Nanociência/Nanotecnologia e Biocombustíveis vistos pelo Modelo SILEX: análise morfolexical de terminologias

Joel Sossai Coleti

(PPGL-UFSCar / FAPESP)

Orientadora:

Profa. Dra. Gladis Maria de Barcellos Almeida

(UFSCar)

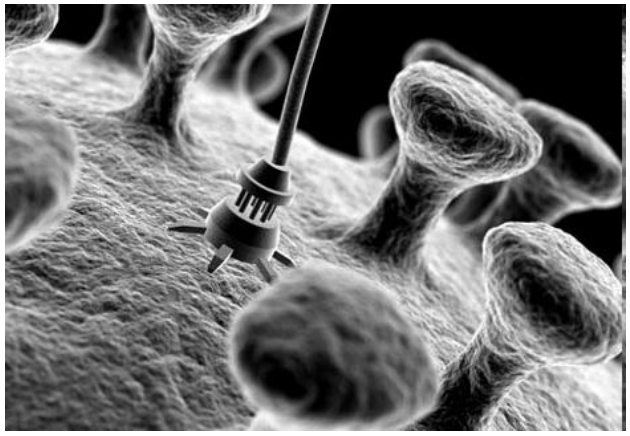
Coorientadora:

Profa. Dra. Margarita Correia

(Universidade de Lisboa)

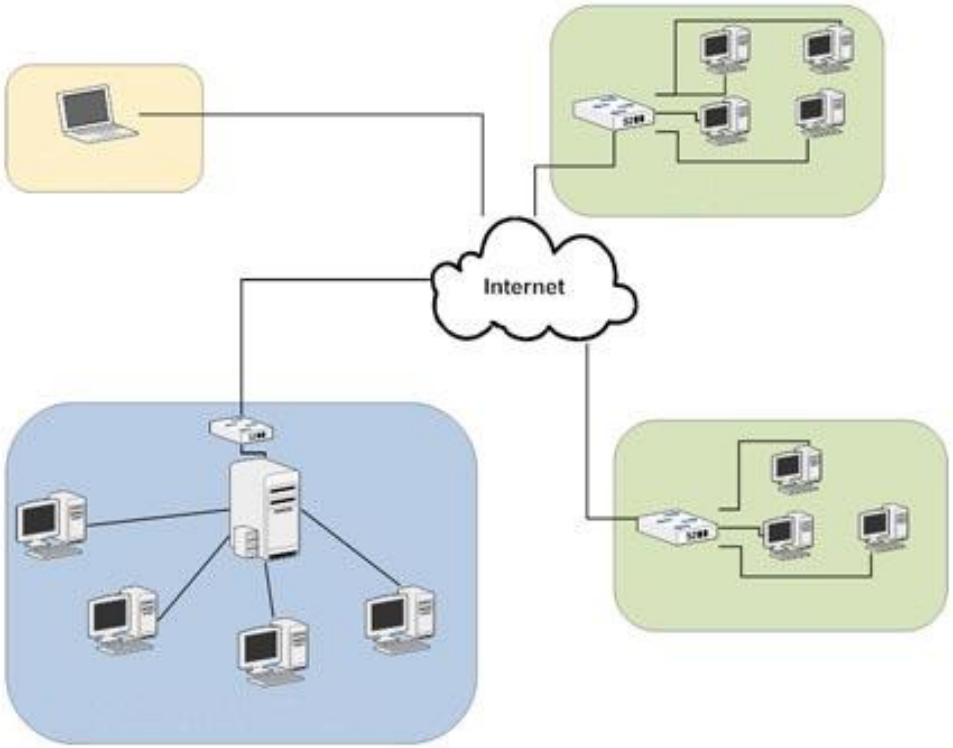
Objetivo 1:

✓ **analisar e descrever** os processos morfológicos possíveis nas terminologias da Nanociência & Nanotecnologia e Biocombustíveis **em português** (variante brasileira), verificando-se os **principais processos de construção dos termos e os morfemas mais produtivos**



Objetivo 2:

✓ organizar uma **base de dados** de maneira que seja possível a **implementação computacional** e a **disponibilização on-line** dos dados obtidos.



Motivação

Mordebe (Portal da Língua Portuguesa, ILTEC)

- Base de dados sobre as características formais do léxico do português
  - ortografia,
  - Flexão,
  - e as relações morfológicas
- Informações sobre semântica e etimologia não é registada.
- Predominância do português europeu, com registros de outras variedades do português.

Repertório terminológico (RT):

Grupo de Estudos e Pesquisas em Terminologia – GETerm:

- Nanociência/Nanotecnologia: **Terminologia em Língua Portuguesa da Nanociência e Nanotecnologia: Sistematização do Repertório Vocabular e Elaboração de Dicionário-Piloto – NanoTerm**; apoio CNPq/Processo n°. 400506/2006-8

- Biocombustíveis: **Terminologia de Biocombustíveis: descrição semântica e morfológica com vistas à sistematização**; apoio CNPq/Processo n°. 473414/2007-4

Corpora - compilação e sistematização

i) Seleção

ii) Compilação e manipulação

iii) Nomeação, cabeçalho e anotação

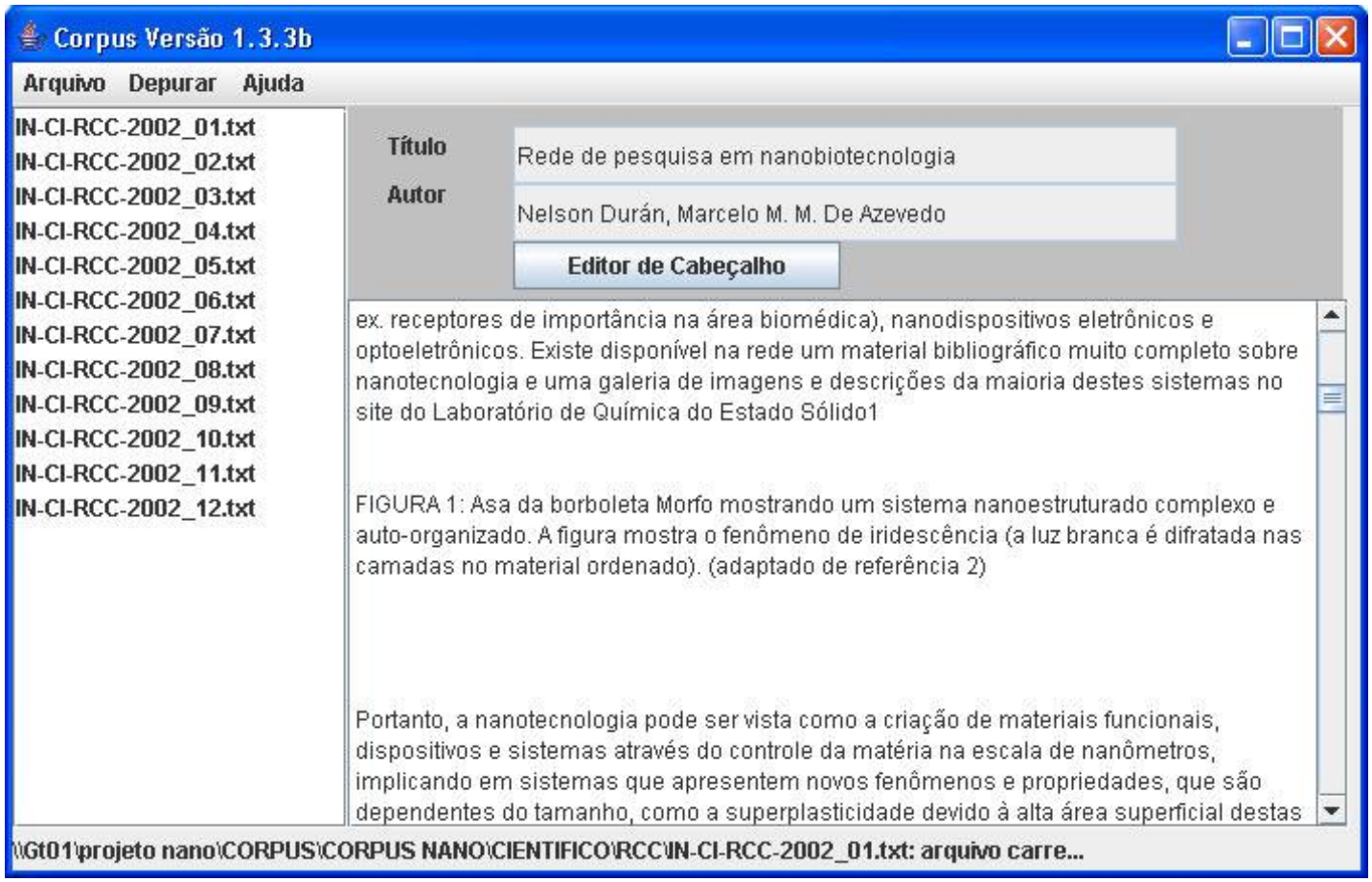
(ALUÍSIO & ALMEIDA, 2006)

*Corpora - **Seleção***

✓ Estudo exploratório dos textos existentes: gêneros, tipos textuais e fontes.

✓ Buscas iniciais motivadas por termos notadamente reconhecidos de cada área.

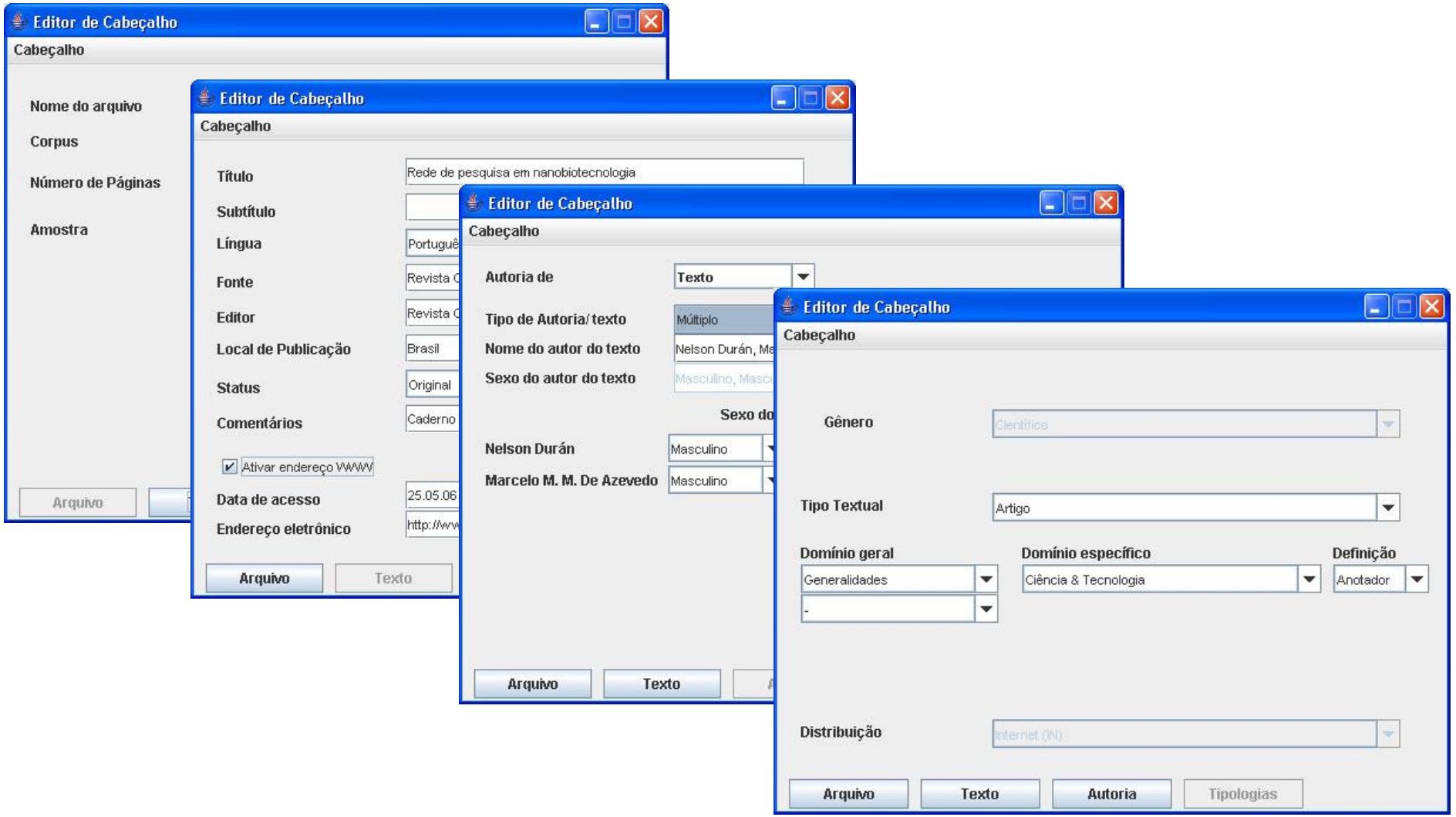
Editor de Cabeçalhos



Editor de Cabeçalhos do Projeto Lacio-*Web*

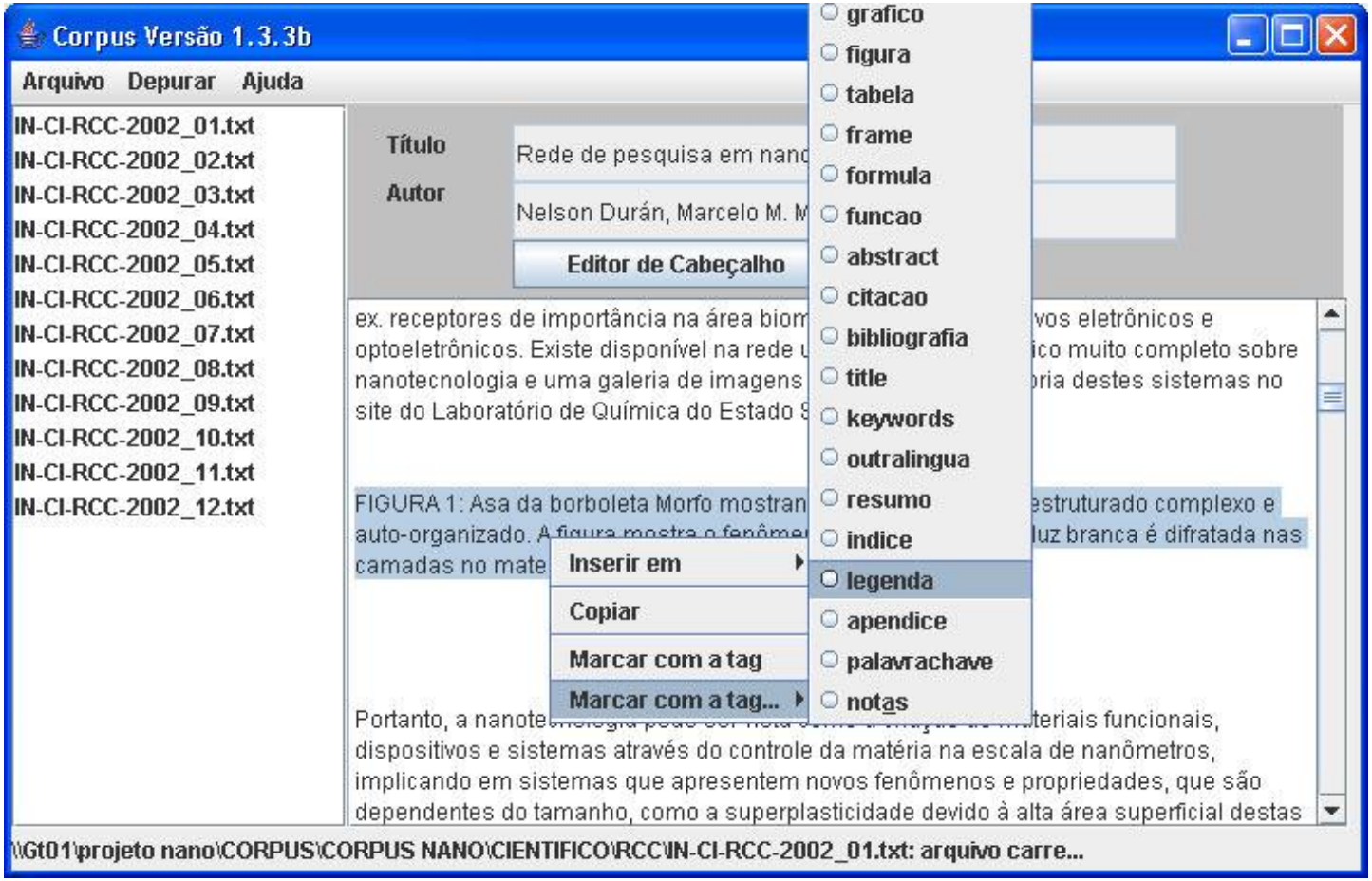
Corpora - Nomeação, Cabeçalho e Anotação

Anotação Estrutural Externa - Cabeçalho



Corpora - Nomeação, Cabeçalho e Anotação

Anotação Estrutural Interna - Etiquetas



Corpora - Nomeação, Cabeçalho e Anotação

Nomeação

- ✓ Por sigla, padronizada por Gênero

Exemplo: Para textos científicos:

IN-CI-Gomes-01abr03_17

***IN**: Texto divulgado pela Internet*
***CI**: Gênero textual Científico*
***Gomes**: Sobrenome do Autor*
***01abr03**: Data de publicação (01 de abril de 2003)*
*_**17:** 17° texto obtido da mesma fonte (Banco de Teses da Capes)*

RT - extração semiautomática dos candidatos a termos

» Consideramos *semiautomática por haver interferência* humana na fase da validação e limpeza da lista gerada pelo programa selecionado para a extração

» Os *candidatos a termos* constituem itens léxicos que se comportam nos seus respectivos contextos como termos, mas cuja autenticidade será validada posteriormente

i) Escolha do software

ii) Elaboração da StopList

iii) Limpeza de falsos candidatos a termos

iv) Validação pelo especialista

RT - extração semiautomática dos candidatos a termos

» Pacote NSP (*N-gram Statistics Package*): implementado por Ted Pedersen, Satanjeev Banerjee e Amruta Purandare, da Universidade de Minnesota, Duluth.

» Resultados: 268.043 candidatos a termos (de um corpus de 2 565 490 palavras), após a revisão pelo linguista e análise pelo especialista de domínio foram efetivamente considerados como termos apenas 3069. (Nanociência/Nanotecnologia)

Terminologia e Morfologia:

Termos são unidades da língua geral associadas a um valor especializado apenas em situações discursivas e pragmáticas específicas, ou seja, quando utilizadas em âmbitos de especialidades (Cabré, 2006).

Morfologia:

Forma de análise: Modelo SILEX

Syntaxe, Interprétation, LEXique

- Criado por Danielle Corbin e posteriormente desenvolvido por sua equipe de trabalho

- O SILEX já possui aplicações profícuas para análise do português: Graça Maria Rio-Torto (Universidade de Coimbra) e Margarita Correia (Universidade de Lisboa)

Modelo SILEX: fase atual

- Modelo de ‘morfologia construcional’: tem como **objeto de estudo a construção de palavras**, não apenas por derivação, mas com recurso a outros processos de construção, tais como a composição, os processos deformacionais ou a lexicalização de sintagmas (CORREIA, 2004a).

Modelo SILEX: fase atual

- Permite o tratamento de:
  - Regras e dos operadores envolvidos na construção de palavras
  - Relação entre a estrutura de uma unidade lexical e a sua capacidade denominativa
  - Mecanismos semânticos associados

Modelo SILEX: associativo e estratificado

O modelo SILEX assume-se como um modelo **associativo e estratificado**:

Por ‘modelo associativo’ entende-se aquele cujas Regras de Construção de Palavras (RCPs) permitem **construir conjuntamente a estrutura morfológica e a interpretação semântica** das palavras construídas

É um ‘modelo estratificado’ porque é **composto por vários níveis**, ao longo dos quais se vai construindo o significado das palavras construídas.

Repertório terminológico:

- 927 termos de Biocombustíveis

- 3069 termos de Nanociência/Nanotecnologia

3996 termos

Recorte: nível morfológico

- 425 Unigramas de Biocombustíveis

- 1794 Unigramas de Nanociência/Nanotecnologia

2219 termos

Delimitação do repertório terminológico:

- Em consonância com o modelo teórico adotado são excluídas do repertório a ser analisado as palavras não-construídas morfologicamente

(palavras não construídas em língua portuguesa, nomes próprios e siglas)

Delimitação do repertório terminológico:

- Importações parcialmente ou completamente opacas:

BOTTOM-UP

COATING

CLUSTER

DIP-COATING

DISPLAY

Delimitação do repertório terminológico:

- Siglas:

MEMS

MET

MEV

MFA

MHZ

Delimitação do repertório terminológico:

- Nomes Próprios:

DIESEL

RAMAN

RIETVELD

RIKEN

VAN DER WAALS

Resultado da Delimitação

- 423 Unigramas de Biocombustíveis

- 1637 Unigramas de Nanociência/Nanotecnologia

2060 termos

Primeiros resultados:

A composição culta (formação a partir de afixos gregos ou latinos)

é demasiadamente frequente em linguagens de especialidade e

também no vocabulário de Nanociência/Nanotecnologia

Composição culta:

Prefixação: ***nano-*** (prefixo grego, adotado na 11ª Conferência Geral de Pesos e Medidas (Resolução nº 12 de 1960) equivalente a um multiplicador 10^{-9}, ou seja, um bilionésimo da unidade indicada)

Composição culta: a prefixação por nano-

- Dimensão Nanométrica:

(nanopartícula, nanossensor, nanotubo, nanotecido, ...)

- Relativo à Nanociência/Nanotecnologia:

(nanoaventura, nano-ética, nanobiotecnologia, nanomecânica, ...)

Composição culta: a prefixação por nano-

- O prefixo nano- ocorre como forma presa (unida ou não por hífen) e como forma livre

- As formas presas podem ser dividas entre as que ocorrem com função substantival daquelas que ocorrem com função adjetival.

- Não há unicidade categorial de base

Composição culta: a prefixação por nano-

Substantivos	Adjetivos
Nanociência	**Nanocristalino**
Nanocompósito	**Nanométrico**
Nanoescala	
Nanoesfera	
Nanoestrutura	
Nanofio	
Nanofita	
Nanomaterial	
Nanômetro	
Nanopartícula	
Nanotecnologia	
Nanotubo	

Composição culta: outros casos

Bio-

Micro-

Infra-

Bootstrapping:

O resultado obtido pela análise dos termos fornecerá, ao final da pesquisa, material linguístico que poderá ser transformado em matéria para um novo processo contribuindo assim para o refinamento/melhoramento do novo resultado e assim sucessivamente tornando o processo de extração de termos cada vez mais eficiente.

Bibliografia

ALMEIDA, G.M.B. A Teoria Comunicativa da Terminologia e a sua prática. **Alfa** (Araraquara), v. 50, p. 81-97, 2006. Disponível em: http://www.alfa.ibilce.unesp.br/download/v50-2/06-Almeida.pdf

ALUÍSIO, Sandra Maria ; ALMEIDA, G. M. B. . **O que é e como se constrói um *Corpus*? Lições aprendidas na compilação de vários corpora para pesquisa lingüística**. Calidoscópio (UNISINOS), v. 4, p. 156-178, 2006.

CABRÉ, M.T. **La terminología:** representación y comunicación – elementos para una teoria de base comunicativa y outros artículos. Barcelona: Institut Universitari de Lingüística Aplicada, 1999.

CABRÉ, M.T. Theories of Terminology: their description, prescription and explanation. **Terminology**, v. 9, n. 2, p. 163-200, 2003.

CORBIN, D. **Morphologie dérivationnelle et structuration du lexique**. 2 vols. Tubinga: Max Niemeyer Verlag, 1987.

Bibliografia

CORBIN, D. Form, structure and meaning of constructed words in an associative and stratified lexical component. In: **Yearbook of Morphology 2**. Dordrecht: Foris Publications, 1989, p. 31-54.

CORBIN, D. Introduction - La formation des mots: structures et interprétations. In: **Lexique 10**. Villeneuve d'Ascq: Presses Universitaires de Lille, 1991, p. 7-30.

CORBIN, D. La représentation d'une "famille" de mots dans le Dictionnaire dérivationnel Du français et ses corrélats théoriques, méthodologiques et descriptifs. In: **Recherches linguistiques de Vincennes**, 1997 pp. 5-37 + errata.

CORBIN, D. Programme de recherche (1997-2003). Le Dictionnaire des affixes et Le Dictionnaire dérivationnel du français: mises en pratique d'une théorie morphologique. In: **Lexique 16**. Villeneuve d'Ascq: Presses Universitaires du Septentrion, 2004, p. 53-66.

Bibliografia

CORBIN, P. Introduction: Lexique 16, treize ans après Lexique 10. In: **Lexique 16**. Villeneuve d'Ascq: Presses Universitaires du Septentrion, 2004, p. 9-52.

CORREIA, M. Introdução. In: **A denominação das qualidades em português – contributos para a compreensão da estrutura do léxico português**. Tese de doutoramento apresentada à Universidade de Lisboa, 1999.

CORREIA, M. **Denominação e construção de palavras**. Lisboa: Edições Colibri, 2004.

CORREIA, M. Terminologia e morfologia: marcas morfológicas da génese do vocabulário da Náutica em português. In: M. T. CABRÉ, R. ESTOPÀ & C. TEBÉ (eds.), **La terminología em el siglo XXI** – Contribución a la Cultura de la Paz, la Diversidad y la Sostenibilidad (Actas del IX Simposio Iberoamericano de Terminología RITERM04). Barcelona: IULA / Universitat Pompeu Fabra, 2006, p. 31-52. Disponível em: http://www.iltec.pt/pdf/wpapers/2004-mcorreia-barcelona.pdf

RIO-TORTO, G. M. **Morfologia derivacional** – teoria e aplicação ao Português. Porto: Porto Editora, 1998.

MUITO OBRIGADO!!

Joel Sossai Coleti (joelscoleti@gmail.com)

Gladis Maria de Barcellos Almeida (gladis@ufscar.br)

www.geterm.ufscar.br