

Introdução ao Processamento de Línguas Naturais

SCC5908 Introdução ao Processamento de Língua Natural

Thiago A. S. Pardo

1

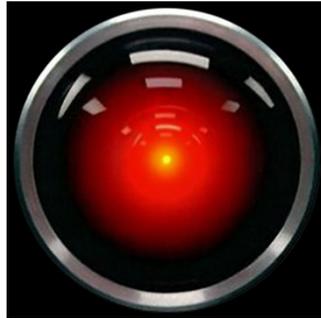
Tópicos

- Processamento de línguas naturais
 - O que é
 - Para que serve
 - Como se faz
 - Um pouco da história

2

[Meta]

- HAL 9.000 (Heuristically programmed ALgorithmic Computer)



[Outras referências]

- 2001: Uma Odisséia no Espaço
- Jornada nas Estrelas
- Guerra nas Estrelas
- IA
- Matrix
- Eu, robô
- O homem bicentenário
- Controle absoluto
- Wall-E



[Para construir um computador como HAL]

- Requer um volume enorme de conhecimento de uma dada língua
 - Reconhecimento (*faz até leitura labial*) e síntese de fala (*fonética e fonologia*)
 - Conhecimento das palavras envolvidas (*morfologia e vocabulário*)
 - Significado (*semântica*) e como combinam (*uso das palavras*)
 - Como grupos de palavras se juntam (*gramática*)
 - Manter um diálogo (*discurso*)
 - É educado responder... mesmo que você queira matar alguém (HAL)
 - É educado ser cooperativo... mesmo que esteja fingindo (HAL)
- O uso de língua natural também pressupõe **conhecimento do mundo e de senso comum**

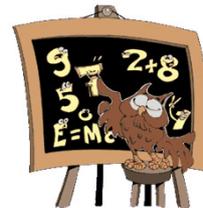
5

[Língua Natural]

- Língua humana



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



[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



7

[PLN]

- Instruir o computador a lidar com a língua
 - Entendimento
 - Análise morfológica e sintática, semântica e discursiva
 - Geração, síntese
 - Tradução, produção de resumos
 - Correção gramatical
 - Busca de respostas para perguntas
 - Recuperação de informação da Internet
 - Auxílio a escrita e ao aprendizado de línguas

- Interdisciplinar
 - Computação
 - Linguística



[PLN: um pouco de história]

- Nascimento na 2ª guerra mundial
 - Tradução automática

- Possíveis nomes
 - *Computational Linguistics*
 - *Mechan linguistics*
 - *Automatic Language Data Processing*
 - *Natural Language Processing*

9

[PLN: um pouco de história]

- Trajetória da Inteligência Artificial
 - Primeiros trabalhos → grande expectativa → resultados pobres → desilusão e hibernação da área → novos horizontes
 - Relatório da ALPAC (*Automatic Language Processing Advisory Committee*): *Languages and Machines – computers in translation and linguistics*

10

[PLN: um pouco de história]

- Dicotomia
 - Simbolismo vs. estatística (*subsimbolismo*)
- Globalização, internet, tecnologia da informação, Google

11

[PLN]

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



VS.



[PLN: Eliza]

- Programa psicólogo dos anos 60
 - Muitas versões na Internet
- Parece muito esperto, mas se perde depois de algum tempo de conversa
 - Muitas pessoas foram enganadas pelo programa!!!
- Muito simples
 - Se baseia na fala da pessoa para formular sua pergunta/resposta

13

[PLN: Eliza (Jackson e Moulinier, 2002)]

Eliza: como vai você hoje?
 P: não muito bem
 Eliza: por que você diz não muito bem?
 P: eu estou com alergia
 Eliza: interessante
 P: eu fico com manchas vermelhas
 Eliza: acho que não sei do que você está falando
 P: manchas vermelhas no meu rosto
 Eliza: seu rosto
 P: isso mesmo
 Eliza: é isso mesmo?

14

[PLN: Eliza (Jackson e Moulinier, 2002)]

P: você é um programa, não é?
 Eliza: te agrada pensar que eu sou um programa?
 P: com certeza
 Eliza: eu percebo
 P: ok
 Eliza: um pouco antes você falou de seu rosto
 P: e sobre o seu rosto?
 Eliza: o que é que você realmente quer saber?
 P: você é meu tipo de garota?

15

[PLN: Eliza (Jackson e Moulinier, 2002)]

Eliza: talvez em suas fantasias eu seja seu tipo de garota
 P: em minhas fantasias você é loira
 Eliza: o que te faz pensar que sou loira?
 P: não sei
 Eliza: você realmente não sabe?
 P: não
 Eliza: você está dizendo não apenas para ser negativo?
 P: não
 Eliza: você está sendo um pouco negativo
 P: não, não estou
 Eliza: por quanto tempo você não tem estado?

16



Ed - CONPET - Converse comigo ! - Microsoft Internet Explorer

Arquivo Editar Exibir Favoritos Ferramentas Ajuda

Endereço <http://www.inbot.com.br/ed/popup.htm> Ir

Ed: Bem vindo ao site do CONPET! Eu sou Ed, o robô. Vim de outro planeta com a missão de ajudar na preservação de energia e de outros recursos naturais. Sobre o que você quer conversar?

Você: Você é um robô?

Ed: Claro que sou um robô. Só agora que você descobriu???

Você: Do que é feito?

Ed: Já digo já. Que tal você contar algo primeiro?

Você: Não sou um robô.

Ed: Não mesmo? Perdão. Pensei que fosse. Às vezes é difícil distinguir quem é de quem não é.

Você:

Internet

Robô Ed: conhecimento especializado

[PLN]

- “Conversar” com uma máquina não é tão difícil
- Fazer a máquina “entender” é difícil, talvez impossível
 - Futuro distante
 - Muitas pesquisas
 - Programas especializados
 - Recursos lingüísticos e lingüístico-computacionais



[PLN]

- Será que é preciso entender de fato a língua para fazer coisas fantásticas?

19

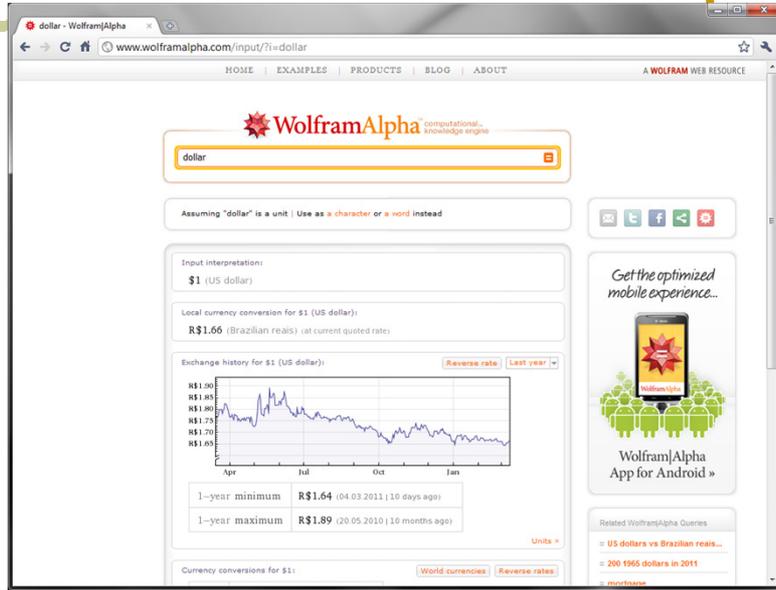
[WolframAlpha]

The screenshot displays the WolframAlpha website interface. At the top, the search bar contains the text "São Carlos". Below the search bar, the results are organized into several sections:

- Input Interpretation:** São Carlos
- Population:** 271 758 people (0.14% of total for Brazil)
- Area:** 3190 km² (square kilometers) (0.037% of total for Brazil)
- Capital city:** São Carlos

Additional elements include a "Give us your feedback" form, a "New to Wolfram|Alpha? TAKE THE TOUR" button, and a "Get the optimized mobile experience..." advertisement for the Wolfram|Alpha App for Android. The footer contains navigation links and copyright information: © 2011 Wolfram Alpha LLC—A Wolfram Research Company | Terms | Privacy | Entry Index.

WolframAlpha



Qwiki



[Qwiki]

- Sumariza informação da web
 - Não há armazenamento de informação
- Sintetiza voz
- Busca informação relacionada
 - “Experiência de informação”
- Inspiração no computador de bordo de Wall-E
 - Vídeo

23

[Watson (IBM)]

- Venceu os melhores participantes humanos no show de perguntas e respostas Jeopardy!
 - *“more than 100 different techniques are used to analyze natural language, identify sources, find and generate hypotheses, find and score evidence, and merge and rank hypotheses”*
 - *“sources of information include encyclopedias, dictionaries, thesauri, newswire articles, and literary works. Watson also used databases, taxonomies, and ontologies. Specifically, DBPedia, WordNet, and Yago were used”*



[PLN]

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

25

[PLN]

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

26

[PLN]

- Será que é preciso ser tão fantástico para ser útil?
 - Comando **wc do UNIX** é um programa de PLN? Usa conhecimento lingüístico?
 - Definição
 - *Short for "word count", wc displays the total number of bytes, words and lines in a text file*
 - Exemplo (em linha de comando/DOS)

```
C:\> wc meu_arquivo.txt
5 (linhas) 13 (palavras) 57 (bytes/caracteres)
```

27

[PLN]

- De que um computador necessita para ser capaz de entender uma fala humana e interagir adequadamente?
 - Como nós, humanos, fazemos isso?

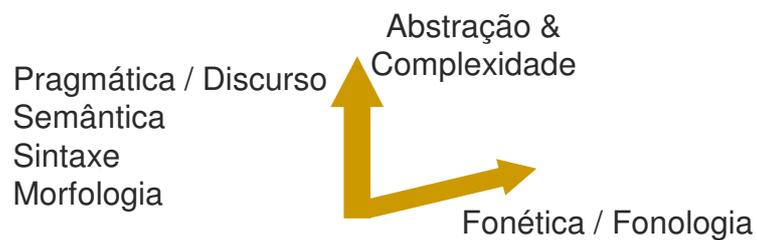


Quem é Lula?
 Que preguiça!
 Está calor aqui.
 linguista: O que Chomsky disse?
 Informata: O que Chomsky disse?
 Quem é Lula? Sei que não é o molusco.

28

[PLN]

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



29

[PLN]

- Fonética: estuda como os humanos produzem, transmitem e recebem sons, independente de língua; sistema físico
- Fonologia: estudo dos sons em uma língua específica, como os sons são construídos
 - Fones, fonemas, local (bilabial, palatal, etc.) e modo de articulação (pausa, nasal, fricativo, etc.), etc.



[PLN]

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

↕

Transcrição fonética

↑

- Pragmática / Discurso
- Semântica
- Sintaxe
- Morfologia
- Fonética / Fonologia**

IPA	ASCII	examples
ʌ	^	cup, luck
ɑ:	a:	arm, father
æ	@	cat, black
ə	..	away, cinema
e	e	met, bed
ɜ:ʔ	e:(t)	turn, learn
ɪ	i	hit, sitting
i:	i:	see, heat
ɒ	o	hot, rock
ɔ:	o:	call, four
ʊ	u	put, could
u:	u:	blue, food
aɪ	ai	five, eye
aʊ	au	now, out
oʊ/əʊ	Ou	go, home
eəʔ	e..(r)	where, air
eɪ	ei	say, eight
ɪəʔ	i..(r)	near, here
ɔɪ	oi	boy, join
ʊəʔ	u..(r)	pure, tourist

[PLN]

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

↕

Transcrição fonética

↑

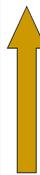
- Pragmática / Discurso
- Semântica
- Sintaxe
- Morfologia
- Fonética / Fonologia**

IPA	ASCII	examples
ʌ	^	cup, luck
ɑ:	a:	arm, father
æ	@	cat, black
ə	..	away, cinema
e	e	met, bed
ɜ:ʔ	e:(t)	turn, learn
ɪ	i	hit, sitting
i:	i:	see, heat
ɒ	o	hot, rock
ɔ:	o:	call, four
ʊ	u	put, could
u:	u:	blue, food
eəʔ	e..(r)	where, air
eɪ	ei	say, eight
ɪəʔ	i..(r)	near, here
ɔɪ	oi	boy, join
ʊəʔ	u..(r)	pure, tourist

Reconhecimento de voz
Síntese de voz

[PLN]

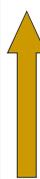
- Palavra: construção, componentes de formação
 - Morfema, grafema, raiz, afixo (prefixo, sufixo, etc.), vogal temática, desinência


 Pragmática / Discurso
 Semântica
 Sintaxe
Morfologia
 Fonética / Fonologia

33

[PLN]

- 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

34

7

[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 [ele] **PERS** M 3S NOM
queria [querer] <fmc> **V** IMPF 3S IND VFIN
jogar [jogar] **V** INF
tênis [tênis] **N** M S/P
com [com] **PRP**
Janete [Janete] **PROP** M/F S
 ,
mas "mas" <co-vfin> <co-fmc> **KC**
também [também] **ADV**
queria [querer] <fmc> **V** IMPF 3S IND VFIN
jantar [jantar] **V** INF
com [com] **PRP**
Suzana [Suzana] **PROP** F S
 .
Sua [seu] <poss 3S> **DET** F S
indecisão [indecisão] **N** F S
o [ele] **PERS** M 3S ACC
deixou [deixar] <fmc> **V** PS 3S IND VFIN
louco [louco] **ADJ** M S
 .

7

[PLN

Ele queria jogar tênis com Janete, mas também queria jantar com Suzana.

Pragmática / Discurso
Semântica
Sintaxe
Morfologia
Fonética / Fonologia

Ele [ele] **PERS** M 3S NOM
queria [querer] <fmc> **V** IMPF 3S IND VFIN
jogar [jogar] **V** INF
tênis [tênis] **N** M S/P
com [com] **PRP**
Janete [Janete] **PROP** M/F S
 ,
mas "mas" <co-vfin> <co-fmc> **KC**
também [também] **ADV**
queria [querer] <fmc> **V** IMPF 3S IND VFIN
jantar [jantar] **V** INF
com [com] **PRP**
Suzana [Suzana] **PROP** F S
 .
Sua [seu] <poss 3S> **DET** F S
indecisão [indecisão] **N** F S
o [ele] **PERS** M 3S ACC
deixou [deixar] <fmc> **V** PS 3S IND VFIN
louco [louco] **ADJ** M S
 .

Analísadores sintáticos
Corretores gramaticais

[PLN]

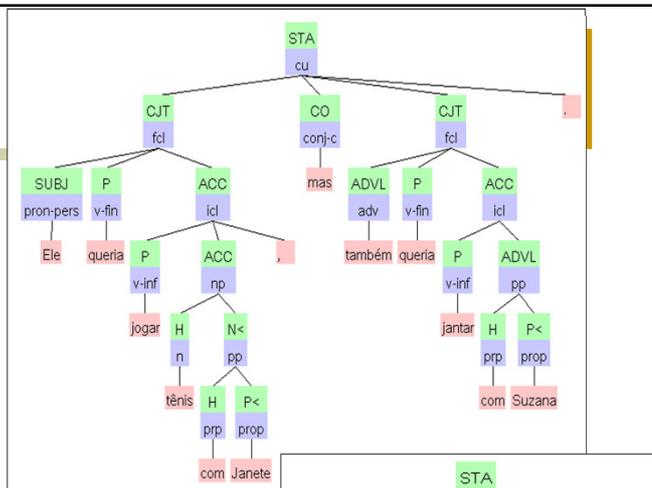
- Como as sentenças são formadas, como as palavras podem se combinar
 - Função: sujeito, predicado, objetos, predicativos, etc.
 - Estruturação: sintagma nominal, sintagma verbal, etc.

↑ Pragmática / Discurso
 Semântica
Sintaxe
 Morfologia
 Fonética / Fonologia

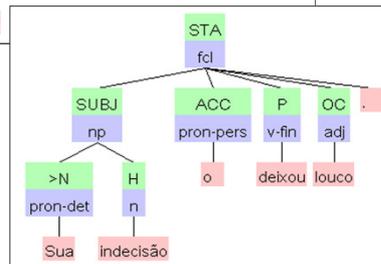
37

[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



[PLN]

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

Análise sintática
Corretores Gramaticais
Sumarizadores
Tradução Automática

↑ Pragmática / Discurso
Semântica
Sintaxe
Morfologia
Fonética / Fonologia

[PLN]

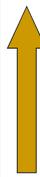
- Como as sentenças são formadas, como as palavras podem se combinar
 - E além da sentença? Há sintaxe?

↑ Pragmática / Discurso
Semântica
Sintaxe
Morfologia
Fonética / Fonologia

40

[PLN]

- Significado
 - Palavras, expressões, orações, sentenças, textos
 - Lexical, composicional, textual

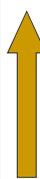

 Pragmática / Discurso
Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia

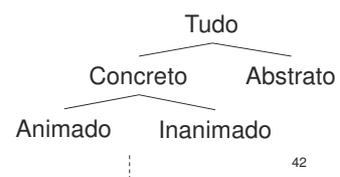
41

[PLN]

- Traços semânticos, classificações ontológicas

	Mesa	Cavalo	Garota	Mulher
Animado	-	+	+	+
Humano	-	-	+	+
Fêmea	-	-	+	+
Adulto	-	+	-	+

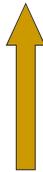

 Pragmática / Discurso
Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia



42

[PLN]

- Papéis semânticos/temáticos
 - Agente, tema, instrumento, experienciador, fonte, etc.
 - [O menino]_{AGENTE} chutou [a bola]_{TEMA}


 Pragmática / Discurso
Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia

43

[PLN]

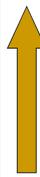
- Classes/categorias/tipos semânticos
 - Humano, local, data, organização, etc.
 - O [menino]_{HUMANO} chutou a bola
 - Entidades nomeadas e mencionadas


 Pragmática / Discurso
Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia

44

[PLN]

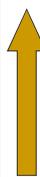
- Relações “lexicais”
 - Sinonímia, antonímia, hiperonímia/hiponímia, meronímia/holonímia, etc.


 Pragmática / Discurso
Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia

45

[PLN]

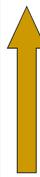
- Diversos fenômenos
 - Metáforas, expressões idiomáticas, polissemia
 - Qual a diferença entre polissemia e homonímia?
 - Banco (assento vs. instituição financeira) é polissêmico, mas manga (camisa vs. fruta) não é


 Pragmática / Discurso
Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia

46

[PLN]

- Semântica composicional
 - Análise de seqüência de palavras, sentenças


 Pragmática / Discurso
Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia

47

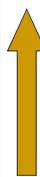
[PLN]

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 ≠ arremessar o calçado.

...


 Pragmática / Discurso
Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia

48

[PLN]

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

queria(exper(ele),objetivo(jogar(tênis),comutativo(Janete)))...

Pragmática / Discurso

Semântica

Sintaxe

Morfologia

Fonética / Fonologia

49

[PLN]

contraste

[Ele queria jogar tênis com Janete], [mas também queria jantar com Suzana].

causa

[Sua indecisão o deixou louco].

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

Pragmática / Discurso

Semântica

Sintaxe

Morfologia

Fonética / Fonologia

50

[PLN]

contraste

[Ele queria jogar tênis com Janete], [mas também queria jantar com Suzana].

causa

[Sua indecisão o deixou louco.]

Analísadores semânticos
Corretores gramaticais
Sumarizadores
Tradução automática

Ele qu...
mas também...
Suzana. Sua indecisão o deixou louco.

- Pragmática / Discurso**
- Semântica**
- Sintaxe
- Morfologia
- Fonética / Fonologia

51

[PLN]

- Discurso
 - Aquilo que está além da sentença
 - Semântica textual
 - Relacionamento proposicional, correferência e expressões referenciais, marcadores textuais, intenções, etc.

- Pragmática / Discurso**
- Semântica
- Sintaxe
- Morfologia
- Fonética / Fonologia

52

[PLN]

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

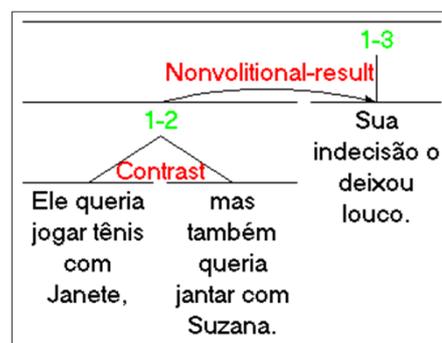
Pragmática / Discurso

Semântica
Sintaxe
Morfologia
Fonética / Fonologia

53

[PLN]

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



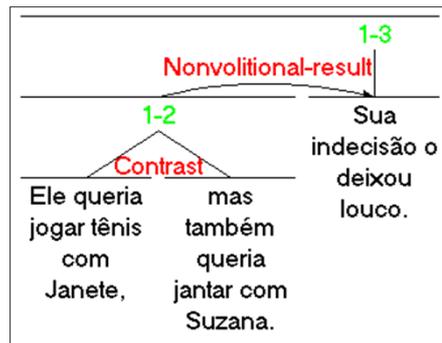
Pragmática / Discurso

Semântica
Sintaxe
Morfologia
Fonética / Fonologia

[PLN]

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

(Intend E (Believe L “o desejo de fazer duas coisas incompatíveis o deixou louco”))

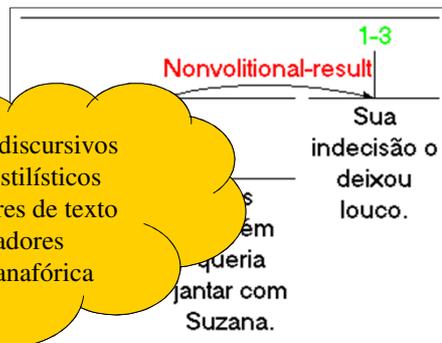


- Pragmática / Discurso**
- Semântica
 - Sintaxe
 - Morfologia
 - Fonética / Fonologia

[PLN]

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

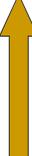
(Intend E (Believe L “o desejo de fazer duas coisas incompatíveis o deixou louco”))



- Pragmática / Discurso**
- Semântica
 - Sintaxe
 - Morfologia
 - Fonética / Fonologia

[PLN]

- Pragmática
 - Língua em uso, interação, contexto
 - Fatores como força, educação, hierarquia, crença, cooperação, atitude
 - Estilos de escrita e de fala
 - Suposições sobre produtor e receptor, nível de conhecimento, interesses
 - Modelagem do usuário

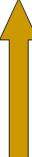


Pragmática / Discurso
 Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia

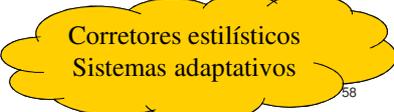
57

[PLN]

- Pragmática
 - Língua em uso, interação, contexto
 - Fatores como força, educação, hierarquia, crença, cooperação, atitude
 - Estilos de escrita e de fala
 - Suposições sobre produtor e receptor, nível de conhecimento, interesses
 - Modelagem do usuário



Pragmática / Discurso
 Semântica
 Sintaxe
 Morfologia
 Fonética / Fonologia



Corretores estilísticos
 Sistemas adaptativos

58

[Terminologia]

- **Cuidado** com a terminologia
 - Palavra vs. conceito
 - Palavras funcionais/gramaticais vs. palavras de conteúdo
 - *Tokens* vs. *types*
 - Hapax legomena: palavras que ocorrem uma única vez
 - *Honorificabilitudinitatibus*, de Shakespeare
 - Segmento textual, frase, sentença, etc. vs. proposição
 - Marcador textual vs. marcador discursivo

59

[Terminologia]

- **Cuidado** com a terminologia
 - Palavra vs. lexema vs. item lexical?

60

[Terminologia]

- **Cuidado** com a terminologia
 - **Palavra vs. lexema vs. item lexical?**
 - Perini (1998)
 - Lexema: forma básica da palavra (*pedra*) – léxico mental
 - Palavra: variações do lexema (*pedra* e *pedras*)
 - Item lexical: lexema + informações léxicas, como pronúncia, flexões, variações, propriedades sintáticas, significado, etc.
 - Outra visão, mais simples, bastante difundida em PLN
 - Lexema = item lexical: o que está no léxico
 - Palavra: variações do lexema, lexema no texto/discurso

61