Hans-Heinrich Lieb  
Sebastian Drude

**Advanced Glossing:**  
* A Language Documentation Format  
(Working Paper, November 2000)

**Table of Contents**

1 **Introduction**  
   1.1 Requirements for a language documentation format 3  
   1.2 Typological Glossing and Advanced Glossing 3  
   1.3 Terminological remarks 4  
   1.4 Methodological remark 4  
   1.5 Presentation of Advanced Glossing 5

2 **Advanced Glossing**  
   2.1 Syntactic glossings: Table 1 5  
   2.2 Notations used in Table 1 5  
   2.3 Morphological Glossings: Table 2 6  
   2.4 Notations used in Table 2 6

3 **General comments**  
   3.1 Shared features of the syntactic and morphological glossing tables 6  
   3.2 Incomplete glossings 7  
   3.3 Nature of glossings 7

4 **The syntactic glossing table (1): Representing basic information (Lines I to IX, XII, and XIII)**  
   4.1 Line I: Number and order of phonological words 8  
   4.2 Lines II and III: Segmental phonetic form. Phonetic intonation 8  
   4.3 Lines IV and V: Phonological words. Phonological intonation 8  
   4.4 Line VI: Orthographic representation of phonological words 9  
   4.5 Lines VII and VIII: Word categories and word form categories 10  
   4.6 Line IX: Meanings and semantic effects 10  
      a. Lexical meanings 10  
      b. Semantic effects: the auxiliary part of complex word forms 11  
      c. Many-word word forms without an auxiliary part 11  
      d. Semantic effects: particles 12  
   4.7 Line XII: Rendering of the sentence in an established orthography 12  
   4.8 Line XIII: Sentence meaning paraphrases 12
5 The syntactic glossing table (2): Representing structural and relational information (Lines X and XI) 13
5.1 General remarks 13
5.2 Line X: Surface constituent structure 14
5.3 Line XI: Relational information 14

6 The morphological glossing table 15
6.1 General remarks 15
6.2 Lines I to V: Number and order of morphs — Segmental phonological form — Phonological intonation — Morphs — Morphological intonation 16
6.3 Lines VI to VIII: Orthographic representation of morphs — Stem and morpheme categories — Stem form and morpheme form categories 17
6.4 Line IX: Meanings and semantic effects 17
6.5 Line X: Representing structural information 18
   a. Overall character of Line X 18
   b. Dealing with morphological discontinuity 19
6.6 Line XI: Relational information 19
6.7 Lines XII and XIII: Rendering the phonological word in an established orthography — Word meaning 20

Appendix 1. Table 1: Syntactic Glossing Table 21
Appendix 2. Table 2: Morphological Glossing Table 22
Appendix 3. Overall structure of the format 23
   1: A glossing table 23
   2: A glossing 24
   3: Morphological and syntactic glossings of a sentence 24
   4: A documentation of a text 25
Informally, *unübersichtlichen* modifies *die probleme*.

Information on the nucleus or head relation in the sentence was represented already in Line X; from a systematic point of view, it should appear in Line XI. If another sentence had been chosen, we might have had entries on other relations, too. In most if not all cases, such entries could be constructed on the basis of information contained in Lines VII to X provided Line VII also contains entries on valency categories.

Even relational ambiguity is not beyond the format adopted for Line XI: A single sequence of digit sequences may combine with different relation names.

Once again, the relations assumed for Line XI in a given syntactic glossing table are theory dependent, but this does not hold of the line format itself; whatever the relations assumed, as long as these are 'surface relations' in a traditional sense whose occurrences are coded by means of numerals as indicated, the line format remains unaffected.

### 6 The morphological glossing table

#### 6.1 General remarks

Table 2 is arranged in a form strictly analogous to Table 1, and many explanations for Table 1 simply carry over to Table 2. Lines VI to XIII of a morphological glossing table are strictly analogous to the corresponding lines of a syntactic one; the syntactic entities in the syntactic table are simply replaced by corresponding morphological ones.

There is, however, a major difference in Lines II and III: In Table 1, the entries in Lines II and III are phonetic; in Table 2, they are phonological. This is due to the fact that our morphological glossing is for a phonological not a phonetic word, is for a word listed in Line VI of Table 1. The phonological words that figure in a syntactic glossing table are subjected to morphological not to phonetic glossing in a morphological glossing table.

It may seem that this prevents us from providing phonetic information on phonological words when only individual phonological words, outside a sentence context, are available, for example, words contained in a word list. But suppose that we are dealing with raw data that appear to be realizations not of sentences but of individual words. Even in this case, what we are really confronted with is realizations of elliptic sentences. For example, a single word may be realized in answering a question such as, "What is this word?" The answer elliptically realizes the sentence (in English): "This word is . . . " We may well be interested only in the phonetic and morphological properties of the phonological word that is realized. We would then fill in only the first four lines of a one-column syntactic glossing table, and all lines of a corresponding morphological one.

This characterises the documentation situation from a systematic point of view, touching on various rather subtle and controversial points in phonology. Obviously, in such a documentation situation, shortcuts may and will be used.

The second phonological word in Table 1, *unübersichtlichen*, is chosen as an example for the morphological Table 2. Individual correspondences between the two tables will as a rule not be pointed out.
6.2 Lines I to V: Number and order of morphs — Segmental phonological form — Phonological intonation — Morphs — Morphological intonation

Line I is on the number and order of morphs in *unübersichtlichen*, specified by Arabic numerals; there are five morphs.

Line II gives the syllabified sound sequence of *unübersichtlichen*, and Line III its (phonological) intonation. Lines II and III are jointly equivalent to the entry in cell IV.2 of Table 1 which names the phonological word in phonological notation; Lines II and III differ from this entry only by making the intonational properties of *unübersichtlichen* explicit. It may come as a surprise that in German, too, pitch contours are assumed to be a major part of word intonations. If different assumptions are made, the content of Line III changes accordingly. Making the assumptions on German happens to have a fortunate by-effect for our presentation: It also demonstrates how, in a phonological word of a tone language, tones would be represented explicitly by naming pitches (level pitches or glides).

Lines II and III of a morphological glossing table contain only information that would already be represented in a corresponding syntactic glossing table, even though Line III makes explicit the intonational properties of the phonological word. This may allow for shortcuts in filling in the table.

Line IV differs from Lines II and III in naming not the phonological word but the individual morphs, as indicated by means of slashes before and after each entry in Line IV. (Using slashes in Line IV in this way is vital from a systematic point of view. Naturally, such repetitive features of an entry are obvious candidates for automatization.) It should be noted that the entries in Line IV specify morphs completely, including their intonational properties. In a tone language the word stress symbols would be replaced by tone symbols. So-called free tones in a tone language may be construed — in agreement with their treatment in Autosegmental Phonology — as morphs without a sound sequence and syllable structure, and represented by means of a separate column that has a bar symbol in various lines.

Line III, which gives the pitch contour of the intonation of the phonological word, is not yet sufficient to specify the pitch properties of all morphs named in Line IV. For example, the representation of the *sicht*-part of *unübersichtlichen* in Line II is marked in Line III by "H,L", indicating secondary word stress, whereas the morph *sicht* in Line IV has the symbol for primary word stress. It is only in the context of the entire phonological word that the pitch for primary word stress (H) is replaced by the pitch characterization of secondary word stress (H,L). This shows that the pitch contour of the intonation of *un über sicht lichen* — a "morphological word" — is obtained from the pitch contours of the individual morphs but need not be identical to the sequence of these contours.

This may appear as a very subtle point specific to one analysis of German word intonation. However, anybody who has ever studied a tone language will immediately remember a basic phenomenon in such languages, namely, the expression of morphological relations by means of systematic changes of the tones of the relevant morphs. Any format for morphological glossing must provide for such phenomena.

What is represented in Line V is the pitch contour for the morphological word *un über sicht lichen*, denoted in an abbreviated way through which Line V becomes identical to
Line III, which names the pitch contour for the phonological word *unübersichtlichen*. For this reason, Line V may simply be left empty, with blanks in the cells.

### 6.3 Lines VI to VIII: Orthographic representation of morphs — Stem and morpheme categories — Stem form and morpheme form categories

Line VI contains a sequence of orthographic names for the morphs in Line IV. The sequence must agree with the orthographic name chosen for the phonological word in the syntactic glossing table.

In a tone language we may have a non-segmental morph (‘free tone’), which may be hard to represent orthographically. If no orthographic name is chosen, a bar sign appears in Line VI in the column for this morph.

Line VII supplies new information of a properly morphological kind: information on the stem and morpheme classes associated with the various morphs. These classes are indicated schematically (see Sec. 2.4). The expression “SubStₙ/AdjStₙ” may be read as “the set of morphemes that combine with a form of a substantitive Stem of class l to yield a form of an adjective Stem of class m”. Line VII presupposes a wide-spread conception in morphology by which a Stem, usually, a so-called word stem, may have several stem forms, and is different from any of its forms if only trivially so. The distinction may also be extended to grammatical morphemes.

Distinguishing Stems (word stems) and grammatical morphemes from their forms, we have, in particular, morphological categories for Stems (Line VII) and morphological categories for stem forms (Line VIII). Categories for stem forms are widely assumed, for example, in speaking of Preterite stems (i.e., Preterite stem forms) as opposed to Present tense stems. In Table 2, there are no relevant stem form or morpheme form categories, and this is indicated by a bar in each cell of Line VIII.

The Stem and morpheme categories in Line VII may have to be characterized by also referring to stem form categories; for example, the *bar*-suffix in German combines mainly with forms of Stems of transitive verbs, and these forms must be Present tense forms. Coding such requirements in the name of a Stem or morpheme class is non-trivial and not easily subjected to standardization.

Once again, the details for filling in Lines VII and VIII may be theory dependent; the fundamental distinction between, say, Stem classes and stem form classes is made more or less universally.

### 6.4 Line IX: Meanings and semantic effects

This Line again specifies lexical meanings and semantic effects. The lexical meanings associated with stem morphs are of exactly the same type as the meanings associated with phonological words.

Among the semantic effects associated with individual morphs, we again have two types. One is exemplified by the effect associated with *lich* in column 4 and denoted by “suitable-for” in Line IX. Intuitively, suitable-for takes a meaning associated with *über sicht*, say, the concept ‘overview’, and transforms it into a corresponding ‘suitability meaning’, say, the concept ‘suitable for overview’. The precise nature of such a semantic effect depends
on the presupposed theoretical framework. Ontologically, the effect associated with un in
column 1 and denoted by “not” in Line IX is of the same kind. Intuitively, not takes a
meaning associated with über sicht lich and ‘negates’ it. The two effects are typical of
derivation affix effects.

The semantic effect associated with en in column 5 is given in Line IX as: Unmc Pl
UnmQ Wk. This expression is to be understood exactly as in Table 1, that is, the effect
associated with the inflection morph en is a set of syntactic, not morphological categories.
This is typical of the semantic effects associated with inflection morphs.

The example chosen for Table 2 does not tell us how to deal with morphological intrans-
parency. Suppose that we wish to analyze sicht into two morphs, sicht’ and t, thus creating
one column for sich and one column for t. This would resume the original derivation of
Stems of verbal nouns from Stems of verbs by means of a suffix t, which is no longer pro-
ductive but has left behind a number of semi-transparent stem forms. How are we to indi-
cate that the meaning of sicht is ‘view’ if we associate the verbal concept ‘seeing’ with sich
but do not associate the name of a semantic function (or similar entity) with t? Moreover,
in some cases we may not even be able to find a concept for a stem morph such as sich.

We propose the following solution. If a suitable concept for a stem morph such as sich is
available, a name for the concept is associated in Line IX of a morphological glossing table
with a stem morph such as sich, in our case, this would be “seeing” in single quotes. (If
there is no suitable concept a bar symbol would be entered in Line IX.) The affix morph
(t) would normally be associated in Line IX with the name for its semantic effect (which
can never be a lexical meaning). In place of this name, we now fill in a name (‘view’) of
the concept that is the meaning of the stem and affix morphs together (sicht). Since we
now have a concept name in a column for an affix morph and not a name for a semantic
affix effect, this concept name is now interpreted as naming the meaning of a complex part
of the morphological word. The information on the category of the affix morph in
Line VII and the structural information in Line X will tell us what this part is.

Cases of derivation by conversion are covered by means of different morphological glossing
tables that are linked by means of their Comments and may be compared for purpose of
grammar writing.

6.5 Line X: Representing structural information

a. Overall character of Line X

Virtually all morphological frameworks (at least since Nida 1945) explicitly or implicitly
provide for constituent structures in the morphological analysis of phonological words.
The details vary from one framework to another. We submit that for the morphological
analysis of phonological words structural information must be available that is of the same
type as the information supplied in Line X of a syntactic table. Therefore, the entries in
Line X of Table 2 again form an unordered list, and are of the same kind as the entries in
Line X of Table 1, and are read in the same way. For example,

“2,3,4: Sf” for “the part of the morph sequence consisting of the morphs in columns 2, 3,
and 4 [i.e. über sicht lich] is associated with the category Stem form”.

Once again, bold face is interpreted separately to indicate the (morphological) nucleus- or
head-relation m-nuc: “the part of the morph sequence consisting of the morphs in col-
umn 2 and 3 [i.e. über sicht] is the morphological nucleus of the part of the morph sequence consisting of the morphs in columns 2, 3, and 4 [i.e., of über sicht lich]

There are many more entries in Line X of Table 2 than in Line X of Table 1, which correctly renders the fact that the morphological complexity of unübersichtlichen is much greater than the syntactic complexity of die unübersichtlichen probleme.

The details of the entries, such as the assumption of Affix form or Stem Group as specific morphological constituent categories, are theory dependent; information on (surface) constituent categories in this line is not.

Again, automatic generation of a tree diagram from the entries in Line X should be fairly easy and can use the same algorithm as in the case of Line X in Table 1.

b. Dealing with morphological discontinuity

There is no discontinuity in the case of unübersichtlichen, and this is typical of German morphology. However, morphological discontinuity is a basic linguistic phenomenon and must be accounted for. The following proposals are, to the best of our knowledge, in agreement with what is implicit in the descriptive formats used for languages that typically exhibit discontinuity at the morphological level.

There are two types of discontinuity, created, on the one hand, by so-called circumfixes and, on the other, by infixes. A form of a circumfix may be construed as a sequence of two or more morphs (somewhat stretching the usual sense of "morph"), and the occurrences of a circumfix form are dealt with in a morphological glossing table in exactly the way occurrences of a form of a circumposition are treated in a syntactic table, see above, Sec. 5.2. In many cases where infixes could be postulated an alternative treatment via stem form alternation may be preferable but it seems unwise to exclude infixes quite generally. A true infix creates 'split stem forms'. Discontinuity of a stem form can also be treated by construing the stem form as a sequence of morphs, each with its separate column in the morphological glossing table, where only the first column may be filled in completely and where Line X of the glossing table indicates that these morphs belong together.

6.6 Line XI: Relational information

Although not as wide-spread in morphology as the use of constituent structures, morphological relations — largely patterned on grammatical relations in syntax — are included in most contemporary frameworks and should therefore be provided for in morphological glossing.

Line XI in Table 2 has been construed as precisely analogous to Line XI in Table 1, in particular with respect to the form and interpretation of individual entries. The only entries that must be explained are the ones with "m-qual", to be understood as follows:

"m-qual: 4 2,3 3" for "the part of the morph sequence consisting of the morph in column 4 [lich] qualifies morphologically the 2,3-part of the morph sequence [über sicht] with respect to the 3-part [sicht]"

This relation is patterned on syntactic relations such as negation and is here assumed for German for reasons that need not concern us in the present context.
6.7 Lines XII and XIII: Rendering the phonological word in an established orthography — Word meaning

In Line XII the rendering of the phonological word in the established German orthography ("unübersichtlichen") differs from the orthographic name ("unübersichtlichen") in Line V of Table 1 only by not being in italics. In other cases, differences may be less trivial. In particular, one phonological may require two orthographical words in the established orthography, and conversely, two phonological words may have to be rendered by one orthographic word. Because of such deviations, a separate Line XII is justified even in a morphological glossing table.

The entry in Line XIII of Table 2, "involved" in single quotes, is identical to the name of the meaning of unübersichtlichen in Table 1. Explanations for the name of the meaning (such as "more precisely: 'hard to analyze in all respects'") may have been given already in the Comment of the syntactic Table 1 but are more naturally introduced in the Comment of the morphological Table 2, in a part of the Comment correlated with Line XIII; this line of Table 2 can be linked to the relevant cell in Table 1.
### Appendix 2. Table 2: Morphologic Glossing Table

<table>
<thead>
<tr>
<th>I</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>/?un. /</td>
<td>?yy.bə.r-</td>
<td>ziX.t-</td>
<td>liX</td>
<td>o.n /</td>
</tr>
<tr>
<td>III</td>
<td>H</td>
<td>H,L</td>
<td>H,L</td>
<td>H,L</td>
<td>L</td>
</tr>
<tr>
<td>IV</td>
<td>/?un. /</td>
<td>/?yy.bə.r /</td>
<td>/ziX.t /</td>
<td>/liX. /</td>
<td>/o.n /</td>
</tr>
<tr>
<td>V</td>
<td>H</td>
<td>H,L</td>
<td>H,L</td>
<td>H,L</td>
<td>L</td>
</tr>
<tr>
<td>VI</td>
<td>un</td>
<td>über</td>
<td>sicht</td>
<td>lich</td>
<td>en</td>
</tr>
<tr>
<td>VII</td>
<td>Pref</td>
<td>PrepStj</td>
<td>SubStk</td>
<td>SubSt/AdjStm</td>
<td>AdjFlexn</td>
</tr>
<tr>
<td>VIII</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>IX</td>
<td>not</td>
<td>‘over’</td>
<td>‘view’</td>
<td>suitable-for</td>
<td>UnmC Pl UnmG Wk</td>
</tr>
<tr>
<td>X</td>
<td>1: Af</td>
<td>2: Stf</td>
<td>3: Stf 4: Af</td>
<td>5: Af 2,3: Stf 2,3,4: Stf</td>
<td>1,2,3,4: Stf 1,2,3,4,5: StGromp</td>
</tr>
<tr>
<td>XI</td>
<td>m-mod: 2 3 m-qual: 4 2,3 3</td>
<td>m-mod: 1 2,3,4</td>
<td>m-qual: 5 1,2,3,4 1,2,3,4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XII</td>
<td>unübersichtlichen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XIII</td>
<td>‘involved’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>