JÓZSEF HERMAN

LATE LATIN DATA BASE
GUIDELINES FOR DATA COLLECTION

INSTITUTE FOR LINGUISTICS OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST, P. O. BOX 19, 1250  T.: (36-1)-1-75-82-85, 1-56-12-44
LATE LATIN DATA BASE
GUIDELINES FOR DATA COLLECTION

I. Basic concepts and procedures
1. For the purposes of this project, *data* are defined as any linguistic phenomena that can be isolated at text level (in terms of surface manifestation, such phenomena can be phonetic [orthographic], morphological, syntactic, lexicographical, or lexicosemantic) and that deviate from what is called the ‘classical’ norm. For the data collector, ‘datumhood’ is based on mere difference, defined as above, irrespective of whether that difference corresponds to his/her expectations in terms of diachronic processes that he/she may be aware of, whether it is due to linguistic or extralinguistic (i.e. graphic or technical etc.) factors. Whenever the collector is uncertain if a phenomenon deviates from the norm or corresponds to it, or when such a decision would require special investigations, the phenomenon has to be treated as datum and recorded; doubts are to be indicated under “REMARK” by a question mark or the words “forsitan recte”.
2. The following figure represents — in a reduced format — a data form of the Late Latin Data Base:

LLAT.DBF data form
1. CODE_1 .................................. CODE_2 ..................................
2. TEXT ..................................................................

..................................................................
3. BIBL ..................................................................
4. LOC_1 ......... LOC_2 ......... LOC_3 ........
5. YEAR .......... CENTURY ........ QUARTER ....
6. SPEC_1 ........ SPEC_2 ........ SPEC_3 ....
7. EVAL ........................................
8. REMARK ..................................................................

The specific rules for filling in rubrics are given heading-by-heading in section II. below. General rules are as follows:
a/ each occurrence of each datum is to be recorded on a separate data form, i.e. each data form should contain exactly one occurrence of a datum. Exceptions are deviations coded as 2.1 or 2.5 /æ > E,
ae > E/; these are to be recorded on collective data forms, cf. I.3. below;

b/ if an inscription or, in general, a bibliographically identifiable minimal text unit contains several data, headings 4., 5., and 6. of the current data need not be filled in from the second datum onwards, as these would be repetitions of identical information in each form; the program automatically supplies the missing code numbers;

c/ if several identical data occur in a single inscription or text unit, a separate data form is to be devoted to each occurrence, with the restrictions noted under b/ above; to avoid misunderstanding, however, heading 2. (TEXT) will have to contain the serial number, written in Arabic numerals, of the occurrence; this number has to be preceded by three spaces.

Example: if a particular inscription contains two occurrences of mese (for mense), rubric 2. of the two data forms will have to contain 'MESE 1' and 'MESE 2', respectively.

d/ if a minimal text unit (= word) contains two or more data, each has to be recorded in a separate form, with the same word appearing in rubric 2. (TEXT) in both forms.

Example: if a word AGOSTO (for agosto) is found, the collector has to fill out a data form with the code 2.11 /au > A/ and another with 1.9. /ī > O/; rubric 2. will contain AGOSTO in both cases; rule b/ above applies as usual.

3. Certain types of data that are extremely frequent in all areas of the Empire and throughout the post-classical period will be recorded, for simplicity's sake, by number of occurrence only, within larger groups and territorial units. Such data, obviously, are aē > E (code numbers: 2.1 and 2.5); it may turn out later that further code categories are also to be treated in this way. The tool for recording such data is the collective data form, containing the following headings:

II. How to fill in data forms and collective forms

Data forms
Instructions are given separately for each heading.

1. CODE
This rubric will contain the code number that identifies the datum in terms of the linguistic phenomenon involved. Code numbers are to be looked up in the tables of Appendix I. (list of code numbers).

This heading contains two subheadings:

   CODE.1       CODE.2

If the linguistic characterization of the datum is obvious or non-problematic, the collector selects the appropriate code number, inserts it under CODE.1, leaving subheading CODE.2 empty.

Example: FECIRVNT can only be interpreted as containing the deviation ē > ı; the appropriate code number is 1.4; number 1 is to be written before, and number 4 after the dot found following the subheadings.

Whenever a datum is ambiguous, the collector will look up the code numbers of both interpretations and write one next to CODE 1, the other next to CODE.2. The order does not suggest preference or degrees of probability as determined by the collector.

Examples: ANNVS for annos, in dates, can be interpreted as deviation of the type ó > V (1.18) or as an instance of case substitution nom. pro
acc. (7.2). Both code numbers are to be recorded. VISO for visu abl. can be ü > O (1.20) or declination shift (decl. II pro IV, 7.15).

It may be the case that one possible interpretation of a particular form is deviation from the classical norm, hence a relevant datum, but following the other interpretation the form is regular, hence irrelevant. In such cases datum interpretation is written under CODE.1 as usual, but the other possibility is not recorded under the subheading CODE.2 but in rubric 8 (REMARK); see below for the way this is done.

Example: AGET can be interpreted as standing for agit, code number 1.15 /i > E/, but if the context makes this possible, it can also be a regular future form.

2. TEXT
The relevant stretch of text containing the datum. The amount of text to be recorded is the minimal amount unambiguously showing what type of deviation from the classical norm is involved. For a phonetic/orthographical phenomenon, it is usually enough to record the word in which the phenomenon occurs. E.g. FECIRVNT or CARISIMO unambiguously shows what the deviation is, further elements of the text would be superfluous. For morphological or syntactic phenomena, it is normally necessary to record those elements of the context (omitting irrelevant material) that determine the interpretation of the datum.

Examples: VISO in itself does not show that it is a variant of classical visu abl., the syntagm EX VISO does; therefore, the latter has to be recorded. If IVCVNDO PVERO is intended as dat. poss., this fact can be indicated by quoting the full syntagm ARCA IVCVNDO PVERO.

With respect to lexical innovation (neologisms, apparently novel borrowings), it is sufficient to record the word itself since in such cases further detailed investigations are called for, investigations that the collector cannot undertake during data collection. Lexicological data are thus mere reminders: the collector is not required to do lexicological research to ascertain the relevance of such data. If the datum appears to be relevant from a lexico-semantic point of view, recording the context may be useful for bringing out the intended meaning.

Example: credulitas, in addition to its usual pejorative/depreciative meaning, occurs — though extremely rarely — with a specifically Christian meaning ‘persistance in one’s faith’; to indicate this, longer context is to be quoted: HVC IBIMVS ET NOS SIT MODO SANCTA FIDES SIT PIA CREDULITAS.

Technicalities of transliteration:

a/ texts are to be transliterated in capital letters, in accordance with Latin epigraphic writing; each letter is to be written in a separate slot , , , , ,

b/ transliteration is to be reproduced faithfully as it appears in printed text editions, with ligatures resolved (i.e. replaced by distinct characters).

c/ in transliterated texts, the following conventions are to be used:

[...] [ ] IVVENI [s]... square brackets indicate illegible portion of the original text, with the number of dots corresponding to that of missing illegible letters; if the number of letters can not be determined, without dots. The editor’s conjecture is only reproduced (in lower-case letters) if it is necessary for the correct interpretation.

... ( ) ................. parentheses indicate portions of text omitted by the collector as irrelevant for the linguistic phenomenon under consideration.

(...)
e.g. SV(MV?)S for an obscure MV ligature

portion of text whose reading the collector finds obscure or questionable are enclosed in parentheses and followed by a question mark.

e.g. PICTORIBVS D V OBIS

within a longer text, the word or word group containing the datum is highlighted by double space.

(e.g. LAVNIONI/VIXIT

a single slash indicates the beginning of new line on the stone.
3. BIBL

This heading is to be completed with a reference to the bibliographical source of the datum.

a/ For data taken from Corpus Inscriptionum Latinarum, only the volume number is recorded (in Arabic numerals), followed by a dot, followed by the inscription number. The latter is not followed by a dot, except for inscriptions of more than ten lines, where the inscription number is followed by another dot and the line number. Thus, '3.4210' appearing under BIBL indicates the inscription CIL III.4210, '3.4150.21' indicates the 21st line of CIL III.4150. If the inscription number is a collective number of several inscriptions each having its own sub-number, the inscription number is followed by '//' and the sub-number (without a following dot).

b/ If the datum is taken from some other collection of inscriptions, the abbreviations found in Appendix II (list of sources) is to be used, followed by the inscription number (without an intervening dot), thus ‘RIU 200’ stands for Die römischen Inschriften Ungarns, Budapest 1972, inscription No 200.

c/ For data taken from other sources (annals, periodicals), the abbreviation taken from the list of periodicals (Appendix II) is followed by volume and page number (no dot preceding the volume number), thus 'AA 4, p.25' for Archaeologischer Anzeiger, Volume 4, page 25. If the inscription quoted is identified by number in the source, the page number is followed by a dot and the number of inscription as above.

d/ If the collector collects data from a collection or periodical not appearing in the lists of Appendix II, he/she should consult the director of this project concerning the abbreviation to be used, and/or submit the abbreviation he/she wants to use for inclusion in the lists.

e/ For data taken from Corpus Inscriptionum Latinarum, earlier text editions need not be referred to. For data taken from other sources, an = sign followed by the numbers of CIL refers to the fact that the inscription also occurs in the Corpus. Thus e.g. 'RIU 196 = 3.4252' means that the datum was taken from RIU but the inscription was previously published in the Corpus under the number given. Antecedents other than the Corpus need not be referred to.

4. LOC

This heading is for localizing the datum geographically. The three subheadings are as follows:

LOC.1 province
LOC.2 larger administrative unit within a traditional province
LOC.3 town or village

For subheadings LOC.1 and LOC.2 use the abbreviations of Appendix III.

III. Further rules to observe are the following:

A. For completing LOC.3:
   The place name is only given if the Latin name is known; in that case, the first five letters of the name are to be recorded, e.g. AQVIN = Aquincum. If the name is less than five letters long, superfluous slots are left empty, e.g. VEII = Veii. If the Latin name of the town is unknown, the procedure is the following:
   a/ if, according to the collection used, the inscription is found in the vicinity of a place whose Latin name is known, that Latin name is given between slants, e.g. /AQVIN/ = in the vicinity of Aquincum (e.g. in Törökbálint, but this is not recorded);
   b/ if there is no such nearby place with a Latin name, the rubric is left empty.

B. For completing LOC.2:
   This rubric is invariably filled in on the basis of Appendix III, irrespective of whether LOC.3 is also filled in or not. There are two types of exception when LOC.2 remains empty:
   a/ when LOC.3 has ROMA in it;
   b/ when Appendix III does not provide a LOC.2 unit within the given province, e.g. AD.

C. For completing LOC.1:
   Since LOC.2 always implies LOC.1, the latter rubric is usually left empty, to be filled in by the program automatically. Thus LOC.1 is only filled in by the collector in the following cases:
   a/ in the case defined in B.b/ above;
   b/ when both LOC.2 and LOC.3 are unfillable, i.e. the exact provenance of the inscription (within the province) is unknown. Attention: this rule always applies for inscription found on movables (mobilier), since the places where these were made and found are not necessarily identical.
5. YEAR – CENTURY – QUARTER

These rubrics are for the date of origin of the inscription or text unit containing the datum. Instructions for completing them are as follows:

A. The subheading YEAR can only be filled in if the inscription is textually dated, i.e. contains the exact year when it was written. Consular, indictional, or other calendar date of the inscriptions are given in the sources in terms of the current calendar style: this date is to be recorded. The text may contain further details of the date (month, day); these are not to be recorded. If the transposition of a dated inscription in terms of modern calendar style is uncertain (e.g. an incompletely recorded consul’s name or an indiction may correspond to several years), all possibilities are to be noted. If the date refers to a span of several years (e.g. 506-8), this is to be given, as in the printed source.

B. The subheading CENTURY is to be completed only if YEAR remains empty. This is done as follows:

a/ if the source gives an estimated century, it is recorded (in Arabic numerals). E.g. among the notes to DIEHLCHR 1593 we find the following: descr. Gsell, qui saec. attributed III. This is sufficient basis for recording 4 for CENTURY. In cases of estimates like “saec.IV-V”, the subheading will contain 4-5.

b/ on the basis of historical hints (emperors or other historical figures mentioned in the text, reference to events, or other historical considerations), the collector can also supply the estimated century number. This is indicated by a “/” following the number; 5/ thus means that it is the collector who thinks the text is from the fifth century.

Attention: estimates are not to be based on linguistic considerations; this would imply circularity.

c/ if the century cannot be estimated, the subheading remains empty.

C. The subheading QUARTER (of a century, numbered 1 to 4) can be completed if CENTURY contains an estimated number. The sources of quarter estimates are, mutatis mutandis, the same as for centuries. If a plausible estimate is not possible, the rubric remains unfilled.

6. SPEC

The purpose of these three subheadings is to indicate the type of inscription in order to facilitate linguistic analysis or interpretation. Instructions are as follows:

SPEC.1 Serves to indicate whether the inscription is Christian or not: if it is (this fact is generally noted by the collectors and the text itself nearly always reveals anyway if this is the case), the rubric is to be filled by a + sign. Otherwise it remains empty.

SPEC.2 Serves to distinguish verse from prose. If the inscription is in verse, the rubric is filled by a letter P (for poema); if it is in prose, the rubric remains empty.

SPEC.3 Official texts are identified by O, milestones as a special case by OM. For private inscriptions, the rubric remains empty. The distinction is usually trivial; tombstone inscriptions, votive inscriptions installed by private individuals are obviously private; building inscriptions, or those referring to emperor worship, legal texts, etc., are official. In dubious cases, the installer is the clue: if it is a state or city administrative body, a high official, a collegium, or a military unit, the inscription can be identified as official.

7. EVAL

Serves to indicate factors determining the evaluation of the datum, if any. The following abbreviations can be used:

MS = the inscription is known from a manuscript copy, not in the original.
FR = fracture on the stone (in the datum!);
ABR = the letters are abraded, difficult to read (in the datum!);
CONF = the inscription contains confuse sets of letters; CONF.1: in one or two words; CONF.2: extensive confuse portions; CONF.3: the inscription is almost completely confuse.

8. REMARK

Any further remark facilitating linguistic interpretation can be made here. Possible types:

A. Further coding possibilities (other than those noted under heading 1; thus: ADDITIONAL CODE/S/;....

B. The way the datum fits into the linguistic character of the text: it is especially important to note if the deviation concerned is isolated within the text it occurs in. For instance, if we find a datum FECIRVNT, it
may be important to note that the same word occurs in the text with no deviation from the norm, thus: "ibid.FECERVNT".

C. Since peculiarities of the overall appearance of the text cannot be reflected by the present method of collection, remarks facilitating linguistic evaluation can be made in English or Latin, e.g. "litt. I et E simil."

D. If the datum can be interpreted as corresponding to the classical norm, this is to be noted here; referring back to the example AGET in II.1 above, this can be done as recte futurum verbi?

If there are several remarks, supply serial numbers.

Collective data forms

For collective data forms, CODEs are as above (cf. I.1,a); LOC corresponds to LOC.2 units or large cities. CENTURY should possibly contain broad estimates (2-3, 4-5, 6). FREQ contains the number of data occurring in the given territory and period. The simplest way of recording that number is to draw lines in groups of five, and give the sum in Arabic numerals on closing the sheet.

APPENDIX I

1.1 ā > E 1.11 ā > E 2.1 ae > E
1.2 ā > E 1.12 ā > E 2.2 ĕ > AE
1.3 ā > I 1.13 ĕ > I 2.3 ĕ > AE
1.4 ā > I 1.14 ĕ > I 2.4 ae > I
1.5 ī > E 1.15 ī > E 2.5 ae > E
1.6 ī > E 1.16 ī > E 2.6 ĕ > AE
1.7 ē > V 1.17 ē > V 2.7 ĕ > AE
1.8 ē > V 1.18 ē > V 2.8 ae > I
1.9 ū > O 1.19 ū > O 2.9 au > O
1.10 ū > O 1.20 ū > O 2.10 o > AV
1.21 ū > I 2.11 au > A
1.22 i > V 2.12 oe > E

3.1 syncope
3.2 epenthesi
3.3 prosthesis
3.4 contractio ii
3.5 contractio uu
3.6 confusio i ~ e (ante voc.)
3.7 commutaciones vocalium variae
3.8 litterae perperam incisae (voc. aut cons.)
3.9 ie ~ e
3.10 ū ~ ō
3.11 litterae graecae (voc. aut cons.)
3.12 i (+voc) > ō
3.13 e (+voc) > ō
3.14 Y pro i
3.15 V pro y
4.1 p > B
4.2 t > D
4.3 k > G
4.4 desonorisatio
4.5 b > V
4.6 v > B
4.7 k > yod
4.8 g > yod
4.9 ti commut.
4.10 di commut.
4.11 ni commut.
4.12 L ~ R
4.13 H > ō
4.14 PH ~ F
4.15 c > K
5.1 -m > ō
5.2 -s > ō
5.3 -t > ō
5.4 -ns > S
5.5 -vu > V
5.6 x > S
5.7 x > SS
5.8 x > CS
5.9 x > XS
5.10 x > XX

Remarks:
1. Sign > is to be interpreted 'represented in the text as', e.g. "1.11 ā > E" is to be read: "a sound which should be a short unaccented a following usual norm is represented in the text by a letter E".
2. Sign ~ means "alternates with".
5.7 commut. variae consonantium
5.8 suppressio syllabae
5.9 syllaba perperam adiuncta (dittogr.)
5.10 cons. (+cons) > Ø

6.1 simplex > duplex L
6.2 simplex > duplex R
6.3 simplex > duplex M
6.4 simplex > duplex N
6.5 simplex > duplex S
6.6 simplex > duplex aliae
6.7 duplex > simplex L
6.8 duplex > simplex R
6.9 duplex > simplex M
6.10 duplex > simplex N
6.11 duplex > simplex S
6.12 duplex > simplex aliae

7.1 acc. pro nom.
7.2 nom. pro acc.
7.3 dat. pro gen.
7.4 gen. pro dat.
7.5 abl. pro acc.
7.6 acc. pro abl.
7.7 dat.-ē > E
7.8 abl.-ē > I
7.9 abl.-ē > E
7.10 nom.pl. -AS
7.11 fem. pro masc.
7.12 masc. pro fem.
7.13 neutr. pro masc.
7.14 masc. pro neutr.
7.15 fem. pro neutr.
7.16 neutr. pro fem.
7.17 decl. II pro IV
7.18 decl. I pro V
7.19 cett. ad decl. nom. et adiect. pertin.
7.20 -ABVS pro -is
7.21 -(I)BVS pro -is

8.1 praes pro fut.
8.2 -is pro -es (in 2. pers. verborum)
8.3 -es pro -is (in 2. pers. verborum)
8.4 -it pro -et (in 3. pers. verborum)
8.5 -et pro -it (in 3. pers. verborum)
8.6 cett. ad conjugationem pertin.

9.1 commut. in decl. pron. is
9.2 commut. in decl. pron. ille, iste
9.3 commut. in decl. pron. qui
9.4 commut. in decl. pron. hoc
9.5 cett. ad decl. pron. pertin.

10.1 praepositionum permutatio
10.2 praep. > casus sine praep.
10.3 casus sine praep. > praep.
10.4 varia ad congruentiam nomin. et adi. pertin.
10.5 quod, quia etc. pro acc. c. inf.
10.6 syntactica varia

11. lexica et lexico-semantica
11.1 recompensio
12. onomastica (nomina nec latina nec graeca; nomina graeca vel latina valde rara aut alter conspicua)

APPENDIX II

7. BUECHERM = F. Buecheler: Carmina Latina epigraphica. Lipsiae, 1895. /Supplementum: Lommatzsch, Lipsiae, 1926./
AFRICA

HISPANIA
15. IGALIC = Inscriptiones romanarum de Galicia. Santiago.

BRITANNIA

GALLIA

GERMANIA

RHAETIA and NORICUM


ITALIA

SARDINIA

GRAECIA

Dalmatia and Illyricum

PANNONIA

MOESIA

SCYTHIA
42. PIPPSCYTH = D.M. Pippidi: Inscriptiile din Scythia Minor grecesti si latine. Bucuresti.

ASIA

SYRIA

APPENDIX III

<table>
<thead>
<tr>
<th>LOC</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispania</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>Gallaecia</td>
<td>GAL</td>
<td></td>
</tr>
<tr>
<td>Cathagiensis</td>
<td>CAR</td>
<td></td>
</tr>
<tr>
<td>Tarraconensis</td>
<td>TAR</td>
<td></td>
</tr>
<tr>
<td>Lusitania</td>
<td>LUS</td>
<td></td>
</tr>
<tr>
<td>Baetica</td>
<td>BAE</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gallia</th>
<th>GL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aquitania</td>
<td>AQU</td>
</tr>
<tr>
<td>Lugdunensis</td>
<td>LUG</td>
</tr>
<tr>
<td>Belgica</td>
<td>BEL</td>
</tr>
<tr>
<td>Narbonensis</td>
<td>NAR</td>
</tr>
<tr>
<td>Britannia</td>
<td>BR</td>
</tr>
<tr>
<td>B.Superior</td>
<td>BRS</td>
</tr>
<tr>
<td>B.Inferior</td>
<td>BRI</td>
</tr>
<tr>
<td>Germania</td>
<td>GE</td>
</tr>
<tr>
<td>G.Superior</td>
<td>GES</td>
</tr>
<tr>
<td>G.Inferior</td>
<td>GEI</td>
</tr>
</tbody>
</table>

/Alpes/
| A.Maritimae | ALM |
| A.Cottiae  | ALC |
| A.Graiae   | ALG |

<table>
<thead>
<tr>
<th>Raetia</th>
<th>RA</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.Prima</td>
<td>RA1</td>
</tr>
<tr>
<td>R.Secunda</td>
<td>RA2</td>
</tr>
</tbody>
</table>

Italia
| I_Regio I | IT1 |
| I_Regio II | IT2 |
| etc.      |    |
| Sardinia  | SAR |
| Corsica   | COR |
| Sicilia   | SIC |

Noricum
| N.Ripense | NOR |
| N.Mediterraneum | NOM |

<p>| Pannonia | PA |
| P.Superior | PAS |
| P.Inferior | PAI |
| Dalmatia  | DL |
| Dacia     | DA |</p>
<table>
<thead>
<tr>
<th>Region</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>AF</td>
</tr>
<tr>
<td>Mauretania Tingitana</td>
<td>AFMT</td>
</tr>
<tr>
<td>Mauretania Caesariensis</td>
<td>AFMC</td>
</tr>
<tr>
<td>Numidia</td>
<td>AFN</td>
</tr>
<tr>
<td>A.Proconsularis</td>
<td>AFP</td>
</tr>
<tr>
<td>A.Byzacena</td>
<td>AFB</td>
</tr>
<tr>
<td>Tripolitania</td>
<td>AFT</td>
</tr>
<tr>
<td>Moesia</td>
<td>MO</td>
</tr>
<tr>
<td>M.Superior</td>
<td>MOS</td>
</tr>
<tr>
<td>M.Inferior</td>
<td>MOI</td>
</tr>
<tr>
<td>Agri Decumates</td>
<td>AD</td>
</tr>
<tr>
<td>Reliqua /Graecia et Oriens/</td>
<td>RE</td>
</tr>
</tbody>
</table>