

CLIPP

Christiani Lehmanni inedita, publicanda, publicata

titulus	Interlinear morphemic glossing
huius textus situs retis mundialis	http://www.christianlehmann.eu/publ/img.pdf
dies manuscripti postremum modificati	06.10.2011
ocasio orationis habitae	—
volumen publicationem continens	Booij, Geert & Lehmann, Christian & Mugdan, Joachim & Skopeteas, Stavros (eds.), <i>Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung. 2. Halbband.</i> Berlin: W. de Gruyter (Handbücher der Sprach- und Kommunikationswissenschaft, 17.2)
annus publicationis	2004
paginae	1834-1857

comparable in this with a phonetic transcription. Just as the latter has been successfully standardized by the IPA, so interlinear morphemic glossing should be standardized.

This will be done in the present article in the form of a set of rules, which are listed in section 6.1. Such a standardization only concerns linguistic science. Linguistic data are often presented to a lay public, with the purpose of education, entertainment or divulgation of the achievements of our science. Here some kind of interlinear glossing may be necessary, too. However, scientific formalism tends to damage rather than serve the good cause. An example how interlinear glossing has been handled in a book directed to a non-specialist public is quoted in the next section (Finck 1909). The present article is biased in favor of a more formalized treatment, on the assumption that it will be easier to derive a less formal representation from the proposals made here than the other way round. The treatment is, however, not fully formal, since it focuses on interlinear glossing in printed texts. In the annotation of texts by markup languages for automatic retrieval, the same conceptual problems, but very different technical problems arise which will not be dealt with here.

Data are commonly quoted from sources in which they are already provided by an analysis. In linguistic publications, it has been wide-spread usage to quote data together with their IMG and their translation, even if their form or language is different from the one used in the quoting context. That is, such composite data representations have been treated as indecomposable blocks. Such scruples do not seem to be warranted. Primary data may be quoted and provided with the quoting author's analysis and translation (cf. Bickel et al. 2004:1).

1.2. Precursors

Interlinear glossing has precursors in the descriptive tradition which link it up not with some kind of morphological representation, but with efforts to bring out the spirit of the language. The point there is not to provide a formal representation of a piece of linguistic data, but to render the language-specific construal of the world intelligible. To this end, literal translations were provided. For instance, G. Gabelentz (1901:460), in a passage arguing that the personal verb suffixes in Semitic languages are possessive pronouns, gives the following Arabic example: “*ya-kfī-ka-hūm* er genügt dir gegen sie (eig. er-genügt-dein-ihr)”.

The IMG is a late-comer in linguistics. Early grammars were intended as primers, the user was expected to work through them and learn the morphemes; so no glossing was necessary. Many scientific grammars, e.g. of Latin, Greek, Arabic etc., were meant for the initiated who needed no glossing either (not seldom even the free translations were spared). Even comparative studies, historical or typological, left the analysis of the examples of diverse languages to the reader. H. C. Gabelentz, in the middle of a discussion of Lule, Osage and other languages, presents the following passage:

"Im Dakota (meine Grammatik der Dakota-Sprache § 34) dient die 3 Pers. Plur. Act. dazu, das Passivum auszudrücken, sogar wenn ein Actor im Singularis hinzuzudenken ist, z.B. *Jesus Jan eñ hi q ix Jordan watpa ohna baptizapi*, Jesus kam zu Johannes und sie taufte ihn (st. er wurde getauft) im Jordanfluss." (H. C. Gabelentz 1861:465)

Here the reader who does not have the grammar mentioned on his desk is given no chance.

Pace Gabelentz, IMGs are needed when two conditions coincide: the level of analysis is above morphology, and the reader is not expected to be familiar with the languages under discussion (which is generally the case in typology, but not in descriptive or historical-comparative linguistics). W.v. Humboldt (1836[1963]:534) invented his own device to help the reader identify L2 meaningful elements with L1 morphemes. He gives the following example from Classical Nahuatl:

1 2 3 4 5 6 7 8 9 1 3 2 4 5 6 7 8 9
ni- c- chihui -lia in no- piltzin ce calli ich mache es für der mein Sohn ein Haus

While dispensing with the IMG proper, this method fails for L1 elements which cannot be rendered by L2 words.

Beside the literal translation illustrated above, G. Gabelentz (1901) uses a variety of techniques. He also has interlinear glosses, as when he says: ‘Der Satz “Ich bin Dein Sohn” heißt im Maya:

a – mexen – en.
Dein Sohn ich,’ (Gabelentz 1901:383)

and occasionally (e.g. Gabelentz 1901:400) he uses Latin as L2 in IMGs.

Finck (1909) is one of the first linguistic publications that illustrate the working of a language with a sizable text provided with a free translation and an IMG. The following sentence from his Turkish text (Finck 1909:83) illustrates his glossing style:

<i>xodža-da</i>	<i>esbāb-in</i>	<i>dzümle-si-ni</i>	Der Meister warf nun
Meister=auch	Kleider=(der)	Gesamtheit=ihre=die	sämtliche Kleider ins Feuer
<i>ateš-e</i>	<i>vur-up</i>	<i>yak-ar</i>	und verbrannte sie.
Feuer=zü	werf=enderweise	verbrenn=end	

As may be seen, these forerunners have no grammatical category labels yet. Finck glosses Turkish *-in* ‘GEN’ by Germ. *der* because this word displays a morphological trace of the genitive. Similarly, Turkish *-up* ‘GER’ is glossed by *-enderweise*, maybe the closest to a gerund that German can muster. This procedure is a tribute to the non-specialist readership that the booklet aims at, but necessarily falsifies the working of the language by attributing lexical meanings to its grammatical morphemes.

It took a long time until interlinear morphemic glossing became firmly established. In Bloomfield’s *Language*, of 1933, examples abound, but they are presented like this:

“Some languages have here one word, regardless of gender, as Tagalog [kapa’tid]; our *brother* corresponds to a Tagalog phrase [kapa’tid na la’la:ki], where the last word means ‘male’, and our *sister* to [kapa’tid na ba’ba:ji], with the attribute ‘female’” (Bloomfield 1933:278).

IMGs that fulfill most of the requirements set out below appear first in the sixties of the 20th century. From the eighties on, they become standard in publications dealing with languages whose knowledge is not presupposed. Editors and publishers increasingly require them even for languages like Latin, French and German that used to be well-known to linguists. The development is towards (not only providing translations for, but even) glossing every language except English. This is apparently a symptom of a global development in

which every language except English becomes exotic.

Good IMGs are relatively costly, both for the scientist and for the typesetter. Authors and publishers are therefore not too eager to produce them (well). There is at least one software on the market that aids the linguist in generating systematic IMGs for his texts, the interlinearizer that comes with the program Shoebox, from the Summer Institute of Linguistics (cf. Simons & Versaw 1988; Art. 168).

Since IMGs are fairly recent in linguistics, they have seldom been treated by linguistic methodology. The first treatise of the present subject is Lehmann (1982). Subsequent work includes Simons & Versaw (1988), Lehmann et al. (1994), Lieb & Drude (2000), Bickel et al. (2004). They have been freely made use of in the present treatment.

1.3. Levels of representation

Interlinear morphemic glossing must be seen in the larger context of representation of linguistic data and, even more comprehensive, of the documentation of a language (cf. Lieb & Drude 2000). On such a background, an isolated example given in a descriptive context is a particularly constrained case of the edition and annotation (also called ‘markup’ for technical purposes) of a piece of primary linguistic data for posterity. In other words, a general-purpose edition of a linguistic corpus is a kind of maximum model, subject to the full set of rules for explicitness, detail and elaboration, from which the quotation of an isolated example in the context of some grammatical discussion represents a subset delimited by considerations of feasibility, usefulness and the like.

Every linguistic representation of some piece of raw data, even if it limits itself to a phonetic transcription, involves some linguistic analysis (Lehmann 2004). Insofar, no sharp boundary is to be drawn between the sheer representation of data and their analysis. Bearing this in mind, we can speak of various levels at which linguistic data may be represented. Presupposing spoken language data, at least the following are relevant:

- (a) raw data recording (video or audio tape),
- (b) phonetic transcription of the utterance,
- (c) orthographic representation of the utterance,
- (d) morph(ophon)emic representation of the utterance,
- (e) IMG of the utterance,
- (f) free translation of the utterance into the background language,
- (g) descriptive and explanatory comment on pragmatic or cultural aspects of the utterance.

This set may be supplemented by even more representations (cf. Lieb & Drude 2000). There may be a phonological representation distinct from both levels (b) and (d). There may be a syntactic representation, e.g. in the form of a labeled bracketing. And there may be a semantic representation instead of, or in addition to, representation (f). In such representations, the portion of linguistic analysis is probably even stronger than in the seven levels enumerated.

The raw data have a temporal structure which is projected onto a spatial line in written representations. These representations are synchronized more or less closely. For instance, representation (f) generally matches L1 sentences, units of level (g) may be associated with L1 units of any size, and representation (e) may match representation (d) morpheme by morpheme. This has different consequences for the typographic layout. For instance, units of

level (g) may be associated with the running text by making full use of a multidimensional display, while representation (f) may be in a lateral column at the same height as its original, as is usual in synoptic editions and also practiced in the example from Finck (1909) given in section 1.2. Other representations should be arranged in lines one of which is beneath the other and runs in parallel with it.

For the purposes of descriptive and typological grammatical analysis and exemplification, the seven-level set is generally reduced to only three. What may be called the ‘canonical trilinear representation’ of linguistic examples involves:

- a representation of L1 at one of the levels (b), (c) or (d),
- an IMG in L2 (level e),
- an idiomatic translation into L2 (level f).

An IMG will seldom be paired with a phonetic representation, because this serves phonetics, while an IMG serves grammar. They therefore form an unequal pair. If both are required, they will generally be mediated by another representation, a morphophonemic or orthographic one.

It makes a difference for the glossing whether L1 is rendered in a morphophonemic representation or in conventional orthography. In the former case, the rules of orthography do not apply, and the linguist may dress up the representation in such a way that a biunique mapping onto the IMG is facilitated. In the latter case, morpheme boundaries may be obscured by the orthography, and there will be delimiters such as blanks, hyphens and punctuation marks which do not necessarily represent grammatical boundaries and may interfere with the glossing. However, the choice between an orthographic and a scientific representation of a text is generally a higher-order choice which cannot depend on glossing requirements. In particular, an example may be quoted unchanged from a primary source (think of Sanskrit examples). It may then not be possible to insert boundary symbols and the like in the L1 text. Glossing conventions therefore have to be adjusted to use with orthographic representations.

If the first line representing the L1 text differs too much from a morphophonemic representation, then it is advisable to expand the canonical trilinear representation by an additional morphophonemic representation. It will then be this line that the IMG refers to.

The two languages involved will be called L1 and L2 throughout. However, it should be clear that the relationship between them is asymmetric: L1 is the object language, L2 is the metalanguage. The symbols occurring in an IMG have a different status from the elements of the text line that they gloss: For present purposes, the L1 text line consists of morphs, while the IMG consists of names of L2 morphemes and of grammatical categories (cf. section 3.2). There can, thus, be no question of “mirroring” the structure of the L1 expression by the sequence of the L2 elements. Instead, an element in an IMG serves as a kind of mnemonic hint to the meaning or function of its corresponding L1 element.

1.4. Delimitation

The complete set of representations rendering an L1 text may be sufficient to derive a grammatical description from it (as postulated in Lieb & Drude 2000, §1.1). However, given its inherent restrictions, an IMG cannot by itself compensate for a grammar (or just a morphology). Apart from the form of presentation, the most important substantive difference between a grammatical description and an IMG lies in the fact that the grammar treats of

categories in the sense of classes, while the IMG identifies individual morphemes. For instance, a grammar treats of the verbal category of aspect. An IMG contains a gloss for an individual aspect morpheme, e.g. PERF, neglecting the question of whether this is actually an aspect morpheme or rather a tense morpheme, and also leaving unanswered questions concerning other members of the paradigm, let alone the construction and use of the PERF morpheme. Some of these kinds of information may be given in other representations, e.g. a syntactic representation.

By the same token, the IMG does not indicate the syntactic category of a word form. For instance, the IMG of Germ. *laufend* is ‘run:PART.PRS’, showing that the form contains a morpheme whose function it is to mark a present participle. The gloss is not ‘run(part.prs)’ or anything of the sort, meaning that *laufend* is a present participle. While the latter is true, it is not the task of an IMG to give this information.

Moreover, the type of morphological unit is not an object of an IMG. Thus, concepts like ‘stem’, ‘root’, ‘prefix’ do not appear in IMGs. Such information may, to a large extent, be inferred from a proper IMG, since the gloss of a root differs typographically from the gloss of a grammatical formative.

Similarly, an IMG cannot replace a lexicon. Here again, elements appearing in an IMG are but names of elements appearing in the L1 line. They are not meant to exhaust the meaning of such an element.

Finally, an IMG is not meant to replace an idiomatic translation. Thus, it cannot and should not render closely the sense of an L1 item in the given context. An IMG is regularly accompanied by a free translation which fulfills precisely this purpose.

2. Prerequisites of morphological analysis

Interlinear glossing might appear to be just an elementary form of representing data. As a matter of fact, it presupposes a morphological analysis. The following analytic problems are directly reflected by the glosses.

2.1. Unmarkedness and zero morphemes

Where the L1 text contains a morph, the IMG contains an element rendering it. Where the L1 text contains nothing, the issue of rendering it is complicated by markedness theory. Germ. *Herr* may be glossed by ‘master’ or by ‘master(NOM.SG)’. Latin *mone-t* may be glossed by ‘warn-3.SG’ or by ‘warn(IND.ACT)-3.SG’ (according to R16). Moreover, one may believe that such forms contain zero morphemes and put thus: *Herr-Ø* ‘master-NOM.SG’, *mone-Ø-Ø-t* ‘warn-IND-ACT-3.SG’. All of these IMGs are formally correct. The choice among them is not a matter of appropriate glossing, but of morphological theory. For interlinear glossing, only the general rule R1 is relevant.

2.2. Allomorphy

If the L1 representation to be glossed corresponds to standard orthography, the analyst has no decisions to make in its regard. Otherwise, a good option for the representation (as well as for

any writing system) is a morphophonemic representation which steers a middle course as far as allomorphy is concerned: Phonologically conditioned allomorphy is resolved (ignored), morphologically conditioned allomorphy is not resolved (is rendered).

The IMG, on the other hand, shows morphemes, not allomorphs. In order to understand what this implies, consider three examples. Modern Yucatec Maya expresses completive and incompletive aspect by suffixes on transitive and (one conjugation class of) intransitive verbs as follows:

aspect	completive	incompletive
valence		
transitive	-ah	-ik
intransitive	-Ø	-Vl

Tab. 169.1: Aspectual suffixes in Yucatec Maya

For instance, *t-u hats'-ah* 'PAST-SBJ.3 beat-CMPL (he beat it)'. One might think that the table contains four morphemes. Actually, however, transitivity is inherent in the verb stem and conditions allomorphy in the aspect suffix. The conditioning factor should not make part of the gloss. That is, the correct gloss for *-ah* is not 'TR.CMPL', but simply 'CMPL'. See also 4.5.

Yucatec Maya also has personal clitics that precede nouns as possessive cross-reference markers and verbs as subject cross-reference markers. If the noun or verb starts with a vowel, a glide is inserted in its front. The choice between the two glides *w* and *y* is morphologically conditioned: If the pronoun is of first person singular or of second person, it is *w*; if the pronoun is of third person, the glide is *y*. For instance, *in watan* 'POSS.1.SG Ø:wife (my wife)', *u yatan* 'POSS.3.SG Ø:wife (his wife)'. It is also possible to regard the noun forms modified by the initial glide as stem allomorphs, in which case the glide would not even receive the gloss by 'Ø'. However, in the third person, a pronominal clitic followed by the glide can be omitted. Thus, *yatan* by itself means 'his wife'. (Historically, the glide is indeed a reflex of an older cross-reference marker). We therefore have *u y-atan* 'POSS.3 Ø-wife' ~ *y-atan* 'POSS.3-wife', and we face the problem that the same element is not even a morph in one context, but a full-fledged morpheme in another. Whatever the correct morphological analysis may be, the IMG presupposes it and brings it out.

Last, consider gender marking in a language such as Latin (cf. Art. 48). *Puellae bonae* means 'good girls', *pueri boni* 'good boys'. Apart from motion, gender is inherent in a noun stem. It is, however, recognizable by the declension suffixes. Nevertheless, the gloss of the morph in question does not contain the conditioning category. The noun forms will be glossed 'girl.F:NOM.PL', 'boy.M:NOM.PL', implying that gender is a category of the stem, not of the suffix. What about the adjectives? Gender is not inherent in an adjective stem. We may therefore gloss them by 'good:NOM.PL.F' and 'good:NOM.PL.M'. Then one and the same element would be a morpheme on adjectives, but a conditioned allomorph on nouns, and therefore it would get two different glosses. Since two different glosses for the same element are not admissible in interlinear glossing (R4), this would entail that there are two homonymous declension suffixes *-ae* in Latin, which is obviously undesirable. We may stop this consideration here, since the problem is obviously not one of glossing, but one of morphological analysis. R2 codifies the convention that IMG expressions represent morphemes, not allomorphs.

3. Principles of interlinear glossing

3.1. General

In the canonical trilinear representation, one L1 text line is matched by two L2 lines, the IMG and the free translation. This entails a division of labor between the two L2 representations. The free translation is the idiomatic semantic equivalent of the L1 line, the IMG is a representation of its morphological structure. There is consequently no need for the translation to be particularly literal, just as there is no need for the IMG to repeat the morphemes that appear in the translation. For instance, a polysemous L1 item will be rendered by its contextual sense in the free translation, but by its basic meaning in the IMG (R8). Unnecessary parallelism between the two L2 lines is redundant; the trilinear canonical representation offers an occasion to provide additional information.

In principle, the degree of detail displayed in an IMG depends on the purpose the example with its gloss is meant to serve. However, the author cannot foresee the purposes to which others will want to use his examples. A morphological detail that is not at stake in the current discussion may be essential for the argument another linguist may wish to base on the example. For this reason, the principle is to allow for as much precision and detail as seems tolerable (R3). The following rules specify the properties of a complete IMG. They do not exclude less detailed IMGs where they suffice. Cf. R13 and R23 for possibilities of under-specifying morphological structure.

The IMG of a morpheme is some sort of name for it, a name that alludes to its meaning or function and is insofar mnemonic or, at least, more helpful to the non-specialist than the L1 morph itself. It must therefore have a certain recognition value. R4, which actually is a tightening of R1, therefore requires that given a particular L1 morpheme, its gloss will be the same in all contexts; and apart from full synonymy, no two morphemes of L1 will have the same gloss. These points will be elaborated in the following subsections.

3.2. Glossing vocabulary

Glosses are taken from a language L2 that serves as a metalanguage of L1. L2 is based on a natural language – in this article, English –, but with far-reaching deviations from natural language use. The glossing vocabulary consists of the following kinds of symbols:

- vocables:
 - L2 morphemes and stems
 - grammatical category labels
- boundary symbols.

The difference between the two kinds of vocables is the following: Morphemes and stems are taken from natural L2 vocabulary and are meant to be translation equivalents (in a sense to be made precise below) of L1 items. For instance, the notation “Germ. *Schreib-tisch* ‘write-table (desk)’” is to be interpreted thus: The German word form *Schreibtisch* ‘desk’ consists of two morphs, of which *schreib-* means ‘write’ and *tisch* means ‘table’. Grammatical category labels, on the other hand, are taken from scientific terminology and are meant to categorize the function of L1 items. For instance, “Germ. *schreib-en* ‘write-INF (write (inf.))’” is to be interpreted thus: The German word form *schreiben* ‘write (inf.)’ consists of two morphs, of

which *schreib-* means ‘write’, while *-en* is an infinitive marker (that is, *-en* does not mean ‘infinitive’; it is the German word *Infinitiv* which means ‘infinitive’). To bring out this essential difference between the two kinds of IMG vocables, L2 morphemes and stems are written in straight orthography, while grammatical category labels are written in (small) capitals (R29).

A grammatical category label represents (i.e. is the name of) the value of a grammatical category (the latter being taken, technically, as a parameter or attribute). For instance, the label ‘ACC’ is the name of the value ‘accusative’ of the morphological category ‘case’. Just as a grammatical category label is a name of a value of a grammatical category, what is called ‘L2 morphemes and stems’ are actually names of L2 morphemes and stems. In the following, we will abide by the simpler way of speaking. The choice and use of vocables are treated in the following subsections; boundary symbols are treated in section 4.

3.3. Lexemes

An L1 lexeme is, in principle, glossed by an L2 lexeme (R5(a)). Sometimes more than one L2 word is necessary, for instance in Germ. *fabulieren* ‘invent.stories’. However, profusion is to be avoided. Adjectives that do not require a copula in predicative function are often glossed by adding a copula, e.g. West Greenlandic *anurli* ‘windy’ is glossed as ‘be.windy’ in Fortescue (1984:65). This is only correct if a word of this class requires an attributor in attributive function. Otherwise it wrongly implies that there is no difference between adjectives and verbs, and it tends to obscure the fact that the language does not use a copula with adjectival predicates.

L1 cardinal numerals are glossed by Arabic numbers. An issue arises for proper names, which are often not glossed at all. However, there is no room here for an exception to the general rule: a proper name is rendered by its counterpart in L2. Some proper names have conventional counterparts that are specific to L2; Engl. *John* corresponds to Germ. *Hans*, and Engl. *Munich* corresponds to Germ. *München*. These then appear in the IMG. Whenever there is no such language-specific convention, the counterpart of an L1 name is usually the same word in L2.

If L2 is English, no problem arises for the form in which L2 lexemes are quoted in the IMG. In other languages, lexemes have a citation form in conformity with L2 conventions. If this is an inflected form, like the nominative for nouns or the infinitive for verbs, then it is excluded from an IMG by R5(b), and instead the bare stem must be used. The reason is that such a gloss would seem to imply that there is a nominative, or an infinitive, in the L1 line where actually just a stem is being glossed.

3.4. Grammatical formatives

L1 morphs are, in principle, glossed by citation forms of L2 morphemes. However, interlinear morphemic glossing crucially revolves around grammatical properties of L1 items. These will differ between L1 and L2. Even if, in a number of cases, the L2 stem appearing in a gloss has the same grammatical properties as the L1 morph that it represents, this cannot be expected and therefore not be relied upon. For instance, Latin *eum* could be glossed by Engl. *him*, and at the typological level, they do share a number of features. However, *eum* is accusative and

can thus not be indirect object, while *him* is the form for direct and indirect object. Therefore, grammatical items of L1 are generally not glossed by grammatical items of L2, but by a configuration of symbols taken from the scientific metalanguage and representing their grammatical features, i.e. by grammatical category labels (R6). Thus, Latin *eum* may be glossed by ‘ANA:ACC.SG.M’.

No bound grammatical or derivational morphemes should appear in IMGs. Free grammatical morphemes may be used to render free grammatical morphemes. However, use of those in the second column of Tab. 169.2 is discouraged unless L1 happens to exhibit the same ambiguity as English:

word class	instead of	use
copulas, auxiliaries	<i>be</i>	COP, PASS, PROG ...
	<i>have</i> (except to mean ‘possess, own’)	PF, OBLG ...
prepositions	<i>by</i>	AG, ERG ...
	<i>with</i>	INST, COM, ASSOC ...
	<i>for</i>	BEN, DEST ...
	<i>as</i>	EQT, ESS ...
	<i>from</i>	ABL, DEL ...
	<i>to</i>	DAT, ALL, DEST, TERM, INF ...
	<i>of</i>	GEN, ASSOC ...
subordinators	<i>that</i>	COMP, SR (, D3)
	<i>if</i>	INT, COND.SR
relativizers	<i>that</i>	REL
	<i>who</i>	REL.HUM.NOM ...
	<i>which</i>	REL.NHUM.NOM ...

Tab. 169.2: Free grammatical morphemes

Some morphemes are extremely deeply entrenched in the semantic or pragmatic system of the language and simply have no translation equivalent in L2. Two common ways out are a) to repeat the significans of the item in the gloss, and b) to indicate the class of the item instead of its meaning. Thus, we find the German modal particle *eben* glossed either as ‘EBEN’ or as ‘PTL’. Both glosses are inadequate. If there is no translation equivalent in natural L2, then the linguist has a specialized metalanguage to describe such functions. For the sake of an IMG that is not devoted to modal particles in particular, a gloss like ‘REAFF’ (reaffirmed) will be fully sufficient and more helpful than either of the aforementioned.

A gloss is a proper name of an L1 morpheme. It does not give information on the grammatical class of the morpheme in question other than what is implied by the name itself. If a gloss is ‘ACC’, one assumes that the morpheme belongs to the grammatical class of the case morphemes. It is the task of the grammar to clarify whether or not this implication is correct in a particular case. The gloss will not be ‘CASE.ACC’ or anything of this sort. For the same reason, the gloss of the perfective aspect is simply ‘PFV’ and not ‘PFV.ASP’, and so on.

From this it follows that the gloss will not be ‘ASP’ either. In the literature, one frequently encounters glosses such as ‘PTCL’ (particle), ‘AGR’ (agreement), ‘ART’ (article). If L1 possesses only one particle, agreement morpheme (hardly imaginable) or article (this is possible), then these glosses are sufficient. In all other cases, this kind of gloss is not helpful because it does not give the information on the meaning or function of the morpheme that a

gloss is supposed to give. Moreover, the whole glossing becomes inconsistent, as some glosses name particular morphemes, while others name the class a morpheme belongs to. More on this in section 3.9.1.

3.5. Ambiguity

Each morpheme of L1 should be recognizable by its gloss. The reader is supported in this task if glosses are consistent within one publication. It will rather confuse him if Yucatec Maya *k'iin* is once glossed 'sun' and the next time 'day'. **Polysemy** is resolved in the idiomatic translation. The gloss renders neither the contextual sense nor the full meaning range of an item. Naturally, this does not apply to **homonymy**. Homonymous L1 morphs represent different morphemes and therefore receive different glosses. This is stipulated by R7, which follows from R4.

If the senses of an item are reducible to a **Gesamtbedeutung**, then this should be used in the gloss (R8). For instance, the Turkish dative/allative suffix *-a* is glossed by 'DAT'. The **Gesamtbedeutung** rather than the **Grundbedeutung** should appear in the gloss, because it has better chances to fit all the diverse contexts in which the item occurs. Sometimes, there is either no **Gesamtbedeutung**, or if there is, L2 does not have a term for it. In cases like Yucatec Maya *k'iin* 'sun, day', there are various alternatives. First, the **Grundbedeutung** may be used as the gloss; thus Yucatec Maya *k'iin* 'sun'. However, if all the occurrences of a polysemous morpheme in a particular publication reflect the same (derived) reading, then generally no useful purpose is served if it is consistently glossed by its basic meaning. For instance, all the occurrences of Yucatec Maya *k'iin* in a particular text might mean 'day'. Then this would be the appropriate gloss. Finally, any kind of reduction may seem misleading. Then two or even more senses may be indicated in the gloss, separated by a slash, e.g. Yucatec Maya *k'iin* 'sun/day'. (2) illustrates the same convention.

- (2) Korean
- | | | | |
|----------------------------|-----------------|------------|----------------|
| <i>Toli-nun</i> | <i>kae-hako</i> | <i>cal</i> | <i>non-ta.</i> |
| Toli-TOP | dog-ADD | often/well | play:PRS-DECL |
- 'Toli likes to play with the dog.'

Syncretism often involves extensive polysemy and/or homonymy. If it were to be made explicit in an IMG, then e.g. the gloss for Lat. *ancillae* would have to be 'maid.F:GEN.SG/DAT.SG/NOM.PL'. This may be appropriate if the discussion in the context deals with syncretism. Otherwise, only the category actually required by the context may be shown, e.g.:

- (3) Latin
- | | |
|-----------------|--------------|
| <i>ancillae</i> | <i>orant</i> |
| maid.F:NOM.PL | pray:3.PL |
- 'the maids pray'

In other words, in cases of syncretism the last two bullet points of R8 must be resorted to.

A whole paradigm of markers may be used in two clearly distinct functions. For instance, a set of cross-reference markers may combine with a verb to reference its subject,

3.7. Derived stems

The morpho-semantic structure of a derived stem may be completely regular and transparent, as in Germ. *wolk-ig* ‘cloud-ADJVR (cloudy)’, or it may be opaque, as in Germ. *heil-ig* ‘salvation-ADJVR (holy)’. If the discussion focuses on word-formation, then both of these words will be glossed as indicated. If the internal structure of stems is of no relevance, then it will not be shown in the L1 text line, and consequently the glosses can reduce to ‘cloudy’ and ‘holy’, respectively.

For opaque complex stems, morphological segmentation plus corresponding gloss often amounts more to etymology than to morphological analysis. It also unnecessarily obscures the correspondence of the gloss to the idiomatic translation. This should be borne in mind before one carries it through as a general principle in text editions.

In an ideal methodological situation, an IMG is taken from a lexicon, where the gloss constitutes one of the fields in the microstructure of each lexical entry. The German lexicon may contain, e.g., the three entries *Huf* ‘hoof’, *Eisen* ‘iron’ and *Hufeisen* ‘horse-shoe’. If the latter occurs in an L1 text, then it may either be analyzed or not. In the former case *Huf* and *Eisen* will be looked up in the lexicon and will be matched by their glosses, while in the latter case *Hufeisen* will be looked up and be glossed accordingly.

3.8. Submorphemic units

There are two kinds of submorphemic units: parts of morphemes with a sound-symbolic value and strings of phonemes inserted between morphemes for euphonic or similar reasons. The former kind is not generally subjected to morphemic analysis and may therefore be left out of consideration here. The latter kind may be illustrated by the second element in forms such as French *a-t-il* ‘has he’ and Germ. *Weihnacht-s-gans* ‘Christmas goose’. If the submorphemic unit is not at stake in the context, then the first choice is to abstain from an analysis by regarding the submorphemic unit as part of a stem alternant: *Weihnachts-gans* ‘Christmas-goose’. The second choice is to render the submorphemic unit by \emptyset , e.g. *a-t-il* ‘has- \emptyset -he’. A euphonic submorphemic unit may be glossed by ‘EU’ instead of ‘ \emptyset ’.

3.9. Grammatical category labels

3.9.1. General

As was said in 3.4, the gloss for a grammatical item is generally not a grammatical item of L2, but a grammatical category label (R6). For instance Yucatec Maya *yàan* is not rendered by ‘be’, but by ‘EXIST’, one of the reasons being that L2 ‘be’ is a copula, while Yucatec Maya *yàan* is not. While this poses few problems for such categories for which the European grammaticographic tradition possesses terms, it does pose a problem for certain classes of semi-grammaticalized items such as function verbs and coverbs. Coverbs are words which are grammaticalized from verbs to minor parts of speech, mostly adpositions. If they function as the latter, they may express a semantic role. In Mandarin, for instance, *yòng* has the lexical meaning ‘use’ and the grammatical meaning ‘INSTR’, as in (6).

- (6) Chinese
Tā yòng shǒu zǒu lù.
 he use/INSTR hand walk road
 ‘He walks on his hands.’

This kind of problem is not solved by putting the lexical meaning in upper case (‘USE’), since ‘use’ is neither a grammatical concept in L2 nor a term of the grammatical metalanguage. Applying R8 in such cases would imply opting in favor of the Gesamtbedeutung of the item, which in such cases is the grammatical meaning. The gloss would then be ‘INSTR’ (or some more language-specific grammatical category which may better suit this particular function). The problem remains, however, that the same word can occur as the sole predicate of a clause, in the meaning ‘use’ (e.g. *tā yòng shǒu* ‘he uses his hand’). An IMG ‘INSTR’ would be hardly intelligible there. The alternative of only using the Grundbedeutung – ‘use’ in (6) and throughout – would be in conflict with the principle that morphological analysis must be kept distinct from etymology. Here the third alternative offered by rule R8 may be resorted to, viz. providing both meanings in the gloss of each occurrence of the item, thus: *yòng* ‘use/INSTR’.

An IMG identifies an L1 morpheme. It names a value, not a parameter. Mentioning the name of the generic category in the gloss instead of the specific value is nevertheless widespread usage. One finds both Japanese *yom-i* and *yon-de* glossed by ‘read-CONV’ (converb), which hinders the reader in his attempt to keep the converb forms apart. One finds Onondaga *waʔha-yeʔkwa-hni:-nu?* ‘he bought tobacco’ glossed as ‘TNS:he/it-tobacco:buy-ASP’ (Woodbury 1975:10), which is of no use for somebody studying the interdependence of incorporation with tense and aspect.

IMGs not seldom contain labels that do not correspond to the principles introduced so far. Sometimes, elements without morphological status are separated and glossed. Sometimes, the parameter instead of the particular value of a grammatical category is identified. Sometimes, syntactic or semantic instead of morphological information is given. Here is an incomplete list of labels that have repeatedly been found in glosses but which should be avoided.

label	intended meaning	comment
A	transitive subject	in morphemic glosses, the abbreviation is ERG
ADV	adverb	specify meaning
AGR	agreement	specify agreement categories
AGT	agent	this is not a value of a morphological category
ART	article	only if it has no determinative properties
ASP	aspect	specify particular aspect
AUX	auxiliary	only if there is only one auxiliary morpheme in the language
CARD	cardinal	only if it is a morpheme or grammatical feature
CLF	classifier	this is a word class
CLT	clitic	this is neither a morphological category nor a value of one
EP	epenthetic	has no morphological status, should not be separated in the first place
EVID	evidential	specify particular evidential
PAT	patient	this is not a value of a morphological category
PREP	preposition	this is a word class

PTL	particle	this is (at best) a word class
TNS	tense	specify particular tense

Tab. 169.3: Labels to be avoided

3.9.2. List of grammatical categories and their glossing labels

No list of grammatical category labels can be complete. The list following in Tab. 169.4 (which incorporates the list in Lehmann et al. ²1994) only contains the most widespread categories. When more than one abbreviation is mentioned, they are given in the order of preference. To the extent that these abbreviations are or become wide-spread, they get the status of linguistic abbreviations like ‘NP’, which need not be defined when used. If a publication uses labels not contained in the following list, it must explain them in an individual list of abbreviations.

Grammatical category labels are subject to two conflicting requirements: they must be both distinct and short. The former requirement takes precedence. It is, for instance, not possible to use ‘COMP’ in one and the same publication to mean both ‘completive’ and ‘complementizer’. The list in Tab. 169.4 avoids such clashes. However, in an individual publication that has nothing to do with complementation, the aspect may, of course, be abbreviated by ‘COMP’ (instead of ‘CMP(L)’, as in the list). Parenthesized parts of an abbreviation are only necessary if a distinctness conflict arises.

Tab. 169.4 contains only such terms which may appear in an IMG. In other publications, similar lists of terms for syntactic categories and functions and for semantic and pragmatic functions may be found.

‘Cross-reference position’ means a morphological slot, usually on a verb, occupied by pronominal elements that agree with or refer to a dependent in a specific syntactic function. ‘Case’ means a case relator that may take the form of a case affix or an adposition. Verb derivational morphemes get these glosses only if they are homonymous with nominal case relators.

value	abbrev.	category	comment
1 st person	1	person	
2 nd person	2	person	
3 rd person	3	person	
abessive	(PRV) (AVERS)		use ‘privative’ and ‘aversive’
ablative	ABL	local case	‘from’ (= separative)
absolute	ABSL	nominal	free non-incorporated form of noun
absolutive	ABS	grammatical case or cross-reference position	in ergative system
abstract	ABSTR	nominal	
accusative	ACC	grammatical case	
action nominalizer	ACNNR	deverbal nominal derivation	
active	ACT	voice; case or cross-reference position	in active system
actor	ACR	grammatical case or cross-reference position	
actor topic	A	voice	
additive	ADD	case	
addressee-honorific	2HON	honorification	
addressee-humble	2HML	honorification	
adelative	ADEL	local case	

adessive	ADESS	local case	
adhortative	(HORT)		use 'hortative'
aditive	(ALL)		use 'allative'
adjectiv(al)izer	ADJR	derivational or syntactic	
admonitive	ADM	mood	
adverbializer	ADVR	derivational or syntactic	
adversative	ADRVS	interpropositional relation	'whereas'
affirmative	AFFMT	opposite to negative	normally unmarked
agent nominalizer	AGNR	deverbal nominal derivation	
agentive	AG		
alienable	AL	possessive attribution morpheme	
allative	ALL	local case	'to'
allocutive	ALLOC	honorification	kind of addressee-honorific
anaphoric	ANA	pronominal	
andative	AND	deictic	
animate	AN		
anterior	ANT	tense	relative tense
anticausative	ACAUS	deverbal verb derivation	= deagentive, blocking of actor argument
antipassive	APASS	voice	
aurist	AOR	tense-aspect	perfective past (as opposed to imperfect)
applicative	APPL	deverbal verbal derivation	subtypes may be distinguished by APPL.REC, APPL.INST etc.
apprehensional	APPR	interpropositional relation	'lest'
assertive	ASRT	modality	subtype of declarative: high degree of commitment
associative	ASS(OC)	adnominal case	'with, à'
assumed	ASSUM	evidential	
attenuative	ATTEN	deverbal verb derivation	
attributor	AT	nominal	links an attribute to the head
auditory	AUD	evidential	
augmentative	AUG	denominal nominal derivation	
auxiliary	AUX		if it is the only auxiliary root
benefactive	BEN	case	'for'
cardinal	CARD	numeral	if marked grammatically
caritive	(PRV)		use 'privative'
causative	CAUS	deverbal verb derivation	
circumstantial	CIRC	interpropositional relation	'in, by'
clamative	(EXCL)		use 'exclamative'
classifier	CLF	nominal	followed by class identifier, e.g. HUM
cohortative	(HORT)		use 'hortative'
collective	COLL		
comitative	COMIT	case	'with, in the company of'
common	COMM	gender	either masc. or fem.; cf. 'human' and 'animate'
comparative	CMPR	degree of comparison	
complementizer	COMP	subordinator	= SR
completive	CMPL, CMP	aspect	normally = perfective
conative	CNTV	mood	
concessive	CONC	interpropositional relation	'although'
conditional	COND	interpropositional relation; mood	'if'; 'would'
conjunctural	CONJC	evidential	

conjunctive	CONJ	interpropositional relation	of non-finite predicate
connector, -ive	CONN		if there is only one
consecutive	CONSEC	interpropositional relation	‘so that’
construct	CONST	nominal	construct state
converb	(GER)		use ‘gerund’
continuous	CONT	aspect/aktionsart	
copula	COP		if there is only one
crastinal	CRAS	tense	tomorrow
dative	DAT	grammatical case	
deagentive	(ACAUS)		use ‘anticausative’
debitive	(OBLG)		use ‘obligative’
declarative	DECL	sentence-type	normally unmarked
deferential	DEFR	honorification	~ speaker-humble
definite	DEF	determination	
deictic of 12 person	D12	determination	
deictic of 1 st person	D1	determination	
deictic of 2 nd person	D2	determination	
deictic of 3 rd person	D3	determination	
delative	DEL	local case	‘down from’
demonstrative	DEM	determination	
dependent verb form	(SUBJ)		use ‘subjunctive’
desiderative	DES	deverbal verb derivation	
destinative	DEST	local case; also on non-finite verb forms (= supine)	‘to’; if typically for human destinations, use ‘benefactive’
determiner	DET	pronominal	will normally be DEF, INDEF, GNR, SPEC, NSPEC
detransitivizer	DETR	deverbal verb derivation	see also ‘anticausative’ and ‘introversive’
different subject	DS		
diminutive	DIM	denominal noun derivation	
direct	DR	voice	vs. inverse
direct evidential	DIREV	evidential	
direct object	DO	cross-reference position	
directional	DIR	case or verb derivation	‘towards’; use AND and VEN for deictic directionals
distal	DIST	determination	remote from deictic center
distributive	DISTR	nominal or verbal	
donative	DON		auxiliary of benefactive construction
dual	DU, DL	number	
dual exclusive	DE	number	
dual inclusive	DI	number	
dubitative	DUB	mood	
durative	DUR	aktionsart	
dynamic	DYN	aktionsart	vs. stative
egressive	EGR	aktionsart	
elative	ELAT	local case	‘out of’
emphasizer/emphatic	EMPH	funct. sentence perspective	e.g., class of pronoun
equative	EQT	1. case; 2. predicative	‘as’; feature/marker of adjective in nominal clause
ergative	ERG	grammatical case or cross-reference position	in ergative system
essive	ESS	case	‘as’; see also ‘transformative’
evidential	EVID	verbal	
exclamative	EXCL	mood	

exclusive			use 'dual exclusive', 'plural exclusive'
exist(ential)	EXIST	grammatical verb	
experiential	EXPER	aspect	
extrafocal	EXFOC	verbal	status of subordinate clause of cleft-sentence
extraversive	EXTRV	deverbal verb derivation	transitivization by addition of undergoer
factitive	FACT	denominal/deadjectival verb derivation	A-FACT NP 'make NP A'
familiar	FAM	pronominal	
feminine	F	gender	
finite	FIN	verbal	
first person dual inclusive	I2		if treated as a quasi-singular; otherwise 'dual inclusive'
focus	FOC	funct. sentence perspective	
formal	FRM	mood	
frequentative	FREQ	aktionsart	multiple times on several occasions
future	FUT	tense	
generic	GNR	determination	
genitive	GEN	grammatical case	
gerund	GER	verbal	verbal adverb or converb
gerundive	(OBLG)		use 'obligative'
habitual	HABIT	aktionsart	~ customary
habitual-generic			use 'habitual', 'generic'
habitual-past			use 'habitual', 'past'
hesitative	HESIT	funct. sentence perspective	
hesternal	HEST	tense	yesterday's past
hodiernal future	HODFUT	tense	today's future
hodiernal past	HODPST	tense	today's past
honorific	HON	honorification	
hortative	HORT	mood	1 st person imperative
human	HUM		
humble	HML	honorification	comprises 'speaker-humble, addressee-humble, referent-humble'
hypocoristic	HCR	affect	
hypothetical	HYP	mood	
illative	ILL	local case	'into'
immediate	IMM	tense	specifier of other tenses
immediate/imminent future	IMMFUT	tense	
immediate past	(RECPST)		use 'recent past'
imperative	IMP	mood	
imperfect	IMPF	tense-aspect	imperfective past; vs. aorist
imperfective	IPFV	aspect	
impersonal	IMPR		only if formally distinct from the specific persons
impersonal passive	IPS	voice	passive without promotion to subject in active system
inactive	INACT	grammatical case or cross-reference position	
inalienable	INAL	nominal	possessive attribution morpheme or feature
inanimate	INAN		
inceptive	(INGR)		use 'ingressive'
inchoative	INCH	denominal verbal derivation	N/A-INCH 'become N/A'
inclusive			use 'dual inclusive', 'plural inclusive'
incomplete,	INCMP(L)	aspect	normally = imperfective

noncompletive			
inconsequential	INCONS	interpropositional relation	
indefinite	INDEF	determination	
independent	INDEP	mood	only if distinct from indicative
indicative	IND	mood	
indirect object	IO	cross-reference position	
inessive	INESS	local case	'inside'
inferential	INFR	mood or evidential	
infinitive	INF	verbal	
ingressive	INGR	aktionsart	
injunctive	INJ	mood	
instructive	(MAN)		use 'manner'
instrument nominalizer	INSTNR	deverbal nominal derivation	
instrumental	INST(R)	case	
intensive	INTS	verbal	often aktionsart
interrogative	INT	sentence type	particle or morphological category
intransitive	INTR	verbal	morpheme or grammatical category
intransitive subject	S	cross-reference position	only if opposed to both A and P; use SBJ otherwise
introversive	INTRV	deverbal verb derivation	blocking of undergoer argument
inverse	INV	usually verbal	vs. direct
invisible	INVS	determination	
irrealis	IRR	mood	
iterative	ITER	aktionsart	several times on one occasion
jussive	JUSS	mood	3 rd ps. imperative or dependent mood
lative	LAT	local case	'to ~ from ~ via'
ligature	LIG	nominal	
linker	LNK	nominal	links constituents of a phrase, typically an NP; properly includes 'attributor'
locative	LOC	local case	
locative topic	LT	voice	
logophoric	LOG	pronominal or verbal	
malefactive	MAL	deverbal verb derivation	
manner	MAN	case	also on non-finite verbs
manner nominalizer	MANNR	deverbal nominal derivation	
masculine	M	gender	
masculine personal	MHUM	gender	
medial	MED	determination	medial distance from deictic center
medial	MEDV	verbal	verb form in a chain
mediative	MEDT	case	'between, among; by means of'
mediopassive	MEDP	voice	
middle	MID	voice	excludes passive
motivative	MTV	case	'by'; sometimes called 'causal'
narrative	NARR	tense	
near future	NRFUT	tense	after 'immediate future'
negative	NEG		
neuter	N	gender	
nominalizer	NR	deverbal nominal derivation or syntactic subordination	see also the more specific ones
nominative	NOM	grammatical case	
non-	N		e.g. NPST
non-finite	NFIN	verbal	
non-future	NFUT	tense	
non-human	NHUM	gender	
non-masculine personal	NM	gender	

non-past	NPST	tense	
non-plural	NPL	number	< 3
non-singular	NSG	number	> 1; only if there is a plural for > 2
non-specific	NSPEC	determination	
non-visual	NVIS	evidential	non-eye-witness
non-volitional	NVOL	verbal	
noun class n	CLn		where n is a number or a feature
object	OBJ	cross-reference position	
obligative	OBLG	mood	
oblique	OBL	case	
obviative	OBV	person	vs. proximate
optative	OPT	mood	
ordinal	ORD	numeral	
participle (marker)	PART	verbal	
partitive	PRTV	case	
passive	PASS	voice	
past	PST	tense	
patient nominalizer	PATNR	deverbal nominal derivation	
patient topic	PT	voice	
paucal	PAU	number	
pejorative	PEJ	affect	
perfect	P(R)F	tense-aspect	
perfective	PFV	aspect	
pergressive	(PERL)		use 'perlative'
perlative	PERL	local case	'through'
place nominalizer	LOCNR	deverbal nominal derivation	
pluperfect	PLUP	tense	past or perfect of a past
plural	PL	number	
plural exclusive	PE	number	
plural inclusive	PI	number	
pluritive	(PL)		plural of a singulative; use 'plural'
polite	(FRM)		use 'formal'
positional	POSIT	verbal	
positive	(AFFM)		use 'affirmative'
possessive	POSS	possessive adjective, pronoun and cross-reference position	not for an adnominal case relation; that is GEN or AT
postcrastinal	POCRAS	tense	future after tomorrow
postelative	POSTEL	local case	'from behind'
posterior	POST	relative tense	
postessive	POSTESS	local case	'behind'
post-hodiernal	POHOD	tense	future after today
potential	POT	mood	
precative	PREC	mood	for requesting
predicative	PRED	nominal	predicative form
present	PRS	tense	
preterite	(PST)		use 'past'
pre-hesternal	PRHEST	tense	past before yesterday
primary object	PO	cross-reference position	
privative	PR(I)V	case	'without'
processive, -ual	PROC	denominal verb derivation	
progressive	PROG	aspect	
prohibitive	PROH	mood	negative imperative
prolative	PROLAT	local case	'along, by (way of)'
proprietary	PROPR	case or derivational category	'having, provided with'
prospective	PROSP	tense-aspect	'going to'; opposite of perfect
proximal	PROX	determination	near the deictic center

proximate	PRX	person	vs. obviative
punctual	PNCT	aspect or aktionsart	
purposive	(DEST)		use ‘destinative’
quality nominalizer	QUALNR	deverbal nominal derivation	
quotative	QUOT	marking indirect speech	
realis	RLS	mood	vs. irrealis
recent past	RECPST	tense	= immediate past
reciprocal	REC(P)	voice or pronominal	
reduplicative			gloss by function
referent-honorific	3HON	honorification	
referent-humble	3HML	honorification	
referentive	RFR	case	‘about’
reflexive	R(E)FL	voice or pronominal	
reinforcement	(INTNS)		use ‘intensive’
relational(izer)	RELL	nominal	
relative	REL	subordinative and/or pronominal	in relative clause
relative	(RFR)		use ‘referentive’
remote	(DIST)		use ‘distal’
remote past	REMPST	tense	
repetitive	REP	aktionsart	only if distinct from iterative
reportative	RPRT	evidential	
resultative	RES	aspect or aktionsart	
reversive	RVRS	aktionsart	
same subject	SS		
secondary object	SO	cross-reference position	
semelfactive	SMLF	aktionsart	
sensory	SENS	evidential	
separative	(ABL)		use ‘ablative’
sequential	SEQ	interpropositional relation	vs. simultaneous
simultaneous	SIM	interpropositional relation	vs. sequential
singular	SG	number	restricted
singulative	SGT	nominal	vs. collective
sociative	SOC	verbal	‘together’
speaker-honorific	1HON	honorification	
speaker-humble	1HML	honorification	
specific	SPEC	determination	
speculative	SPECL	evidential	
stative	STAT	aktionsart	
subrelative	SUBEL	local case	‘from under’
subessive	SUBESS	local case	‘under’
subject	SBJ	cross-reference position	
subjunctive	SUBJ	mood	
sublative	SUBL	local case	‘to under’
subordinator	SR	interpropositional relation	only for the single universal subordinator (‘that’)
superdirective	(SUPL)		use super-lative
superrelative	SUPEL	local case	‘from above’
superessive	SUPESS	local case	‘above’
superlative	SUP	degree of comparison	
super-lative	SUPL	local case	‘to above’
terminative	TERM	local case or aktionsart	‘up to’
topic	TOP	funct. sentence perspective	
transformative	TRNSF	case	‘becoming’; dynamic counterpart of essive
transitive	TR	verbal	morpheme or grammatical category
transitive patient	P	cross-reference position	only if opposed to both S and A; use OBJ

transitive subject	A	cross-reference position	otherwise only if opposed to both S and P; use ERG otherwise
transitivizer	TRR	deverbal verb derivation	
translative	TRNSL	local case	‘across’
trial	TRL	number	only if distinct from paucal
undergoer	UGR	cross-reference position	
unrestricted	(PL)		use ‘plural’
unspecified	UNSPEC	person	unspecified argument of relational base
validator			use ‘assertive’, ‘declarative’
venitive	VEN	deictic	
verbalizer	VR, VBZ	verbal derivation	
visible	VS	determination	
visual	VIS	evidential	eyewitness
vocative	VOC	case	
volitional, volitive	VOL	verbal	
zero	∅		making no contribution to sentence meaning

Tab. 169.4: Grammatical category labels

4. Boundary symbols

4.1. Basic rules

Rules R1 and R4 guarantee correspondence between units in the L1 text and in the IMG. They do not, however, insure that the vertical alignment works in a mechanical way. This is desirable in certain contexts such as automatic parsing. It can be guaranteed in a fully formalized representation, which would then take the form of a table (s. Lieb & Drude 2000). In less formal situations, it cannot be fully guaranteed because there may be good reasons not to insert morpheme boundaries in the L1 text while still representing each morph by a separate gloss (cf. R13). Correspondence of boundary symbols in the L1 and the IMG lines is therefore not generally an equivalence, but only an implication: boundary symbols in the L1 line are matched by corresponding boundary symbols in the IMG (R9). We will review the kinds of boundaries and their delimiters in turn.

The **word boundary** is shown by a blank in L1. This is repeated in the IMG, and conversely there is a blank in an IMG only if there is a corresponding blank in the L1 line. This particular rule (R10) is therefore stricter than R9. R10 prohibits two situations: a word being rendered by a sequence of two words; and a sequence of two words being rendered by one word. The first situation will be discussed in section 4.5. Sometimes a sequence of two L1 units (words or morphemes) corresponds to one L2 unit. In principle, this situation should not arise in the IMG because each of the L1 units should have its own gloss. However, it is possible that either the L1 units have no meaning in isolation or else mean something totally different than their combination, the latter being idiomaticized. In such cases, glossing them separately might give a misleading impression of the workings of the grammar. When the bisected L1 unit forms an orthographic unit (e.g. a compound), one may simply dispense with the analysis (cf. section 3.7). For instance, instead of Germ. *be-komm-en* ‘APPL-come-INF’,

one can write *bekomm-en* ‘get-INF’. If the orthography requires a boundary, as in Yucatec Maya *le kah* ‘when’, the first choice is to gloss the items separately (in this case, ‘DEF SR’) and to leave the semantic interpretation to the idiomatic translation. The second choice is to indicate the semantic unity of the two L1 items typographically by replacing the blank by a boundary symbol that does not interfere with the orthography, e.g. by an underscore: *le_kah* ‘when’ (R11). If L1 orthography links the two items by another symbol that is also an IMG boundary symbol, as in Engl. *vis-à-vis* ‘facing’, no satisfactory solution is known.

Apart from special cases to be noted, the **morpheme boundary** is shown by a hyphen in L1 (R12). This is repeated in the IMG; and here again the converse applies, too. Apart from the *vis-à-vis* type exception, this does not pose any problems. It does, however, happen that the L1 text contains a combination of two morphemes, but no boundary is shown between them. Various motivations for this are conceivable, be it that two morphemes are fused in a portmanteau morph, be it that the position of the boundary is not clear or irrelevant, be it that the analyst does not want to disfigure L1 orthography with boundary symbols. In such cases, a colon in the IMG is a hint at a morpheme boundary existing, but not shown in the L1 line (R13). The purpose of R13 is to allow the analyst to forgo a segmentation while still saving R1 and insuring biuniqueness of the other boundary symbols. Several examples may be seen in (1). The colon is also used to render a portmanteau morph, e.g. French *au* ‘DAT:DEF’. More on this in section 4.5.

Special symbols may be introduced to distinguish kinds of morpheme boundaries. For instance, the use of the plus sign to signal a boundary in **compounding**, as in German *Weihnachts+gans* ‘Christmas+goose’ is rather widespread; and occasionally it is also found in derivation, as in German *wolk+ig* ‘cloud+ADJVR (cloudy)’ (R14).

No orthography distinguishes **clitic boundaries** from word and morpheme boundaries. If L1 is represented in conventional orthography, then the simplest solution for an IMG is not to distinguish them either. Thus French *je le sais* ‘I know it’ will be glossed as ‘SBJ.1.SG DO.3.SG.M know.SG’, while Latin *itaque* ‘and so’ will be glossed by ‘so:and’. If clisis is important or the L1 representation is non-orthographic, then the clitic boundary will be shown by an equal sign both in the L1 text and in the IMG, thus: *ita=que* ‘so=and’ (R15).

If a **zero morph** or **morpheme** is represented in L1 by \emptyset (cf. section 2.1), no special measures need be taken. If it is not there represented, then its gloss is enclosed in parentheses (R16), like this: Lat. *timor* ‘fear.M(NOM.SG)’. In this example, a stem is accompanied by two (complexes of) grammatical category labels, ‘M’ and ‘NOM.SG’. The first is separated by a period because it corresponds to an inherent feature of the stem. The second is enclosed in parentheses because it corresponds to a separate morpheme.

4.2. Discontinuity

Discontinuous units – words or morphemes – are like bisected units in that one semantic unit is represented by two expression units. However, they present the added difficulty that their parts are not adjacent, so the IMG has to make it explicit what belongs together. For a **discontinuous stem** or **affix**, diverse solutions have been proposed in the literature. Among them is the proposal (Bickel et al. 2004) to repeat the same gloss under each part of the discontinuous item. However, this seems misleading, as the syntagmatic cooccurrence of synonymous L1 items is not at all rare – e.g. in hypercharacterization – and must be distinguished from discontinuity. An unambiguous solution for a **circumfix** is to set it off by

angled brackets, like this: Germ. *ge>lauf<en* ‘<PART.PRF>run’ (run (part.prf.))’ (R17). Discontinuous words are rare. The first choice is to try and gloss each part independently, as done for the German circumposition *um ... willen* ‘for’ in (7).

- (7) German
um unser-es Heil-es willen
 for our-GEN.SG salvation-GEN.SG sake
 ‘for (the sake of) our salvation’

The second choice is to treat them by the same formalism as for circumfixes. Consider the case of preverbs. In several Indo-European languages, they may be distanced from their host verb to yield a discontinuous verb stem. There are two options for glossing such discontinuous compounds: If the compounding is relatively transparent, one may prefer to provide the preverb and the base each with its gloss. If the compound is completely lexicalized, this might be misleading, and so it may be preferable to treat it as a discontinuous morpheme in the gloss, as in (8).

- (8) German
es hör>-t jetzt <auf
 it <stop>-3.SG now
 ‘it stops now’

Infixes, too, require a special boundary symbol in order to insure that the root bisected by them is perceived as a unit. This is achieved enclosing them in angled brackets as shown in (9)-(10) (R18).

- (9) Latin
vi<n>c-o
 conquer<PRS>-1.SG
 ‘I conquer’

- (10) Indonesian
t<el>unjuk
 <AGNR>point
 ‘forefinger’

The gloss of a left-peripheral infix precedes the gloss of its host, the gloss of a right-peripheral infix follows it (Bickel et al. 2004).

4.3. Reduplication

Reduplicative segments may have the same kinds of grammatical functions as affixes, and sometimes they are formally not easily distinguished from affixes. Therefore they must be glossed just like affixes, but at the same time they must be formally distinguished from affixes. This is achieved by providing the same kind of gloss for them as for grammatical formatives, but separating them by a tilde (R19); Bickel et al. 2004), as in (11)-(12).

- (11) Ancient Greek
gέ~graph-a
 PRF~write-1.SG
 ‘I have written’
- (12) Yucatec Maya
k’áa~k’as
 INTNS~bad
 ‘wicked’

4.4. Other morphological processes

Morphological processes not covered by the above conventions comprise transfixation, internal modification, metathesis, subtraction and suprasegmental processes (cf. ch. VIII). These are like infixation in not being peripheral to the base, but they differ from it in that the grammatical meaning in question is not associated with a single string of segments which, if subtracted, leaves the base. The notation recommended here distinguishes them from the other morphological processes, but not from each other. Such a morpheme can hardly be signaled in the L1 representation. In the IMG, its gloss follows the gloss of the base, separated by a backslash (R20). An example of transfixation is the Arabic broken plural, as in *bujūt* ‘house\PL (houses)’. Apophony, metaphony, e.g. German *säng-e* ‘sing\IRR-1/3.SG (I/he would sing)’, and tone shift, as in Yucatec Maya. *hàats’* ‘beat\INTROV (beat (unspec. object))’ are treated in the same way.

4.5. Semantic and grammatical features

The gloss of a grammatical morph often consists of a set of symbols. They are separated by a period, as in Germ. *Tisch-es* ‘table-GEN.SG’ (R21). The same rule applies in the situation mentioned in section 3.3, where an L1 lexeme is glossed by more than one L2 words. These, too, are separated by a period, as in Germ. *fabulier-en* ‘invent.stories-INF’.

Lexical stems fall into grammatical classes. Noun stems, for instance, have gender; verb stems have valence. If such grammatical categories are covert, this information is not deducible from (the gloss of) the lexical meaning. It therefore makes sense to represent it in the gloss of the stem. The Latin example *puellae* ‘girl.F:NOM.PL’ of section 2.1 shows how this may be done for gender. The same would be possible with transitivity. Instead of Yucatec Maya *hats’-ah* ‘beat-CMPL’ as shown in section 2.2, we might put ‘beat.TR-CMPL’. It does not seem necessary to have a rule here beyond R3 and R21.

The period between values of different morphological categories cumulated in one morpheme is dispensable between person, gender and number, provided the resulting letter sequence is unambiguous. Thus, Latin *lauda-mus* may be glossed as ‘praise(PRS.IND)-1.PL’ or ‘praise(PRS.IND)-1PL’.

Sometimes the period is used as a general-purpose symbol to hide the lack of an analysis, including the function of the colon as regulated by R13. This is not recommendable if – as is usually the case – the period is also used in the function regulated by R21. Given

bracketing seems more natural there (R24). In principle, an IMG may even be combined with a labeled bracketing; but above some rudimentary level, this will soon lead to illegibility.

5. Typographic conventions

IMGs obey a number of typographic conventions all of which aim at facilitating the reader's task. First, if there are more lines of linguistic representation (cf. section 1.3), for instance one of syntactic constituency or lines that show syntactic, semantic or pragmatic functions of the construction, then these follow the IMG, as stipulated in R25. Second, words (neither larger nor smaller units) of L1 are left-aligned with their glosses (R27). Further, since IMGs are generally longer than the L1 text they render, they are printed in a smaller type-face (R28), and grammatical category labels are abbreviated (R29).

For comparison, here is an example of a publication which does not observe these rules (Monod-Becquelin 1976:138 on Trumai):

šyšyk letsi k'ate šy hai-ts šyšy-ka-ke

“avec du piment, je rends le poisson piquant (regarde)”

// piment / avec / poisson / *actualis.* / *lère pers. erg.* / piquant-*causatif-marque d'adjectivisation* //

Furthermore, since IMG lines are not sentences, the relevant orthographic rules of punctuation, initial capitalization and syllabification do not apply (R30 – R32).

6. Summary

Instead of a prose summary, a list of the rules and symbols proposed follows:

6.1. Rules

6.1.1. Glossing principles

- R1. With the exceptions specified below, there is a symbol or a configuration of symbols in the IMG if and only if there is a morph in the L1 text that it corresponds to.
- R2. The IMG represents morphemes, not allomorphs. Therefore, the gloss of a grammatically conditioned allomorph does not contain the grammatical category that conditions it.
- R3. An IMG should be as precise and detailed as tolerable. The limits of precision and detail are defined by practical considerations of complexity and intelligibility.
- R4. There is a biunique mapping of individual L1 morphemes onto glosses.
- R5. (a) An L1 lexeme is glossed by L2 lexemes.
(b) L1 stems are glossed by L2 stems.
- R6. The gloss of a grammatical morph is a configuration of grammatical category labels each of which represents the value of a grammatical category. A grammatical morph

should not be glossed by an L2 bound morpheme. It may be glossed by an L2 word if that has the same function as the L1 morph.

- R7. Homonymy is resolved in the IMG, polysemy is preferably not.
- R8. The gloss of a polysemous L1 item should represent, in the order of decreasing preference,
- its Gesamtbedeutung,
 - its Grundbedeutung,
 - the set of its senses,
 - its contextual sense.

6.1.2. Boundary symbols

- R9. Apart from R30, there is a boundary symbol of a certain type in the IMG if there is a corresponding boundary symbol in the L1 text. More strictly, there is a blank, hyphen, plus, equal sign, angled bracket and tilde in an IMG if and only if there is an identical symbol in the L1 text corresponding to it.
- R10. A word boundary is shown by a blank ().
- R11. Two successive orthographic L1 words which must be glossed by one L2 word are linked by an underscore (_).
- R12. A morpheme boundary is shown by a hyphen (-).
- R13. A morpheme boundary not shown in the L1 text is indicated by a colon (:) in the IMG. This applies also to portmanteau morphs.
- R14. A boundary in a compound stem, and possibly also in a derived stem, may be shown by a plus sign (+).
- R15. A clitic boundary may be shown by an equal sign (=).
- R16. A gloss of a zero morpheme or allomorph is enclosed in round parentheses (()).
- R17. The string enclosed in a discontinuous L1 item P1 ... P2 is enclosed in inverted angled brackets (P1> ... <P2). In the IMG, P1 receives a gloss enclosed in angled brackets; P2 is not glossed.
- R18. An infix is enclosed in angled brackets both in the L1 text and in the IMG. The gloss of a left-peripheral infix precedes the gloss of its host, the gloss of a right-peripheral infix follows it.
- R19. A reduplicative segment is glossed like an affix (i.e. by a configuration of grammatical category labels) and separated from its source by a tilde (~).
- R20. A grammatical meaning expressed by a non-segmentable morphological process (transfixation, internal modification, metathesis, subtraction, suprasegmental process) is not signaled in the L1 representation. Its gloss follows the gloss of the base, separated by a backslash (\).
- R21. Elements of an IMG that represent components of one L1 morph are separated by a period (.).
- R22. As a special case of R21, components of one L1 cross-reference morph that have distinct reference are separated by the ampersand ('&') or, where no conflict with R17 and R18 arises, by the greater-than sign ('>') for actor and undergoer cross-reference.
- R23. An L1 word form whose morphological structure is not represented in the IMG may be represented by a set of symbols whose status as representing morphs or features is ignored and whose sequence has no implications as to L1. Such symbols that jointly correspond to an L1 word form are joined by an underscore (_).

R24. If constituent structure is to be displayed, square brackets ([]) can be inserted in the IMG.

6.1.3. Typographic conventions

- R25. The IMG is in the line immediately below the corresponding L1 text line.
 R26. The distance between an L1 text line and the line immediately preceding it is greater than that between it and the IMG line belonging to it.
 R27. Each L1 word form is left-flush with the L2 word or complex of symbols rendering it. If such an arrangement is impossible, the following is a minimum requirement: If there is, in an IMG, an equivalent to an element of an L1 text line, it is contained in the line immediately below that line.
 R28. The IMG is printed in a smaller type-face than the L1 text. If this is impossible, then at least grammatical category labels are in small capitals.
 R29. Grammatical terms appearing in IMGs are abbreviated, without a period at the end, and set in (small) capitals.
 R30. There is no punctuation in an IMG. Parentheses including optional material in the L1 line are not repeated in the IMG, either (cf. R16).
 R31. There is no sentence-initial uppercase in an IMG.
 R32. There is no syllabication either in the L1 line or in the IMG.

6.2. Symbols

L1	IMG	meaning
xy	$x\ y$	word boundary between x and y
x_y	z	x and y are two orthographic words, but one lexical word
z	x_y	x and y jointly render z without morphological analysis
$x-y$	$x-y$	morpheme boundary between x and y
$x+y$	$x+y$	x and y form a compound or a derivative stem
$x=y$	$x=y$	x and y are joined by clisis
z	x/y	x and y are alternative meanings of ambiguous z
xy	$x:y$	morpheme boundary between x and y not shown in the L1 text
	(x)	x does not have a significans in the L1 text
$a<x>b$	$ab<x>$	x is an infix in ab
$x>a<y$	$<xy>a$	xy is a circumfix around a
z	$x\y$	y is a non-segmentable morphological process on lexeme x
z	$x.y$	x and y are semantic or grammatical components of z
z	$x&y$	x and y are grammatical components of z cross-referencing two
	$(x>y)$	different dependents
x	$[x]$	x is a syntactic constituent
x	$[x]_Y$	x is a syntactic constituent of category Y

7. References

7.1. Specialized literature

- Bickel, Balthasar & Comrie, Bernard & Haspelmath, Martin 2004, *The Leipzig Glossing Rules. Conventions for Interlinear Morpheme by Morpheme Glosses*. Leipzig: Max-Planck-Institut für Evolutionäre Anthropologie
- Lehmann, Christian 1982, "Directions for Interlinear Morphemic Translations". *Folia Linguistica* 16, 199-224
- Lehmann, Christian 2004, "Data in Linguistics". *Linguistic Review* 21.2, 000-000
- Lehmann, Christian & Bakker, Dik & Dahl, Östen & Siewierska, Anna ²1994, *EUROTYP Guidelines*. Strasbourg: Fondation Européenne de la Science (EUROTYP Working Papers)
- Lieb, Hans-Heinrich & Drude, Sebastian 2000, *Advanced Glossing: A Language Documentation Format*. Berlin: Technische Universität (Working Paper)
- Simons, Gary F. & Versaw, Larry 1988, *How to use IT. A Guide to Interlinear Text Processing*. Dallas, Tx.: Summer Institute of Linguistics (Revised edition, Version 1.1)

7.2. Sources of examples

- Bloomfield, Leonard 1933, *Language*. New York etc.: Holt, Rinehart & Winston
- Evans, Nicholas 1997, "Role or Cast? Noun Incorporation and Complex Predicates in Mayali". Alsina, Alex & Bresnan, Joan & Sells, Peter (eds.), *Complex Predicates*. Stanford: CSLI Publications, 379-430
- Finck, Franz Nikolaus 1909, *Die Haupttypen des Sprachbaus*. Leipzig: B. G. Teubner [Nachdr. d. 3., unveränd. Aufl.: Darmstadt: Wiss. Buchgesellschaft, 1965]
- Fortescue, Michael 1984, *West Greenlandic*. London etc.: Croom Helm (Croom Helm Descriptive Grammars)
- Gabelentz, Hans Conon von der 1861, "Über das Passivum. Eine sprachvergleichende Abhandlung". *Abhandlungen der philologisch-historischen Classe der Königlich-Sächsischen Gesellschaft der Wissenschaften* 8, 449-546
- Monod-Becquelin, Aurore 1976, "Classes verbales et construction ergative en trumai". *Amérindia* 1, 117-143
- Woodbury, Hanni 1975, "Onondaga Noun Incorporation: Some Notes on the Interdependence of Syntax and Semantics". *International Journal of American Linguistics* 41, 10-20

Christian Lehmann, Erfurt (Germany)