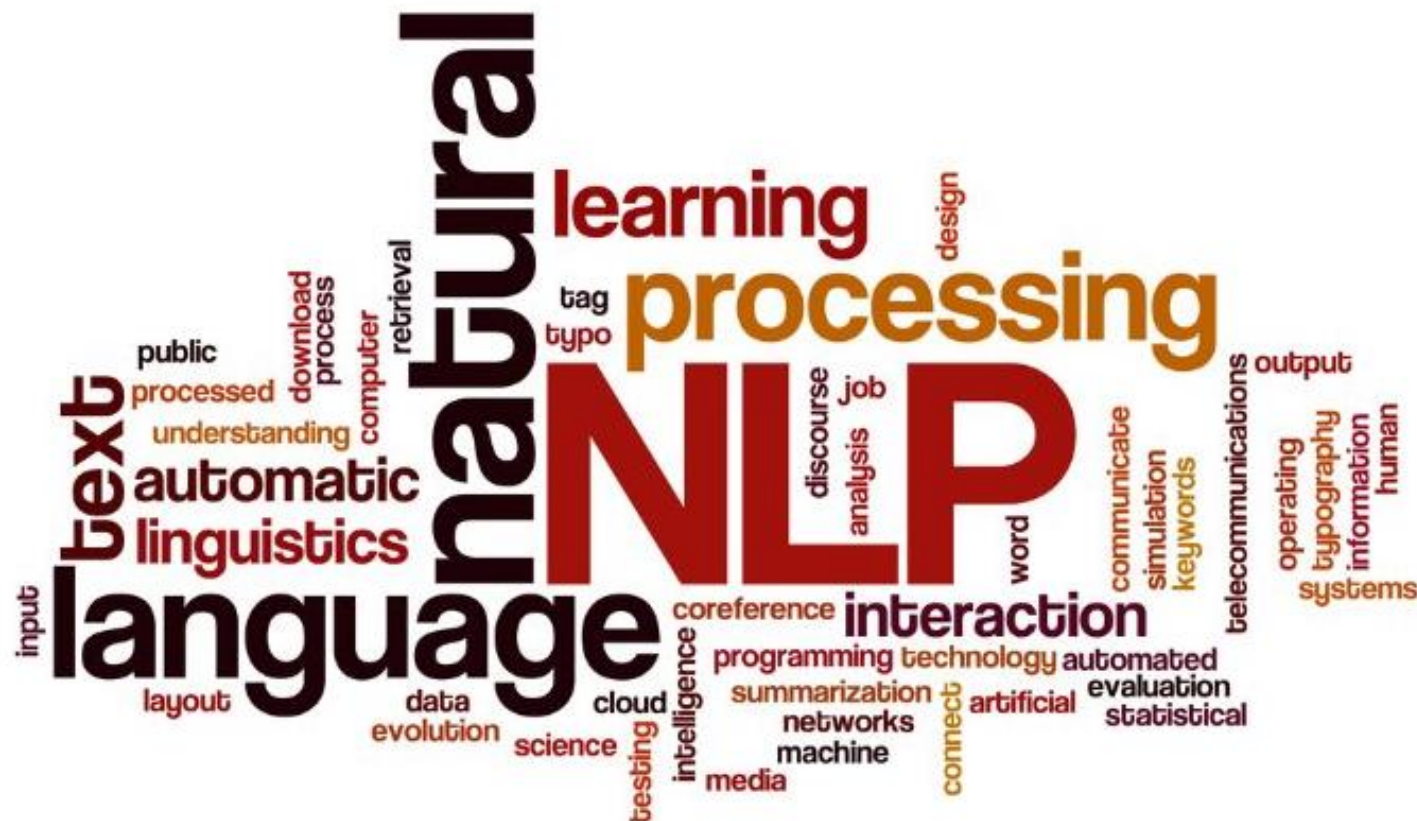


Processamento de Linguagem Natural (PLN)



Histórico

- O processamento de linguagem natural (PLN), nasceu praticamente junto com o computador.
- Na Segunda Guerra os militares americanos tinham interesse de traduzir automaticamente conversações gravadas dos alemães, russos e de outras nações;
- O material científico estrangeiro “obtido” pela comunidade científica precisava de traduções rápidas, já que no curso da guerra, a cada dia fazia-se novas descobertas científicas.

Histórico da tradução automática

- A possibilidade de fazer tradução automática de publicações sobre tecnologia e ciência deixou os cientistas animados.
- Essa foi uma das razões para que pesquisas fossem financiadas na área.
- Os primeiros trabalhos sobre PLN (basicamente tradução automática) começaram em 1946, traduzindo do idioma russo para o inglês.
- A tradução era executada palavra por palavra, baseadas em uma lista de palavras chave.

Histórico da tradução automática

- Em 1948 num trabalho do inglês Pichens, surge a preocupação com as regras de construção de frases.
- Em 1952 acontece o primeiro congresso sobre tradução automática, nos EUA, no MIT (*Massachusetts Institute of Technology*) com a participação de 18 pesquisadores.
- O governo norte-americano encomenda um estudo sobre as pesquisas em tradução automática (Relatório ALPAC -Automatic Language Processing Advisory Comitee).
- O relatório publicado em 1966 teve um teor fortemente negativo que provocou o corte das verbas de financiamento.

Histórico da tradução automática

- Na Europa dos anos 80, surge o renascimento do interesse na tradução automática, em função da criação da Comunidade Europeia.
- Em 1982 é lançado o projeto EUROTRA que consistia em um sistema de tradução automática para as nove línguas de países que constituíam a Comunidade Europeia.
- O ambicioso objetivo do projeto EUROTRA não era apenas criar um programa, mas também estabelecer fundamentos para o estudo da língua natural através do computador.

Áreas de Aplicação

- As principais áreas de aplicação para o processamento de linguagem natural são:
 - Acesso a banco de dados;
 - Recuperação de informação;
 - Extração de informação;
 - Tradução automática;
 - Geração de resumos.

Reconhecimento de Linguagem Natural	Reconhecimento de Voz	Digitadores automáticos; Controle por voz; e Autenticação por voz.	entrada
	Reconhecimento da escrita	OCR - <i>Optical Character Recognition</i> ; Reconhecimento de letra de forma; e Reconhecimento de letra de mão.	
Reconhecimento de Linguagem Natural	Produção de voz	Gerar fala a partir de um texto; Ensino de Idiomas; e Auxílio a deficientes visuais.	saída
	Análise e produção de texto	Gerar textos como análise de outros textos; e Correções de texto; e Tradutores automáticos.	

Acesso a Banco de Dados

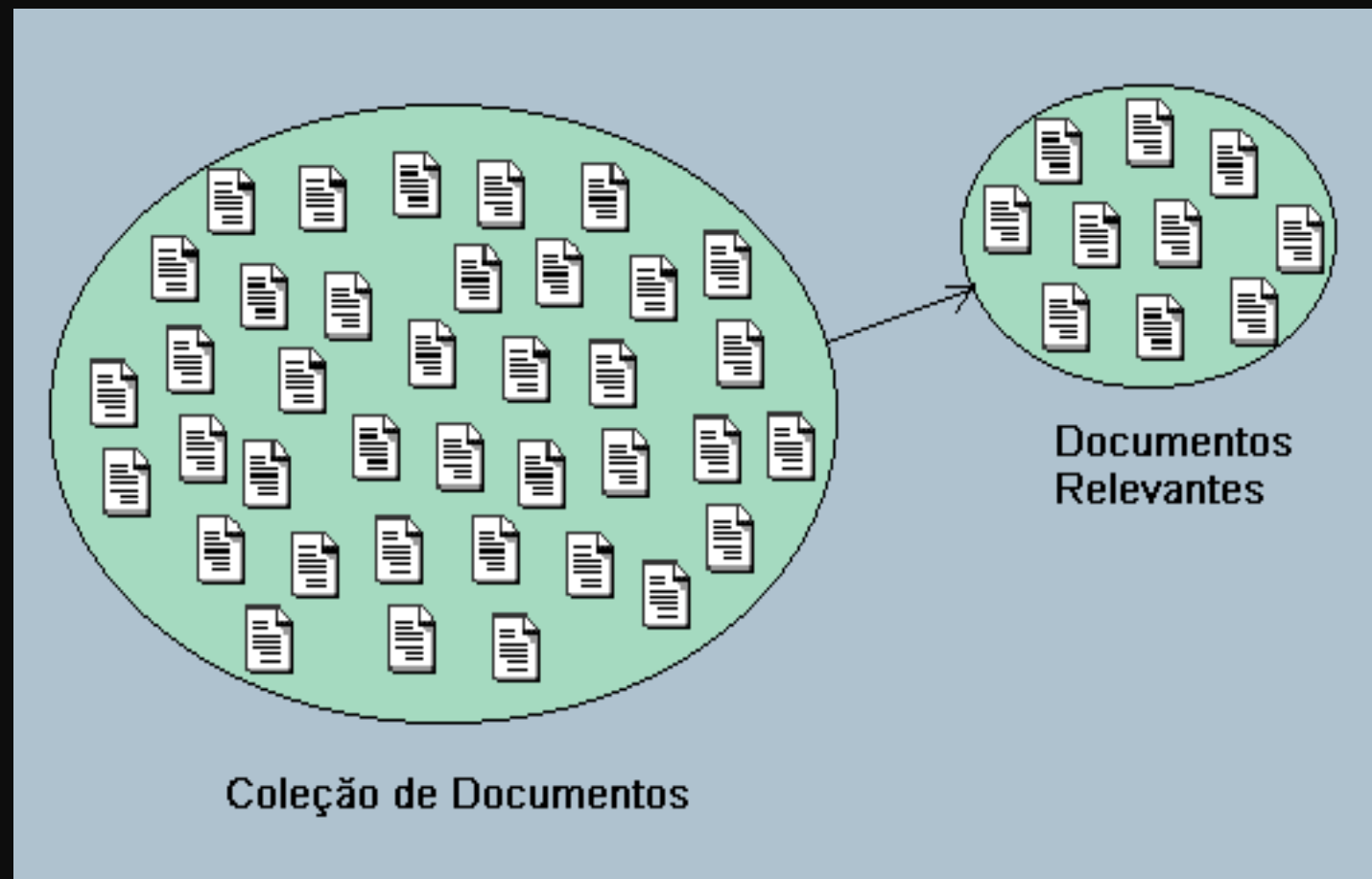
- O acesso as informações em um banco de dados é feito usualmente utilizando-se linguagens que permitem a consulta nesses bancos, conhecidas com *query languages* (SQL, CQL, DMX, XQuery entre outras).



- Exemplo em SQL: `SELECT * FROM tabela_pessoa;`

Recuperação de Informação

É o estudo e desenvolvimento de técnicas que permitam encontrar documentos relevantes de uma coleção de documentos.



Extração de Informação

- Na extração de informação procura-se por informações diretamente nos textos, mostrando a informação ao invés de documentos.
- São utilizadas técnicas baseadas na busca de determinadas palavras-chaves (*tags*), tais como:
 - Nome de pessoas;
 - Nome de empresas;
 - Nome de produtos;
 - Entre outras.

Tradução Automática

- Tradução automática é a tradução por computador de frases dadas numa língua para outra língua. Os primeiros trabalhos utilizavam dicionários bilíngues e faziam tradução palavra a palavra.
- A teoria linguística começou a ser incorporada nas traduções automáticas com os trabalhos de Noam Chomsky. Ele dizia que o conhecimento de uma língua natural poderia ser descrita através de um conjunto finito de regras, e essas seriam universais.
- Essas regras poderiam gerar um número infinito de frase de uma língua, que poderia ser gramatical (pertencente à língua) ou agramatical (não pertencente à língua).

Problemas com PLN

Homonímia Lexical

Ocorre quando uma palavra tem mais de um sentido.



Problemas com PLN

Ambiguidade sintática

Ocorre quando a sentença aceita duas análises sintáticas diferentes, como no exemplo:

Viajando pela primeira vez para a Europa cruzei com um grupo de jovens brasileiros.

Quem viajou pela primeira vez? Eu ou o grupo de jovens brasileiros?

Problemas com PLN

Ambiguidade de Escopo

Percebe-se às vezes que a sentença indica dois ou mais escopos, como no exemplo:

Apesar de ser exímio advogado, o procurador da União não cumpre todas as disposições estatutárias. Essa sentença poderia significar que:

- O procurador descumpre todas as disposições (descumpre todas ou seja, tem por norma violar a legislação);
- O procurador cumpre as disposições, mas não todas (não tem por norma violar a legislação, mas comete falhas).

Problemas com PLN

Diferentes correlações possíveis

É a compatibilidade de um anafórico com dois ou mais antecedentes distintos, também chamada de ambiguidade anafórica, como no exemplo:

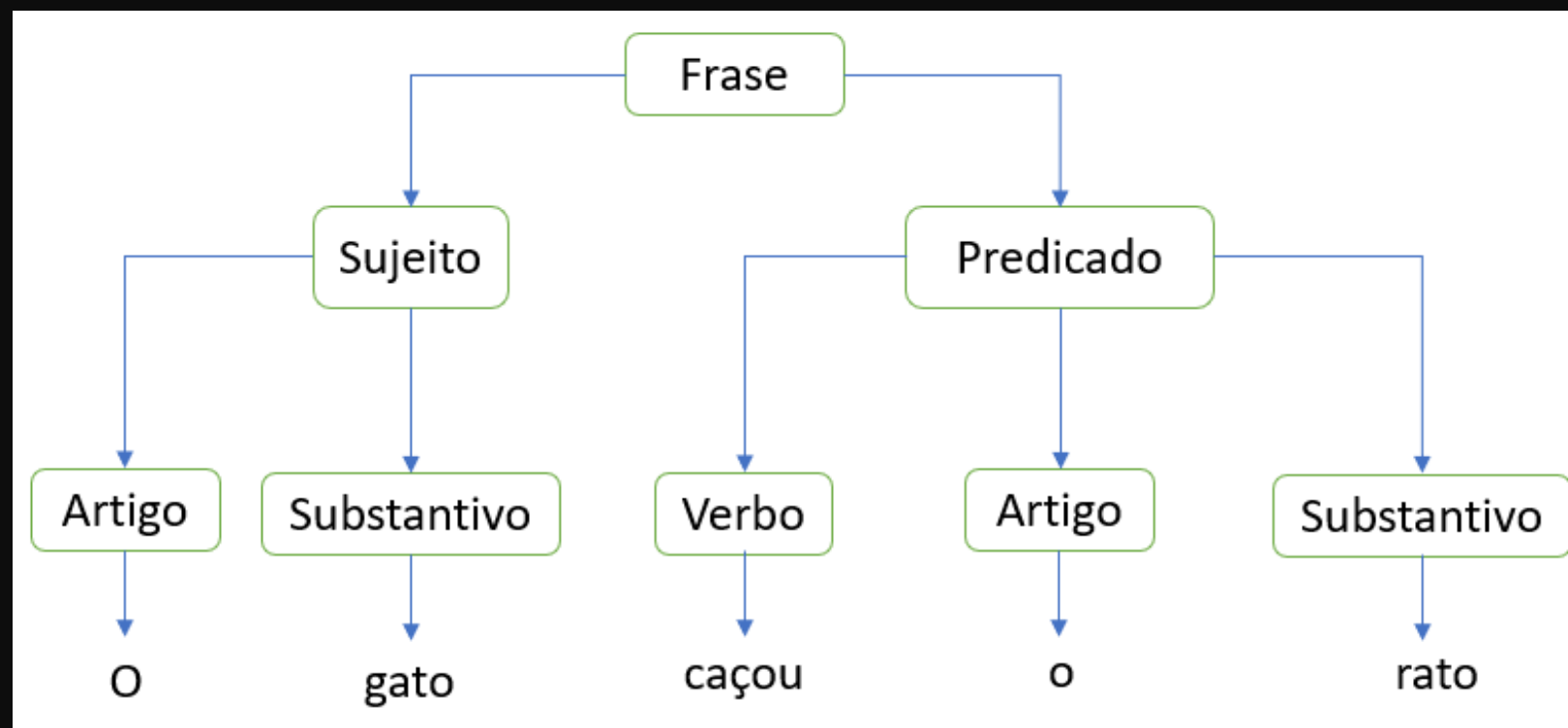
O ladrão entrou na casa do prefeito e tirou toda a sua roupa.

De quem foi tirada a roupa? Do ladrão ou do prefeito?

PNL Desenvolvimento

As fases do desenvolvimento de um sistema de PLN são:

1. Análise morfológica;
2. Análise Sintática;
3. Análise Semântica;
4. Análise do Discurso;
5. Análise Pragmática.



Análise Morfológica

É o estudo da estrutura e da classificação das palavras em função do uso de:

Substantivos	Adjetivos
Pronomes	Verbos
Numerais	Artigos
Advérbios	Preposições
Conjunção	Interjeição

Análise Morfológica

As línguas naturais possuem regras morfológicas que produzem as possíveis variantes de cada palavra.

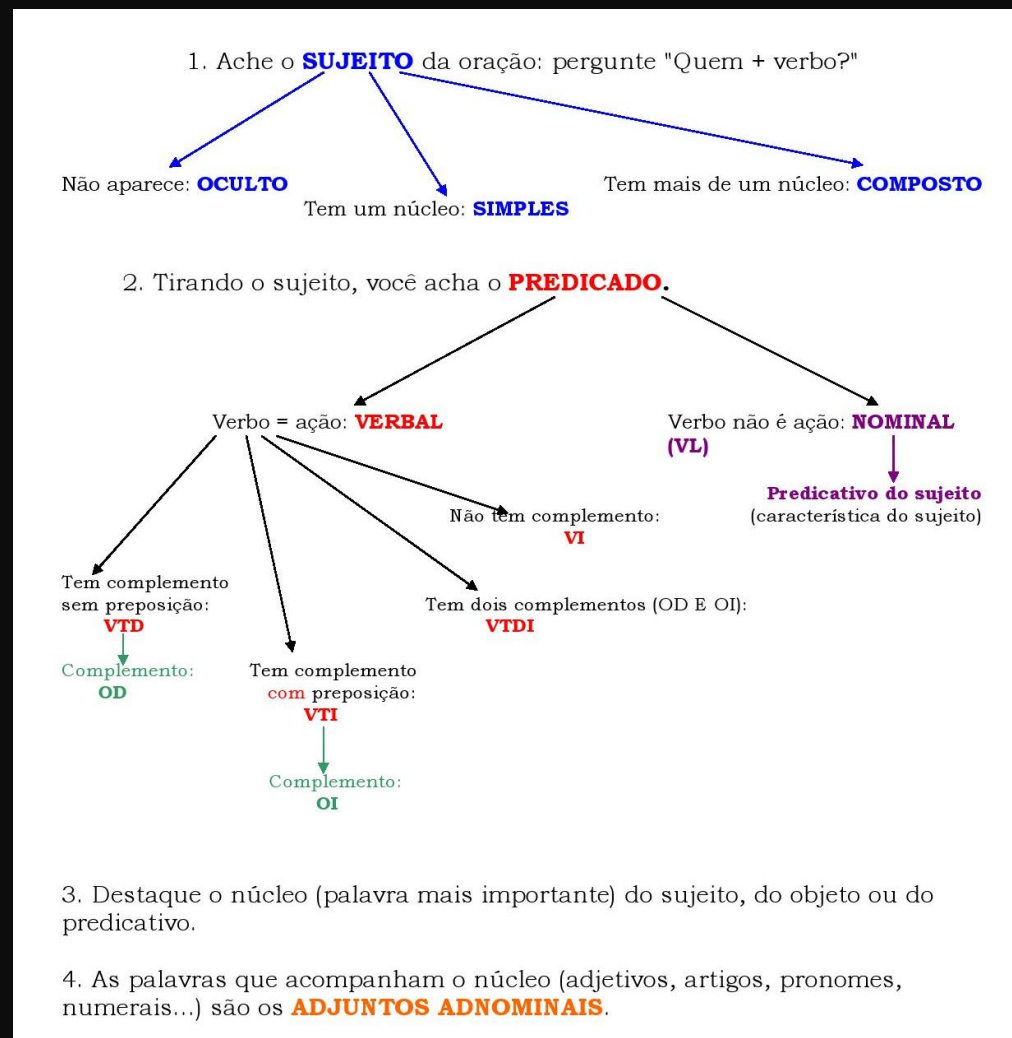
Exemplo:

- **Construir** tem como variantes, entre outras, **construção** e **construído**.
- Um pedaço da palavra (**constru**) se repete nas demais, que receberam a aposição do que chamamos de **sufixos**.
- Este pedaço de palavra que se mantém nas variantes, chamamos de **lexemas**.

Análise Sintática

É o estudo das unidades básicas da linguagem, ou seja, as sentenças.

Na fase da análise sintática, o sistema de processamento de linguagem natural verifica se a sequência das palavras nas sentenças são válidas para a gramática utilizada.



Análise Semântica

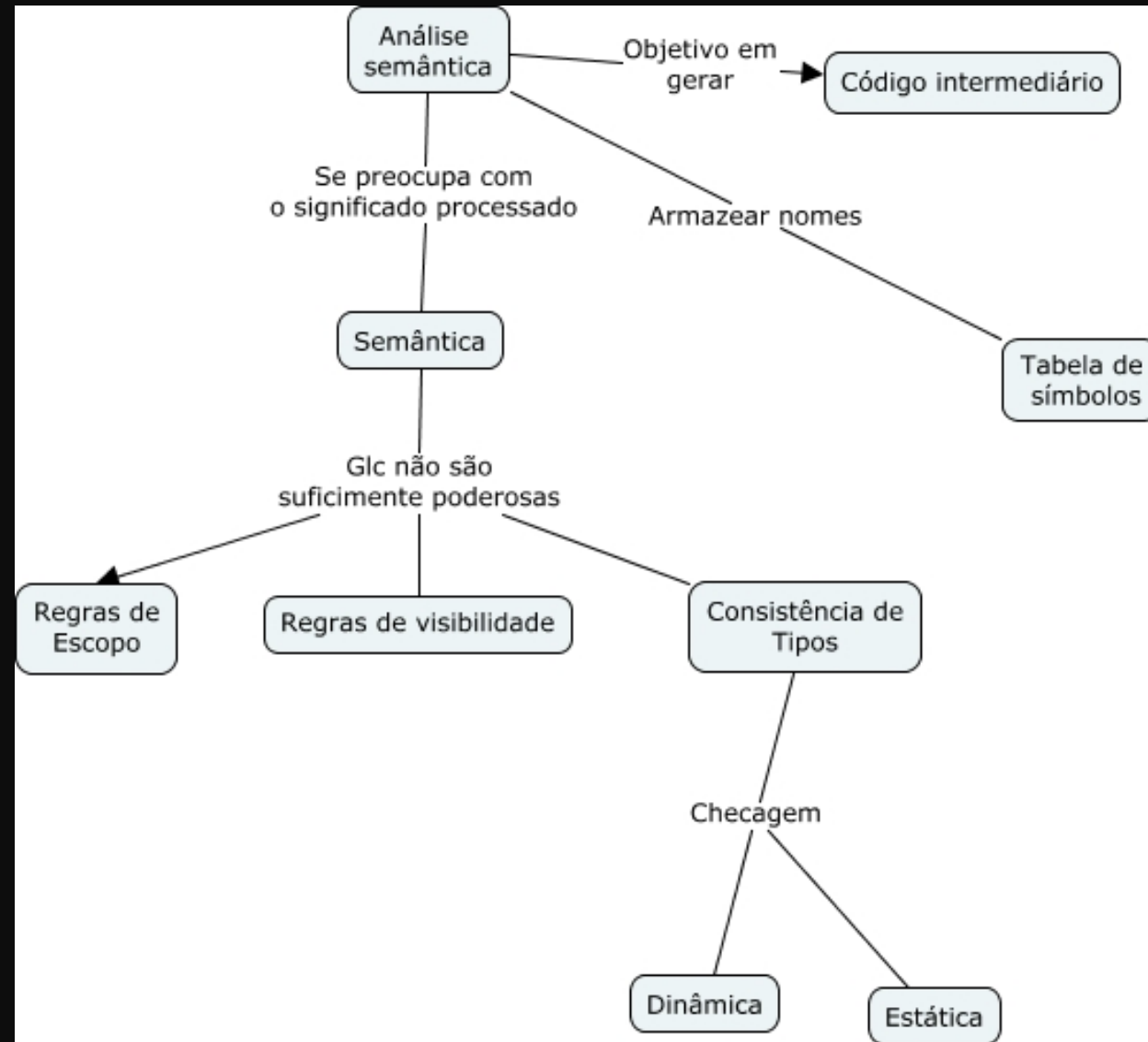
Análise Semântica é o estudo do significado das palavras e da interpretação das frase ou enunciados.

Durante a análise semântica, utiliza-se a estrutura gerada durante a análise sintática para construir outras estruturas que representem o significado das sentenças.

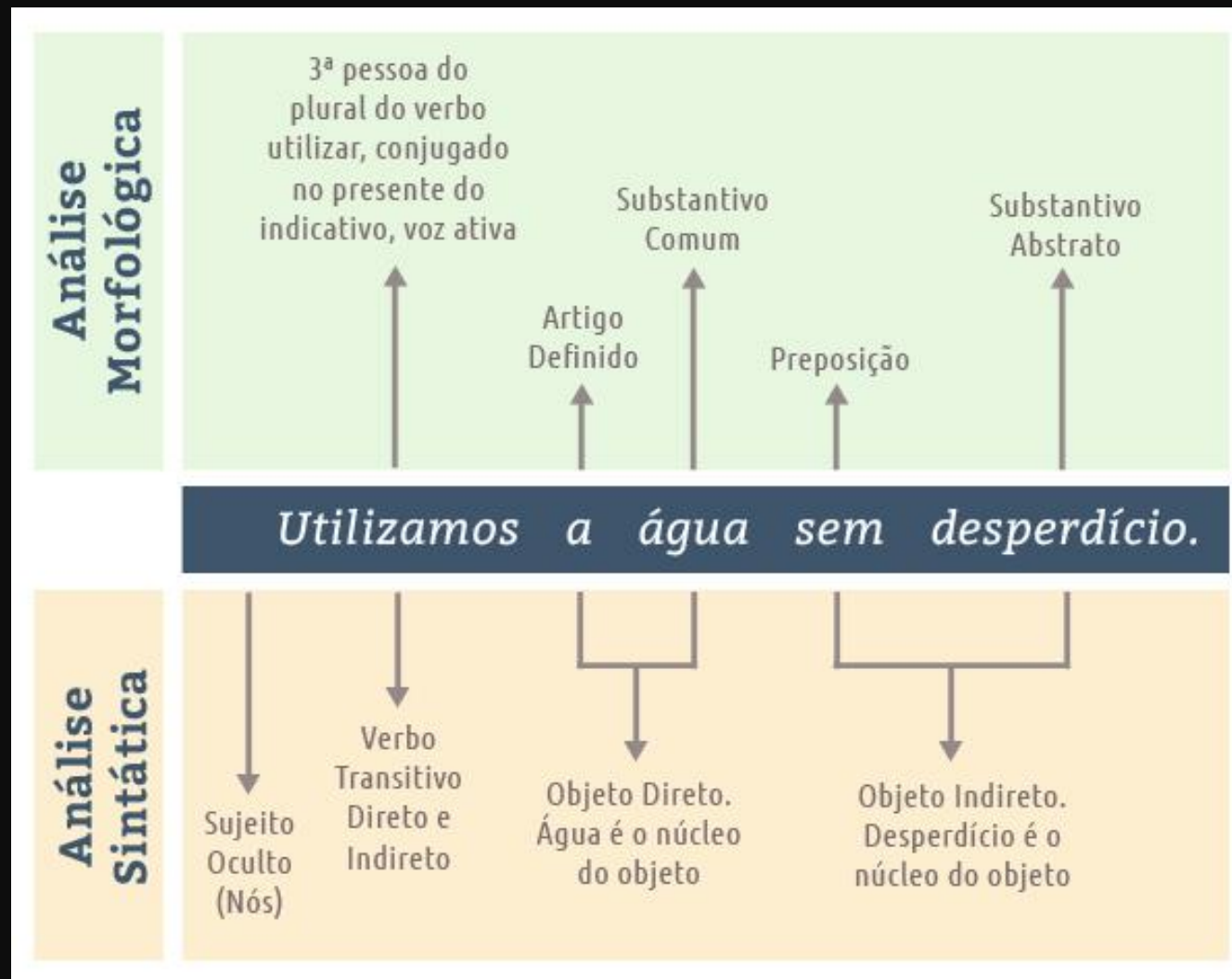
Os formalismos utilizados nesta fase do processamento de linguagem natural podem ser classificados em fracos e fortes:

- Formalismos fracos: redes semânticas e frames;
- Formalismos fortes: gramáticas de casos, dependência conceitual e *scripts*.

Análise Semântica

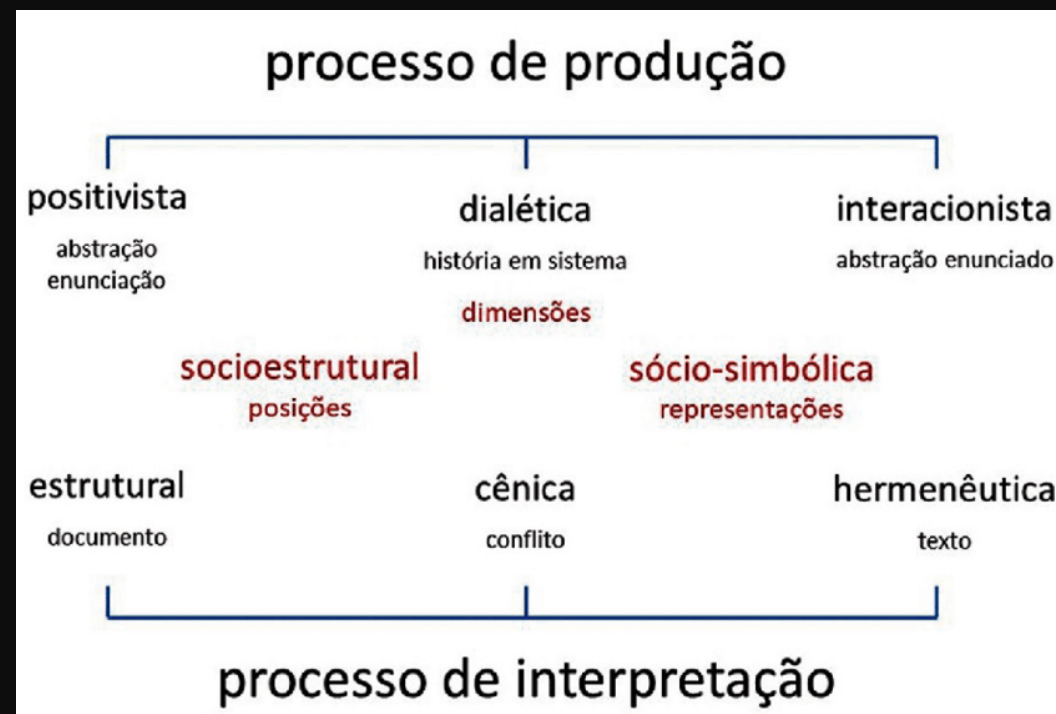


Análise Semântica e Sintática



Análise do Discurso

- É a identificação da estrutura do discurso.
- O discurso é também organizado em unidades constituídas por um mais elementos denominados segmentos do discurso.



Análise Pragmática

- Na análise pragmática são estudados os enunciados, ou seja, os significados das frases, sob o ponto de vista dos interlocutores.
- Esta análise é de suma importância principalmente nos diálogos onde é preciso determinar as intenções dos interlocutores.



Workflow do PLN



Bibliografia

Braga, A. P., Ludermir, T. B., & Carvalho, A. P. (2000). *Redes Neurais Artificiais - Teoria e aplicações*. Rio de Janeiro: LTC - Livros Técnicos e Científicos.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (10 de Novembro de 2007). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, pp. 159-190.

Kovács, Z. L. (1996). *Redes Neurais Artificiais - Fundamentos e Aplicações*. São Paulo: Edição Acadêmica São Paulo.

Michael, N. (2002). *Artificial Intelligence - A guide to Intelligent Systems*. Edinburgh: Pearson.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1983). *Machine Learning*. Berlin: Springer-Verlag.

Minsky, M. (2006). *The Emotional Machine*. New York: Simon & Schuster.

Rich, E., & Kevin, K. (1993). *Inteligência Artificial*. São Paulo: Makron Books.

Suart, R., & Peter, N. (2004). *Inteligência Artificial*. Rio de Janeiro: Elsevier.