

dvGLOSS.sty 0.11: flexible linguistic gloss macros

Waral xchi-qa-tz'ib'a-j *wi*
 here PROSP-A1p-write-ss FOC
xchi-qa-tiki-b'a' *wi ojer tzij,*
 PROSP-A1p-stand-CAUS FOC old word

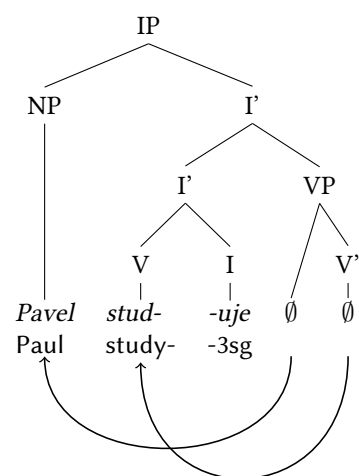
u-tik-ar-ib'al
 A3s-plant-VERS-INSTR

u-xe'-n-ab'al *puch, r-onojel x-b'an* *pa tinamit K'iche'*,
 A3s-root-AP-INSTR PART A3s-all CPL-make:PASS P town K'iche'

r-amaq' K'iche' winaq.
 A3s-nation K'iche' people

“Here we will write, / will erect the ancient words: / the planting, / the rooting, all that was done in the K'ichee' town, / the nation of K'ichee' people.”

(*Popol Wuj*)



Subject	Object	Verb
<i>Wičása ki</i> man the	<i>mathó wə</i> bear a	<i>ktə</i> killed

Subject	Object	Verb
<i>Mathó wə</i> bear a	<i>wičása ki</i> man the	<i>ktə</i> killed

Subject	Object	Verb
<i>Wičása ki</i> man the	<i>ix?é wə</i> rock a	<i>wayále</i> saw

Object	Subject	Verb
<i>Ix?é wə</i> rock a	<i>wičása ki</i> man the	<i>wayále</i> saw

(*Van Valin 1985*)

Contents

1	What	2
2	How	3
2.1	Declaring shortcuts	6

1 What

This is a macro package for setting interlinear glossed text.

Why another one of those? There are already at least two widely used sets of glossing macros floating around: the old reliable midnight macros which are used in `linguex` and `gb4e`, and a custom-written set of macros in `expex`. So why write a new one?

Well, I can explain, but this’ll take a minute.

What glossing macros do is take horizontally organized input — like this

```
\gl{lorem ipsum dolor sit amet}
{1      2      3      4      5}
```

— and pair the words off into a bunch of vertical boxes, like this:

lorem	ipsum	dolor	sit	amet
1	2	3	4	5

But sometimes you want to issue a command that will “escape” from this process: that will not get paired off or boxed up, but will end up on its own and outside of any box. For instance, if you want to force a line break using the `\par` command, or make an adjustment to the spacing between words, it really needs to take place *between* boxes, like so:

lorem	ipsum	\par	dolor	\kern3pt	sit	amet
1	2		3		4	5

The midnight macros don’t have any escape mechanism that will make this possible. In `expex`, there are some built-in shortcuts that work a little like escaped commands (for instance, if you type a forward slash character in the middle of a gloss, it will tell `expex` “start a new line here,” and so this is a little bit like being able to escape a `\par` command.) — but these shortcuts are hardcoded into the package, and it’s difficult to create new ones.

This package has an escape mechanism that can get arbitrary bits of \TeX code out into the space in between words. It also lets you define your own single-character

shortcuts that are equivalent to bits of escaped code. (It comes with a few shortcuts predefined, but you can override these.) Basically nothing is hard-coded, and everything can be redefined. This lets you play all sorts of fancy formatting tricks like the ones on page 1 — but also, more prosaically, it means you can have your line breaks wherever you want them:

```
\gl{lorem ipsum !{\par} dolor !{\kern2pt} sit amet}
  {1      2              3              4      5}
```

2 How

```
1 \newtoks\ta\newtoks\tb
2 \newdimen\glhangindent\glhangindent=2em
3 \newdimen\betweenglspace\betweenglspace=\jot
4 \newdimen\withinglspace\withinglspace=0pt
5 \newdimen\aboveglspace\aboveglspace=0pt
6 \newdimen\glstrutheight\newdimen\glstrutdepth
7 \newdimen\glstrutwidth\glstrutwidth=0pt
8 \newdimen\glsmallstrutheight\newdimen\glsmallstrutdepth
9 \def\glSPACE{\penalty0\hspace{0pt}}
10 \def\everygla{\sl}
11 \def\everyglb{\sf}
12 \def\glESCAPE{!}
13 \long\def\addtokens#1\to#2{\ta={#1}\tb=\expandafter{#2}\edef#2{\the\tb\the\ta}}
14 \def\pop#1\to#2{\expandafter\popoff#1\to#2\remainderin#1}
15 \long\def\popoff#1 #2\to#3\remainderin#4{#3={#1}\def#4{#2}}
16 \long\def\split#1{\expandafter\ssplit#1\xyzy}
17 \long\def\ssplit#1#2\xyzy{\gdef\istchar{#1}\gdef\restchars{#2}}
18 \def\ifnotin#1#2{%
19   \def\@ifnotin##1#1##2##3\zyzy{\ifx\ifnotinfound##2}%
20   \expandafter\@ifnotin#2#1\notfound\zyzy}
21 \newif\ifspecial
22 \def\dvglSpecials{}
23 \def\checkSpecial#1{%
24   \ifnotin#1{\dvglSpecials}\specialfalse\else\specialtrue\fi}
25 \def\dvglpassthru{}
26 \def\checkpassthru#1{%
27   \ifnotin#1{\dvglpassthru}\specialfalse\else\specialtrue\fi}
28 \def\makeSpecial#1{%
29   \edef\dvglSpecials{\dvglSpecials#1}}
30 \def\makepassthru#1{%
31   \edef\dvglpassthru{\dvglpassthru#1}}
32 \def\stacksymbol#1#2{%
33   \mbox{\vtop{\hbox{#1}/\strut}\nointerlineskip\hbox{#2}/\strut}}}%
34 \def\glossword#1#2{%
35   \stacksymbol{\everygla#1}{\everyglb#2}%
36   \glSPACE}
37 \def\glosswordmacro#1#2{\glossword{#1}{#2}}
```

`\merge` The meat and bones of the thing: this is what I called “zipwith-with-exceptions” in the
`bracketcount` explanation up above, but “merge” is shorter and easier to type. All three arguments
`maxbracketcount` should be token list macros (so the usage is `\merge\x\and\y\to\z`). The first two
arguments should contain its input; the last will contain its output at the end. The
input token lists will be destroyed in the course of the macro, so I hope you weren’t
using them for anything.

```
38 \newcount\bracketcount\newcount\maxbracketcount%
39 \def\merge#1\and#2\to#3{%
```

If there’s still stuff in our first token list, pop an item off the front. Then, split the first
character off of that first item.

```
40 \ifx\empty#1\else%
41 \pop#1\to\ta%
42 \def\istchar{}\def\restchar{}\%
43 \edef\temp{\the\ta}\split\temp%
```

If that first character is an escape, add the rest to our output token list.

```
44 \ifx\istchar\glescape%
45 \expandafter\addtokens\restchars\to#3%
46 \expandafter\addtokens\space\to#3%
47 \else%
```

If that first character is a passthrough character, add it *and* all the rest to the output
token list.

```
48 \expandafter\checkpassthru\istchar\ifspecial%
49 \expandafter\addtokens\istchar\to#3%
50 \expandafter\addtokens\restchars\to#3%
51 \expandafter\addtokens\space\to#3%
```

(This is a cheat. For `\gltree` down below, we’ll want to know the deepest bracket
nesting that occurs in the course of a gloss. Here we have some code to calculate that.
It is hard-wired to work only with square brackets, and only when square brackets
have been declared as passthru characters.)

```
52 \if\istchar[\advance\bracketcount by 1\fi%
53 \ifnum\bracketcount>\maxbracketcount%
54 \maxbracketcount=\bracketcount\fi
55 \if\istchar]\advance\bracketcount by -1\fi%
```

Okay, back to the important stuff. If that first character is a special character, gener-
ate the appropriate control sequence name from it, make the remaining characters in
the item its argument, and add that to the output token list. (So if `*` is a special char-
acter, and we see `*argument` as an item in the first argument, we’ll end up sending
`\gl*<argument>` to output.

```
56 \else%
57 \expandafter\checkspecial\istchar\ifspecial%
58 \expandafter\addtokens\csname gl\istchar\endcsname\to#3%
59 \expandafter\addtokens\expandafter<\restchars>\to#3%
60 \expandafter\addtokens\space\to#3%
61 \else%
```

That’s all the special cases. If we’ve gotten this far, we know we’re building a normal glossword. Make sure the second argument isn’t empty; pop off an item from the second argument; package those two items together into a glossword; and away we go. One tricky thing: `\glosswordmacro` is expanded exactly once here. So if we do `\def\glosswordmacro\foo`, and the current items are `first` and `second`, we’ll end up outputting `\foo{first}{second}`.

```

62      \ifx\empty#2\else%
63      \pop#2\to\tb%
64      \edef\temp{\expandafter\noexpand\glosswordmacro{\the\ta}{\the\tb} }%
65      \expandafter\addtokens\temp\to#3%
66      \fi%
67  \fi%
68  \fi%
69  \fi%
```

And recurse!

```

70      \merge#1\and#2\to#3%
71  \fi%
72 }
```

`\putgl` This is just a wrapper around `\merge`, to compensate for the fact that `\merge` wants an extra space at the end of its first two arguments. Ideally, in a future version, we’ll add some more code here to make sure there’s one *and only one* trailing space, since multiple spaces also make `\merge` barf.

```

73 \long\def\putgl#1#2\into#3{%
74   \def\x{#1 } \def\y{#2 } \def#3{ }%
75   \merge\x\and\y\to#3%
76 }
```

`\gl` Here’s the basic user-facing gloss macro. Take two arguments and merge them together. Do a little formatting. execute the output we got from from `\merge`. Do a little more formatting. The stuff having to do with indentation and spacing is at the end, so we can make reasonably sure it will still be in effect when the paragraph is broken. But we don’t actually break the paragraph here, in case the user wants to add a label in the right margin (see `\lb` below).

```

77 \long\def\gl#1#2{%
78   \putgl{#1}{#2}\into\z%
79   \ifvmode\vskip\aboveglSPACE\fi%
80   \z%
81   \hangafter1\hangindent=\glhangindent%
82   \lineskiplimit=\betweenglSPACE\lineskip=\betweenglSPACE%
83   \rightskip=0pt plus 1fil%
84 }
```

`\gltree` This one not only creates a sequence of glossboxes, but feeds it to `tikz-qtrees`’s `\Tree` macro as an argument. To make this go smoothly, we need to define a replacement for `\glossbox` which will create a `tikz` node instead of just a `TeX`box. We also keep track of how many nodes we’ve created so that we can give each one a name — the first one’s named “1,” the second “2,” and so on.

`\tikzglossword`
`nodecount`

(I do not understand why the `\noexpands` in this macro are necessary, or why they work. but they do seem to be necessary, and they do seem to work. So be it.)

```

85 \newcounter{nodecount}%
86 \def\nc{\addtocounter{nodecount}{1}}
87 \def\tikzglossword#1#2{%
88   \node[align=left](\noexpand\arabic{nodecount}){%
89     \noexpand\stepcounter{nodecount}\noexpand\everygla#1\noexpand\\noexpand\everyglb#2};}%

```

Right. Now we can define `\gltree` itself. This is pretty straightforward. We declare left and right brackets to be passthru characters; we initialize the bracket counter; we do a little formatting; and away we go. The tree is set up to be `\maxbracketcount * 30` points tall; that leaves room for each level in the tree to have a height of 30 points.

```

90 \newcommand{\gltree}[3]{%
91   \begingroup%
92   \bracketcount=0\maxbracketcount=0%
93   \makepassthru[\makepassthru]%
94   \begin{tikzpicture}[every leaf node/.append style={inner sep=1}, baseline=base, sibling distance=]
95   \setcounter{nodecount}{1}%
96   \let\glosswordmacro=\tikzglossword%
97   \putgl{#1}{#2}\into\z%
98   \multiply\maxbracketcount by 30%
99   \expandafter\Tree\z%
100  #3%
101  \end{tikzpicture}%
102  \endgroup%
103 }

```

`\gltab` This one wraps our sequence of glossboxes in \TeX 's tabbing environment, and sets up some shortcuts for commonly used commands in that environment. The `minipage` is necessary to keep us from getting a linebreak and extra vertical space at the beginning and end of the tabbing environment. The width of the `minipage` is set arbitrarily; this is dumb and we should do better in a future version.

```

104 \long\def\gltab#1#2{%
105   \begingroup%
106   \makeglshortcut={\=} \makeglshortcut>{\>} \makeglshortcut/{\[\betweenlinspace]}%
107   \putgl{#1}{#2}\into\z%
108   \begin{minipage}[t]{0.9\textwidth} \begin{tabbing} \z \end{tabbing} \end{minipage}%
109   \endgroup%
110 }

```

`\ft` These are for putting a “free translation” after a gloss (`\ft`) and for putting a “label”
`\lb` to the right of a gloss (`\lb`). Nothing too complicated.

```

111 \def\ft#1{\par\nopagebreak[4]\lineskiplimit=0pt\lineskip=1pt\addvspace{\jot}\#1\strut}
112 \def\lb#1{{\unskip\nobreak\hfil\penalty0\hskip2em\mbox{} \nobreak\hfill\mbox{\strut#1}}

```

2.1 Declaring shortcuts

We have a macro `\makespecial` which will tell the parser “treat this character as special.”

But that alone won't get us anywhere interesting. We also need to *define* the macro that will be executed when the special character in question is encountered. And it would be nice if we could wrap both steps — calling `\makespecial` and defining the macro — in a user-friendly package. That's what this last section does.

Suppose we want to declare `/` as a shortcut character which inserts a line break. Step one is simple:

```
\makespecial{/}
```

Now let's consider how the parser will respond after we've done this. Suppose we then execute this command:

```
\gl{a b * c}{1 2 3}
```

The output which is sent to be typeset will look like this

```
\glossword{a}{1}
\glossword{b}{2}
\gl/<>
\glossword{c}{3}
```

where `\gl/` is a control sequence — one which could not normally be input directly. And if we execute the command

```
\gl{a b /argument c}{1 2 3}
```

The output which is sent to be typeset will look like this:

```
\glossword{a}{1}
\glossword{b}{2}
\gl/<argument>
\glossword{c}{3}
```

Step two, then, is to define a macro that will do something useful in those contexts. If the slash were a normal character, this would be easy:

```
\def\gl/<#1>{\ifx#1\empty\else\vskip{#1}\fi}
```

But since the slash isn't a normal character, we need to do it this way instead:

```
\expandafter\def\csname gl/\endcsname<#1>{\par\ifx#1\empty\else\vskip{#1}\fi}
```

`\makeglshortcut` Well, that's just what `\makeglshortcut` does: carries out Step One and Step Two for an arbitrary character and arbitrary macro code.

```
113 \long\def\makeglshortcut#1#2{%
114   \makespecial{#1}%
115   \expandafter\gdef\csname gl#1\endcsname<##1>{#2}}
```

`\makeglsurround` `\makeglsurround` does the same thing, except that in Step Two it creates a delimited macro with several arguments. This requires some unpleasant fucking around with token registers and `\expandafter`.

```

116 \long\def\makeglsurround#1#2#3{%
117   \xdef\dvglspecialchars{\dvglspecialchars#1#2}%
118   \ta=\expandafter{\csname gl#1\endcsname}%
119   \tb=\expandafter{\csname gl#2\endcsname}%
120   \expandafter\expandafter\expandafter\gdef%
121     \expandafter\expandafter\the\ta%
122     \expandafter<\expandafter##\expandafter1\expandafter>%
123     \expandafter##\expandafter2\the\tb<##3>{#3}}

/ And here we have definitions for the built-in shortcuts.
[
124 \makeglshortcut/{\ifx#1\empty\else\vskip{#1}\fi}
]
125 \makeglshortcut[{\stacksymbol{[}{[{}]}
*...* 126 \makeglshortcut[{\stacksymbol{[}{$_{#1}$}{[}{$_{#1}$}}]
<...< 127 \makeglsurround ** {%
128   \rlap{\raisebox{1.5em}{\footnotesize\sf#3}}}%
129   \fbox{#2\unskip}\glspace}
130 \def\doarrow\glossword#1\glossword#2\glossword#3\glossword{\linkto\glossword#1\underset{\glossword#2}{fr
131 \makeglsurround<<{\doarrow#2}

```