



Metadata Guidelines for the Latin Subcorpus *PaLaFraLat*

PaLaFra – The transition from Latin to French: constitution and analysis of a
Latin-French digital corpus

Version 1.0

Sebastian Ortner

April 2018

Contents

1	General principles	4
2	The French subcorpus <i>PaLaFraFro</i>	4
3	The Latin subcorpus <i>PaLaFraLat</i>	4
3.1	Key data of PaLaFraLat – Part I	6
3.2	Key data of PaLaFraLat – Part II	8
4	Metadata-Tags	11
4.1	Identification tag	12
4.2	Title	12
4.3	Discursive domain	12
4.4	Discursive genre	12
4.5	Full name of the author	13
4.5.1	Name of the author	13
4.5.2	Cognomen of the author	13
4.6	Referendary	13
4.7	Recipient	14
4.8	Identifier	14
4.9	Date of composition	14
4.9.1	Start date of composition	15
4.9.2	End date of composition	15
4.9.3	Formal date of composition	15
4.10	Date of the manuscript	15
4.10.1	Start date of the Manuscript	15
4.10.2	Final date of the Manuscript	16
4.10.3	Formal date of the Manuscript	16
4.11	Name of the editor	16
4.12	Bibliography of the edition	16
4.13	Identification of the manuscript	16
4.14	Digitalization source	17
4.15	Funding	17
4.16	Morphosyntactic tagset	17
4.16.1	Tagset	17
4.16.2	Status	18
4.17	Link to the <i>Digital Monumenta Germaniae Historica</i> (dMGH)	18
4.18	Textorder (Identification number which is only visible in Offline-TXM)	18
4.19	Dimensions	19

4.19.1	Words	19
4.19.2	Tokens	19

1 General principles

In these guidelines, the metadata tags used to describe and characterize the Latin texts of the French-German research project *PaLaFra* will be explained in detail. The manual adopts general rules from the French guidelines, established within the context of the *Base de français médiéval* (BFM). The acronym *PaLaFra* stands for the French *Le passage du latin au français*. The project aims for the creation of a digital corpus composed of Old French texts as well as Late Latin texts of the Merovingian epoch.

(a) Access to the digital corpus:

Lyon: <http://txm.bfm-corpus.org/>

Regensburg: <http://rhsslk1.ur.de/txm/>

(b) More information on the project:

<http://www.palafra.org>

(c) *Base de français médiéval* (BFM) guidelines:

http://ccfm.ens-lyon.fr/IMG/pdf/Manuel_Descripteurs_BFM.pdf

2 The French subcorpus *PaLaFraFro*

The French subcorpus *PaLaFraFro*, like the subcorpus *PaLaFraLat*, is hosted by the BFM.

→ For more information about *PaLaFraFro* please consult BFM¹ website.

3 The Latin subcorpus *PaLaFraLat*

In the following, we present the texts co-opted in the Latin subcorpus *PaLaFraLat*. All texts are provided by the *Monumenta Germaniae Historica*² editions and were retrieved from the digitalized versions on dMGH³ and openMGH⁴. These sites offer an OCR performed scan of the original edition as well as both an HTML and XML-TEI version.

The traditional text philology of the 19th century left its mark on many of these editions (Selig 2005, Heinzlmann 1992), because of its principle of establishing an edition out of several manuscripts. This sort of textual criticism was spread by Karl Lachmann (1793–1851) and is also called stemmatology. In terms of choosing/naming *the* manuscript used for building the edition, we always decided to provide the corresponding information of the archetype, the oldest document. The project does not reconsider the total history of every manuscript – so this work still remains reserved to historians.

¹<http://bfm.ens-lyon.fr>

²<http://www.mgh.de>

³<http://www.dmgf.de>

⁴<http://www.mgh.de/dmgf/openmgh>

Nevertheless, in terms of understanding the language and its graphical variants in the Merovingian epoch, the editions are still of great value for diacronic research. Van Acker (2007: 205) points out that the documentation of manuscripts using this methodology was very carefully established (e.g. the critical apparatus) and allows us to still reconstruct the manuscript tradition today. In the TXM Software⁵, a permanent synoptic view permits us to immediately take an even closer look at the critical notes directly linked to the digital edition of the MGH available under <http://www.dmgf.de>, as shown by the following example.

The screenshot displays the TXM software interface for the digital edition of the *Vita Amandi*. The main text on the left is in Latin, and the right side shows the critical apparatus. The apparatus includes a list of manuscripts (A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z) and their respective readings for each line of text. The interface also shows a search bar, a page number (432), and navigation controls.

Figure 1: Synoptic view of the *Vita Amandi* in TXM

For all texts which were integrated into the Latin subcorpus *PaLaFraLat*, the history and quality of the edition was taken into consideration. In the following section, an overview of the chosen texts is presented in detail. All texts date from the Merovingian or very early Carolingian Epoch⁶.

⁵<http://textometrie.ens-lyon.fr>

⁶Date of composition.

3.1 Key data of PaLaFraLat – Part I

ID	Title	Number of texts	Domain	Genre	Editor	Date
1–29	Lives of saints	29	religious	hagiography	Krusch, Levison	6th – 9th century
101–148	Epistolae Austrasicae	48	historical	epistolary	Gundlach	5th – 6th century
201–236	Desiderii ep. Cardurc. ep.	36	historical	epistolary	Arndt	first half of the 7th century
301–319	Epistolae aevi Merovingici coll.	11	historical	epistolary	Gundlach	6th century
401	Gregory of Tours	1	historical	historiography	Krusch	end of the 6th century
402	Fredegar, IVth book	1	historical	historiography	Krusch	end of the 7th century – beginning of the 8th century
403	Fredegar, Continuaciones	1	historical	historiography	Krusch	first half of the 8th century
404	Liber Historiae Francorum	1	historical	historiography	Krusch	beginning of the 8th century
501	Pactus legis Salicae	1	juridical	legislative	Eckhardt	6th century
502	Lex Ribuaria	1	juridical	legislative	Beyerle/Buchner	first half of the 7th century
601	Formulae Andecavenses	1	juridical	formulary	Zeumer	end of the 6th century
602	Formulae Bituricensis	1	juridical	formulary	Zeumer	beginning of the 8th century
603	Formulae Marculfi	1	juridical	formulary	Zeumer	tury – beginning of the 9th century
604	Formulae Marculfi Supplementum	1	juridical	formulary	Zeumer	ca. 650 – 730
					Zeumer	ca. 650 – 730

605	Formulae Pithoei	1	juridical	formulary	Zeumer	before the end of the 7th century
606	Formulae Salica Bignonanae	1	juridical	formulary	Zeumer	769 – 775
607	Formulae Salica Merkelianae	1	juridical	formulary	Zeumer	beginning of the 8th century – 10th century
608	Formulae Salica Lindenbrogianae	1	juridical	formulary	Zeumer	end of the 8th century – middle of the 9th century
609	Formulae Turonenses vulgo Sirmondicae dictae	1	juridical	formulary	Zeumer	middle of the 8th century – 9th century
701–738	Merovingian charters	38	juridical	charter	Kölzer	625 – 768
801–802	Charters of the Arnulfings	2	juridical	charter	Heidrich	751
Total		187				

3.2 Key data of PaLaFraLat – Part II

Title	MGH identifier	Page range	Link
Lives of saints	MGH SS rer. Merov. 3-7	–	http://www.mgh.de/dmgh/resolving/MGH_SS_rer_Merov._3_S._II
Epistolae Austrasicae	MGH Epp. 3	110–153	http://www.mgh.de/dmgh/resolving/MGH_Epp._3_S._110
Desiderii ep. Cardurc. ep.	MGH Epp. 3	191–214	http://www.mgh.de/dmgh/resolving/MGH_Epp._3_S._191
Epistolae aevi Merovingici coll.	MGH Epp. 3	434–468	http://www.mgh.de/dmgh/resolving/MGH_Epp._3_S._434
Gregory of Tours	MGH SS rer. Merov. 1,1	265–321	http://www.mgh.de/dmgh/resolving/MGH_SS_rer_Merov._1,1_S._265
Fredegar, IVth book	MGH SS rer. Merov. 2	118–168	http://www.mgh.de/dmgh/resolving/MGH_SS_rer_Merov._2_S._118
Fredegar, Continuaciones	MGH SS rer. Merov. 2	168–193	http://www.mgh.de/dmgh/resolving/MGH_SS_rer_Merov._2_S._168
Liber Historiae Francorum	MGH SS rer. Merov. 2	238–328	http://www.mgh.de/dmgh/resolving/MGH_SS_rer_Merov._2_S._238
Pactus legis Salicae	MGH LL nat. Germ. 4,1	1–236/273	http://www.mgh.de/dmgh/resolving/MGH_LL_nat_Germ._4,1_S._1
Lex Ribuaria	MGH LL nat. Germ. 3,2	52/73–134	http://www.mgh.de/dmgh/resolving/MGH_LL_nat_Germ._3,2_S._52

Formulae Andecavenses	MGH	Formulae Mero- wingici et Karolini aevi	1–25	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._1
Formulae Bituricensis	MGH	Formulae Mero- wingici et Karolini aevi	166–181	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._166
Formulae Marculfi	MGH	Formulae Mero- wingici et Karolini aevi	32–106	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._32
Formulae Marculfi Supplementum	MGH	Formulae Mero- wingici et Karolini aevi	107–112	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._107
Formulae Pithoei	MGH	Formulae Mero- wingici et Karolini aevi	596–598	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._596
Formulae Salica Bignonanae	MGH	Formulae Mero- wingici et Karolini aevi	227–238	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._227
Formulae Salica Merkelianae	MGH	Formulae Mero- wingici et Karolini aevi	239–264	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._239
Formulae Salica Lindenbrogianae	MGH	Formulae Mero- wingici et Karolini aevi	265–284	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._265
Formulae Turonenses vulgo Sirmondicae dictae	MGH	Formulae Mero- wingici et Karolini aevi	128–165	http://www.mgh.de/dmgh/resolving/MGH_Formulae_Merowingici_et_Karolini_aevi_S._128

Merovingian charters	MGH DD Mer. I	62–64, 76–77, 88–89, 108–110, 185–186, 189–192, 217–220 etc.	http://www.mgh.de/dmgh/resolving/MGH_DD_Mer._1_S._V
Charters of the Amulfings	MGH DD Amulf.	48–50, 50–51	<i>online not available (openMGH/dMGH)</i>

4 Metadata-Tags

Detailed *example* of the metadata published in the current textcorpus version of PaLaFraLat in TXM:

	Tags	Tags (TXM Portal Online)	Values
1	id		001 visio baronti monachi longoretensis
2	title		Visio Baronti Monachi Longoretensis
3	discursivedomain		religious
4	discursivegenre		hagiography
5	authorname	author	anonymous; –
	authorsurname	surname	N/A
6	referendary		N/A
7	recipient		N/A
8	identifier		Baront.
9	compositiondate	cp date	678–79
	compositiondatestart	cp date (not before)	0678-03-25
	compositiondateend	cp date (not after)	0679-12-01
	compositiondateformal	cp date (mean)	0678-12-01
10	manuscriptdate	ms date	around 800
	manuscriptdatestart	ms date (not before)	0775-01-01
	manuscriptdateend	ms date (not after)	0825-01-01
	manuscriptdateformal	ms date (mean)	0800-01-01
11	editorname	editor	Wilhelm Levison
12	editionbibliography	edition	MGH SS rer. Merov. 5, Hannover/Leipzig 1910: 368–394.
13	manuscriptidentifier	manuscript	Bayerische Staatsbibliothek, München, Shelf mark: Clm 6293
14	digitalizationsource	source	dMGH
15	funding		ANR; DFG
16	morphosyntactictagset	morphosynt	comphistsem
17	status	annotation status	verified
18	dmghlink		http://www.mgh.de/dmgh/resolving/ MGH_SS_rer_Merov._5_S._368
19	tokencount	tokens	5160
20	textorder		001

4.1 Identification tag

- (a) Tag Name: *id*
- (b) Definition: a numeral identifier with the text title is used for internal order and storing purposes of TXM. See in detail the tag *title* and *textorder* for this.
- (c) The equivalent in the BFM for this tag name is: *ID*.

4.2 Title

- (a) Tag Name: *title*
- (b) Definition: Complete title of a text as it is indicated by the scientific editor who used mostly the title found in the original manuscript or added later by time. If no title is indicated, like primarily in the epistolary genre and with charters, we have established one recurring to other facts like *author*, *recipient* etc.
- (c) Example: *Epistula domni Remedii episcopi ad domnum Chlodoveum regem*.
- (d) The equivalent in the BFM for this tag name is: *Titre*.

4.3 Discursive domain

- (a) Tag Name: *domainediscursif*
- (b) Definition: Functional trait corresponding to the principal destination of the text and the domain of activity to which it belongs to (cf. similar definition by the (French) Manual of the BFM⁷, p. 8–9).
- (c) In the PaLaFraLat subcorpus three discursive domains are used to group the texts. Texts concerning Christianity are described by *religious* (mainly hagiographic texts like saints' lives and visions), texts telling events in the past or letters to someone (e.g. between the king and a bishop) are *historical* and texts establishing rules to social life or regulating social life fall under the term *juridical*.
- (d) The equivalent in the BFM for this tag name is: *Domaine discursif*.

4.4 Discursive genre

- (a) Tag Name: *genrediscursif*

⁷http://ccfm.ens-lyon.fr/IMG/pdf/Manuel_Descripteurs_BFM.pdf

- (b) Definition: Trait corresponding to formal internal text properties, which are difficult to define, but following common sense (cf. similar definition by the (French) Manual of the BFM, p. 9–10).
- (c) Example: In the PaLaFraLat subcorpus five discursive genres are used to describe the formal internal text property. In the religious domain only *hagiographic* texts are found, the historical domain knows the *epistolary* genre for the letters as well as texts of the *historiography* and the juridical domain differentiates between the *formularium* genre and *charter* genre.
- (d) The equivalent in the BFM for this tag name is: *Genre discursif*.

4.5 Full name of the author

4.5.1 Name of the author

- (a) Tag Name: *authorname*
- (b) Definition: The name of the author of the text. Reference is the MGH edition or where applicable newer research.
- (c) Example: *Alcuin* or *Fredegar*. Although, in the case of the *Chronicle of Fredegar*, the name *Fredegar* was later artificially added and does not refer to the real name of the author which remains anonymous, thus far.
- (d) The equivalent in the BFM for this tag name is: *Auteur*.

4.5.2 Cognomen of the author

- (a) Tag Name: *cognomen*
- (b) Definition: the cognomen of the author, if any (otherwise it is marked by “–”), is recorded separately in this tag, including as well official titles or marks of origin of the author.
- (c) Example: *episcopus, diaconus, dux, monachus, rex/regina* etc.
- (d) There is no equivalent for this tag name in the BFM.

4.6 Referendary

- (a) Tag Name: *referendary*
- (b) Definition: The referendary who recognizes in a charter the validity of the causa. This is a category which is only applicable to the discursive genre of the charters.
- (c) Example: *Chrodinus, Droctoaldus, Tetbertus*, etc.
- (d) There is no equivalent for this tag name in the BFM.

4.7 Recipient

- (a) Tag Name: *recipient*
- (b) Definition: person or institution who should receive the letter (1), who is beneficiary/covenantee of the charters act (2), a convicted or involved person in a charter (3).
- (c) Example: (1) *Chlodoveus rex*, (2) *monasterium sancti Dionysii*, (3) *Amalbercthus*
- (d) There is no equivalent for this tag name in the BFM.

4.8 Identifier

- (a) Tag Name: *identifier*
- (b) Definition: A sort of a short identification title of the author's text. Like in the BFM, an abbreviation of the title is used to identify the text. The reference for the Latin subcorpus is the German Dictionary of Medieval Latin (*Mittellateinisches Wörterbuch*⁸) headed by the *Bayrische Akademie der Wissenschaften* (BAAdW)⁹,
- (c) Example: *Vita Adelph.* for *Vita Adelphii Abbatum Habendensis*
- (d) The equivalent in the BFM for this tag name is: *AuteurRef*.

4.9 Date of composition

- (a) Tag Name: *cpdate*
- (b) Definition: The composition date is noted as it is indicated by the scientific editor or historian of reference. A formal normalization of the date is processed following the *Digital Bodleian*¹⁰.
- (c) Format: circa = *c. yyyy*; period of time = *yyyy – yyyy*, *1rd quarter of the 1th century*, *Mid 1th century*, *1nd half of the 1th century*, *late 1th century – late 2th century*, *1th century – 2th century*, *c. 0912-0914*, *Mid 1th century – 2th century*.
- (d) The equivalent in the BFM for this tag name is: *Date de composition libre*.

For a sorted chronological order, the dates which are in a free format have to be normalized to exact dates, even if they are artificially constructed:

⁸<http://www.mlw.badw.de/>

⁹<https://www.badw.de/>

¹⁰<https://digital.bodleian.ox.ac.uk/>

4.9.1 Start date of composition

- (a) Tag name: *compositiondatestart*
- (b) Definition: The oldest date of the period of time that is indicated in the date of composition field is normalized to the format yyyy/mm/dd. If the exact day is not known, we set 1/1/yyyy.
- (c) Example: For the free formatted date *2nd half of the 8th century*, we use *0750-01-01* for the beginning.
- (d) The equivalent in the BFM for this tag name is: *Date début*.

4.9.2 End date of composition

- (a) Tag name: *compositiondateend*
- (b) Definition: The most recent date of the period of time that is indicated in the date of composition field is normalized to the format yyyy/mm/dd. If the exact day is not known, we set 1/1/yyyy.
- (c) Example: For the free formatted date *2nd half of the 8th century*, we use *0800-01-01*.
- (d) The equivalent in the BFM for this tag name is: *Date fin*.

4.9.3 Formal date of composition

- (a) Tag name: *compositiondateformal*
- (b) Definition: The middle date of the period of time that is indicated in the date of composition field is normalized to the format yyyy/mm/dd. If the exact day is not known, we set 1/1/yyyy.
- (c) Example: For the free formatted date *2nd half of the 8th century*, we use *0775-01-01*.
- (d) The equivalent in the BFM for this tag name is: *Date de composition formalisée*.

4.10 Date of the manuscript

→ Cf. the section about the date of composition.

The equivalent in the BFM for this tag name is: *Ms date libre*.

4.10.1 Start date of the Manuscript

→ Cf. the section about the date of composition.

The equivalent in the BFM for this tag name is: *Ms date début*.

4.10.2 Final date of the Manuscript

→ Cf. the section about the date of composition.

The equivalent in the BFM for this tag name is: *Ms date fin*.

4.10.3 Formal date of the Manuscript

→ Cf. the section about the date of composition.

The equivalent in the BFM for this tag name is: *Ms date formelle*.

4.11 Name of the editor

(a) Tag name: *editorname*

(b) Definition: Name of the scientific editor, composed by: firstname (*editorfirstname*) + surname (*editorsurname*)

(c) Example: *Bruno Krusch*.

(d) The equivalent in the BFM for this tag name is: *Éditeur scientifique*.

4.12 Bibliography of the edition

(a) Tag Name: *editionbibliography*

(b) Definition: the bibliographic reference of the (MGH) edition composed by: the publishing house (internal: *editionbibliography*) + place of publication (internal: *publicationplace*) + year of publication (internal: *publicationyear*) + first page of the chapter (internal: *publicationpages-tart*) + last page of the chapter (internal: *publicationpagestart*).

(c) Example: *MGH SS rer. Merov. 7, Hannover/Leipzig 1920: 19–40*.

(d) The equivalent in the BFM for this tag name is: *Titre complet*.

4.13 Identification of the manuscript

(a) Tag Name: *manuscriptidentifier*

(b) Definition: the manuscript that was mostly used by the editor or/and is the archetype. The tag is composed by: the institution storing the manuscript (internal: *manuscriptidentifierinstitution*) + the name of the city in the national language (internal: *manuscriptidentifiercity*) + the shelfmark for this manuscript (internal: *manuscriptidentifiershelfmark*).

(c) Example: *Bibliothèque nationale de France, Paris, lat. n. 5240*

(d) The equivalent in the BFM for this tag name is: *Manuscrit de base*.

4.14 Digitalization source

- (a) Tag Name: *digitalizationsource*
- (b) Definition: the source of digitalization is provided in this field. The whole PaLaFraLat sub-corpus is derived from the html source files of the *dMGH*, the *digital Monumenta Germaniae Historica*¹¹.
- (c) Example: *dMGH*
- (d) The equivalent in the BFM for this tag name is: *Origine de texte*.

4.15 Funding

- (a) Tag Name: *funding*
- (b) Definition: The PaLaFra research project is funded by the national research agencies *Deutsche Forschungsgesellschaft (DFG)*¹² and *Agence Nationale de la Recherche (ANR)*¹³.
- (c) Example: *DFG* and *ANR*
- (d) There is no equivalent for this tag name in the BFM.

4.16 Morphosyntactic tagset

- (a) Tag Name: *morphosyntactictagset*
- (b) Definition: All texts have been subject to automatical lemmatization, POS-tagging and morphological annotation with subsequent manual correction procedures, using the appropriate tagset *lapos* with 16 pos categories and 13 different morphological features with specific subcategories.
 - see the *Tagset Description* for *Lapos*
 - see the *Annotation Guidelines*

4.16.1 Tagset

- (a) Tag Name: *tagset*
- (b) Definition: the name of the used tagset.
- (c) Example: *lapos*
- (d) There is no equivalent for this tag name in the BFM.

¹¹<http://www.dmgh.de>

¹²<http://www.dfg.de>

¹³<http://www.agence-nationale-recherche.fr>

4.16.2 Status

- (a) Tag Name: *status*
- (b) Definition: gives two values about the actual morphosyntactic process status. It permits to identify the steps of the half-automatic annotation process.
- (c) Example: *verified* vs. *not verified*
- (d) There is no equivalent for this tag name in the BFM.

4.17 Link to the *Digital Monumenta Germaniae Historica* (dMGH)

- (a) Tag Name: *dmghlink*
- (b) Definition: We provide the link to the beginning of the editors text.
- (c) Example: http://www.mgh.de/dmgh/resolving/MGH_SS_rer._Merov._5_S._368
- (d) There is no equivalent for this tag name in the BFM.

4.18 Textorder (Identification number which is only visible in Offline-TXM)

- (a) Tag Name: *textorder*
- (b) Definition: a unique identification number representing groups and the order of all texts
- (c) Chart of the text groups:

ID	text	domain	genre
1–29	Life of Saints	religious	hagiographie
101–148	Epistolae Austrasicae	historical	epistolary
201–236	Desiderii ep. Cardurc. ep.	historical	epistolary
301–311	Epistolae aevi Merovingici coll.	historical	epistolary
401–404	Historiographical Texts	historical	historiography
501–502	Pactus legis Salicae & Lex Ribuarua	juridical	legislative
601–609	Formulae	juridical	formulary
701–738	Merovingian charters	juridical	charter
801–802	Charters of the Arnulfings	juridical	charter

4.19 Dimensions

4.19.1 Words

- (a) Tag Name: *words (only visible in the TXM Online Portal)*
- (b) Definition: The number of all words is provided here.
- (c) Example: *Vita Richarii: 3667 words*
- (d) The equivalent in the BFM for this tag name is: *Nombre de mots.*

4.19.2 Tokens

- (a) Tag Name: *tokencount (TXM Software)/tokens (TXM Online Portal)*
- (b) Definition: The number of all tokens is provided here. Tokens are counted by the total number of words and punctuation marks as well.
- (c) Example: *Vita Richarii: 4280 tokens*
- (d) The equivalent in the BFM for this tag name is: *Nombre de mots.*