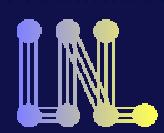


**SWIIVRE – A WEB SITE FOR THE INITIATION,
INFORMATION, VALIDATION, RESEARCH AND
EXPERIMENTATION ON UNL (UNIVERSAL
NETWORKING LANGUAGE)**

**Wang-Ju TSAI
GETA, CLIPS-IMAG
BP53, F-38041 Grenoble cedex 09 France
Wang-Ju.Tsai@imag.fr**



OBSERVAÇÕES PRELIMINARES

- ESCLARECIMENTOS
METODOLOGIA: língua # linguagem
- OS 2 pontos de contato do paper com idéias do grupo.



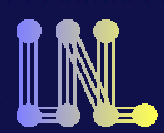
ABSTRACT

- O projeto UNL vem progredindo por 5 anos; vimos vários componentes sendo criados pelas várias equipes (“times”) de desenvolvimento, mas não ainda um ambiente integrado para usuários que queiram experimentar e usar a UNL.
- A PROPOSTA.....



O que é o SWIIVRE ?

- É a proposta de criação de Web Site para :
 - iniciação,
 - informação,
 - validação,
 - pesquisa e
 - experimentação em UNL (Universal Networking Language)
- ou seja, um ambiente integrado.



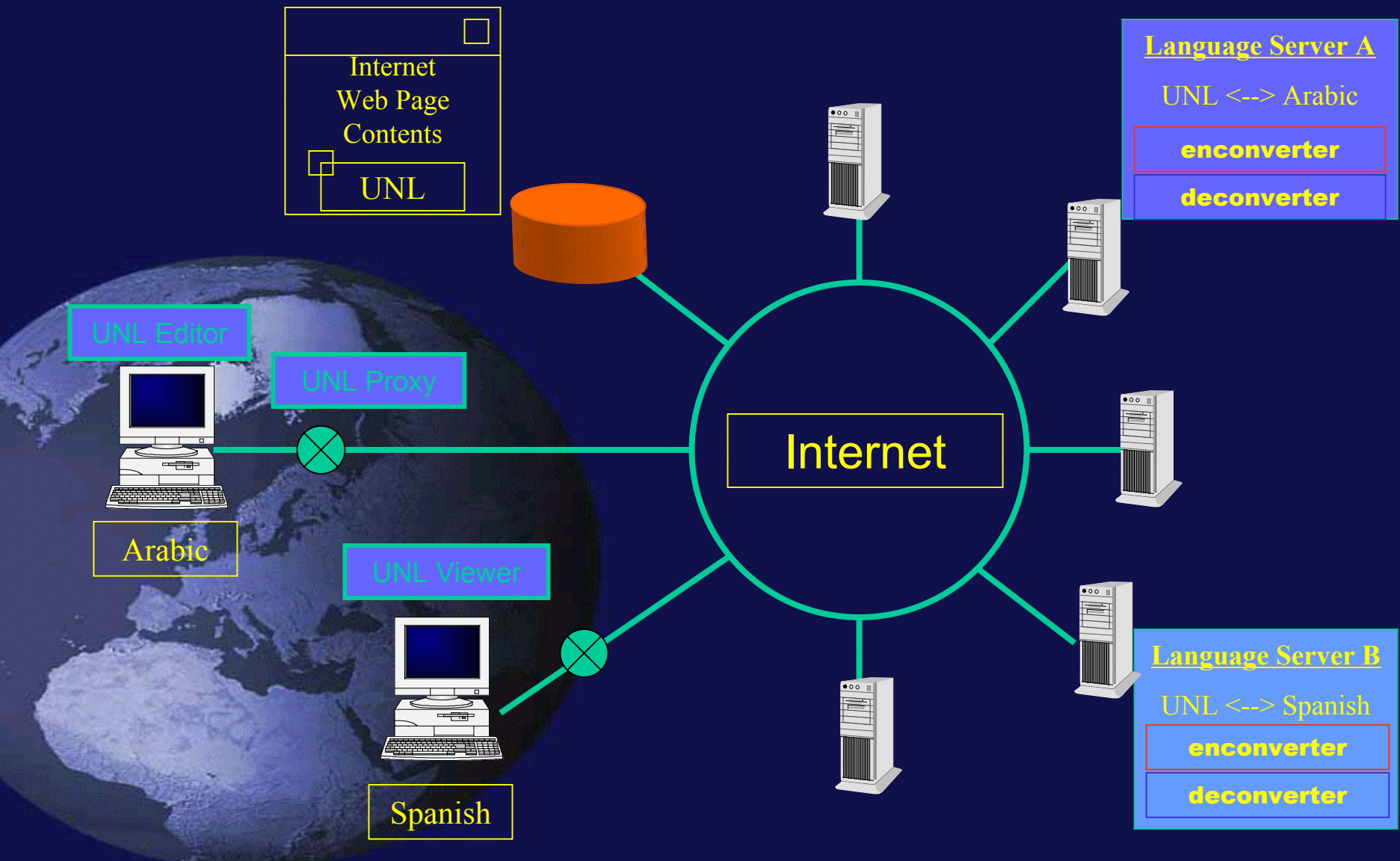
Propõe ainda:

O formulário/formato XML para os documentos UNL servindo de base para uma futura implementação da UNL na Internet.





The UNL System





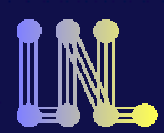
A ideia em torno da UNL

- Seguindo os fundamentos de Boitet (no sentido de defender a democratização das MT- Machine Translators- e TA's –Translation Aids- propõe o autor o uso do formato XML para os documentos como base para a implementação da UNL na Internet, como ambiente para os usuários comuns (não lingüistas, não treinados no código UNL).



UM DOCUMENTO UNL MULTILINGUAL, no formato XML

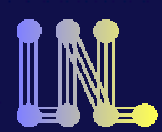
- A UNL entendida como “ a html do conteúdo linguístico), representada por um hipergráfico e está contida entre parenteses (tags), enquanto que a desconversão deste hipergráfico é colocada entre as tags da linguagem correspondente. A
- A UNL lembra em muito a estrutura da Mark up language. Como corolário seria muito natural implementar a UNL no formato XML, suportado por todos os browsers e por outras aplicações. Simplesmente aplicando filtros XSL, o usuário terá condições de capturar a parte do documento que deseja ver.



• Este documento XML-UNL tem as seguintes características:

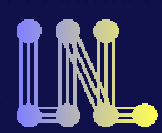
➤ **MULTILINGUAL:** a versão deconvertida de cada linguagem alvo poderá ser armazenada entre as tags especificando o nome da língua.

➤ **MULTIVERSÃO:** cada língua/alvo permite várias versões, sendo que torna-se possível existir mais de um deconverter e enconverter para cada língua. Além do que, o código UNL para cada sentença poderá ter várias versões aperfeiçoadas.



“TRACEABILITY” (localizar, rastrear): através de dois atributos, “Versionno” e “Parentno”, do elemento “version”, nós podemos rastrear a mudança de versão do código de cada sentença e versões de desconversão, já que podem vir de qualquer língua e qualquer versão.





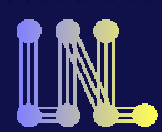
QUAIS AS OPERAÇÕES QUE PODEM SER EXECUTADAS PELOS USUÁRIOS NESTE DOCUMENTO MULTILINGUAL:

- **ENCONVERSÃO:** enconverter da NL para gráfico UNL
- **DECONVERSÃO:** deconverter do gráfico UNL para uma versão específica NL.
- **BROWSING:** navegar o documento (página) na versão NL específica.
- **POST EDITING:** se o usuário não está satisfeito com o resultado da deconversão, ele poderá “pós editar” o resultado da deconversão, ou inserir a versão NL manualmente.



O roteiro seguinte dará maiores explicações sobre como este formulário XML-UNL deverá trabalhar.

- O autor do paper (e first user) WJT criou um documento intitulado “A salinidade” que inclui um parágrafo e duas sentenças: “ A salinidade triplicou. E comi fruta”.
- **PASSOS:**



- 11) Ele então converteu estas duas sentenças em um gráfico UNL, que produziu os seguintes códigos UNL e versões originais de texto:

```
<unl>
```

```
<arc>
```

```
obj(triple(icl>do).@present.@complete.@entry,  
salinity.@def.@topic)
```

```
</arc>
```

```
</unl>
```

```
<text Langno="e101" Comment="org" >
```

```
The salinity has tripled. </text>
```

```
<unl>
```

```
<arc>
```

```
agt(eat(icl>do).@present.@complete.@entry, i)
```

```
</arc>
```

```
<arc>
```

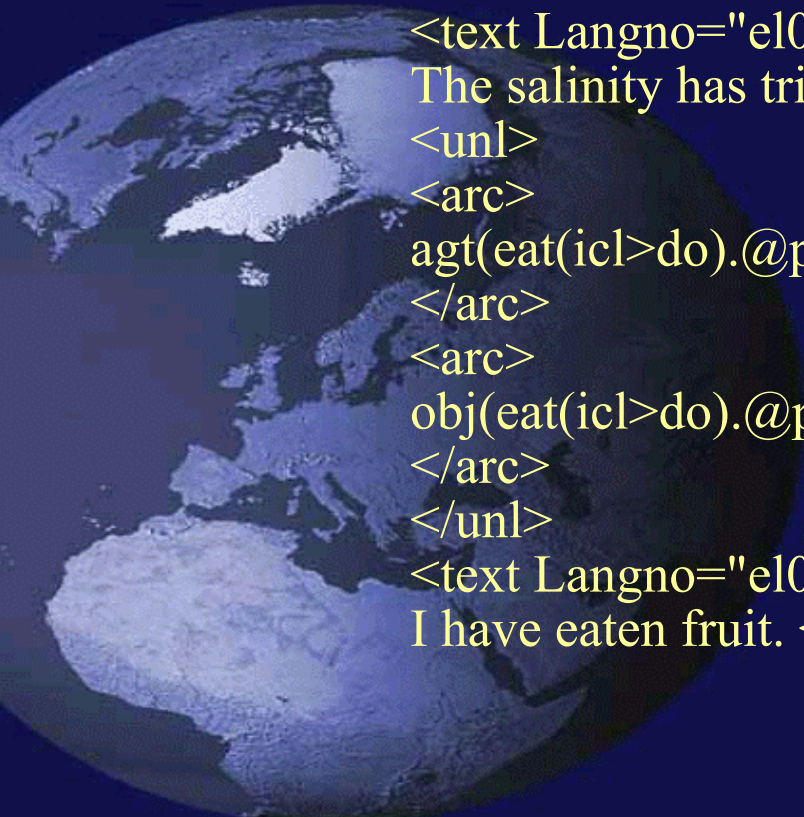
```
obj(eat(icl>do).@present.@complete.@entry, fruit)
```

```
</arc>
```

```
</unl>
```

```
<text Langno="e101" Comment="org" >
```

```
I have eaten fruit. </text>
```





2) Ele deconverteu estes gráficos UNL da primeira sentença para o espanhol e francês, e a segunda (sentença) para o francês e o russo. Ele também inseriu manualmente outra versão em francês para a segunda sentença. Então colocou este documento na internet.

- 3) Outro usuário com um browser francês lerá este documento da seguinte maneira:
- “La salinité s’est triplée. Jái mangé fruit (Jái mangé des fruits)”, que significa que a segunda sentença possui duas traduções possíveis. Isto não significa que uma delas seja melhor que a outra, simplesmente mostra ao usuário o máximo de informações possível e a evolução deste documento.



4) Um terceiro usuário, EB, constatando existir um erro no resultado da desconversão para o russo da segunda sentença, decidiu então modificar o gráfico UNL. Ele criou uma segunda versão adicionando o atributo “@pl” para a UW “fruit”, e deconverteu as duas sentenças para o russo e francês. Ele também passou a primeira sentença para dois deconversores em francês, então os dois resultados de desconversão foram gravados.

- Os resultados estão no apêndice II. [Slide 20](#)



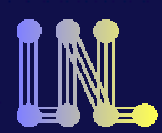
INTRODUÇÃO AO WEB SITE SWIIVRE

- Da discussão anterior podemos concluir que para tornar a UNL “usável”, é importante ter um ambiente integrado para os usuários comuns. Mas a situação corrente é que todos os componentes dos diversos LC estão espalhados por todo o lugar. Para um membro iniciante na “sociedade UNL” terá que dispender vários esforços para se familiarizar com a especificação UNL. Terá que ir ao UNL center, utilizar-se de alguns tutoriais; entretanto, se finalmente precisar uma desconversão para o Tailandês ou Indonésia, não será capaz de encontrar.
- (vide apêndice III onde listamos as fontes acessíveis para os membros da sociedad. UNL na internet.)



O usuário comum é aquele que não é linguista, nunca foi treinado no código UNL mas tem necessidade de ler um documento multilingual através da internet. Ele acessa a internet mas está equipado tão somente com um browser básico. O ambiente proposto o capacitará a acessar outros UNL-NL dicionários ou UNL-NL enco/deco “engines” na internet. Este ambiente também deverá prover toda a informação e o status atual do sistema UNL.





Initiation, Information, Vérification, Recherche, et expérimentation sur UNL).

- Para experimentar esta idéia e facilitar a pesquisa na UNL, o autor implementou um site denominado SWIIRE. A construção deste web site é um projeto que está em andamento.
- A seguir apresentaremos uma imagem da página inicial deste site:
- <http://clips.imag.fr/geta/User/wang-ju.tsai/welcome.html>
-

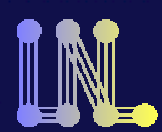


Projeto SWIIVRE-UNL Project

un Site Web pour l'initiation, l'Information, la Validation, la Recherche et l'Expérimentation sur UNL

a Web Site for the Initiation, Information, Validation, Research and Experimentation on UNL (Universal Networking Language)

| | |
|-------------------------------------|--|
| About This Site | Initiation on UNL |
| UNL Resources | Create UNL Graph |
| Post Edit UNL graph | UNL Corpus |
| Comments | Links & References |



CONCLUSÕES:

- **O Autor afirmar ter dado o primeiro passo rumo a integração de todos os componentes da UNL em um website. O próximo passo será simplificar/modernizar o procedimento da edição dos gráficos, especificações e consulta a dicionários, bem como a pós-edição do gráfico.**



APÊNDICE I

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<!-- ?xml-styleheet type="text/xsl" href="unlxsl01.xsl"? -->
<!-- XML pour UNL 4/10/2001 -->
<!DOCTYPE UNLDOC [
<!ELEMENT D (P+) >
<!ELEMENT P (S+)>
<!ELEMENT S (version+)>
<!ELEMENT version (unl, text*)>
<!ELEMENT unl (arc*)>
<!ELEMENT text (#PCDATA)>
<!ELEMENT arc (#PCDATA)>
<!ATTLIST D DocName CDATA #REQUIRED>
<!ATTLIST P Paragraphno CDATA #REQUIRED>
<!ATTLIST S Setenceno CDATA #REQUIRED>
<!ATTLIST version Versionno CDATA #REQUIRED
Parentno CDATA #REQUIRED
Date CDATA #REQUIRED
Author CDATA #REQUIRED>
<!ATTLIST text Langno CDATA #REQUIRED
Comment CDATA #REQUIRED>
```

Appendix (III)

The resources accessible at each LC for UNL society members

| | Enco | Deco | Dico | Introduction of UNL system | Linked by UNLC | remarks |
|------------|------|------|------|----------------------------|----------------|----------------|
| Arabic | √ | √ | √ | Arabic | √ | |
| Chinese | | √ | | English | √ | |
| French | | √ | | | | |
| Indonesian | | | | Indonesian | | |
| Italian | | √ | | Italian | √ | |
| Russian | | √ | | English | | |
| Spanish | | √ | | English | √ | Tutorials also |
| Thai | | | | Thai | √ | |
| UNLC | | | √ | English | | |