



[Home](#) [Input](#) [Output](#) [Usage Manual](#) [Demo](#) [Why archive?](#) [Resources](#)

The Vaultese Falcon: an interlinear text archiving tool.

Developed by Jeremy Fahringer, 2005.

Swarthmore College, Department of Linguistics.

The Vaultese Falcon was developed to bridge the gap between theoretical research on archiving interlinear texts digitally and the linguist in the field (see [Why Archive?](#)). While in no way comprehensive, this tool allows someone with prepared interlinear text data to save that data in an archival form, to take the next step in thinking about data archiving and longevity.

Workflow

You may input standard interlinear text in plain text (using Unicode (UTF-8) and not specialized fonts) separated by whitespace and common morpheme separators ([enter your text here](#)). This plain text input is then processed by the Vaultese Falcon IGT parsing engine, and is output in archival XML (also encoded in Unicode) ([download and view the output here](#)). For a demonstration of the workflow with a provided interlinear text snippet, try the [demo](#).

```
<word> .
  <item type="txt"> </item> .
  <morphemes> .
    <morph> .
      <item type="txt"> kajgamčik </item> .
      <item type="gls"> wonderful </item> .
    </morph> .
    <morph> .
      <item type="txt"> ti-g </item> .
      <item type="gls"> ADJ </item> .
    </morph> .
  </morphemes> .
</word> .
```

1. A snippet from an archival interlinear text data model (in [XML](#)).

1. A Tuvan people's tale called Boktu-Kiriš Bora-Šeelei,

t̪iʋa ulustu - ŋ boktu - kiriš, bora - šeelej dep.
 Tuvan people- GEN Boktu- Kiriš Bora- Šeelei QUOT

2. of which there are many different versions.

d̪ika x̪oj janzi̪ tool- dar- i̪ bar.
 very many different tale - PL- 3 COP

2. A surface display of archival interlinear text data, through an XSL stylesheet.

Scope of Project

The Vaultese Falcon has been developed to handle generally accepted interlinear text forms, but the scope of the tool is purposefully limited (see the [Usage Manual](#) for more detailed information and printer-friendly documentation). Words are separated by whitespace (either tabs or any amount of spacing). It currently separates morphemes in the following way:

- normal morpheme separation by "-"
- clitics by "="
- infixes by "<>"
- compounds by " =" (where the first morpheme of a compound is preceded by "=").

This tool is not intended to solve all of the existing problems in the display and formatting of interlinear text, including non-Western directionality and line overflow.

Compatibility

Cross-browser compatibility is currently limited to input and downloadable output. However, display through XSL stylesheets is still not uniformly supported in all browsers. This site has been designed for optimal functionality in the [Mozilla Firefox](#) browser (tested in versions 1.0 and 1.5). For complete functionality, please enable cookies and Javascript in Firefox.

v1.0 Copyright 2005, Jeremy Fahringer. Swarthmore College, Department of Linguistics.