

O enConverter

UFSC – PPGEP - UNL

Mestrado/Doutorado em Engenharia da Produção

Participantes:

José Roberto dos Santos

César Stradiotto

Everson

Tales

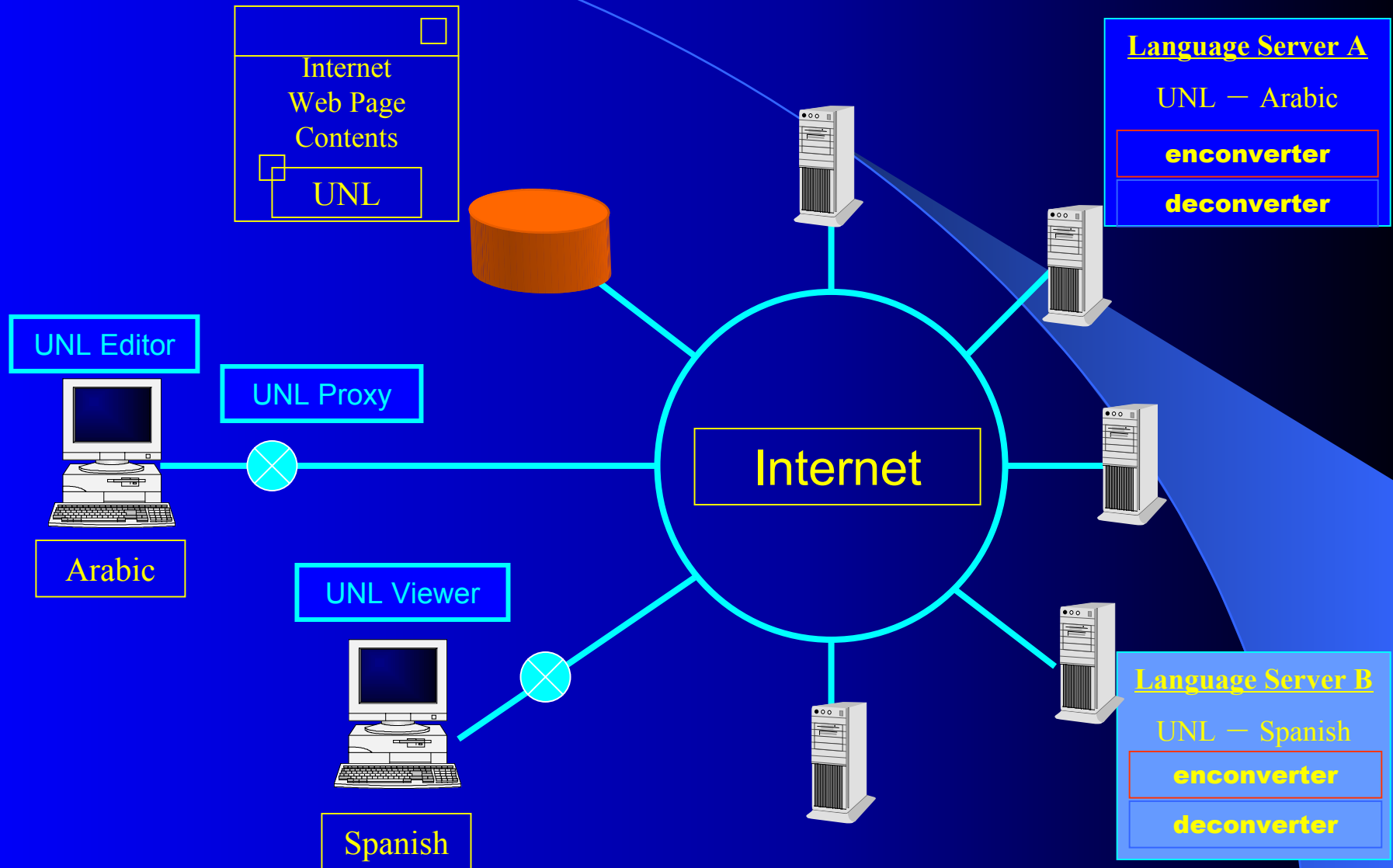
Enconverter

Definição:

Analizador de sentenças com a capacidade de gerar comandos de UNL tendo como base:

- Dicionário de Words,
- KnowleadeBase,
- Enconversion Rule

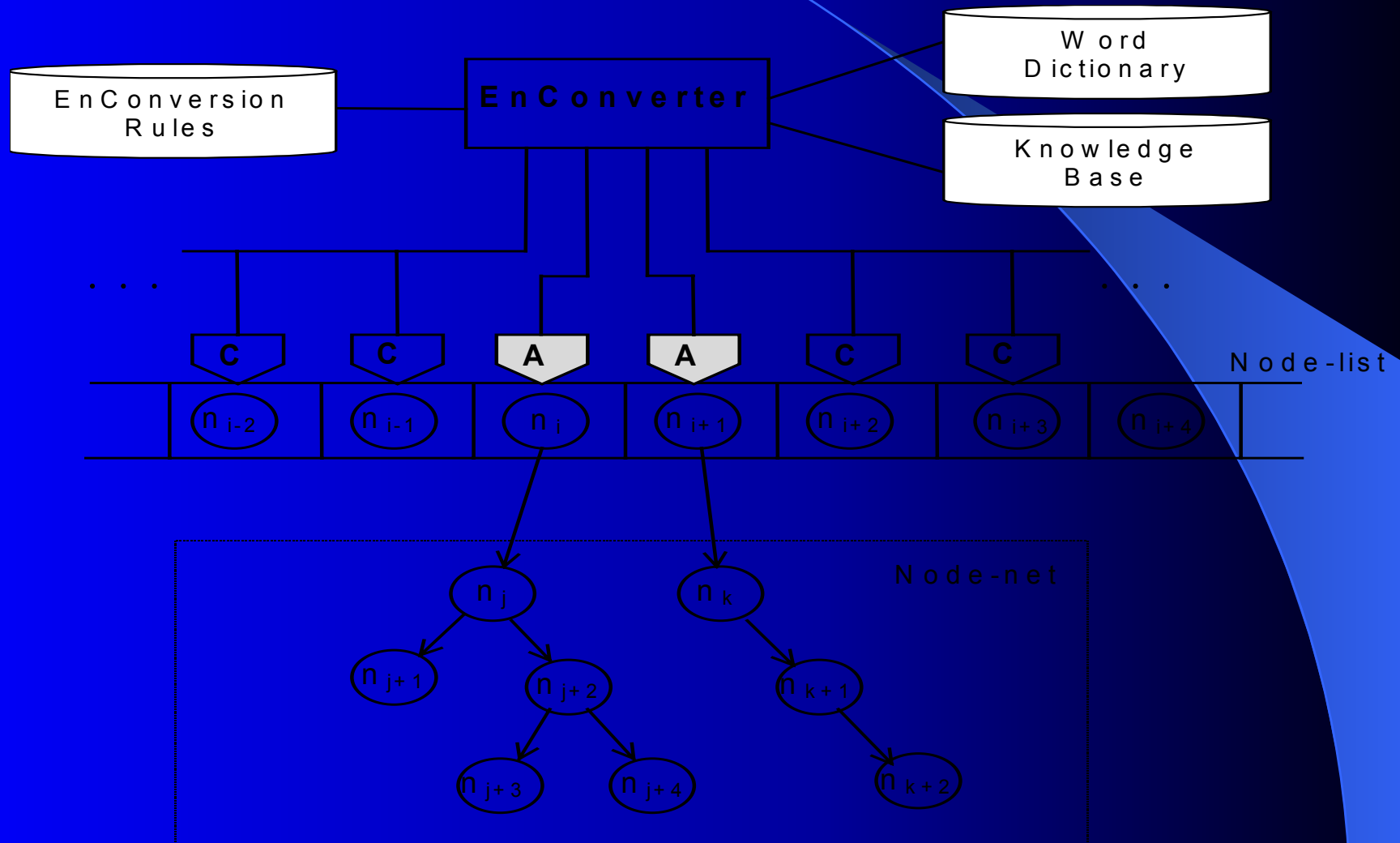
The UNL System



Necessidades Básicas

- Sentenças em acordo com a sintaxe morfológica;
- Tempos verbais coerentes;
- Correções gramaticais prévias;
- Regras de conversão da língua natural para uma correta geração de comandos UNL;
- Dicionário de Words atualizado;
- Knowledge Base compatível com a língua natural

Estrutura do Enconverter



Arquivos necessários para a Conversão

-Arquivo de Entrada:

“Elas também tratam de qualquer problema de radiocomunicação de caráter mundial.”

Arquivos necessários para a Conversão

-Dicionário de Words:

[elas] {} "they" (PPR,3PP,FEM,NOM);

[também] {} "also" (ADV,MAN);

[trat] {} "address(icl>treat)" (VER,P05,VI2,STM,ACT);

[am] {} "" (SFX,TEN,ET1,MOO,IND,NBR,PER,3PP,P05);

[de] {} "" (PRE);

[qualquer] {} "any(mod<thing)" (NDP,COM,SGN);

[problema] {} "matter(icl>problem)" (NOU,MCL,^NBR);

[radiocomunicação] {} "radiocommunication(icl>communication)"
(NOU,FEM,STM,SGN(ão),PLN(ões));

Arquivos necessários para a Conversão

-Dicionário de Words (continuação):

[caráter] {} "character(icl>nature)" (NOU,MCL,SGN);

[mundia] {} "worldwide(mod<thing)"

(ADJ,COM,STM,^NBR,SGN(I),PLN(is));

[I] {} (SFX,NBR,SGL);

[is] {} (SFX,NBR,PLN);

[ão] {} (SFX,NBR,SGL);

[ões] {} (SFX,NBR,PLN);

Arquivos necessários para a Conversão

-Enconverter Rules:

R{}P001;

+{:}{BLK:}P255;

+{STM:@:}{SFX:}P255;

:{VER,ET1,^&@present:&@present:}{:}P200;

+{VER,VI2,!de:}{[de]:}P200;

-{PRE:}{NOU:@:}P200;

<{:}{[.]:}P200;

Arquivos necessários para a Conversão

-Enconverter Rules (continuação):

:(SHEAD){^&@entry:&@entry::}{STAIL:::}P200;

>{ADV,MAN::man:}{VER:::}P100;

>{PPR::agt:}{VER,ACT:::}P100;

>{NDP::mod:}{NOU:::}P100;

<{NOU:::}{ADJ,SGN::mod:}P100;

<{VER,VI2,ACT:::}{NOU::obj:}(STAIL)P100;

<{NOU:::}{NOU,PRE,SGN::mod:}(^BLK,^ADJ)P50;

Arquivos necessários para a Conversão

-Arquivo de Saída:

```
[>>]{} "" (STAIL) <.,0,0>;
```

```
[S]
```

;Elas também tratam de qualquer problema de radiocomunicação de caráter

```
{unl}
```

```
obj(address(icl>treat):0C.@entry.@present, matter(icl>problem):0V)
```

```
agt(address(icl>treat):0C.@entry.@present, they:00)
```

```
man(address(icl>treat):0C.@entry.@present, also:05)
```

```
mod(matter(icl>problem):0V, character(icl>nature):1R)
```

```
mod(matter(icl>problem):0V, radiocommunication(icl>communication):17)
```

```
mod(matter(icl>problem):0V, any(mod<thing>):0M)
```

```
mod(character(icl>nature):1R, worldwide(mod<thing>):1Z)
```

```
{/unl}
```

```
[/S]
```

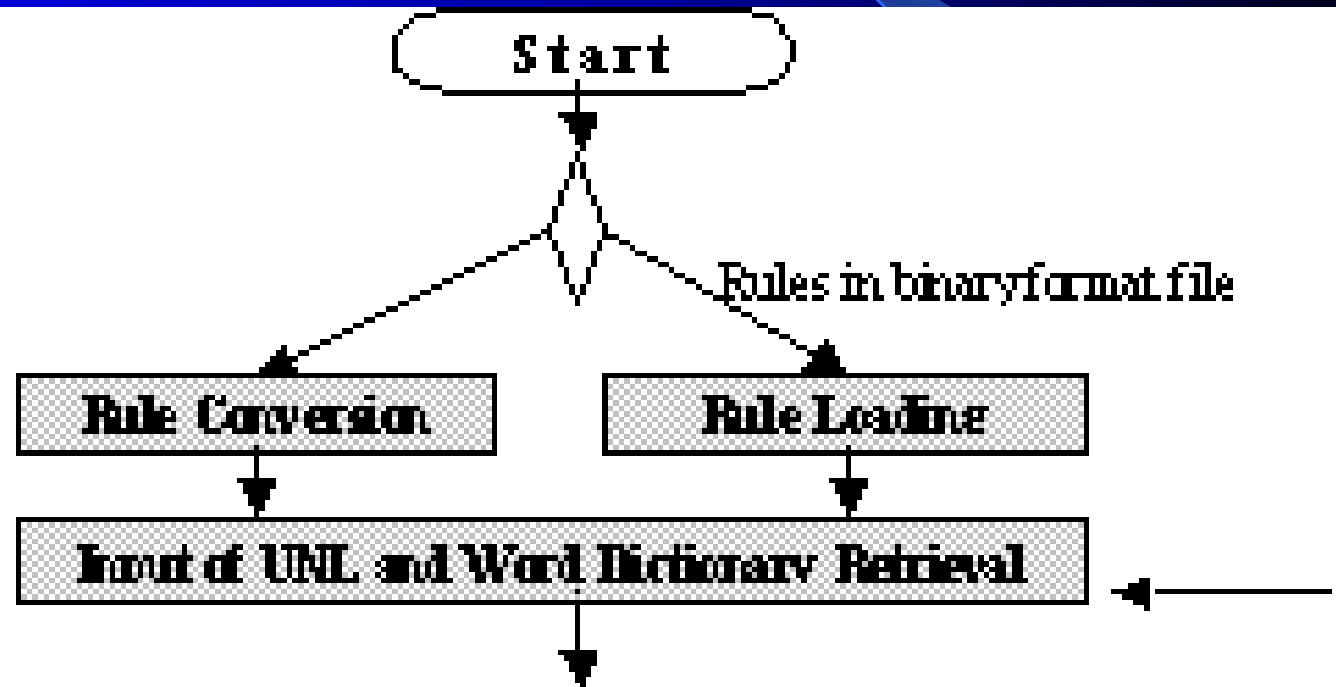
Arquivos necessários para a Conversão

-Arquivo de Saída (continuação):

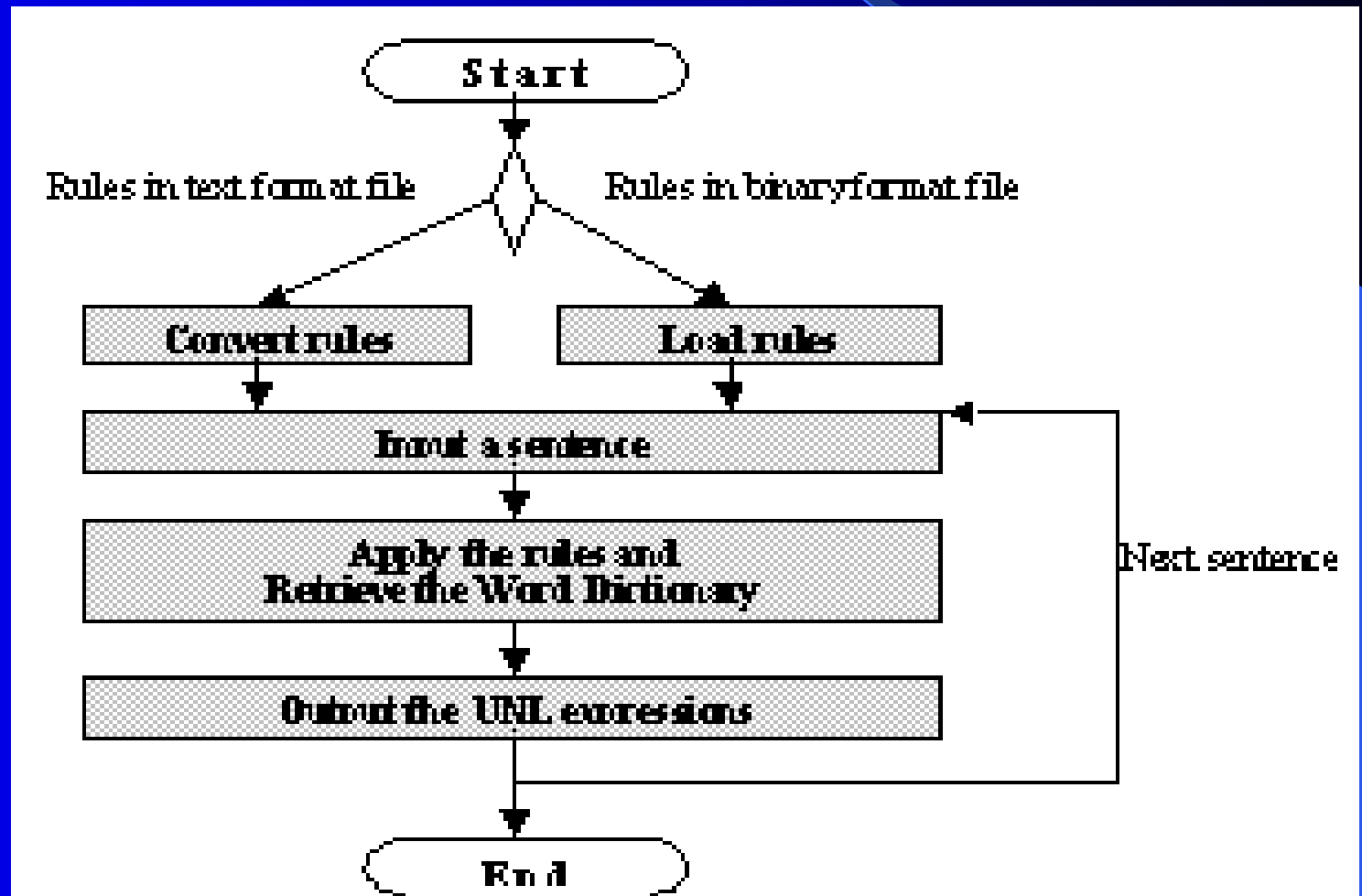
```
===== NODE LIST =====  
/[<<]/"Elas também tratam de qualquer problema de  
radiocomunicação de caráter mundial."/>>/  
=====  
Inode >>>>>  
-- word candidates for Inode --  
[<<]{} "" (SHEAD) <.,0,0>;  
rnode >>>>>  
-- word candidates for rnode --  
[elas]{} "they" (PPR,NOM,3PP,FEM) <p,0,0>;  
===== APPLIED RULE =====  
R{:::}{:::}P1;
```

Atividade Previas do Enconverter

Carga de Regras e Dicionário de Words



Atividade de Enconverter



Convenção do Enconvert

- Inicia-se pela primeira palavra do texto;
- Procura uma word correspondente;
- Cria a node-list;
- Aplica a regra conforme a sua prioridade

===== NODE LIST =====

```
/[<<]/"Elas também tratam de qualquer problema de radiocomunicação de caráter mundial."/>>/
```

```
=====
```

```
Inode >>>>
```

```
-- word candidates for Inode --
```

```
[<<]{} "" (SHEAD) <.,0,0>;
```

```
rnode >>>>
```

```
-- word candidates for rnode --
```

```
[elas]{} "they" (PPR,NOM,3PP,FEM) <p,0,0>;
```

```
===== APPLIED RULE =====
```

```
R{:}{:}P1;
```

```
-- other applicable rules --
```

```
=====
```

```
>>>> Inode
```

```
[<<]{} "" (SHEAD) <.,0,0>;
```

```
>>>> rnode
```

```
[elas]{} "they" (PPR,NOM,3PP,FEM) <p,0,0>;
```

===== NODE LIST =====

/<</[elas|uw=they]/" também tratam de qualquer problema de radiocomunicação de caráter mundial."/>>/

[<<]} "" (SHEAD) <.,0,0>;

[elas]} "they" (PPR,NOM,3PP,FEM) <p,0,0>;

=====

Inode >>>>>

[elas]} "they" (PPR,NOM,3PP,FEM) <p,0,0>;

rnode >>>>>

-- word candidates for rnode --

[]} "" (BLK) <.,0,0>;

===== APPLIED RULE =====

+{:::}{BLK:::}P255;

-- other applicable rules --

R{:::}{:::}P1;

=====

>>>>> Inode

[<<]} "" (SHEAD) <.,0,0>;

>>>>> rnode

[elas]} "they" (PPR,NOM,3PP,FEM) <p,0,0>;

Pesquisa no Dicionário de Words

Dicionário de Entries

- Pesquisa todos os morphemes, iniciando pelo string escolhido;
- Baseia-se nas prioridades, freqüência e tamanho definidas na UW
- Da prioridade ao morpheme de maior freqüência;
- Caso exista mais de um morphemes candidato assume o de maior prioridade de entrada.

Formato do Dicionário de Words:

[HW]{} "UW" (ATTR1, ATTR2, ...) <FLG, FRE, PRI>;

[elas]{} "they" (PPR,NOM,3PP,FEM) <p,0,0>;

Pesquisa no Dicionário de Words

Numerais Arábicos

-No caso de valores arabicos, o Enconverter cria uma entrada automaticamente e assinala os atributos “NUM,ANUM”

Formato do Dicionário de Words:

```
[HW]{} “HW” (NUM, ANUM) <., 0, 0>;
```

Exemplo:

```
[2000]{} “2000” (NUM, ANUM) <., 0, 0>;
```

Entradas Temporárias

- Existe apenas duas regras para aplicar valores temporários,
 - 1) Valores numéricos arábicos;
 - 2) Definidas como regras, exemplo ABC, John Smith, os valores devem ser mantidos no formato original.

```
[HW]{} "HW" (TEMP) <., 0, 0>;
```

Exemplo

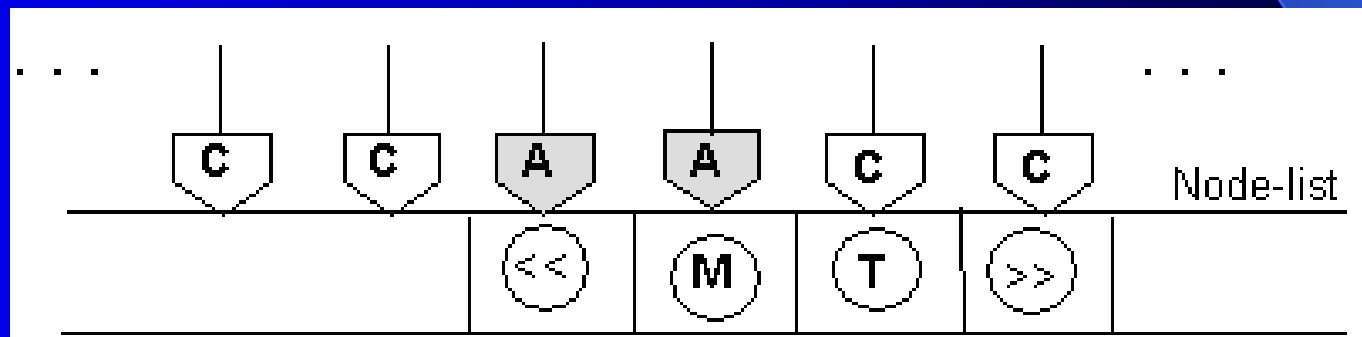
```
[ABC]{} "ABC" (TEMP) <., 0, 0>;
```

Espaços em branco

-Cria um atributo automaticamente com a abreviação BLK

```
[ ]} "" (BLK) <., 0, 0>;
```

Quando qualquer entrada de dicionário pesquisada ou criado um string, as entradas de dicionário leva a posição do string.



Aplicação da Regra baseado em Word Decision

- Depois de recuperação do dicionário de palavras,
- uma tentativa é feita para aplicar as regras de enconversion para a lista das entradas pesquisadas da morfema.
- Se tem sucesso, uma regra é encontrada e selecionada, aplica-se a regra.

Aplicação da Regra baseado Análise Sintática ou Semântica

-Análise sintática é levada em conta aplicando regras de tipos que são qualquer dos " +", " -", " < ou ">". Estes tipos de regras das condições dos nodes nas janelas para gerar uma árvore sub-sintática que usa os dois nodos na Análise o Windows

Aplicação da Regra baseado Análise Sintática ou Semântica (cont)

- **Análise semântica** é ativada pela indicação de relações no campo <RELATION> de uma regra que tenha uma condição "<" ou ">".
- Satisfazendo tal regra é aplicada, o EnConverter gera uma relação binária na cadeia de UNL que usa a relação indicada no campo <RELATION> da regra e duas UWs de ambos os nodes.
- Se o resultado do Knowledge Base falhar, nenhuma relação binária é gerada e esta regra não é aplicada.

Aplicação da Regra baseado Análise Sintática ou Semântica (cont)

EnConverter também provê uma função para a criação ou Modificação de um UW que usa o dois UWs dos nodes na Análise o Windows.

- É levado em conta aplicando um "+" ou "-" tipo de regra em combinação com certos operadores.
- Esta função é útil quando criamos UW de números arábicos em uma expressão de idioma nativo, por exemplo.
- Copiando ou inserindo um node faz com facilidade um looping ou multi-referência para uma UW na cadeia de UNL, e também é levado em conta aplicação das regras.

Backtracking

Backtracking acontece quando uma regra aplicável não pode ser achada ou a regra falhou a sua aplicação. Isto também ocorre quando um Backtrack Rule do tipo '?' é aplicado. Quando isto ocorre, o processo retorna a fase prévia quando a última regra de análise aplicada, ou um word/morpheme foi selecionado, e ela tenta outra possibilidade.

Backtracking

Quando backtrack acontece, o sistema se comportará diferentemente nas situações seguintes.

1. quando o sistema backtrack para uma fase o word/morpheme foi selecionado:
2. quando o sistema regressa à fase quando uma regra simplesmente não aplicada a seleção:
3. Outro tipo de backtrack, é p Node Assignment Backtracking. É ativado aplicando uma regra do tipo "?L " ou "?R ". Regressa diretamente à fase quando o node da esquerda ou direito foram criados e rejeitado.

Fim de Enconversion / Regra Aplicação

Se o sistema retorna à fase do estado inicial e nenhuma outra regra aplicável é trabalhada, e o processo termina em erro.

Normalmente, o processo de aplicação de regra termina em Sentence Tail Node, o node tem o atributo " STAIL " move Left Analysis Window, ou o node Sentence Head tem o atributo " SHEAD " move Right Analysis Window.

O EnConverter então começa a produção das expressões de UNL.

Enconversion Rules

Teoria

As regras de Enconversion descrevem as condições para aplicação da regra:

O modo para reescrever os atributos dos nodes para satisfazer estas condições, e o caminho para construir uma árvore sintática.

Enquanto aplica as regras de enconversion, o EnConverter facilita a análise morfológica, sintática e semântica usando um método de interleaf. Isto significa que toda a informação pode ser usada eficazmente nos processos de enconversion.

Finalmente, gera as árvores sintáticas e cadeias de UNL para o introduzir as orações.

Teoria

EnConverter olha, e opera os nodes na Node-list por suas janelas, e são descritas as condições e ações para cada janela nas regras de enconversion.

Cada parte de uma regra expressa as condições de ações, os nodes adjacentes na Node-List. Como descreve:

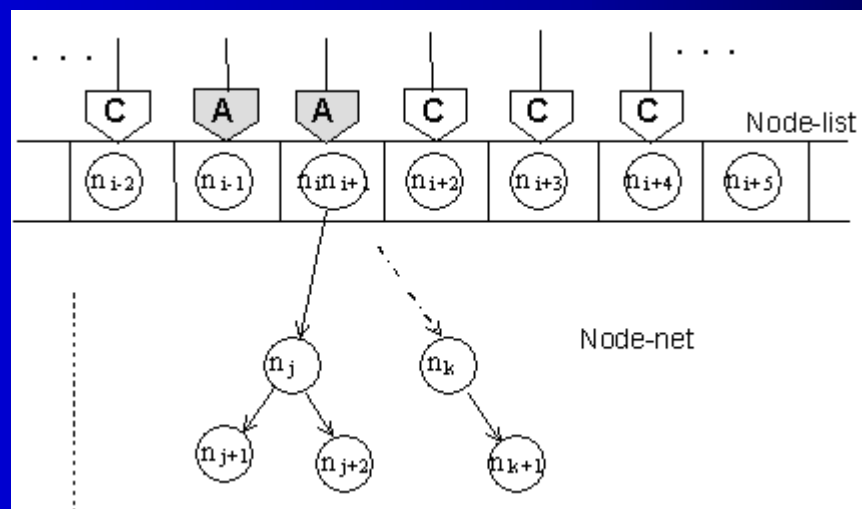
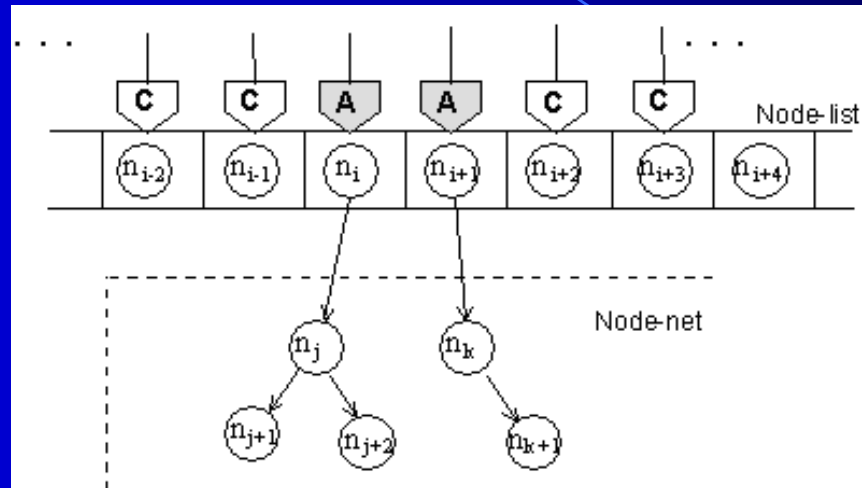
- the Left Condition Windows (LCW or PRE),
- the Left Analysis Window (LAW),
- the Middle Condition Windows (MCW or MID),
- the Right Analysis Window (RAW),
- and the Right Condition Windows (RCW or SUF).

Tipos de Regras de Enconversion

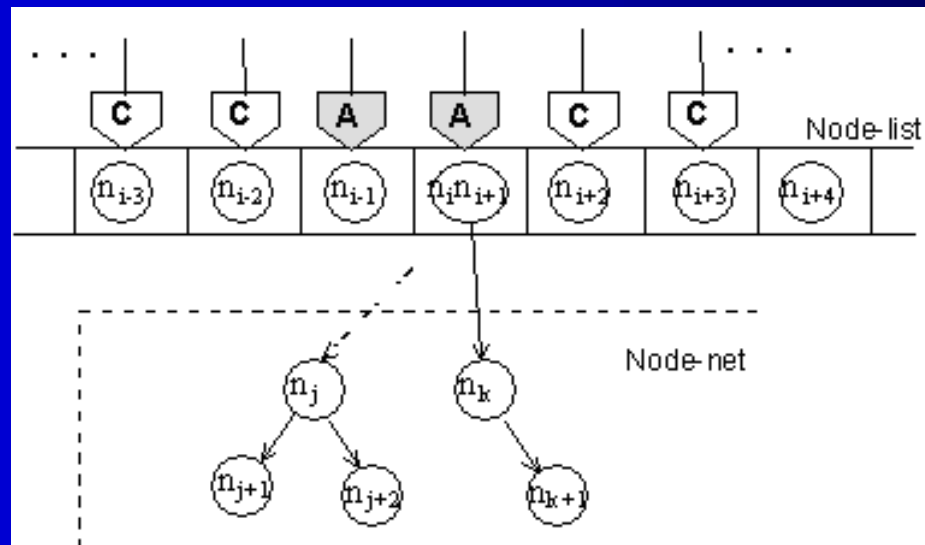
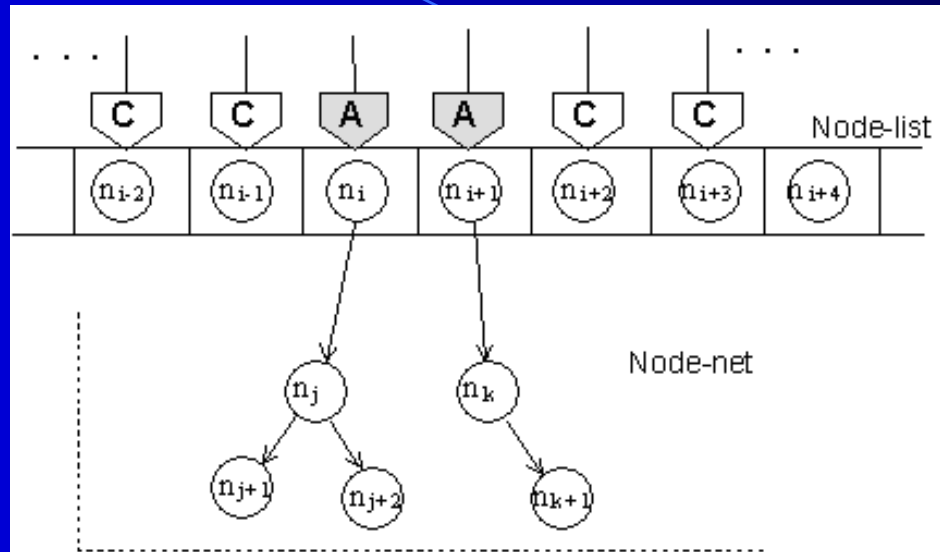
Esta seção descreve os tipos e funções de regras de enconversion. Há 15 tipos de regras de enconversion, cada tipo que é representado através de um a dois símbolos como '+', '-', '<', '>', 'L', 'R', ':', '?', '?L', '?R', 'C', 'G', 'X', 'DL' e 'DR' no qual será usado o <TIPO> campo de regras. O '+' e '-' tipos de regra também podem ter extensional funciona prendendo certos operadores.

Enconversion de uma oração nativa para UNL é levado a cabo aplicando estes tipos de regras de enconversion.

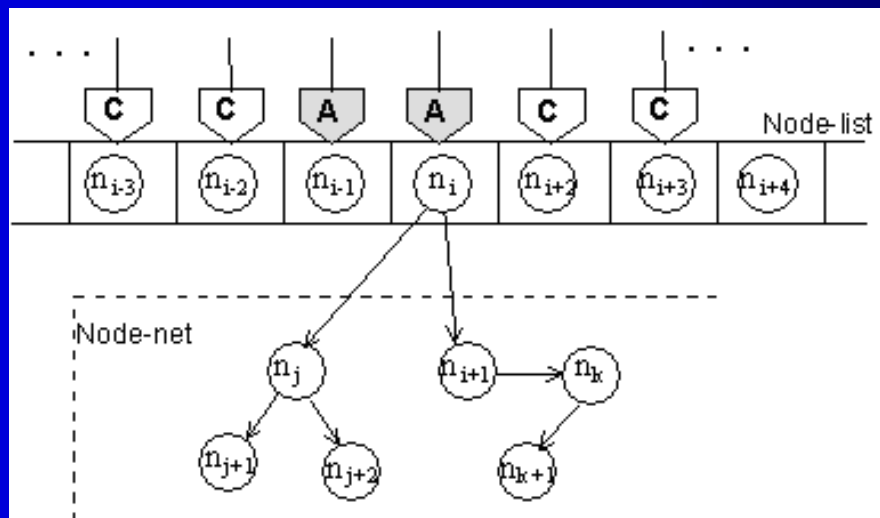
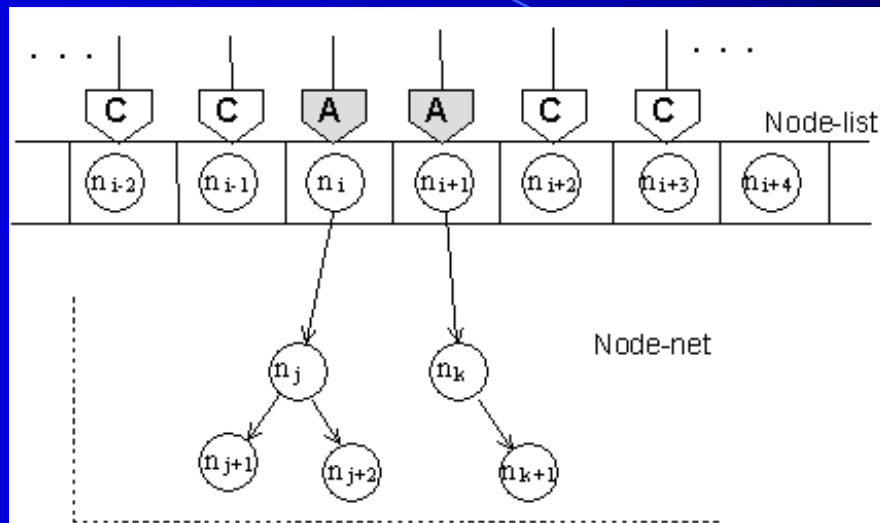
Left Composition " + | +: + | +: c | +: * "



Right Composition " - | -:+ | -:c | -:* "

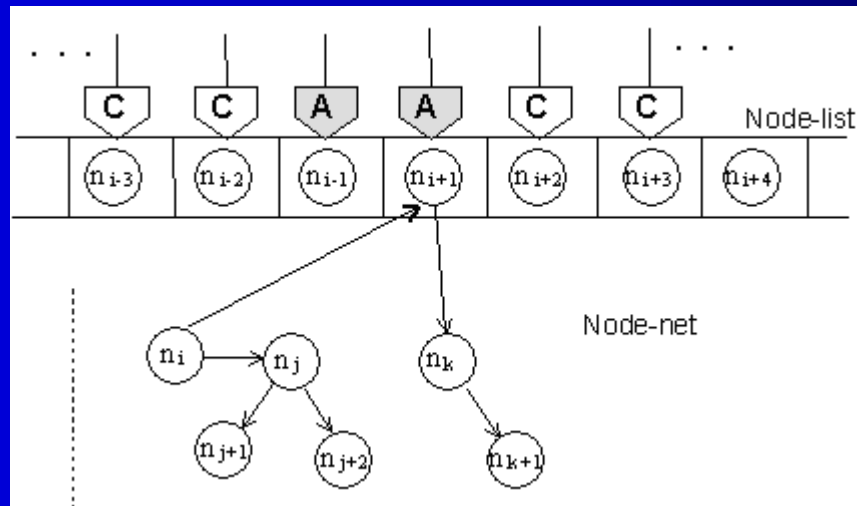
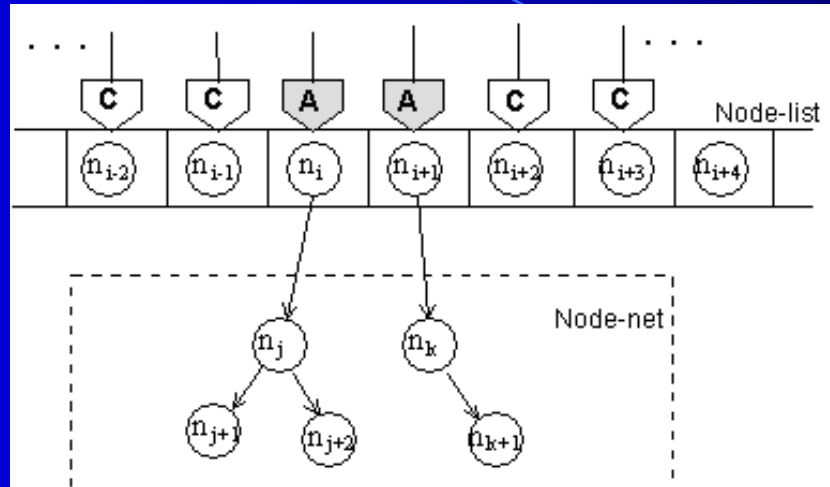


Left Modification "<"



Right Modification

">"



- Left Shift "L"
- Right Shift "R"
- Attribute Changing ":"
- Backtrack "?"
- Left Node Assignment Backtrack "?L"
- Right Node Assignment Backtrack "?R"

-Node Copy

"C"

- Syntactic Tree Copy

"G"

- Node Exchange

"X"

- Left Node Deletion

"DL"

-Right Node Deletion

"DR"

Look-ahead Process in Morphological Analysis

Teoria

Análise morfológica é levada em conta consultando o Word Dictionary enquanto aplicando uma regra quando string de carácter aparece em uma janela (the right Analysis Window ou um de the right side Condition Windows) de EnConverter. Em primeiro lugar, uma lista de candidatos de morfema que emparelham o introduza desde o princípio do string é criado. Se uma morfema que satisfaz as condições requeresse para o node na janela é achado da lista, a morfema é seleccionada e a regra é aplicada.

Um Backtracking pode mudar a morfema com o próximo candidato para quaisquer dos nodes que são criados neste processo de análise morfológica.

Teoria

Quando aplicando uma regra de enconversion, se um string de carácter aparece em um right hand Condition Windows, o sistema pesquisa automaticamente todas as morfemas que compõem o string de carácter.

Tenta achar uma lista constituinte de morfemas que satisfaçam as condições para uma certa Condição de mando onde o string esta localizado.

Se existe na lista, o EnConverter cria os nodes para todas as morfemas constituintes no Node-list e estes nodes são usados em processos de enconversion adicionais. O olhar-à frente processo de análise morfológica acha as morfemas constituintes para um string de carácter na Condição o Windows e cria os nodes deles para enconversion adicional.

