INSTITUT FÜR SPRACHWISSENSCHAFT - UNIVERSITÄT KÖLN

Arbeitspapier Nr. 37

GUIDELINES FOR INTERLINEAR MORPHEMIC TRANSLATIONS

A proposal for a standardization

Christian Lehmann

Mai 1980

Herausgeber der Reihe:

Prof. Dr. Hansjakob Seiler

Institut für Sprachwissenschaft
Universität
D - 5000 Köln 41

[Summary at the end, sect. 6.]

0. PRELIMINARIES

There are a few aspects of linguistic work which are susceptible to standardization. They concern mainly notational matters; the International Phonetic Alphabet is an example, or the transliteration of foreign scripts by Latin letters, e.g. the transliteration of Greek script proposed by André Martinet ("A project of transliteration of Classical Greek", *Word* 9 (1953), 152-161) and now widely accepted. The LSA Style Sheet constitutes an attempt to normalize the form of linguistic publications. The form of interlinear morphemic translations shares aspects with both of the examples cited; it concerns both the notation of linguistic representations and the outer form of linguistic publications. Guidelines for the making of interlinear morphemic translations constitute, therefore, a chapter of grammaticography, a practical discipline analogous to lexicography, but badly neglected in linguistics; I am aware of no treatment of the present topic in the literature.

The form of interlinear morphemic translations is so technical an issue that it seems to be susceptible to a theory-free treatment and, thus, to a standardization which has some chance of being agreed upon by the majority of the specialists of whatever persuasion. And it is something that ought to be standardized because there is a great deal of variation in the literature which may lead to a certain amount of misunderstanding or at least difficulty of comprehension. The deterrent examples adduced below are by no means pure inventions of mine; similar examples could easily have been cited from published sources. On the other hand, many of the conventions postulated below are already observed in well-edited morphemic translations. Thus, the treatment that follows only sanctions what is already good usage, being less of an exploratory than of a didactic nature (as is, in fact, the whole discipline of grammaticography).

I have to ask the reader two favors: First, that he regard the following proposals as a preliminary version of something which will certainly be greatly modified before it can be called anything like final. Second, that he contribute from his own ex-

perience to the final version. My colleagues and I at the Linguistic Department of the University of Cologne have found it useful to use a common format in morphemic translations,[1] and if good guidelines are available, we would certainly like to stick to them. The reader who informs me of his criticism and his suggestions for improvements will not only do a favor to us; he will also better the chances that such conventions as here proposed can become standard among a wider linguistic public. Therefore, I wholeheartedly invite the reader's comments.

1. BASIC IDEA

I shall use the precise term interlinear morphemic translation (IMT) to designate the object of this study. If necessary, the neologism transmorphemization may be used; informally, it will suffice to call it morpheme translation. An IMT is a translation of a text in a language L1 into a string of elements taken from L2 where, ideally, each morpheme of the L1 text is rendered by a morpheme of L2 or a configuration of symbols representing its meaning and where the sequence of the units of the translation corresponds to the sequence of the morphemes which they render. It is called interlinear a) because it should be arranged, typographically, beneath the line of text which it translates, and b) because it is normally used in addition to and before a normal translation.

An IMT is needed whenever it is essential that the reader grasp the grammatical structure of the L1 text but is presumed to be so unfamiliar with L1 that he will not be able to do so merely with the aid of a normal translation and the context in which the text is cited. This means that the necessity of an IMT depends on three factors: 1) the linguistic knowledge of the readership addressed, 2) the language L1 - assuming that L2 is "Standard Average European" -, and 3) the context in which the L1 text is cited.

As for 1), the reader is supposed to be familiar with L1 either if it is also Standard Average European or if he is a philo-

[1] I greatfully acknowledge the suggestions that I have received in discussing this paper with Hansjakob Seiler and my colleagues of UNITYP at Cologne; and I thank Roger Barron for correcting my English grammar and style.

logist of L1. Therefore, the decision for or against IMTs depends partly on whether the publication is to be read only by philologists of L1 or by the general linguistic public. I suggest, therefore, that every general linguistic journal indicate, in its directions for contributors, those languages for which IMTs are desirable by way of exclusion, i.e. by enumerating those for which IMTs seem superfluous.

Pending this, I may turn now to condition 2) and give a provisional list of such latter languages, which I regard as maximal. That means, in non-philological publications, every foreign language text ought to be accompanied by an IMT, with the possible exception of the following languages: English, German, French, Italian, Spanish, Latin, Russian as well as L2 (= L1) and, possibly, languages genetically closely related to L2.

As for condition 3), the context will help the reader either if it contains a complete morphological analysis of the L1 text (which is almost never the case) or if it is a grammar of L1 plus dictionary. This means that IMTs are dispensable in complete grammars; but even here, the reader will be grateful for them.

An IMT will not usually make good sense in L2. This is why it is normally followed by a translation in normal L2 speech; and, moreover, whenever there is an IMT, there is usually no further need to give a narrow translation, so that the non-morphemic translation may then be quite idiomatic. There are certain cases where an IMT, though necessary for whatever particular reasons, differs so little from an idiomatic translation that the average reader can fully understand the meaning of the L1 text with the help of the IMT alone. In such cases, the IMT may be given as the only translation of the L1 text. (There are, of course, often quite similar cases where, though comprehension of the grammatical structure is essential, a normal narrow translation (perhaps a somewhat forced one) may be offered as the only one; these do not concern us here.)

2. PRIMARY AIM

Since an IMT is regularly accompanied by an idiomatic translation, there is no need to make it particularly "readable" by

giving only rough morphemic equivalents which, though not rendering exactly the meaning or function of the L1 morphemes, help to give a good L2 text.

(1) Paul kaufte die Birne, Peter aß sie.
GER "Paul bought the pear, Peter ate it."

In (1), e.g., *sie* may be rendered by *her* in an IMT; it need not be rendered by *it*, since that will appear in the idiomatic translation, anyway.

(2) Wir laufen Schlittschuh. "We skate/are skating."

Similarly, an IMT of (2) might contain *slide-shoe* for *Schlittschuh*. The primary aim of an IMT is to make the grammatical structure of the L1 text transparent. The textual fluency of the IMT by standards of the L2 grammar is a subordinate aim, at best. It may be pursued, for the benefit of the reader, to the extent that the primary aim is satisfied.

But the primary aim is, of course, not absolute. It is clear that there may be acceptable IMTs of (1) and (2) containing *it* and *skates* (noun), respectively. This depends entirely on the purpose for which the L1 text is being studied. Depending on whether there is an interest in pronoun gender with respect to (1) or in nominal composition with respect to (2), the IMTs will or will not contain *her* and *slide-shoe*. There seems no reasonable purpose to be served by imposing, in a general way, the rigid requirement that each morpheme of the L1 text be rendered separately and with full precision. Thus, the degree of detail displayed by an IMT is relative to the purpose it is meant to serve and varies between the upper bound just mentioned and a lower bound where the units of the L1 text rendered separately and with order conservation are words, not morphemes. The upper bound may be reached in general purpose linguistic editions of texts in exotic languages, while the lower bound may be approached by studies concentrating on questions of word order.

3. Rendering lexical and grammatical elements

3.1. Lexical elements

Lexical formatives or lexematic roots of L1 are rendered by lexical formatives of L2. The following points are to be observ-

ed:

1. If L2 is an inflecting language, care must be taken not to use the nominative and the infinitive to render noun and verb roots (or stems) of L1; instead, roots (or stems) of L2 have to be used, simply because the nominative and the infinitive are or may be inflected forms displaying a specific grammatical category not (necessarily) present in the L1 original. Thus, if L2 is German, (3)(i) is wrong and (ii) is right.

(3)(i) Or-e-mus!
LAT beten-KONJ.PRS-1.PL

(ii) Or-e-mus!
bet-KONJ.PRS-1.PL
"Lasset uns beten!" ("Let us pray!)

2. Homonymy must, of course, be resolved in IMTs, leaving the irrelevant meanings unmentioned. But polysemy should not be resolved. It is advisable to regard the lexical morphemes of L1 as context-independent and translate them in the IMT by a cover term, i.e. that one L2 word considered the nearest equivalent to the L1 morpheme qua lexical entry. Thus, Hittite *es-* in (8) is transmorphemized by *eat* and not by *celebrate*. This procedure has two advantages. First, it contributes to making the grammatical structure of the L1 text explicit because it lets the reader see what meaningful elements there are and which aspects of the total meaning of the text are to be attributed to the grammatical construction. Second, it allows giving an identical translation for repeated occurrences of the same lexeme in different contexts, which facilitates the reader's task to identify the elements. For the linguistically trained reader, there is no danger here of obliterating the relation between the IMT and the normal translation.

3. When the L1 lexeme is compound, this may or may not be made explicit by morphological analysis, putting either *Schlittschuh* or *Schlitt-schuh* (or *Schlitt+schuh*, see sect. 4.5. below). In the first case, the composition may, in the second it must be reflected in the IMT; see sect. 4.1. below.

3.2. Grammatical elements

If a grammatical formative of L1 has a sufficiently exact counterpart in L2, this may be used in an IMT. Thus, there is no

problem with the transmorphemization of the free grammatical formatives in (4).

(4) Io non lo trov-o.
ITA I not it find-1.SG
"I do not find it."

But such a simple situation does not always obtain. Quite often there is either no morpheme at all in L2 to correspond to the relevant L1 morpheme, as in English there is none for any of the grammatical formatives in (5).

(5) Time-o ne veni-a-t.
LAT fear-1.SG NEG.VOL come-SBJV.PRS-3.SG
"I am afraid he might come."

Or there is an equivalent, but its distribution is so different that it would be unrecognizable, as might be the case with *-en* (as in *wooden*) in (6).

(6) ferr-eus
LAT iron-en:NOM.SG.M
"iron (adj.)"

Or the existing equivalents are not exact enough, as might be the case with *sie* vs. *her* in (1), because *her* contains a feminine sex component which *sie* does not. Or there is a special emphasis on the constitution of the L1 morpheme in terms of semantic or grammatical components, which, though also present in the L1 counterpart, is not made explicit by it. Thus, in (7), there is often good reason not to render Latin *-t* by English *-s*, but to decompose it in the IMT.

(7) Manu-s manu-m lava-t.
LAT hand-NOM.SG hand-ACC.SG wash-3.SG
"One hand washes the other."

In all these cases, it is necessary or desirable to render an L1 morpheme by a label or configuration of labels taken from some grammatical metalanguage and representing the semantic or functional components. Since such terms have no chance to reappear in an idiomatic translation, since they have, in fact, a different status from common L2 morphemes in IMTs, they are put in upper case in order to facilitate the reader's understanding.

In general, such grammatical category labels are preferable to natural L2 morphemes even in cases like (4), since they are more precise. It is chiefly by this virtue that an IMT can give

grammatical information which becomes distorted in a normal translation. Only by means of category labels as, e.g., HAB, PRF, IMPF, COMPL etc. can we show such aspectual information which either disappears in a Standard Average European translation or has to be hinted at by such clumsy circumlocutions as "habitually, used to, just, finished" etc.[1]

It is an abuse of this device to render a grammatical morpheme by its own capitalization instead of an informative grammatical category label. The widespread custom to put words in upper case in order to represent their meaning has led, even in works of high quality, to such morpheme "translations" where, e.g., *ne* in (4) would be rendered by NE. Such bankruptcy declarations of grammatical analysis should be inadmissible in IMTs.

In what follows, I shall give a provisional list of abbreviations that may be used as grammatical category labels in IMTs.[2] This list has to be accompanied by two caveats. First, it is not intended to be complete, nor could it be, for reasons requiring no explanation. Second, the abbreviations proposed are not necessarily optimal for any individual publication; for if an author needs only part of them, he may go farther in abbreviating without obliterating distinctions, while if he has to extend the list, he may have to lengthen some of the labels in order to guarantee distinctiveness.

3.3. Grammatical category labels

ablative	ABL	active	ACT
absolutive 1. unmarked syntactic function in ergative systems	ABS	adjectivalizer	ADJR
		adverbializer	ADVR
2. non-relational form of the noun	ABSOL	adversative	ADVERS
accusative	ACC	affective	AFFECT
		affirmative	AFF

[1] The reader may convince himself how unhelpful the exclusive use of non-terminological L2 elements in IMTs can be by examining the IMTs offered by Nikolaus Finck in his *Die Haupttypen des Sprachbaus*, Leipzig : Teubner, 1909.

[2] The list is a selective compilation of those in Hansjakob Seiler, *Cahuilla grammar*, Banning : Malki Museum, 1977; Ronald W. Langacker, *Studies in Uto-Aztecan grammar*. Vol. 1: An overview of Uto-Aztecan grammar, Dallas, Tex.: SIL, 1977:8-10; and Christian Lehmann, *Der Relativsatz*, Köln : Institut für Sprachwissenschaft (*akup* 36), 1979:491f.

agentive	AG
alienable	AL
allocutive	ALL
animate	AN
antipassive	ANT
applicative	APPL
(article	ART)
(aspect	ASP)
assertive	ASS
attributor	AT
augmentative	AUG
auxiliary	AUX
benefactive	BEN
causative	CAUS
circumstantial	CIRC
classifier	CLFR
comitative	COM
complementizer	CMPLR
completive	COMPL
concessive	CONC
conditional	COND
conjunction	CNJ
connective	CONN
continuative/-ous	CONT
copula	COP
correlative	CORR
dative	DAT
declarative	DECL
definite	DEF
deictic of 1.ps.	D1
2.ps.	D2
3.ps.	D3
demonstrative	DEM
desiderative	DESID
different subject	DS
diminutive	DIMIN
directional	DIR
distal	DIST
distributive	DISTR
dual	DU
dubitative	DUB
durative	DUR
emphatic	EMPH
ergative	ERG
evidential	EVID
exclamation	EXCLM
exclusive	EXCL
exist(ence)	EXIST
factitive	FACT
feminine	F
first person	1
focus	FOC
frequentative	FREQ
future	FUT
genitive	GEN
gerund	GER
habitual	HAB
honorific	HON
hortative	HORT
human	HUM
imperative	IMP
imperfect(ive)	IMPF
impersonal	IMPRS
inalienable	INAL
inanimate	INAN
inceptive	INCEPT
inchoative	INCH
inclusive	INCL
indefinite	IND
inferential	INFR
infinitive	INF
injunctive	INJ
instrumental	INST
intensifier	INTNS
interrogative	INT
intransitive	ITR
invisible	INVIS

irreal	IRR
jussive	JUSS
locative	LOC
manner	MAN
masculine	M
modal	MDL
modifier	MOD
narrative	NARR
negative	NEG
neuter	N
nominalizer	NR
nominative	NOM
non-human	NH
noun-class *n*	CL*n*
numeral classifier	CLFR
object (personal verb affix)	OBJ
obligational	OBLIG
oblique	OBL
optative	OPT
participle	PART
particle	PTL
partitive	PRTV
passive	PASS
past	PAST, PRT
pejorative	PEJ
perfect(ive)	PRF
(person	PS)
plural	PL
possessive	POSS
potential	POT
present	PRS
preterite	PRT
progressive	PROG
proximal	PROX
question	INT
quotative	QUOT
realized	REAL
reciprocal	REC
reduplication	RDP
referential	REF
reflexive	REFL
relative particle, pronoun, affix	REL
remote	REM
resultative	RSLTV
same subject	SS
second person	2
singular	SG
specific	SPEC
stative	STAT
subject (personal verb affix)	SBJ
subjunctive	SBJV
subordinator	SR
substantivizer	SBSTR
temporal	TEMP
third person	3
topic	TOP
transitive	TR
verbalizer	VR
visible	VIS
vocative	VOC
volitional	VOL

4. Boundaries between units

4.1. Word and morpheme boundaries

Next, we need conventions for the graphic separation of the text units. The obvious thing to begin with is the separation of

the L1 words by blank spaces. Thus we require that there be a one-to-one correspondence between the spaces in the L1 text and those in the IMT, i.e. that a blank space be left in an IMT if and only if there is a corresponding blank space in the L1 text. The reason for this and the following similar convention is, evidently, to facilitate the reader's task of mapping the L1 units onto the IMT units.

(8) n-an apedani mehuni essandu.
HIT CONN-him that:$^{DAT}_{SG}$ time:$^{DAT}_{SG}$ they shall eat
"They shall celebrate him on that date."

In (8), where this convention has been violated, the reader has no mechanical way to do the mapping (unless there are other graphic aids as will be discussed in sect. 5).

An analogous convention can be established with respect to the symbols used for separating bound morphemes. Suppose hyphens are used for this purpose. Then we require that a hyphen be put in an IMT if and only if there is a corresponding hyphen in the L1 text. This convention has often been neglected in the literature, but its usefulness becomes apparent when we look at a version of (6) that does not conform to it:

(6') ferr-eus
iron-en-NOM.SG.M

For the innocent reader, there is no hint in (6') as to whether the suffix rendered by *-en* is located before or after the hyphen in the Latin text (more difficult examples are conceivable).

The problem illustrated by (6') and (8) could be remedied by accepting the following convention: There is a boundary symbol in an IMT if and only if there is a corresponding boundary symbol in the L1 text; i.e., there is a biunique mapping of units between the L1 text and the IMT.

But it happens that this, though rather straightforward at first sight, would be no reasonable convention to accept. For all purposes related to levels of grammar above morphophonemics, there is usually a much greater interest in the morphemic composition of the L1 units as such than in the exact location of the morpheme boundaries. That is, it often suffices to know which morphemes there are, and in which order, without knowing their

(morpho-)phonemic shape. Second, in highly fusional languages, it is often plainly impossible to indicate morpheme boundaries (in "surface" forms). Thus one might, with good reason, not care to take a stand on the question as to where exactly the boundary between the second and the third morphemes in (6') is located. Third, it might seem desirable to introduce, into a rather "broad" IMT as (8), auxiliary units which would be necessary in an idiomatic L2 translation and which serve, therefore, to make the IMT more readable. And last, there are several languages with a well established orthography which one would like to use in the reproduction of a (L1) text, but which does not lend itself easily to the introduction of morpheme boundary symbols, since it already makes extensive use of hyphens, or is not directly related to morphophonemics, or its extension by such symbols would be simply unusual or unesthetic.

The result of all this is that it is perfectly normal for an IMT to contain more separate units than appear separate in the corresponding L1 text. We should, therefore, establish a convention like the following: Each unit of the L1 text is rendered by at least one unit in the IMT. If there is a boundary symbol in the L1 text, there must be a corresponding boundary symbol in the IMT. It is clear that this convention does not allow for separations in L1 texts which have no counterpart in the IMT, i.e. it does not allow to render two units by one. Situations in which this might seem desirable arise with some frequency in languages with excessive compounding (e.g. Chinese), but yielding them would be against the spirit of the IMT and, in fact, against the spirit of linguistics: an analysis which is not to be accounted for should not be made in the first place. The same must be said about L1 elements that have no meaning and thus no counterpart in L2, e.g. empty derivative affixes or submorphemic units like connecting vowels and the like. If it is necessary to separate such an element in an L1 text, either its function has to be indicated (say, by CONN) or it has to be explicitly rendered by $\emptyset$.

The existence of such one-to-many correspondences between the L1 text and its IMT as mentioned above requires that we introduce special boundary symbols which do not appear in L1 texts and whose appearance therefore signals that such a situation ob-

tains. Here, the following cases have to be kept apart: 1) There may be distinct sequential morphemes in the L1 text which are rendered by separate units in the IMT, but morpheme boundaries are, for one or other of the reasons mentioned, not shown in the text. 2) The IMT contains an element whose morphemic equivalent in the L1 text is not merely not shown separately, but does not exist. 3) There is an element in the IMT which does have a counterpart in the L1 text, but it is not a sequential morpheme so that the L1 text cannot be parsed. 4) One L1 morpheme is rendered by a configuration of units, normally grammatical category labels, in the IMT. Let us consider each of these cases in turn.

4.2. Morpheme boundaries not shown in the L1 text

The problem of a sequence of units in an IMT which correspond jointly to a string smaller than or equal to a word in the L1 text and not morphemically analyzed has already arisen and been taken account of in examples (6) and (8). There, the colon was used to join such units. In that case, the units were morphemes. Can the same device be used to remedy the deficiencies remaining in (8), where the surplus elements in the IMT are words? Let us assume that all the elements in the phrase *they shall eat* have their morphemic equivalents in the Hittite text. I suggest that we use the colon here too. This has the disadvantage that one is not accustomed to see one's L2 words joined by colons, but has the advantage of not unduly augmenting our symbol inventory. We thus put a colon in (8), observing, at the same time, morpheme order correspondence:

(8') n-an apedani mehuni essandu.
CONN-him that:$^{\text{DAT}}_{\text{SG}}$ time:$^{\text{DAT}}_{\text{SG}}$ eat:they:shall

We thus establish the following convention: If there are separate elements in an IMT - no matter whether they are morphemes, grammatical category symbols or words - which do have distinct morphemic counterparts in the L1 text but the latter are not separated, such morpheme boundaries shown in the IMT but not in the L1 text are represented by a colon.

4.3. Elements without significans in the L1 text

The situation where an IMT contains an element which does

not have a distinct morphemic counterpart in the L1 text arises mainly with zero morphemes and unmarked grammatical categories. How do we transmorphemize, e.g., a nominative or absolutive case which is morphologically zero? There are two main solutions here. Either, we may follow markedness theory and simply recognize no morpheme at all in such cases. The maxim would then be: what is not there in an L1 text is not rendered in the IMT. Thus, a possible IMT of (9) might run as follows:

(9) ŋatjulu-ɭu ŋa yankiri pantu-ɳu.
WAL I-ERG SBJ.1 emu spear-PRT
"I speared the emu."

In this solution, the transmorphemization problem does not arise. Or we may recognize a morpheme with zero expression whose meaning is indicated in the IMT. Then we have the following alternative:

(9') ŋatjulu-ɭu Ø-ŋa yankiri pantu-ɳu.[1]
I-ERG AUX-SBJ.1 emu-ABS spear-PRT

(9'') ŋatjulu-ɭu ŋa yankiri pantu-ɳu.
I-ERG (AUX) SBJ.1 emu (ABS) spear-PRT

While the two main solutions are theoretically distinct, so that the choice between them is not simply a matter of adequate IMT, the two alternatives available within the second case are notational variants. Those who do not want to overload their L1 texts with matters of linguistic analysis will prefer the parenthesis version exhibited in (9''). If we wanted to render, e.g., the present tense of the verb in (7), we might prefer the parenthesis notation in the IMT, as in (7'), to the indication of a zero in the Latin text:

(7') Manu-s manu-m lava-t.
hand-$^{NOM}_{SG}$ hand-$^{NOM}_{SG}$ wash (PRS)-$^{3}_{SG}$

Disregarding the markedness solution, which does not pertain to writing IMTs, we can now establish a convention for zero morphemes: If there is an element in an IMT which has no significans in the L1 text, it is put between parentheses. Zero morphemes may, alternatively, be represented by 'Ø' in the L1 text and treated as regular morphemes in the IMT.

[1] The personal clitics are suffixed to auxiliaries expressing tense if such are overtly present.

4.4. Non-sequential morphemes

Parsing the L1 text is near to impossible whenever significantia are not in linear order but are superimposed to one another. This regards internal modifications of morphemes, for instance tone change or vowel alternations like apophony. Normally authors will refrain from the attempt of analyzing such morphophonemic structure in the L1 text, but some may like to distinguish such processes from affixation in the IMT. We therefore introduce a special boundary symbol for IMT configurations rendering such alternations, and choose the arrow, which is reminiscent of the American structuralists' 'replacive morpheme':

(10) cep-i
LAT catch→PRF-1.SG
"I caught"

The convention proposed reads: If a grammatical meaning is expressed by internal modification of a morpheme in the L1 text, the IMT contains first the counterpart of the affected morpheme, then a rightward arrow, then the elements representing the meaning of the grammatical process.

4.5. Configurations of grammatical category labels

The situation where one L1 morpheme is rendered by a configuration of units in the IMT arises whenever its semantic and/or grammatical components are to be represented in an IMT. Normally, the grammatical category labels introduced in sect. 3.2. above will appear here. First, as for their arrangement, linear ordering among elements is absent in meaning decomposition. They may, therefore, be arranged below each other, as shown in (8) and (7'). This seems entirely natural, since they correspond to those widely recognized semantic or grammatical features for which such an arrangement is common enough. However, depending on considerations of space disposition, they may, alternatively, be arranged linearly. In neither case is their ordering relevant.

Second, as for the boundary symbol used to separate the units of such a configuration, there should be no separation at all between such units because they do not correspond separately to any morphosyntactic unit in the L1 text. Therefore, if they are arranged below each other, a boundary symbol can be dispensed with. If they are arranged linearly, the best symbol to sepa-

rate them which does not signal a boundary is the period, because - for reasons to be discussed in sect. 5 - such grammatical category labels will normally be abbreviated without a period at the end. Thus, the period used to separate them may be taken for the period we normally expect after abbreviations. This use has been observed in the examples. It pertains also to those rare cases in which the semantic components of a morpheme are not rendered exclusively by grammatical category labels, as in (11).

(11)

LAT	nolu-mus NEG want(PRS)-1 PL	=	nolu-mus NEG.want(PRS)-1.PL "we do not want"

We arrive, therefore, at the following convention: If elements representing semantic or grammatical components of an L1 morpheme are arranged below each other, there is no boundary symbol between them; if they are arranged in line, a period is put between them.

In order to illustrate the difference between the various boundary symbols introduced so far, let us consider their applicability in one and the same example:

(12)

LAT	(i) puer boy-NOM.SG	(iv) puer boy → NOM.SG
	(ii) puer boy:NOM.SG	(v) puer boy.NOM.SG
	(iii) puer boy(NOM.SG)	

With the exception of the first one, these alternative IMTs imply different claims as to the morphological structure of Latin *puer*. (12)(i) is notationally wrong and uninterpretable, because a hyphen may be put in an IMT only if there is a corresponding morpheme division in the L1 text equally represented by a hyphen. (ii) implies that *puer* could be parsed into a sequence of two overt morphemes - factually wrong. (iii) implies that *puer* is paradigmatically nom.sg. without having an overt expression for this category - correct. (iv) implies that *puer* contains a replacive morpheme, some sort of alternation, expressing the nom.sg. - wrong. And (v) implies that *puer* is not amenable to morphological analysis and can only be semantically decomposed into the components indicated - wrong.

4.5. Kinds of juncture

By now, the only morpheme boundary symbol used in L1 texts is the hyphen. One might ask whether it would not be desirable to distinguish various kinds of morpheme boundaries. We might wish, e.g., to distinguish word formation from inflexion, and possibly composition from derivation. This can be done as in the following example.

(13) Auto+fahr=er-n
GER car+drive=AG-DAT.PL
"to car drivers"

I suggest that we make the use of such refinements optional, because we cannot foresee which kinds of morpheme boundaries an author will want to distinguish in his L1 or whether he will need any such distinctions at all for his purpose. Thus, I propose the following convention: The hyphen as a general purpose morpheme boundary symbol may be partly substituted, in L1 texts and likewise in IMTs, by symbols signalling specific kinds of morpheme concatenation, e.g. composition, derivation or cliticization. The use of the plus sign for composition and the equal sign for derivation is recommended.

4.6. Infixes and circumfixes

A special problem is posed by infixes and circumfixes, because the discontinuous L1 morphemes involved here will only by rare coincidence be able to be rendered by discontinuous L2 morphemes. Normally one L2 morpheme or a (configuration of) grammatical category label(s) will correspond to the discontinuous morpheme. This means that in this one case we are forced to violate the principle stated in sect. 4.1. that each unit of the L1 text is rendered by at least one unit in the IMT. In order to forestall misunderstandings, such violations must be marked off by special boundary symbols. I suggest that we proceed as in the following examples.

(14) a‹wá›?u
DAK bring‹($^{\text{OBJ}}_{3}$) $^{\text{SBJ}}_{1}$›
"I am bringing it"

(15) ge›seh‹en
GER ›$^{\text{PRT}}_{\text{PART}}$‹see
"seen"

The conventions proposed are, accordingly: Infixes as well as the elements rendering them in an IMT are enclosed in angled brackets (outward-directed arrowheads) as their boundary symbols. In the IMT, the equivalent of the split-up morpheme precedes the equivalent of the infix. Circumfixes as well as the elements rendering them in an IMT are set off by interchanged angled brackets (inward-directed arrowheads) as their boundary symbols. In the IMT, the equivalent of the circumfix precedes the equivalent of the enclosed sequence.

4.7. Constituent structure

A word must be said about syntax in IMTs. The primary aim of an IMT is to show the morphemes and their sequence. It cannot make explicit the syntactic construction, i.e. it cannot replace a constituent structure diagram. Naturally, nothing prevents the use of labelled brackets in L1 texts or in IMTs; but this device is typographically cumbersome and space-consuming. One thing one can recommend is the use of different syntactic boundary symbols for different constituents, which dispenses the indexes. We have such symbols as [ ], # #, / /, // //, ¤ ¤, and maybe some more at our disposal, but extensive use of them is impossible without incommodating the reader. Therefore, they are less recommendable in general purpose works and more in works where all text examples illustrate a restricted set of syntactical problems. Incidentally, such symbolism need not be used both in L1 texts and in IMTs since the mapping of units is already guaranteed by all the other devices. In accordance with the general tendency to place grammatical information in IMTs and not to disfigure L1 texts with it, it seems reasonable to show syntactic structure exclusively in IMTs. Compare example (16), where clauses are enclosed in [ ] and NPs in # #.

(16) o filho d-o professor que morr-eu
POR $\#^{\text{DEF}}_{\text{M}}$ son of-$\#^{\text{DEF}}_{\text{M}}$ teacher [REL die-$^{\text{PRT}}_{\text{3.SG}}$] ##
"the son of the (that) teacher who died"

The same reservations must be made about showing constituency within words. I think we would overdo IMTs if we displayed morpheme and position classes, say by labels as STEM, PREFIX 1, PREFIX 2 etc. IMTs are not meant to fill the place of grammars.

4.8. Orthography

Finally, two restrictions should be imposed on orthography. First, syllabication is inadmissible in morphemically analyzed L1 texts, because the hyphen means morpheme boundary. This is especially relevant if the L1 text is a transcription of a syllabic script: the syllabic transcription favored by philologists of such languages is irreconcilable with morphemic analysis.

Second, there is no punctuation in IMTs, for the obvious reason that there are no natural language sentences and clauses in IMTs.

5. TYPOGRAPHIC PRESENTATION

With the conventions so far established, an unambiguous mapping of each element of the L1 text onto one or more elements of the IMT is guaranteed. But it would still be a tiresome procedure for the reader if the IMT were not arranged in the line immediately beneath the corresponding L1 text line. A minimum requirement would be the following: If there is, in an IMT, an equivalent to an element of an L1 text line, it is contained in the line immediately below that line.

The reason why this rather obvious convention is so often neglected in current IMTs is that a well-done IMT line is generally much longer than the corresponding L1 text line, because the rendering elements are generally longer than the rendered ones. The following conventions may be suggested in order to alleviate this problem:

1. IMTs are composed in a smaller type than L1 texts.
2. Grammatical category labels are generally abbreviated, without a period at the end.
3. Several such elements belonging to one configuration (which corresponds to one L1 morpheme) may be arranged below each other.
4. If such components consist of a specific and a generic category, the generic one is omitted from the IMT. We do not write, e.g., COMPL.ASP (completive aspect) or 1.PS (first person), but simply COMPL and 1.

Given that we can, following these conventions, attain a

reasonable approximation of the length of an IMT line to that of the corresponding L1 text line, we may think of imposing the following, much stricter convention, which entails the above-mentioned minimum requirement: Each word of an L1 text line has the set of elements of an IMT rendering it centered below itself.

It is clear that this greatly facilitates the mapping task to be fulfilled by the reader. Several comments about this convention are in order. First, given the unequal length of the two lines in question, it is clear that this convention requires introducing additional blank spaces between words, mostly in the L1 text line. Even more of such additional blanks would be needed if we required, alternatively, that each of the L1 text words be arranged left-flush with the set of elements rendering it. This would reduce the amount of text placeable in one line and render a more discontinuous text, which is a disadvantage for various reasons. This disadvantage would be even more blatant if we restricted the convention still more by requiring that not the words but the morphemes should be left-flush with the elements rendering them; or, what amounts to the same, that each boundary symbol have the boundary symbol corresponding to it arranged exactly below itself. It is questionable whether such refinements really serve the interest of the reader, because even if the task of mapping between L1 text elements and IMT elements were thus facilitated, a fluent reading of the L1 text would be impeded. Compare the following four versions of the same IMT, whose first version fulfills the minimum requirement, while the second has centering of words, the third left-flushness of words and the fourth left-flushness of morphemes.

(17) (i) Esa-ida-zu zer ari z-era-n.
BAS say(IMP)-Dat.1-ERG.2 [what do ABS.2-AUX-NR]

(ii) Esa-ida-zu zer ari z-era-n.
say(IMP)-$^{DAT}_{1}$-$^{ERG}_{2}$ [what do $^{ABS}_{2}$-AUX-NR]

(iii) Esa-ida-zu zer ari z-era-n.
say(IMP)-$^{DAT}_{1}$-$^{ERG}_{2}$ [what do $^{ABS}_{2}$-AUX-NR]

(iv) Esa -ida-zu zer ari z- era-n.
say(IMP)-$^{DAT}_{1}$-$^{ERG}_{2}$ [what do $^{ABS}_{2}$-AUX-NR]
"Tell me what you are doing!"

While I can see a sensible improvement from (i) to (ii), (iii) and (iv) do not seem to carry further advantages with them.

But there is a different factor involved, too, which is the cost of production of such sophisticated IMTs, both from the point of view of the author who has to type the manuscript and, possibly, from that of the type setter. It seems to me that while version (i) is easiest to produce, version (ii) should be most difficult. We might, therefore, leave it to the author to decide whether he prefers to have his L1 text words centered [(ii)] or left-flush [(iii)] with the units rendering them. Even this might prove impossible to be agreed upon, for economic reasons.

There is one final convention to be established whose usefulness is, apparently, not so obvious that it should not have been frequently disregarded in the literature: The distance between an L1 text line and the line immediately preceding it is greater than that between it and the IMT line belonging to it. This somewhat awkward formulation takes into account the possibility that the line following an IMT line may be not the next L1 text line but a free translation still referring to the preceding.

6. Conventions restated

The following conventions, which emerge as the result of the preceding discussion, are hereby submitted to the specialists for judgement on their possible integration in a general standardization of IMTs:

I. An interlinear morphemic translation (IMT) is a translation of a text in a language L1 into a string of elements taken from a language L2, where, ideally, each morpheme of the L1 text is rendered by a morpheme of L2 or a configuration of symbols representing its meaning, and where the sequence of the units of the translation corresponds to the sequence of the morphemes which they render.

II. In non-philological publications, every foreign language text ought to be accompanied by an IMT, with the possible exception of the following languages: English, German, French, Italian, Spanish, Latin, Russian as well as L2

(= L1) and, possibly, languages genetically closely related to L2.

III. The primary aim of an IMT is to make the grammatical structure of the L1 text transparent.

IV. The degree of detail displayed by an IMT depends on the purpose it is meant to serve.

V. Lexical formatives of L1 are rendered by lexical formatives of L2.

VI. L1 roots (or stems) are not rendered by inflected (nominative or infinitive) forms of L2, but by roots (or stems).

VII. Homonymy is resolved in IMTs, polysemy is not. A polysemous L1 lexeme is constantly rendered by its nearest context-independent L2 equivalent.

VIII. If a grammatical formative of L1 has a sufficiently exact counterpart in L2, this may be used to render it in an IMT. In general, it is preferable to render a grammatical formative of L1 by a label or configuration of labels taken from some grammatical metalanguage and representing the semantic or grammatical components. Such grammatical category labels are put in upper case.

IX. In L1 texts and in IMTs, the word boundary symbol is the blank space, and the principal morpheme boundary symbol used is the hyphen.

X. Each unit of the L1 text is rendered by at least one unit in the IMT. If there is a boundary symbol in the L1 text, there is a corresponding boundary symbol in the IMT. (In particular: There is a blank space, hyphen, plus sign, equal sign or angled bracket in an IMT if and only if there is an identical symbol in the L1 text corresponding to it.)

XI. If there are separate elements in an IMT - no matter whether they are morphemes, grammatical category symbols or words - which do have distinct morphemic counterparts in the L1 text but the latter are not separated, such morpheme boundaries shown in the IMT but not in the L1 text are represented by a colon.

XII. If there is an element in an IMT which has no significans in the L1 text, it is put between parentheses. Zero morphemes may, alternatively, be represented by 'Ø' in the L1 text and treated as regular morphemes in the IMT.

XIII. If a grammatical meaning is expressed by internal modification of a morpheme in the L1 text, the IMT contains first the counterpart of the affected morpheme, then a rightward arrow, then the elements representing the meaning of the grammatical process.

XIV. If elements representing semantic or grammatical components of an L1 morpheme are arranged below each other, there is no boundary symbol between them; if they are arranged in line, a period is put between them.

XV. The hyphen as a general purpose morpheme boundary symbol may be partly substituted, in L1 texts and likewise in IMTs, by symbols signalling specific kinds of morpheme concatenation, e.g. composition, derivation or cliticization. The use of the plus sign for composition and the equal sign for derivation is recommended.

XVI. If constituent structure is to be displayed, it is recommended to insert different syntactic boundary symbols for different constituent types in the IMT.

XVII. Syllabication is inadmissible in morphemically analyzed L1 texts.

XVIII. There is no punctuation in IMTs.

XIX. For each pair of an L1 text word and the set of elements rendering it, the latter is arranged below the former in such a way that the boundaries of the shorter complex are, on a vertical line, within the boundaries of the longer one. Ideally, the two are centered; the greatest admissible deviation is to put them left-flush. If such an arrangement is impossible, the following minimum requirement must be observed: If there is, in an IMT, an equivalent to an element of an L1 text line, it is contained in the line immediately below that line.

XX. 1. IMTs are composed in a smaller type than L1 texts.

2. Grammatical category labels are generally abbreviated, without a period at the end.

3. Several such elements belonging to one configuration (which corresponds to one L1 morpheme) may be arranged below each other.

4. If such components consist of a specific and a generic category, the generic one is omitted from the IMT.

XXI. The distance between an L1 text line and the line immediately preceding it is greater than that between it and the IMT line belonging to it.

ABBREVIATIONS

The following abbreviations of language names have been used:

BAS Basque	ITA Italian
DAK Dakota	LAT Latin
GER German	POR Portuguese
HIT Hittite	WAL Walbiri

Die Arbeitspapiere des Instituts für Sprachwissenschaft der Universität Köln erscheinen seit September 1968 in unregelmäßigen Abständen. Die mit einem Stern bezeichneten Arbeitspapiere sind noch vorrätig.

1. Seiler, H. 1968, Grundzüge der Allgemeinen Sprachwissenschaft. Erster Teil: Generative Grammatik. Ausarbeitung der Vorlesung SS 1967, 147 S
2. 1969, Zur Gestaltung eines Studienführers für Studenten der Sprachwissenschaft unter Berücksichtigung einer sprachwissenschaftlichen Grundausbildung für Studenten benachbarter Disziplinen, 5 S
3. Seiler, H., Scheffcyk, A. 1969, Die Sprechsituation in Linguistik und Kommunikationswissenschaft. Referat einer Diskussion, 12 S
4. Katičić, R., Blümel, W. 1969, Die sprachliche Zeit, 12 S

*5. Brettschneider, G. 1969, Das Aufstellen einer morphophonemischen Kartei (illustriert an der Morphophonemik des japanischen Verbs) 43 S

6. Penčev, J. 1969, Einige semantische Besonderheiten der bulgarischen Geschmacksadjektive, 17 S
7. Seiler, H. 1969, Zur Problematik des Verbalaspekts, 19 S
8. Gottwald, K. 1970, Auswahlbibliographie zur Kontrastiven Linguistik, 42 S
9. Ibañez, R. 1970, Emphase und der Bereich der Negation Satz- vs. Satzgliednegation, 9 S
10. Penčev, J. 1970, Die reflexiven, medialen und passiven Sätze im Bulgarischen, 53 S
11. Untermann, J. 1970, Protokoll eines Kolloquiums über die Situation des Faches Indogermanistik, veranstaltet auf Einladung des Instituts für Sprachwissenschaft, Köln am 30.1.70, 11,15 - 13,00 Uhr, 26 S
12. Seiler, H. 1970, Abstract Structures for Moods in Greek 18 S
13. Bäcker, J. 1970, Untersuchungen zum Phonemsystem und zur Nominalflexion im Litauischen (unter besonderer Berücksichtigung des Akzentwechsels in der Nominalflexion) 13 S
14. Rosenkranz, B. 1970, Georg von der Gabelentz und die Junggrammatische Schule, 14 S
15. Samuelsdorff, P. 1971, Problems of English-German Automatic Translation, 16 S
16. Rosenkranz, B. 1971, Zur Entstehungsgeschichte der idg. Verbalflexion, 30 S
17. Babiniotis, G. 1971, Phonologische Betrachtungen zum Wandel a zu e im Ionisch-Attischen, 40 S
18. Seiler, H. 1971, Possessivität und Universalien. Zwei Vorträge gehalten im Dezember 1971: I. Zum Problem der Possessivität im Cahuilla (Uto-Aztekisch, Südkalifornien) II. Possessivität und Universalien

19. Maas, U. 1972, Semantik für Sprechakte

20. Seiler, H. 1972, Zum Problem der sprachlichen Possessivität

21. Leys, O. 1972, Nicht-referentielle Nominalphrasen

22. Pisarkowa, K. 1973, Possessivität als Bestandteil des polnischen Sprachsystems, 23 S

*23. Brettschneider, G., Lehmann, Ch. 1974, Der Schlagwortkatalog des Instituts für Sprachwissenschaft d. Univ. Köln, 32 S

*24. Wiesemann, U. 1974, Time Distictions in Kaingang

25. Untermann, J. 1975, Etymologie und Wortgeschichte

*26. Seiler, H. u.a. 1975, Deskriptive und etikettierende Benennung; Relativkonstruktionen, 155 S (Becker, Katz, Walter, Habel, Schwendy, Kirsch, Clasen, Seip)

27. Lehmann, Ch. 1975, Sprache und Musik in einem Schumann/Heine-Lied, 14 S

28. Stephany, U. 1975, Linguistic and Extralibguistic Factors in the Interpretation of Children's Early Utterances, 28 S

29. van den Boom, H., Samuelsdorff, P. 1976, Aspects-Kommentar Protokolle eines Seminars aus dem WS 1975/76

30. Walter, H. 1976, Gapping, Wortstellung und Direktionalitätshypothese

31. Ojo, V. 1976, Linguistische und soziolinguistische Aspekte der Entlehnung

32. 1976, Diskussion von Roman Jakobson mit Professoren u. Studenten der Uni Köln, 18 S

33. Samuelsdorff, P. 1977, On Describing Determination in a Montague Grammar, 12 S

*34. Auer, P., Kuhn, W. 1977, Implikative Universalien, linguistische Prinzipien und Sprachtypologie, 21 S

35. Lehmann, Ch. 1978, Der Relativsatz im Persischen und Deutschen; ein funktional-kontrastiver Vergleich, 22 S

*36. Stephany, U. 1978, The Modality Constituent - A Neglected Area in the Study of First Language Acquisition, 20 S