

## Index lexical

### Mode d'emploi

25 octobre 2020

#### 1. Explications générales

##### 1.1. Un fichier mère assorti de quatre fichiers par période

L'index lexical de la Partie 3 est livré en cinq versions qui proposent des présentations différentes, mais dont le contenu est identique (toutefois, voir section 3 *infra*) :

1. par étymologie (fichier mère) : liste des étymologies (une étymologie par ligne)
2. par étymon : liste des étymons (un étymon par ligne)
3. par formes AF : liste des formes AF (une forme AF par ligne)
4. par formes MF : liste des formes MF (une forme MF par ligne)
5. par formes FC : liste des formes FC (une forme FC par ligne)

Les versions 2 à 5 sont des reconfigurations du fichier mère (version 1, par étymologie). Le mode d'emploi explique d'abord le fonctionnement du fichier mère (section 2), puis celui des autres versions (section 3).

Présentant des vertus pour la recherche automatique, le fichier mère par étymologies permet d'appréhender toutes les formes diachroniques d'un mot. Pareille présentation n'existe pas dans les grammaires car elle est impossible à réaliser en version papier. A l'intention des usagers moins familiers avec les recherches automatiques et qui préfèrent la liste alphabétique classique, le fichier mère par étymologies est complété par les index habituels qui organisent l'accès par les quatre périodes 1) étymon, 2) AF, 3) MF, 4) FC. Ils répliquent ainsi ce que serait une version papier de l'index : il y aurait un index des étymons (éventuellement séparé en formes latines, germaniques, grecques etc., que dans la version électronique on peut regrouper en triant par la colonne "étymon attribut") qui fournit la liste alphabétique de tous les étymons ; il y aurait ensuite un index des formes AF qui produit la liste alphabétique de toutes les formes AF, un index des formes MF avec la liste alphabétique de toutes les formes MF, et enfin un index des formes FC qui propose la liste alphabétique de toutes les formes FC. Dans ces fichiers les différentes formes d'un mot occupent des lignes différentes, alors que dans le fichier mère elles sont placées sur le même ligne mais dans des colonnes différentes.

Enfin, indiquons que les abréviations pratiquées dans les divers fichiers (ainsi que dans le présent mode d'emploi) sont celles de la Partie 3 de la GGHF, rappelées en annexe.

##### 1.2. Deux formats électroniques

L'index est livré sous deux formats électroniques : Excel (.xlsx) et texte (.txt). La version Excel est contenue dans un seul fichier (un classeur avec cinq onglets), alors que la version .txt (colonnes délimitées par des tabulations, encodage Unicode) comporte cinq fichiers

indépendants.

Il est à noter que le soulignement des voyelles toniques est perdu dans le format .txt, qui donc ne permet pas cet affichage : la tonicité ici n'est marquée que par l'accent aigu (voir les explications en section 2).

Le format .xlsx permet des recherches automatiques ainsi que le tri par colonne : par exemple, afin de regrouper les étymons par langue (latin, germanique, francique, gotique, celtique etc.), on triera en fonction de la colonne "étymon attribut", ce qui affichera tous les étymons latins regroupés, tous les étymons franciques regroupés, etc.

Universel, le format .txt permet l'importation de l'index dans le logiciel de choix de l'utilisateur, par exemple dans une base de données, où des recherches automatiques avec des expressions régulières et sur plusieurs colonnes à la fois sont possibles, ainsi que l'affichage des seules lignes qui correspondent à la recherche.

### 1.3. Index des inscriptions et citations de l'Appendix Probi

La Partie 3 de la GGHF fait encore état d'inscriptions latines, d'attestations tardives de mots latins ainsi que de conseils prodigués par l'Appendix Probi. Un index séparé répertorie ces formes dans un fichier Excel ("inscriptions et Appendix Probi Partie 3 GGHF.xlsx") comportant deux onglets qui présentent, respectivement, les inscriptions et attestations tardives (classées en fonction du numéro de paragraphe) et les citations de l'Appendix Probi (classées en fonction de leur numérotation officielle).

## 2. Fonctionnement du fichier mère (par étymologie)

Le but de l'index est de permettre des recherches automatiques multi-critères qui puissent prendre en compte toutes les formes des mots en même temps. Ainsi par exemple "lister les mots présentant en latin i long initial non précédé de vélai qui ont e comme première voyelle en AF". Afin de permettre de telles recherches, il est nécessaire que chaque mot se trouve seul dans une cellule particulière, ses attributs (langue à laquelle il appartient, forme grammaticale, signifié etc.) se trouvant dans une cellule séparée.

L'index est raisonné. Les occurrences d'un mot donné ne sont pas toutes prises en compte : dans la Partie 3 un mot peut être cité à titre d'exemple d'un processus (ratiōne > *raison* pour la métathèse de j, par exemple), mais on n'apprend rien sur le mot à cet endroit. Plutôt que son occurrence dans le texte, ce qui décide donc de l'inscription d'un mot dans l'index est la question *est-ce qu'à cet endroit on apprend quelque chose sur l'évolution de ce mot?*

L'index repose sur une partition en quatre périodes : 1) étymon (latin, germanique, celtique etc.), 2) ancien français (AF), 3) moyen français (MF) et 4) français contemporain (FC). Cette périodisation est l'objet des colonnes.

Chacune des quatre périodes peut être représentée par plusieurs formes du même mot : des formes grammaticales différentes du même mot (frīgīdu masc., frīgīda fém. par exemple) lorsqu'il s'agit de l'étymon, des variantes pour les autres périodes (par exemple AF *genoil* et *genouil* à partir de *genūculu*). Le tableau comporte ainsi, successivement

- 3 colonnes pour l'étymon :           étymon 1, étymon 2, étymon 3
- 9 colonnes pour l'AF :           AF1, AF2, AF3, ..., AF9
- 3 colonnes pour le MF :           MF1, MF2, MF3
- 3 colonnes pour le FC :           FC1, FC2, FC3

Les lignes quant à elles répertorient les diverses formes évolutives qu'un mot donné a connu durant ces périodes : une ligne est une étymologie.

L'index renvoie à la Partie 3 de la GGHF au moyen des numéros de paragraphes de ladite partie. Ceux-ci apparaissent dans la colonne "renvois" placée en lieu central dans le tableau entre les étymons et l'AF. Le renvoi aux remarques des paragraphes se fait par leur numéro, séparé du paragraphe par un point : 324.6 par exemple renvoie à la remarque 6 du paragraphe 324.

étymon 1	étymon 2	étymon 3	renvois	AF1	AF2	MF1	MF2	FC1	FC2
cœmētēriu	cīmītēriu	°cīmīteriu	179.1, 288.1	cemetire	cimitire			cimetiēre	
cōquīna	°cocīna		97, 139, 324.6, 388	cuisine		cuisine	cuirene	cuisine	
crīspu			256, 372, 374	crep	crep			crēpu	crēpé

S'ajoutent à cela trois types de contenu :

- en tête du tableau, une colonne fournissant, le cas échéant, la source grecque de l'étymon ;
- une colonne "germ. mod." indiquant des représentants en anglais, allemand ou néerlandais modernes des étymologies en question ;
- en queue de tableau, quatre colonnes hébergeant les représentants des étymologies dans d'autres langues romanes (occitan, provençal, catalan, italien, espagnol, portugais, roumain, toscan, gascon). Ces mentions sont peu nombreuses.

grec	étymon 1	étymon 2	AF1	AF2	AF3	germ. mod.	FC	lang. rom.
κόλαφος	cōlaphum	°cōlopu	colp	coup			coup	
	māsculu	māscel	mascle	masle		angl male	mâle	
	ōculu		ueil	uel	oil		œil	it. occhio

En principe chaque mot contenu dans une de ces colonnes peut requérir la spécification d'un attribut (langue à laquelle il appartient, forme grammaticale, signifié etc.). C'est la raison pour laquelle à chacune des colonnes contenant un mot est associée une colonne "attribut" (sauf pour l'AF à partir de AF5, pour MF3 et pour FC3). On aura par exemple :

étymon 1	étymon 1 attr.	étymon 2	étymon 2 attr.	AF1	AF1 attr.	AF2	AF2 attr.	FC
botellu	acc sg	botellōs	acc pl	boel	sg	bouiaus	pl	boyau
meum		°mem	atone	mon				mon
ōculu		ōclu	AppPr 111		ueil	oil	entrée GDC, TL	
artemīsia	lc	°artemīsia		armoise		ermoise		armoise
recipiō	1s	°recipo		receif		reçoif		(je) reçois
brogilos	celt	brogilu		brueil	entrée TL	bruil	entrée TL	breuil
°blettjan	frk	°blettiāre		blecier				blessen

Les différentes formes d'un mot sont séparées par une colonne qui indique leur relation, qui peut être de trois types :

	étymon 1	étymon 1 attr.		étymon 2	étymon 2 attr.		étymon 3	étymon 3 attr.
énumération : virgule	bot <u>ellu</u>	acc sg	,	bot <u>ellōs</u>	acc pl			
"devient" : >	°bū <u>kōn</u>	frk	>	°būc <u>āre</u>				
"croisé avec" : X		v <u>adu</u>	X	°w <u>ad</u>	frk	>	°w <u>adu</u>	

Les étymologies ne sont pas toujours complètes, i.e. ne comportent pas toujours des représentants de toutes les périodes. Il y a ainsi des lignes qui montrent seulement des formes latines, des formes AF, des formes MF ou des formes FC. Cela est fonction du contenu de la GGHF, qui peut par exemple étudier les seules formes du latin d'un mot (lorsqu'il est question de l'apophonie interne). De même il est parfois question de la seule forme AF, MF ou FC d'un mot.

Lorsqu'un mot n'a pas de représentant en FC, son signifié est fourni dans la colonne "Sé (AF ou MF)". Il en va de même pour les cas où le mot a un représentant moderne, mais dont le sens n'est pas le même qu'en AF ou MF.

étymon 1	étymon 1 attr.		AF	Sé (AF ou MF)	germ. mod.	germ. mod. attr.	FC
f <u>uscina</u>		>	foisne	'trident'			
°app <u>ariculāre</u>		>	apareillier	'préparer'			appareiller
°hrok	frk	>	froc	'jupe'	Rock	nha	froc

Enfin, l'index comporte une colonne "type changement" qui indique, au premier chef, si l'évolution de l'étymon vers l'AF est de nature phonétique et populaire ou non. Lorsqu'elle ne l'est pas, il s'agit le plus souvent de changements analogiques, mais d'autres cas peuvent se présenter et cette colonne cherche à fournir des renseignements à leur égard. Il est à noter que seul le rapport entre la forme la plus évoluée de l'étymon et l'AF est évalué : l'évolution ultérieure n'est pas décrite (sauf mention explicite).

étymon 1	étymon 1 attr.		étymon 2	type chgt	AF1	AF2	FC
causa		>		phonét.	chose		chose
lūna		>		phonét.	lune		lune
clāmo	1s	>		analogique	claim		(je) clame
ca <u>ule</u>		>		chou analogique	chol		chou
ministeri <u>āle</u>		>	°minister <u>āle</u>	chute de i / yod	menestrel		ménestrel
sapio	1s	>		chute précoce de la labiale	sai		(je) sais
spōnsu		>		dialectal	espos		époux
aug <u>ustu</u>		>	°ag <u>ustu</u>	dissimilation	aost		août
fav <u>ōre</u>		>		emprunt (au latin)	favor		faveur
fīn <u>īre</u>		>		finir analogique	fenir	finir	finir
ex <u>īre</u>		>		issir analogique	eissir	issir	'sortir'
Nāt <u>āle</u>		>		obscur	noel		Noël

Reste deux problèmes techniques : l'astérisque et la tonicité des voyelles. Celle-ci ne peut pas faire l'objet d'une recherche automatique lorsqu'elle est indiquée par le soulignement. L'index comporte pour cette raison, pour chacune des trois colonnes d'étymon, une version "tonique soulignée" et une autre "tonique accent aigu". Dans cette dernière la tonicité des

voyelles est indiquée par un accent aigu (le cas échéant superposé au macron). Il existe ainsi quatre symboles distincts par voyelle qui peuvent être recherchés indépendamment (par exemple a bref atone = "a", a long atone = "ā", a bref tonique = "á" et a long tonique = "ǎ"). A noter que la diphtongue latine au lorsqu'elle est tonique est en mode accent aigu marquée par cet accent sur ses deux membres : au = áú. La colonne "tonique soulignée" permet d'afficher l'aspect graphique habituel.

Enfin, le fait que l'astérisque précède le premier caractère du mot dans les formes non attestées d'une part empêche le classement par ordre alphabétique, d'autre part complique des recherches automatiques visant le début du mot : par exemple, pour obtenir la liste des mots commençant par pa-, il faudra chercher #pa- et #°pa-. C'est la raison pour laquelle les trois colonnes étymon qui marquent la tonique par l'accent aigu afin de pouvoir servir à la recherche automatique ne comportent pas d'astérisque, qui apparaît dans une colonne à part à leur gauche.

étymon 1 attr.	étymon 1 tonique soulignée	étymon 1 astérisque	étymon 1 tonique accent aigu	étymon 2 tonique soulignée	étymon 2 astérisque	étymon 2 tonique accent aigu
1s pf	prehendi °vērācu vētulu	°	prehēdi vērācu vétulu	> °prēnsi °vēclu	°	°prēnsi véclu

### 3. Fonctionnement des fichiers organisant l'accès par période

La différence entre le fichier mère et les fichiers qui organisent l'accès par les quatre périodes 1) étymon, 2), AF, 3) MF, 4) FC est l'inscription des différentes formes d'un mot : dans des colonnes différentes pour le fichier mère, dans des lignes différentes pour les autres fichiers.

A titre d'exemple, prenons le cas d'un mot AF qui dans l'index connaît des formes multiples. Voici le cas concernant les trois mots examinés en section 2, tel qu'ils sont affichés dans le fichier mère : les colonnes AF1, AF2, AF3 etc. contiennent les différentes formes du même mot, selon le principe "une ligne, une étymologie".

grec	étymon 1	étymon 2	AF1	AF2	AF3	germ. mod.	FC	lang. rom.
κόλαφος	colaphum	°colopu	colp	coup			coup	
	masculu	mascel	mascle	masle		angl male	mâle	
	oculu		ueil	uel	oil		œil	it. occhio

Le même contenu affiché par ordre alphabétique des formes AF (fichier "par formess AF") se présente de la manière suivante.

AF	grec	étymon 1	étymon 2	germ. mod.	FC	lang. rom.
colp	κόλαφος	colaphum	°colopu		coup	
coup	κόλαφος	colaphum	°colopu		coup	
mascle		masculu	mascel	angl male	mâle	
masle		masculu	mascel	angl male	mâle	
ueil		oculu			œil	it. occhio
uel		oculu			œil	it. occhio
oil		oculu			œil	it. occhio

Ici les différentes formes AF d'un mot occupent donc des lignes différentes (et non pas des colonnes différentes). Par conséquent les colonnes AF1, AF2, AF3, ..., AF9 ont disparu. Toutes les autres colonnes du fichier mère en revanche sont maintenues : la ligne de chaque

forme AF précise les différents étymons, les différentes formes MF, FC etc. Si donc il existe trois formes AF pour un mot donné, comme pour ocolu dans le tableau ci-dessus, il y aura trois lignes qui sauf pour la première colonne seront identiques : dans les autres colonnes la même l'information est répliquée trois fois puisqu'elle se rapporte de la même manière aux trois formes AF.

Il en va de même pour les autres fichiers qui organisent l'accès par période. A titre d'exemple, considérons le même contenu tel qu'il apparaît dans le fichier par étymon.

étymon	grec	AF1	AF2	AF3	germ. mod.	FC	lang. rom.
<u>col</u> aphum	κόλαφος	colp	coup			coup	
° <u>col</u> opu	κόλαφος	colp	coup			coup	
<u>mas</u> culu		mascle	masle		angl male	mâle	
<u>mas</u> cel		mascle	masle		angl male	mâle	
<u>ocu</u> lu		ueil	uel	oil		œil	it. occhio

Ici ce sont les colonnes étymon 1, étymon 2, étymon 3 qui, fusionnées dans la colonne étymon, ont disparu. Le contenu de la ligne concernant l'étymologie du FC *coup* apparaît ainsi deux fois, sur la ligne colaphum et sur la ligne °colopu.

Il est à noter que le contenu des quatre fichiers qui organisent l'accès par période est identique à celui du fichier mère, sauf pour ce qui concerne les mots qui n'ont pas de forme répertoriée pour la période en question. En effet, il existe de nombreux mots dans le fichier mère qui n'ont pas de forme AF répertoriée, tels celt Caturigas > FC *Chadorges*, ou encore FC *humus*, *bêtise* etc. Ces mots sont supprimés dans le fichier "par formes AF" puisqu'ils n'y auraient pas d'entrée. Il en va de même pour les trois autres fichiers par période : dans le fichier par étymon, les mots du fichier mère qui n'ont pas d'étymon répertorié n'existent pas, dans le fichier par MF les mots qui n'ont pas de forme MF répertoriée n'apparaissent pas, et dans le fichier par FC les mots qui n'ont pas de forme FC répertoriée sont absents (par exemple ceux qui n'ont pas de représentant moderne tel °vënditiōne > *vençon* 'vente')

Enfin, le fichier par formes FC a une particularité dont il est utile de faire état. D'ordinaire, lorsqu'une portion seulement de la forme FC correspond à l'étymon, cela est indiqué par des parenthèses : (à) *jeun*, (bête de) *somme*, (em)*preindre*, (re)*gain*, (ré)*soudre*, (il) *aime*, (j') *espère*, (je) *bois*, (qu'il) *chevauche*, (que j') *aille*, (tu) *sers*, etc. Cela empêchant le classement alphabétique de la portion concernée, une colonne "préfixe" a été ajoutée au fichier par formes FC qui dans le cas de mots séparés contient ceux en parenthèses :

préfixe	FC	FC attr.
(à)	jeun	
(bête de)	somme	
(il)	aime	
(qu'il)	chevauche	
(je)	bois	

Au cas où la portion entre parenthèses fait partie du mot, celui-ci est inscrit sans parenthèses dans la colonne FC, et sa version avec parenthèses apparaît dans la colonne FC attribut :

préfixe	FC	FC attr.
	absoudre	(ab)soudre
	conduire	(con)duire
	émincer	(é)mincer
	empreindre	(em)preindre
	regain	(re)gain

Cela permet le classement entièrement alphabétique de ces entrées dans la colonne FC

#### 4. Annexe

Rappel des abréviations pratiquées dans la Partie 3 de la GGHF qui sont également applicables à l'index lexical.

objet	signifié par	exemple
<i>général</i>		
tonicité de la voyelle	voyelle soulignée	p <u>a</u> ter
	mots provenant d'autres langues que le latin (notamment germaniques) : c'est l'accent dans la forme romanisée qui compte pour l'évolution, non celui de la forme germanique. (frk °plegan > °plegāre > <i>plegier</i> FC <i>pleiger</i> ). Par conséquent la voyelle tonique des formes germaniques demeure non marquée (frk °plegan). Elle l'est en revanche dans les formes romanisées (got / vha <i>magan</i> > °exmagāre > <i>esmaiier</i> 'troubler').	°plegan °exmagāre
longueur vocalique	macron	ratio <u>ne</u>
brévitité vocalique	aucune indication particulière	ratio <u>ne</u>
diphthongue	aucune indication particulière	p <u>i</u> ra > <i>peire</i> > <i>poire</i>
hiatus	si utile, les deux voyelles sont séparées par un point : V.V. La notation à tréma pour l'AF ( <i>chäi<i>ere</i></i> ) n'est pas pratiquée.	°cath <u>e</u> dra > <i>cha.iere</i>
formes AF	sauf indication contraire, les formes citées sont pour les noms le cas régime singulier, pour les verbes l'infinifit présent	
<i>langues, étapes évolutives</i>		
lat	latin (indifférencié pour lc, lt, ltm)	
lc	latin classique	
lt	latin tardif (= latin vulgaire selon la tradition)	
ltm	latin médiéval	
pfr	proto-français : étape évolutive qui n'est plus le latin et non encore l'AF	
AF	ancien français	
MF	moyen français	
FC	français moderne (début 21 <sup>e</sup> siècle)	
germ	germanique	
frk	francique	

objet	signifié par	exemple
got	gotique	
gr	grec	
a occ	ancien occitan	
vha	vieux haut allemand (8 <sup>e</sup> -11 <sup>e</sup> siècles)	
mha	moyen haut allemand (11 <sup>e</sup> -14 <sup>e</sup> siècles)	
nha	nouveau haut allemand (depuis le 15 <sup>e</sup> siècle)	
all	allemand moderne	
celt	celtique	
pic	picard	
it	italien	
angl	anglais	
prov, aprov	provençal, ancien provençal	
hbr	hébreu	
arab		arabe
jud.fr	judéo-français	
<i>indications</i>		
<i>grammaticales</i>		
sg, pl	singulier, pluriel (si survenant seuls)	
masc, fém, neut.	masculin, féminin, neutre (si survenant seuls)	
m, f, n	masculin, féminin neutre (si survenant accompagnés)	CSm, 3f
nom, gen, acc, abl	nominatif, génitif, accusatif, ablatif	
1s, 3p, 2s, 6p etc.	pour les verbes, indication de la personne (chiffre) et du nombre (sg, pl)	1s vīdeō
CS, CR	en AF : cas sujet, cas régime	
CSs, CSp, CRs, CRp	en AF : cas sujet sg / pl, cas régime sg / pl	
pf, imp, subj, fut, inf, ipf	parfait, impératif, subjonctif, futur, infinitif, imparfait	
#	début ou fin du mot	#d (d initial) d# (d final)
X	croisement croisement entre syncope et épenthèse FC <i>chignon</i> résultat d'une influence par FC <i>tignon</i> ?	[syncope X épenthèse] <i>chignon</i> 'collier' (FC <i>chignon</i> X <i>tignon</i> ?)
[...]	transcription phonétique	
(...)	dans les formes latines, item(s) perdu(s) dans l'évolution ultérieure	lep(o)re > <i>lievre</i> , gal(bi)nu > <i>jaune</i>
<...>	forme graphique	
pop.	évolution populaire, par opposition à des évolutions savantes ou par emprunt	
o.i.	origine inconnue	
<i>dictionnaires</i>		
App.Pr. 13	Appendix Probi, mot n° 13	
TL6-1203 : 45	Tobler et Lommatzsch (1925-1995), tome 6, page 1203, ligne 45	



objet	signifié par	exemple
GD	Godefroy (1881-1902)	
GDC	Compléments de Godefroy (1881-1902)	
TLFi	Trésor de la Langue Française informatisé	
EM	Ernout et Meillet (2001)	
TACF	Tableaux de conjugaison de l'ancien français (Okada et Ogurisu 2007-2012)	
DELG	Dictionnaire Etymologique de la Langue Grecque (Chantraine 1968-1977)	
REW	Romanisches Etymologisches Wörterbuch (Meyer-Lübke 1935)	
DECT	Dictionnaire Electronique de Chrétien de Troyes ( <a href="http://www.atilf.fr/dect/">http://www.atilf.fr/dect/</a> )	
FEW 3 : 123b	Französisches Etymologisches Wörterbuch (von Wartburg 1934-2002), tome 3, page 123, colonne b.	
Nègre 10234	n° 10234 de Nègre (1990-1991)	
<i>textes</i>		
StLég	Saint Léger	
Eul	Eulalie	
Rol	Roland	
Alex	Alexis	
Serm	Serments	
Pass	Passion	
ChTr	Chrétien de Troyes	