Introduction

Interlinear morpheme-by-morpheme glosses are common in linguistic texts to give information about the meanings of individual words and morphemes in the language being studied. A set of conventions called the **Leipzig Glossing Rules** was developed to give linguists a general set of standards and principles for how to format these glosses. The most recent version of these rules can be found in PDF form at **this link**, provided by the Department of Linguistics at the Max Planck Institute for Evolutionary Anthropology.

There is a staggering variety of LaTex packages designed to properly align and format glosses (including `gb4e`, `ling-macros`, `linguex`, `expex`, and probably even more). These modules vary in the complexity of their syntax and the amount of control they give to the user of various aspects of formatting. The `typst-leipzig-glossing` module is designed to provide utilities for creating aligned Leipzig-style glosses in Typst, while keeping the syntax as intuitive as possible and allowing users as much control over how their glosses look as is feasible.

This PDF will show examples of the module's functionality and detail relevant parameters. For more information or to inform devs of a bug or other issue, visit the module's Github repository **https://github.com/neunenak/typst-leipzig-glossing**

Basic glossing functionality

As a first example, here is a gloss of a text in Georgian, along with the Typst code used to generate it:

from "Georgian and the Unaccusative Hypothesis", Alice Harris, 1982

ბავშვ-ი	ატირდა
bavšv-i	aṭirda
child-NOM	3S/cry/INCHO/II

The child burst out crying

```
#import "leipzig-gloss.typ": gloss
#gloss(
    header: [from "Georgian and the Unaccusative Hypothesis", Alice Harris, 1982],
    source: ([ბავშვ-ი], [ატირდა]),
    transliteration: ([bavšv-i], [aṭirda]),
    morphemes: ([child-#smallcaps[nom]], [3S/cry/#smallcaps[incho]/II]),
    translation: [The child burst out crying],
)
```

And an example for English which exhibits some additional styling, and uses imports from another file for common glossing abbreviations:

I'm	eat-ing	your	head
1SG.SBJ=to.be	eat-PROG	2SG.POSS	head

"**I'm eating your head!**"

```
#gloss(
  source: ([I'm], [eat-ing], [your], [head]),
  source-style: (item) => text(fill: red)[#item],
  morphemes: ([1#sg.#sbj\=to.be], [eat-#prog], [2#sg.#poss], [head]),
  morphemes-style: text.with(size: 10pt, fill: blue),
  translation: text(weight: "semibold")[I'm eating your head!],
  translation-style: (item) => ["#item"],
)
```

The `#gloss` function has three pre-defined parameters for glossing levels: `source`, `transliteration`, and `morphemes`. It also has two parameters for unaligned text: `header` for text that precedes the gloss, and `translation` for text that follows the gloss.

The `morphemes` param can be skipped, if you just want to provide a source text and translation, without a gloss:

Trato de entender, debo comprender, qué es lo que ha hecho conmigo
I try to understand, I must comprehend, what she has done with me

```
#gloss(
    source: ([Trato de entender, debo comprender, qué es lo que ha hecho
conmigo],),
    source-style: emph,
    translation: [I try to understand, I must comprehend, what she has done with
me],
)
```

Note that it is still necessary to wrap the `source` argument in an array of length one.

Here is an example of a lengthy gloss that forces a line break:

Ich arbeite ein Jahr um das Geld zu verdienen, das dein Bruder an einem Wochenende ausgibt.
I work one year to the money to earn, that your brother on one weekend spends.
“I work one year to earn the money that your brother spends in one weekend”

```
#gloss(
    source:    ([Ich],[arbeite],[ein],[Jahr],[um],[das],[Geld], [zu],[verdienen,],
[das], [dein],[Bruder], [an],[einem],[Wochenende],[ausgibt.]),
    source-style: text.with(weight: "bold"),
    morphemes: ([I],  [work],[  one], [year],[to],[the],[money],[to],[earn,],
[that],[your],[brother],[on],[one],  [weekend],   [spends.]),
    translation: ["I work one year to earn the money that your brother spends in
one weekend"]
)
```

To add more than three glossing lines, there is an additional parameter `additional-lines` that can take a list of arbitrarily many more glossing lines, which will appear below those specified in the aforementioned parameters:

Hunzib (van den Berg 1995:46)

ождиг	хо^{н}хе	мукъер
oʒdig	χõχe	muq’er
ož-di-g	xõxe	m-uq’e-r
boy-OBL-AD	tree(G4)	G4-bend-PRET
at boy	tree	bent

“Because of the boy, the tree bent.”

```
#gloss(
    header: [Hunzib (van den Berg 1995:46)],
    source: ([ождиг],[хо#super[н]хе],[мукъер]),
    transliteration: ([oʒdig],[χõχe],[muq’er]),
    morphemes: ([ož-di-g],[xõxe],[m-uq'e-r]),
    additional-lines: (
        ([boy-#smallcaps[obl]-#smallcaps[ad]], [tree(#smallcaps[g4])],
[#smallcaps[g4]-bend-#smallcaps[pret]]),
        ([at boy], [tree], [bent]),
    ),
    translation: ["Because of the boy, the tree bent."]
)
```

Numbering Glosses

The `gloss` function takes a boolean parameter `numbering` which will add an incrementing count to each gloss. A function `numbered-gloss` is exported for convenience; this is defined as simply `#let numbered-gloss = gloss.with(numbering: true)`, and is called with the same arguments as `gloss`:

(1) გვ-ფრცქვნ-ი
gv-prtskvn-i
1PL.OBJ-peel-FMNT
You peeled us

(2) მ-ფრცქვნ-ი
m-prtskvn-i
1SG.OBJ-peel-FMNT
You peeled me

```
#gloss(
    source: ([გვ-ფრცქვნ-ი],),
    transliteration: ([gv-prtskvn-i],),
    morphemes: ([1#pl.#obj\-peel-#fmnt],),
    translation: "You peeled us",
    numbering: true,
)

#numbered-gloss(
    source: ([მ-ფრცქვნ-ი],),
    transliteration: ([m-prtskvn-i],),
    morphemes: ([1#sg.#obj\-peel-#fmnt],),
    translation: "You peeled me",
)
```

The displayed number for numbered glosses is iterated for each numbered gloss that appears throughout the document. Unnumbered glosses do not increment the counter for the numbered glosses.

The gloss count is controlled by the Typst counter variable `gloss-count`. This variable can be imported from the `leipzig-gloss` package and manipulated using the standard Typst counter functions to control gloss numbering:

(21) from *Standard Basque: A Progressive Grammar* by Rudolf de Rijk, quoting P. Charriton
Bada beti guregan zorion handi baten nahia.
There always is in us a will for a great happiness.

```
#gloss-count.update(20)

#numbered-gloss(
    header: [from _Standard Basque: A Progressive Grammar_ by Rudolf de Rijk,
quoting P. Charriton],
    source: ([Bada beti guregan zorion handi baten nahia.],),
    translation: [There always is in us a will for a great happiness.],
)
```

Styling lines of a gloss

Each of the aforementioned text parameters has a corresponding style parameter, formed by adding `-style` to its name: `header-style`, `source-style`, `transliteration-style`, `morphemes-style`, and `translation-style`. These parameters allow you to specify formatting that should be applied to each entire line of the gloss. This is particularly useful for the aligned gloss itself, since otherwise one would have to modify each content item in the list individually.

In addition to these parameters, Typst's usual content formatting can be applied to or within any given content block in the gloss. Formatting applied in this way will override any contradictory line-level formatting.

This text is about eating your head.
I'm *eat-ing* *your* *head*
1SG.SBJ=to.be eat-PROG 2SG.POSS head
I'm eating your head!

```
#gloss(
    header: [This text is about eating your head.],
    header-style: text.with(weight: "bold", fill: green),
    source: (text(fill:black)[I'm], [eat-ing], [your], [head]),
    source-style: text.with(style: "italic", fill: red),
    morphemes: ([1#sg.#sbj\=to.be], text(fill:black)[eat-#prog], [2#sg.#poss],
[head]),
    morphemes-style: text.with(fill: blue),
    translation: text(weight: "bold")[I'm eating your head!],
)
```

Standard Abbreviations

The Leipzig Glossing Rules define a commonly-used set of short abbreviations for grammatical terms used in glosses, such as ACC for "accusative (case)", or PTCP for "participle" (see "Appendix: List of Standard Abbreviations in the Leipzig Glossing Rules document)

By convention, these are typeset using SMALLCAPS. This package contains a module value `abbreviations`. Individual abbreviations may be accessed either with Typst field access notation or by importing them from `abbreviations`:

(from *Why Caucasian Languages?*, by Bernard Comrie, in *Endangered Languages of the Caucasus and Beyond*)
[qálɐ-m ∅-kw'-á] ɬ'ə́-r
city-OBL 3SG-go-PRF man-ABS
The man who went to the city.

```
#import "leipzig-gloss.typ": abbreviations
#import abbreviations: obl, sg, prf

#gloss(
    header: [(from _Why Caucasian Languages?_, by Bernard Comrie, in _Endangered
Languages of the Caucasus and Beyond_)],
    source: ([\[qálɐ-m], [ø-kw'-á\]], [ɬ'ə́-r]),
    morphemes: ([city-#obl], [3#sg\-go-#prf], [man-#abbreviations.abs]),
    translation: "The man who went to the city."
)
```

The full list of abbreviations is as follows:

Full list of abbreviations

1 - `1` - first person
2 - `2` - second person
3 - `3` - third person
A - `a` - agent-like argument of canonical transitive verb
ABL - `abl` - ablative
ABS - `abs` - absolutive
ACC - `acc` - accusative
ADJ - `adj` - adjective
ADV - `adv` - adverb(ial)
AGR - `agr` - agreement
ALL - `all` - allative
ANTIP - `antip` - antipassive
APPL - `appl` - applicative
ART - `art` - article
AUX - `aux` - auxiliary
BEN - `ben` - benefactive
CAUS - `caus` - causative
CLF - `clf` - classifier
COM - `com` - comitative
COMP - `comp` - complementizer
COMPL - `compl` - completive
COND - `cond` - conditional
COP - `cop` - copula
CVB - `cvb` - converb
DAT - `dat` - dative
DECL - `decl` - declarative
DEF - `def` - definite

DEM - `dem` - demonstrative
DET - `det` - determiner
DIST - `dist` - distal
DISTR - `distr` - distributive
DU - `du` - dual
DUR - `dur` - durative
ERG - `erg` - ergative
EXCL - `excl` - exclusive
F - `f` - feminine
FOC - `foc` - focus
FUT - `fut` - future
GEN - `gen` - genitive
IMP - `imp` - imperative
INCL - `incl` - inclusive
IND - `ind` - indicative
INDF - `indf` - indefinite
INF - `inf` - infinitive
INS - `ins` - instrumental
INTR - `intr` - intransitive
IPFV - `ipfv` - imperfective
IRR - `irr` - irrealis
LOC - `loc` - locative
M - `m` - masculine
N - `n` - neuter
N- - `n-` - non- (e.g. NSG nonsingular, NPST nonpast)
NEG - `neg` - negation, negative
NMLZ - `nmlz` - nominalizer/nominalization
NOM - `nom` - nominative
OBJ - `obj` - object
OBL - `obl` - oblique
P - `p` - patient-like argument of canonical transitive verb
PASS - `pass` - passive
PFV - `pfv` - perfective
PL - `pl` - plural
POSS - `poss` - possessive
PRED - `pred` - predicative
PRF - `prf` - perfect
PRS - `prs` - present
PROG - `prog` - progressive
PROH - `proh` - prohibitive
PROX - `prox` - proximal/proximate
PST - `pst` - past
PTCP - `ptcp` - participle
PURP - `purp` - purposive
Q - `q` - question particle/marker
QUOT - `quot` - quotative
RECP - `recp` - reciprocal
REFL - `refl` - reflexive
REL - `rel` - relative

RES - `res` - resultative
S - `s` - single argument of canonical intransitive verb
SBJ - `sbj` - subject
SBJV - `sbjv` - subjunctive
SG - `sg` - singular
TOP - `top` - topic
TR - `tr` - transitive
VOC - `voc` - vocative

Building used-abbreviations pages

A user of `leipzig-glossing` might wish to generate an introductory page displaying which abbreviations were actually used in the document.

Further Example Glosses

These are the first twelve example glosses given in **https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf**. along with the Typst markup needed to generate them:

(1) Indonesian (Sneddon 1996:237)
Mereka di Jakarta sekarang.
they in Jakarta now
They are in Jakarta now

```
#numbered-gloss(
    header: [Indonesian (Sneddon 1996:237)],
    source: ([Mereka], [di], [Jakarta], [sekarang.]),
    morphemes: ([they], [in], [Jakarta], [now]),
    translation: "They are in Jakarta now",
)
```

(2) Lezgian (Haspelmath 1993:207)
Gila abur-u-n ferma hamišaluğ güğüna amuq'-da-č.
now they-OBL-GEN farm forever behind stay-FUT-NEG
Now their farm will not stay behind forever.

```
#numbered-gloss(
    header: [Lezgian (Haspelmath 1993:207)],
    source: ([Gila], [abur-u-n], [ferma], [hamišaluğ], [güğüna], [amuq'-da-č.]),
    morphemes: ([now], [they-#obl\-#gen], [farm], [forever], [behind], [stay-#fut\
-#neg]),
    translation: "Now their farm will not stay behind forever.",
)
```

(3) West Greenlandic (Fortescue 1984:127)
palasi=lu niuirtur=lu
priest=and shopkeeper=and
both the priest and the shopkeeper

```
#numbered-gloss(
    header: [West Greenlandic (Fortescue 1984:127)],
    source: ([palasi=lu], [niuirtur=lu]),
    morphemes: ([priest=and], [shopkeeper=and]),
    translation: "both the priest and the shopkeeper",
)
```

(4) Hakha Lai
a-nii -láay
3SG-laugh-FUT
s/he will laugh

```
#numbered-gloss(
    header: [Hakha Lai],
    source: ([a-nii -láay],),
    morphemes: ([3#sg\-laugh-#fut],),
    translation: [s/he will laugh],
)
```

(5) Russian

My	s	Marko	poexa-l-i	avtobus-om	v	Peredelkino
1PL	COM	Marko	go-PST-PL	bus-INS	ALL	Peredelkino
we	with	Marko	go-PST-PL	bus-by	to	Peredelkino

Marko and I went to Perdelkino by bus

```
#numbered-gloss(
    header: [Russian],
    source: ([My], [s], [Marko], [poexa-l-i], [avtobus-om], [v], [Peredelkino]),
    morphemes: ([1#pl], [#com], [Marko], [go-#pst\-#pl], [bus-#ins], [#all],
[Peredelkino]),
    additional-lines: (([we], [with], [Marko], [go-#pst\-#pl], [bus-by], [to],
[Peredelkino]),),
    translation: "Marko and I went to Perdelkino by bus",
)
```

(6) Turkish
çık-mak
come.out-INF
to come out

```
#numbered-gloss(
    header: [Turkish],
    source: ([çık-mak],),
    morphemes: ([come.out-#inf],),
    translation: "to come out",
)
```

(7) Latin
insul-arum
island-GEN-PL
of the islands

```
#numbered-gloss(
    header: [Latin],
    source: ([insul-arum],),
    morphemes: ([island-#gen\-#pl],),
    translation: "of the islands",
)
```

(8) French

aux	chevaux
to-ART-PL	horse.PL

to the horses

```
#numbered-gloss(
    header: [French],
    source: ([aux], [chevaux]),
    morphemes: ([to-#art\-#pl],[horse.#pl]),
    translation: "to the horses",
)
```

(9) German

unser-n	Väter-n
our-DAT-PL	father.PL-DAT.PL

to our fathers

```
#numbered-gloss(
    header: [German],
    source: ([unser-n], [Väter-n]),
    morphemes: ([our-#dat\-#pl],[father.#pl\-#dat.#pl]),
    translation: "to our fathers",
)
```

(10) Hittite (Lehmann 1982:211)

n=an	apedani	mehuni	essandu.
CONN=him	that.DAT.SG	time.DAT.SG	eat.they.shall

They shall celebrate him on that date

```
#numbered-gloss(
    header: [Hittite (Lehmann 1982:211)],
    source: ([n=an], [apedani], [mehuni],[essandu.]),
    morphemes: ([#smallcaps[conn]=him], [that.#dat.#sg], [time.#dat.#sg],
[eat.they.shall]),
    translation: "They shall celebrate him on that date",
)
```

(11) Jaminjung (Schultze-Berndt 2000:92)

nanggayan	guny-bi-yarluga?
who	2DU.A.3SG.P-FUT-poke

Who do you two want to spear?

```
#numbered-gloss(
    header: [Jaminjung (Schultze-Berndt 2000:92)],
    source: ([nanggayan], [guny-bi-yarluga?]),
    morphemes: ([who], [2#du.#A.3#sg.#P\-#fut\-poke]),
    translation: "Who do you two want to spear?",
)
```