

Interlinear Text XML and XSL Demonstration

Baden Hughes, Steven Bird and Cathy Bow

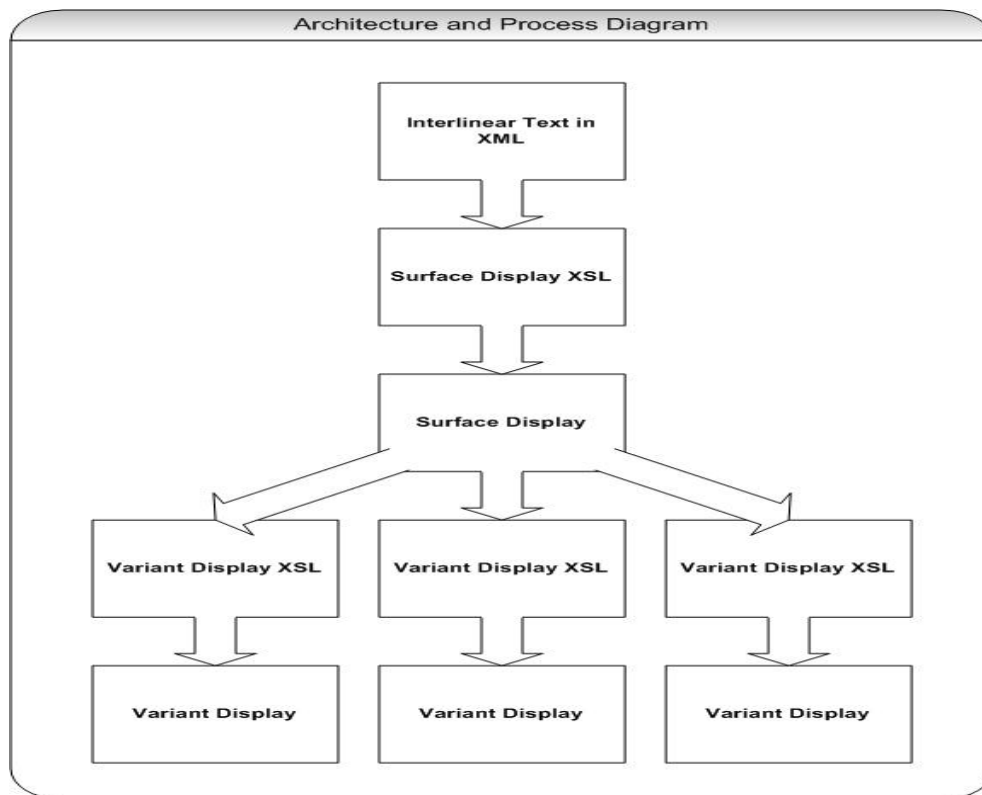
Department of Computer Science and Software Engineering

University of Melbourne, Victoria 3010, Australia

{badenh, sb, cbow}@cs.mu.oz.au

1. Architecture

In demonstrating our model we have adopted a three level architecture which consists of an underlying data representation, a surface display format, and a variant display format. Essentially the processes involved in demonstrating the flexibility of this architecture are the conversion of the underlying data to a surface display, and then converting the surface display to a variant display (see illustration below).



2. Underlying Data

The underlying data is interlinear text structured according to our model and expressed in XML. This underlying data can be validated against an existing DTD.

3. Surface Display

The surface display is a basic display format corresponding to “traditional” interlinear text, which is enabled by the application of an XSL stylesheet to the underlying data. There are two types of surface display we have identified. The first is a simple type, namely direct application of a single XSL stylesheet to an underlying XML document. The second is more complex, including the parameterization of user selected display input which in turn affects the XSL stylesheet and the

corresponding display of the underlying XML document. In the next section, these distinctions form the basis of the categorization of functions.

4. Variant Display

The variant display is a customized display format which demonstrates the flexibility of manipulating the underlying data for different display purposes. For this demonstration we have identified a number of desirable variants based on common linguistic data structures.

4.1 Simple Variant

We have identified two simple variant displays: 1) free translation as separate block and 2) frame interface based expansion of free translation. In the first variant, we manipulate the surface display to format the free translation as a separate block of text from the interlinear content. In the second, we manipulate the surface display to provide the gloss in a separate frame from the interlinearization, and allow synchronized scrolling and linking between the segments of the gloss and the relevant interlinear segments.

4.2 Complex Variant

We have identified a number of complex variant displays based on parameterized input. These are 1) tree view or metastructural view; 2) line re-ordering; 3) optional line display; 4) wordlist linkage, including context; and 5) concordance linkage, including context.

- 1) The *tree-view or metastructural display* essentially allows navigation of the interlinear text using a tree view display format. Individual branches of the tree can be expanded or compressed. This may be useful for structural analysis of the text.
- 2) The *line re-ordering display* allows the selection of a preference for the order of the lines of interlinear text, eg display source text first, display source text last. This may be useful in evaluation contexts for backglossing.
- 3) The *optional line display* allows a selection of preference for how many interlinear lines are displayed. This may be useful for context where features of interest are identified in the interlinear text (eg syntactic vs morphological vs phonological annotation).
- 4) The *word list linkage display* allows the selection of a particular word, and the corresponding display of interlinear content for that word. This may be useful for contexts where particular words are of intermittent interest and detected whilst browsing the source or translation text.
- 5) The *concordance linkage display* allows the selection of a particular, and the corresponding display a list of of all other occurrences of the particular word within the text, including the surrounding words. Any context can be selected, and the complete interlinearization displayed for that context.