

Workshop de Tecnologias Adaptativas – WTA 2015

Semântica no Reconhecedor Gramatical Linguístico

Ana Contier, Djalma Padovani, João José Neto

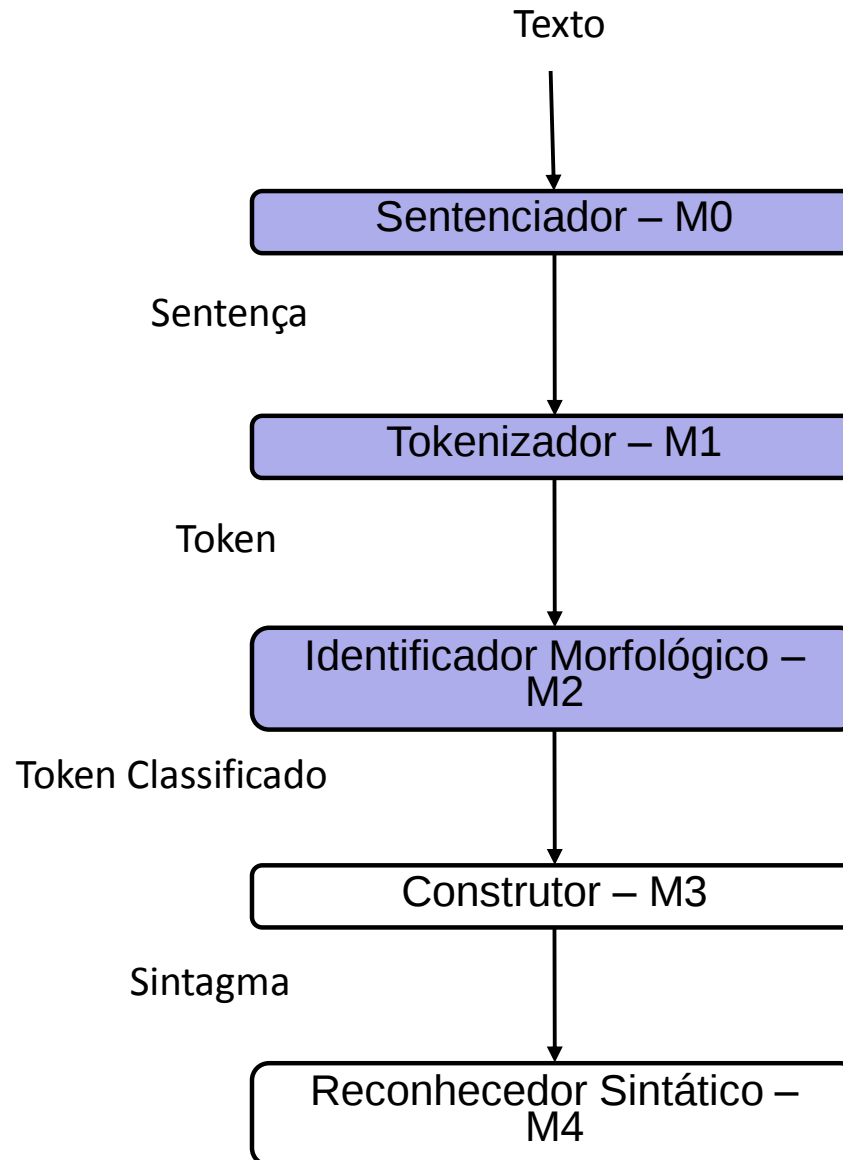
Linguagem Natural - Desafios

- Crescente quantidade de documentos disponíveis apenas em Linguagem Natural
- Recuperação das informações relevantes deste grande volume é difícil e custosa
- É necessário determinar e interpretar a estrutura das sentenças em muitos níveis de detalhes
- As linguagens naturais exibem um intrincado comportamento estrutural visto que são profusos os casos particulares a serem considerados

Linguagem Natural – Aplicações

- Geração automática de resumos
- Representação de conhecimento
- Geração de taxonomias hierárquicas
- Mecanismos de buscas
- Tradutores, revisores e corretores gramaticais
- Extratores de informação

O Reconhecedor Gramatical Linguístico



Módulo M0 - Sentenciador

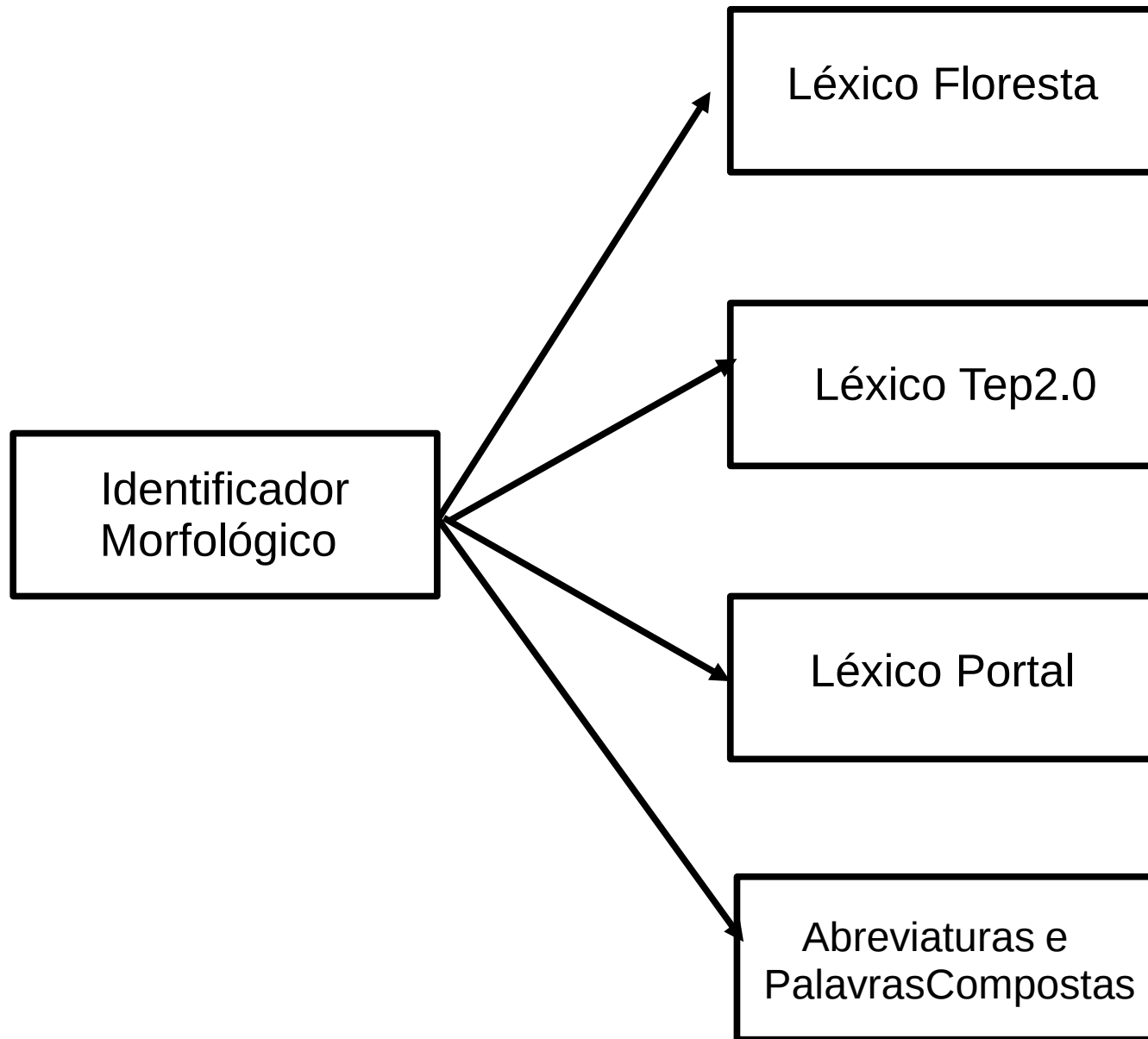
- Quebra do texto em linhas
- Limpeza de caracteres indesejados
- Identificação de caracteres de finais de sentença
- Identificação de abreviaturas
- Separação de sentenças

Módulo M1 - Tokenizador

- Identificação de palavras compostas
- Identificação de abreviaturas
- Identificação de números, caracteres monetários, hora e percentuais
- Separação de palavras

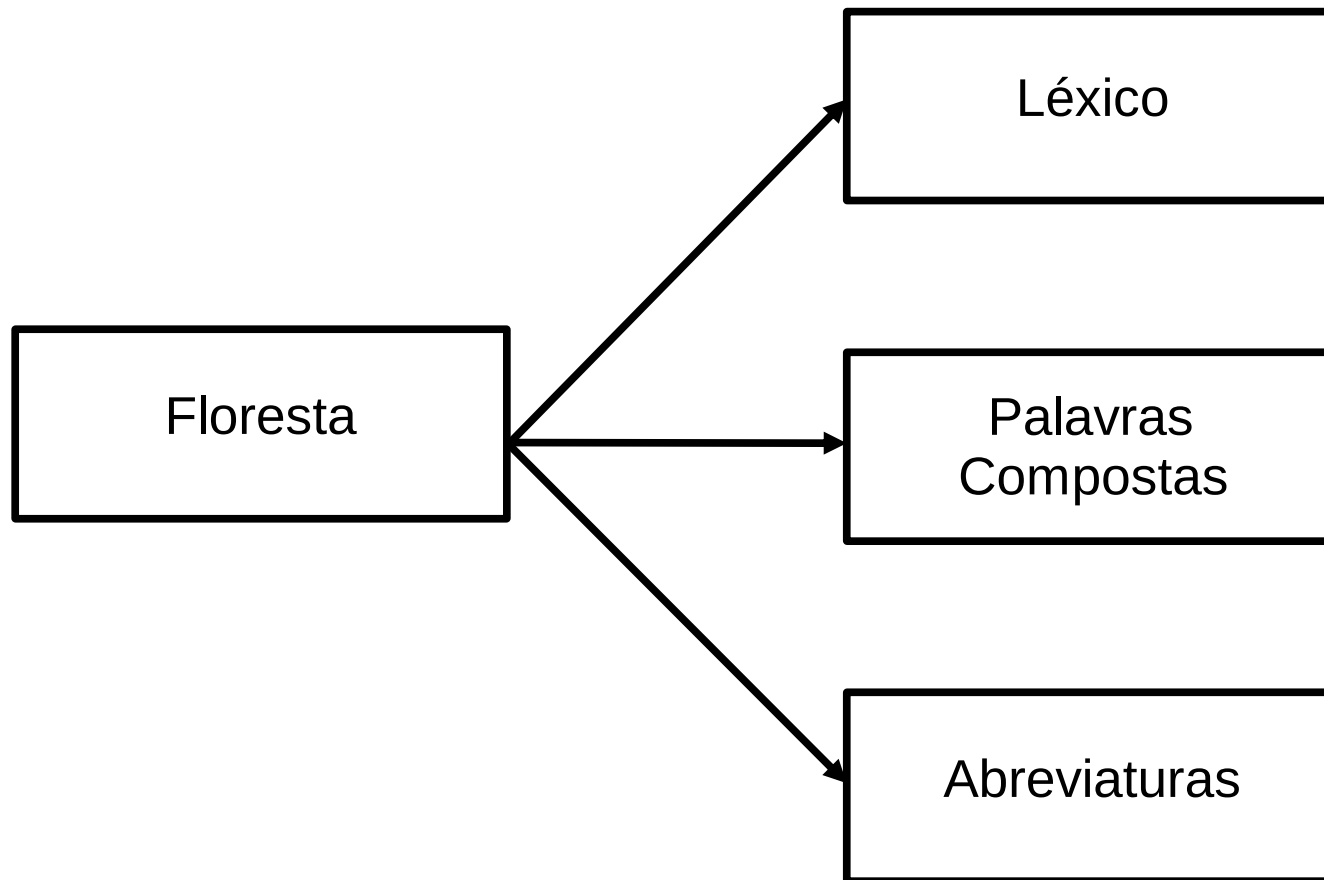
Módulo M2 – Identificador Morfológico

Sequência de Consultas



Arquivos de Apoio - Floresta

Desdobramento do Arquivo



Arquivos de Apoio - Floresta

Exemplo de Frase Etiquetada do Corpus Floresta

#9363 CF997-3 Segundo declarações do próprio diretor, ele vive até hoje de forma angustiada:
[STA+fcl (ADVl+pp

```
    (H+prp Segundo)
    (P<+np
    (H+n declarações)
    (N<+pp
    (H+prp de)
    (P<+np
    (>N+art o)
    (>N+pron-det próprio)
    (H+n diretor)))]
[.]
(SUBJ+pron-pers ele)
(P+v-fin vive)
(ADVl+advp
 (>A+prp até)
 (H+adv hoje))
(ADVl+pp
 (H+prp de)
 (P<+np
  (H+n forma)
  (N<+v-pcp angustiada)))]
[:]
```

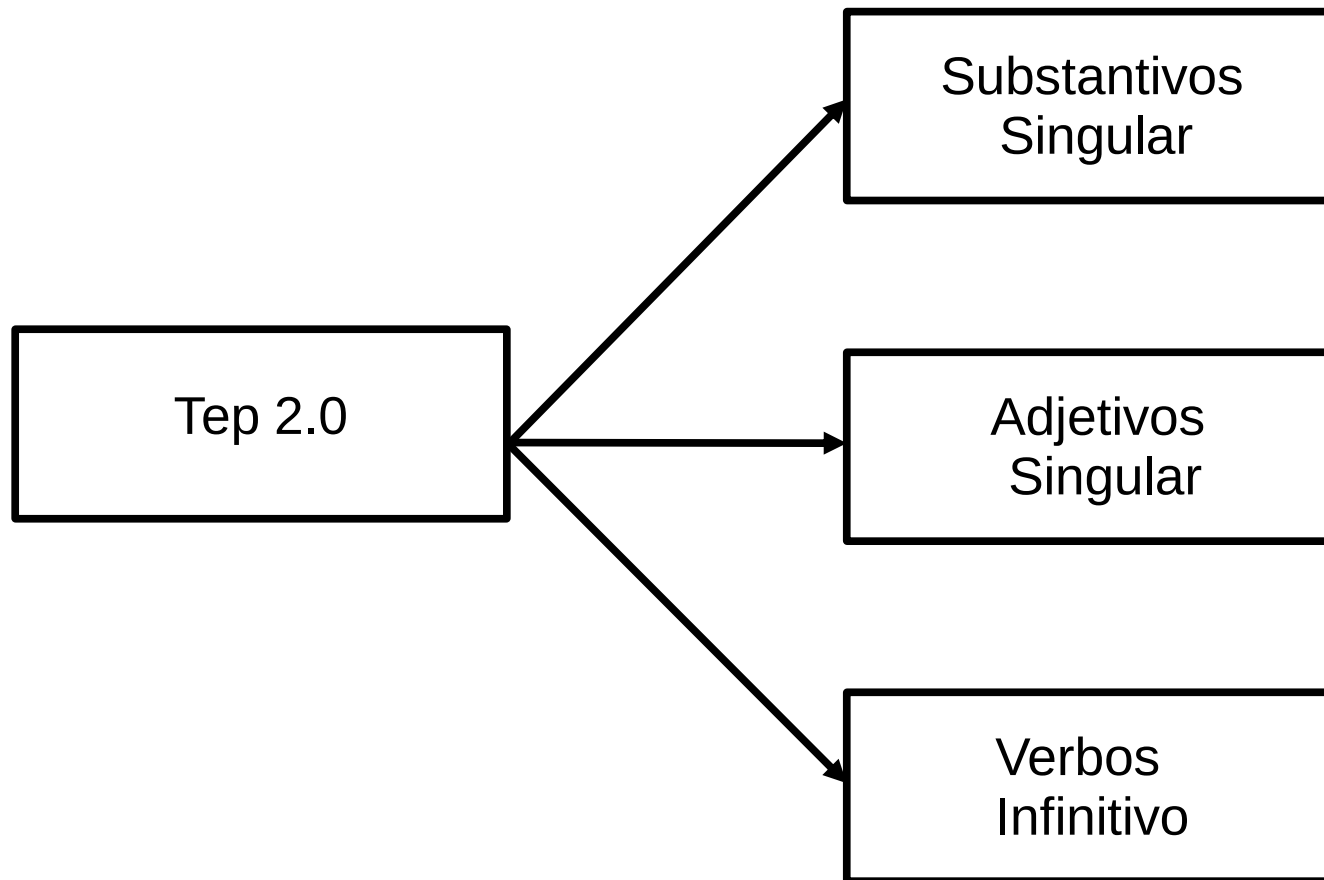
Arquivos de Apoio - Floresta

Fragmento do Léxico

Hill	CJT+prop			
Hillary	SUBJ+prop		CJT+prop	
Hillary Clinton		CJT+prop	P<+prop	
Hilma	H+prop			
Hilton	H+prop			
Hiléia	H+prop			
Hirano Design International Inc.			H+prop	
Hirohisa Fujii		N<+prop		
História	H+prop	P<+prop		
Hitachi	H+prop			
Hitler	CJT+prop			
Hizbollah	H+prop			
Hocine	N<+prop			
Hockenheim		P<+prop		
Hofheim	CJT+prop			
Hoje	ADVL+adv		CJT+adv	>N+adv H+adv
Hoje de manhã		SUBJ+adv		
Hoje em dia		ADVL+adv		
Hoje mesmo		ADVL+adv		

Arquivos de Apoio - Tep2.0

Desdobramento do Arquivo



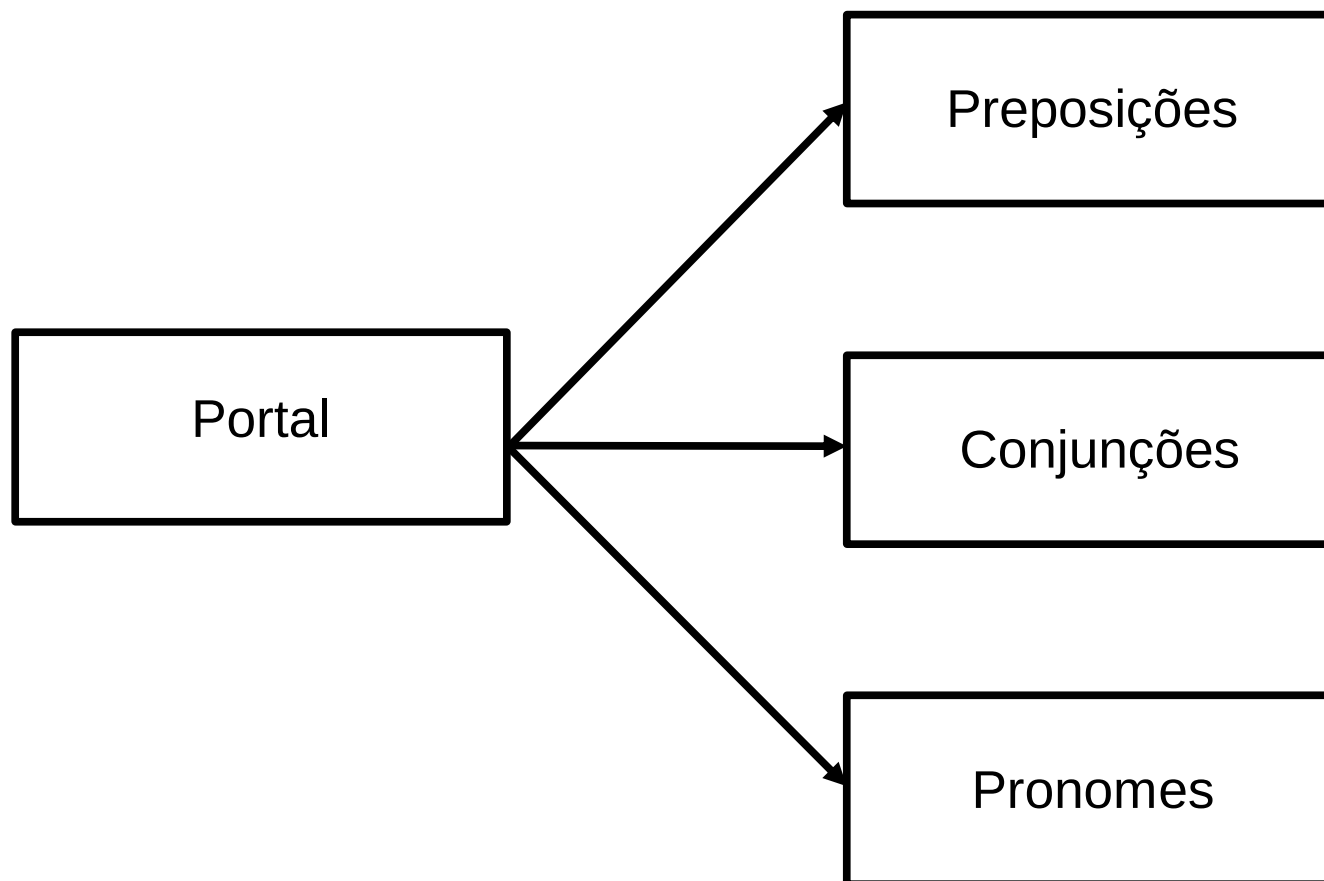
Arquivos de Apoio - TEP2.0

Fragmento do Léxico

1. [Verbo] {exagerar, exceder, quinta-essenciar, rebuscar, refinar, requintar}
2. [Verbo] {ababalhar, babar, escumar, espumar}
3. [Verbo] {ababalhar, sujar}
4. [Verbo] {aclara, alumiar, clarear, iluminar}
5. [Verbo] {enegrecer, escurecer} <4>
6. [Verbo] {limpar, purgar, purificar}
7. [Verbo] {impurificar, sujar} <6>
8. [Verbo] {apurar, clarificar, limpar, purgar, purificar}
9. [Verbo] {purgar} <7>
10. [Verbo] {coar, destilar, insinuar-se}
11. [Verbo] {animar} <1746>
12. [Verbo] {açalmar, fornecer-se, prover-se}
13. [Verbo] {acamaradar, amigar, amistar}
14. [Verbo] {embarricar, embarrilar}
15. [Verbo] {bastar, satisfazer}
16. [Verbo] {depravar-se}
17. [Verbo] {armar-se, munir-se}
18. [Verbo] {abceder, absceder, aporismar, supurar}
19. [Verbo] {abecar, abotoar, abufelar}
20. [Verbo] {abençoar-se, benzer-se}
21. [Verbo] {conformar-se, sujeitar-se}
22. [Verbo] {abater-se, acamar-se}
23. [Verbo] {abater, acamar}

Arquivos de Apoio – Portal São Francisco

Desdobramento do Arquivo



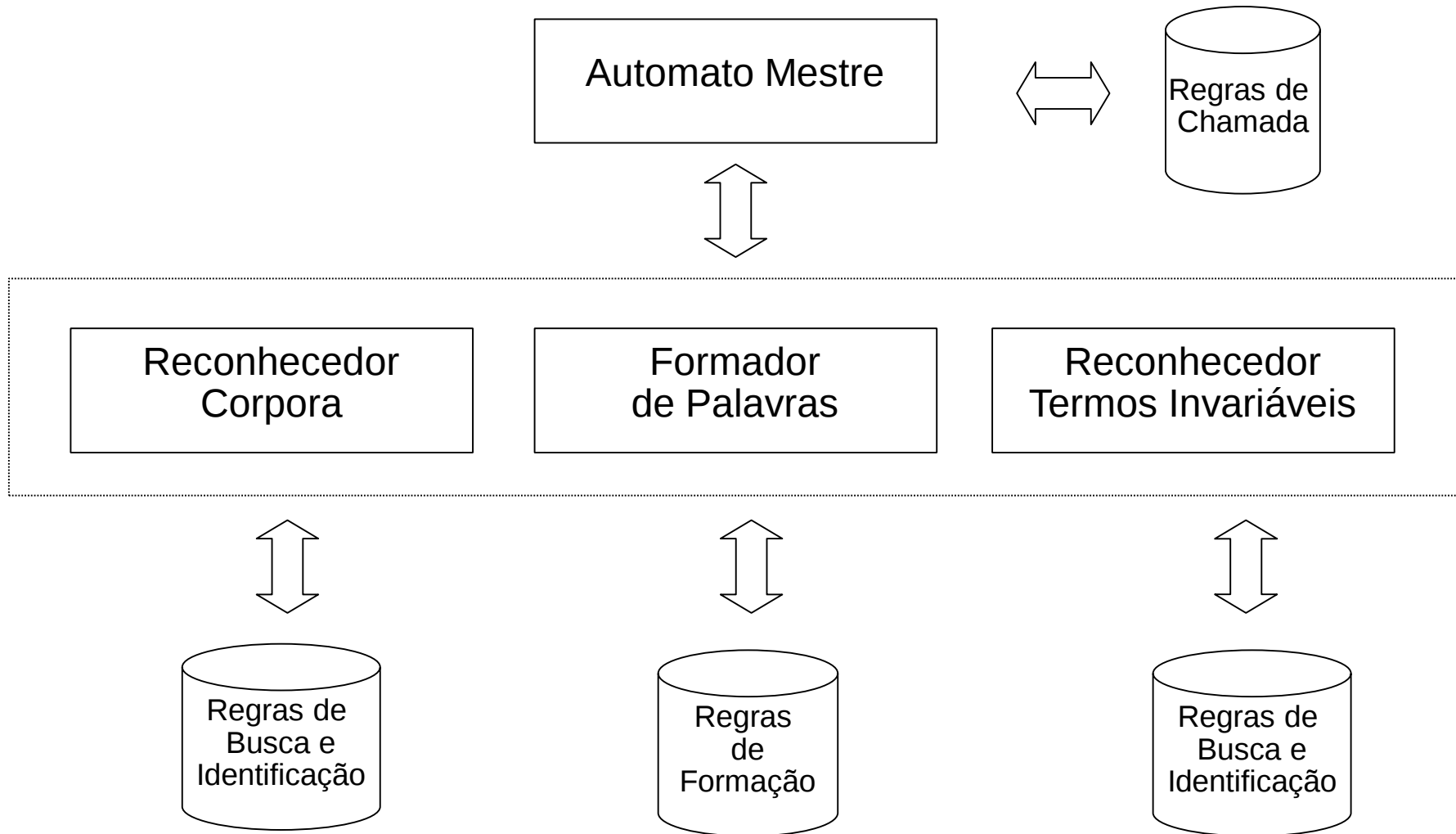
Arquivos de Apoio – Portal São Francisco

Fragmento do Léxico

a não ser que	conj-s	condicional
ademaís conj-c	explicativa	
ainda que conj-s	concessiva	
antes que conj-s	temporal	
ao passo que	conj-s	final
apesar de que	conj-s	concessiva
assim conj-c	conclusiva	
assim como	conj-s	comparativa
assim que	conj-s	temporal
até que conj-s	temporal	
bem como	conj-s	comparativa
bem como	conj-c	aditiva
bem que conj-s	concessiva	

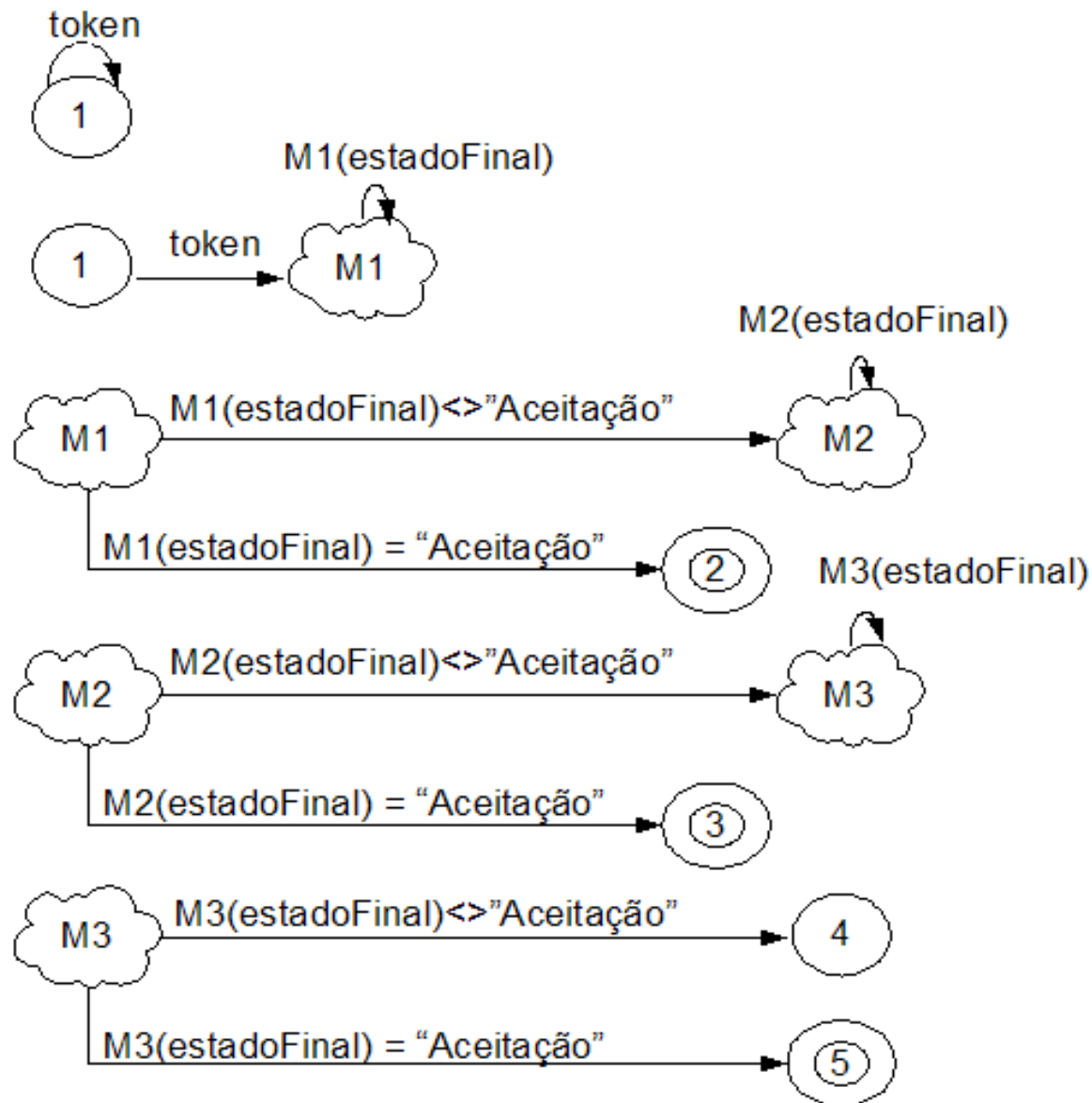
Módulo M2 – Identificador Morfológico

Estrutura do Autômato Adaptativo



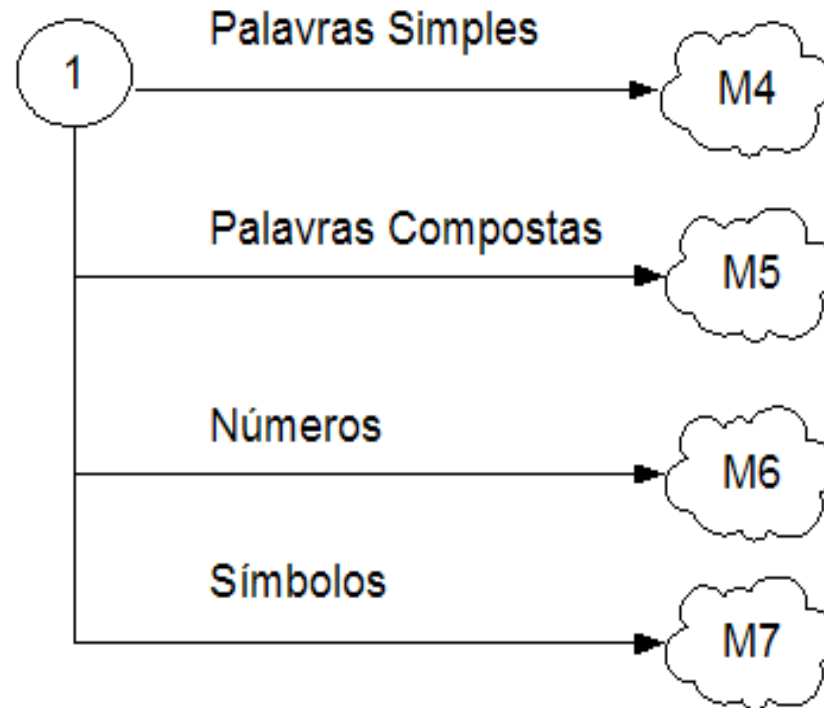
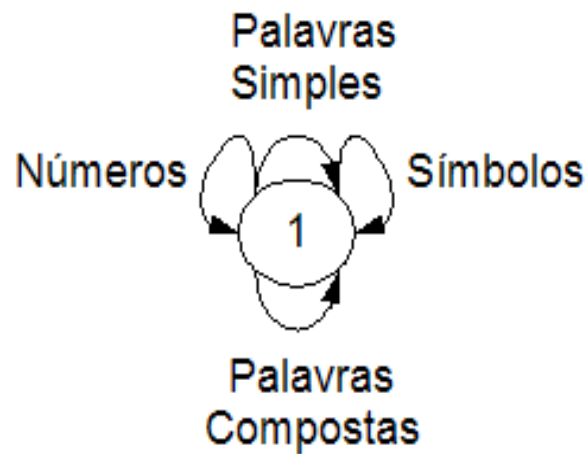
Módulo M2 – Identificador Morfológico

Estrutura do Autômato Adaptativo



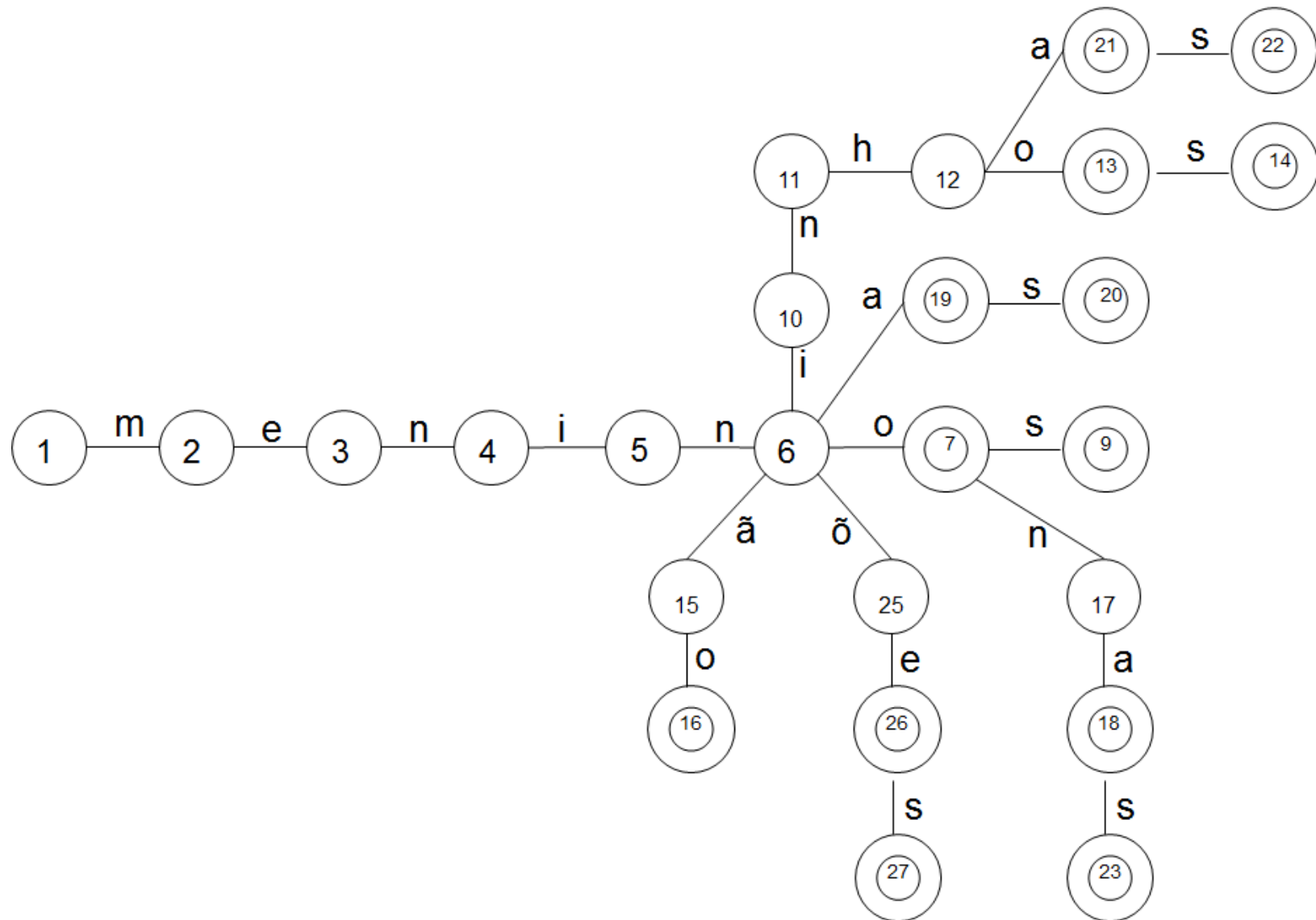
Módulo M2 – Identificador Morfológico

Estrutura do Autômato Adaptativo



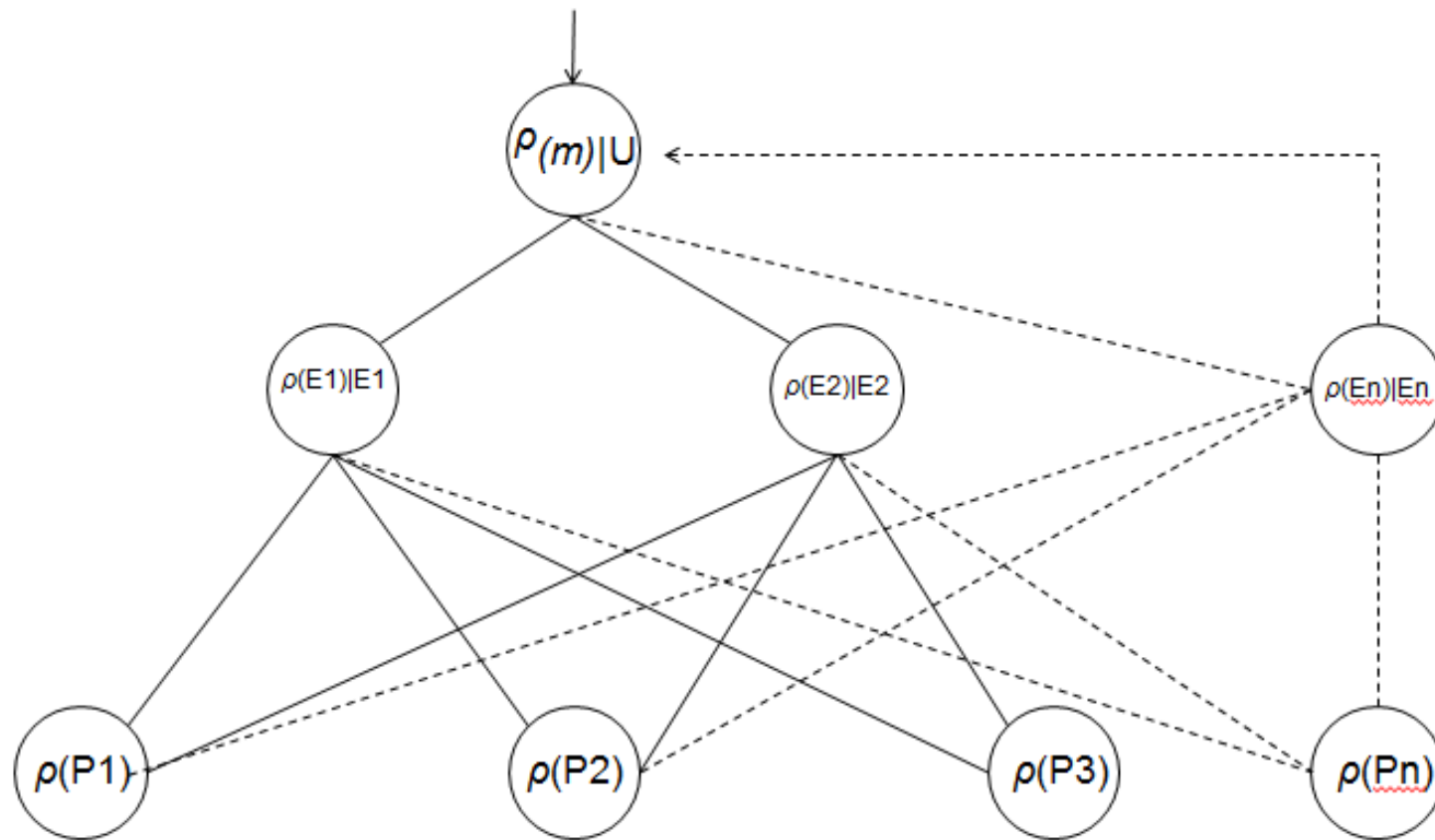
Módulo M2 – Identificador Morfológico

Estrutura do Autômato Adaptativo



Módulo M2 – Identificador Morfológico

Desambiguação Morfológica



$\rho(m)|U$ – Máxima probabilidade | 2 Últimas Etiquetas

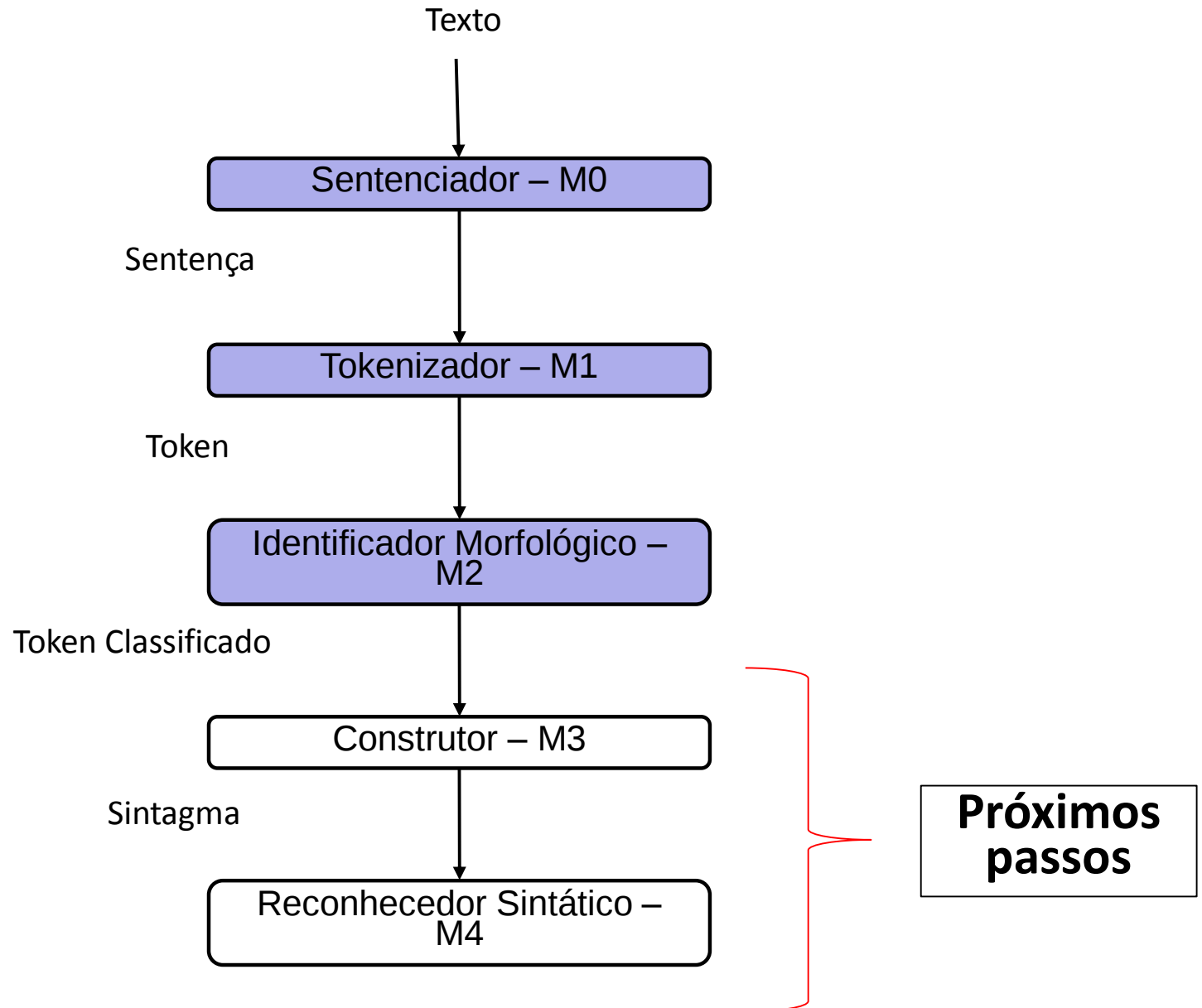
$E(n)$ – Etiqueta do conjunto de etiquetas

$P(n)$ – Classificação morfológica do conjunto de palavras do léxico

$P(\underline{En})$ – Probabilidade de ocorrência da etiqueta, dado que as 2 últimas foram U

$P(\underline{Pn})$ – Probabilidade da etiqueta da palavra analisada ser $E(n)$

O Reconhecedor Gramatical Linguístico



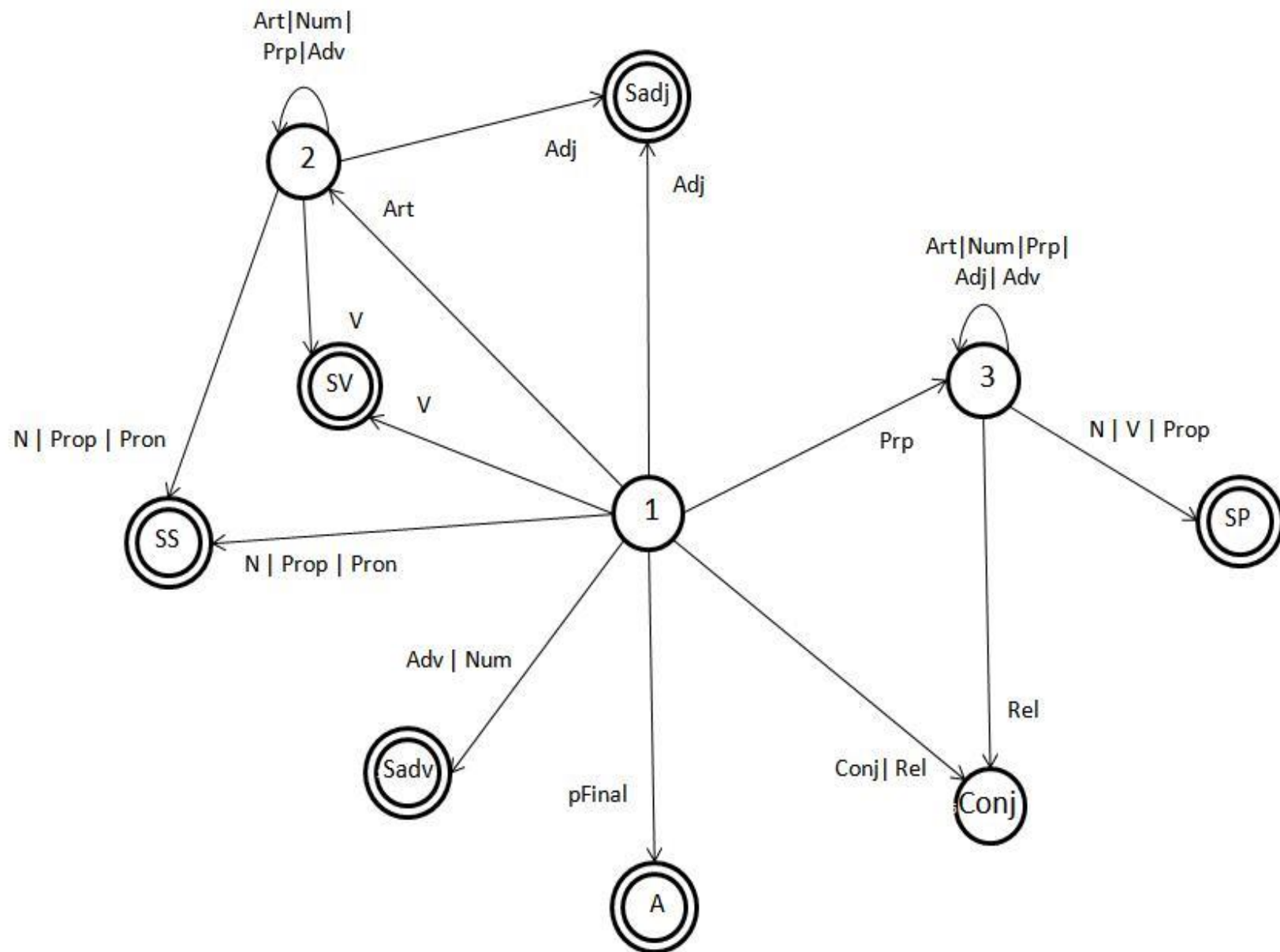
Padrões Oracionais Gramática de Luft

TABELA 2
PADRÕES ORACIONAIS DE LUFT

Padrões Pessoais Nominais				
SS	Vlig	SS		
SS	Vlig	Sadj		
SS	Vlig	Sadv		
SS	Vlig	SP		
Padrões Pessoais Verbais				
SS	Vtd	SS		
SS	Vti	SP		
SS	Vti	Sadv		
SS	Vti	SP	SP	
SS	Vtdi	SS	SP	
SS	Vtdi	SS	Sadv	
SS	Vtdi	SS	SP	SP
SS	Vi			
Padrões Pessoais Verbo-Nominais				
SS	Vtpred	SS	SS	
SS	Vtpred	SS	Sadj	
SS	Vtpred	SS	SP	
SS	Vtpred	SS	Sadv	
SS	Vtpred	SS		
SS	Vtpred	Sadj		
SS	Vtpred	SP		
Padrões Impessoais Nominais				
	Vlig	SS		
	Vlig	Sadj		
	Vlig	Sadv		
	Vlig	SP		
Padrões Impessoais Verbais				
	Vtd	SS		
	Vti	SP		
	Vi			

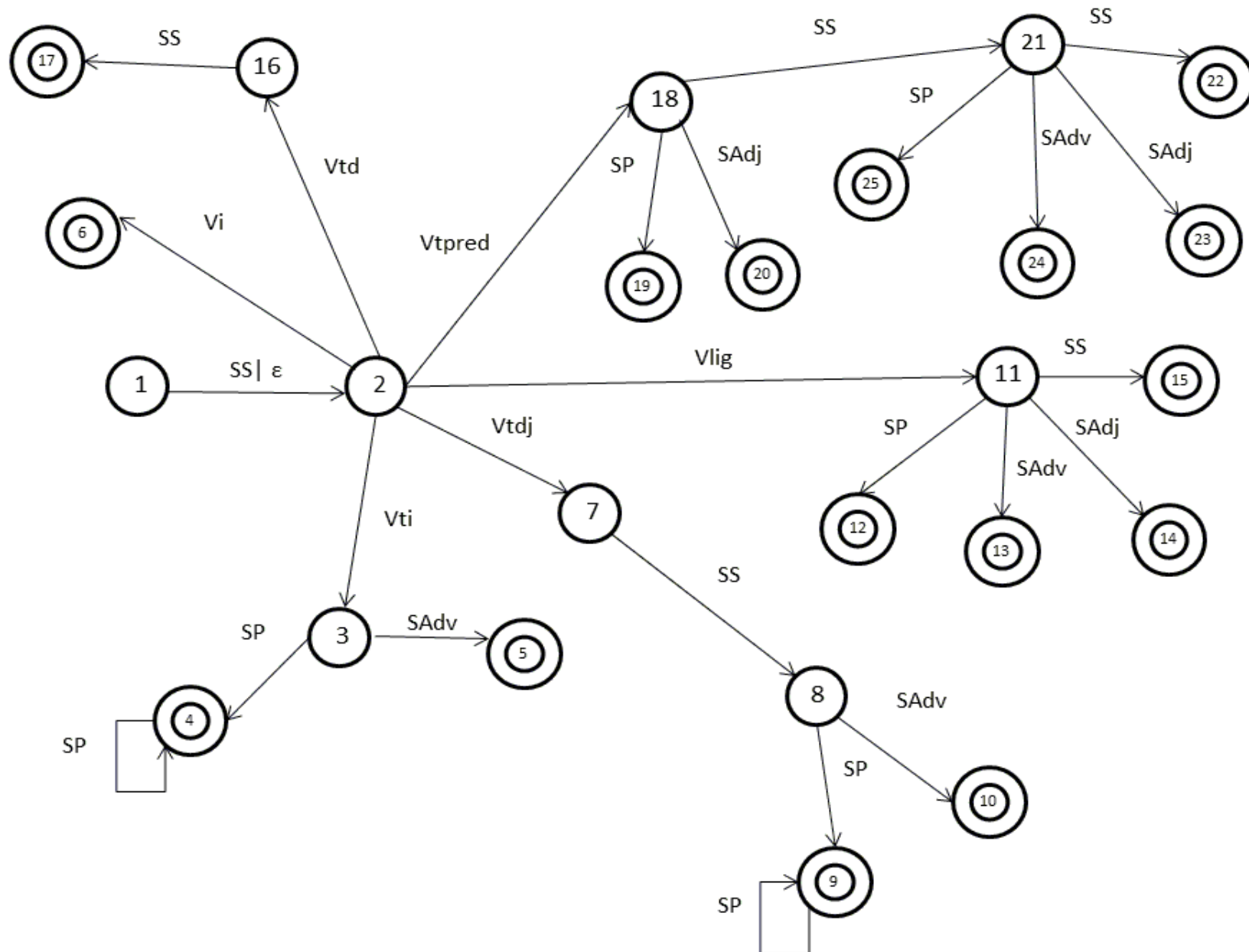
Módulo M3 – Construtor

Autômato Construtor de Sintagmas



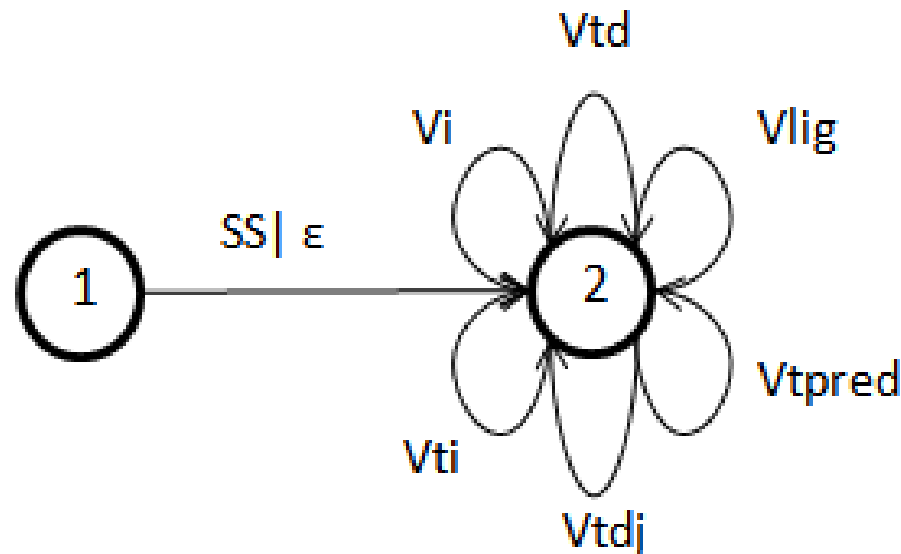
Módulo M4 - Reconhecedor Sintático

Configuração Completa



Módulo M4 - Reconhecedor Sintático Adaptativo

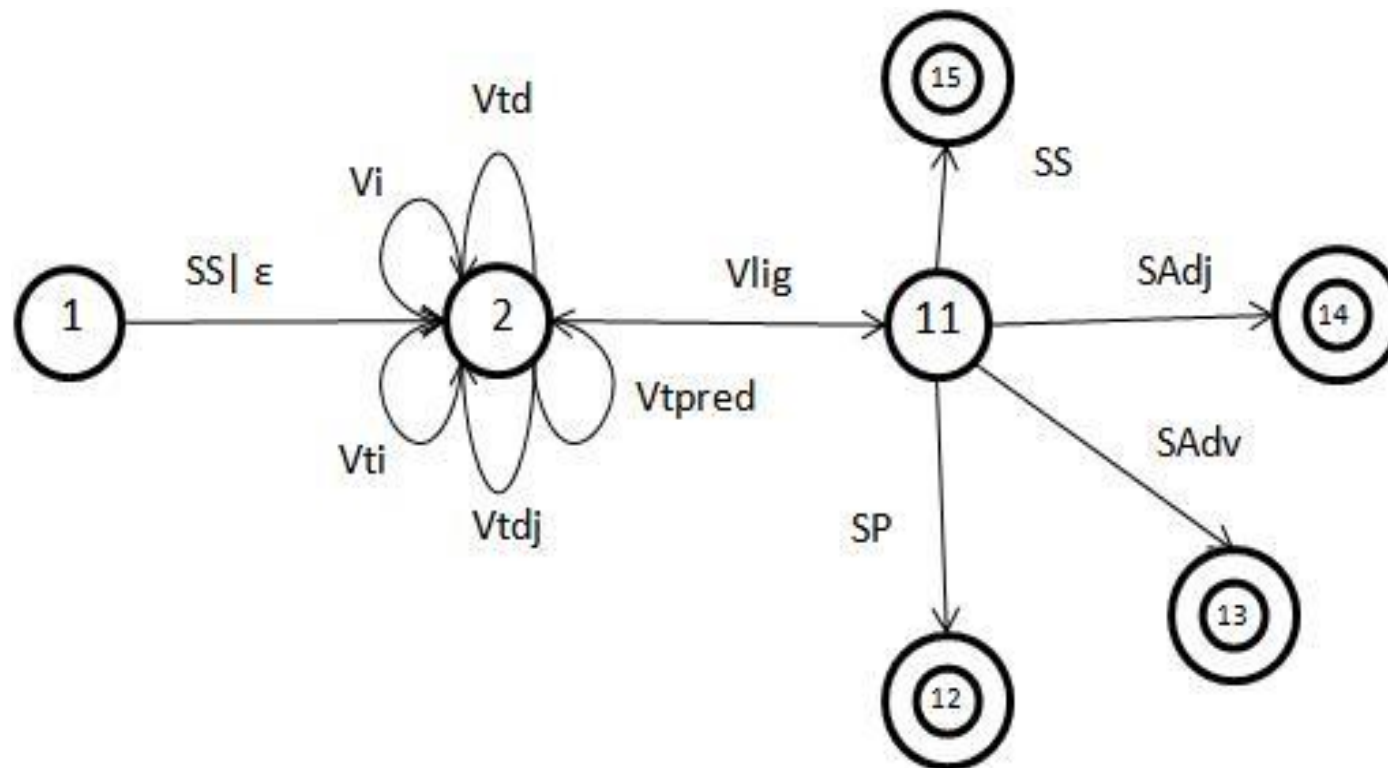
Configuração Inicial



$$\alpha(j): \{ o^* : \\ - [(j, \underline{Vlig})] \\ + [(j, \underline{Vlig}) \rightarrow o, \beta(o)] \\ \}$$

$$\beta(o): \{ t^*u^*v^*x^* : \\ + [(o, \underline{SP}) \rightarrow t] \\ + [(o, \underline{Sadv}) \rightarrow u] \\ + [(o, \underline{Sadj}) \rightarrow v] \\ + [(o, \underline{SS}) \rightarrow x] \\ \}$$

Módulo M4 - Reconhecedor Sintático Adaptativo Configuração Final



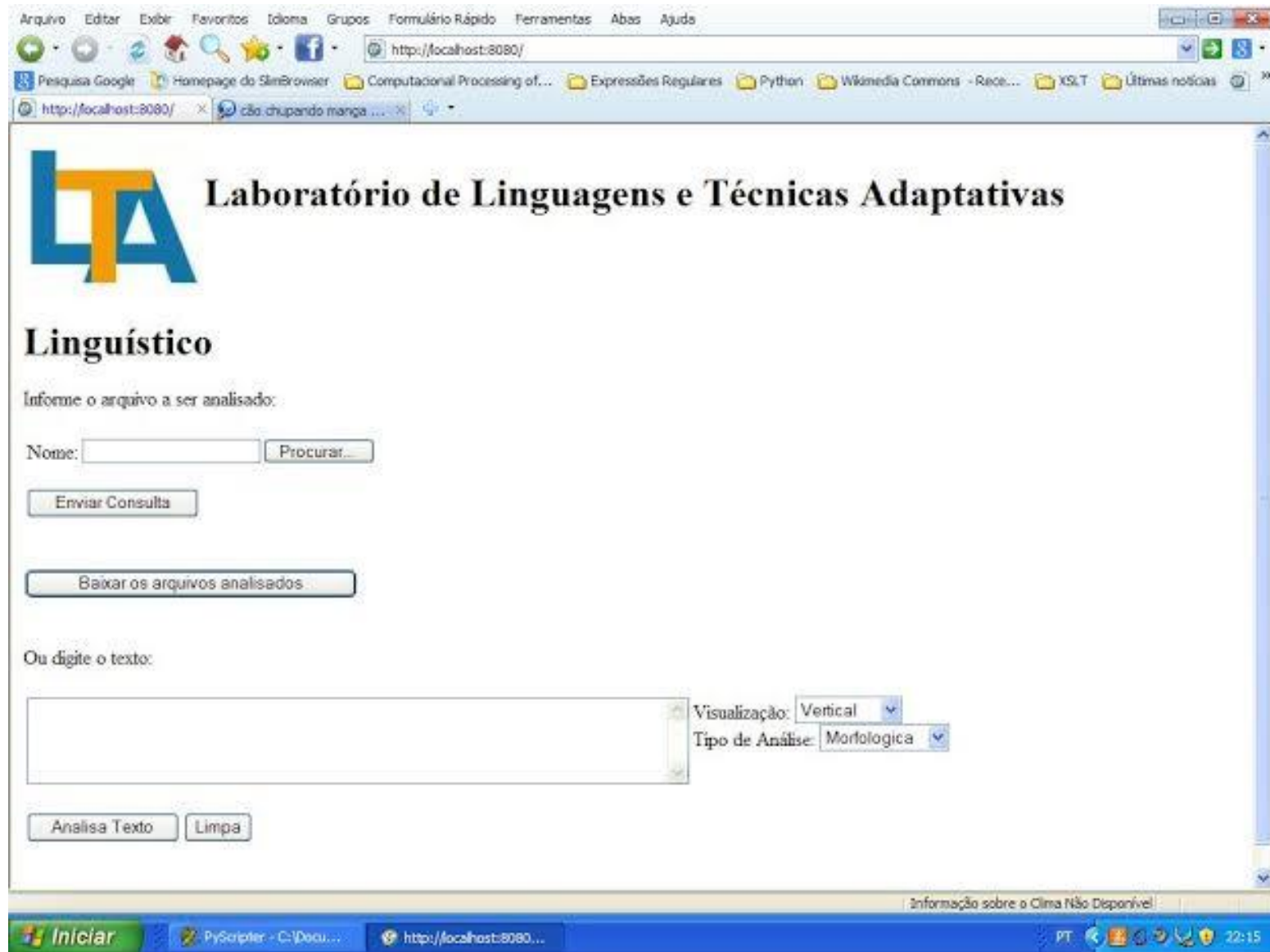
Semântica Aplicada ao Linguístico

- Classificação Morfossintática
 - Estrutura do Token Analisado
 - Automato identifica propriedades durante leitura do token
 - Sequência de Etiquetas X Token Analisado
 - Viterbi – Automato com Características Probabilísticas
 - Regência X Token Analisado
 - Características de tokens anteriores é usada para inferir a classificação do token analisado

Semântica Aplicada ao Linguístico

```
# #####
# Grammar Productions
# #####
# S expansion productions
S -> NP[NUM=?n] VP[NUM=?n]
# NP expansion productions
NP[NUM=?n] -> N[NUM=?n]
NP[NUM=?n] -> PropN[NUM=?n]
NP[NUM=?n] -> Det[NUM=?n] N[NUM=?n]
NP[NUM=pl] -> N[NUM=pl]
# VP expansion productions
VP[TENSE=?t, NUM=?n] -> IV[TENSE=?t, NUM=?n]
VP[TENSE=?t, NUM=?n] -> TV[TENSE=?t, NUM=?n] NP
# #####
# Lexical Productions
# #####
Det[NUM=sg] -> 'this' | 'every'
Det[NUM=pl] -> 'these' | 'all'
Det -> 'the' | 'some' | 'several'
PropN[NUM=sg]-> 'Kim' | 'Jody'
N[NUM=sg] -> 'dog' | 'girl' | 'car' | 'child'
N[NUM=pl] -> 'dogs' | 'girls' | 'cars' | 'children'
IV[TENSE=pres, NUM=sg] -> 'disappears' | 'walks'
TV[TENSE=pres, NUM=sg] -> 'sees' | 'likes'
IV[TENSE=pres, NUM=pl] -> 'disappear' | 'walk'
TV[TENSE=pres, NUM=pl] -> 'see' | 'like'
IV[TENSE=past] -> 'disappeared' | 'walked'
TV[TENSE=past] -> 'saw' | 'liked'
```

O Linguístico



Arquivo Editar Exibir Favoritos Idioma Grupos Formulário Rápido Ferramentas Abas Ajuda

http://localhost:8080/

Pesquisa Google Homepage do SlimBrowser Computacional Processing of... Expressões Regulares Python Wikimedia Commons - Rece... XSLT Últimas notícias

http://localhost:8080/ x cão chupando manga ...

LTA Laboratório de Linguagens e Técnicas Adaptativas

Linguístico

Informe o arquivo a ser analisado:

Nome:

Ou digite o texto:

Visualização:

Tipo de Análise:

Informação sobre o Clima Não Disponível

Iniciar PyScripter - C:\Docu... http://localhost:8080... PT 22:15

O Linguístico – Exemplo

Arquivo Editar Exibir Favoritos Idioma Grupos Formulário Rápido Ferramentas Abas Ajuda

http://localhost:8080/index

Pesquisa Google Homepage do SlimBrowser Computacional Processing of... Expressões Regulares Python Wikimedia Commons - Rece... XSLT Últimas notícias

http://localhost:8080/...

Ou digite o texto:

As eleições transcorreram pacificamente.

Visualização: Vertical

Tipo de Análise: Morfológica

Analisa Texto Limpa

As:
classificação selecionada: >N+art
classificações do léxico: ['>N+art', '>N+pron-det']

eleições:
classificação selecionada: H+n
classificações do léxico: ['H+n', 'P<+n']

transcorreram:
classificação selecionada: H+v-fin-2
classificações do léxico: ['H+v-fin-2']

pacificamente:
classificação selecionada: ADVL+adv
classificações do léxico: ['ADVL+adv']

.:
classificação selecionada: Pontuação
classificações do léxico: ['.']

Concluído

Informação sobre o Clima Não Disponível

Iniciar PyScripter - C:\Docu... http://localhost:8080... Imagem - Paint PT 22:50

Considerações Finais

- Autômatos adaptativos apresentam forte potencial de aplicação nas diversas etapas do processamento de linguagens naturais
- Gramática de Luft pode ser usada como suporte teórico linguístico
- O Linguístico está em fase de desenvolvimento. A primeira versão dos módulos Sentenciador e Tokenizador estão prontas e está sendo desenvolvido o módulo Desambiguador Morfológico
- Nas próximas etapas da pesquisa serão aperfeiçoados os módulos atuais e incorporados autômatos de construção de sintagmas e reconhecimento sintático