

Python para Processamento de Linguagem Natural

Introdução ao Processamento de Linguagem Natural

Roney Lira de Sales Santos
Rogério Figueredo de Sousa
Prof. Thiago A. S. Pardo

[PLN: Definição]

- Instruir o computador a lidar com a língua, ou, como se diz informalmente, a “ler e escrever”
 - Interpretação de textos
 - Tradução automática
 - Revisão gramatical
 - Busca de respostas para perguntas
 - Sumarização
 - Auxílio a escrita e ao aprendizado de línguas
 - Etc.

- Multidisciplinar
 - Computação
 - Linguística

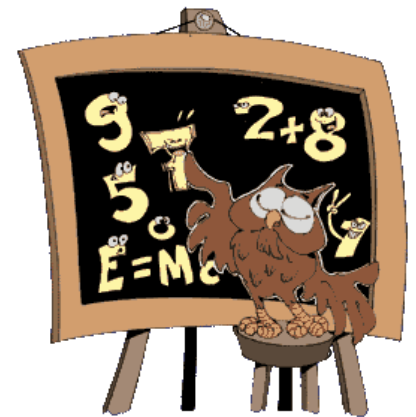


[Língua Natural]

- Língua humana



- Em oposição às linguagens artificiais
 - Matemática, lógica, linguagens de programação de computadores



Questão

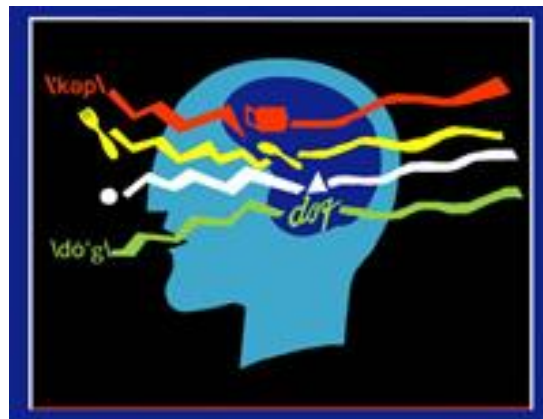
- Qual a diferença entre “língua” e “linguagem”?
- É Processamento de Linguagem Natural ou Processamento de Línguas Naturais?

[Linguagem & língua]

- **Linguagem**: capacidade humana de comunicação e suas manifestações, de forma verbal ou não
 - Fala, gestos, música, dança, pintura, *um sorriso*
 - Envolve nosso aparato físico e mental/cognição
- **Língua**: código de comunicação utilizado por uma comunidade, com suas regras específicas
 - Português, Inglês, LIBRAS, etc.

[PLN]

- Processamento de Língua Natural
 - Linguística Computacional
 - Processamento de Linguagem Natural
 - Engenharia das Línguas Naturais
- No Brasil, tradicionalmente visto como subárea da Inteligência Artificial & Computação
 - Habilidade linguística é um tipo de inteligência



[PLN: um pouco de história]

- Nascimento na 2ª guerra mundial
 - Tradução automática
- Possíveis nomes
 - *Computational Linguistics*
 - *Mechanolingustics*
 - *Automatic Language Data Processing*
 - *Natural Language Processing*

[PLN: um pouco de história]

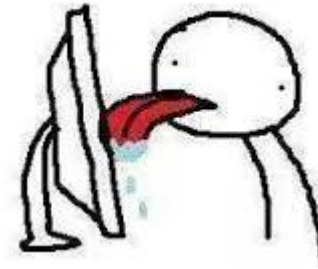
- Globalização, internet, tecnologia da informação, Google
- Mais recentemente, *web 2.0/3.0*, *redes sociais*, *smartphones*, *big data* e ciência de dados, *deep learning*

[PLN]

- Auxílio às tarefas humanas
 - (Ainda) Não substitui o humano
 - Não é possível “automatizar” toda a língua, apenas aspectos dela
 - O computador (ainda) é uma máquina estúpida!



VS.



[PLN]

- “Conversar” com uma máquina não é tão difícil

- Fazer a máquina “entender” é difícil, talvez impossível
 - Futuro distante
 - Muitas pesquisas
 - Programas especializados
 - Recursos linguísticos e linguístico-computacionais



[PLN]

- Será que é preciso ser tão fantástico para ser útil?
 - Exemplos de programas simples que são úteis?

[PLN]

- Será que é preciso ser tão fantástico para ser útil?
 - Sugestão de possíveis sinônimos
 - Revisão ortográfica e gramatical
 - Outros?
 - Simples? Ou mais claros e facilmente automatizáveis?

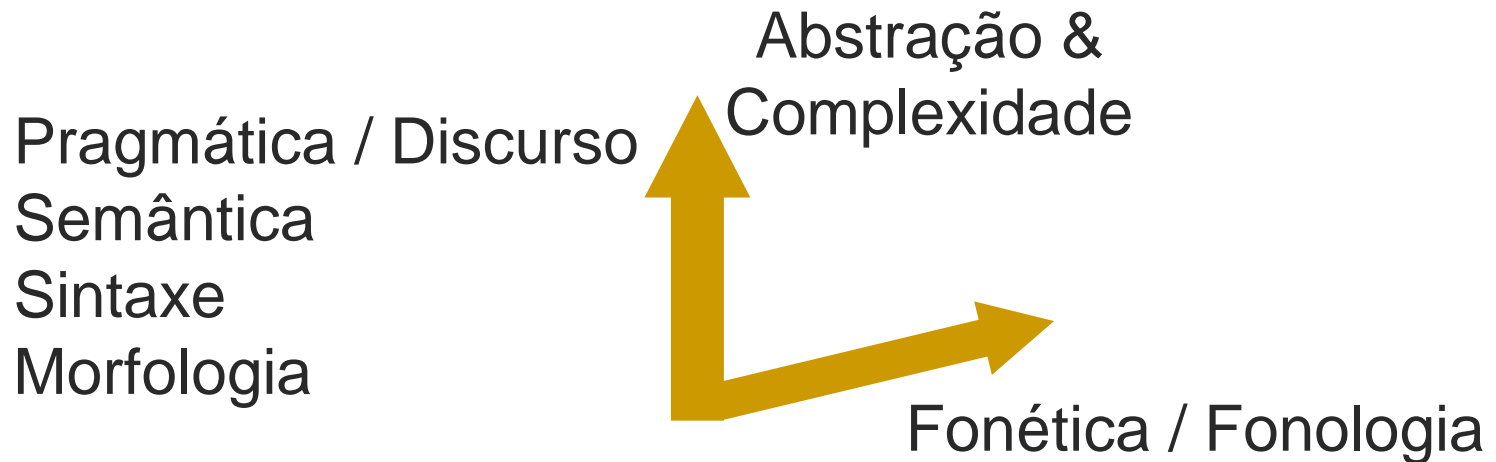
O que é necessário para aprender uma língua?

- Conhecer as palavras e como elas são formadas
- Saber o significado das palavras
- Como compor frases
- Como referenciar entidades do mundo
- Como conectar frases
- Protocolos de comunicação na língua/cultura
- Etc.

E como ensinar isso às máquinas?

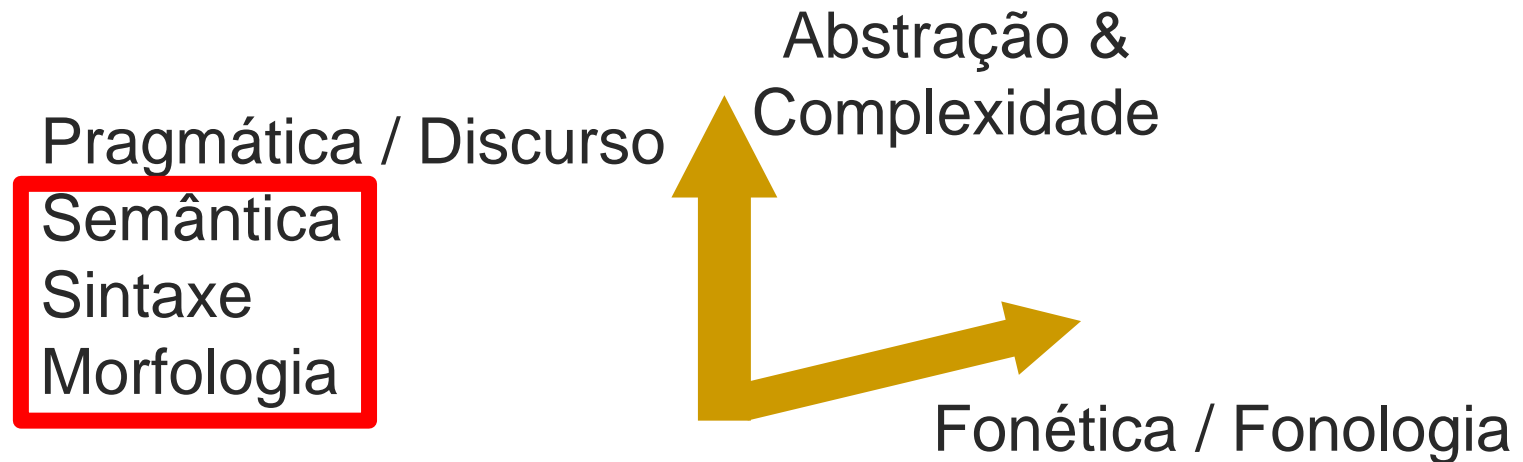
PLN: Níveis de Conhecimento

- Vários níveis de conhecimento
 - Tradicionalmente distinguidos em PLN, apesar dos limites entre eles serem nebulosos na maioria dos casos



PLN: Níveis de Conhecimento

- Vários níveis de conhecimento
 - Tradicionalmente distinguidos em PLN, apesar dos limites entre eles serem nebulosos na maioria dos casos



[PLN: Níveis de Conhecimento]

- Palavra: construção, componentes de formação
 - Morfema, raiz, afixo (prefixo, sufixo, etc.), vogal temática, desinência
 - Stemming
 - Lematização



Pragmática / Discurso

Semântica

Sintaxe

Morfologia

Fonética / Fonologia

[PLN: Níveis de Conhecimento]

- Interação entre morfologia e sintaxe: classes gramaticais ou etiquetas morfossintáticas
- Substantivo/nome, verbo, adjetivo, advérbio, pronome, preposição, conjunção, interjeição, etc.



Pragmática / Discurso

Semântica

Sintaxe

Morfologia

Fonética / Fonologia

[PLN

Ele queria jogar
tênis com Janete,
mas também queria
jantar com Suzana.
Sua indecisão o
deixou louco.

Pragmática / Discurso
Semântica

Sintaxe

Morfologia

Fonética / Fonologia

Ele/PRS#ms3

queria/QUERER/V#ii-3s

jogar/JOGAR/V#inf-nInf

tênis/? ,TÊNI/PPA#??

com/PREP

Janete/PNM

,*//PNT

mas/CJ

também/ADV

queria/QUERER/V#ii-3s

jantar/JANTAR/CN#ms

com/PREP

Suzana/PNM

.*//PNT

Sua/PNM

indecisão/INDECISÃO/CN#fs

o/DA#ms

deixou/DEIXAR/V#ppi-3s

louco/LOUCO/ADJ#ms

./PNT

[PLN: Níveis de Conhecimento]

- Como as sentenças são formadas, como as palavras podem se combinar
 - Função: sujeito, predicado, objetos, predicativos, etc.
 - Estruturação/constituição: sintagma nominal, sintagma verbal, etc.



Pragmática / Discurso

Semântica

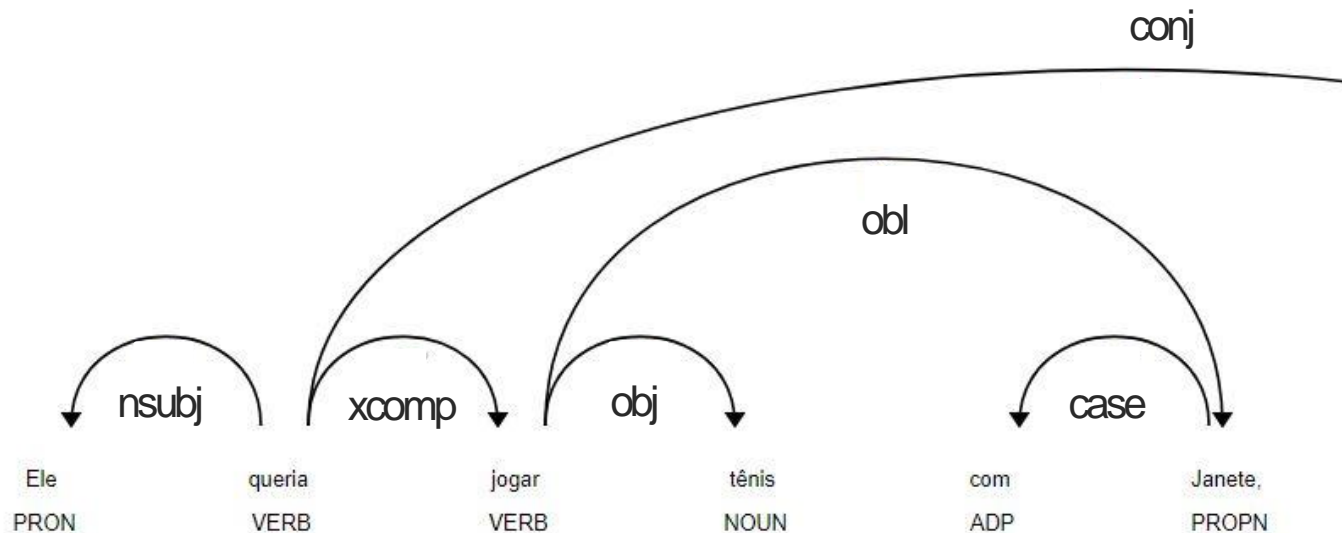
Sintaxe

Morfologia

Fonética / Fonologia

[PLN: Níveis de Conhecimento]

Ele queria jogar tênis com Janete, mas também queria jantar com Suzana. Sua indecisão o deixou louco.



Pragmática / Discurso

Semântica

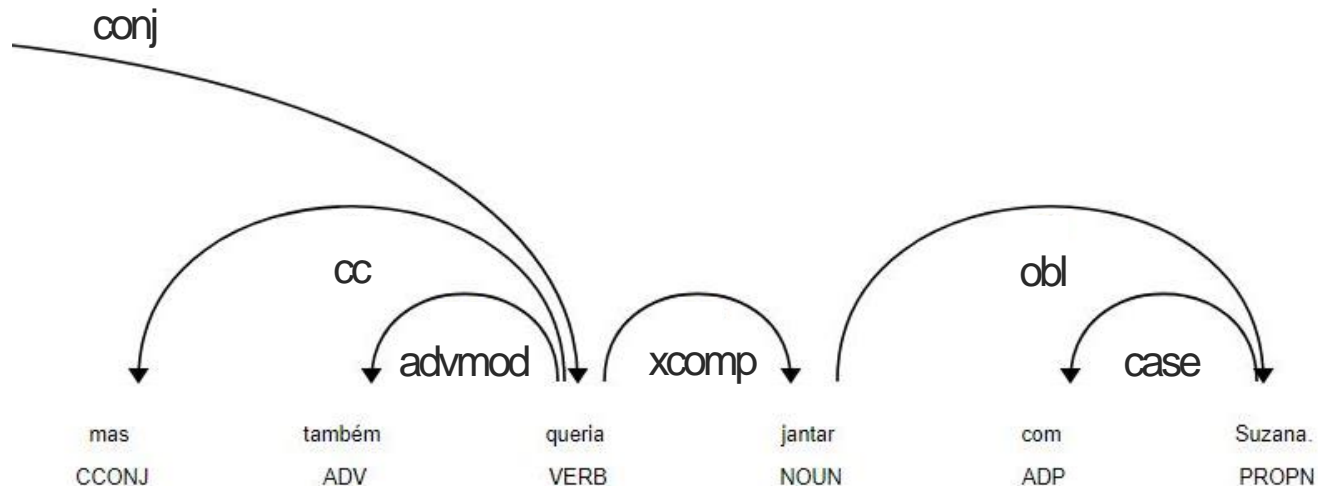
Sintaxe

Morfologia

Fonética / Fonologia

[PLN: Níveis de Conhecimento]

Ele queria jogar tênis com Janete, mas também queria jantar com Suzana. Sua indecisão o deixou louco.



Pragmática / Discurso

Semântica

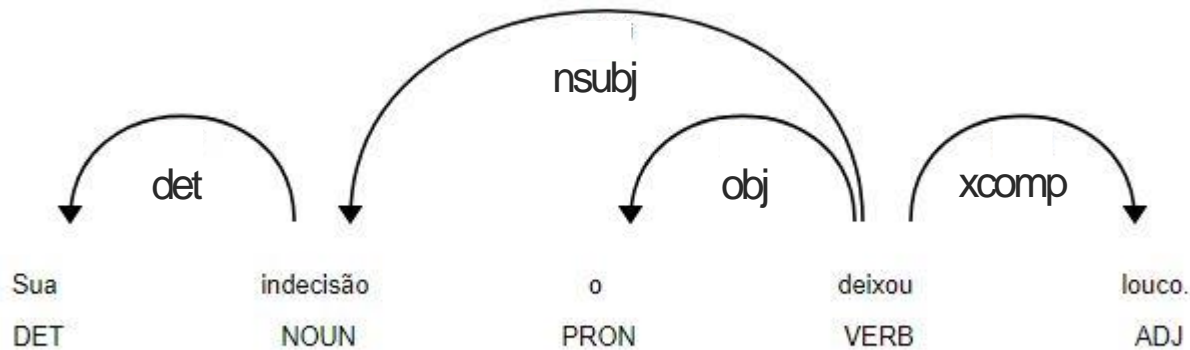
Sintaxe

Morfologia

Fonética / Fonologia

[PLN: Níveis de Conhecimento]

Ele queria jogar tênis com Janete, mas também queria jantar com Suzana. Sua indecisão o deixou louco.



Pragmática / Discurso

Semântica

Sintaxe

Morfologia

Fonética / Fonologia

[PLN: Níveis de Conhecimento]

- Significado
 - Palavras, expressões, orações, sentenças, textos
 - Metáforas, expressões idiomáticas, polissemia
 - Qual a diferença entre polissemia e homonímia?
 - Banco (assento vs. instituição financeira) é polissêmico, mas manga (camisa vs. fruta) não é



Pragmática / Discurso

Semântica

Sintaxe

Morfologia

Fonética / Fonologia

[PLN: Níveis de Conhecimento]

Ele queria jogar tênis com Janete, mas também queria jantar com Suzana. Sua indecisão o deixou louco.

“Ele”, “Janete” e “Suzana” = humanos.

Jogar tênis = praticar o esporte tênis \neq arremessar o calçado.

...



Pragmática / Discurso

Semântica

Sintaxe

Morfologia

Fonética / Fonologia

[PLN: Níveis de Conhecimento]

- Pragmática
 - Língua em uso, interação, contexto
 - Fatores como força, educação, hierarquia, crença, cooperação, atitude
- Discurso
 - Aquilo que está além da sentença



Pragmática / Discurso

Semântica

Sintaxe

Morfologia

Fonética / Fonologia

[PLN]

- Considerações para uso por um computador
 - Os níveis de conhecimento precisam ser representados (**formalizados**) e manipulados automaticamente
 - **Interação** entre os níveis
 - Morfologia e sintaxe
 - Sintaxe e semântica
 - Semântica e discurso

[PLN]

- Considerações para uso por um computador
 - Os níveis de conhecimento precisam ser representados (**formalizados**) e manipulados automaticamente
 - Interação entre **níveis mais distantes**
 - Morfologia e semântica (goleiro e porteiro vs. padeiro)
 - Morfologia e pragmática (são carlense vs. são carlino, laranjada e limonada vs. cajuada)
 - Sintaxe e discurso (subordinadas)

[PLN e humanos]

- Humanos lidam naturalmente com
 - Ambiguidade
 - Irregularidade
 - Vagueza
 - Variedade
 - Etc.
- ... máquinas (ainda) não!

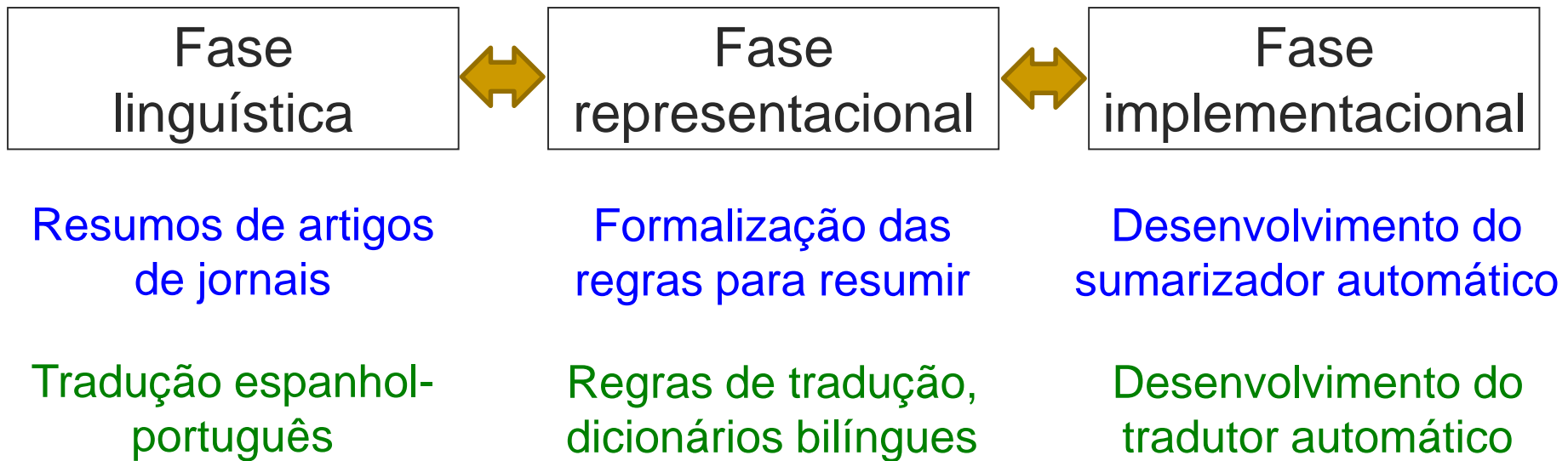
[PLN: Fases do Trabalho]

- Trabalho em PLN



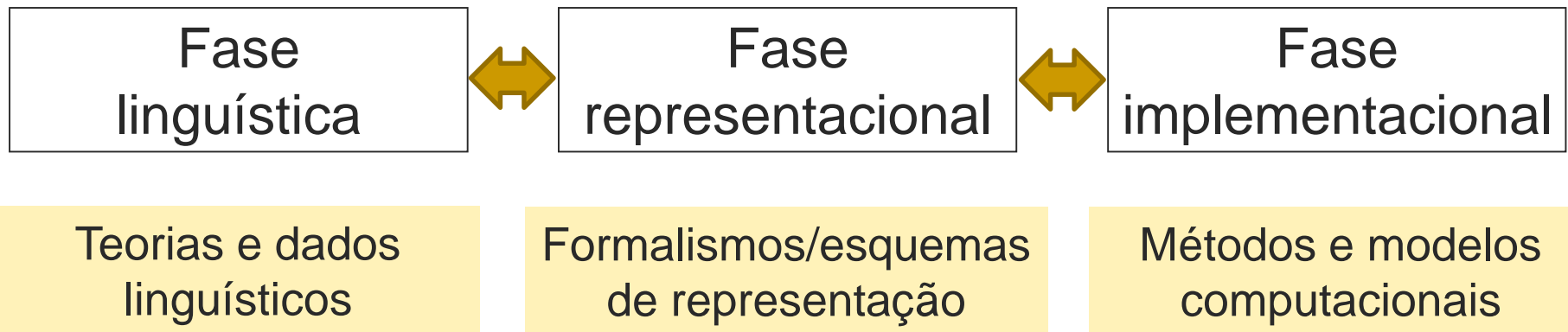
[PLN: Fases do Trabalho]

■ Trabalho em PLN



[PLN: Fases do Trabalho]

- Trabalho em PLN



- Aspectos da língua que são possíveis capturar e automatizar

[PLN]

- Classificação
 - Recursos
 - Ferramentas
 - Aplicações

[Recursos]

- **Cópus**
 - Anotação: humana e/ou automática
 - XML, XCES, TEI, etc.
 - Paralelo, comparável, alinhado, etc.

- **Dicionários monolíngues e bilíngues**
 - *Machine readable vs. machine tractable*

- **Léxicos**
 - Vários paradigmas

[Ferramentas]

- Segmentadores textuais: palavras (*tokenizador*), sentenças, parágrafos, tópicos
- Stemmers, lematizadores, nominalizadores
- Etiquetadores morfossintáticos (*taggers*)
- Analisadores sintáticos *shallow* (*chunkers*) e *deep* (*parsers*)
- Analisadores semânticos e discursivos
- Alinhadores textuais: lexicais, sentenciais, etc.
- Concordanceadores, *word counting*, ...
- Classificadores de polaridade
- Etc.

[Aplicações]

- Tradutores automáticos
- Revisores ortográficos e gramaticais
- Ferramentas de auxílio à escrita
- Sumarizadores automáticos
- Simplificadores textuais
- Minerador de opinião
- Etc.

[Recursos, ferramentas e aplicações]

■ Atenção

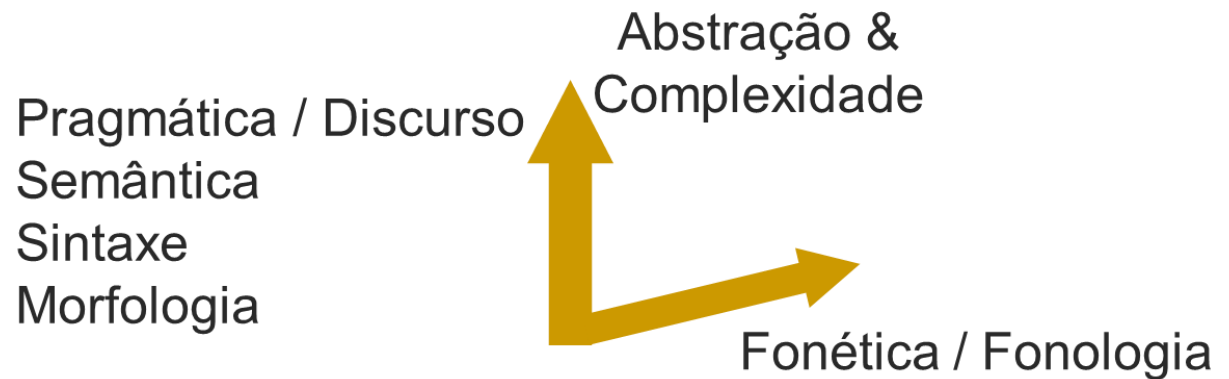
- Classificação difusa, às vezes
- Dependente do uso
 - Sumarizador como passo intermediário para recuperação da informação → ferramenta
 - Dicionário eletrônico para consulta → aplicação

[PLN e áreas correlatas]

- Limites cada vez mais suaves entre PLN e outras áreas
 - Recuperação de informação
 - Banco de dados
 - Interação humano-computador
 - Mineração de textos
 - Linguística de córpus

PLN - resumo

- Vários níveis de conhecimento



- Formalização e interação entre os níveis
- Etapas do trabalho em PLN
- Classificação

[PLN - resumo]

- Vários níveis de conhecimento

**BORA
VAMO
VAMOS PRATICAR!!**

Fonética / Fonologia

- Relação e interação entre os níveis
- Etapas do trabalho em PLN
- Classificação