

The Digital Representation of Contracted Script

The Anglo-Norman Hub is the home of the online version of the Anglo-Norman Dictionary, which is supported by and cross-referenced to a text corpus, which, as is not the case for all digitisation projects, will include a full critical apparatus for every document.

This corpus is currently being developed to include many bilingual documents of ecclesiastical or governmental origin, mostly from the 13th. 14th. and 15th. centuries. These include some of the publications of the Record Commission set up in 1800 on the authority of King George III, which scrupulously attempt to reproduce the typography of the official Manuscripts preserved in the Tower of London and elsewhere.

Facsimile of part of the Statute Roll

*Entry of the Inspeximus of Magna Carta:
upon the Statute Roll 25 Edw. I.*

Willelmus dei gra Rex Anglie hunc hunc et hunc hunc, omnes ad quos prefatus latus pervenit ad hunc Inspeximus in eodem cartam de h. quendam Regis
hunc pro omni de libertatibus hunc in hac carta. Ipseque dei gra Rex hunc, hunc hunc, hunc assensum, hunc et Comes Londonie, Monachos. Episcopus
Willelmus. Priores, Comites. Baronibus, Decanibus, Praepositis. Armigeris, et omnes milites et fideles sue piam etiam inspeximus soliti. Scitis quod nos
mandatu dei et p salute anime nre et animarum antecessorum et successorum nrorum ad custodiam pte ecclesie et modicam regni nri spontanea et bona voluntate
nra deditur et concessimus hunc. Episcopus Willelmus. Priores, Comites. Baronibus et omnes de regno nro hunc libertates subscriptas, tenendas in regno nro
libet impetunt. In primis concessimus deo et hac pferet etiam nra confirmacione p nob et heredibus nris impetunt quod carta de libertatibus libere sit et habeat
omnia iura sua integra et libertates suas illas. Concessimus etiam et deditur omnes libere hominibus regni nri et nob et heredibus nris impetunt hunc libertates
subscriptas tenendas et tenendas eis et heredibus nris et nob et heredibus nris impetunt. Si quis contra voluerit nroque hunc alioque tenentem de nob in capite /

The same passage in the 1800 printed edition

Ex magno Rot. Stat. in Turr. Lond. m. 40, 39, 38.

(¹) DWARDUS dei gr̃a Rex Angl̃ Dñs Hib̃n &
Dux Aquit̃: Omnibꝫ ad quos p̃sentes litt̃e p̃ve-
nint saltm. Inspeximus magnam cartam dñi H. quon-
dam Regis Angl̃ p̃ris ñri de libertatibꝫ Angl̃ in hec
verba. Henricus dei gr̃a Rex Angl̃, Dñs Hib̃n, Dux
Normañ, Aquit̃, & Comes Andegãṽ, Archiep̃is, Ep̃is,
Abbatibꝫ, Prioribꝫ, Comitibꝫ, Baronibꝫ, Vicecomitibꝫ,
Prep̃õitis, Ministris, & om̃ibꝫ Ballis & fidelibꝫ suis
p̃sentem cartam inspectũr saltm. Sciatis qđ nos in-
tuitu dei, & p̃ salute anime ñre, & animaz̃ antecessoz̃
& successoz̃ ñroz̃, ad exaltaçõem sc̃e eccl̃ie, & emenda-
çõem regni ñri, spontanea & bona volũtate ñra, dedimꝰ
& concessimꝰ Archiep̃is, Ep̃is, Abbatibꝫ, Prioribꝫ, Co-
mitibꝫ, Baronibꝫ, & om̃ibꝫ de regno ñro, has lib̃tates
subscriptas, tenendas in regno ñro Angl̃ imp̃petuũ.

In the introduction to "The Statutes of the Realm" it is stated "in printing the Text of the antient Statutes, the same Mode has been adopted as with respect to the Charters, of exhibiting the Contractions as near as may be to the likeness of the Original Characters."

The remit of the Anglo-Norman Hub requires that all the texts in the corpus be encoded in Unicode-compliant XML.

XML stands for Extensible Markup Language, which is increasingly replacing HTML as a standard language for marking up pages for the world-wide web, because it can record the significance, not merely the format, of the content, allowing far more sophisticated searches to be undertaken, and potentially storing more information than may be visible through a particular portal.

Unicode is fast becoming the most widely-accepted international standard for the representation of both alphanumeric and ideographic characters on computers and is sufficient for the representation of a large number of

modern languages. Many fonts available in Word-Processing software now include the whole of the Unicode character set.

Part of the Appendix to the Introduction to the Statutes

A P P E N D I X. D.

EXPLANATION OF THE CONTRACTIONS,

Used in printing the Records and Manuscripts copied in this Collection.

THE general Character and Appearance of the Charters and Records copied in this Collection, are exhibited in engraved Specimens.* With respect to the PRINTED CONTRACTIONS, it is to be observed, that wherever the Manuscript is abbreviated, the Print has a Mark of Contraction, as similar to that of the Manuscript as the Types will admit. The different Manuscripts vary a little from each other in the Mode of making the several Contractions. The following Explanation of the Marks of Contraction used in the Print (applying both to the Latin and French Text, though the latter is generally less abbreviated than the former), may serve as well to render the printed Copy intelligible, as also to explain the Contractions in the Manuscripts, and to make the reading and consulting of them, when necessary, more easy to Persons not used to antient Records.

A strait Line over a Vowel denotes the Omission of the Letter m or n following :

quā - - quam		nō - - non		statī - - - statim
antiq ^a - - antiquam		cōmun - - commun		volūtātē - voluntatem
hōiū - - hominum		avaūdiz - - avaunt ditz		

The strait Line over m in the Middle of a Word denotes the Omission of the Letter n following :

oīnes - - - omnes		oīnia - - - omnia
-------------------	--	-------------------

A Crooked Line over some Letter, or a Line through some Letter, of the Word contracted, denotes the Omission of one or more Letters of the Word :

G ^r ā - - Gratia		Epīs - - Episcopis		co, tō - { cō or tio
Ballīs - - Ballivis		oīns - - omnes		cion or tion
noī - - nobis		libtates - libertates		corona ^{cōis} - coronacionis
heat - - habeat		Dñs - - Dominus		expedita ^{cō} - expeditacio
Saī - - Salutem		dicī - - dictum, &c.		malicōse - maliciose

A small superior Letter denotes an Omission, of which such Letter forms a Part :

q ^m } quam		occ ^o ne - occasione		p ^m is - primis		s ⁱ - - sibi
q ^t		imp ^s onet ^r - imprisonetur		p ⁱ ^o - - prius		n ⁱ - - nisi

In the Appendix to the Introduction to the Statutes, the use of contractions is explained: the vowels with nasal bars seen at the top are all within the Unicode character set. The "m" with a horizontal bar is not a Unicode character but can be represented by an m followed by a diacritical mark which the rendering software will automatically re-position above the letter to which it applies. The same applies to the various tildes, except

that these cannot be made to apply to the "co" letter combination as shown. Superscript characters do not have separate Unicode representations but they can be represented through formatting instructions embedded in the XML. The superscript character representing the "us" of "prius" has been represented with a "9", but this is not considered satisfactory as it could upset some search software.

While Unicode includes many of the non-standard characters used in medieval manuscripts and early printed books, it unfortunately excludes many widely-used abbreviations which are familiar to medieval Latinists.

SourceForge.net have addressed this by producing a font called Junicode which includes many characters used by medievalists, giving them unallocated codes within the "Private Use Area" of Unicode. While this will allow scholars to produce printed materials containing these characters, it is of no use when it comes to displaying text on the internet.

The Medieval Unicode Font Initiative is attempting to rectify this limitation and has produced the MUFI recommendation for the use of the Unicode "Private Use Area" for the representation of additional characters used in print editions of medieval texts in a variety of languages. Unfortunately the codes used in this recommendation are not the same as those in Junicode. In January of this year the MUFI made a formal proposal to the Unicode authority for the addition of new characters to the next version of the Unicode standard. However, neither the recommendation nor the proposal include all the characters we have found in the publications of the Record Commission.

The MUFI proposal is accompanied by various printed examples of the characters they propose to add to the Unicode standard. These include the character resembling a "9" which generally stands for "con", a superscript character also resembling a "9" as well as the frequently-occurring contractions for "per" and "pro"

Continuation of the explanation published with the Statutes, exemplifying the two forms resembling 9s.

<i>The following Characters or Abbreviations have certain explicit Significations, viz.</i>		
Character. 3	Signification. us	<i>At the End of Words in the Dative or Ablative Plural :</i>
		Abbatib ₃ - - Abbatibus quib ₃ - - - quibus
	sometimes et	s ₃ - - - set (sed) poss ₃ - - - posset
		<i>Or as a comprehensive Mark of Abbreviation :</i>
		quib ₃ - - quilibet videl ₃ - - - videlicet
		qn ₃ - - quandoque
9 above the Line.	us	<i>Not being the Dative or Ablative Plural :</i>
		huj ⁹ modi - } hujusmodi dedim ⁹ - - - dedimus
		h ⁹ i - - - } exc ⁹ are - - - excusare
		huj ⁹ - - - hujus
	sometimes os, ost.	p ⁹ - - - } post p ⁹ tea - - - postea
		p ⁹ t - - - }
9 even with the Line.	com, or con	gtenta - contenta gmuni - communi gseil - conseil

The following page shows the abbreviation for "rum" which is not in the MUFI recommendation or proposal. It also show various characters with flourishes representing "er" or "re". While some flourishes are included in the MUFI proposal, they do not look anything like this and it is unclear whether they are to be considered as the same character

"rum"-abbreviation and flourishes

Character.	Signification.		
q̃ or q̄	que	absq̄ - - - - absque	usq̄ - - - - usque
		<i>Frequently used also for quod and quia.</i>	
}	er	ex̄citum - exercitum	infreḡiat - infregerint
		itin̄e - - itinere	den̄s - - deners
	sometimes re	eḡssum - - - egressum	transḡssum - transgressum
ꝛ	rum	antecessoꝛ - - - antecessorum	nꝛoꝛ - - - nostrorum
		<i>Sometimes to mark an Abbreviation beginning with r:</i>	
		coꝛ - - - - coronacionem	
&	et	<i>The & appears in the Print wherever an Abbreviation for et appears in the Manuscript.</i>	
⁊	and	<i>In the Copies of some Scotch Records in Appendix H. to the Introduction.</i>	
&	eciam		
+	est		
Ꝛ	cer	fec̄it - - - - fecerit	c̄tum - - - - certum
m̄	mer	am̄cietur - - - amercietur	m̄icator - - - mercator
ꝛ	ser	sv̄icium - - - servicium	sv̄ir - - - - servir
ꝛ	ter	ceṯa - - - - cetera	t̄ram - - - - terram
	sometimes tre	t̄spasours - - - trespassours	
ū	ucr	fūit - - - - fuerit	hūit - - - - habuerit
ū	ver	estov̄iū - - - estoverium	ov̄tes - - - - overtes

We now turn to the three main "p"-contractions, which are perhaps the most important of the proposed additions to Unicode. the processing of which we will examine shortly.

P-contractions

p	} <i>As the Prepositions per and pro, and in forming Words :</i>	cepit - - - - ceperit		pp ^a - - - - propria
p		pbū - - - - probum		psone - - - - persone
		aptemēt - - - - apertement		pfit - - - - profit
		pmisit - - - - permisit		pmisit - - - - promisit
		pchein - - - - prochein		p ^a - - - - persona
	also for			
p	par and por	patus - - - - paratus		pceles - - - - parceles
		tempe - - - - tempore		corpum - - - - corporum
p̄	pre	p̄sentem - - - - presentem		p̄laz - - - - prelatz
		p̄d̄cm - - - - predictum		p̄ndront - - - - prendront
z	tz	fiz - - - - fitz		fez - - - - fetz
		establisemenz - - - - establissementz		tenaunz - - - - tenauntz
<i>In the Copies of some Records the following Characters occur.</i>				
℄	is or s	Scott℄ - - - - Scottis		<i>As also the Saxon Characters,</i>
		Lord℄ - - - - Lords		p - th ȝ - y

The MUFI proposal states, "Contemporary philologists and linguists who want to study the graphemic conventions in use in medieval times. . . must rely on *bona fide transcriptions* of the texts. . . only possible when the elemental character set used in the manuscripts is encoded uniquely. . . Transliteration-based editions are virtually useless for those scholars who are interested in the study of medieval writing systems. . ."

The proposal continues: "Transcription is an editorial process which does not entail the replacement or the distortion of the original character set. . . The practice of *expanding abbreviations* . . . is not *transcription*, but simply *transliteration*"

The Editorial Body of the Anglo-Norman Hub has decided, that in the absence of a complete character set for representing the contractions, and in order to produce material which can be indexed and searched and linked to the Anglo-Norman dictionary, we should go down the road of transliteration and expand the contractions. This presents more of a problem for the Anglo-Norman, which has no standard spellings, than for Latin, whose spelling is relatively standardized. Rather than apply editorial interpretations which can be disputed, wide use is being made of the apostrophe, wherever more than one expansion is possible.

On the other hand, in order to digitise material with contracted characters, it was necessary, before attempting any transliteration, to create some kind of transcription. Standard codes not being yet available for all the characters, some kind of interim codes would have to be invented, and

these must sufficiently resemble the original characters to allow the entire text to be re-keyed by an agency.

Examples of the "per" "pro" and "pre" contractions as represented by the agency

Category/Signification	To be keyed as	Expanded example	Rekeyed example
per/pro	p!!	ceperit	cep!!it
	\$p\$	propria	\$p\$pⁱa
	\$p\$	probum	\$p\$bū
	p!!	persone	p!!sone
	p!!	apertement	ap!!tement
	\$p\$	profit	\$p\$fit
	p!!	permisit	p!!misit
	\$p\$	promisit	\$p\$misit
	\$p\$	prochein	\$p\$schein
	\$p\$x	proximo	\$p\$x
	p!!	persona	p!!^a
par and por	p!!	paratus	p!!atus
	p!!	parceles	p!!celes
	p!!	tempore	temp!!e
	p!!	corporum	corp!!um
Where unclear which of above	\$p!!\$		
Capitals equivalent to pro above	\$P\$		
Capitals equivalent to per above	P!!		
pre	p?~	presentem	p?~sentem
		prelatz	p?~laz!!
		predictum	p?~d!!c~m
		prendront	p?~ndront

Text sample from the Statutes

℞ Collectoribz custume sue lanaꝝ in Civitate London, saltm. Cum nup p nos & consiliū nrm ordinatum fuisset, qđ lane, coria, pelles lanute, plumbum & stagmen nō dimittent^r seu quomodolibet venderent^r, nisi p bonis sterlingis seu aliis mcandisis legalibz, put in statuto inde edito plenius continet^r; p quod vice^{ba} nris

Same example as it came back from the re-keying process

R!! Collectorib\$z\$ custume sue lana\$za\$ in Civitate London!!, saſtm. Cum nup!! p!!
 nos & consiliū nr~m ordinatum fuisset, qd!! lane, coria, pelles lanute, plumbum &
 stagmen nō dimitt?~ent^r seu quomodolibet venderent^r, nisi
 \$p\$ bonis sterlingis seu aliis m?~candisis legalib\$z\$, \$p\$ut in statuto inde edito
 plenius continet^r; p!! quod vice^{bz} nr~is

The transcriptions produced with these interim codes still preserved the full graphemic information of the original, and this has been preserved in the XML, using another set of codes, cross-referenced through a table to their equivalents in Unicode, Junicode and MUFI codes wherever such codes exist.

Examples of the per pro and pre contractions with the various codes used to represent them

per/pro	p!!	E670	F145	p-
	\$p\$	E67D	F147	p\$
	\$p\$	E67D	F147	p\$
	p!!	E670	F145	p-
	p!!	E670	F145	p-
	\$p\$	E67D	F147	p\$
	p!!	E670	F145	p-
	\$p\$	E67D	F147	p\$
	\$p\$	E67D	F147	p\$
	\$p\$x	E67D;x	F147;x	p\$x
par and por	p!!	E670	F145	p-
	p!!	E670	F145	p-
	p!!	E670	F145	p-
	p!!	E670	F145	p-
	p!!	E670	F145	p-
Where unclear which of above	\$p!!\$			p-\$
Capitals equivalent to pro above	\$P\$	E27D	F146	P\$
Capitals equivalent to per above	P!!	E270	F144	P-
pre	p?~			p? p? p? p?

Part of the XML file corresponding to the previous text example

```
<p><expan abbr="R-">Rex</expan> Collectorib<expan  
abbr="z$">us</expan> custume sue lana<expan abbr="z£">rum</expan> in  
Civitate Londo<expan abbr="n-">nii</expan>, sa<expan abbr="l-  
t">lute</expan>m. Cum nu<expan abbr="p-">per</expan> <expan abbr="p-  
>per</expan> nos & consili<expan abbr="u-">um</expan> n<expan  
abbr="r~">ostru</expan>m ordinatum fuisset, q<expan abbr="d-  
>uod</expan> lane, coria, pelles lanute, plumbum & stagmen n<expan  
abbr="o-">on</expan> dimit<expan abbr="t?">ter</expan>ent<expan  
abbr="rsup">ur</expan> seu quomodolibet venderent<expan  
abbr="rsup">ur</expan>, nisi <expan abbr="p$">pro</expan> bonis sterlingis  
seu aliis <expan abbr="m?">mer</expan>candisis legalib<expan  
abbr="z$">us</expan>, <expan abbr="p$">pro</expan>ut in statuto inde  
edito plenius continet<expan abbr="rsup">ur</expan>; <expan abbr="p-  
>per</expan> quod vice<expan abbr="bzsup">comitibus</expan> n<expan  
abbr="r~">ostr</expan>is
```

This contains both the codes representing the graphemic information and the expansions

Rendered output

Rex Collectorib***us*** custume sue lanar***um*** in Civitate London***ii***,
sal***utem***. Cum nu***per per*** nos & consili***um nostrum*** ordinatum
fuisset, ***quod*** lane, coria, pelles lanute, plumbum & stagmen ***non***
dimit***terentur*** seu quomodolibet venderent***ur***, nisi ***pro*** bonis
sterlingis seu aliis ***mercandisis legalibus***, ***pro***ut in statuto inde
edito plenius continet***ur***;

This is what the user will actually see, fully expanded text, with only italics used to show where a contraction has been expanded.

Although this graphemic information will not be visible to the user when the text appears on the web, because of the versatile nature of XML, it will be possible, in the future, once a complete medieval character set exists, for the user of the electronic text to choose between viewing the original characters or the expansions, and to switch between these two

modes at the click of a mouse button. This can be achieved without any alteration being made to the XML files.

The preparation of those parts of the Statutes of the Realm, which are to be included in the corpus has required the expansion of nearly 73,000 contractions. Within the scope and timescale of the project, such a task could not be performed without some degree of automation. Automation however has not been straightforward, as so many contractions can represent more than one sequence of letters.

As a first step towards expanding the contractions within the digitised text, a digital concordance was created, using Rob Watt's Concordance Software version 3.0, of all the word-forms found in the text, showing the context of each occurrence.

Extract from the Concordance

de tenem~tis q~ ten~t^r in Gavelkind!! in eod!!	Com!!	, de cet?~o non teneant^r magne Assise p!! xij Milites, sic~ ali...
in Gavelkind!! in eod!! Com!! de cet?~o non teneant^r	Com!!	, & infra eund~ cor~ Justic~ nr~is itin?~antib~z\$ in adventu eo\$...
in Gavelkind!! in eod!! Com!! de cet?~o non teneant^r	Com!!	cor~ Justic~ nr~is itin?~antib~z\$ in adventu eo\$za\$. Et mand!! est Vic~...
in Gavelkind!! in eod!! Com!! de cet?~o non teneant^r	Com!!	suo legi & firmit?~ ten?~i faciat sc~d!!m qd!! Carta illa c~tinet. T. ut s<s...</td>
c~.</note><lang="LA">Rex c~cessit \$p\$bis ho~ib~z\$ de	Com!!	Kanc~ qd!! loco magna\$za\$ Ass~ q~ capi c~suever~t p!! xij Milites de...
p~r[~] ~ Gavelkind & int?~ Gavelkindeis; de cet?~o c	Com!!	suo legi fac~ & ten?~i. T. ut s^a [R!! ap~ Westm!! vij die Nov...
inde precipiet^r qd!! nich!! [carius efficiatur in C	Com!!	infra libertatem nec ex^a \$3] teneantur s\$za\$ qd!! Com!! ten...
Justic~ qe nous leaument elirrom, auxibien de nus meymes	com!!	teneat^r solu~modo ad profros faciendos, set in eo nullum pla...
g="LA">Hoc Sacrament~ p?~stabitur ab omnib~z\$ Ballivis	Com!!	des autres [xij Chevalliers e autres \$9] prodes homes de Nostre hundre...
ing="LA">Deinde liberabunt^r om!les Articuli de	Com!!	, alio m~o q^a prius.</lang>
\$2 efficiet~ in villa vel in	Com!!	xij Jurator~ cuj⁹ libet Hundredi, et deinde assignabitur eis te...
	Com!!	, nisi sit coram Justic~, vel

This example from the concordance represents the abbreviated form "Com" with a bar. As you can see, the contexts shows that in one case this represents the French word "come" and the other examples are a variety of Latin forms of "Comitatus"

The next stage was to arrange the contracted words in order of frequency of occurrence. From the concordance, a table was created of all the contracted word-forms which occur more than twice within the text.

Table of expansions, example 1

Contracted word	Occurrences	Action	Expansion	DAT comments
&	6264	Leave		leave
Q~	2989	Wait		wait
P^R	803	Wait		(impossible)
SEIGN^R	325	Expand	seignour	wait
AUT?~S	289	Expand	autres	(impossible)
MAN?~E	250	Expand	manere	wait
S^R	224	Expand	sur	(impossible)
P!!LEMENT	215	Expand	parlement	
T?~RE	199	Expand	terre	
T?~RES	196	Expand	terres	
DENGLET?~RE	191	Expand	dengleterre	
G^ANT	181	Expand	grant	
LO^R	167	Expand	lour	
P!!TIE	152	Expand	partie	
AP?~S	142	Expand	apres	
S?~RONT	139	Expand	serront	
H<BAR>O</BAR>ME	137	Expand	homme	
QD!!	130	Expand	quod	
CO~E	128	Expand	commune	
G^ANTZ	121	Expand	grantz	
S?~RA	110	Expand	serra	
Q^ANT	105	Expand	quant	
NO⁹	104	Expand	nous	
WESTM!!	94	Expand	Westmonster	
DEV?~S	88	Expand	devers	
CO~ES	87	Expand	communes	
\$P\$FIT	82	Expand	profit	
P?~NDRE	80	Expand	prendre	
\$F\$	74	Expand	seignour	

This is the start of the table, arranged in descending order of frequency of occurrence, showing the most common contracted words, most of which have been unambiguously expanded, all French apart from "quod". Of the first three, the ampersand has been left as it is and the following two defy consistent expansion. The "q" with a tilde can be expanded in the Latin, from its context, but in French can represent "qi" or "qe", which are not used with any grammatical consistency in Anglo-Norman as one might expect, and which consequently have been left as "q"+apostrophe, throughout the French. The next example, "p" with a superscript "r" again can represent either "par" or "por" in French, and scholars can debate endlessly which of these is correct in any given context. It has been represented by "p"+apostrophe+"r".

The words in the table were considered one by one in order to establish, wherever possible a standard transliteration of the word. This was done was by examination of the contexts to establish the meaning and, where possible, by finding uncontracted examples of the same words, and selecting, in the case of Anglo-Norman variant spellings, the version which occurred most frequently. Where no uncontracted example could be found, a search was made in the Anglo-Norman dictionary or the Medieval Latin Word List, as appropriate. Further consultation of other Medieval Latin Dictionaries enabled almost all the Latin to be established. Anglo-Norman forms not found in the Anglo-Norman dictionary have been referred to the project leader, Professor David Trotter. The words listed were then replaced "en bloc" on a computer, leaving only a few thousand.

Table of expansions, example 2

MEMORAND!!	8	Expand	memorandum	
MEST?~	8	Expand	mestre	
P!!DRE	8	Expand	perdre	
P?~DC~M	8	Wait		wait probably predictum/am
P?~MER	8	Expand	premer	
P?~SCHEINS	8	Expand	proscheins	
P?~SENTZ	8	Expand	presentz	
PASCH!!	8	Expand	Pasche	
T?~MINEZ	8	Expand	terminez	
T?~RA	8	Expand	terra	
TEN!!	8	Wait		wait
TEN~TZ	8	Expand	tenementz	
V?~GE	8	Expand	verge	
V?~ROIE	8	Expand	verroie	
V?~TUE	8	Expand	vertue	
\$P\$PRES	7	Expand	propres	
\$Q\$	7	Wait		impossible!
1^O	7	Leave		leave

This part of the table shows a mixture of unambiguous Latin and French, but with a couple of examples from each language which must be expanded with reference to context, as with "predictum" or "predictam"

The forms left to be resolved were those which occurred only once or twice, identical contracted forms representing different words in different contexts and those so abbreviated that they could only be elucidated from the context of each occurrence. From amongst these, somewhat less than half involved standard contractions with only one possible expansion, which again could be treated "en bloc", on a rule-based system, leaving the remainder to be examined individually within their contexts and to be expanded with reference once again to corresponding uncontracted forms within the text or in the dictionaries, or, as a last resort, with the aid of the seventeenth and eighteenth-century English translations. The handful which still remain unresolved are either place-names, which, it is clear from "Orbis Latinus" are subject to a high degree of variance, even in Latin, and which have been left abbreviated, and previously unlisted Anglo-Norman forms, which are currently being referred to Professor Trotter.

At the present there are a number of very large library digitisation programmes afoot. As far as we have been able to discover, the principal players in this endeavour are either producing facsimiles of the printed page, which cannot be searched for particular words, or they are producing text by OCR, that is, by an Optical Character Recognition

Process, which while it can be searched for words, cannot be considered as accurate without thorough proofreading, and which, as we have found, is completely incompetent when it comes to non-standard characters, The Google-Bodleian project has decided, because of outstanding copyright and intellectual property issues, to focus on older materials. The European Library Project is committed to "digitising the European Patrimony". All this is planned while still there is no standard way of fully encoding the graphemic content of many early printed materials, let alone manuscripts. If these projects go ahead without resolving these issues and without minute attention to detail, then there will be a serious loss of information which will make it impossible for future generations of scholars, to produce any serious work through the consultation of these web archives, or, where an OCR process has been employed, to reinterpret text which has been hastily expanded, or worse still, expanded simply through the application of a rule-based system by a computer. Present-day scholars need therefore to take an active interest in the current library digitisation projects to ensure that text files and not simply facsimiles are produced, that digital records are created to the highest standards of quality, and that scholarly criteria override commercial imperatives. The Anglo-Norman Hub Digitisation Project is entering its final six months. The project is not interested in the remainder of the prolific output of the Record Commission, because it almost all in Latin. It would be far better were the remainder of this material to be digitised with the same degree of care and attention to detail as has been applied at the Anglo-Norman Hub, rather than being treated on the basis of Google's "ten dollars a book" rule of thumb.

Postscript

This paper was delivered to the 8th. International Conference on Late and Vulgar Latin, at St. Catherine's College Oxford in September 2006. Since then a second MUFI recommendation has appeared. All references in this paper are to the first recommendation. For further information please visit <http://gandalf.aksis.uib.no/mufi/>