

AHRC / University of Sheffield / University of Liverpool

The Online Froissart Project

Manual for transcription and markup

Godfried Croenen & Natasha Romanova

Version 1.2
July 2010

Table of contents

1. Introduction	5
2. General transcription rules	6
3. Word separation	8
3.1. <u>General Rule</u>	8
3.2. <u>Forms in a period of transformation</u>	8
3.3. <u>Place names and personal names</u>	9
3.4. <u>Numerals</u>	9
3.5. <u>Other Cases</u>	10
4. Use of Capitals	12
4.1. <u>Geographical Names and Derived Adjectives</u>	12
4.2. <u>Titles</u>	13
4.3. <u>Definite Articles with Placenames</u>	14
4.4. <u>References to God</u>	14
4.5. <u>References to Countries and Institutions</u>	14
4.6. <u>'Saint'</u>	14
4.7. <u>Names of Feasts and Holidays</u>	15
4.8. <u>Numerals</u>	16
5. Use of Diacritics	16
5.1. <u>General Rules</u>	16
5.2. Acute Accent	16
5.3. Cedilla	18
5.4. Diaeresis	18
6. Punctuation marks	20
7. General encoding rules	22
8. Encoding the logical structure of the document	23
9. Encoding the physical appearance of the texts	25
9.1 Layout	25
9.2 Palaeography	26
9.3 Decoration	27
10. Canonical references	28
10.1. General	28
10.2. References to page numbers of printed editions	30

11. Gaps, errors, additions, corrections	31
11.1. Gaps	31
11.2. Scribal errors	33
11.3. Additions	34
11.4. Scribal corrections	35
12. Names and referring strings	35
12.1. General	35
12.2. The elements <name> and <rs>	37
12.3. The ‘type’ attribute	38
12.4. The ‘key’ attribute	40
13. Linguistic features	40
13.1. General	40
13.2. Direct speech	40
13.3. Quoted text	41
13.4. Proverbs	41
13.5. Technical terms	41
13.6. Text in other languages	41
13.7. Heraldry	41
13.8. Dates	41
14. The TEI header	41
Appendix 1: Elements allowed in the transcription files	42
Appendix 2: Values for rend@	42
Appendix 3: Values for type	42
Appendix 4: A sample TEI header	42

1. Introduction

Every edition or transcription, whether hand-written, printed or electronic, reflects a number of editorial decisions that necessarily have to be taken when trying to represent a medieval source. The present manual contains a detailed statement of the transcription and markup policies used by the **Online Froissart Project**. These take into consideration the main aims of the project and the specific nature of the texts which are its focus.

The **Online Froissart Project** is concerned with Jean Froissart's *Chronicles*, written in the last quarter of the 14th century, probably in a Picard dialect of Middle French. They survive in a relatively large number of — mostly 15th-century — manuscripts, some of which are incomplete or damaged. Book I survives in 57 manuscripts (representing 5 different redactions), Book II survives in 33 manuscripts (probably representing 3 redactions), Book III in 24 manuscripts (representing 2 redactions) and Book IV in 21 manuscripts (probably a single redaction). Apart from these there survive 3 manuscripts of the *Chronique de Flandre*, and 8 manuscripts containing an abridged version of one or more books of the *Chronicles*. To the text tradition should further be added the incunable and post-incunable editions (starting with the edition by Vérard, ca. 1499), which contain all four books of the *Chronicles*.

The *Chronicles* are a very substantial text, and each of the four books contains several hundreds of thousands of words. Book I is by far the longest part of the *Chronicles*, containing about 500,000 words; Book II contains about 270,000 words; Book III about 320,000 words and Book IV about 300,000 words.

Froissart's prose works are an important witness of 14th-century French language. The dialectal features of his language deserve scholarly attention, but the study of Froissart's language is complicated by the nature of the textual tradition of the *Chronicles*. Only a very small number of manuscripts, whose origins seem to be close to the author, both chronologically and geographically speaking, might be directly useful for studying the original language of the author. A substantial proportion of the surviving manuscripts of the *Chronicles*, however, were produced in Paris, during the first decade of the 15th century, and copied in a language which has been made to conform more to *francien* by removing many of the original picard features. Many of the later manuscripts, dating from the second half of the 15th century, originated in the Southern Low Countries. They were often probably copied after Parisian exemplars, at times possibly by Flemish-speaking scribes, in a language that also shows many signs of standardisation.

The time-lag between the writing of the *Chronicles* and the production of the surviving manuscripts, and the dialectal differences between the language of the author and those of the scribes, have undoubtedly had a substantial effect on the dialectal features found in the manuscripts. This is reflected in the transcription policies used by the **Online Froissart Project**. The transcription rules can be easily applied to manuscripts which reflect a wide chronological and dialectal range of linguistic features.

The transcriptions to be produced by the **Online Froissart Project** are mainly aimed at historians and literary scholars. As such the level of detail they contain on palaeographical and linguistic

features of the manuscripts is limited. A level of dense encoding, which would retain or try to represent many details of the palaeographical and orthographic features of each of manuscript, would require a substantial additional investment of time and effort, especially given the size of the textual tradition of the *Chronicles*. The potential pay-off of such a rich encoding would not justify implementing it for a substantial corpus like the *Chronicles*, and would be unfeasible at present. Scholars who want to study these features can directly access the manuscript images in order to do so.

The transcription policy which has been adopted allows the project team to transcribe and encode efficiently a substantial amount of text, while at the same time benefitting the largest possible audience. In practice, this means that texts are on the whole transcribed as they would be printed in modern critical editions. The XML encoding reflects two main aspects of the manuscripts: the logical structure of the text, and the appearance on the physical page. Additional markup gives access to historical commentaries (names).

In formulating the transcription guidelines we have followed in many instances the advice to editors of French medieval texts in:

P. BOURGAIN and F. VIELLIARD, *Conseils pour l'édition des textes médiévaux*, fasc. III. *Textes littéraires*, Paris, 2002.

A. FOULET and M.B. SPEER, *On Editing Old French Texts*, Lawrence, 1979.

[F. VIEILLARD, O. GUYOTJEANNIN (eds.)], *Conseils pour l'édition des textes médiévaux*, fasc. I. *Conseils généraux*, Paris, 2001.

The encoding guidelines found here reflect the advice given in the *TEI P5 guidelines for encoding of electronic texts*.

2. General transcription rules

The text of the manuscripts is transcribed as it would appear in modern critical editions. This means that abbreviations are to be expanded and that the orthography of the original manuscripts is retained.

Note. The rule governing the expansion of abbreviations also affects Latin abbreviations which may be occasionally found in the manuscripts. Thus “etc.” or “&c.” should be expanded to “et cetera”.

Certain aspects of the medieval orthography, however, require standardisation. For many letters of the alphabet medieval scribes used different letter forms. As far as late medieval book scripts are concerned, these include amongst others the long ‘s’ and different shapes of short ‘s’, two main forms of ‘d’ (straight ‘d’ and slanted ‘d’), and different forms of the letters ‘r’. All these

forms have no equivalents in modern type faces, and are thus transcribed using their single modern equivalent.

The situation is slightly more complicated as regards the medieval letters ‘u’ and ‘i’. Both medieval letters were used to represent different sounds: consonants and vowels (and semi-vowels). Moreover, both letters appear in several forms, resembling the modern letters ‘u’ or ‘v’, and ‘i’ or ‘j’ (short ‘i’, written as a minim sitting on the baseline, or long ‘i’, resembling a modern ‘j’). Like is the case with the other letters of the medieval alphabet the difference between different forms of the same letter found in the manuscript need not be recorded in the transcription and both letters should be represented by their modern equivalent. However, as both letters were used by medieval scribes to represent vowels as well as consonant, modern editors have long adopted the usage to distinguish between the two cases by using ‘u’ and ‘i’ to represent medieval ‘u’ and ‘i’ when used as vowels (or semi-vowels), and ‘v’ and ‘j’ when these same letters (in either form) are used as consonants. The spellings as vowel or consonant should of course reflect the sound value of the medieval word, not that of the modern French equivalent, in case of a sound change.

Note. The spelling of the medieval French verb ‘pouvoir’ (and its conjugated forms) is ‘pouvoir’, given the medieval sound value. The third person plural, however, is transcribed as ‘pevent’.

Although ‘s’ and ‘z’ are different letters in the medieval alphabet, in some late medieval book scripts ‘z’ is sometimes used as an equivalent of final ‘s’. An obvious example in the corpus relevant to the **Online Froissart Project** is the so-called Rome manuscript of Book I of Froissart’s *Chronicles* (Vatican City, Biblioteca Apostolica Vaticana, Reg. Lat. 869), where final ‘s’ at the end of line is often written as ‘z’. The phenomenon of writing ‘z’ instead of final ‘s’ is always linked to particular scribes and manuscripts. While it occurs in a number of manuscripts such as the ‘Rome’ manuscript, there are also a substantial number of other manuscripts containing the *Chronicles* in which this practice is not represented. As such, no uniform rule about the standardisation of final ‘z’ can be adopted, as it would be inconsistent to standardise final ‘z’ in some manuscripts to ‘s’ and not in others. Because of this, and because ‘s’ and ‘z’ are different characters in the medieval alphabet, the transcription of the letters ‘s’ and ‘z’ always reflects the spelling of the individual manuscripts and no standardisation from ‘z’ to ‘s’ is applied when ‘z’ appears in final position in a word.

The same principle applies to the transcription of the letters ‘i’ and ‘y’. In a small number of manuscripts these letters are both used to represent the sound [i], while in some other manuscripts ‘y’ is used as a variant form of ‘ii’. Since there is no consistency in these practices across the corpus of manuscripts, no standardisation is to take place. In the transcriptions both letters are therefore represented as they appear in the manuscript and no normalisation of the manuscript spelling is carried out (except that ‘i’ can be standardised to ‘j’, see *supra*)..

A similar situation arises as regards the letter ‘x’ in final position. In many manuscripts containing Old French texts final ‘-x’ seems to be used as an equivalent for ‘-ls’ or ‘-us’, and some editors of Old French text therefore treated final ‘x’ as an abbreviation and expand them. The same spelling practice is also found in many manuscripts containing the *Chronicles*, but in the transcriptions final ‘x’ is always transcribed as ‘x’.

3. Word separation

3.1. General Rule

The Online Froissart Project follows the established usage of most editors of medieval French texts to standardise word separation, thereby ignoring most of the time the presence or absence of spaces between strings of letters in the manuscript. The editors transcribe as one word each group of orthographic words ('word' here defined as 'a string of letters between two spaces in the manuscript') where only the group has a meaning and each individual word could not have a semantic and grammatical function outside the group. Such a practice facilitates the transcription (especially where it is not easy to determine whether indeed the words are separated in the manuscript) and is useful for concordancing and collation. In most cases, the decisions presented below are consistent with the decisions taken by the editors of the *Dictionnaire du Moyen Français*. In difficult cases of word separation (where there is no agreement among editors and lexicographers), editors on the Project adopt a consistent practice of transcribing a word (or a collocation of two or more words) in the same way across the manuscripts.

As a rule, the transcription does not reproduce the word separation as found in the manuscripts but imposes a system of word separation that is dictated by the logic of the language, adopted conventions and, where needed, the readings of other manuscripts:

Ex. Je me vueil fonder et ordonner sur les vrayes croniques jadis faictes et rassemblés par reverend homme et discret seigneur, monseigneur le Bel, chanoyne de Saint Lambert du Liege, qui grant cure et toute bonne diligence mist en ceste matiere, et la continua tout son vivant au plus justement qu'il peut. Et moult luy cousta a l'aquerre [*not* la querre] et a l'avoir.

3.2. Forms in a period of transformation

There are, however, two exceptions to this general rule, where the editors reproduce the word separation of the manuscript. This makes it possible to register certain morphological features of the language of the manuscript and make available substantial data for research into the evolution of Middle French morphology.

3.2.1. 'tres': a prefix (Lat. trans-) that evolved to become an adverb.

The transcription reflects the word separation found in the manuscript.

Ex. "Tresnoble sire" or "Tres noble sire" (depending on manuscript)

3.2.2. Relative Pronouns

Word separation of the manuscript is always respected in the case of relative pronouns "le quel/lequel", "la quelle/laquelle", "les quels/lesquels" and elided forms "au quel/auquel", "du quel/duquel" etc. Where the first element of the compound is written at the end of the line, the

transcriber makes a judgement about the scribe's practice.

3.3. Place names and personal names

3.3.1. Compound placenames and names of persons which in modern French are hyphenated, are transcribed as separate words:

Ex. Saint Malo, Saint Leu en Constantin
sire de Saint Aubin

3.3.2. Words which in their modern French form start with an elided form of the definite article are transcribed as separate words using an apostrophe.

Ex. la ville de L'Aigle (the historical spelling is Laigle, but this was changed to L'Aigle in 1961)

Note. Froissart seems to have been confused by the English name 'Derby', the first letter of which he seems to have understood as the elided form of the preposition 'de'. This usage is found in nearly all cases in the manuscripts. The phrase 'le conte derby' found in the manuscripts is therefore consistently transcribed as 'le conte d'Erby':

Ex. Et lui fist le roy faire son obseque aussi sollempnellement comme se feust esté son cousin germain le conte d'Erby

3.3.2. Compounds which in modern French usage are written as one orthographical word are also transcribed as one word:

Ex. Languedoc (placename)
langue d'oc (language)
Neufchastel, le Neufchastel (Newcastle in England)

3.4. Numerals

When transcribing compound numerals (written as words and not Roman numeral) modern French usage is followed in the case of hundreds, thousands and millions, i.e. elements of the compound are transcribed separately:

Ex. trois cents, deux millions

However, in the case of compound numerals which use 'twenty' ('vint') the second element is treated as a separate word and is not connected to the first element by a dash, as in modern French:

Ex. quatre vints, six vints

3.5. Other Cases

3.5.1 In most cases, compound terms are transcribed following modern French usage. Where a compound is hyphenated in modern French, it is transcribed as one word. The exceptions are compounds like ‘beau frere’, ‘belle seur’ etc., which are not considered compound words but rather noun phrases in Middle French, and combinations of the definite article followed by a form of the past participle ‘dit’.

- Ex.** aujourd’hui, aujourd’hui
arrieregarde
bienvenu,e (adj. ‘welcome’)
doresnavant (however, where words can be separated: d’ores en avant)
gentilhomme, gentilzhommes
horsmis (prep. ‘except for’, unless there is an agreement with ‘mis’)
le dit, la dite, du dit etc
madame, monseigneur (except where not followed by a name or a title and referring to a speaker’s wife, husband, lover or feudal lord, etc.)
malaise
nonchaloir
outrepasser (cf. passer outre)
saufconduit
tousjours, toutzjours
derechief
toutefois, toutesfois, touteffois (adverb)
sitost
sitrestost

Note: Froissart uses the word ‘endemain’ to mean ‘the next day’. Adverbial expressions which contain this word are transcribed as ‘l’endemain’, ‘a l’endemain’, except where the noun used clearly must start with the letter ‘l’ because it is preceded by another article: ‘le lendemain’.

3.5.2 In compounds which contain the preposition ‘a’, the preposition is normally transcribed separately:

- Ex.** **a certes** (adv. ‘surely, certainly’)
a dire (in the expression *savoir à dire* ‘être en mesure d’affirmer, de dire qqc. à qqn’)
a tant (adverb of time, ‘then, at that moment’)
a tant que (conjunction ‘until’)
a venir, ad venir (adj. ‘future’)
au dessus, au dehors (‘above’, ‘outside’)

As a rule, in Froissart's texts **a tout/atout** is used as a preposition 'with' where the lexical value of 'tout' ('all') is either lost or becomes superfluous. This is confirmed by the fact that in some cases 'tout' does not agree with the following noun in gender and number and the preposition is often used with non-quantifiable nouns. We treat **atout** as one lexicographical word (cf. DMF) and spell it as one word where there is no explicit agreement with the following noun (in feminine gender and/or plural number) and as two words when there is an agreement:

Ex. atout son fils, atout son ost
il s'en partit **atout** le grant tresor et s'en vint droit a Nantes
ilz yroient avec lui et avecques leur seigneur le duc de Normendie, chascun **atout**
tant de gens qu'il pourroit
Droit a ce point se repairoit ung chevalier qui avoit le guet fait celle nuit pardevers
l'ost, **atout** sa compaignie

Or vint le conte d'Erby devant la Riolo **a toutes** ses gens
et revenoit le soir **a toute** sa compaignie

Where the first consonant of the word following the preposition is doubled, the expression is transcribed as one word:

Ex. a faire, affaire (in the expression *avoir [moult, peu] a faire/affaire*)
a savoir, c'est assavoir
aucunes fois, aucuneffois

3.5.3 Compound prepositions and adverbs with the prefix **par-** are transcribed as one word:

Ex. paravent (preposition 'avant, antérieur à' adverb 'auparavant')
pardela
pardedela
pardevant
pardevers
parmi

Where the form of the adverb is even more complex, it is transcribed as separate words:

Ex. depar de dela

3.5.4. Compound conjunctions which contain prepositions **par** and **pour** are normally transcribed as separate words:

Ex. par quoi, par quoy ('this is why, because of which')
pour tant que ('since')

However, **pourquoi** (adverb in questions; conjunction) is be transcribed as one word:

Ex. **pourquoi** ...?
la raison **pourquoi**
ce **pourquoi**

3.5.5. As in modern French, the conjunction **quoique** is transcribed as one word, but when **quoi** is a pronoun and the object of the verb of subordinate clause, it is treated as one word:

Ex. Quant messire Bourcicault s'en parti de Paris, **quoyqu'**il donnast a entendre que il aloit celle part, il se faigny de prendre le droit chemin de Roleboise.

Si ne fut mie sitost ce traittié acompli, **quoyqu'**il feust entamé, mais fut longuement demené et tousjours aloit le roy d'Angleterre avant querant le grans païs.

Ex. Si commanda le duc que le pont feust fait, quoy qu'il coustast.

3.5.6. There is also a number of specifically 'medieval' words or expressions which do not have a modern counterpart. In these cases, the transcription standardises the word separation taking into account the decisions taken by lexicographers of Middle French and editors of Middle French texts:

Ex. **depar** (prep. 'from': 'le filz de la femme ne puet avoir droit de succession depar sa mere')
si comme conj. 'as, in the manner in which'
si que
siques (adv. 'so, thus': 'Siques après la mort du derrenier roy Charles, les XII pers et les barons de France ne donnerent point le royaume a la sueur qui estoit royne d'Angleterre')

4. Use of Capitals

The use of capitals follows modern French usage. Capitals are used at the start of a sentence and for proper names.

4.1. Geographical Names and Derived Adjectives

Names of places are spelled with a capital letter, but adjectives derived from geographical names are spelled with a small letter unless they are used as a noun.

Ex. Cy commencent les croniques de sire Jehan Froissart contenant les nouvelles guerres de France, d'Angleterre, d'Escoce, d'Espagne, d'Alemaigne et de Bretagne.

Ex. la garnison françoise
et tindrent les François frontiere
Car il estoit françoys de nom et d'armes

The exception, as in modern French, concerns the names of languages which are spelled with a small letter.

Ex. me dist en riant, en bon françois

References to nationality or allegiance placed after verbs **estre**, **se rendre**, **tourner** are adjectives and therefore begin with lower case letters:

Ex. Quant il aperchurent le bannierre dou connestable yssir hors et les Bretons ossi et congnuent tantost qu'il y avoit trayson de leurs pillars et qu'il estoient tourné franchois

Si le volloit tenir em prison et li pronmetoit bien que jamais n'en partiroit de la se il ne se tournoit françois mais se il volloit y estre franchois, il li donroit en Franche si grant terre et si belle revenue que bien li deveroit souffire et le marieroit hautement et ricement

References to nationality or allegiance, placed in apposition to a name, are treated as nouns and are therefore spelled with a capital:

Ex. Raymond des Sors, Gascon

4.2. Titles

The further use of capitals is kept to a minimum. Capitals are not used for titles:

Ex. le roy
madame
le pape
le dauphin
bourg ('bastard')

Where titles are used as part of a nickname or sobriquet, however, they are considered to be a name and transcribed with a capital:

Ex. le Captal de Buch
le Chanoine de Robertsart

4.3. Definite Articles with Placenames

Placenames which include a definite article in Froissart's texts are treated differently depending on whether these names are used with an article in modern French or not:

Ex. La Rochelle
le Liege
Le Qesnoi

4.4. References to God

Nouns and pronouns referring to God are spelled with capital letters. This includes the word 'Grace' used in medieval dates when it is not followed by an explicit reference to God.

Ex. l'an de Grace M CCC L
l'an de grace Nostre Seigneur M CCC L
l'an de Nostre Seigneur Jhesu Crist

je requier au Sauveur de tout le monde, qui de neant crea toutes choses, que Il vueille creer et mettre en moy, sens et entendement

4.5. References to Countries and Institutions

In accordance with the modern French usage, when certain words are used as reference to a unique institution or country, they begin with a capital letter:

Ex. Eglise [as Christian Church, not a building]
Empire [in references to the Holy Roman Empire]
College [College of Cardinals]

BUT: la grant eglise de Nostre Dame de Courteray

4.6. 'Saint'

4.6.1. In saints' names the name of the saint has a capital letter, but not the word 'saint':

Ex. saint Paul, saint Pierre
monseigneur saint Denis

4.6.2. Similarly, in reference to Christian community and Christian church, 'saint' is not capitalized:

Ex.: sainte Eglise, sainte chrestienté

4.6.3. There is, however, a small number of cases where the compound phrase, rather than the main noun on its own, represents the reference to a person, a saint, or divinity:

Ex. Sainte Vierge, Saint Esperit, Saint Sauveur
Saint Pere, Saint Pere le pape, nostre Saint Pere le pape

4.6.4. By contrast, geographical names and institutions that are derived from saints' names or contain the word 'saint' in their name are spelled with a capital letter for both the adjective 'saint' and the following first name:

Ex. Saint Omer, Porte Saint Antoine
Sainte Terre, Terre Sainte, Saint Sepulcre
Saint Siege
Saint College de Rome

4.6.5. If a saint's feast is referred to by the saint's name only, 'Saint' is capitalized as in modern French usage:

Ex. la Saint Jean, la Saint Martin

Otherwise 'saint' is not capitalized as in other references to saints:

Ex. le jour de saint Nicolas, la Conversion saint Paul
le jour saint Denis en octobre

Environ la Saint Jehan Baptiste, l'an mil III^C XLV, se parti le roy de madame la royne sa femme

4.7. Names of Feasts and Holidays

Names of feasts and holidays are capitalized. In the names of saints' feasts, 'saint' is only capitalized when the day is referred by the saint's name only (see 4.4.5):

Ex. Ainsi passa l'iver et Karesme jusques a Pasques
Karesme ensuivant
le vigile de saint Nicolas
le jour du Saint Vendredi
la Saint Nicolas

4.8. Numerals

Roman numerals are transcribed as capital letters. In order to indicate that certain letters should not be read as letters but rather as numbers, medieval scribes often used isolation points before, after, and sometimes in between the Roman numerals. These isolation points are not transcribed, as they have no function in a modern edition. White space is added between thousands, hundreds, tens and after any characters that are superscript.

Ex. M CC LX et V
III^{xx} III

5. Use of Diacritics

5.1. General Rules

The diacritics used in the transcriptions published by the Project are limited to three: the acute accent (é), the cedilla (ç), and the diaeresis (ï, ü). The general rules about the use of these three diacritic signs have been followed for a long time by editors of medieval French texts. They were first formulated by Mario Roques in 1926¹ and are elaborated in F. VIEILLARD et al. (eds.), *Conseils pour l'édition des textes médiévaux*.

Diacritics are used to help the modern reader with the interpretation of the text by providing indications about the sound value of the written words. They further help to avoid confusion between similar or identical medieval graphes, which had different sound values or were used to represent different lexicographical words. The decision to add diacritics, then, is always based on some knowledge about the sound value of the medieval word, but the information the diacritics offer in this respect is limited as they only make clear the distinction between stressed and unstressed final 'e' (acute accent), and whether or not a graphe 'c' represents a fricative or occlusive consonant (cedilla), and the number of syllables in some words (diaeresis).

5.2. Acute Accent

¹M. ROQUES, "Règles pratiques pour l'édition des anciens textes français et provençaux", in: *Bibliothèque de l'École des Chartes*, 87 (1926), 453-459; ID., "Établissement de règles pratiques pour l'édition des anciens textes français et provençaux", in: *Romania*, 52 (1926), 243-249.

The acute accent is only used on the letter ‘e’ (not on ‘a’ and ‘u’ as in modern French).

5.2.1. General Use

Acute accent is only placed on the ‘e’ in the final syllable of a word to indicate that the ‘e’ is stressed. As a rule, it is only needed in words with more than one syllable:

Ex. *advisié but gre, ble*

There exists, however, a number of homonym monosyllabic words. For these ‘é’ is used to distinguish between them. As a general rule, (Cf. EDC Manual, p. 48), ‘é’ is used in words semantically charged in order to separate them from the homographs that are grammatical tools (articles, prepositions, conjunctions). Necessarily, words of this latter group will be more frequently encountered in the texts.

Ex. *pres (‘near’) vs prés (fields)*

In the cases where a monosyllabic word has three different meanings and uses, two of which are as grammatical words and one as a noun, we only distinguish the latter one (a substantive meaning) by an ‘é’:

Ex. *des (indefinite article), des (since) vs dés (dice)
mes (but), mes (my) vs més (dishes)
tres (very), tres (since) vs trés (tents)*

5.2.2. In the transcriptions published by the Project, acute accent does not give any further indication about the sound value of ‘e’, for example whether or not it is open or closed. Grave accent (è) is not used, therefore, the spelling does not always follow conventions adopted in modern French or by some of the editors of medieval French texts:

Ex. *après not après, entrués not entruès/entrues*

5.2.3. Final Syllable Ending in ‘s’, ‘z’ and ‘t’

In the spelling system of many manuscripts covered by the Project the word ending ‘-ez’ often implies an emphatic ‘e’, but this is not systematically the case, and many manuscripts use ‘z’ as equivalent spelling to ‘s’ (especially at the end of the line, see above 2.). There is therefore a need to distinguish emphatic and non-emphatic ‘e’ when followed by ‘z’ and ‘s’ in final syllable. In the 14th- and 15th- century manuscripts, the word ending ‘-et’, however, always implies an emphatic ‘e’, therefore no accent is added in these cases.

Ex. *travilliet, adviséz, après*

In the transcriptions of those manuscripts where a consistent programme of corrections was implemented by a later hand to remove or attenuate dialectal traits (e.g. Paris, BnF, ms. fr. 5006),

both the original and the corrected versions are transcribed and encoded with a <subst> tag. Where alternative transcriptions could require different spellings of ‘e’ (with or without acute accent) depending on the reading chosen, the original reading, not the corrected one, is always given precedence and use of diacritics is predicated upon the original reading:

Ex. si furent grandement festiie<subst>t<add>s</add></subst>
Ils furent mout esmerveille<subst>t<add>s</add></subst>
tant furent mene<subst>t<add>s</add></subst>
la truie que amené<subst>s<add>e</add></subst> avoient
je vous ordonne a estre chié<subst>s<add>f</add></subst> de nos
gens

5.2.4. Double ‘e’ at the End of a Word

In the combination ‘ee’ at the end of a word, ‘é’ is only added when the second vowel is emphacized:

Ex. pensee vs créé

5.3. Cedilla

Cedilla (‘ç’) be used to indicate that the letter ‘c’ represents [s] sound before *a, o, u*.

Ex.: Si fut esbay et commença a sonner et a corner en sa buisine

5.4. Diaeresis

5.4.1. General Rule

The texts edited by the project are written exclusively in prose (with the exception of a very small number of quoted verses in Book I), therefore the number of syllables in individual words is not critical for the enjoyment and performance of these works. Therefore, diaeresis, or “tréma”, is used very sparingly by the Project following the EDC guidelines² and when used indicates that vowels are placed in hiatus, in order to distinguish between different words and different forms

²‘Dans les textes en prose, le tréma doit être utilisé avec prudence et parcimonie’, p. 51.

of a verb spelled identically. The diaeresis is only used on the letters ‘i’ (also ‘y’) and ‘ü’.

5.4.1. Diaeresis on i (y)

5.4.1.1. Diaeresis is used to indicate difference in pronunciation and avoid confusion between homographs:

Loïs (form of the proper name ‘Louis’) and **lois** (‘laws’)
oïr (infinitive form of the verb) and **oir** (‘heir’)

5.4.1.2. Diaeresis is used to indicate the third person singular preterit forms of the verbs ‘obeir’ and ‘desobeir’ where they are spelled identically to the past participles:

Ex. Le Captal, qui desiroit les armes et qui estoit par lignage tenuz de servir son cousin le roy de Navarre, obeÿ et parti du conte de Foix [preterit]

Ex. Quant messire Galleas tint son oncle, comme dit est, et qu’il se vey seigneur obey de toute Lombardie [past participle]

5.4.1.3. Diaeresis is used to indicate preterit forms of the verbs ‘oyr’ (‘ouyr’), ‘joir’ (and ‘resjoir’) and ‘traire’ (and ‘retraire’) where there exists a possibility of confusion with the past participle:

Ex. La oÿ messe le roy et disna’ [preterit]
si comme vous avéz oy [past participle]

Ex. tantost qu’il ouÿ nouvelles de la venue du roy anglois [preterit]
Vous avéz ouy parler de la bataille de Crecy [past participle]

Ex. Adont se retraït chascun en sa maison [preterit]
Et puis lui fut le ventre ouvert et le cuer trait hors [past participle]

When the preterit form of ‘oir’ ends with a ‘t’, diaeresis is not used as there is no possibility of confusion:

Ex. Quant il vit que les François ardoient ainsi le païs, et oyt les povres gens crier et plourer

There is no need for a diaeresis in other forms or when the third person singular preterit is ended in ‘st’:

Ex. tout considere c’estoit bien qu’il se parteist et se traist vers son royaume
Tantost se trairent devers lui Guisebrecht Mahieu et ses freres

In other cases, diaeresis is not added on i and y:

- Ex.** e+o+i(y): escheoir, veoit, veoir
Ex. e+i(y): empreis, meist, preist, deist, meschey, pourvey, logeis, obeissance
Ex. ou+i/y: enfouy, resjouye, ouie

Note: the forms of the word ‘yes’ **oui, ouy, ouil, ouyl, oil, oyl** are equally be spelled without a diaeresis

5.4.2 Diaeresis on ü

The combination e+u: diaeresis is only added above the letter ‘u’ in this group in the case of homographs to avoid confusion. It is only used in those words where confusion may otherwise be possible, but it is not normally use in related words:

- Ex.** veü, veüs (past participle of ‘veoir’, ‘to see’) vs. veu, veus (form of the masculine noun veu ‘vow’, ‘wish’)
 [but ‘veue’, ‘pourveu’]

seür (adj.) vs. seur (form of ‘soeur’)
 [but “asseurer”, “asseuree”]

peü (past participle of pooir ‘to be able to’) vs. peu (= ‘peu’; cf. “pou”)
 [but “peurent”, “peussent”, etc.]

meür (adj. = ‘mûr’) vs. meur (= ‘moeur’)

creü (p.p. ‘croire’) vs. creux (‘hollow’)

In most other cases where the combination ‘eu’ represent two vowel sounds the ‘tréma’ is not used as there is no possibility of confusion:

- Ex.** “sceu”, “sceue”, “ramenteu”, “eu”, “resceue”, “escheu”, “apparceu”, “pleu”, “eust”, “apparceust”, “feust”, “sceust”, “peust”, “geurent”, “peurent”, “feussent”, “receut”, “apparceut”

6. Punctuation marks

Medieval punctuation is of interest to palaeographers and historical linguists, but their investigations of this aspect of medieval manuscripts are highly specialised. The transcriptions produced as part of the **Online Froissart Project** punctuation are not meant to assist with research of this nature. The punctuation signs used by the project in the transcriptions are mainly there to help the modern reader by indicating syntactic units (phrases and sentences). Palaeographers and historical linguists who are interested in the punctuation systems used by the scribes of the Froissart manuscripts should refer to the original manuscripts or digital facsimiles.

Medieval punctuation is only vaguely related to modern punctuation, although many of the signs used in medieval manuscripts show a certain resemblance to modern punctuation marks. Incunabula use diacritical signs whose appearances are much closer to modern print characters, but like the contemporary manuscript signs the usage and meaning of these differ from the modern punctuation signs. The latter are indeed derived from their historical counterparts, but they differ from them both in their outward appearance and in their usage and meaning.³

Because of the difference between medieval and modern systems of punctuation, modern critical editions of French medieval texts usually do not try to represent the medieval punctuation. Editions or transcriptions which attempt to represent medieval punctuation via typographical means are sometimes called ‘diplomatic’ editions or transcription, but these are often very inadequate representations of the medieval palaeographical reality. For an adequate representation of these signs the normal typographical punctuation symbols would not suffice and more complex markup would be needed.

The **Online Froissart Project**, for methodological as well as practical reasons, does not attempt to represent the medieval punctuation in an unambiguous way (by representing each scribal phenomenon in a different unambiguous manner). Neither does it adopt the attitude of many editors, which often completely ignore the manuscript punctuation, and apply instead their own punctuation, based on their own editorial judgment. Instead, the approach adopted by the **Online Froissart Project** aims for something between these two extremes in order to represent to certain degree the medieval punctuation, while at the same time providing a minimum level of punctuation in order to render the text intelligible to the modern reader.

In providing punctuation in our transcriptions we have followed a number of general guidelines. A useful approach to medieval punctuation is to think of these markers in terms of weak and strong/emphatic signs, which indicate weaker or stronger caesuras in the grammatical structural, as well as in the rhythm of the text.⁴ In the transcriptions the presence of a *punctus* (whether or not it is followed by a capital letter) or words starting with capital letters or other emphatic forms of letters (whether or not preceded by a *punctus*) is interpreted as an indication that the scribe wanted to signal a strong or emphatic caesura. Whenever the grammatical structure of the text allows this, the presence of this emphatic caesura is signalled in the transcription by a full stop (or an equivalent sign like an exclamation mark). The presence of oblique strokes (resembling commas) have been interpreted as weak caesuras and have, if the grammatical structure allowed this, are rendered as commas in the transcription. In many cases linebreaks act as the equivalent to punctuation signs, insofar as that whenever they coincide with weak caesuras, scribes usually did not add oblique strokes to their transcriptions. In such cases no commas have been added to the text.

Whenever a manuscript has no such punctuation at all in a passage which clearly contains more

³For an exhaustive study see Malcolm B. Parkes, *Pause and Effect: An Introduction to the History of Punctuation in the West* (Aldershot: Scolar Press, 1992).

⁴Christiane Marchello-Nizia, ‘Ponctuation et « unités de lecture » dans les manuscrits médiévaux ou : je ponctue, tu lis, il théorise’, *Langue Française*, 40 (1978), 32-44.

than one syntactic unit equivalent to a modern sentence, punctuation has been introduced to reflect the sentence structure as far as the modern transcriber can interpret this.

Certain letters, when they appear at the start of word, are often rendered in an emphatic form ('r' in form of a capital in many of the 15th century manuscripts). These emphatic letter forms are normally not interpreted as signalling an emphatic caesura, and therefore in such no new sentence is started.

Isolation points around Roman numerals are also not rendered in the transcription because their function is not syntactic.

Whenever it is impossible to align the medieval punctuation with the syntactic structure as interpreted by the transcriber, medieval punctuation signs (or capitalisation) are simply ignored and modern punctuation is imposed. In all cases punctuation has always been added sparingly.

When consulting the transcriptions users should therefore keep the following in mind. Whenever the transcription signals the start of a new sentence, this usually means that there is an indication of an emphatic caesura in the manuscript (punctus and/or capitalised word). In some cases there is no such punctuation in the manuscript, but punctuation is needed for the modern reader. Whenever there is a comma in the transcription, this indicates that there is either an oblique stroke in the manuscript or a syntactic need for an added punctuation sign.

Questions marks and exclamation marks have no graphic equivalent in the Froissart manuscripts but have been rendered like the other emphatic caesuras by the medieval scribes. It is therefore the interpretation of the transcriber which guides when they are to be used in the transcription. The use of other modern punctuation signs is kept to a minimum. The use of colons is avoided, if possible, unless at the end of short phrases of narratorial voice introducing direct speech. The use of semi-colons is also kept to an absolute minimum, although they are sometimes used to separate items in long enumerations which are syntactically part of the same sentence.

7. General encoding rules

All the transcriptions made for the **Online Froissart Project** are encoded in XML and follow the TEIP5 schema and guidelines. The TEI encoding of the transcriptions provides information about three different aspects of the original manuscript and the text it contains: 1. the logical structure of the manuscript text, 2. the codicological features of the manuscript and the appearance of the text on the manuscript page (layout), 3. basic interpretations of linguistic and historical features of the text. Moreover, the TEI header also contains metadata about the manuscript transcribed, as well as documentation about the electronic file (work log).

The encoding of the logical structure establishes a sequence of sections (manuscript chapters), each of which contains a number of generic (paragraphs) and specific (rubrics, titles) elements which contain the transcribed text. The segmentation of the text is normally indicated in the manuscript by various elements of decoration, which follows a more or less complex hierarchy (rubrics, initials, miniatures, paragraph signs). Apart from the encoding of highlighted letters or text segments (different scripts, decorated letters), the encoding of the physical appearance of the

text also relates to the precise distribution of words and characters (in quires, pages, columns, lines). Linguistic features of the text that are encoded include direct speech, quoted text, the use of technical words, and words or phrases in other languages. Historical information present in the texts is encoded through the markup of names and other referring strings.

Each electronic file normally corresponds to a single manuscript volume. When a manuscript volume has at a later point in time been split into two or more parts and bound as separate volumes, the original state of the manuscript is given precedence and the whole transcription is kept in one electronic file (this applies for example to Brussels, Royal Library, ms. IV 251, vols. 1 and 2, and Paris, BnF, ms. fr. 6477–6479).

In the opposite case, where (parts of) separate manuscripts have been bound together, each transcription file corresponds to what is likely to have been originally a single separate volume. This applies for example to the single folios which survive from an otherwise lost set of Books I–III, now Brussels, Royal Library, ms. II 88. The transcription of the text of these fragments is divided up into three electronic files, each of which correspond to what is likely to have been the three manuscript volumes making up the original lost set (separate files corresponding to each book of the *Chronicles*).

If a manuscript volume contains more than one Book (or parts of more than one Book) of the *Chronicles*, it is transcribed in a single electronic file (although for practical reasons such transcriptions may be held in separate files until the transcription and markup processes have been completed).

If a manuscript contains texts other than Froissart's *Chronicles*, these other texts are omitted from the transcription, unless they have been integrated into the text of the *Chronicles* in such a way that they cannot be easily distinguished from the *Chronicles*. Thus, in most of the manuscripts of the A redaction of Book I, the interpolations from the *Grandes Chroniques* for the years 1350–1356, have been copied by the medieval scribes as if they formed part of the text of Froissart's *Chronicles*. Therefore, the transcription files corresponding to these manuscripts contain these interpolations. Similarly, the 'lettres de défi' exchanged between Duke Louis of Orléans and King Henry IV of England, transcribed by the scribes of Brussels, Royal Library, ms. IV 1102 and Paris, BnF, n.a.f. 9606 following the text of Book II of the *Chronicles*, are included in the transcriptions of these manuscripts. (To allow for comparison, the same 'lettres de défi' are also included in the transcription of Leiden, University Library, VGGF 9-2, even though in the latter manuscript they are clearly a later addition).

8. Encoding the logical structure of the document

The encoding scheme proposed by TEI favours the underlying logical structure of the text over other possible views of the same source, including encodings which focus on the physical appearance and layout of the text, or encodings which are optimised for displaying or querying the documents. The logical structure of the document is encoded through a hierarchically organised structure of XML elements which contain either further XML elements, or text, or a combination of XML elements and text. The alternative views of the document (for example codicological structure of the codex), essentially alternative hierarchies, cannot be encoded as

such using current XML technologies, which only allow a single hierarchy of XML elements. The solution proposed by TEI and adopted in the encoding used by the **Online Froissart Project**, is to use empty elements, which are essentially allowed anywhere in the encoded text, to mark the beginning and ending of the elements constituting these alternative hierarchies (for example, the beginning and ending of a page is encoded through the <pb/> element).

Each transcription file contains a <teiHeader> (with metadata) and a <text>. In its most simple form a <text> contains a <body>, with one or more sections (encoded with the generic element <div>, for ‘division’), each of which contains one or more paragraphs (<p>). More complex hierarchies can be created through the use of nested divisions (for example, ‘parts’ containing ‘chapters’) or the addition of titles or headings at the start of divisions. <text>s can also contain sections which do not belong to the <body> of the <text>, but to the <front> or <back> matter. These are typically tables of rubrics, either placed in the front or the back of the manuscript.

To establish the logical structure of the text as it appears in the manuscript transcribed, the transcriber takes into account elements of layout and formatting. Each book of the *Chronicles* is encoded as a sequences of sections or ‘manuscript chapters’. In manuscripts which have been rubricated, each rubric normally introduces a new chapter. In such cases the chapter is marked up with <div type=“chapter”>, and the text of the rubric marked up as <head type=“rubric”>. If chapters are numbered, then the number should be recorded as the value of the ‘n’ attribute on <div>, using Arabic numbers.

Each chapter contains one or more paragraphs. The start of paragraphs is indicated in the manuscripts either by large or decorated initials (which are marked up accordingly), or by paragraph signs, which are transcribed as ‘¶’ (Unicode 00B6) and marked up if they are in any way decorated.

Ex. <div type="chapter" n="2">
<head type="rubric">Comment le prince et la princesse vindrent a Tarbe et de la requeste que le conte d’ Armignac fist au prince et a la princesse de Galles et comment le paÿs de Gascongne estoit lors traveillié et guerroyé par garnisons. Chappitre II. </head>
<p><hi rend="champie, 5 lines">E</hi>nre la conté de Foix ...</p>
<p><hi rend="red and blue">¶</p> Pour ce temps que ...</p></div>

In manuscripts without rubrics, or with only a single rubric at the start of the whole text, the sequence of manuscript chapters is determined on the basis of the sequence of large or decorated initials, whereby each such initial is interpreted as signalling the start of a new manuscript chapter. In manuscripts without rubrics these manuscript chapters (interpreted in this way) are usually shorter than in manuscripts where the chapters are indicated by rubrics. In the latter, the manuscripts often correspond to two, three or four manuscript chapters found in the manuscripts without rubrics.

Tables of rubrics are marked up as a <div type="table">. Such tables may have their own rubric, marked up with the normal <head type="rubric">. The list of rubrics in a table are not marked up as <head> elements, but instead as <item>s in a <list> element. If they are numbered, the numbers are recorded as the value of the ‘n’ attribute, equivalent to the way this is done for numbered chapters.


```

Ex. <back>
<div type="table">
<head type="rubric">Cy commence l'autre partie principale des cronicques. </head>
<list>
<item n="1">et premierement parle comment il se parti de France. Chappitre <num
value="1" type="ordinal" rend="Roman">I<hi rend="superscript">e</hi></num></hi>.
</item>
<item n="2">Comment le prince et la princesse ...</item>
....
</list>
</div>
</back>

```

9. Encoding the physical appearance of the texts

Some elements of the layout and physical appearance of the text in the manuscript inform the encoding of the logical structure of the text and are as such recorded implicitly in the markup that identifies the logical structure of the document (start of chapters and paragraphs). There are however further aspects which are not directly linked to the logical structure of the text, such as the distribution of the writing on the manuscript pages and aspects of decoration. These elements are also recorded in the encoding with the use of appropriate elements.

9.1 Layout

The layout of the text, the distribution of the writing on the manuscript page, and the codicological structure of the manuscript are encoded using a set of empty elements, which in each case immediately precede the respective unit (<lb/> at the start of a line, <cb/> at the start of a column, <pb/> at the start of a page, <qb/> at the start of a quire, <milestone type="volume"/> at the start of a volume). Each type of a higher order will normally trigger the insertion of at least one occurrences of the elements of a lower order (unless no text whatsoever appears before the next occurrence of the same element). Thus a column break will immediately be followed by a line break; a page break will be followed possibly by a column break (if the page layout has two columns) and a line break; a quire break will be followed by a page break, possibly a column break, and a line break.

```

Ex. ‘françois, <pb n="2 r"><cb n="A"><lb>ne oncques ne s'en pot excuser'
'...<trailer lang="la"><hi rend="colour-stroked yellow">E</hi>xplicit tabula.
</trailer></div><pb n="5 v"><pb n="6 r"><pb n="6 v"><pb n="7 r"><pb n="7 v"><pb
n="8 r"><pb n="8 v"></front><body><pb n="9 r">...'

```

No default expectations should be attached to the presence of any of these empty elements in relation to word boundaries. Given that the transcriptions of the **Online Froissart project** do not have explicit encoding of individual (orthographic or lexicographical) words the transcriber follows a practice of standardising inter-word spaces to indicate word boundaries (cf. *supra* 3. Word separation). The presence of spaces between words helps the reader but is also necessary to allow for easy further processing of the text (tokenisation, lemmatisation). In the encoding of

the **Online Froissart project** no spaces are left between the empty elements indicating line breaks etc. and the letters or (non-space) characters which follow it, nor between the successive elements which mark the break indicated by the empty elements. If a line break (column break, page break, ...) occurs inside an orthographic word no space is left in front of the empty element either. When the line break (column break, page break, ...) coincides with a word division a space is left in front of the line break but not after.⁵

Although most manuscripts transcribed by the **Online Froissart Project** have numbered folios rather than numbered pages, folio breaks as such are not encoded. Instead each side of the folio is marked up as a page and a page break element (<pb/>) is inserted at the start of each.⁶ The label attribute ('n') is used to record the page numbers. These are composed of the folio number followed by a space and 'r' or 'v' (for 'recto' and 'verso'). When no folio numbers are present in the manuscript a numbering is supplied in square brackets. Corrections to the numbering present in the manuscript are also indicated by square brackets.

Ex. 'françois, <pb n="[2] r"><cb n="A"><lb>ne oncques ne s'en pot excuser'
'françois, <pb n="2 [bis] r"><cb n="A"><lb>ne oncques ne s'en pot excuser'

When more than one folio numbering is found in the manuscript, only the most authoritative one (usually the most modern) is recorded in <pb/>. Roman numerals are converted to arabic numerals for the value of 'n', although a faithful rendering of the medieval numbering can be transcribed using the <fw> element (see *infra*).

For the other elements the traditional designations are used. Columns are labelled 'A' or 'B' on each side of the folio (the values 'C' and 'D' would in theory be used for the third and fourth column on a page, but no four-column manuscripts of the *Chronicles* are known to exist). Quires are numbered continuously in arabic numbers. Lines are not numbered.

In some cases the 'rend' attribute is used on <lb/> to record particular aspects of the distribution of the text on ruled lines. One such case is when the medieval scribe used a horizontal or oblique line (linking stroke) to indicate that an orthographic word is being split over the line. When in this part of the transcription the word division of the manuscript is followed, the <lb> element will occur in the middle of a word. The 'rend' attribute is then used to record that a linking stroke occurs in the manuscript. This is only used for line breaks which occur inside words that are transcribed as orthographic words in the transcription.

Ex. '... ne onc<lb rend="hyphen">ques ne s'en pot excuser ...'

Other values for the 'rend' attribute on <lb/> clarify more complex cases where a single ruled line contains text segments which do not logically follow on from each other. This is the case when

⁵If no space is left it would be impossible without further encoding (in particular the use of rend or type attributes) to distinguish linebreaks occurring inside words from those occurring in between orthographic words.

⁶The TEI has indeed no equivalent 'folio break' element.

the text of a rubric is distributed over several lines but does not take up the complete lines, while the text of the following chapters fills the rest of these ruled lines.

9.2 Palaeography

The encoding does not record palaeographical features of the writing, but does normally identify changes of scribal hand. This is achieved by the insertion of the empty element `<handShift>` just before the first character written by the new hand. The `<handShift>` element requires a reference to a definition or description of the hand found elsewhere in the transcription (normally in the header).

Ex. `<head type="rubric" rend="rubricated"><handShift new="#TNYG">Ci commentent ...'`

Note: in order to be able to use the `handShift` element one needs a corresponding `handNote` statement in the header: `<handNote xml:id="TNYG">One scribal hand</handNote>`

9.3 Decoration

9.3.1. Decoration of textual elements

The decoration in the manuscripts can be classified as decoration of textual elements (e.g. initials) and non-textual decoration (border illumination, miniatures). Decoration of textual elements is encoded through the use of the 'rend' attribute, mainly on `<hi>` elements, unless the text to be marked up is already completely inclosed in another element to which the 'rend' attribute can logically refer (typically `<head>` or `<trailer>`). The presence of the general purpose `<hi>` element indicates that a particular string of letters is set off from the rest of the text because it has been executed differently or because it is in some way decorated. The 'rend' has as value a comma-separated list of values which, taken together, indicates in what way the text is set apart. The values in this list refer to the execution of the letters (initial, lombard, display script), the size (in ruling units), colours and placement on the page (centred). Each atomic value is meaningful on its own and its interpretation is normally not predicated on the other elements in the comma-separated list.

Ex. `<hi rend="illuminated initial, 5 lines">A</hi>fin que ..'`
`<hi rend="champie, 2 lines">V</hi>ous avéz bien ouy ...'`

Colours indicated should refer to the whole letter or string of characters. In case of colour-stroking the indication of the colour used to highlight the letter should not be separated from the rest by a comma (which would indicate that the letter was executed in such colour, with colour-stroking in another unspecified colour) but should be part of the statement.

Ex. `<hi rend="colour-stroked red">d</hi>essoubz la ...'` **not** `<hi rend="colour-stroked, red">d</hi>essoubz la ...'`

When initials have been executed in several colours these colours should not be separated by commas, but a single phrase should be used, with first the primary colours, and then introduced by ‘with’ an indication of any secondary colours. Champ initials, which are executed in a particular combination of red, blue, gold, black and white, are simply encoded as ‘champie’, without any further indication of the disposition of the colours.

Ex. ‘<hi rend="pen-flourished initial, blue with red, 2 lines">Q</hi><hi rend="display script">uant</hi> ...’
‘<hi rend="champie, 2 lines">B</hi>ien est ...’

A complete list of acceptable values of the ‘rend’ attribute, with their interpretation, is given in Appendix 2.

9.3.3. Miniatures

The TEI encoding scheme is not particularly well-suited for the encoding of non-textual decoration in the transcriptions. Descriptions of the miniatures are therefore not included in the transcription at the junction where they are found but are found in the codicological descriptions of the manuscripts. In the transcription only the presence and location of the miniatures is indicated by the use of an <anchor> element which allows the logical reference to the miniature descriptions elsewhere. These anchor points have as value for the ‘type’ attribute ‘miniature’ and have an identifier which is unique across the corpus of transcriptions, composed of three components: the MS sigil (here ‘Bes-2’), followed by ‘-M’, followed by the rank number of the miniature inside the manuscript volume.

Ex. ‘<qb n="2"><pb n="1 r"><lb><anchor type="miniature" xml:id="Bes-2-M1"><lb><lb>...’

A line break element precedes the <anchor> point to indicate the top line covered by the miniature. A further number of <lb> elements is inserted to a total which is equal to the height of the miniature (in ruling units).

10. Canonical references

10.1. General

The markup of the transcriptions using the TEI P5 schema aims at representing closely the manuscript copy transcribed. Chapter divisions follow those present in the manuscript. However, in order to enable the alignment of the transcriptions from one manuscript to the next needed for the collation of witnesses and for easy navigation of the text across different manuscript copies, canonical references are also added to each transcription file. This system of canonical references provides a grid that allows each section of text to be identified with reference to the standard editions of the *Chronicles*. The empty element used to record these canonical references is <milestone/>, with the ‘ed’, ‘n’ and ‘unit’ attributes set to appropriate values.

The canonical references are as far as possible based on the Société d'Histoire de France edition of Froissart's *Chronicles*, which covers Books I to III (for Book IV the chapter division of the Buchon 1835 edition should be used). The SHF edition divides the *Chronicles* into a series of numbered 'chapters' (not the *chapitres* indicated in the *sommaire* of each volume, but rather the *paragraphes* indicated in the printed edition with the § signed followed by a number). In the transcription file the beginning of each such chapter, when present in the transcribed manuscript, is indicated with a milestone, where the value of the 'ed' attribute is "SHF", the value of 'unit' is "chapter", and the value of the 'n' attribute is a character string composed of a number between 1 and 3 (indicating which book of the *Chronicles* this belongs to), followed by a hyphen, followed by the number of the chapter in the SHF edition.

Note. The Prologue of Book I in the SHF edition has not been numbered, but has been given the value '1-0'.

Note. In the SHF edition of Book III the number '42' has been used twice; for the second chapter numbered '42' the label '3-42bis' is used instead.

Ex. <milestone ed="SHF" unit="chapter" n="2-63"/>
<milestone ed="SHF" unit="chapter" n="3-42bis"/>

Note. Some chapters are only present in particular versions of the *Chronicles*, which are not always the versions used by the SHF as the base version for their edition. In such cases bis numbers have been used to number these additional chapters. These include the following chapters: 1-223bis in the 'Amiens' version of Book I, 1-780bis in the 'A' version of Book I, 2-100bis in the *Chronique de Flandres* (prologue of this version of Book II), 2-300bis in the *Chronique de Flandres*.

The mechanism explained above is sufficient to align text blocks across manuscript witnesses in most cases, but in some circumstances extra alignment points are needed. At times it is necessary to insert extra milestones for the processing with the *Collate* software, which is sometimes unable to process long chapters of the SHF edition. These extra alignment points have often been inserted where the SHF edition starts a new paragraph (indicated typographically by a carriage return and tab). The extra alignment points are encoded using the same <milestone/> element, but 'unit' is set to "subchapter". The value for the 'n' attribute is formed in the same way as above, but is followed by a hyphen and a number indicating the paragraph (beginning with the second paragraph; there is no need to insert an extra milestone element for the first paragraph, as this coincides with the start of the chapter).

Ex. <milestone ed="SHF" unit="subchapter" n="2-63-4"/>
<milestone ed="SHF" unit="subchapter" n="3-42bis-2"/>

A similar situation occurs when *Collate* is unable to align textual witnesses, because they diverge significantly from the base text, either because of damage which has resulted in text loss, or because textual variation makes the alignment difficult. The boundaries of these gaps or text blocks which display higher than normal variance, which normally do not align neatly with the SHF chapters, need to be precisely indicated in the parallel witnesses. In these cases extra <milestone/> elements are inserted at the start and end of each passage where alignment is

possible. In the case of damaged witnesses, appropriate `<gap/>` elements are inserted, and a `<milestone/>` element is inserted in front of these to indicate where the text loss starts. The values for the ‘n’ attribute is in these cases constructed using the number of the Book, followed by the number of the SHF chapter, followed by a sigil for the manuscript, and, if more than one such milestone needed for this witness, an extra number, all separated by hypens.

Note. If the passage which has been lost in a witness spans text between more than two `<milestone/>` elements, only the first and last `<milestone/>` have been added to the encoding.

Ex. `<milestone n="3-1" ed="SHF" unit="chapter"/>`
`<div type="chapter">`
`<p><gap reason="fragmentary"/> <milestone n="3-1-Turin-1" ed="SHF" unit="subchapter"/>portees, tant comme il me pleust a y estre, et <milestone n="3-1-fr6475-2" ed="SHF" unit="subchapter"/>la fus informé de la greigneur partie des besongnes qui estoient advenues ou royaume de Castille, de Portugal, de Navarre, d'Arragon <milestone n="3-1-Turin-2" ed="SHF" unit="subchapter"/><gap reason="fragmentary"/> </p>`
`</div>`

The construction of the values for the canonical references based on the SHF edition of Froissart’s *Chronicles* poses problems for chapters which are textually different in different redactions of the *Chronicles*. This applies especially to certain series of chapters in Books I and III (possibly also to Book II). In order to distinguish chapters whose texts belong to a particular redaction of the *Chronicles*, the values for the canonical references should be modified. The values starting with “1Am-”, “1A-”, “1B-” or “1Rom-” indicate that chapters belong to the *Amiens*, *A*, *B* and *Rome* redactions of Book I. For Book III, the values starting with “3A-” or “3B-” indicate that these chapters belong respectively to the “*first*” and “*second*” redactions of Book III.

Apart from the variations between different versions of the *Chronicles*, the manuscripts of the *A* redaction of Book I of the *Chronicles* also contain extraneous material borrowed from the *Grandes Chroniques de France*. Chapters borrowed from the *Grandes Chroniques de France* for the reign of Philippe VI of Valois follow on from the middle of SHF § 108 in the Oxford manuscript, and in mss. fr. 2647, 2651 and 2677. Another block of chapters borrowed from the *Grandes Chroniques de France*, covering the years 1350 to 1356, replaces SHF chapters § 320-371 in most manuscripts of the *A* redaction of Book I. Like for the rest of the transcriptions, these chapters are identified with `<milestone/>` elements, but the references are to editions of the *Grandes Chroniques de France* (the value of ‘ed’ is set to “GCF”). For the chapters covering the reign of Philippe VI of Valois, reference is made to the SHF edition by Viard.⁷ The value of ‘n’ is “PhVI-” followed by the chapter number in the SHF Viard edition. For the chapters covering the period 1350-1356, reference is made to the edition by Paris.⁸

⁷Jules Viard, (ed.), *Les Grandes chroniques de France*, 10 vols. (Paris, 1920–1953).

⁸Paulin Paris, (ed.), *Les Grandes chroniques de France, selon que elles sont conservées en l’église de Saint-Denis en France*, 6 vols. (Paris, 1836–1838).

Ex. <milestone n="PhVI-45" ed="GCF" unit="chapter"/>

10.2. References to page numbers of printed editions

Apart from the <milestone/> elements, which are included in all the transcriptions which contain the relevant chapters of the text, some transcription files also contain page references to printed editions. These are limited to the transcriptions of those manuscripts which have been used as base manuscripts for the respective editions. In cases where this is a multi-volume edition, an additional <milestone/> element is used to record the volume number.

Book I

Paris, BnF, mss. fr. 6477–6479: references to vol. 1–8 of the SHF edition of Book I.

Ex. <milestone n="1" ed="SHF" unit="volume"/><pb ed="SHF" n="1"/>

Book II

Paris, BnF, fr. 5006: references to vol. 9–11 of the SHF edition of Book II (§ 1–82, 272–289, 392–406, 500–502).

Leiden, University Library, VGGF 9-2: references to vols. 9–11 of the SHF edition of Book II.

Ex. <milestone n="12" ed="SHF" unit="volume"/><pb ed="SHF-12" n="1"/>

Book III

Berlin, Staatsbibliothek, Ms. Rehdiger 3 (Depot Breslau, 1, Bd. 3): references to vols. 11–13 of the edition by Kervyn de Lettenhove, and for the final chapter to vol. 15 of the SHF edition of Book III.

Ex. <milestone n="11" ed="KAR" unit="volume"/><pb ed="KAR-11" n="1"/>

Paris, BnF, ms. fr. 2650: references to vols. 12–15 of the SHF edition of Book III.

Ex. <milestone n="14" ed="SHF" unit="volume"/><pb ed="SHF-14" n="1"/>

The order in which these empty elements are inserted is in principle not very important since they can be thought of as the representation in the transcription of independent dimensions to the text. Nevertheless, whenever more than one of these elements appear together in the text, we have tried to insert elements of a higher unit first (quire before page, page before column, column before line) and elements relating to the manuscript itself before elements relating to secondary instantiations after (folio breaks before edition chapters).

11. Gaps, errors, additions, corrections

11.1. Gaps

The gaps recorded in the transcriptions fall into two major kinds: gaps in the transcription because text present in the manuscript has not been transcribed, and gaps in the text because the manuscript transcribed is incomplete. Two elements (<gap> and <space>) are used to encode these omissions. When a non-standardised description of the omission is needed, transcribers can use the <desc> attribute, which is allowed as a child of <gap> or <space>.

Ex.

The first kind of gap is always encoded with the <gap> element. The required ‘reason’ attribute is used to indicate why text has been omitted from the transcription. The following values are allowed: ‘damaged’, ‘deleted’, ‘fragmentary’, ‘illegible’, ‘inaccessible’, ‘sampling’, ‘trimmed’, and ‘wanting’. It is possible to indicate the approximate size of the missing text by an appropriate combination of values for the ‘extent’ and ‘unit’ attributes. The possible values for ‘unit’ are: ‘letters’, ‘words’, ‘lines’, ‘columns’, ‘pages’, ‘folios’, and ‘quires’. Whenever the ‘unit’ attribute has a value, the ‘extent’ attribute should contain a numerical value indicating the approximate number of units omitted from the transcription.

The value ‘sampling’ for the ‘reason’ attribute on <gap> indicates that the transcription is not complete because only a sample of the manuscript has been transcribed.

Ex. Quant le duc d’Anjou vey <gap reason="sampling">

When it has been impossible to transcribe a portion of text because of damage to the manuscript one of the following values is used: ‘damaged’, ‘deleted’, ‘fragmentary’, ‘illegible’, or ‘trimmed’. A value of ‘fragmentary’ indicates that only a small proportion of a manuscript has survived, typically a few folios or less; ‘trimmed’ refers to damage where the edges of the page have been trimmed resulting in textual loss; ‘damage’ indicates other non-specified damage, typically one or several missing pages in an otherwise complete manuscript.

Ex. Comment le roy ^d<gap reason="fragmentary">
... des pays voisins ^l<gap reason="damaged" extent="4" unit="words"/><desc>strip
of parchment cut out</desc></gap>n ...

The values ‘deleted’ and ‘illegible’ for the ‘reason’ attribute refer to text which is present in the manuscript but has not been transcribed because the text has been deleted (made illegible by crossing out) or is otherwise illegible. Whenever text has been omitted from the transcription because the manuscript witness is inaccessible (for example, because its current location is unknown and the transcription is based on reproductions in sales catalogues).

Ex. pour la cause de ce que ^{299 r}^A<gap reason="inaccessible">

In situations where text is missing from the manuscript itself two elements are used to encode two different situations. When text has been omitted unwittingly by the scribe of the manuscript (or the scribe of its exemplar) the element <gap> is used with a value of ‘wanting’ for its reason attribute.

Ex. d'icelles maisons <gap reason="wanting"> <milestone n="2-499-10" unit="subchapter" ed="SHF"><gap reason="wanting"> <milestone n="2-499-1e2-2" unit="subchapter" ed="SHF">sens <lb>rentes

When text is missing from the manuscript because a letter or a word, or a number of letters or words have been deliberately skipped with the intention of adding them later, the element <space> is used. A typical situation is where initials or other decorative elements that have not been executed.

Ex. <space unit="characters" extent="1">e <num type="ordinal" rend="words" value="4">quart</num> chapitre

11.2. Scribal errors

The transcriptions made for the **Online Froissart Project** are not fully edited texts, but aim to represent closely the text as transmitted in individual manuscripts. In order to make these transcriptions as readable as possible, the transcriber provides editorial corrections whenever the text is incomprehensible, or logically or linguistically faulty. In those cases the transcriber provides a reading which renders the text intelligible. The corrected reading will often be based on readings from other witnesses, but its aim is not to restore an original authorial reading: it simply helps the reader with the interpretation of the particular text version present in the manuscript.

In these cases the manuscript reading is marked up with the <sic> element and the proposed correct reading with the <corr> element. The value of the resp attribute on the latter points to the transcriber responsible for the corrected reading. The pair of <sic> and <corr> elements is nested in a <choice> element, which should not have any other content. The contents of the <sic> and <corr> elements should always be complete words; the use of this mechanism to correct single letters or parts of a word is not permitted.

Ex. messires Ainmos, et estoit <choice><sic>comptes</sic><corr type="editorial" resp="#GC">comtes</corr></choice> de Kent

Whenever a word or combination of words enclosed in a <sic> element contains a milestone-like element (<lb/>, <cb/>, <pb/>, <milestone/>) this element should also be repeated in the corresponding <corr> element, so as to insure that in a rendering of the text based on the corrected readings, the distribution of the text on the page still corresponds to the manuscript.

Ex. au <choice><sic>instru<lb>ment</sic><corr type="editorial" resp="#GC">juste<lb>ment</corr></choice> considerer

When the <sic>-<corr> mechanism is used to indicate superfluous words in the manuscript reading, the encoding should be used in such a way that the reading in the <corr> element contains at least one word.

Ex. l'ordonnance du <choice><sic>conseil <lb>du conseil</sic><corr type="editorial" resp="#GC"><lb>conseil</corr></choice>

The <corr> element can be used to clarify particular unusual spellings of words which, although maybe not strictly uncorrect, may be confusing for readers. The use of the <corr> encoding in these cases does therefore not necessarily reject the manuscript reading, but offers help to the reader.

Ex. j'ay ce livre <lb><choice><sic>hystroiuet</sic><corr type="editorial" resp="#GC">hystroiuet</corr></choice>
... tous li <choice><sic>peules</sic><corr resp="#GC" type="editorial">peuples</corr></choice> estoit ...

Finally, the <corr> element can be used to provide missing text whenever a manuscript witness lacks a phrase or passage that has been skipped by the scribe and that has been encoded with the <gap> or <space> element.

Ex. et l'appelloit on <choice><sic><space extent="1" unit="words"/></sic><corr type="editorial" resp="#GC">Jehan d'Iorque</corr></choice>

11.3. Additions

Text added to the manuscript witness is marked up with the <add> element. Scribal additions which merely add letters or words which the scribe (or a later corrector) added to complete the text are transcribed in the place where it was meant to be inserted, even if the addition is found outside the normal stream of text (for example in the margin, with a reference sign). The required attribute 'place' indicates where the addition is found relative to the rest of the text. The optional attribute 'hand' allows the transcriber to indicate which scribal hand wrote the addition if it is not the same as the normal scribe. The possible values for the 'place' attribute are: 'inline', 'supralinear', 'infralinear', 'intercolumnar', 'margin-bot', 'margin-left', 'margin-right', 'margin-top'.

Ex. messire Ge<add place="supralinear" hand="#ScribeC">o</add>rge <lb>Belle

The Froissart manuscripts often contain added text which was not meant to be a simple addition of text to complete the text, and therefore cannot be simply added to the text stream. At least three different types can be found: added rubrics, readers' notes, instructions to illuminators.

When a rubric has been added at the start of a manuscript chapter, it should be marked up in the normal way, with an <add> element nested inside a <head> element.

Ex. <head type="rubric"><add place="margin-left">Cy fait mention des parens du roy d'Engleterre nommé Edouart.</add> </head>

This method, however, is not suitable for marking up rubrics that have been added in the middle of manuscript chapters, since the <head> element is not allowed inside a <p> element, and is also not allowed in the middle of a chapter. Added rubrics that appear in the middle of a chapter cannot therefore be marked as in the above example. Instead they are marked up with simply with an <add> element on which an 'type' attribute with the value 'rubric' indicates that the added text

is meant to be a rubric and therefor splits the current manuscript chapter into two.

Ex. ... ne le seuist.<add type="rubric" place="margin-left">Coment mesure Heu le
Despensier pourchassa que la roine Ysabel fist deboutee de France.</add> Si fist ...

Unlike the additions discussed so far, readers' notes and instructions to illuminators were not intended to become part of the text. They are therefore transcribed and marked up in such a way that signals clearly that they are outside the normal textual sequence. The element used to do so is <fw> (forme works, indicating paratextual elements outside the text block).

The content of the <fw> element is transcribed at an appropriate place, and certainly within the page or column boundaries which are relevant for its relative location. It has two required attributes: 'type' and 'place'. The two values for the 'type' attribute that are relevant here are: 'readers-note' and 'illumInstr'. If a reader's note is written in an identified hand, then the text will need to be marked up with a nested <add> element with an appropriate reference on the 'hand' attribute.

Ex. ... qui tenoit terre <lb><fw type="readers-note" place="margin-left"><add
place="margin-left" hand="#CurA">l'impression ditVII<hi rend="superscript">C</hi>
livres</add></fw>grant en <name type="place">Angleterre</name> ...

11.4. Scribal corrections

Scribal corrections in the form of added letters or words are marked up with simple <add> elements (see above). When the correction involves only the deletion of letters, this is marked up with a element. A 'rend' attribute on the element informs how the deletion is signaled in the manuscript. Valid values are: 'erased', 'expunged', 'instruction' or 'struck-through'; or one of the combinations 'struck-through-expunged', 'struck-through-red', 'struck-through-red-expunged'. A value of 'expunged' indicates that dots are placed under the letters that need to be deleted; 'instruction' indicates that the text is not actually deleted, but that a 'deleatur' or equivalent sign was added.

In cases where a corrections involves the deletion of letters or words a more complex mechanism is needed. This takes the form of a combination of a combination of an <add> and a element nested inside a <subst> element which marks up the whole substitution of one series of letters by another. Normally the element appears first, followed by the <add> element. The same attributes on these elements are required ('rend' on ; 'place' on <add>). Since such corrections often involve the new text being written over the old text, two additional values for the 'rend' attribute are relevant: 'overwritten' and 'erased-overwritten'. With the latter two values for the 'rend' attribute on the deletion, the corresponding value for the 'place' attribute on the addition is 'inline'.

Ex. l<subst><del rend="erased-overwritten">i<add place=
"inline">e</add></subst> prince

12. Names and referring strings

12.1. General

The *Chronicles* contain a large number of proper names. Most of these refer to persons and places. Further references belong to categories which in some way can be related to persons or places, such as references to institutions, peoples, families, collectives. Finally, a relatively small number of references belong to categories which may, or may not, be defined in relation to persons or places, such as events like ‘la bataille de Poitiers’. The retrieval of all these occurrences of names through simple text searching tools would not be straightforward because the scribes of the manuscripts adopt a range of variant spellings for the same name. Furthermore, references to persons, places and institutions are not always by proper names, but often also by simple nouns or noun phrases (‘le roy’, ‘ceste ville’, ‘l’eglise’). The latter would not be accessible through a simple text-search system.

In a printed edition the problems of locating and accessing relevant text passages are solved by using an index of persons and places. An index provides a list of canonical versions of the names of persons and places referred to in the text, possibly with some summary information (biographical dates for persons, geographical location for places), accompanied by a complete list of references to all the passages in the text where reference is made to the index entries, regardless of the variant spellings, or whether or not the reference in the text consists of a complete version of the proper name. An index also contains cross-references under alternative versions of the names, and between entries which have recognisable and formal interrelationships (cross-references between rulers and the territories over which they rule; cross-references between different persons whose family relationships are explicitly referred to in the text). Finally, a well-organised index may subdivide long and complex entries into a series of subentries, each of which refer to a particular aspect of the main entry (a city, its inhabitants, a major event which took place in the city, etc.).

In order to allow a retrieval tool to locate efficiently all the references to persons and places in the transcriptions, proper names and noun phrases referring to such index terms are explicitly marked up. Two separate elements are used to mark up references to index items in the transcriptions: the specialised <name> element, which is used to mark up references consisting exclusively of proper names, and the more generic <rs> (“referencing string”) element, which is used to mark up all the other references which would be included in an index of persons and places .

In theory these two elements <name> and <rs>, in combination, would allow for detailed markup of all the onomastic material and the elements which make up names and references to persons or places (*TEI P5: Guidelines for Electronic Text Encoding and Interchange*, 3.5.1 and 13.2). However, such detailed markup of the relevant text segments, especially when <name> and <rs> are nested and combined with further specialised elements provided by TEI P5, would lead to an exponential increase in the complexity of the markup and the time necessary to implement it across all transcribed manuscripts. Nesting of these elements is also likely to lead to conflicts with other classes of elements (like <hi>, <choice>, <corr>). Since the purpose of this aspect of the markup is effectively to build search functionality equivalent to, and not necessarily more complex than that of an index in a printed edition, the <name> and <rs> elements are not nested. Instead, for those occurrences where a reference to a person or place contains further references to other

persons or places, cross-references should be added to the database of canonical forms.

The <rs> and <name> elements have two attributes. The required 'type' attribute allows the references to be assigned to a discrete category. The 'key' attribute allows the unambiguous identification of each reference by means of a pointer which links each marked up reference to an entry in a list of canonical forms of the index terms (persons, places, etc.).

Since the use of the <name> and <rs> elements is meant to facilitate the automated retrieval of references to index terms and is not, for example, to assist with the study of all onomastic material present in the transcriptions, certain proper names appearing in the text are not marked up. References to saints in medieval dates based on saints' feasts are not marked up as names (but references to saints in general are marked up, as are the place names or names of ecclesiastical institutions which include saints' names). References to names which are mentioned in the text simply as names, and not as references to identifiable persons, are not marked up either. References to personal names (rather than persons), when they do refer to identifiable persons, however, are marked up.

Ex. 'je vins en <rs type="place" key="KEY1">son chastel a Ortais ou paÿs de Berne</rs>, le jour sainte Katherine'
'ne ne les nommoit ne Jehan, ne Gautier, ne Guillaume'
'bel enfant estoit et avoit a nom <name type="person" key="KEY3">Jehan</name>, et la <rs type="person" key="KEY4">dame fille du roy</rs> <name type="person" key="KEY4">Betrix</name>'

12.2. The elements <name> and <rs>

The <name> element is used for marking up references which consists solely of onomastic material. For personal names this means references made up of one or more occurrences of the following: first name, surname or family name, and nickname. Other references are only marked up with the <name> tag if the whole reference is a name.

Ex. <name type="place" key="KEY5">Poitiers</name>
<name type="person" key="KEY6">Edouart</name>
<name type="person" key="KEY7">Guy de Chastillon</name>
'la <name type="event" key="KEY8">Bataille des Trente</name>'

The <rs> element is used to mark up all other references to index terms which cannot be marked up as <name>. This includes references containing proper names but also additional elements ("son frere le duc de Lancastre", "un heraut nommé Derby", "la bonne ville de Gaind"), as well as references consisting only of nouns or noun phrases without a proper name, but whose identification is clear from the context ("le roy"). Pronominal references to index items are not marked up, as they necessarily refer to other instances of references to persons, places etc. mentioned in the close vicinity of the occurrences, each of which in turn is marked up and hence retrievable. When references to an index items contain proper name but also include other informative text, like titles, forms of address or blood relationships, they are marked up with <rs> instead of <name>.

Ex. ‘le <rs type= “person” key=“KEY9”>conte</rs>
 ‘le <rs type= “person” key=“KEY9”>conte de Foix</rs>
 <rs type= “person” key=“KEY10”>Charles, roi de France</rs>
 ‘la <rs type=“person” key=“KEY11”>fille du roi de France</rs>
 <rs type=“person” key=“KEY12”>maistre Jehan Froissart</rs>
 ‘la <rs type=“place” key=“KEY13”>bonne ville de Gaind</rs>
 <rs type=“person” key=“KEY7”>messire Gui de Chastillon, conte de Bloys, seigneur
 d’Avesnes, de Beaumont de Sconnehone et de la Geude</rs>
 ‘l’<rs type= “person” key=“KEY14”>evesque de Saintes en Poitou</rs>
 <rs type="institution" key=“KEY15”>une abbeie de dames au dehors de Burghez, que
 on dist le Sorghes</rs>

The marked-up text should exclude preceding and trailing spaces and punctuation marks, as well as preceding definite articles. The exception to the latter rule are the names of cities, towns, villages and rivers, where the definite article should be considered part of the name and hence included within the <name> tag, unless it is part of an elided word form. Preceding definite articles with names of provinces and countries, however, are not included in the markup (‘la France’, ‘la Picardie’).

Ex. <name type= “place” key=“KEY16”>l’Ecluse</name>
 ‘du <name type=“place” key=“KEY17”>Liege</name>
 ‘pour la passer <name type=“place” key=“KEY18”>l’Escaut</name>’

Noun phrases consisting of an indefinite article followed by a noun, followed by a relative subclause which determines the noun, should be marked up as one unit, including the indefinite article.

Ex. ‘et arriverent a <rs type=“place” key=“KEY19”>une autre ville que on appelle ou païs
 Ville Vesieuse</rs>’

When a single phrase, which cannot be conveniently split up, combines references to two or more index terms, the whole phrase is marked up as one <rs> element. To indicate that two or more items are being referred to, the ‘key’ attribute contains a list of values separated by white space. If the phrase contains references to index items belonging to different types, it is nevertheless split into two phrases, each of which has its own value for the ‘type’ attribute and series of values for the ‘key’ attribute.

Ex. ‘les <rs type="person" key=“KEY20 KEY21”>contes de Northombrelande et de
 Northingen</rs>
 ‘<rs type="person" key=“KEY6 KEY22”>deux roys</rs> et <rs type="place"
 key=“KEY23 KEY24”>royaulmes d’Angleterre et d’Escoce</rs>’

12.3. The ‘type’ attribute

All occurrences of the <name> and <rs> elements require a value for the ‘type’ attribute, which provides for a convenient typology of the index terms. Seven different categories of index terms

are marked up, each of which corresponds to a different value for the required ‘type’ attribute. The possible values are: ‘collective’, ‘event’, ‘family’, ‘institution’, ‘people’, ‘person’, ‘place’, ‘ship’. Most common are references to individually identifiable places and persons.

All geographical names, including names for rivers, harbours, seas and landmark buildings, are marked up with the value ‘place’ for the ‘type’ attribute.

Ex. <name type= “place” key=“KEY25”>Poitiers</name>
la <name type=“place” key=“KEY26”>Porte des Malades</name>
'en <rs type=“place” key=“KEY1”>son chastel a Ortais ou paÿs de Berne</rs>’

References to individually identifiable persons marked up with the value ‘person’ for the ‘type’ attribute. Titles, which often include geographical names, are marked up as part of the reference to a person, using <rs> rather than <name> for the whole string.

Ex. <name type= “person” key=“KEY7”>Guy de Chastillon</name>
<rs type=“person” key=“KEY7”>messire Gui de Chastillon, conte de Bloys, seigneur d’Avesnes, de Beaumont de Sconnehone et de la Geude</rs>

References to titles on their own, when they are not referring to specific identifiable persons, should not be marked up using <rs> with the value ‘institution’ for the attribute ‘type’.

Ex. ‘nous lui devons foy et loyaulté porter comme <rs type=“institution” key=“KEY27”>duc de Aquitaine</rs> et <rs type=“institution” key=“KEY28”>per de France</rs> et <rs type=“institution” key=“KEY29”>conte de Pontieu</rs>’

The type ‘institution’ should also be used to mark up references to institutions of royal government and references to churches and other ecclesiastical institutions.

Ex. ‘pour parlementer au <rs type=“institution” key=“KEY37”>conseil du roy</rs>’
‘au <name type=“institution” key=“KEY38”>Saint Sepulcre</name>’
‘tenoit le <rs type=“institution” key=“KEY39”>Saint Siege de Romme</rs>’

When a surname or family name is used not as a reference to an individual member of that family but to the family as a whole the ‘type’ attribute is given the value ‘family’.

Ex. ‘que tous devoient crier : <q><name type="family" key=“KEY30”>Douglas</name>
! <name type="place" key="KEY31 ">Saint Gille</name> !</q>.’

References to the inhabitants of a country, region or town, are marked with ‘people’ as the value for the ‘type’ attribute. This value should be used even when these are references not to the entire population but to a more or less restricted group.

Ex. ‘contre les <name type="people" key="KEY32">Anglois</name>’
<rs type="people " key="KEY33">ceulx de Gand</rs>

References to groups of people who are not identified geographically but through some other

form of group identity are marked up as ‘collective’. The type ‘collective’ should only be used for groups which have such a group identity, and not for *ad hoc* groups. The latter are not marked up as such, but any names they contain are marked up with the appropriate tags.

Ex. ‘les <name type="collective" key="KEY34">Blans Chapperons</name>
‘que le <rs type="person" key="KEY35">pappe Urbain</rs> envoya pour destruire les
<name type="collective" key="KEY36">Clementins</name>
‘les barons d’Angleterre qui furent presens au siege’
‘le <rs type="person" key="KEY6">roy</rs> et les <rs type="collective"
key="KEY40">barons d’Angleterre</rs>’

The ‘event’ type is used for marking up noun phrases which are references to recognisable events. The labels thus marked up are limited to major military events which can be identified both in time (as taking place on a particular date) and place (as taking place in a particular location). When such index terms contain geographical or personal names, they are marked up as <rs>. When they consist of a noun phrase without a proper name, the whole phrase is considered to be the name for the event, and hence marked up as <name> (with the nouns capitalised in the transcription). Other references to events connected to placenames which lack the precision which would allow them to be identified as single events are not marked up as events, but the place names are marked up. References to sieges, although some of these are to modern scholars well known episodes of the Hundred Years War, such as the siege of Tournai, are not marked up as events, because they did not take place on a single day. Also, in Froissart’s usage, references to such siege seem to be more geographical references than chronological ones.

Ex. ‘la <name type="event" key="KEY8">Bataille des Trente</name>
‘la <rs type="event" key="KEY41">bataille de Poitiers</rs>
‘les incidensses de <name type="place" key="KEY42">Bretaigne</name>’

12.4. The ‘key’ attribute

The ‘key’ attribute provides an external means of locating a full definition for the index term being named. Index terms for which no value for the ‘key’ attribute can be assigned (because they cannot be identified in any way) should not be marked up with the <name> or <rs> elements.

As TEI P5 places no further limitations on this attribute, it may contain white space and as such provides a convenient method of linking two or more index terms to one marked up phrase.

Ex. <rs type="person" key="KEY43 KEY44">monseigneur de Beaugeu, le pere et le
filz</rs>

13. Linguistic features

13.1. General

**

13.2. Direct speech

**

13.3. Quoted text

**

13.4. Proverbs

**

13.5. Technical terms

**

13.6. Text in other languages

lang=la lang=fr lang=en (also in <s> or <seg>)

13.7. Heraldry

**

13.8. Dates

date (when needs checking)

14. The TEI header

Appendix 1: Elements allowed in the transcription files

Appendix 2: Values for rend@

not executed

partially executed

style of initial: champie, lombard, illuminated initial, decorated initial

size (in number of lines): 1 line, 2 lines, 3 lines etc.

rubricated

colour-stroke

Appendix 3: Values for type

(@name, @rs, @quote)

Appendix 4: A sample TEI header