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:

Lehmann, Christian. 1983. "Directions for interlinear morphemic translations". *Folia Linguistica* 16: 193-224.

Croft, William. 2003. *Typology and universals.* 2nd ed. Cambridge: Cambridge University Press, pp. xix-xv.

The rules
(slightly revised version of January 2004)

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.

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   hamišaluǧ güǧüna   amuqʼ-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 5 (and especially 5C).
    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   niuirtur=lu
priest=and   shopkeeper=and
'both the priest and the shopkeeper'

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).

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 (4) may be chosen, depending on the purpose of the gloss.

(4) Russian
My   s   Marko   poexa-l-i   avtobus-om   v   Peredelkino.
1PL   COM   Marko   go-PST-PL   bus-INSTR   ALL   Peredelkino
we   with   Marko   go-PST-PL   bus-by   to   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.

(5) Turkish
çık-mak
come.out-INF
'to come out'

(6) Latin
insul-arum
island-GEN.PL
'of the islands'

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

(8) Hittite (Lehmann 1983:211)
n=an   apedani   mehuni   essandu.
CONN=him   that.DAT.SG   time.DAT.SG   eat.they.shall
'They shall celebrate him on that date.'

(9) Jaminjung (Schultze-Berndt 2000:92)
nanggayan   guny-bi-yarluga?
who   2DU.A.3SG.P-FUT-poke
'Who do you two want to spear?'

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-D.

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.

(10) Turkish   (cf. 5)
 çık-mak
 come_out-INF
 'to come out'

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

(11) Latin   (cf. 6)
 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.

(12) Hittite (Lehmann 1983:211) (cf. 8)
 n=an apedani mehuni essandu.
 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 of the stem (ablaut, mutation, etc.), the backslash is used to separate the category label and the stem gloss.

(13) German   (cf. 7)
 unser-n Väter-n
 our-DAT.PL father\PL-DAT.PL
 'to our fathers'

(14) Irish
 bhris-is
 PST\break-2SG
 'you broke'

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.

(15) Jaminjung   (cf. 9)
 nanggayan guny-bi-yarluga?
 who   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.

(16) Italian
　　and-iamo
　　go-PRES.1PL　　　　　　(not: *go*-PRES.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 single-letter abbreviations without a period. If this option is adopted, then the second gloss is used in (17).

(17) Belhare
　　ne-e　　　*a-khim-chi*　　　*n-yuŋŋa*
　　DEM-LOC　1SG.POSS-house-PL　3SG.N-be.NPST
　　DEM-LOC　1s.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.

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

Rule 7: Bipartite elements

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

(i) The gloss may simply be repeated:

(19)　Lakhota
　　　na-wíčha-wa-xʔų
　　　hear-3PL.UND-1SG.ACT-hear
　　　'I hear them'

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

(20) Lakhota
      na-wíčha-wa-xʔų̄
      hear-3PL.UND-1SG.ACT- STEM
      'I hear them'

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

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

Rule 8: Infixes

Infixes are separated by angle brackets, both in the object-language text and in the gloss.

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

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

Infixes are generally easily identifiable as left-peripheral (as in 22) or as right-peripheral (as in 23), and this determines the position of the gloss corresponding to the infix with respect to the gloss of the stem.

Rule 9: 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.

(24) Hunzib (van den Berg 1995:46)
     ož-di-g   xõxe   m-uqʼe-r
     boy-OBL-AD   tree(G4)   G4-bend-PRET
     'Because of the boy the tree bent.'

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.

(25) Hebrew
    yerak~rak-im
     green~ATT-M.PL
    'greenish ones'   (ATT = attenuative)

(26) Tagalog
    bi~bili
     IPFV~buy
    'is buying'

(27) Tagalog
    b<um>i~bili
     <ACTFOC>IPFV~buy
    'is buying'

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
DEM	demonstrative
DET	determiner
DIST	distal
DISTR	distributive
DU	dual
DUR	durative

ERG	ergative
EXCL	exclusive
F	feminine
FOC	focus
FUT	future
GEN	genitive
IMP	imperative
INCL	inclusive
IND	indicative
INDF	indefinite
INF	infinitive
INS	instrumental
INTR	intransitive
IPFV	imperfective
IRR	irrealis
LOC	locative
M	masculine
N	neuter
N-	non- (e.g. NSG nonsingular, NPST nonpast)
NEG	negation, negative
NMLZ	nominalizer/nominalization
NOM	nominative
OBJ	object
OBL	oblique
P	patient-like argument of canonical transitive verb
PASS	passive
PFV	perfective
PL	plural
POSS	possessive
PRED	predicative
PRF	perfect
PRS	present
PROG	progressive
PROH	prohibitive
PROX	proximal/proximate
PST	past
PTCP	participle
PURP	purposive
Q	question particle/marker
QUOT	quotative
RECP	reciprocal
REFL	reflexive
REL	relative
RES	resultative
S	single argument of canonical intransitive verb
SBJ	subject
SBJV	subjunctive
SG	singular
TOP	topic
TR	transitive

References

Fortescue, Michael. 1984. *West Greenlandic.* (Croom Helm descriptive grammars) London: Croom Helm.

Haspelmath, Martin. 1993. *A grammar of Lezgian.* (Mouton Grammar Library, 9). Berlin - New York: Mouton de Gruyter.

Lehmann, Christian. 1983. "Directions for interlinear morphemic translations". *Folia Linguistica* 16: 193-224.

Schultze-Berndt, Eva. 2000. *Simple and complex verbs in Jaminjung: A study of event categorization in an Australian language.* Katholieke Universiteit Nijmegen Ph.D. Dissertation.

Sneddon, James Neil. 1996. *Indonesian*: *A comprehensive grammar.* London: Routledge.

van den Berg, Helma. 1995. *A Grammar of Hunzib.* (Lincom Studies in Caucasian Linguistics, 1.) München: Lincom Europa.