Interactive alignment of Parallel Texts – *a cross browser experience*

(standards in practice)

**Gavin Brelstaff** (gjb@ crs4.it)
CRS4 09010 Pula (CA) – Sardinia, Italy

**Francesca Chessa**
University of Sassari, Italy

Multilingual Web Workshop Pisa
April 2011
Introduction
Alignment of parallel texts; multi-lingual; minority languages; poetry

But why?

Dante’s was a minority language.
"Think global, act local"

He was the cat that walked by himself and all places were alike to him.  
*Kipling*

"Think local, act global"

**Genius loci**
the creative spirits of place – *geolocated*.

**Minority language**
a seed-bed for poetic expression, beyond mere communication.

Whenever we lose a language the “genetic basis” for such expression diminishes, globally.
Echo Chamber

*Minority language*

Island language

(song, verse, prose)
Echo Chamber
in poet’s head
Echo Chamber inside the head
(ear,tongue,thought)
Echo Chamber inside the head
(ear, tongue, thought, eye)
Nel mezzo del cammin di nostra vita mi ritrovai per una selva oscura ché la diritta via era smarrita

Language Barrier
Cultural context $A$

Cultural context $B$

cf R.Jakobson

Language Barrier
Minority language

Global language

“cellular membrane”

osmosis

Assist

avoiding dilution, shrivelling, bursting.

Language Barrier
Nel mezzo del cammin di nostra vita mi ritrovai per una selva oscura ché la diritta via era smarrita

Language Barrier
Nel mezzo del cammin di nostra vita mi ritrovai per una selva oscura, ché la diritta via era smarrita.

**Parallel text alignment** ↔ to communicate *semantics*

- standards-based markup
- web delivery, cross-browser
- non-verbal interactivity
- beyond *GoogleTranslate* →

Midway this way of life, we’re bound upon I woke up and found myself in a dark wood.

Midway del cammin di nostra vita mi ritrovai per una selva oscura, ché la diritta via

MLW Pisa 2011
G. Brelstaff & F. Chessa
Beyond *GoogleTranslate*:

- SMT not going to translate poetry well any time soon.
- We allow the translator to clarify by alignment
- Point-&-click interface to modify standard markup
- Colour-code: formal & dynamic equivalence [Nida-Taber]

*Demo*

Parallel text alignment web interface
Demo (a desktop browser: IE8-9, FF3-4, Opera11, Chrome, Safari)
Demo:
selection by click

Una peraula bia;
cando est offerta e nada
ed est sa prima orta
chi dæ bucca bessit,
in s’istante matessi est accabada.
Est cunsumida e morta.

A word lives
when offered up and said
and once issued
from the mouth
it vanishes in that moment:
Consumed and dead.

Una peraula bia;
cando est offerta e nada
ed est sa prima orta
chi dæ bucca bessit,
in s’istante matessi est accabada.
Est cunsumida e morta.

A word lives
when offered up and said
and once issued
from the mouth
it vanishes in that moment:
Consumed and dead.
## Demo: selection & alignment

<table>
<thead>
<tr>
<th>Sardinian Original</th>
<th>English Version - gb 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peraula bia</strong> - Antonino Mura Ena</td>
<td><strong>Word Lives</strong> - Antonino Mura Ena</td>
</tr>
<tr>
<td>Una <em>peraula</em> bia; cando est offerta: e nada ed est sa prima - orta chi dae bucca bessit, in s'istante matessi est accabada. Est cunsumida e morta.</td>
<td>A <em>word</em> lives when offered up and said and once issued from the mouth it vanishes in that moment: Consumed and dead.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sardinian Original</th>
<th>English Version - gb 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peraula bia</strong> - Antonino Mura Ena</td>
<td><strong>Word Lives</strong> - Antonino Mura Ena</td>
</tr>
<tr>
<td>Una <em>peraula</em> bia; cando est offerta: e nada ed est sa prima - orta chi dae bucca bessit, in s'istante matessi est accabada. Est cunsumida e morta.</td>
<td>A <em>word</em> lives when offered up and said and once issued from the mouth it vanishes in that moment: Consumed and dead.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sardinian Original</th>
<th>English Version - gb 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peraula bia</strong> - Antonino Mura Ena</td>
<td><strong>Word Lives</strong> - Antonino Mura Ena</td>
</tr>
<tr>
<td>Una <em>peraula</em> bia; cando est offerta: e nada ed est sa prima - orta chi dae bucca bessit, in s'istante matessi est accabada. Est cunsumida e morta.</td>
<td>A <em>word</em> lives when offered up and said and once issued from the mouth it vanishes in that moment: Consumed and dead.</td>
</tr>
</tbody>
</table>
Standards in practice

<table>
<thead>
<tr>
<th>Presentation</th>
<th>Content</th>
<th>Structure</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>XHTML</td>
<td>XML</td>
<td>TEI-p5</td>
<td>not RDF</td>
</tr>
</tbody>
</table>

Unicode

Interact

- http put
- REST/ajax
- XMLSchema
- eXist XML db
- XQL
- "put"
- "ajax"
- "Schema"
- "db"
- "XQL"

Pros & Cons

- Interact
- "jQuery"
- "XSL"
- "w3cRange"
- "not"
- "DOM"
- "Javascript"
- "XQL"
- "not"
- "RDF"
**Cons: #1**

We can’t interact directly with *Semantics*

Browsers only **bind events** to XHTML (why not XML?) elements

Incurs two degrees of messy indirection.
Cons: #2

**w3cRange** is not “road worthy”.

We resort to *Click to Select*

Selection within words still lacking.
**Cons: #3**

**TEI-p5** must be subsetted to avoid overlapping markup.

We prioritise alignment tags over \{verse-line,paragraph\} hierarchy.
Unicode in XML attributes permits our novel alignment scheme:

The verbatim source text is simply assigned as an attributed of an enclosing tag in the translated text.

Pros: #1

**Unicode** in XML attributes permits our novel alignment scheme:
Pros: #2

CSS selection mechanism as embraced in jQuery helps tame the complexity of cross-browser DOM programming.
Pros: #3

RESTful archiving is a reality due to:

- **Ajax** in the browser
- **Http PUT** on the wire, &
- **eXist XML db** on the server
## Conclusion

### Standards in practice

<table>
<thead>
<tr>
<th>Cons</th>
<th>Pros</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Can’t bind to XML</td>
<td>• Unicode in attributes</td>
</tr>
<tr>
<td>• W3cRange not ready</td>
<td>• CSS&amp;jQuery vs. DOM</td>
</tr>
<tr>
<td>• Must subset TEI-p5</td>
<td>• RESTful reality</td>
</tr>
</tbody>
</table>
Browser issues

- **Opera**: no transparent cursor in text
- **Firefox**: synchronous scroll down bug
- **IE**: onselectstart issue
- **Google Chrome**: run from disk fix
- **Safari/Chrome/IE** Form Enctype: validation
That’s all folks:

Gavin Brelstaff (gjb@crs4.it)
CRS4 09010 Pula (CA) – Sardinia, Italy

Francesca Chessa
University of Sassari, Italy