

*A Gateway to the Network:
International Standards for
Multilingual Lexical Resources*

Alessandro Lenci, Antonio Zampolli, Nicoletta Calzolari, Irina Prodanof

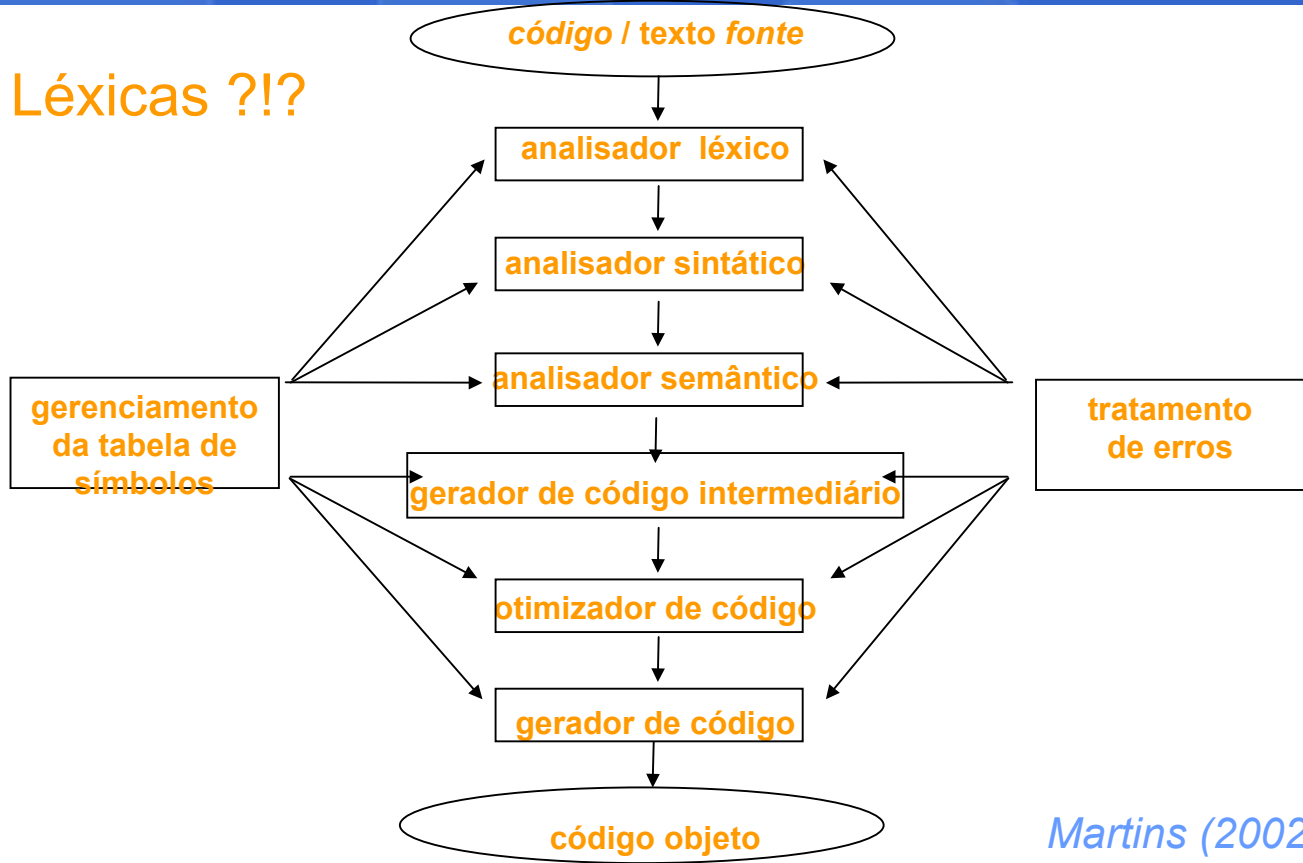
Dipartimento di Linguistica, Università di Pisa

Gabriela Tissiani, M.Eng.
www.lrv.ufsc.br/~gabi

O que é MILE ?

- MILE = Multilingual ISLE Lexical Entry:
 - Definição de um esquema geral para codificação de informações léxicas, desenvolvido pelo ISLE ¹ / CLWG ².
1. **ISLE:** International Standards for Language Engineering
 2. **CLWG:** Computational Lexicon Working Group

Léxicas ?!?



Martins (2002)

FIGURA: estrutura geral de um compilador

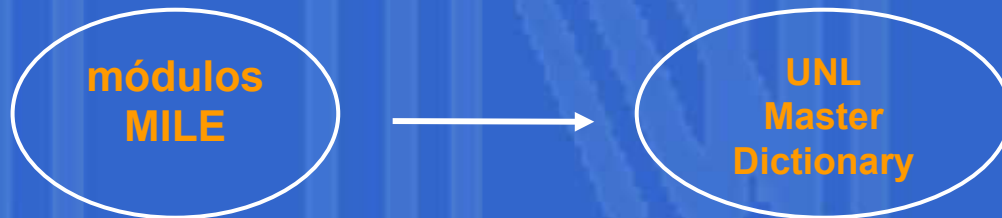
Compilador = tipo de tradutor



- O **analisador léxico** separa a seqüência de caracteres que representa o código ou texto fonte em símbolos básicos da linguagem;
- O **analisador sintático** agrupa os símbolos fornecidos pelo analisador léxico em estruturas sintáticas e o ...
- ... **analisador semântico** utiliza a árvore sintática para determinar o significado do código / texto fonte.

O que tem a ver com UNL ?

- O **ISLE CLWG** propõe utilizar os módulos MILE para “mapear” o conteúdo e a a estrutura dos recursos léxicos multi-línguas da UNL.



Objetivos gerais do MILE

- Ser um “framework” padrão para aplicativos computacionais que utilizam recursos léxicos multi-línguas;
- Foco nas MT (Machine-Translations);
- Desenvolvimento em parceria com:
 - Systran;
 - Microsoft;
 - Outros Laboratórios da indústria e academia (EU & USA)

Objetivos específicos do MILE

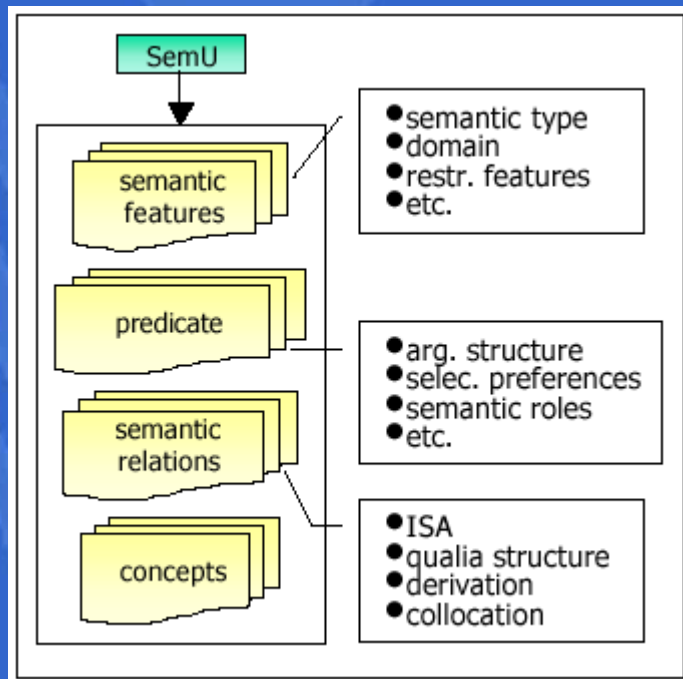
- Listar as principais aplicações que utilizam recursos léxicos e suas recomendações;
- Criar parâmetros de descrições léxicas de conteúdos e estruturas destes aplicativos, como por exemplo, de seus dicionários;
- Sistematizar as informações léxicas para troca de línguas e explicar como elas acontecem nos dicionários.

Arquitetura do MILE

- O MILE foi projetado para:
 - Ser independente do fator língua (serve para qualquer modelo);
 - Oferecer informações claras;
 - Basear-se em análises léxicas que têm um alto grau de embasamento teórico;
 - Prevenir contra soluções de representação específicas.

Arquitetura do MILE

- É centrado na noção de “Unidades semânticas” (SemU);
- Cada “SemU” está relacionada com uma “SynU” (Unidade sintática) e deve conter uma descrição da sintática particular da palavra;
- As SynU” estão associadas às “UM” (Morphological Unit), que descrevem as propriedades morfológicas da palavra.



Relações com o UNL Master Dictionary

- Segundo o CLWG, as noções básicas do MILE podem representar a estrutura e organização do UNL Master Dictionary * .
- * Função do **UNL Master Dictionary**: “linkar” os elementos de uma determinada língua com as UWs - Universal Words (Uchida, 1999).

Conteúdo do UNL Master Dictionary

- *Lemma*: “multi-word expression” em determinada língua;
 - UW (conceito) que o *lemma* expressa;
 - Informações na língua específica sobre as características relacionadas ao *lemma*, expressando seu conceito.
- * Função do **UNL Master Dictionary**: “linkar” os elementos de uma determinada língua com as UWs - Universal Words (Uchida, 1999).

Exemplo do UNL Master Dictionary

- *Ex:
Master
Dict.
Italiano*

```
[programm] {} programma  
"programme(icl>abstract thing)"  
(INMANNER, NOUN, SGA, PLI, MASC, INSCN);
```

```
[programm] {} programma  
"programme(icl>functional thing)"  
(INSCN, NOUN, SGA, INMANNER, PLI, MASC);
```

```
[programm] {} programmare  
"plan(icl>do)"  
(AUXAV, OBJDIINF, SAMESSUBJ, DO, Vamare3, VER  
B, TRANS);
```

```
[programm] {} programma  
"schedule(icl>plan{>abstract thing})"  
(SGA, OBJDIINF, PLI, NOUN, MASC, INSCN);
```

UNL Master Dictionary & MILE

- Sugestão de relação entre o UNL Dic. e o MILE:

UNL Dictionary	MILE
Uws	multilingual concepts
relations between Uws	multilingual relations between concepts
morphosyntactic features	SynU and MU components (inflect. classes, descriptions, positions, grammatical functions, etc.)

- Exemplo de representação XML-based MILE de uma das entradas UNL do exemplo anterior:

```
<MILE_entry>
  <MU>
    <lemma>programma</lemma>
    <gender>MASC</gender>
  </MU>
  <SynU>NOUN</SynU>
  <SemU id="programm1">
    <Concept_Rep status="MULTILINGUAL">
      @abstract_thing
    </Concept_Rep>
  </SemU>
</MILE_entry>
```

Considerações Finais dos Autores

- A necessidade crescente de desenvolvimento de recursos léxicos é um desafio para a comunidade de engenharia lingüística, que retorna em uma contribuição importante para a sociedade da informação...

* Referências:

Martins, Joyce. “Linguagens Formais e Compiladores”, Apostila Univali, maio 2002.

Uchida, H., Zhu, M., Della Senta, T. (1999). The UNL. A Gift for a Millennium, Institute of Advanced Studies, The United Nations University.