Going beyond Google Translate?

Ciclo 2012 di seminari interni CRS4  
29/02/2012

Francesca Chessa  
DLS, University of Sassari, Italy

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

CHItaly2011  
13-16 September 2011  
Alghero, Italy
Going beyond Google Translate?
Francesca Chessa  
DLS, University of Sassari  
Sassari (SS)  
Italy.  
fch @ uniss.it

Gavin Brelstaff  
CRS4, Loc. Piscina Manna, Ed. 1  
09010 Pula (CA)  
Italy.  
gjb @ crs4.it

Natural Exploration of 3D Models
Samuel A. Iacolina  
samueliacolina@gmail.com

Alessandro Soro  
asoro@crs4.it

Riccardo Scateni  
riccardo@unica.it

1 Department of Mathematics and Computer Science  
University of Cagliari  
Via Ospedale, 72 09124 - Cagliari, Italy

2 CRS4 - Center for Advanced Studies, Research and Development in Sardinia  
POLARIS Science Park - Ed. 1 - 09010 Pula - Italy
W3C Workshop Report: Content on the Multilingual Web
4-5 April 2011, Pisa, Italy

http://videolectures.net/w3cworkshop2011_brelstaff_interactive/
poetry 2011
poetry corner
poetry corner
poetry corner
poetry corner
poetry corner (UK, 2011)
The Waste Land iPad app earns back its costs in six weeks on the App Store
The Guardian 8 Aug 2011
Los Fills - Maria Chessa Lai
Ailgerese Catalan original

Una llavor de mela
és caiguda
a un fos
entre la terra negra.
La terra
ha tengut
estujada
al seu ventra
per los dies
del temps,
la llavor amagada.
Un arbre
és eixit
però la mela
margant i agr, agra,
és diferent
també que en tot
assimitgi als parcs.

Children - Maria Chessa Lai
English version - gb&ch 2008

An almond seed
into furrow
fell
amongst dark earth.
The earth
she held
it safe and well
within her girth –
day on day
seed
hid away.
A tree
was born
its almonds rather
bitter sharp
differing from
yet similar
to their father.
Echo Chamber –
natural resonances
Sardinia

*Genius loci*
the
creative spirits
of place
– *not just geolocation*.

*Minority language*
a
seed-bed
for poetic expression,
- *not just ICT communication*.

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

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

- Alghero
- Poetry
- HCI
Los Fills - Maria Chessa Lai
Aigüeres Catalan original

Una llavor de mela
és caiguda
a un fos
entre la terra negra.
La terra
ha tengut
estujada
al seu ventra
per los dies
del temps,
la llavor amagada.
Un arbre
és eixit
però la mela
margant i agra,
éss diferent
també que en tot
assimitgi als parcs.

Children - Maria Chessa Lai
English version - gb&ch 2008

An almond seed
into furrow
fell
amongst dark earth.
The earth
she held
it safe and well
within her girth –
day on day
seed
hid away.
A tree
was born
its almonds rather
bitter sharp
differing from
yet similar
to their father.
Los Fills - Maria Chessa Lai
Algherese Catalan original

Una llavor de mela
és caiguda
a un fós
entre la terra negra.
La terra
ha tengut
estujada
al seu ventra
per los dies
del temps,
la llavor amagada.
Un arbre
és eixit
però la mela
margant i agra,
és diferent
també que en tot
assimitgi als pares.

Children - Maria Chessa Lai
English version - gjb.fch 2008

Una → An de_mela → almond llavor → seed
a → into fos → furrow
es_caiguda → fell
entre → amongst negra → dark terra → earth.
La → The terra → earth
she held
estujada → it safe and well
al → within seu → her ventra → girth → per_los_dies_del_temps → day on day
llavor → seed
amagada → hid away.
Un → A arbre → tree
és_eixit → was born
la_mela → its almonds però → rather
margant → bitter agra → sharp
és_diferent → differing from
també_que_en_ito → yet assimitgi → similar
als → to their pares → father.
Una llavor de mela és caiguda a un fos entre la terra negra.
An almond seed into furrow fell amongst dark earth.
Poetry is an extreme challenge

Google Translate: is pitiful when read aloud as poetry

Echo Chamber (ear, tongue, thought, eye)
HCI and SMT

We are not doing statistical *machine* translation (SMT) here

A *human* translates each poem and marks up the equivalences at three different levels: *word, phrase, idea*.

Spatio-visual cues are thus obtained.
HCl and SMT?

Spatio-visual cues are thus available.

Q: Can such cues provide an **intermediate representation** that can be usefully manipulated by SMT?

Q: Might an algorithm compute **emergent semantic context** from them?

---

<table>
<thead>
<tr>
<th>del temps,</th>
<th>seed</th>
</tr>
</thead>
<tbody>
<tr>
<td>la llavor amagada.</td>
<td>hid away.</td>
</tr>
<tr>
<td>Un arbre</td>
<td>A tree</td>
</tr>
<tr>
<td>és eixit</td>
<td>was born</td>
</tr>
<tr>
<td>però la mela</td>
<td>its almonds</td>
</tr>
<tr>
<td>margant i agra,</td>
<td>rather</td>
</tr>
<tr>
<td>és diferent</td>
<td>bitter sharp</td>
</tr>
<tr>
<td>també que en tot</td>
<td>differing from</td>
</tr>
<tr>
<td>assimitgi als pares.</td>
<td>yet similar</td>
</tr>
<tr>
<td>Los fills</td>
<td>to their father.</td>
</tr>
<tr>
<td></td>
<td>Children,</td>
</tr>
</tbody>
</table>
Algorithm

Common spatio-visual representation

Out

In

Emergent semantic context?

Translator

Semantic context

view

point & click

HCI and SMT?
**HCI: design factors**

- non-verbal interface – *non-linguistic thinking*
- facilitate *smooth switching of gaze* between parallel texts via the secondary highlighting
- text-selection by *point & click* at words
- *see through cursor*-sprite when on text
interactive alignment: steps

- atomic segmentation \{word, phrase, idea\}
- restoring state
- *select* by click
- *merge* operation
- *alignment* between parallel texts
- *split* operation
- review

jQuery implementation
Demo (a desktop browser: IE8-9, FF3-10, Opera11, Chrome, Safari)
Demo: selection by click

Una peraula 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.

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

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

F.Chessa & G.Brelstaff
### Presentation
- XHTML
- CSS

### Content
- XML

### Structure
- TEI-p5

### Semantics
- not RDF

---

#### Interact
- not w3cRange
- XSL
- jQuery

#### Javascript & DOM

---

**now go mobile...**
The Study.

Museum.

The Study. 1. is a place. 2. a part from men.

addicted to his studies, whilst he readeth

Books, 2. which having within his reach, he turneth upon a Desk 4. and picketh all the best things out of them into his own Manual, 5.

Museum. 1. est locus, ubi Studiorum, 2.-secretus ab hominibus, folius sedet, Studiorum spectat, dum lectitatur Libros, 3. quos posset certe etiam in Manuscrib 5. suum optima quas excerptit.
Conclusion

- *Alghero* html app for aligning parallel texts
- genric *multilingual web* application
- text highlights = *spatio-visual* cues
- intermediate representation for semantic context?
- poetry: extreme challenge for SMT
- poetry: a new market?
Follow up:

Francesca Chessa
DLS, University of Sassari, Italy

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