

Coedition of texts and UNL graphs to share text revision across languages and improve MT a posteriori

*International Conference on Universal Knowledge and
Language (ICUKL2002), Goa, 25-29 November 2002*

Christian BOITET & TSAI Wang-Ju

GETA, CLIPS, IMAG, Grenoble

Christian.Boitet@imag.fr, Wang-Ju.Tsai@imag.fr

<http://www-clips.imag.fr/geta/User/wang-ju.tsai/welcome.html>

Plan

- ✍ Motivation : to achieve quality MT in a multi-source & multi-target environment
- ✍ Why a pivot architecture and why UNL?
- ✍ The prototype under construction :
 - principle : divide the correspondence (graph \leftrightarrow text) in 3 parts.
 - interface : necessity of modes, from « naive » to « expert »
- ✍ A problem : computing the 2nd correspondence
 - UNL-tree \leftrightarrow morpho-syntactic lattice
- ✍ A « static » demo on one simple sentence

Motivation : build multisource & multitarget "quality MT"

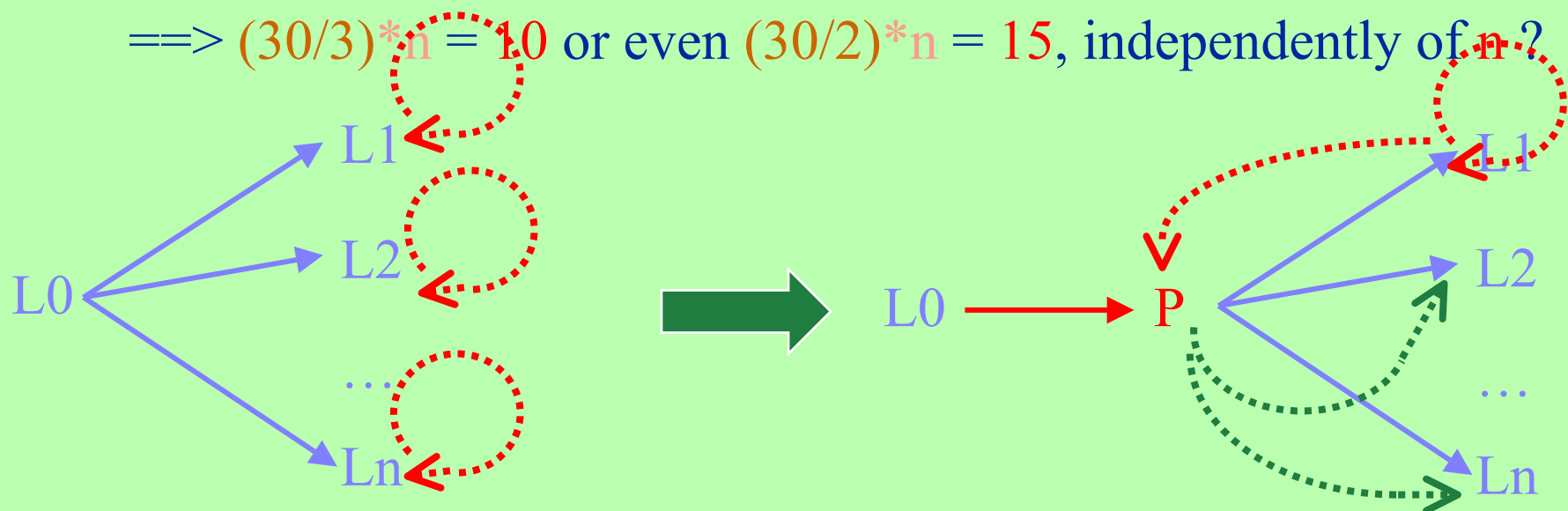
✍ intrinsic limits of the classic procedure

30 standard pages of 250 words, in hours (trans+rev):

✍ $40*n$ by THum, $(12/3+12+12/3)*n = 20 n$ by THAM
($6=, 12\approx, 12\neq$)

✍ $8*n$ with THAM + MAT-R, and $(30/3)*n = 10*n$ if MAT-R only

$\implies (30/3)*n = 10$ or even $(30/2)*n = 15$, independently of n ?



Basic Idea : to edit the « pivot » indirectly, by editing the text

✍ examples of successful « co-edition »

Ambassador (old), MULTIMETEO

✍ 2 problems:

✍ To build the correspondence

abstract structure ↔ concrete text

Without building new and costly resources

✍ To construct a really usable interface

Why use UNL as the pivot ?

✍ brief reminder

UNL is

- ✍ a project
- ✍ an artificial language
- ✍ a format of a multilingual document

✍ the UNL language has unique features
even if it is perfectible !

Language : a simple UNL graph (a) linear form, read by computer

Ronaldo has headed the ball into the left corner of the goal

```
agt(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,  
    Ronaldo(icl>proper noun))  
obj(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,  
    goal(icl>thing))  
ins(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,  
    head(pof>body).@def)  
plt(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,  
    corner(icl>thing).@def)  
pos(corner(icl>thing).@def, goal(icl>thing))  
pos(head(pof>body).@def, Ronaldo(icl>proper noun))  
mod(corner(icl>thing).@def, left(aoj<thing))
```

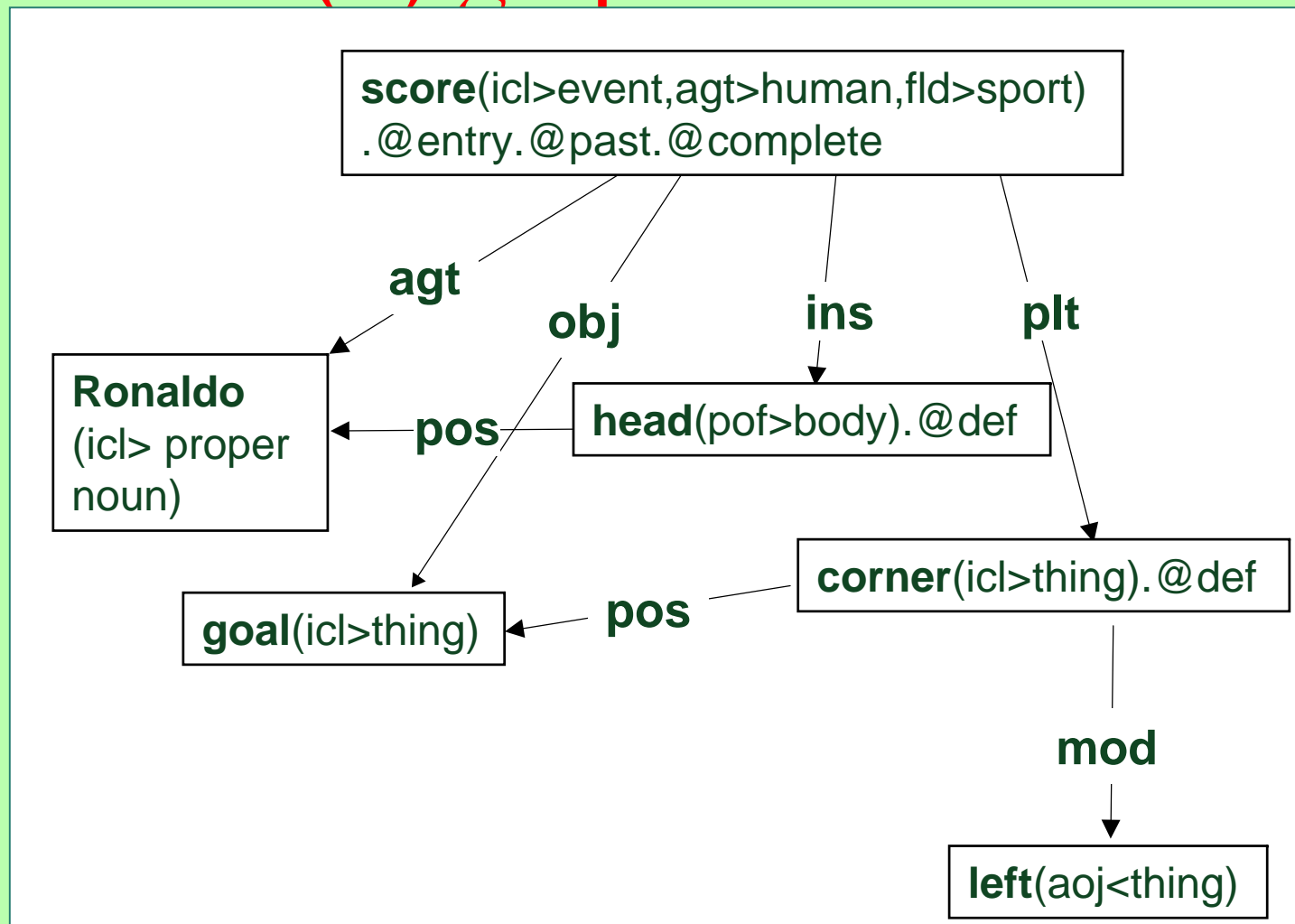
Language : a simple UNL graph

(b) linear form, with colors

Ronaldo has headed the ball into the left corner of the goal

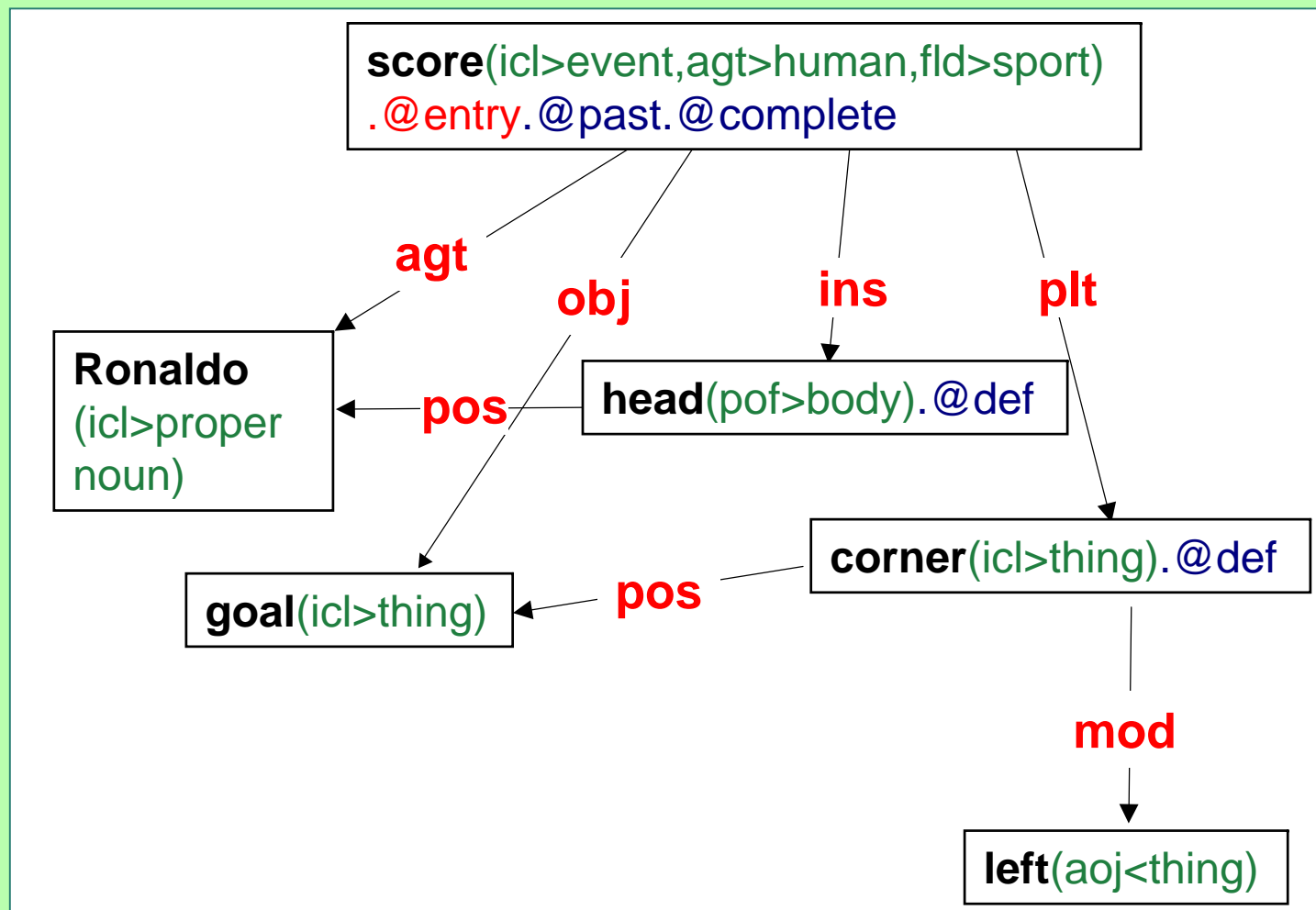
- **agt**(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,
Ronaldo(icl>proper noun))
 - **obj**(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,
goal(icl>thing))
 - **ins**(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,
head(pof>body).@def)
 - **plt**(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,
corner(icl>thing).@def)
 - **pos**(corner(icl>thing).@def, goal(icl>thing))
 - **pos**(head(pof>body).@def, Ronaldo(icl>proper noun))
 - **mod**(corner(icl>thing).@def, left(aoj<thing))
- Is this UNL graph OK? Let us draw it.

Language : a simple UNL graph (c) graphical form



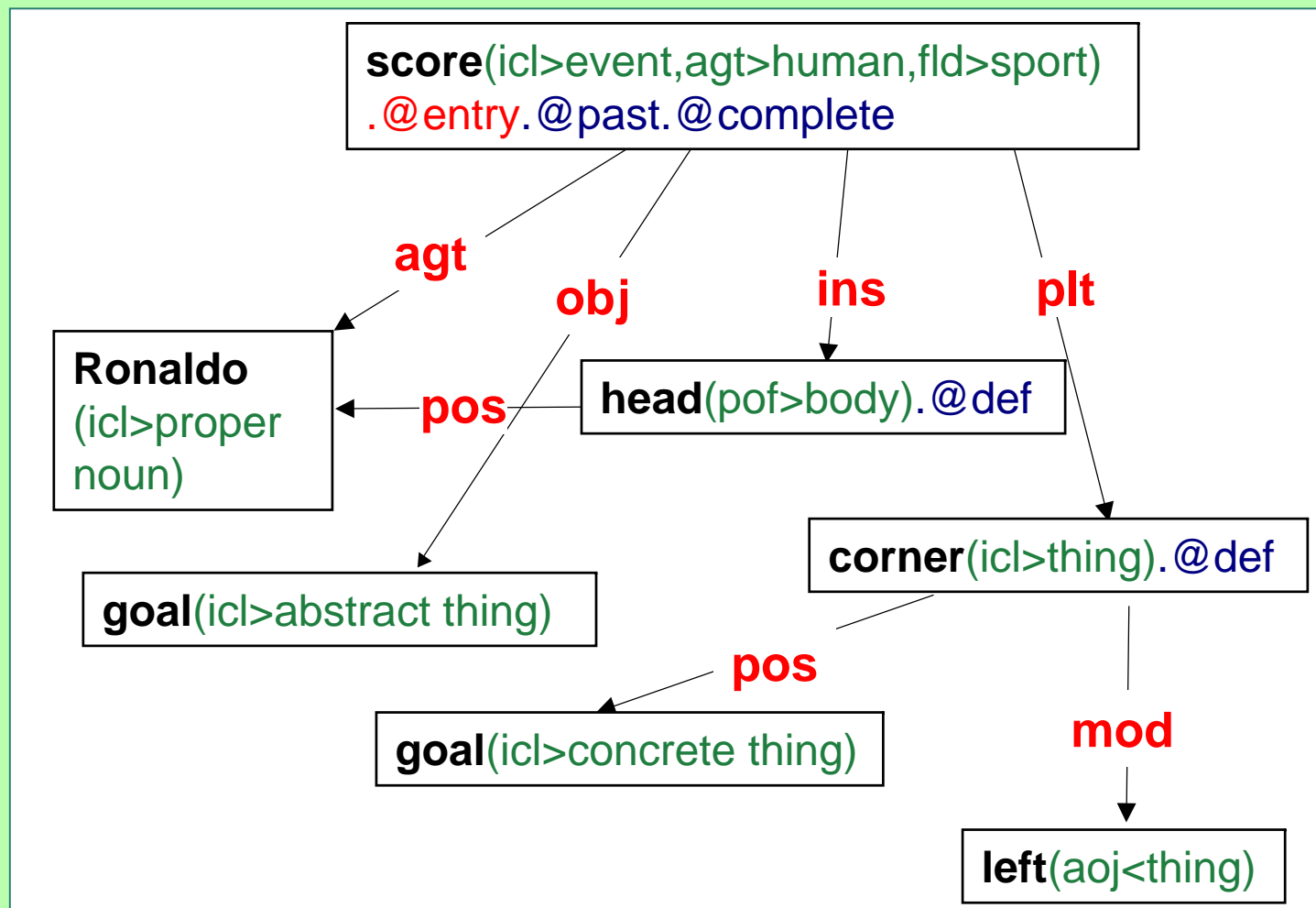
•==> Ronaldo has headed the ball into the left corner of it

Language : a simple UNL graph (c) graphical form, colored



•==> Ronaldo has headed the ball into the left corner of it

Language : a simple UNL graph (d) graphical form, corrected!



•Ronaldo has headed the ball into the left corner of the goal <== OK!

Language : a simple UNL graph (e) linear form, corrected

Ronaldo has headed the ball into the left corner of the goal

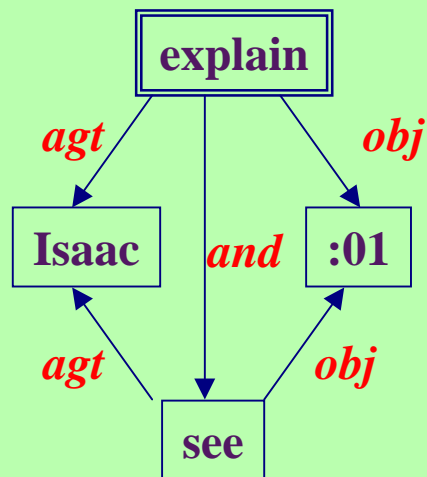
agt(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,
Ronaldo(icl>proper noun))
obj(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,
goal(icl>thing))
ins(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,
head(pof>body).@def)
plt(score(icl>event,agt>human,fld>sport).@entry.@past.@complete,
corner(icl>abstract thing).@def)
pos(corner(icl>thing).@def, goal(icl>concrete thing))
pos(head(pof>body).@def, Ronaldo(icl>proper noun))
mod(corner(icl>thing).@def, left(aoj<thing))

a UNL graph with recursion and its auxiliary UNL-tree

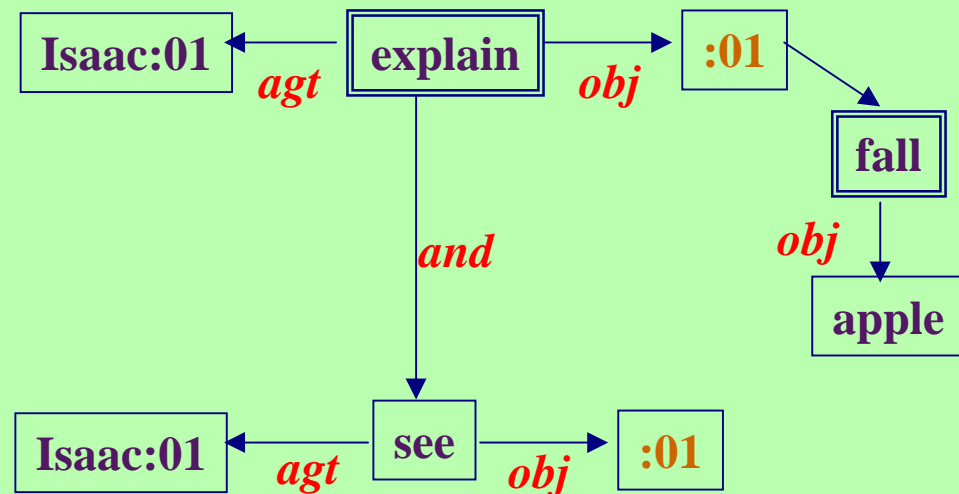
Isaac sees that an apple falls and he explains it.

```

agt(explain (ic>do) @ entry, Isaac (ic>proper noun))
obj(explain (ic>do) @ entry, :01)
obj:01 (fall (ic>occur) @ entry, apple)
and (explain (ic>do) @ entry, see (ic>do))
agt(see (ic>do), Isaac (ic>proper noun))
obj(see (ic>do), :01)
    
```



UNL (hyper) graph



UNL tree (auxiliary)

What is the UNL language ?

- ✍ Small ongoing controversy...
- ✍ A way to look at a UNL (hyper)graph :
 - it corresponds to an utterance **U-L** in language L
 - by representing the abstract structure
 - ✍ of an equivalent English utterance **U-E**
 - ✍ as « viewed from L »
- ==> the semantic attributes not necessarily expressed in L may be absent : **frequent underspecification**
 - ✍ aspect coming from French,
 - ✍ determination or number coming from Japanese,
 - ✍ etc.

The reasons of using UNL in MT

(and it can be used in many other ways!)

- ✍ Technical success of pivot MT does exist
(ATLAS, PIVOT, ULTRA, KANT)
- ✍ UNL derives from the pivot of ATLAS-II (Fujitsu)
& was designed by the same author (H. Uchida)
- ✍ Possible quality & coverage :
ATLAS-II has been the best E ↔ J system since > 10 years
It had 586,000 terms in each dictionary in Sept. 2001
- ✍ Gives certainly a lesser asymptotic quality than
transfer of/via « multi-level structure»
BUT UNL can be « coedited » from any source language

The original UNL-html format

```
<HTML><HEAD><TITLE>
```

Example 1 El/UNL

```
</TITLE></HEAD><BODY>
```

```
[D:dn=Mar Example 1, on= UNL French,  
mid=First.Author@here.com]
```

```
[P]
```

```
[S:1]
```

```
{org:el}I ran in the park yesterday. {/org}
```

```
{unl}
```

```
agt(run(icl>do).@entry.@past,i(icl>person))
```

```
plc(run(icl>do).@entry.@past,park(icl>place)
```

```
.@def)
```

```
tim(run(icl>do).@entry.@past,yesterday)
```

```
{/unl}
```

```
{cn dtime=20020130-2030, deco=man}
```

```
我昨天在公園裡跑步 {/cn}
```

```
{de dtime=20020130-2035, deco=man}
```

```
Ich lief gestern im Park. {/de}
```

```
{es dtime=20020130-2031, deco=UNL-SP}
```

```
Yo corri ayer en el parque. {/es}
```

```
{fr dtime=20020131-0805, deco=UNL-FR}
```

```
J' ai couru dans le parc hier. {/fr} [/S]
```

```
[S:2]
```

```
{org:el}My dog barked at me. {/org}
```

```
{unl}
```

```
agt(bark(icl>do).@entry.@past,dog(icl>animal))
```

```
gol(bark(icl>do).@entry.@past,i(icl>person))
```

```
pos(dog(icl>animal),i(icl>person))
```

```
{/unl} {de dtime=20020130-2036, deco=man}
```

```
Mein Hund bellte zu mir. {/de}
```

```
{fr dtime=20020131-0806, deco=UNL-FR}
```

```
Mon chien aboya pour moi. [/S] [/P][D]
```

```
</BODY></HTML>
```

The equivalent UNL-xml format

✍ As simple as UNL-html

✍ Open to all XML-related tools

```
<unl:D on="WJT" dt="04032002">
<unl:P number="1">
<unl:S number="1">
<unl:org: lang="cn">
我昨天在公園裡跑步
</unl:org>
<unl:unl sn="Ariane" pn="WJT" dt="04032002">
agt(run.@entry.@past,i)
plc(run.@entry.@past,park.@def)
tim(run.@entry.@past,yesterday)
</unl:unl>
<unl:GS lang="in">
今日は公園に走っています。 </unl:GS>
```

```
<unl:GS lang="cn">
我昨天在公園裡跑步 </unl:GS>
<unl:GS lang="de">
Ich lief in den Park gestern. </unl:GS>
<unl:GS lang="el">
I ran in the pary yesterday.</unl:GS>
<unl:GS lang="es">
Yo corri ayer en el parque.</unl:GS>
<unl:GS lang="fr">
J'ai couru dans le parc hier. </unl:GS>
</unl:S>
</unl:P>
</unl:D>
```


Output of the UNL-viewer and display in a browser

Output from the viewer (for French)	Display (in IE or Netscape or...)
<pre><HTML><HEAD><TITLE> Exam ple 1 EI/UNL </TITLE></HEAD><BODY> J' aicouru dans le parc hier. <p>M on chien aboya pourm oi. </BODY></HTML></pre>	<p>Example 1 EI/UNL</p> <p>J' ai couru dans le parc hier. Mon chien aboya pour moi.</p>

Scenario for coedition

- ✍ User reads a multilingual document in lang. L_i (IE)
- ✍ User wishes to correct some errors in L_i
- ✍ User switches to the coedition environment
- ✍ User's corrections will be executed
 - later on the text
 - immediately on the graph
- ✍ User asks for deconversion into L_i
- ✍ User iterates corrections if not satisfied, asks for deconversion into $L_1 \dots L_n$ when OK
- ✍ User returns to reading mode

Example of « FB2004 » (cited in the article)

- ✍ FB2004 = Forum Barcelona 2004
- ✍ project of « UNL usage normalisation »
- ✍ Spanish, Italian, Russian, French, Hindi
- ✍ 6-8 months in year 2001

Example

Une cité retrouvera une zone côtière après un forum

A city will retrieve a coastal zone after a forum

→ **La** cité retrouvera une zone côtière après **le** Forum

*The city will retrieve a coastal zone after **the** Forum*

Initial UNL graph (textual form)

agt(retrieve(icl>do).@entry.@future.@complete,
city(icl>entity))

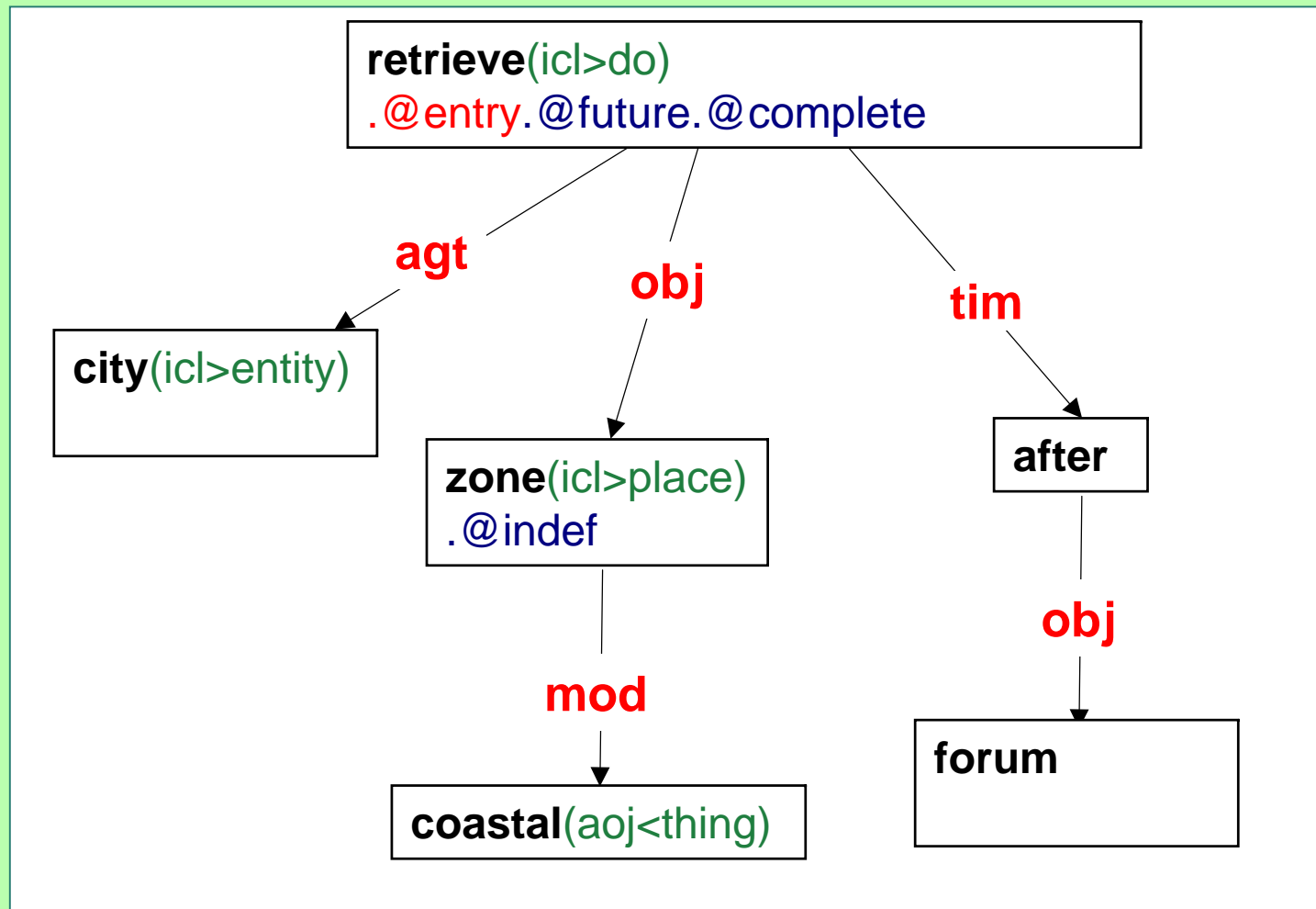
tim(retrieve(icl>do).@entry.@future.@complete, after)

obj(retrieve(icl>do).@entry.@future.@complete,
zone(icl>place).@indef)

obj(after, forum)

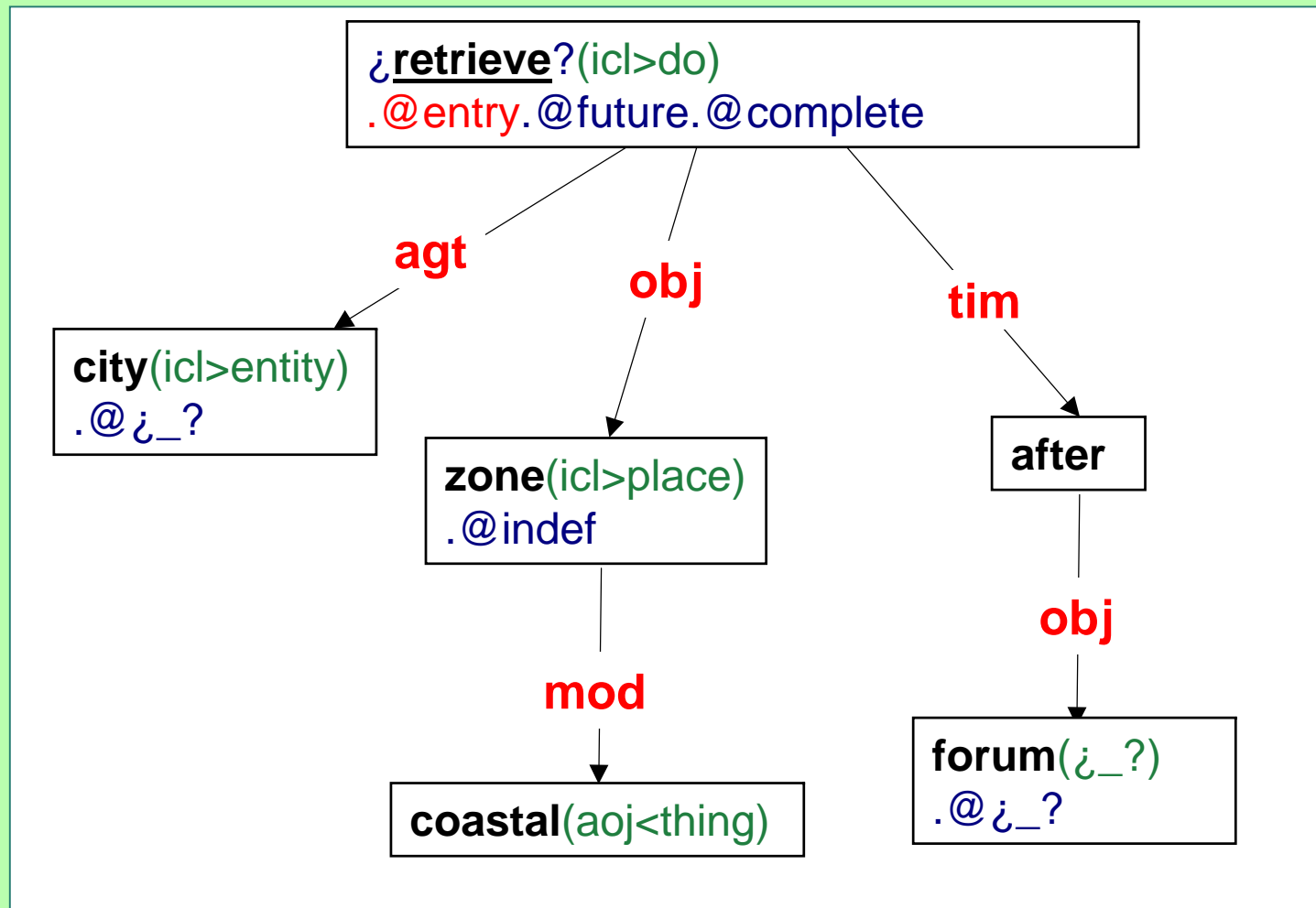
mod(zone(icl>place).@indef, coastal(aoj<thing))

Initial UNL graph (as is)



•A city will retrieve(?) a coastal zone after a forum(?)

Initial UNL graph (annotated)



•;A/The? city will retrieve(?) a coastal zone after ;a/the? forum(?)

Modifications needed

- ✍ add ".@def" on the nodes containing "city", "forum".
- ✍ add "icl>proper name" as a restriction of the UW "forum".
- ✍ Replace "retrieve" by "recover"

Transformed UNL graph

agt(recover(icl>do).@entry.@future.@complete,
city(icl>entity).@def)

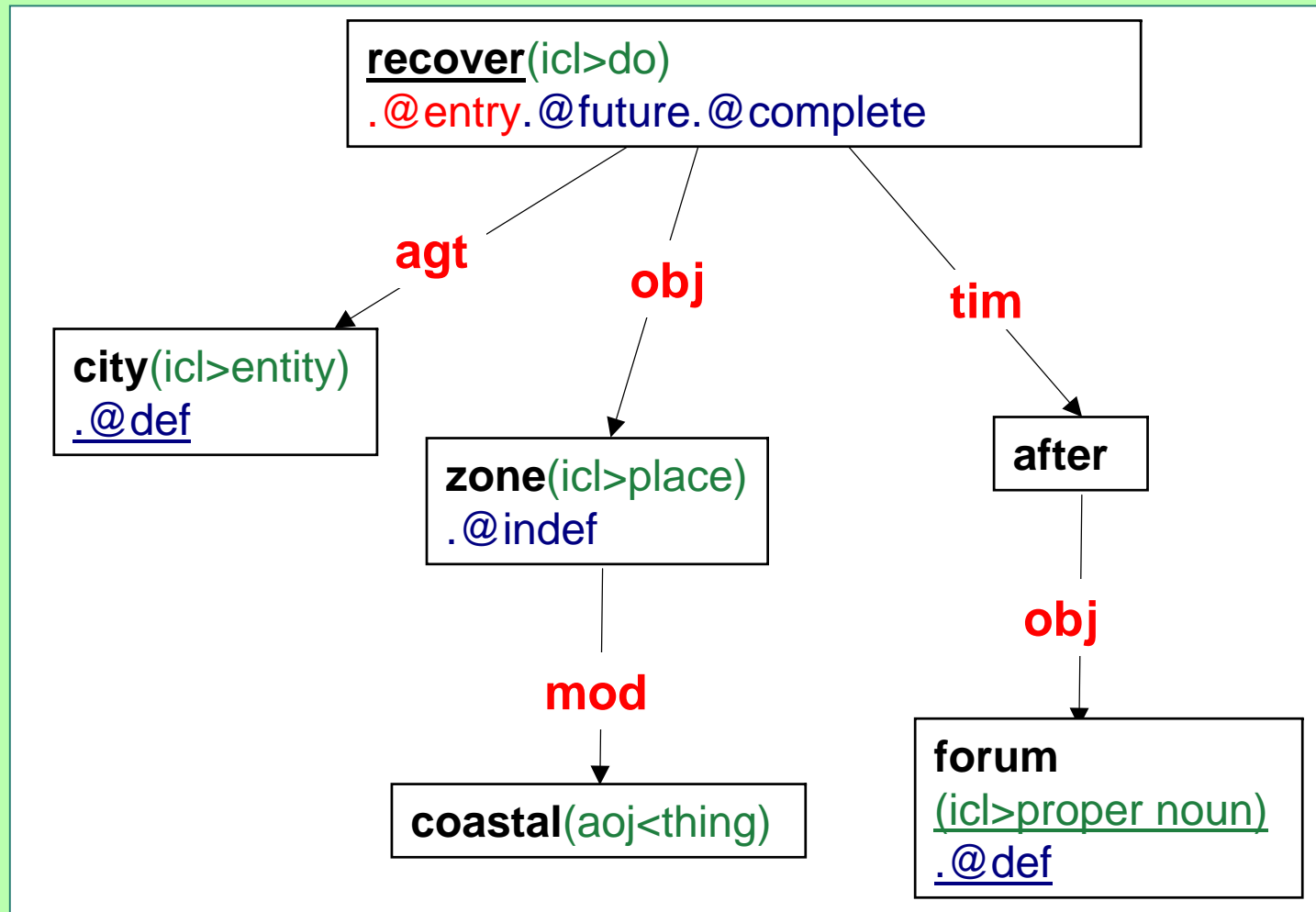
tim(recover(icl>do).@entry.@future.@complete, after)

obj(recover(icl>do).@entry.@future.@complete,
zone(icl>place).@indef)

obj(after, forum(icl>proper noun).@def)

mod(zone(icl>place).@indef, coastal(aoj<thing))

Transformed UNL graph



•The city will recover a coastal zone after the Forum

How to do it? First, the interface

- ✍ Accessible directly from a browser
- ✍ Normal mode
 - without seeing the intermediate structure(s)
- ✍ Expert mode (for hackers, fans, young people...)
 - with direct manipulation of structures

Idea of a GUI (n-2th version)

Show Graph	Deconversion	Find Lemma	Find Correspondence	Save Graph
<p>Une cité retrouvera une zone côtière après un forum.</p>				
English				
<p>After a Forum, a city will retrieve a coastal zone.</p>				
<p>After the Forum, the city will have recovered a coastal zone.</p>				
Spanish				
<p>Ciudad recobrarà una zona de costal después Foro.</p>				
<p>La ciudad habrá recobrado una zona de costal después el Foro</p>				
Italian				
<p>Città ricuperarà une zona costiera dopo Forum.</p>				
<p>La città ha ricuperarà une zona costiera dopo il Forum.</p>				
Japanese				
<p>フォーラムの後で、都市は沿岸水域を取り出</p>				
<p>フォーラムの後で、都市は沿岸水域を取り出すことを持っている。</p>				
Chinese				
<p>在博覽會之後，城市將獲得一片海岸域</p>				
<p>在博覽會之後，城市將獲得一片海岸域</p>				

Original text	Une cité retrouvera une zone côtière après un forum.
Possible Modifications	
Second Deconversion	
Manual Insertion	La cité retrouvera une zone côtière après le Forum.

Show Graph

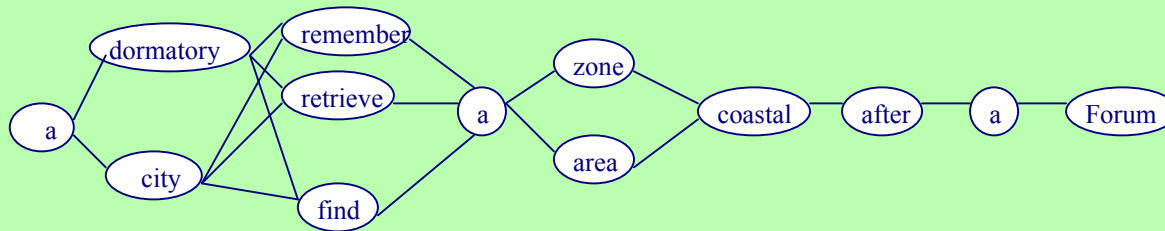
Deconversion

Find Lemma

Find Correspondence

Save Graph

Une cité retrouvera une zone côtière après un forum.



Original text

Une cité retrouvera une zone côtière après un forum.

Possible Modifications

Second Deconversion

Manual Insertion

English

After a Forum, a city will retrieve a coastal zone.

After the Forum, the city will have retrieved a coastal zone.

Spanish

Ciudad recobrarà una zona de costal después Foro.

La ciudad habrá una zona de costal después el Foro

Italian

Città ricuperarà una zona costiera dopo Forum.

La città ha ricuperarà una zona costiera dopo il Forum.

Japanese

フォーラムの後で、都市は沿岸水域を取り出

フォーラムの後で、都市は沿岸水域を取り出すことを持っている。

Chinese

在博覽會之後，城市 將獲得一片海岸域

在博覽會之後，城市 將獲得一片海岸域

Graph : lemma, lattice

Save

Quit

Show Graph

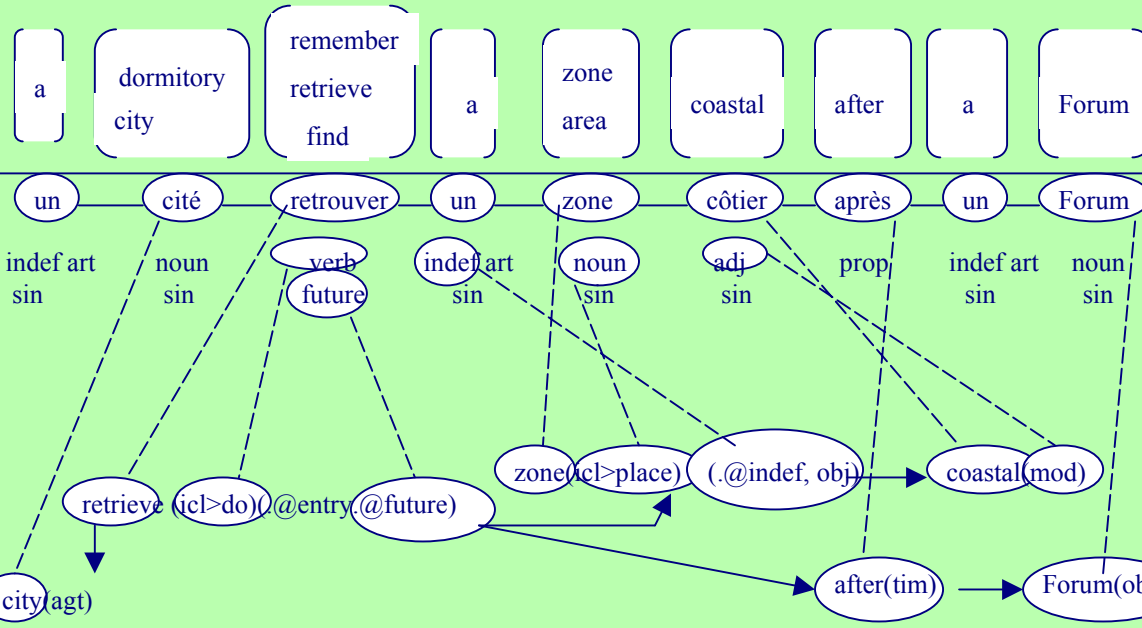
Deconversion

Find Lemma

Find Correspondence

Save Graph

Une cité retrouvera une zone côtière après un forum.



Original text

Une cité retrouvera une zone côtière après un forum.

To Do

la le Maj

Second Deconversion

La cité retrouvera une zone côtière après le Forum.

Manual Insertion

English

After a Forum, a city will retrieve a coastal zone.

After the Forum, the city will have recovered a coastal zone.

Spanish

Ciudad recobrará una zona de costal después Foro.

La ciudad habrá recobrado una zona de costal después el Foro.

Italian

Città ricupererà una zona costiera dopo Forum.

La città ha ricuperato una zona costiera dopo il Forum.

Japanese

フォーラムの後で、都市は沿岸水域を取り出

フォーラムの後で、都市は沿岸水域を取り出すことを持っている。

Chinese

在博覽會之後，城市 將獲得一片海岸域

在博覽會之後，城市 將獲得一片海岸域

Graph : correspondence

Simple text view

Multiple text view

Save

Quit

Principles of coedition

- ✍ In general, it is impossible to deduce a modification on the graph from a modification on the text
 - For example, to replace "un" ("a") by "le" ("the") doesn't imply that the following noun is determined (.@def)
 - ✍ "il aime la montagne" = "he likes mountains"
- ✍ The revision is not made directly by modifying the text, but by using a system of menus
- ✍ The menu items have a « language side » and a hidden « UNL side »

Construction of the correspondence

- ✍ Division in three sub-correspondences

UNL graph

=1=

UNL tree obtained from unfolding the graph

=2=

SMS lattice or morpho-syntactic automaton

=3=

Text

- ✍ For =2=, computation of the links by using

1 or 2 dictionaries L0↔English or UNL

Known relations between L0 attributes and UNL attributes

Maximal projectivity principle (tree rotation)

Demo of mockup on Internet Explorer...

...is not available here, sorry,
but we can see the idea with images from the demo.

✍ Situation

User is in (co)edit mode and has chosen a sentence

✍ What will be illustrated

IE screen shots (from March obsolete mockup!)

Given UNL-graph

Computed morphosyntactic lattice

Steps in the construction of the correspondence



Show Graph

Deconversion

Find Lemma

Find Correspondence

Save Graph

Une cité retrouvera une zone côtière après un forum.

English

After a Forum, a city will retrieve a coastal zone.

After the Forum, the city will have retrieved a coastal zone.

Spanish

Ciudad recobrarà una zona de costal después Foro.

La ciudad habrá recobrado una zona de costal después el Foro

Italian

Città ricupererà una zona costiera dopo Forum.

La città ha ricupererà una zona costiera dopo il Forum.

Japanese

フォーラムの後で、都市は沿岸水域を取り出

フォーラムの後で、都市は沿岸水域を取り出すことを持っている。

Chinese

在博覽會之後，城市將獲得一片海岸域

在博覽會之後，城市將獲得一片海岸域

Original text

Une cité retrouvera une zone côtière après un forum.

Possible Modifications

Second Deconversion

Manual Insertion

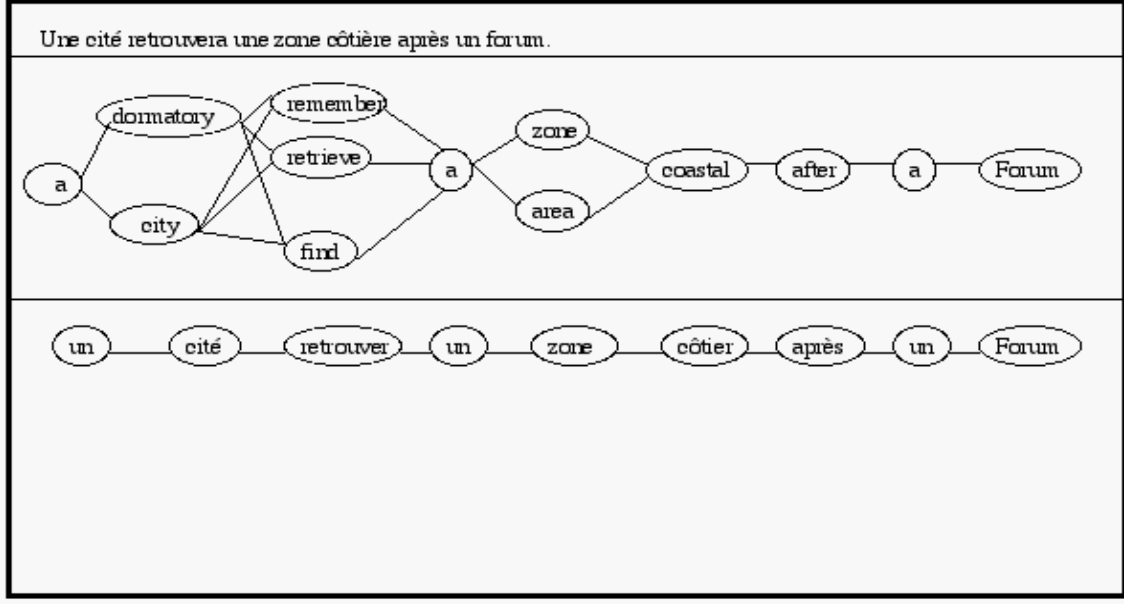
La cité retrouvera une zone côtière après le forum.

Graph : manual insrtion

Save

Quit

- Favoris
- Historique
- Recherche
- Album
- Garde-pages



Original text	Une cité retrouvera une zone côtière après un forum.
Possible Modifications	
Second Deconversion	
Manual Insertion	

English

After a Forum, a city will retrieve a coastal zone.

After the Forum, the city will have retrieved a coastal zone.

Spanish

Ciudad recobrarà una zona de costal después Foro.

La ciudad habrá una zona de costal después el Foro

Italian

Città ricuperarà una zona costiera dopo Forum.

La città ha ricuperarà una zona costiera dopo il Forum.

Japanese

フォーラムの後で、都市は沿岸水域を取り出

フォーラムの後で、都市は沿岸水域を取り出すことを持っている。

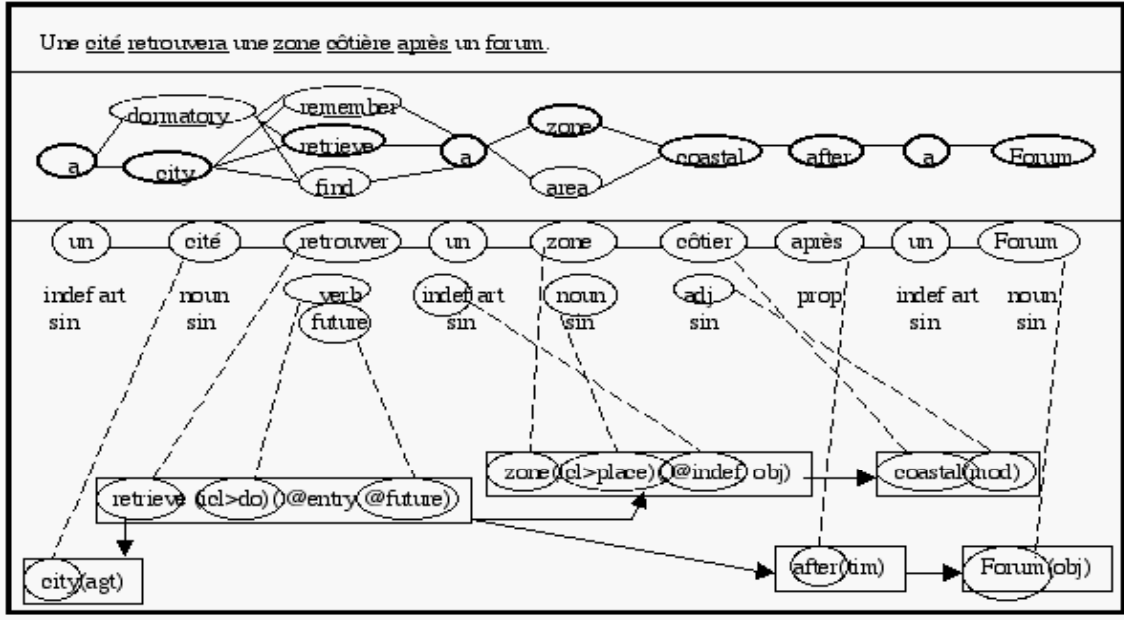
Chinese

在博覽會之後，城市將獲得一片海岸域

在博覽會之後，城市將獲得一片海岸域

Graph : lemma, lattice

- Favorites
- Historique
- Recherche
- Album
- Garde-pages



Original text	Une <u>cit</u> retrouvera une zone <u>côti</u> ère après un forum.
Possible Modifications	<u>cit</u> : (sin pl)(indef, def)(delete)(change)(negate)(add adj)
Second Deconversion	La cité retrouvera une zone côtière après le Forum.
Manual Insertion	

English

After a Forum, a city will retrieve a coastal zone.

After the Forum, the city will have retrieved a coastal zone.

Spanish

Ciudad recobriará una zona de costal después Foro.

La ciudad habrá una zona de costal después el Foro

Italian

Una città ricupererà una zona costiera dopo Forum.

La città ha ricupererà una zona costiera dopo il Forum.

Japanese

フォーラムの後で、都市は沿岸水域を取り出

フォーラムの後で、都市は沿岸水域を取り出すことを持っている。

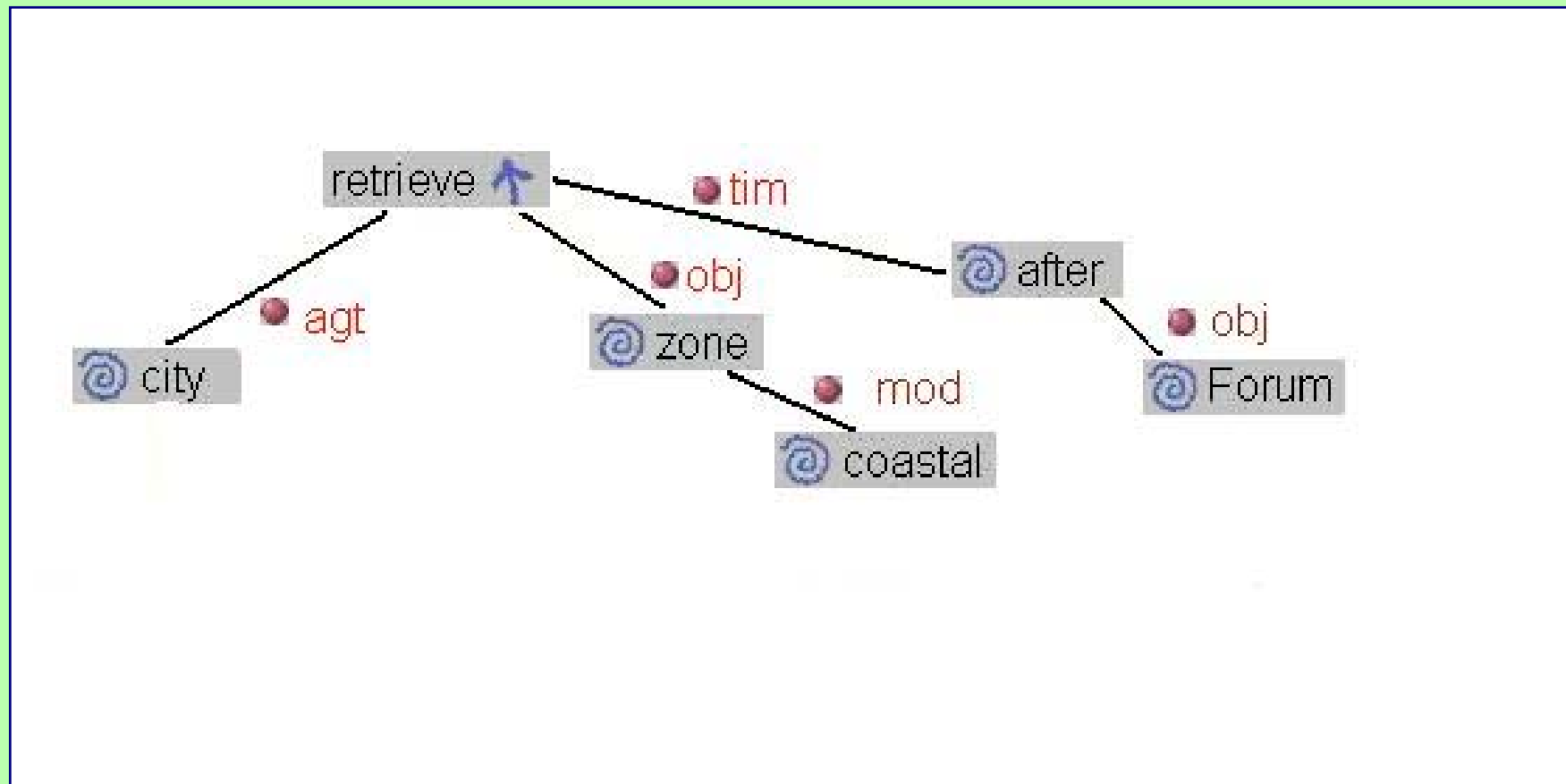
Chinese

在博覽會之後，城市將獲得一片海岸域

在博覽會之後，城市將獲得一片海岸域

Graph : correspondance

Input UNL-graph (here a tree)



UNL-French tree + French lemmas

QuickTime™ et un décompresseur TIFF (non compressé) sont requis pour visionner cette image.

UNL-French disambiguated tree

"zone-N" remains, but "diviser-V" disappears

QuickTime™ et un décompresseur TIFF (non compressé) sont requis pour visionner cette image.

On-line free French lemmatizer: PILAF

QuickTime™ et un décompresseur
BMP sont requis pour visualiser
cette image.

Morphosyntactic lattice

- 1. Output of French morphological analysis
- 2. Enrichment by FRE-ENG dictionary

QuickTime™ et un décompresseur
BMP sont requis pour visualiser
cette image.

SMS ↔ UNL-tree lexical links

3. Enrich the UNL-tree (UNL/ENG-FR) & link

QuickTime™ et un décompresseur TIFF (non compressé) sont requis pour visionner cette image.

SMS ↔ UNL-tree attributive links

4. Link UNL & French attribute values or places

QuickTime™ et un décompresseur TIFF (non compressé) sont requis pour visionner cette image.

New UNL-tree ↔ Chinese text

✍ Alignment of tree-text is not always projective

QuickTime™ et un décompresseur TIFF (non compressé) sont requis pour visionner cette image.

New UNL-tree ↔ Spanish text

✍ Here, 2 errors remain (de costal, el Foro)

==> need for feed-back to the deconversion sites

QuickTime™ et un décompresseur TIFF (non compressé) sont requis pour visionner cette image.

Conclusion

- ✍ « Shared revision » seems possible
 - No claim that all corrections could be done/shared this way!
 - Needs further prototyping (1 more Ph.D. year + students)
 - Idea of small experiment formed at ICUKL-2002
- ✍ « Coedition » is a concept unifying
 - Machine Translation
 - and
 - Multilingual Generation
- ✍ Thanks a lot to our guests & colleagues for
 - the high scientific interest of the conference
 - the excellent organization
 - and the fabulous hospitality