The Leipzig Glossing Rules:  
Conventions for interlinear morpheme-by-morpheme glosses

About the rules

The Leipzig Glossing Rules have been developed jointly by the Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology (Bernard Comrie, Martin Haspelmath) and by the Department of Linguistics of the University of Leipzig (Balthasar Bickel). They consist of ten rules for the "syntax" and "semantics" of interlinear glosses, and an appendix with a proposed "lexicon" of abbreviated category labels. The rules cover a large part of linguists' needs in glossing texts, but most authors will feel the need to add (or modify) certain conventions (especially category labels). Still, it will be useful to have a standard set of conventions that linguists can refer to, and the Leipzig Rules are proposed as such to the community of linguists. The Rules are intended to reflect common usage, and only very few (mostly optional) innovations are proposed.

We intend to update the Leipzig Glossing Rules occasionally, so feedback is highly welcome.

Important references:


The rules  
(revised version of February 2008)

Preamble

Interlinear morpheme-by-morpheme glosses give information about the meanings and grammatical properties of individual words and parts of words. Linguists by and large conform to certain notational conventions in glossing, and the main purpose of this document is to make the most widely used conventions explicit.

Depending on the author's purposes and the readers' assumed background knowledge, different degrees of detail will be chosen. The current rules therefore allow some flexibility in various respects, and sometimes alternative options are mentioned.

The main purpose that is assumed here is the presentation of an example in a research paper or book. When an entire corpus is tagged, somewhat different
considerations may apply (e.g. one may want to add information about larger units such as words or phrases; the rules here only allow for information about morphemes).

It should also be noted that there are often multiple ways of analyzing the morphological patterns of a language. The glossing conventions do not help linguists in deciding between them, but merely provide standard ways of abbreviating possible descriptions. Moreover, glossing is rarely a complete morphological description, and it should be kept in mind that its purpose is not to state an analysis, but to give some further possibly relevant information on the structure of a text or an example, beyond the idiomatic translation.

A remark on the treatment of glosses in data cited from other sources: Glosses are part of the analysis, not part of the data. When citing an example from a published source, the gloss may be changed by the author if they prefer different terminology, a different style or a different analysis.

**Rule 1: Word-by-word alignment**

Interlinear glosses are left-aligned vertically, word by word, with the example. E.g.

(1) Indonesian (Sneddon 1996:237)

Mereka di Jakarta sekarang.

they in Jakarta now

'They are in Jakarta now.'

**Rule 2: Morpheme-by-morpheme correspondence**

Segmentable morphemes are separated by hyphens, both in the example and in the gloss. There must be exactly the same number of hyphens in the example and in the gloss. E.g.

(2) Lezgian (Haspelmath 1993:207)

Gila abur-u-n ferma hamıšaluğ gügüna amuq’-da-č.

now they-obl-gen farm forever behind stay-fut-NEG

'Now their farm will not stay behind forever.'

Since hyphens and vertical alignment make the text look unusual, authors may want to add another line at the beginning, containing the unmodified text, or resort to the option described in Rule 4 (and especially 4C).

Clitic boundaries are marked by an equals sign, both in the object language and in the gloss.

(3) West Greenlandic (Fortescue 1984:127)

palasi=lu niuïurtur=lu

priest=and shopkeeper=and

'both the priest and the shopkeeper'
Epenthetic segments occurring at a morpheme boundary should be assigned to either the preceding or the following morpheme. Which morpheme is to be chosen may be determined by various principles that are not easy to generalize over, so no rule will be provided for this.

Rule 2A. (Optional)
If morphologically bound elements constitute distinct prosodic or phonological words, a hyphen and a single space may be used together in the object language (but not in the gloss).

(4) Hakha Lai
  a-nii -láay
  3SG-laugh-FUT
  's/he will laugh'

Rule 3: Grammatical category labels
Grammatical morphemes are generally rendered by abbreviated grammatical category labels, printed in upper case letters (usually small capitals). A list of standard abbreviations (which are widely known among linguists) is given at the end of this document.

Deviations from these standard abbreviations may of course be necessary in particular cases, e.g. if a category is highly frequent in a language, so that a shorter abbreviation is more convenient, e.g. CPL (instead of COMPL) for "completive", PF (instead of PRF) for "perfect", etc. If a category is very rare, it may be simplest not to abbreviate its label at all.

In many cases, either a category label or a word from the metalanguage is acceptable. Thus, both of the two glosses of (5) may be chosen, depending on the purpose of the gloss.

(5) Russian
  \textit{My s Marko poexa-l-i avtobus-om v Peredelkino.}
  \begin{tabular}{llll}
  1PL & COM & Marko & go-PST-PL \hline
  we & with & Marko & go-PST-PL \hline
  \end{tabular}
  bus-INS \hspace{1cm} ALL \hspace{1cm} Peredelkino

  'Marko and I went to Perdelkino by bus.'

Rule 4: One-to-many correspondences

When a single object-language element is rendered by several metalanguage elements (words or abbreviations), these are separated by periods. E.g.

(6) Turkish
  \textit{çık-mak}
  come.out-INF
  'to come out'
(7) Latin
   *insul-arum*
   island-GEN.PL
   'of the islands'

(8) French
   *aux chevaux*
   to.ART.PL horse.PL
   'to the horses'

(9) German
   *unser-n Väter-n*
   our-DAT.PL father.PL-DAT.PL
   'to our fathers'

(10) Hittite (Lehmann 1982:211)
    \[ n=an \quad \text{apedani} \quad \text{mehuni} \quad \text{essandu}. \]
    \[ \text{CONN=him that.DAT.SG time.DAT.SG eat.they.shall} \]
    'They shall celebrate him on that date.' (\text{CONN} = \text{connective})

(11) Jaminjung (Schultze-Berndt 2000:92)
    \[ \text{nanggayan} \quad \text{guny-bi-yarluga?} \]
    who 2DU.A.3SG.P-FUT-poke
    'Who do you two want to spear?'

The ordering of the two metalanguage elements may be determined by various principles that are not easy to generalize over, so no rule will be provided for this. There are various reasons for a one-to-many correspondence between object-language elements and gloss elements. These are conflated by the uniform use of the period. If one wants to distinguish between them, one may follow Rules 4A–E.

**Rule 4A. (Optional)**
If an object-language element is neither formally nor semantically segmentable and only the metalanguage happens to lack a single-word equivalent, the underscore may be used instead of the period.

(12) Turkish
    (cf. 6)
    \[ \text{çıkmak} \]
    come_out-INF
    'to come out'

**Rule 4B. (Optional)**
If an object-language element is formally unsegmentable but has two or more clearly distinguishable meanings or grammatical properties, the semi-colon may be used. E.g.

(13) Latin
    (cf. 7)
    *insul-arum*
    island-GEN.PL
    'of the islands'
Rule 4C. (Optional)
If an object-language element is formally and semantically segmentable, but the author does not want to show the formal segmentation (because it is irrelevant and/or to keep the text intact), the colon may be used. E.g.

(15) Hittite (Lehmann 1982:211) (cf. 10)
\[ n=an \quad apedani \quad mehuni \quad essandu. \]
\[ \text{CONN=him that:DAT:SG time:DAT:SG eat:they:shall} \]
'They shall celebrate him on that date.'

Rule 4D. (Optional)
If a grammatical property in the object-language is signaled by a morphophonological change (ablaut, mutation, tone alternation, etc.), the backslash is used to separate the category label and the rest of the gloss.

(16) German (cf. 9)
\[ \text{unser-}n \quad \text{Väter-}n \]
\[ \text{our-DAT:PL father\PL-DAT:PL} \]
'to our fathers' (cf. singular Vater)

(17) Irish
\[ bhris-is \]
\[ \text{PST\break-2SG} \]
'you broke' (cf. nonpast bhris-)

(18) Kinyarwanda
\[ mú-kòrà \]
\[ \text{SBJV\1PL-work} \]
'that we work' (cf. indicative mú-kòrà)

Rule 4E. (Optional)
If a language has person-number affixes that express the agent-like and the patient-like argument of a transitive verb simultaneously, the symbol "->" may be used in the gloss to indicate that the first is the agent-like argument and the second is the patient-like argument.

(19) Jaminjung (Schultze-Berndt 2000:92) (cf. 11)
\[ \text{nanggayan} \quad \text{guny-bi-yarluga?} \]
\[ \text{who} \quad \text{2DU->3SG-FUT-poke} \]
'Who do you two want to spear?'
Rule 5: Person and number labels

Person and number are not separated by a period when they cooccur in this order. E.g.

(20) Italian
   and-iamo
   go-FRS.1PL (not: go-FRS.1.PL)
   'we go'

Rule 5A. (Optional)

Number and gender markers are very frequent in some languages, especially when combined with person. Several authors therefore use non-capitalized shortened abbreviations without a period. If this option is adopted, then the second gloss is used in (21).

(21) Belhare
   ne-e a-khim-chi n-yuNNa
   DEM-LOC 1SG.POSS-house-PL 3NS-be.NPST
   'Here are my houses.'

Rule 6: Non-overt elements

If the morpheme-by-morpheme gloss contains an element that does not correspond to an overt element in the example, it can be enclosed in square brackets. An obvious alternative is to include an overt "Ø" in the object-language text, which is separated by a hyphen like an overt element.

(22) Latin
   puer or: puer-Ø
   boy[NOM.SG] or: boy-NOM.SG
   'boy'
   'boy'

Rule 7: Inherent categories

Inherent, non-overt categories such as gender may be indicated in the gloss, but a special boundary symbol, the round parenthesis, is used. E.g.

(23) Hunzib (van den Berg 1995:46)
   oz#-di-g xõxe m-uq'e-r
   boy-BNL-AD tree(G4) G4-bend-PRET
   'Because of the boy the tree bent.'
   (G4 = 4th gender, AD = adessive, PRET = preterite)

Rule 8: Bipartite elements
Grammatical or lexical elements that consist of two parts which are treated as distinct morphological entities (e.g. bipartite stems such as Lakhota na-xʔu 'hear') may be treated in two different ways:

(i) The gloss may simply be repeated:

(24) Lakhota
    na-wičha-wa-xʔu
    hear-3PL.UND-1SG.ACT-hear
    'I hear them' (UND = undergoer, ACT = actor)

(ii) One of the two parts may be represented by a special label such as STEM:

(25) Lakhota
    na-wičha-wa-xʔu
    hear-3PL.UND-1SG.ACT-STEM
    'I hear them'

Circumfixes are "bipartite affixes" and can be treated in the same way, e.g.

(26) German
    ge-seh-en or: ge-seh-en
    PTCP-see-PTCP PTCP-see-CIRC
    'seen' 'seen'

Rule 9: Infixes

Infixes are enclosed by angle brackets, and so is the object-language counterpart in the gloss.

(27) Tagalog
    b<um>ili
    <ACTFOC>buy
    'buy'

(28) Latin
    reli<n>qu-ere
    leave<PRS>INF
    'to leave'

Infixes are generally easily identifiable as left-peripheral (as in 27) or as right-peripheral (as in 28), and this determines the position of the gloss corresponding to the infix with respect to the gloss of the stem. If the infix is not clearly peripheral, some other basis for linearizing the gloss has to be found.

Rule 10: Reduplication
Reduplication is treated similarly to affixation, but with a tilde (instead of an ordinary hyphen) connecting the copied element to the stem.

(29) Hebrew

\textit{yerak-rak-im}

\textit{green-ATT-M.PL}

'greenish ones' \hfill (\textit{ATT} = attenuative)

(30) Tagalog

\textit{bi-bili}

\textit{IPFV-buy}

'is buying'

(31) Tagalog

\textit{b\textsuperscript{-um}i-bili}

\textit{<ACTFOC>IPFV-buy}

'is buying' \hfill (\textit{ACTFOC} = Actor focus)

Appendix: List of Standard Abbreviations

1  first person
2  second person
3  third person
A  agent-like argument of canonical transitive verb
ABL  ablative
ABS  absolutive
ACC  accusative
ADJ  adjective
ADV  adverb(ial)
AGR  agreement
ALL  allative
ANTIP  antipassive
APPL  applicative
ART  article
AUX  auxiliary
BEN  benefactive
CAUS  causative
CLF  classifier
COM  comitative
COMP  complementizer
COMPL  completive
COND  conditional
COP  copula
CVB  converb
DAT  dative
DECL  declarative
DEF  definite
DEMONSTRATIVE
DETERMINER
DISTAL
DISTRIBUTIVE
DUAL
DURATIVE
ERGATIVE
EXCLUSIVE
FEMININE
FOCUS
FUTURE
GENITIVE
IMPERATIVE
INCLUSIVE
INDICATIVE
INDEFINITIVE
INFINITIVE
INSTRUMENTAL
INTRANSITIVE
IMPERFECTIVE
IRRREALIS
LOCATIVE
MASCULINE
NEUTER
NON- (E.G. NNS NONSINGULAR, NPST NONPAST)
NEGATION, NEGATIVE
NOMINALIZER/NOMINALIZATION
NOMINATIVE
OBJECT
OBLIQUE
PATIENT-LIKE ARGUMENT OF CANONICAL TRANSITIVE VERB
PASSIVE
PERFECTIVE
PLURAL
POSSESSIVE
PREDICATIVE
PERFECT
PRESENT
PROGRESSIVE
PROHIBITIVE
PROXIMAL/PROXIMATE
PAST
PARTICIPLE
PURPOSIVE
QUESTION PARTICLE/MARKER
QUOTATIVE
RECIPIROCAL
REFLEXIVE
REL  relative
RES  resultative
S    single argument of canonical intransitive verb
SBJ  subject
SBJV subjunctive
SG   singular
TOP  topic
TR   transitive
VOC  vocative

References