessary alphabetical division within the file and adding in an empty holding entry for the cross-referred word. As other letters are encoded, proper cross referencing across files can be carried out. This can be illustrated with the letter ‘B’ within the ‘A’ file.

<div type="B">  
 <entry xml:id="Badiner" xml :lang:= "fre">  
 <form><orth rendition="#uc">badiner</orth></form>  
 </entry>  
 <entry xml:id="Bouement" xml :lang:= "fre">  
 <form><orth rendition="#uc">bouement</orth></form>  
 </entry>  
 </div>

As will be detailed after, each entry gets a unique identifier using the orthographic word, normally without accentuated characters. The latter proviso has brought in difficulties as sometimes two identical headwords may occur or one may have an accent to differentiate it. Consequently, this procedure may have to be revised or refined. Insofar as all the coding is in UTF 8, this is not a problem.

### Homonyms

Basnage groups words by their spelling and this can mean some strange groupings of homonyms. Close words as *calamite* and *calamité* are clearly two entries, but in the case of *canarie* we find two words with different genders under the same heading. In this case, we use two <hom> with form and grammatical information as below:

<entry xml:lang="fra" xml:id="Canarie">  
 <hom><sense n="1"><form type="lemma"><orth rendition="#uc"  
 >canarie</orth><gramGrp><pos expand="Substantif"  
 >s.</pos><gen expand="Feminin">f.</gen></gramGrp></form>  
 <def>Espece d’ancienne danse que quelques-uns croyent venir des  
 Isles Canaries, &amp;. qui, selon d’autres, vient d’un balet ou  
 mascarade dont les Danseurs etoient habillez en Rois de  
 Mauritaine, ou Sauvages.</def>  
 <note>En cette danse on s’approche , &amp; on se recule les uns des  
 autres, en faisant plusieurs partages gaillards , étranges &amp;  
 bizarres qui representent des Sauvages,</note></sense></hom>  
 <hom><sense n="1"><form type="lemma"><orth rendition="#sc"  
 >canarie</orth><gramGrp><pos expand="Substantif"  
 >s.</pos><gen expand="Masculin"  
 >m.</gen></gramGrp></form>  
 <def>Sorte de petit oiseau qui chante bien qu’on apporte  
 ordinairement des Isles Canaries. </def>  
 <note>On l’appelle autrement <hi rendition="#i">ferin</hi>. Un <hi  
 rendition="#i">canarie</hi> mâle, un <hi rendition="#i"  
 >canarie</hi> femelle.</note>  
 </sense></hom>  
 </entry>

In some cases, two fully separate entries may occur with the same headword spelling. This raises problems with the id, solved by adding \_2 and the type ‘hom’.

## Entry-level organisation

Basnage groups entries by orthographic word. Thus grouped under a single headword, we find a variety of senses, related entries and even homonyms. In order to handle this, we have decided to use only entry and then to divide the rest into senses and related entries whenever there is a change of part of speech. Thus, differentiation is achieved by showing a change in part of speech, but otherwise we respect the structure given by Basnage. In all cases, Basnage uses large capitals for the main entry and designates senses using small capitals, that is to say that any homograph would also be in small capitals and we would use an appropriate attribute to distinguish them from other senses. We encode all characters in lower case and use rendition to give a WC3 style attribute. The basic structure is thus as follows:

<entry>

<sense>… </sense>

<sense>

<sense>… </sense>

<sense>… </sense>

</sense>

<sense>… </sense>

</re>

<xr>… </xr>

</entry>

All entries contain a <sense> with the exception of those that only cross reference, in which case <xr> is used.

In this way, all the headwords are contained within entry and with one or several senses, <sense>, and related forms, <re>, for when he gives derived forms for modifiers or for sub-sense with a variation of the headword, for instance ‘capitaine’ as entry and ‘capitaine de port’ as <re>. The headword is given both an xml:lang and an xml:id attribute, the latter enabling it to be cross-referenced throughout the dictionary as in the following example:

<entry xml:id="Abaisser" xml:lang:"fre" >

In the original paper version, the main entry is always in large capitals and the senses in smaller capitals, if and when the headword is repeated. However, in data entry only lower case is used throughout and the appropriate rendition, upper case – uc – or small caps – sc – is added using the <rendition> element and a set of css values given in the header. The actual presentation on screen can then be handled through the css. The entry also receives an xml:lang attribute which is needed when comparing with other dictionaries as even if the word is not of French origin, Furetière considers that the word has entered the French language if it is to be recorded in the dictionary. The attribute ‘main’ is default and is only changed if the entry is a homonym.

<form type="lemma"><orth rendition="#uc">cadi</orth></form>

When an entry has a spelling variation, the attribute changes to ‘variant’:

<form type="variant"><orth rendition="#uc">cadi</orth></form>

Grammatical information is placed in the <gramGrp> element, which normally follows <form>. The most frequently used elements within gramGrp are for the part of speech, <pos>. When gender is given, as is generally the case with nouns, then the element <gen> is added to <gramGrp>. Subcategory, <subc> is used when this information is supplied. In the Basnage project, we are signalling the categories using the ‘expand’ attribute using a set of values defined in the header. This has been chosen so as to have a standard reference as we can find ‘*verbe’*, ‘*verb.*’ or just ‘v’. All parts of speech, *verbe, substantive, adjective, adverbe* are expanded. Similarly, the gender is also expanded to feminine or masculine. <subc> is expanded to ‘actif’, ‘neutre’ etc. for verbs. For an explanation of the choice of values, see Furetière’s entry ‘verbe’ for categories. It is hoped to develop an analytical framework later. When a plural is indicated <num> is used. The structure is thus:

<gramGrp><pos expand="substantif">ſ.</pos><gen expand="Masculin">m.</gen></gramGrp>

An example for a verb is given below, followed by one for a noun and an adjective.

<entry xml:id="Abaisser" xml:lang="fre">  
 <form type="lemmaGrp">

<form type="lemma"><orth rendition="#uc">abaisser</orth></form>  
 <gramGrp>

<pos expand="Verbe"> verbe </pos>

<subc expand="Actif">actif</subc></gramGrp></form>

<entry xml:id="Cage" xml:lang="fra">  
 <form type="lemmaGrp">  
 <form type="lemma">  
 <orth rendition="#uc">cage</orth></form>  
 <gramGrp>  
 <pos expand="Substantif">s.</pos>  
 <gen expand="Feminin">f.</gen></gramGrp></form>

<entry xml:lang="fra" xml:id="Calamiteux">  
 <form type="lemmaGrp"></form>  
 <form type="lemma"><orth extent="full">calamiteux</orth></form>  
 <form><orth rendition="#sc" extent="part">euse</orth>  
 <gramGrp>  
 <pos expand="Adjectif">adj.</pos>  
 </gramGrp></form>

Basnage does not always give the part of speech but this is useful when extracting data by part of speech. Thus, at the moment, we are declaring nouns (substantif), verbs and adjectives and also gender when this information is given, but are also adding an empty noun element for <pos> when the information is not specified.

<entry xml:id="Abadir" xml:lang="fre">  
 <form type="lemmaGrp">  
 <form type="lemma">  
 <orth rendition="#uc">abadir</orth></form>  
 <gramGrp>  
 <pos expand="Substantif"/>  
 </gramGrp></form>

Verb entries may contain a section detailing a change of part of speech as with the inclusion of derived forms as modifiers. These are encoded using the <re> element.

<re><form type="derivative" corresp="#Abaisser"><orth extent="full" type="pp"  
 >abaissé</orth><orth extent="part" type="fem">ée</orth><pc>.</pc>  
 <gramGrp><pos expand="Adjectif">participe passif &amp;  
 adjectif.</pos></gramGrp></form>

<re><form type="derivative" corresp="#Abandonner"><orth extent="full"  
 type="pp">abandonné</orth>  
 <orth extent="part" type="f">ée</orth><pc>.</pc>  
 <gramGrp><pos expand="Adjectif">part. pass. &amp;  
 adj.</pos></gramGrp>  
 </form>

The related entry often has the past participle in the masculine as the main entry and the feminine as an abbreviation. This is handled by given <orth> twice, once with the attribute extent as *full* and once with extent as *part*. In addition, we add the type as being *pp* for past participle and *fem* for the feminine. As a related entry entails a change in part of speech, the GramGrp element may be required and would contain the part of speech element <pos>. The related entry may be fully detailed with examples and citations as in the cases detailed above.

Occasionally, Basnage comments his headwords. In such cases label, <lbl> may be used as in the following examples:

<entry xml:lang="fra" xml:id="Cable">  
 <form type="lemmaGrp">  
 <form type="lemma"><orth rendition="#uc">cable</orth><pc>.</pc>  
 <lbl>Quelques-uns écrivent</lbl>  
 <form type="variant"><orth rendition="#uc">chable</orth></form>  
 <gramGrp><pos expand="Substantif">s.</pos><gen expand="Masculin">m.</gen></gramGrp></form></form>

<entry xml:lang="fra" xml:id="Cacao">  
 <form type="lemmaGrp">  
 <form type="lemma"><orth rendition="#uc">cacao</orth>  
 <pc>(</pc><lbl> D'autres disent </lbl>  
 <form><orth type="variant">cacaoyer</orth></form>  
 <usg>pour l'arbre &amp;</usg>  
 <form><orth/>cacao</form><usg>pour le  
 fruit.</usg><pc>)</pc></form>  
 <gramGrp><pos expand="Substantif">s.</pos><gen expand="Masculin"  
 >m.</gen></gramGrp></form>

<entry xml:lang="fr" xml:id="Cayer">  
 <form type="lemmaGrp">  
 <form type="lemma"><orth rendition="#uc">cayer</orth></form>  
 <gramGrp><pos expand="Substantif"  
 >s</pos><gen>m</gen></gramGrp></form>  
 <usg>Les Anciens écrivoient <hi rendition="#i">Cahier</hi></usg>

<entry xml:lang="fra" xml:id="Cabestan">  
 <form type="lemmaGrp"><form type="lemma"><orth rendition="#uc"  
 >cabestan</orth>  
 <gramGrp>

<pos expand="Substantif">s.</pos>

<gen expand="Masculin">m.</gen></gramGrp></form>  
 <pron>L'<hi rendition="#i">s</hi>se prononce.</pron>

<lbl> Quelques-uns écrivent </lbl>

<form type="variant"><orth rendition="#i">Capestan</orth>

</form></form>

In the first case, Basnage simply introduces a spelling variation, while in the second he informs us that the word may have two senses, the fruit and the tree, or have a separate word for tree. He gives the other word, but does not supply an entry so <lbl> and <usg> give important semantic information. In a fourth case, we have pronunciation information.

Within the main entry and elsewhere, we may have issues of punctuation. When needed, we use <pc>.</pc>. However, generally, we shall not be using <pc> in the manually encoded version as it will take up a lot of time with little added value.

## Sub-entry level organisation

### Defining

As described above, each entry is given its headword information using <form> and is then divided into senses and related entries. Senses and related entries may have sub-senses. Within the sense element, Basnage generally gives a definition, encoded using the <def> element. Definitions come in a variety of forms and a future analytical task will be to differentiate these and also locate introductory markers. With nouns, a short phrase forms a definiendum with a genus proximum as above. For verbs, a synonyms or series of synonyms may be found. Sometimes the definition is signalled by the verb ‘*signifier’* (to mean) or by other formulae as ‘*c’est à dire’* (that is to say). It is never easy to differentiate between the definition, comment and encyclopaedic information. The consequence is that definitions may appear long, but this will be formalised once more data has become available.

<entry xml:lang="fra" xml:id="Caaroba">  
 <form type="lemmaGrp"><form type="lemma"><orth rendition="#uc"  
 >caaroba</orth></form>  
 <gramGrp><pos expand="Substantif">s.</pos><gen expand="Feminin"  
 >f.</gen></gramGrp></form>  
 <sense n="1"><def>Arbre fort commun dans les Indes  
 Occidentales.</def></sense>  
 </entry>

Basnage uses a variety of means to give a definition, amongst which the phrase form ‘*c’est [def]*’ (this is …).

<def> C'eſt le nom d'une pierre que Saturne devora.</def>

The definition is not always given at the opening of the entry and may follow a discussion of the word sense. Once the work is further advanced, it will be possible to classify definition types and text format.

In some cases he gives synonyms as in:

<def > s'affaisser,</def>

He sometimes tells the reader to look elsewhere for a synonym of more correct usage in which case the phrase is commented using <xr>

<xr><lbl> Voyez </lbl><ref corresp="#Rabaisser" rendition="#sc">rabaisser</ref></xr>

When cross-referencing, the new word is added to the appropriate section of the dictionary, but only as the headword with its appropriate xml:id so as to enable further cross-referencing should this be necessary. We are not treating as synonyms paraphrasing as in “signifie encore, Quitter un lieu,”.

As said above, we then add a dummy entry so as to declare the referred headword without writing the full e,try at this stage. The cross referred entry for #rabaisser is thus:

<div type="R">  
 <entry xml:id="Rabaisser"><form><orth rendition="#uc">rabaisser</orth></form></entry>  
 <entry xml:id="Ravaller"><form><orth rendition="#uc">ravaller</orth></form></entry>  
 </div>

This aspect will be dealt with in more detail when looking at complex entries.

### Terms

This study is particularly interested in what Basnage declares as being terms. This is done in a number of ways using a set of formulae that designate a sense as being a term and relating it to a domain. This is done with the usage element, <usg>, which usually precedes the definition. The set of formulae that we have discovered shows that sometimes the word ‘terme’ may not be used, but only reference to a domain. However, we have decided not to encode these for the moment, but to concentrate on words that have been clearly designated. Sometimes the term is the main headword in a single entry, in other cases it is part of a more complex entry, which may even contain a number of terms each related with different domain. In such cases, the entire entry is given as information in the main entry is obviously pertinent sub-entry containing the term.

In most cases, the term is given as one sense of a given word as in the entry for CACHRYS, where one of the three senses is designated as a term.

<sense n="1"><usg type="terme" ana="#botanique">Terme de  
 Botanique</usg><def>une plants qu'on appelle autrement <hi  
 rendition="#i"/>Armarinte</def>  
 <xr> Voyez <ref corresp="#Amarinthe" rendition="#sc"> armarinte </ref></xr></sense>

<sense n="2">  
 <form>  
 <orth rendition="#sc">parc,</orth>  
 </form><usg type="terme" ana="#marine"> en termes de Marine,</usg>  
 <def> se dit des pêcheries construites sur les greves de  
 la mer. </def>

<sense n="7">  
 <form>  
 <orth rendition="#sc">parc,</orth>  
 </form><usg type="terme" ana="#guerre"> en termes de Guerre, </usg>  
 <def>est un poste hors de la portée du canon, où est le  
 magasin des munitions qui regardent le service du canon, des armes à feu  
 &amp; des feux d'artifice. Le parc de l'artillerie est bien fortifié, &amp;  
 gardé avec des piquiers seulement. </def>

In such cases, <usg> is used to hold the statement of the function of the word. It is supplied with two attributes, a type so as to be able to extract all the words so designated and ‘ana’ to relate the term to a list of domains. The domain list is held in the header as a class declaration with a taxonomy so that a gloss and more detailed information can be added. The initial list was established using Atlas ti, and has been expanded by running a concordancer over the text file, however as the work progresses, other domains will be added. Each domain is identified using the xml:id attribute as follows:

<classDecl>  
 <taxonomy n="domains">  
 <desc>Domains et métiers designés par des termes</desc>  
  
 <gloss><name/></gloss>  
 <gloss xml:id="marine"><name>marine</name></gloss>  
 <gloss xml:id="blason"><name>blason</name></gloss>  
 <gloss xml:id="charpenterie"><name>charpenterie</name></gloss>  
 <gloss xml:id="mythologie"><name>mythologie</name></gloss>  
 <gloss xml:id="architecture"><name>architecture</name></gloss>

The usage tag then encloses the entire statement and also is declared as a term in the attribute so as to allow extraction of all the terms or only those from a given domain.

<sense><usg type="term" ana="#mythologie">Terme de Mythologie.</usg>

Not all terms are designated as such and no attempt is being made at this stage to add other non-designated terms.

### Comment and encyclopaedic information

**Spelling, grammatical and pronunciation comments**

As has been seen above with the entry for *Cacoa*, Basnage sometimes comments words in the <form> section. This may be grammatical information or a guide to pronunciation. In such cases <lbl> is used.

In the case of *cabestan*, Basnage gives information as to a variation in spelling as well on pronunciation.

<form type="lemmaGrp"><form type="lemma"><orth rendition="#uc"  
 >cabestan</orth>  
 <gramGrp><pos expand="Substantif">s.</pos><gen expand="Masculin"  
 >m.</gen></gramGrp></form>  
 <pron>L'<hi rendition="#i">s</hi>se prononce.</pron><lbl>  
 Quelques-uns écrivent </lbl><form type="variant"><orth  
 rendition="#i">Capestan</orth></form></form>

**Encyclopaedic information**

Basnage also comments in text adding other usage information, often with very encyclopaedic entries. This information is currently being encoded using <note>:

<note>Les Marchands ont d'ordinaire un <hi rendition="#i">abajour</hi> dans leurs magasins. La lumiere sombre qui entre par là, efface moins le lustre de leurs étoffes.</note>

Encyclopaedic information can be cursory or highly detailed with copious examples and references.

### Citations and examples

The TEI treats all quoted text as citations with <cit> acting as container for the <quote> element, and in the case of citations the bibliographical link with <ref>. The latter will then contain either <persName> or <title>. This allows us to distinguish between reference to a cited author, as Molière, or a work, as Oeuvres Mêlées by Saint Evremond. The reference to the header is coded using the ref attribute.

<cit type="citation">  
 <quote>La <hi rendition="#i">cabale</hi> est une science  
 sérieuse, &amp;amp; il n'y a que les mélancoliques qui s'y  
 adonnent.</quote>  
 <ref><persName ref="#Villars\_ISN0000000121185233"  
 rendition="#sc">ab de vilars</persName></ref>  
 </cit>

As with citations (below), no work should be cross-referenced if it is not listed in the header. Each example and each citation must be treated separately and never part of a batch.

<cit type="citation">  
 <quote>Dans les Princes le <hi rendition="#i">penchant</hi> au peché est  
 fortifié par la facilité de le commettre, &amp; par l'impunité après  
 l'avoir commis.</quote>  
 <ref><persName ref="#Flechier\_ISN0000000108744010" rendition="#sc">flec.</persName></ref>  
 </cit>  
 <cit type="citation">  
 <quote>Je fais gloire d'avoüer le <hi rendition="#i">panchant</hi> qui me  
 porte à l'amour.</quote>  
 <ref><persName ref="#Voit\_ISN0000000115879407" rendition="#sc">voi.</persName></ref>  
 </cit>  
 <cit type="citation">  
 <quote>Un aveugle <hi rendition="#i">panchant</hi>.</quote>  
 <ref><persName ref="#Corn\_ISN0000000121296128" rendition="#sc">corn.</persName></ref>  
 </cit>

#### Examples

The text is heavily illustrated with examples, which illustrate word usage, including in some cases the declination of the verb. The latter are always referenced to the source also, but also examples of usage rather than being for prestige purposes. Examples and citations are differentiated by a type attribute as either ‘example’ or ‘citation’. Attributes are in the language of the text, French, so that an example is ‘exemple’.

<cit type="exemple"><quote>Dieu n'<hi rendition="#i">abandonne</hi> jamais les siens au besoin.</quote></cit>

As with citations, when encoding several examples, each one is encoded separately.

<cit type="exemple">  
 <quote> Il l'a pris en trahison, il l'a pris <hi rendition="#i"  
 >parderriere.</hi> </quote></cit>  
 <cit type="exemple">  
 <quote>Il est bossu pardevant &amp; <hi rendition="#i"  
 >parderriere.</hi>. </quote>  
 </cit>

#### Citations

As with examples, citations illustrate usage, but more importantly they give a scientific caution to the encyclopaedic content as well as bringing in additional information. By giving bibliographical information, we have access to the knowledge network upon which Basnage was drawing.

<cit type="citation">  
 <quote>La haine entre les Grands se <hi>calme</hi>  
 rarement.<ref><persName ref=="#Corn\_ISN0000000121296128" rendition="#sc" >corn. </persName></ref></quote>

In the case of citations, the <bibl> element is used as above. The ‘ref’ identifier consists of the abbreviation of the authors name linked to the appropriate ISN code refers back to the information in the header. Basnage gives an alphabetical list of the authors cited, generally with the name of the authors and an abbreviation. The abbreviation is not always fully respected in the text, but it is still possible to link to the author in most cases. In other cases, non-indexed authors appear and these are added to a separate list in the header as they appear and are disambiguated. Again, the author is indexed along with an ISNI reference.

The reference generally comes after the quote and must be placed there. In some cases the person is introduced before in which case an <lbl> is included in the <cit> and the person referenced at this point.

<cit type="citation"><lbl><ref><persName ref="#Breb\_ISN0000000115563031">Brebeuf</persName></ref> fait dire par Cesar gourmandant ses soldats qui  
 vouloient l'abandonner;</lbl>  
 <quote> <lg rendition="#i"><l>De guerriers genereux changez en <hi rendition="#n"> populace,</hi></l>  
 <l>Allez, allez croupir  
 dans un calme odieux.</l></lg>  
 </quote>

As with domain names, this list has been added into the text header in the <sourceDesc> with appropriate identifiers as below.

<listPerson>  
 <head>Personnes indexés par Basnage comme source de citations</head>

<person xml:id="Aba\_ISN0000000083391828">  
 <persName>  
 <forename>Jacques </forename>  
 <surname>Abbadie</surname>  
 <roleName>Mr.</roleName>  
 <addName>Abadie</addName>  
 <abbr>abbrévié comme ab ou abs</abbr></persName>  
 <birth when="1654">né en 1654 à <placeName>Nay</placeName><country  
 corresp="#FRE">France</country></birth>  
 <death when="1727-09-25">mort le 25 septembre 1727 à <placeName>Marylebone  
 </placeName><country corresp="#GBR">Royaume-Uni</country></death>  
 <sex>Masculin</sex>  
 <nationality>Française </nationality>  
 <faith>protestant</faith>  
 <education>Éduqué aux académies de Puylaurens, de Saumur et de Sedan </education>  
 <occupation>pasteur, </occupation>  
 <affiliation/>  
 <socecStatus/>  
 <note>théologien protestant français, </note>  
 <event when="1699">  
 <label>Dean of Killaloe and Clonfert</label>  
 </event>  
 <idno type="ISNI">http://www.isni.org/isni/0000000083391828</idno>  
 <idno type="URI">https://fr.wikipedia.org/wiki/Jacques\_Abbadie</idno>  
 </person>

This is now a template and all the information is being updated to this format. Full details of the information located in the header will be given in the header section.

However, Basnage’s list of cited authors is incomplete and so a second list is being compiled with these additional authors as they appear in the text. Proper names may be encoded using <name>, but only cited authors and works are being indexed.

Other persons are encoded simply with the <persName> element. Disambiguation along with an ISNI when possible will be added later. If a person is indirectly cited, they are encoded as <persName>, direct quotations require the use of <ref>< persName>.

When a citation contains verse, this must be shown using the line group and line elements, <lg> and <l>. Given that the whole quote is in italics, rendition element is given on <lg> and the repeated headword the receives <hi> element with rendition as “n” for ‘normal’.

### Idiomatic data

The dictionary includes a large number of idiomatic formulae. Because the <colloc> element is restricted, we are using the <seg> element. As with other analytical issues, the differentiation is done using the attribute, in this case ‘collocation’. <seg> is for interpretation purposes and is not used when a headword is simply repeated in an example, or <seg> as this is highlighting with <hi> and the attribute for italics.

<seg type="collocation">Virer le <hi rendition="#i" >cabestan</hi></seg>

The collocations are predominantly verbal, with naval manoeuvres being a prime subject. For the moment, such collocations are being tagged and will be analysed in more detail when sufficient data is available. The BBI classification system provides a neutral descriptive system for collocation types that can be used in the present.

<seg> is being used for a variety of idiomatic data, and as often in such cases the dividing line between collocation and broader idiom is a fuzzy one. The following example could be seen as a colocation, but is here encoded as being a complete opaque idiomatic unit:

<seg type="#idiome"><hi>cacher</hi> son jeu </seg>

The same policy is being used for proverbs:

<seg type ="#proverbe">Chaud comme une <hi>caille</hi></seg>

Some proper names are close to being noun collocations, but obvious denominations are treated as whole units, such as:

<seg type="Nom-Propre"><hi rendition="#i" >Caisse</hi> des Empruns</seg>

The same procedure is applied to term variants, which are marked as such :

<seg type="term\_variant" rendition="#i" >demi-cabriole</seg>

<seg type="term\_variant" rendition="#i" >caisse de poulie</seg>

As these are analysable compounds, this information will be added later so as to extract terms and their variants by type.

**Etymologies**

Strictly speaking, these are more>less they are encoded with etymStrictly word histories than etymologies, nevertheless they are encoded with <etym>. At this stage, detailed mark-up is not being carried out other than naming the languages using <lang> and an appropriate xml:lang attribute, <foreign> for the words or wordings and <gloss> where a clear gloss is given. We use the three letter code throughout so French, fre, Spanish, spa, Italian, ita and English, eng. As the Greek quoted is ancient Greek, we use grc.

<etym> En <lang xml:lang="lat">latin</lang> <foreign rendition="#i" xml:lang="lat"  
 >abomasum</foreign><pc>.</pc></etym>

The most frequent etymological sources are Ménage and Du Cange:

<etym>Ce mot vient de <foreign xml:lang="lat" rendition="#sc"> capsa</foreign>  
<lang xml:lang="lat"> Latin</lang><pc>.</pc>.<ref><persName ref="#Men\_ISN0000000080815971" rendition=”#sc”>men. </persName></ref>&amp;  
 <ref><persName ref="#Saum\_ISN0000000108910154" rendition=”#sc”>>saum.</persName></ref>. </etym>

These are documented with bibliographic references, the second reference here being Claude Saumaise.

### Other elements

A number of other elements are used, amongst the most frequent <hi> and <foreign>.

<hi> is used for all text in italics other than foreign words or expressions. Whenever the headword is repeated in an example or citation, <hi> is used.

quote>A Rome, comme aujourd'huy, la <hi rendition="#i">cabale </hi>l'emportoit souvent sur le merite, &amp;amp; decidoit du sort des Ouvrages.</quote>

Foreign words are also in italic, but the use is clearly motivated and so they are contained in <foreign> accompanied by the relevant language code in the xml:lang attribute.

<foreign xml:lang="heb" rendition="#i">cabala</foreign>

Some of the citations are from verse. These are coded as a group of lines in <lg> with the individual lines in <l> as below with the citation from Molière.

<quote><lg rendition=”#i”>  
 <l> Son <hi rendition="#n">cagotisime </hi>en tire à toute heures des  
 sommes,</l>  
 <l> Et prend droit de gloser sur tous tant que nous  
 sommes.</l>  
 </lg></quote>  
 <ref><persName ref="#Mol\_ISN0000000123197131" rendition=”#sc”>mol.</persName></ref>

In the above case, the whole is in italics and so the entire line group receives the rendition attribute with <hi rendition="#n" > being used for the repeated headword using ‘n’ to signal normal.

## Header

As each letter is in a separate file, the header will be the same for each letter so as to group all in a full header when the entirety, or a large part, of the dictionary is available. This means that any change made in one file must be carried over to the others. Although the header is largely stable, it is the taxonomic information in the source and encoding descriptions that are changing as the project develops.

### File Description

### Bibliographic information

The reference list has been standardised so as be able to trace the knowledge network of Basnage. Thus, each entry is defined as <person> with <perName> containing the main nominative information with attributes for the xml:id used in the text. This consists of the abbreviation used by Basnage accompanied by the ISNI number. Strangely, xml:id cannot contain the colon character ‘:’ so the ISNI code is repeated in a ref attribute as this can be used to go directly to the ISNI database. ISNI provides unique identifiers for persons and so avoids ambiguity. The xml:base attribute leads to a recognised encyclopaedic entry for the person, generally in French, but if French Wikipedia has no entry then in the appropriate language. As many of Basnage’s authors were Protestants, entries may be in English or German, wherever the protestant diaspora had taken them. Other information concerns their dates of birth and death, their nationality and notes about their activities. Interestingly, the Trévoux Jesuits accused Basnage of not using French talents whilst <nationality> shows the accusation to be false. <event> describes important events in their lives, notably for those who were members of the *Académie Française*, whose date of entry is given.

<person xml:id="Abl\_ISN0000000107866286">  
 <persName>  
 <forename>Nicolas</forename>  
 <surname>Perrot d'Ablancourt</surname>  
 <roleName/>  
 <addName>Mr.</addName>  
 <abbr>Abl</abbr></persName>  
 <birth when="1606-04-06">Né le 6 April 1606 à  
 <placeName>Châlons-en-Champagne</placeName><country corresp="#FRE"  
 >France</country></birth>  
 <death when="1664-11-17">17 novembre 1664 à  
 <placeName>Paris</placeName><country corresp="#GBR"  
 >France</country></death>  
 <sex>Masculin</sex>  
 <nationality>Française</nationality>  
 <faith>protestant, puis catholique</faith>  
 <education>Sedan</education>  
 <occupation>academicien, traducteur</occupation>  
 <affiliation when="1637">Académie française</affiliation>  
 <socecStatus/>  
 <note>Protestant de temps à autre. Traducteur contesté - "les Belles  
 infidèles"</note>  
 <event type="Acad\_Fr" when="1637">  
 <label>entre à l'Académie Française</label>  
 </event>  
 <idno type="ISNI">http://www.isni.org/isni/0000000107866286</idno>  
 <idno type="URI"  
 >https://fr.wikipedia.org/wiki/Nicolas\_Perrot\_d'Ablancourt</idno>  
 </person>

The above example for ‘*abandonné’* cites Corneille. ‘CORN’ generally refers to Pierre Corneille, but as his brother, Thomas, is also cited, Basnage may refer to them both as Messires Corneille, Mrs Corneille. Consequently the header entry brings them together as below.

<personGrp xml:id="Mrs\_Corn">  
 <persName corresp="#CornT\_ISN000000012281648X"><forename>Thomas</forename><surname>Corneille</surname></persName>  
 <persName corresp="#Corn\_ISN0000000121296128"><forename>Pierre</forename><surname>Corneille</surname></persName>  
</personGrp>

The abbreviations used by Basnage are given using both what he gives in the index and what is actually found in the text.

## Input from VICAV Project (« VICAV Dictionary Encoding Guidelines » 2014)

**List of elements used**

**List of attributes used**

**Definition formulae**

* C’est le …
* Espèce de …
* On apelle aussi …
* Signifie aussi…
* Il signifie encore
* [paraphrase : <def>Diminuer le prix. </def>
* On dit