

Glossing of Ella Deloria's unpublished Lakota texts

Paul Kroeber

American Indian Studies Research Institute, Indiana University, Bloomington
23 October 2006

Components of the glossed texts: Basic information

Texts are divided into short numbered units each beginning with a line labeled “\ref”; thus the second unit of “Geo. Schmidt’s Vision Experience” begins with a line reading \ref **ED-Aut 1 s 2**. These are the units into which the texts were divided for purposes of the glossing program. As far as possible, these correspond to ED’s own numbered units (sentences or short paragraphs), but some of ED’s units had to be broken up, since they were longer than the glossing program could handle. ED’s own numbering is retained in the text (not in the “\ref” field).

The “\ref” line provides three kinds of information. In order:

- (1) An abbreviation indicating which subcollection of texts the text belongs to. Abbreviations used for this purpose are as follows. (All subcollections are in the American Philosophical Society Library, unless otherwise indicated; titles of the subcollections are due to either the APS or to Raymond DeMallie):

ED-Aut = Ella Deloria: [Dakota autobiographical texts]

ED-Col = Ella Deloria: Dakota tales in colloquial style

ED-Eth = Ella Deloria: Dakota ethnographic and conversational texts

ED-FC = Additional texts from the Deloria Family Collection (not at American Philosophical Society; location??)

ED-Inf = Ella Deloria: Dakota informal texts and conversations

ED-Leg = Ella Deloria: [Dakota legends]

ED-Msc = Ella Deloria: [Miscellaneous Dakota texts]

ED-Sp = Ella Deloria: Dakota speeches

ED-Tal = Ella Deloria: [Dakota tales]

- (2) An arabic number indicating the number of the text within its subcollection. (Some texts were long enough that they had to be divided into parts for purposes of glossing, the parts being labeled by “A”, “B”, etc., after the number of the text. The parts have been rejoined into single PDF documents, but these references remain. Thus, in “The Buffalo People” [text 4 in the Colloquial Narratives subcollection] parts 4A, 4B, etc., can be seen.)

(3) “s” plus an arabic number for the unit of the text (or the part of the text, for those long texts that were divided into parts A, B, etc.).

Thus, `\ref ED-Aut 1 s 2` means (1) that the text is from the “Autobiographical Narratives” subcollection, (2) that it is the first text within that subcollection, and (3) that this is the second unit (paragraph) of that text.

The “`\itm`” notation at the beginning of texts can be ignored -- it relates just to the computer program used to produce the interlinear texts.

The lines in the glossed text itself are as follows:

- `\p` phonemic text (a phonemic rendition of the Lakota text, essentially as transcribed by ED, though we use a slightly different orthography)
- `\g` Gloss (word-for-word glosses — largely ED’s when she provided one, though sometimes with modification)
- `\m` Morph (breakdown of words into -- roughly speaking -- component morphemes, as far as it seemed useful to do so; see comments below)
- `\a` Analysis (morpheme glosses; glosses of grammatical morphemes are in capitals)

There are also two “freeform fields” (that is, fields associated with a paragraph as a whole, not aligned as glosses of individual words or morphemes):

- `\ft` Free_transln (A free translation. Usually this is ED’s, but occasional modifications have been made. If a lengthy comment or major modification or addition has been made, this is sometimes put in square brackets [] or angle brackets < >. But absence of such brackets does not mean that the translation is exactly ED’s.)
- `\cm` Comments (Any notes or comments that seemed useful. Some of these are the editor’s (PK’s) remarks on Lakota forms that were difficult to read or to identify or gloss — these notes are usually in square brackets []. Other notes and comments (not in square brackets) are ED’s; these represent only a very small sample of those in her typescript, typically just remarks that bear directly on the grammatical or semantic interpretation of Lakota forms. ED’s prose in notes and comments has been considerably edited for compactness.)

A parenthesized question mark “(?)” indicates some doubt about either

transcription (where the original typescript is hard to read) or about the analysis.

Detailed discussion of lines in the glossed text

The \m line (the “morphemic” breakdown of words) requires the most discussion, but it may also be helpful to have some additional remarks on what policies were adopted for the other lines.

\p: The phonemic text line

This line essentially reproduces ED’s transcription. Note that ED consistently transcribes syllable-final stops as voiceless **p k** rather than as voiced **b d**; thus, **t’epyá** ‘devour’ rather than **t’ebyá**, **it’ókáp** ‘in front of’ rather than **it’ókab**, **líklíla** ‘very (reduplicated)’ rather than **liglíla**. Transcription here follows her usage.

Certain minor changes have been made:

(i) Minor character substitutions. In particular, characters with a “wedge” diacritic (ǧ ħ š ž) have been substituted for characters with a “dot” diacritic (ğ ħ š ž). For the palatal affricate (the “ch” sound), we have used č rather than plain c.

(ii) Glottal stop symbol (ʔ) is inserted between vowels; no glottal stop character is used word-initially.

(iii) Secondary stress (˘) is not directly marked; vowels marked by ED with secondary stress are simply given an acute accent (´), the same as primary-stressed vowels. It can be assumed that in compound words written with hyphens, where stress is marked both on the expression preceding the hyphen and on the one following, that the second stress is secondary. ED also occasionally writes secondary stress on separate words; this is handled in two ways: (a) Vocative expressions, which ED often marks as having secondary stress, are simply marked as having primary stress here. (b) In other cases where ED writes secondary stress on a separate word, this is indicated by placing “[-]” before the secondary-stress-bearing word. In general, these versions should not be relied on for information about primary vs. secondary stress. (See De Reuse, in IJAL 1994, for some comments about ED’s marking of secondary vs. primary stress; she probably often marked secondary stress as primary.)

(iv) In general ED’s hyphens have been retained, except in Lexical Compounds (those where the compound elements lose their own stresses, and the whole has stress placed on its second syllable by the Dakota Accent Rule). A hyphen is sometimes retained (or inserted) in Lexical Compounds too, when it helps distinguish a sequence of stop + glottal stop across the compound boundary from a glottalized stop.

(v) ED writes a number of enclitics together with the preceding word: these

include =hA, =pi, =ktA, =ħci/A, =šni, =la, =kA, =laħci/A. In the present version, some of these are usually separated out and marked with a double hyphen (=): =ktA, =ħci/A, =šni, =laħci/A, and sometimes =kA. (=hA and =pi are always written together with the preceding word (with no hyphen), since they are difficult to gloss separately on the \g line.)

(vi) The notation “(?)” in the “\p” line means that there was some uncertainty about what the reading of the original was.

\g: The word-for-word translation line

When ED supplied a word-for-word translation of the text, the \g (“Gloss”) line mostly reproduces it. Minor departures from her translation were made for various reasons. For example, some enclitics have been separated off that ED wrote together with the preceding word, and minor changes in glosses were needed to accommodate this. Occasional attempts were also made to increase the consistency of glosses.

The notation “(?)” in this line (provided that it does not correspond to “(?)” on the \p line) means that there is some uncertainty about the appropriate word-for-word translation, either because the original was hard to read, or because the original lacked a word-for-word translation and there was some ambiguity or difficulty about interpreting the word.

The notation “--” on this line is used for \p-line items for which no separate gloss was supplied.

\m: The morpheme line

This line looks like it consists of morphemes in a morphophonemic or underlying-phonological representation, but that is not quite what it is. The original purpose of the \m and \a lines was to facilitate linguists’ searches for Lakota words and sentences of potential grammatical or lexical interest. On the \m line, that meant providing a recognizable consistent representation for each verb-stem, prefix, or whatever; it was less essential that the representation be a plausible underlying form, and in several cases it is not. The Reflexive prefix (RFL-), for instance, is given the representation ič’i- on the \m line, reflecting its most distinct surface allomorph. A more plausible underlying form for this prefix might be ik’i-, the velar helping to account for its other surface allomorphs ik- (before p) and ig- (before k and y, which become l).

Words are morphologically analyzed on the \m line as far as it seems useful. The pronominal affixes are always separated out, and the Dative, Benefactive, Reflexive Possessive, and Reflexive prefixes almost always are (except in a few lexicalized stems). Elements of compounds are normally separated out unless the compound is highly lexicalized. Derivational affixes, notably the instrumental and locative prefixes, are sometimes separated out, if they seem

to be productive or to make a clear semantic contribution in a given stem. Deciding whether it is worthwhile separating out derivational elements involves many judgment calls; there are doubtless many cases where some users might disagree with our decision to, say, treat an affix as an unanalyzable part of the stem or to separate an affix out. Also, there are doubtless quite a few instances where the same or similar stems have been treated inconsistently, sometimes being analyzed and sometimes not. Comments, queries, and corrections are welcomed!

Nouns (other than compounds) are normally treated as unanalyzable. That is, derivational affixes, even ones that are very common in nouns, such **wó-** and **o-**, are not separated out.

Treatment of infixes in the \m line: Infixes and other infixes (namely the Dative, Benefactive, and Reflexive Possessive prefixes) are written as prefixes in the \m line. We follow Patterson (1990) in recognizing two distinct types of infixing stem.

(i) Verb stems beginning with a single vowel (whether that vowel is a derivational prefix or synchronically unanalyzable) normally insert all the infixable prefixes except 1st person dual/plural **u(k)-** after the vowel. For such stems, there is no need to mark where the infix goes. Thus, \p **owáglake** ‘I tell (about my own affairs)’ is given the \m-line analysis **wa- ki- oyak(A)** (‘1.AG- RFL.PS- tell.of’). This includes stems with single-vowel prefixes (**a-**, **i-**, **o-**); infixable prefixes are written before the single-vowel prefix on the \m line.

(ii) Some stems insert all infixable prefixes including 1st person dual/plural **u(k)-** at some point within the stem. Such stems are marked with “^” (read: “unexpected position of prefix”) at the point where infixes go. Thus, \p **c’ewát’i** ‘I build a fire’ is given the \m-line analysis **wa- c’e^t’i** (‘1.AG- build.fire’).

Other uses of ^. The symbol ^ is used before stems beginning with a single vowel that, exceptionally, prefix all pronominal prefixes (and Dative, etc.). (Here too, one can think of ^ as marking an unexpected position for a pronominal prefixes.)

Stems with double pronouns (duplicate Agent pronouns simultaneously in two places): The few stems that do this (**iyayA** ‘start thither’ [1.AG **ibláble**], **^iyak(A)** ‘run’ [1.AG **wa’imnake**], and a few other lower-frequency ones) are simply regarded as irregular — no attempt is made in the \m form to show where the second pronominal prefix goes.

In a few stems, “..” is used to connect final elements that used to be enclitics, are now lexicalized as part of the stem, but retain the position of enclitics. E.g., **cík’a..la** ‘small’ (where the **la** is etymologically the diminutive enclitic, but is now normally required by the stem), which inserts the enclitic **=pi** ‘Plural’

before its final **la**: **cík’apila** ‘they are small’. Occasionally, too, “<” is used as the gloss of **=šni** (‘not’) when it is lexicalized as part of a stem but is separated from it on the \p line, as in (Col. 5.139; **waḥte..šni** is a lexicalized form for ‘bad’):

\p	waḥtémala		=kta	=šni
\g	she dislike me		=will	--
\m	waḥte..šni	+ ma-	la	=ktA <
\a	bad	+ 1S.PAT-	consider	=IRR <

Transcription in the \m line: juncture symbols (these draw heavily on, e.g., Patricia Shaw’s discussion of Lakota phonology, but are not necessarily identical to Shaw’s usage):

hyphen (-) for affixes (bound elements to which stress can potentially be assigned by the Dakota Accent Rule).

double hyphen (=) for enclitics (bound elements which are outside the domain to which stress can be assigned by the Dakota Accent Rule). (In general, we have used “=” only with enclitics that ED writes together with the preceding stem: =hA, =pi, =ktA, =ḥci/A, =šni, =la, =kA. Other enclitics are written as separate words.)

plus sign (+), normally with spaces around it, connects the elements of Lexical Compounds (compounds where the compound elements lose their own stresses, and the whole has stress placed on its second syllable by the Dakota Accent Rule). “Bound verbs” — verbs which, though inflected, must be phonologically compounded with a preceding infinitival verbs — are treated as lexical compounds. (The commonest bound verbs are **ya** ‘Causative’, **k’iyA** ‘Intentional Causative’, and **ši** ‘command, tell to ...’.)

double plus sign or crosshatch (#), normally with spaces around it, connects the elements of Syntactic Compounds (compounds where the compound elements retain their own stresses, the stress of non-initial elements merely being reduced to secondary).

Transcription in the \m line: pronominal prefixes. The first and second person active pronominal prefixes (whose usual forms are respectively **wa-** and **ya-** in most verbs) are given the consistent forms **w@-** and **y@-** in the \m line. This helps distinguish them from such other prefixes as **wa-** ‘Unspecified Object’, **wa-** ‘by cutting’ (instrumental prefix, **INSTR5**), **ya-** ‘by mouth’ (instrumental prefix, **INSTR7**), and also serves as a reminder that the agent prefixes sometimes have special contracted forms (e.g., **b-** and **l-** with y-verbs).

Transcription in the \m line: other affixes. Similarly, the ‘Reflexive Possessive’ prefix has been given the form **kl-** in the \m line; this helps to distinguish it from the Dative prefix, which appears as **ki-**, and also serves as a reminder that

the Reflexive Possessive prefix sometimes appears in a contracted form (**gl-** with y-verbs).

Transcription in the \m line: ablaut. Stem-final vowels (also some enclitic-final vowels) **a** and **ǣ** that are replaced by **e** or **i** (“ablaut”) before certain enclitics are represented on the \m line as **A** and **Ǻ** (following fairly standard practice in Sioux linguistics). Note that stem-final **a** or **ǣ** in the \m line may mean either that the stem is known not to ablaut, or that we could not determine whether the stem ablauts or not. (Some stems may have been treated inconsistently, and we would appreciate having them pointed out.)

Transcription in the \m line: epenthetic a. Final parenthesized (**a**) or (**A**) on a stem in the \m line indicates that the final vowel of the stem is known to be epenthetic. Final unparenthesized **a** or **A** indicates that the final vowel either is known not to be epenthetic, or that it is not known to be epenthetic (on the basis of data in the text collection and Buechel’s dictionary). (Again, some stems may have been treated inconsistently, and we would appreciate having them pointed out.)

Transcription in the \m line: stems. Two stems have been given somewhat arbitrary forms to help distinguish them.

(1) The stem glossed on the \a line as ‘go’ (1st person sing. **blé**, [**mní** before =**ktA**]; 1st person pl. **uyápi**; 3rd person sing. **yé**, [**yí** before =**ktA**]) is represented on the \m line as **YA**. This helps distinguish it from the also very common causative bound root, represented as **yA**.

(2) The stem **u** glossed on the \a line as ‘stay’ (locational/existential copula) is represented on the \m line as **’u**, helping distinguish it from other **u**’s (article ‘**DEF.PST**’, postpositions ‘because.of’ and ‘using’ [=instrumental ‘with’]), all of which are represented on the \m line as simple **u**.

Transcription in the \m line: stress. Stress is only marked in the \m line when it is unpredictable — that is, on certain forms whose first syllable is stressed, overriding the Dakota Accent Rule which normally places stress on the second syllable of a word. (Certain stems, like **náži** ‘stand’, **máni** ‘walk’, have stress on their first syllable when they have no infixal pronominal prefixes, but shift stress to an infixal pronominal prefix: **nawáži** ‘I stand’, **ma’únipi** ‘we walk’. These stems are simply marked with initial stress in the \m line: **wa- ná^ži**, **uk- má^ni =pi**, in the same way as stems with completely fixed initial stress — the user will just have to remember the stress shift.)

Other notations in the \m line. “(?)” on this line (provided it does not correspond to “(?)” on the \p line) means that there was some uncertainty about the morphological analysis of the Lakota form.

\a: The morpheme gloss line (“Analysis”)

Glosses of stems are sometimes rather ad hoc, and may not always reflect what a more considered semantic or grammatical analysis would consider the best gloss for the stem. Also note that the same stem may be given more than one gloss, if its meaning seems to vary with context. (And sometimes different stems may have the same gloss.)

Glosses of grammatical morphemes (affixes, enclitics, particles) are less ad hoc. (For explanations of the abbreviations, see the link from the Siouxan Corpora web page.) Nonetheless, some convenience-based decisions were made. For example, the pronominal prefix **uk-** (\m-line form; corresponds to **uk-** or **u-** on the \p line) is given the single consistent gloss '1.PL' ('first person plural'), rather than different glosses representing its various functions as 1st person dual inclusive (unaccompanied by **=pi**) vs. 1st person plural or exclusive (accompanied by **=pi**), agent vs. patient, etc. And the enclitic **=pi** is always the gloss 'PL' ('plural'), even in contexts where it clearly codes an impersonal passive rather than a plural agent. (In 'he was thrown by a horse' [Informal Texts 8.26], the **=pi** on the verb clearly cannot be expressing plurality of the agent. But there are also many contexts where it is not clear whether **=pi** ought to be interpreted as plural or as passive; to distinguish glosses for 'plural' and 'passive' for this enclitic would have required many arbitrary decisions as to which gloss to use.) Also, glosses for a number of the sentence-final enclitics and some other particles are only provisional.

Other notations in the \a line. "(?)" on this line (provided that it does not correspond to "(?)" on the \m line or the \p line) indicates that there was some uncertainty about the gloss for this morpheme or stem. Often all that this means is that the morpheme or stem does not appear separately in Buechel's dictionary, so that the morpheme or stem gloss had to be supplied on the basis of ED's free or word-for-word translation.