

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



A.D. MDLXII



F.Chessa & G.Brelstaff

1



Data and text mining

Bag of Peaks: interpretation of NMR spectrometry

Gavin Brelstaff^{1,*}, Manuele Bicego^{2,†}, Nicola Culeddu³ and Matilde Chessa^{4,‡}

¹Biocomputing, CRS4, 09100 Pula (CA), Sardinia, ²DEIR, University of Sassari, via Torre Tonda 34, 07100 Sassari,

³ICB-CNR, 07040 Li Punti, Sassari and ⁴Porto Conte Ricerche, Loc. Tramariglio, Alghero, Italy

<http://bioinformatics.oxfordjournals.org/content/25/2/258.full>

PERCEPTION

[Search](#) [Current issue](#) [Forthcoming](#) [All volumes](#) [Perception homepage](#) [ECVP](#) [Pion homepage](#)

2011 volume 40(11) pages 1282 – 1289

doi:10.1068/p6916

Cite as:

Lorusso L, Brelstaff G, Brodo L, Lagorio A, Grosso E, 2011, "Visual judgments of kinship: An alternative perspective" *Perception* 40(11) 1282 – 1289

[Download citation data in RIS format](#)

Visual judgments of kinship: An alternative perspective

Ludovica Lorusso, Gavin Brelstaff, Linda Brodo, Andrea Lagorio, Enrico Grosso

today

CHIItaly2011

13-16 September 2011
Alghero, Italy

F.Chessa & G.Brelstaff

CHIItaly2011

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 ¹
samueliacolina@gmail.com

Alessandro Soro ^{1,2}
asoro@crs4.it

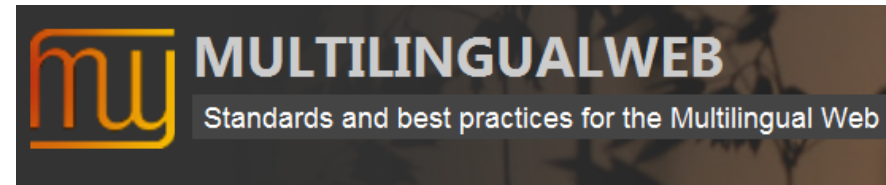
Riccardo Scateni ¹
riccardo@unica.it

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/



ugr

May 2011

Universidad
de **Granada**

poetry 2011



poetry corner

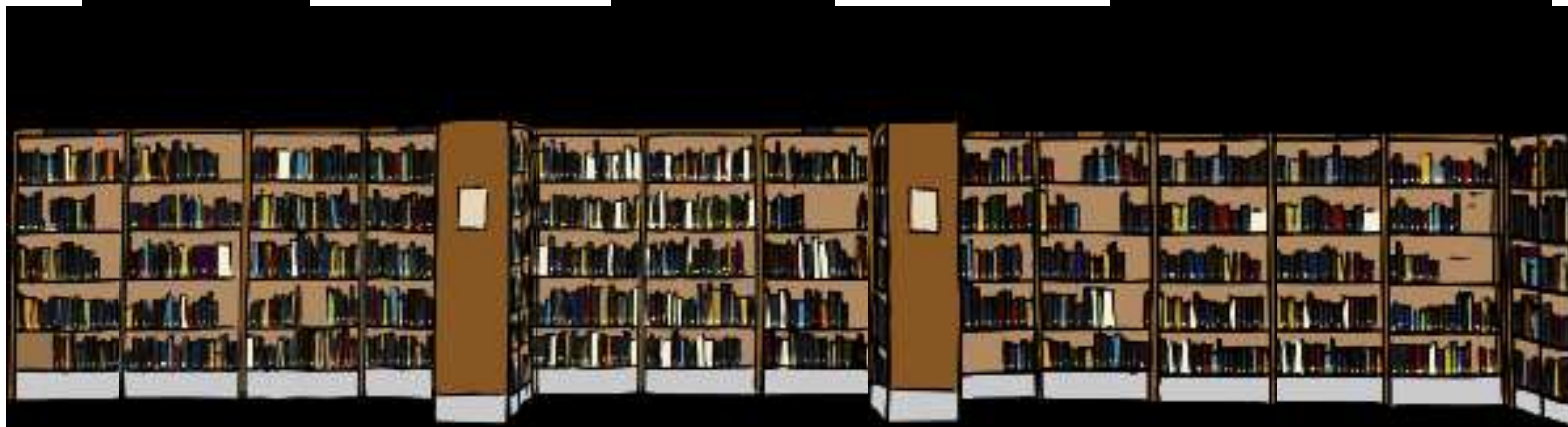


poetry corner

Fiction

Poetry

Healthy Living

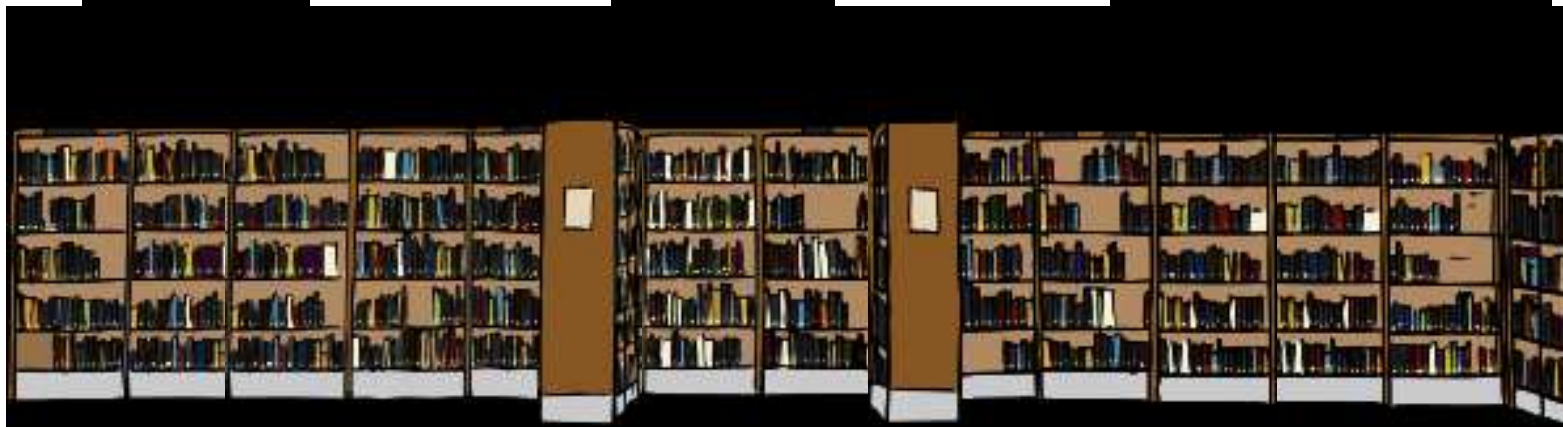


poetry corner

Fiction

Poetry

Healthy Living

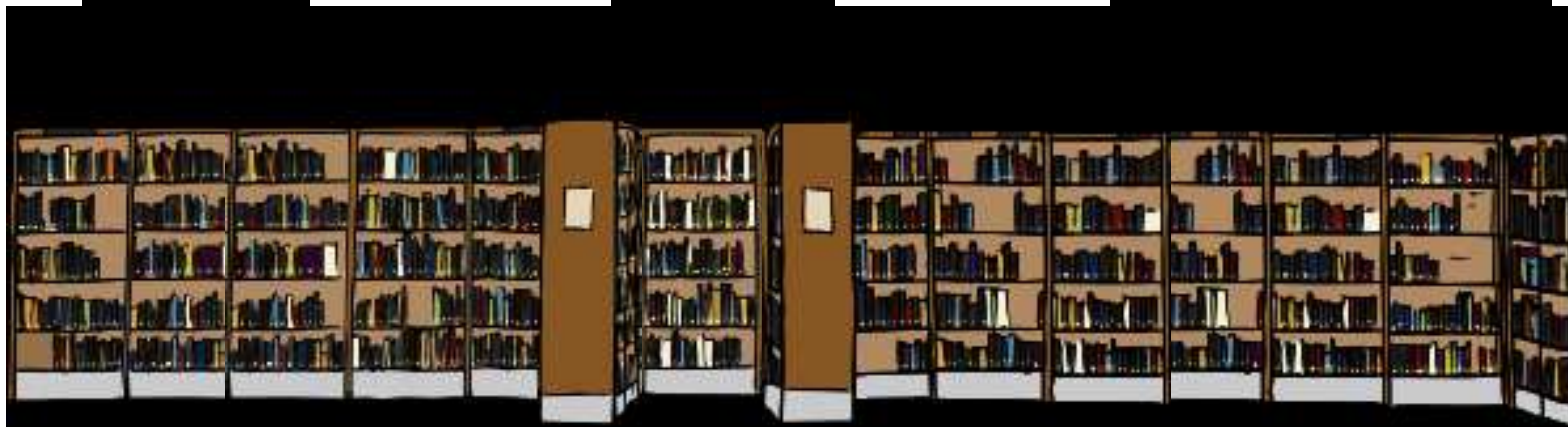


poetry corner

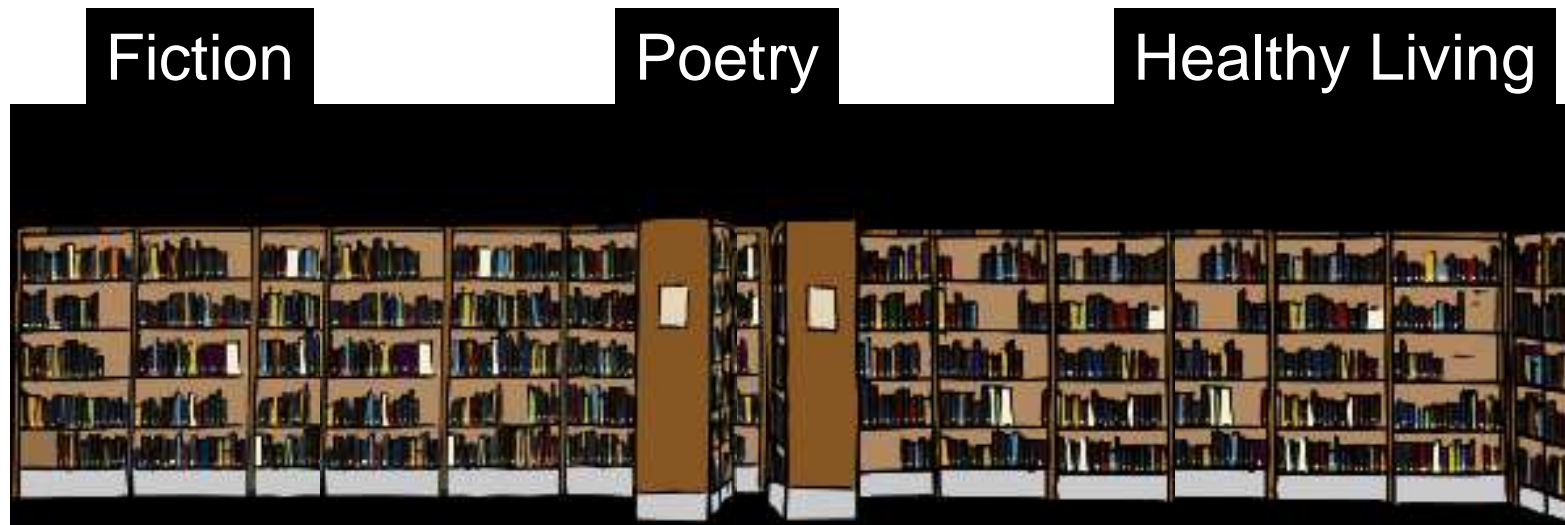
Fiction

Poetry

Healthy Living



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



Alghero app - html

The screenshot displays a web application interface with two columns of text. The left column contains the original Catalan poem, and the right column contains its English translation. The word 'fos' in the Catalan text and 'furrow' in the English text are highlighted in green. A mouse cursor is positioned over the word 'furrow'. The interface includes navigation icons at the top of each column and a vertical scrollbar between them.

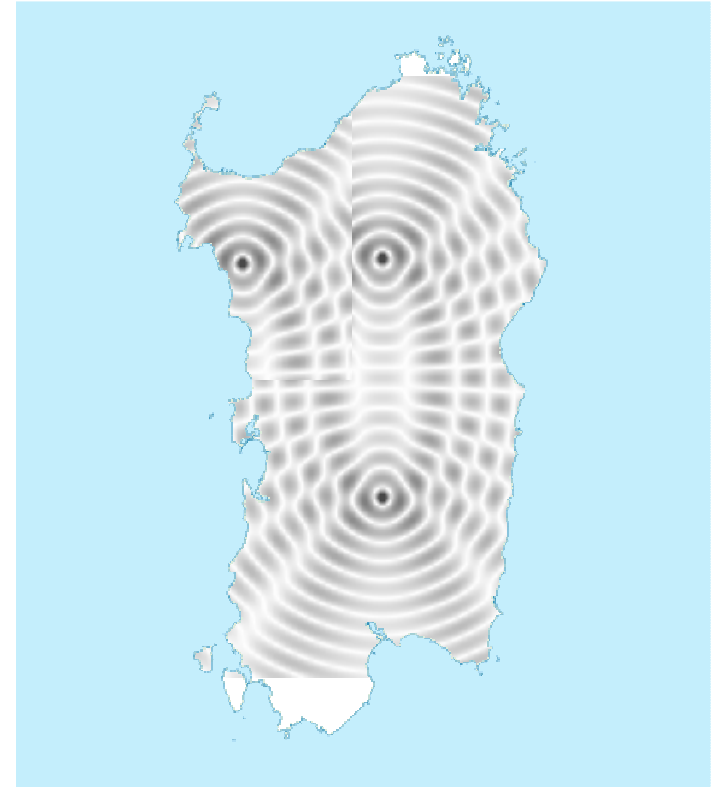
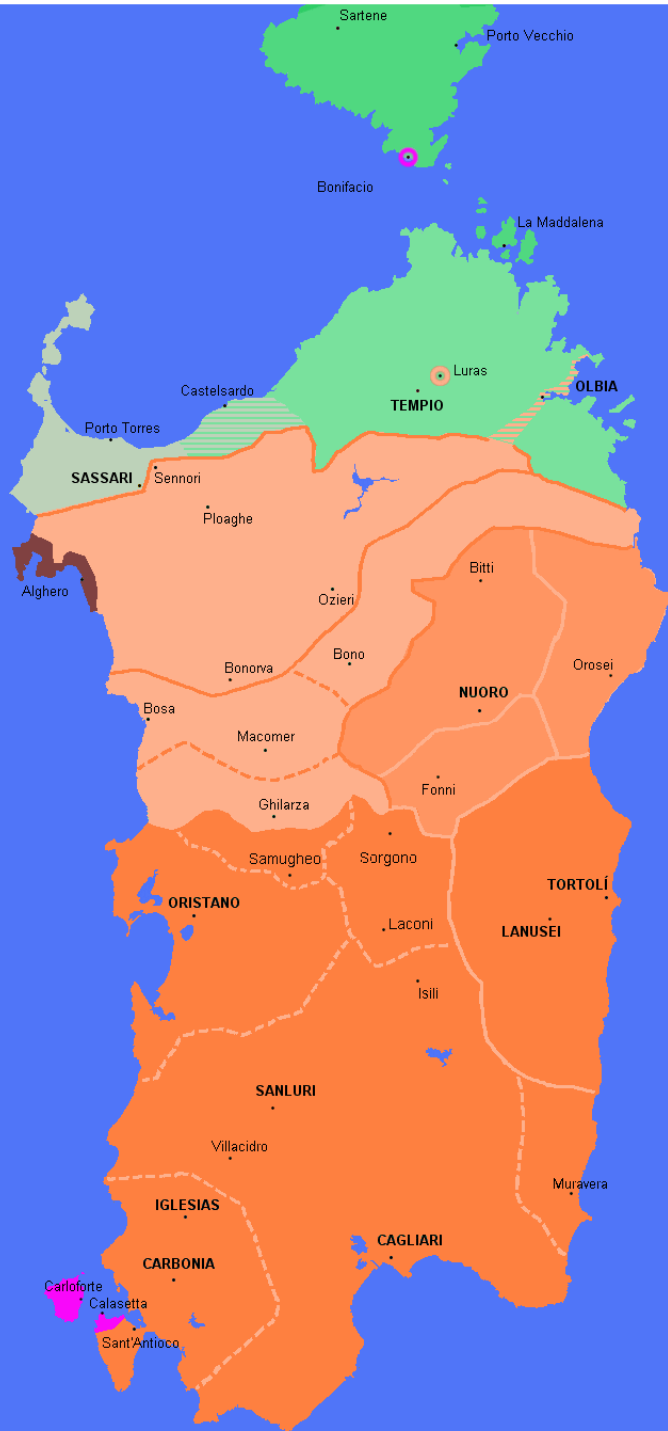
Los Fills - Maria Chessa Lai
Algherese Catalan original

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,
és diferent
també que en tot
assimitgi als pares.

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

Le lingue della Sardegna



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

Internet economy - www



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

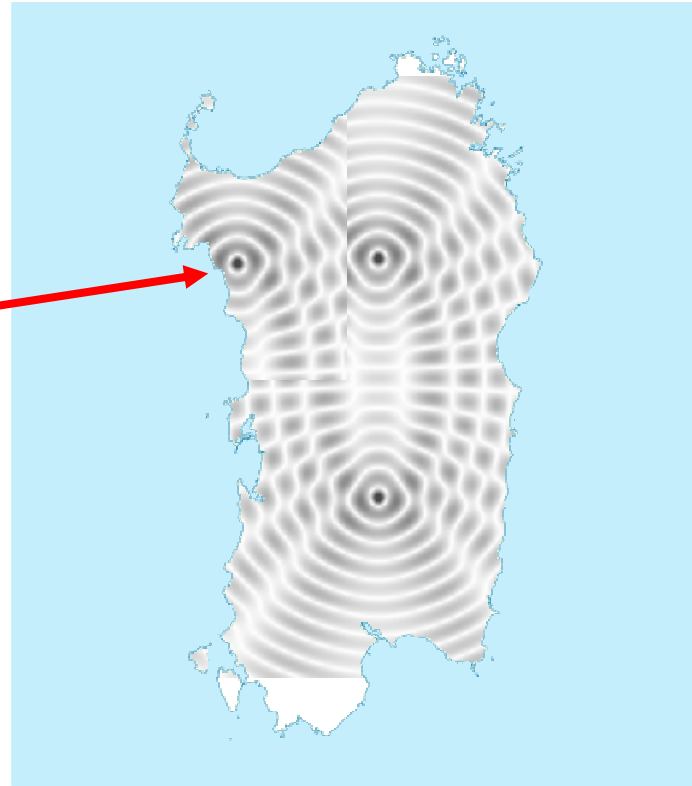
Kipling

This talk:

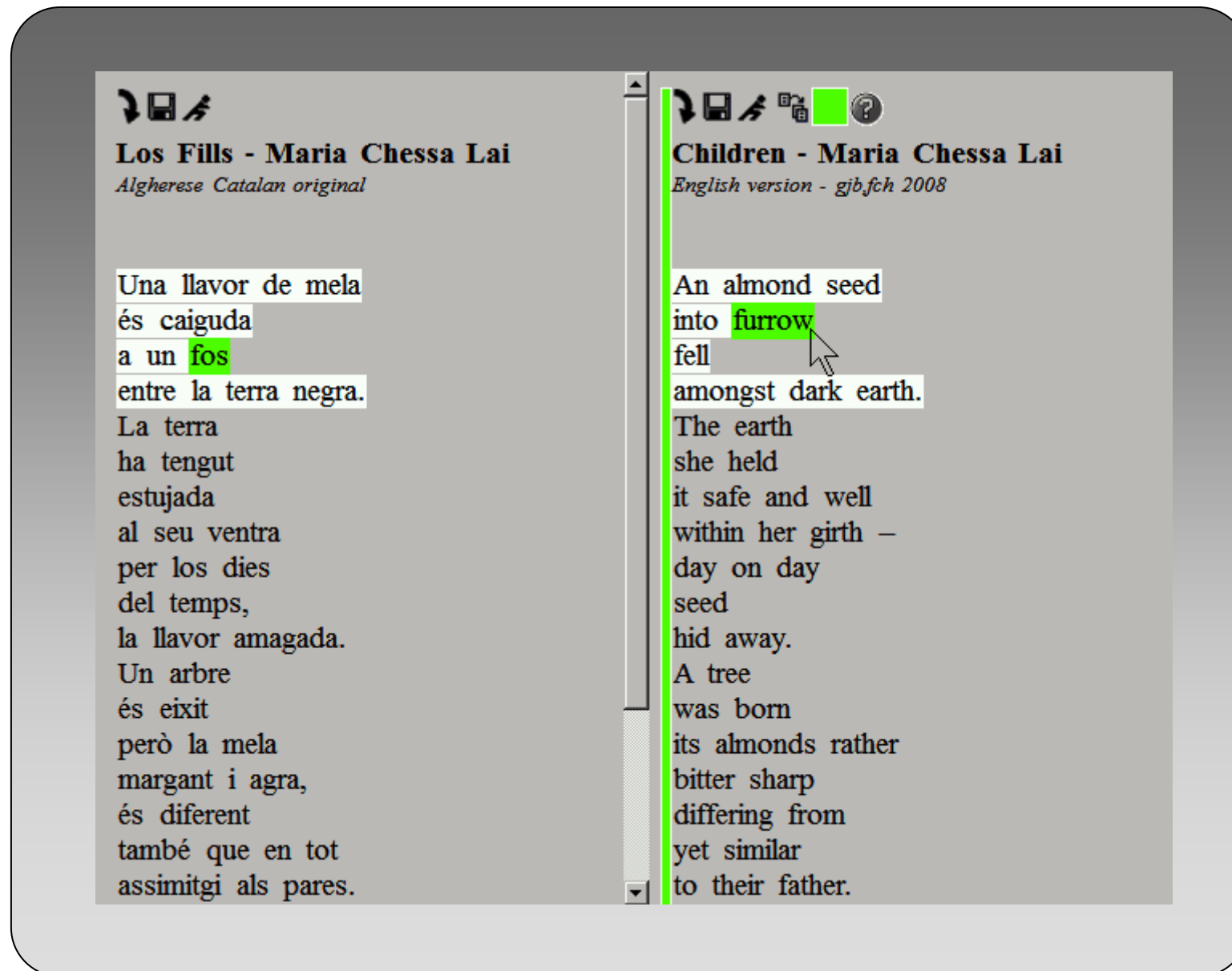
Alghero

Poetry

HCI

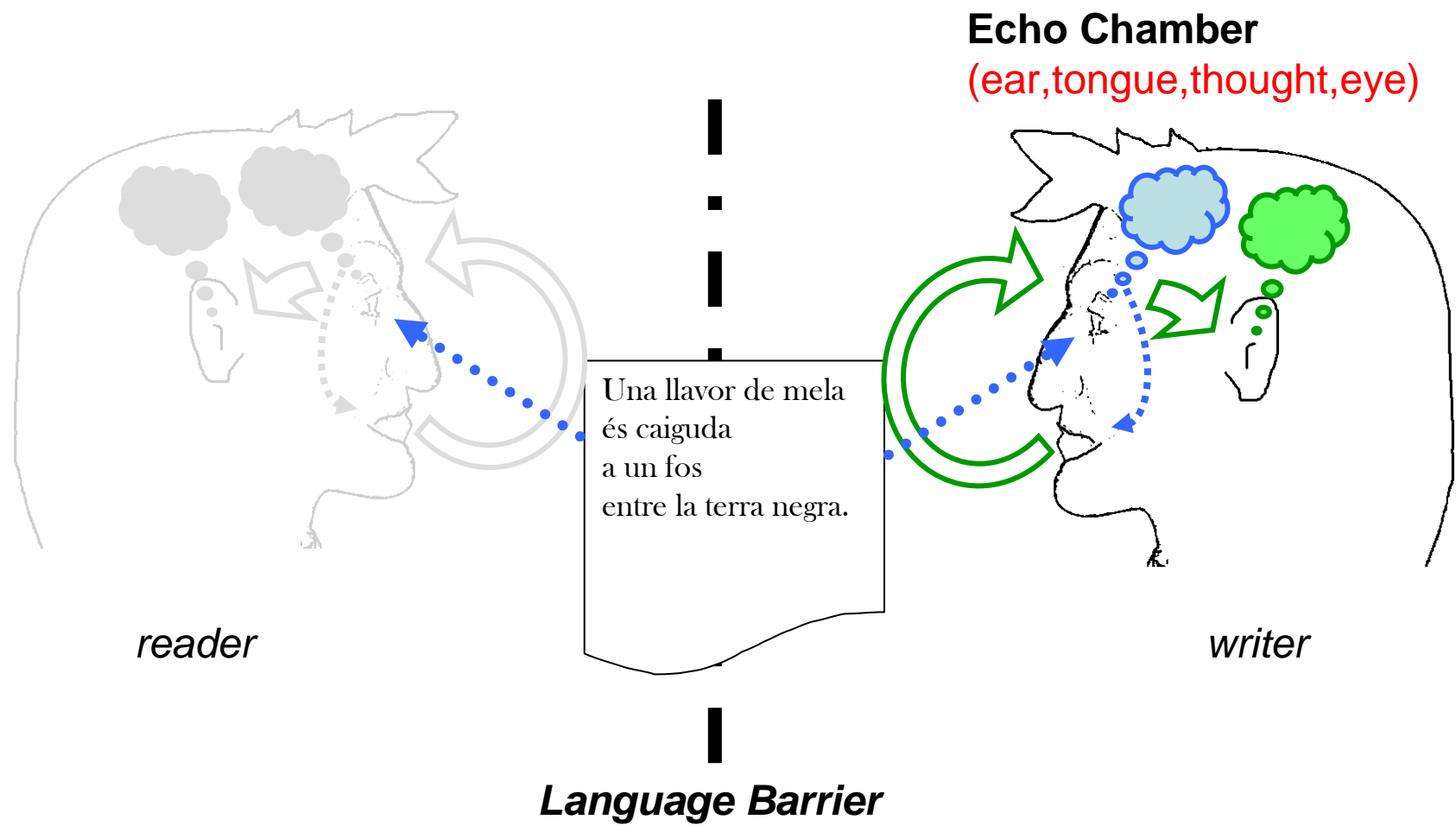


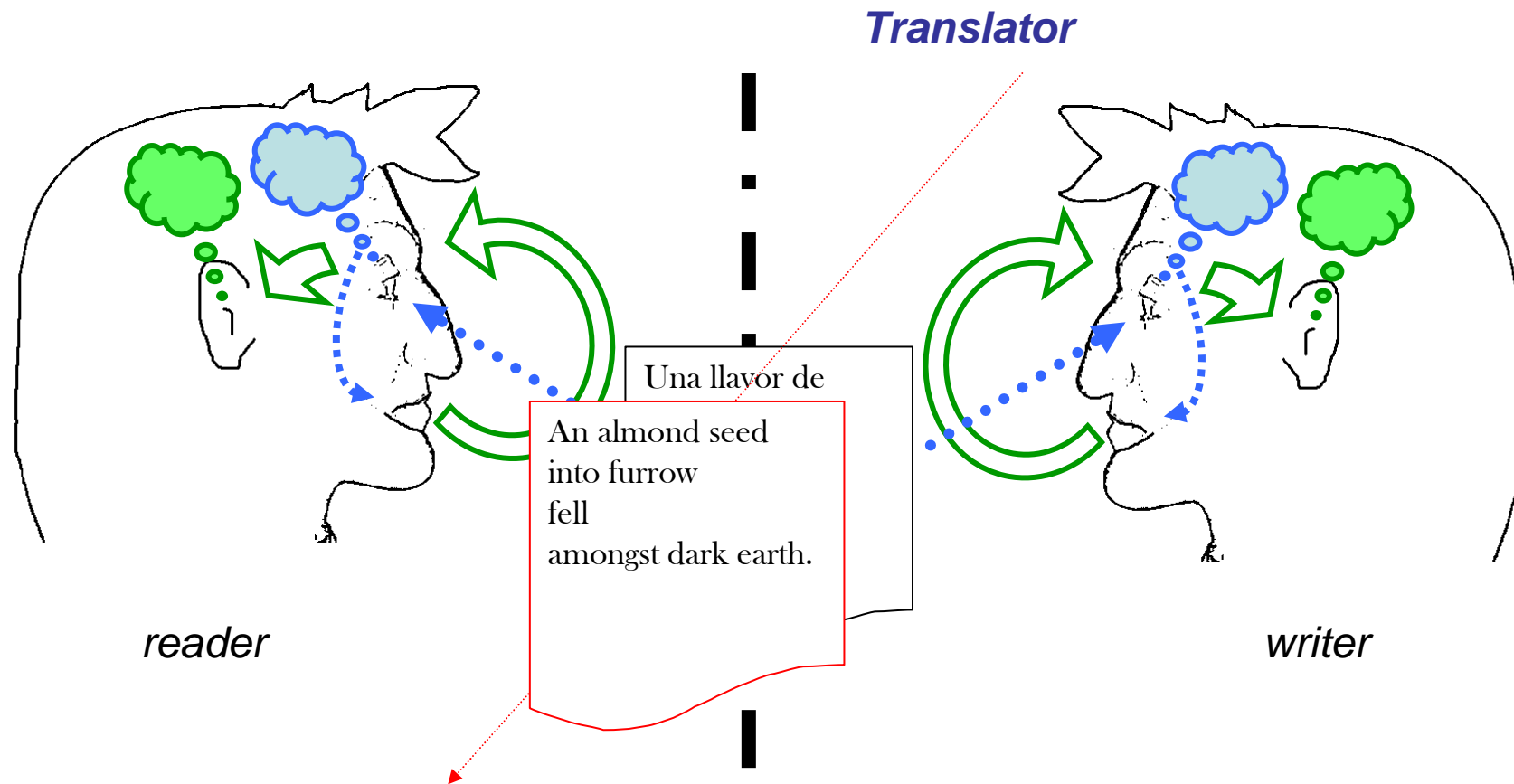
Screenshot



Resulting alignment

<p>↶ ↷ ↵</p> <p>Los Fills - Maria Chessa Lai <i>Algherese Catalan original</i></p> <p>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, és diferent també que en tot assimitgi als pares.</p>	<p>↶ ↷ ↵ ↶ ↷ ?</p> <p>Children - Maria Chessa Lai <i>English version - gjb.fch 2008</i></p> <p><i>Una</i>→An <i>de_mela</i>→almond <i>llavor</i>→seed <i>a</i>→into <i>fos</i>→furrow <i>és_caiguda</i>→fell <i>entre</i>→amongst <i>negra</i>→dark <i>terra</i>→earth. <i>La</i>→The <i>terra</i>→earth she held <i>estujada</i>→it safe and well <i>al</i>→within <i>seu</i>→her <i>ventra</i>→girth – <i>per los dies del temps</i>→day on day <i>llavor</i>→seed <i>amagada</i>→hid away. <i>Un</i>→A <i>arbre</i>→tree <i>és_eixit</i>→was born <i>la_mela</i>→its almonds <i>però</i>→rather <i>margant</i>→bitter <i>agra</i>→sharp <i>és_diferent</i>→differing from <i>també_que_en_tot</i>→yet <i>assimitgi</i>→similar <i>als</i>→to their <i>pares</i>→father.</p>
---	---





Parallel text alignment ↔ to communicate **semantics**

- standards-based markup (TEI, XML), html delivery
- going beyond *Google Translate* →

<p>F.Chessa</p> <p>An almond seed into furrow fell amongst dark earth.</p>	<p>Una llavor de mela és caiguda a un fos entre la terra negra.</p>
---	--

Translate

From: Catalan ▼ ↔ To: English ▼ Translate

Click for alternate translations

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, és diferent també que en tot assimitgi als pares. Los fills que neixen quan venen al món la mare los fa mes són altra gent.	A seed mela is fallin' to be between the black earth. land has had estujada its ventro for these days time, seed hidden. a tree is left there but the mela margins and sour, different also that all assimitgi parents. the sons born when sold to the world mother makes them are most other people.
--	---

Poetry
is an
extreme challenge

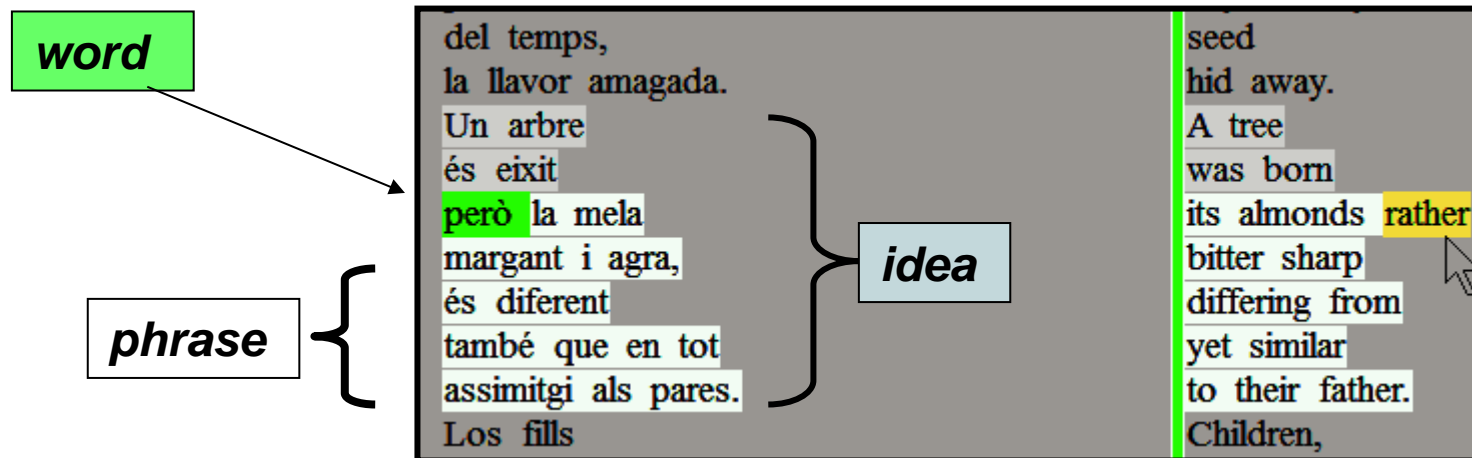
Google *Translate*:
is pitiful when read
aloud as poetry

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.

HCI and SMT ?

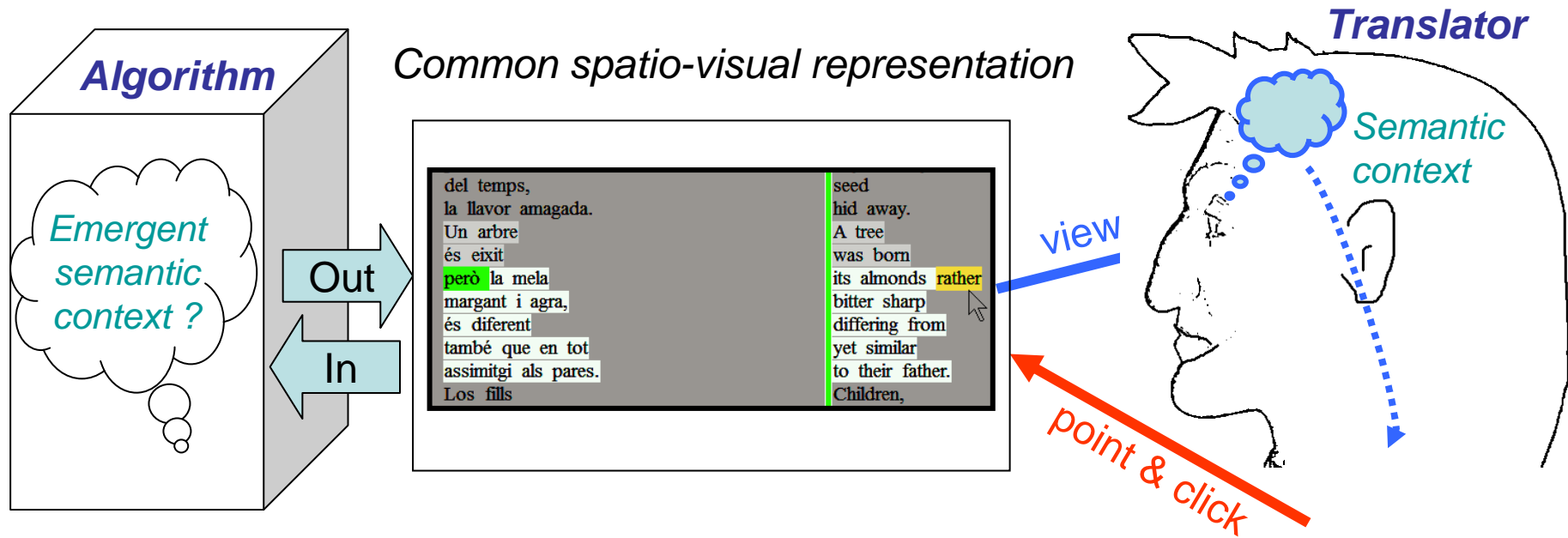
del temps,	seed
la llavor amagada.	hid away.
Un arbre	A tree
és eixit	was born
però la mela	its almonds rather
margant i agra,	bitter sharp
és diferent	differing from
també que en tot	yet similar
assimitgi als pares.	to their father.
Los fills	Children,

Spatio-visual cues are thus available.

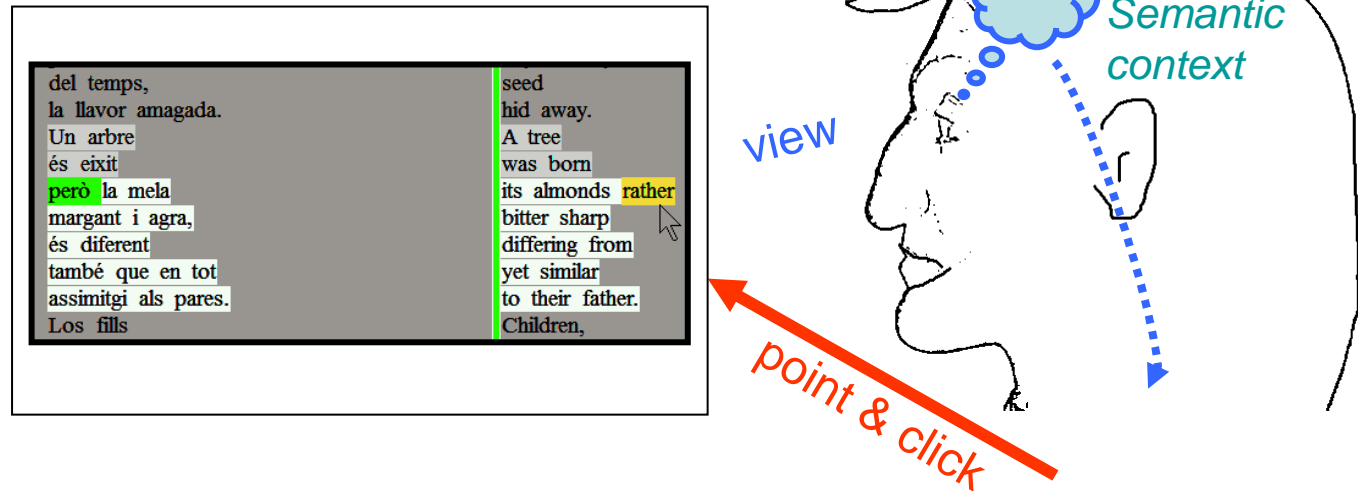
Q: Can such cues provide an *intermediate representation* that can be usefully manipulated by SMT?

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

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)

MLW 2011 demo: Brelstaff, Chessa, Interactive Alignment Of Parallel Text
Gavin Brelstaff and Francesca Chessa, 2011



**Antonino
Mura Ena**

N.Tanda (editor)

[Sardinian-English](#)

(aligned)



**Ольга
Седакова**

F.Chessa

[Russian-Italian](#)

(aligned)



**师涛
(Shi Tao)**

PEN Sardinia

[Chinese-English](#)

(aligned)



**Enrica
Ferreccio**

Fch,Gjb

[Italian-English](#)

(aligned)



**Imp.
Hadrianus**

Fch,Gjb

[Latin-English](#)

(aligned)

**Demo:
selection by click**

Peraula bia - Antonino Mura Ena
Sardinian original

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.

Word Lives - Antonino Mura Ena
English version - gjb 2009

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

Peraula bia - Antonino Mura Ena
Sardinian original

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.

Word Lives - Antonio Mura Ena
English version - gjb 2009

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

Peraula bia - Antonino Mura Ena
Sardinian original

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.

Word Lives - Antonino Mura Ena
English version - gjb 2009

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

Peraula bia - Antonino Mura Ena

Sardinian original

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.

Word Lives - Antonino Mura Ena

English version - gjb 2009

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

Peraula bia - Antonino Mura Ena

Sardinian original

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.

Word Lives - Antonino Mura Ena

English version - gjb 2009

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

Peraula bia - Antonino Mura Ena

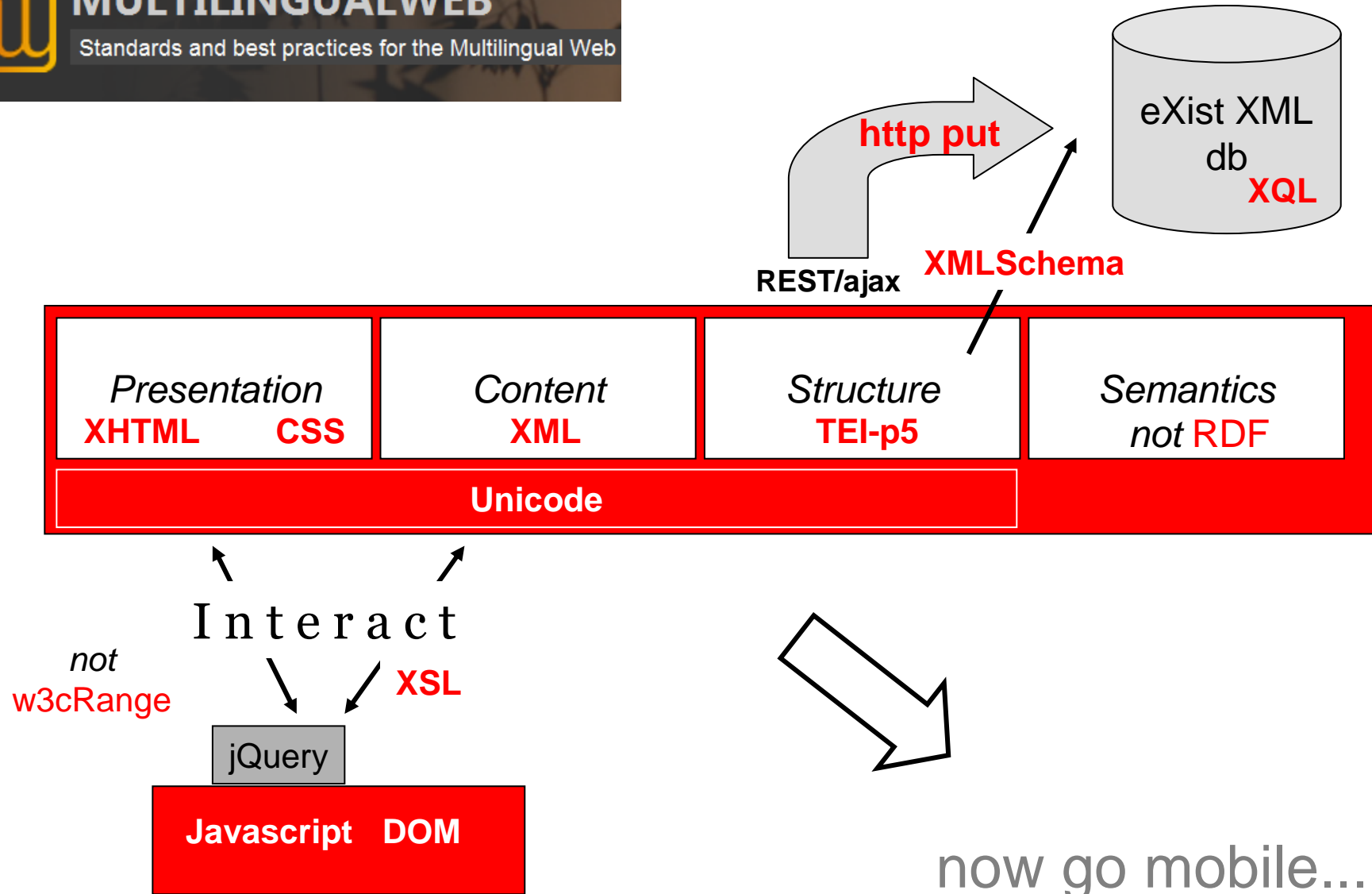
Sardinian original

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.

Word Lives - Antonino Mura Ena

English version - gjb 2009

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



Text Encoding Initiative – the TEI way

(200)

XCV III.
The Study. *Muséum.*



The Study 1.
fo a place
where a Student, 2.
a part from men,
sitteth alone,
addiccd to his Studies,
whilst he readeth
Books, 3.
which being within
his reach, he layeth
open upon a Desk 4.
and picketh all the
best things out of
them into his own
Manual, 5.

Muséum 1.
est locus,
ubi *Studiosus*, 2.
secretus ab hominibus,
solus sedet,
Studiis deditus,
dum lectitat
Libros, 3.
quos penes se
super *Platenum* 4,
exponit, & ex illis
in *Manuale* 5. suum
optima quæq; excerpit,
02

```
<div xml:id="e98" xml:lang="en" type="lesson">
  <head>The Study</head>
  <p>
    <seg xml:id="e9801">The Study</seg>
    <seg xml:id="e9802">is a place</seg>
    <seg xml:id="e9803">where a Student,</seg>
    <seg xml:id="e9804">a part from men,</seg>
    <seg xml:id="e9805">sitteth alone,</seg>
    <seg xml:id="e9806">addicted to his Studies,</seg>
    <seg xml:id="e9807">whilst he readeth</seg>
    <seg xml:id="e9808">Books,</seg>
  </p>
</div>
<div xml:id="l98" xml:lang="la" type="lesson">
  <head>Muséum</head>
  <p>
    <seg xml:id="l9801">Museum</seg>
    <seg xml:id="l9802">est locus</seg>
    <seg xml:id="l9803">ubi Studiosus,</seg>
    <seg xml:id="l9804">secretus ab hominibus,</seg>
    <seg xml:id="l9805">solus sedet,</seg>
    <seg xml:id="l9806">Studiis deditus,</seg>
    <seg xml:id="l9807">dum lectitat</seg>
    <seg xml:id="l9808">Libros,</seg>
  </p>
</div>
```

```
<linkGrp type="alignment">
  <link target="#e9801 #l9801 #p981"/>
  <link target="#e9802 #l9802"/>
  <link target="#e9803 #l9803 #p982"/>
  <link target="#e9804 #l9804"/>
  <link target="#e9805 #l9805"/>
  <link target="#e9806 #l9806"/>
  <link target="#e9807 #l9807"/>
  <link target="#e9808 #l9808 #p983"/>
</linkGrp>
```

Conclusion

- *Alghero* html *app* for aligning parallel texts
- generic *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



A.D. MDLXII



F.Chessa & G.Brelstaff



aqua mare

33