



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
DEPARTAMENTO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

**MAIKSON ADALTO BALDAN**

**UM AMBIENTE PARA CONSTRUÇÃO DE PERFIS A  
PARTIR DE TEXTOS PESSOAIS**

**VITÓRIA, AGOSTO 2012**

**MAIKSON ADALTO BALDAN**

**UM AMBIENTE PARA CONSTRUÇÃO DE PERFIS A  
PARTIR DE TEXTOS PESSOAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Departamento de Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Informática, na área de concentração em Informática na Educação.

Orientador: Prof. DSc. Crediné Silva de Menezes.

**VITÓRIA, AGOSTO 2012**

**MAIKSON ADALTO BALDAN**

**UM AMBIENTE PARA CONSTRUÇÃO DE PERFIS A  
PARTIR DE TEXTOS PESSOAIS**

**Dissertação apresentada ao Programa de Pós-Graduação em Informática do Departamento de Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Informática, na área de concentração de Informática na Educação.**

Aprovada em        de agosto de 2012.

COMISSÃO EXAMINADORA

---

Prof. Dr. Crediné Silva de Menezes

Universidade Federal do Espírito Santo  
(UFES) (Orientador)

---

Prof. Dr. Orivaldo de Lira Tavares

Universidade Federal do Espírito Santo  
(UFES)

---

Prof. Dr. Alberto Nogueira de Castro Junior

Universidade Federal (UFAM)  
(Membro Externo)

**VITÓRIA, AGOSTO 2012**

## **DEDICATÓRIA**

Dedico a mim mesmo e aos meus pais.

## **AGRADECIMENTOS**

Neste momento ímpar da minha vida, agradeço sinceramente:

Aos meus pais, por terem me proporcionado, principalmente no início de minha vida acadêmica, todos os recursos de que necessitei para conquistar minha autonomia e por meio dela alcançar meus objetivos.

A minha noiva Waldirlene que por muitas vezes me motivou, incentivou, aconselhou e ajudou nos momentos de que necessitei, sempre com uma dose sensata de cada atitude no momento certo.

A minha irmã (Maikisilene), familiares e amigos pela compreensão em minha investida acadêmica, principalmente pelo fato de que, em muitos momentos, não pude estar junto.

Agradeço ao Fernando Caetano Trivilin pelo apoio que me proporcionou.

Ao grande educador, professor e orientador Crediné Silva de Menezes pela paciência, pelas palavras e intervenções sempre construtivas.

Aos grandes Professores, mas acima de tudo amigos, Orivaldo de Lira Tavares e Dede Cury pelos seus conhecimentos compartilhados e pela energia positiva que transmitem aos seus aprendizes.

Aos colegas de laboratório (LIED), por apoiarem discussões sem fim sobre nosso projeto MORFEu, cujas contribuições foram extremamente importantes no desenvolvimento de minha pesquisa.

Aos meus amigos conterrâneos (Itaguaçuenses) que, embora distantes, mantiveram firme e inabalada nossa amizade.

Agradeço a Deus, grande criador, por ter me proporcionado a oportunidade e me mantido forte para alcançar o objetivo.

*Maikson A. Baldan.*

Vitória, agosto de 2012.

“Tudo o que um sonho precisa para ser realizado é alguém que acredite que ele possa ser realizado.”  
Roberto Shinyashiki

## RESUMO

O ambiente proposto tem como ponto de partida a modelagem e implementação de uma ferramenta para aquisição de características dos usuários a partir do processamento de suas produções. A abordagem a ser dada é o uso de técnicas de processamento de linguagem natural para identificação progressiva de características pessoais para a composição de perfil. O ambiente constitui-se de um conjunto de regras que são aplicadas sobre o Processamento de Linguagem Natural (PLN) para a construção do perfil. As regras de PLN delimitam o tipo de característica que é de interesse para a composição do perfil, assim como a composição de uma base de conhecimento modelada por lógica de primeira ordem, criada com o propósito de apoiar a identificação de características pessoais que estão implícitas nas produções individuais. Esta base de conhecimento é construída e consultada por um Sistema multiagente (SMA). O SMA utiliza esta base para realizar esclarecimento de dúvidas, dúvidas estas que surgiram durante o processamento das produções individuais e que não puderam ser sanadas com o processamento linguístico. O SMA necessita da base para identificar características pessoais implícitas nas produções individuais. O objetivo da ferramenta é identificar características pessoais dos usuários que utilizam o ambiente MOrFEu (MENEZES, 2008), tendo a noção de flexibilidade à qual o MOrFEu se propõe, permitindo que seja utilizada para diversos fins e tornando necessária a obtenção de características pessoais de forma também flexível. Esta flexibilidade está restrita às especificações das regras de PLN e lógicas de primeira ordem.

**Palavras-chave:** Construção de Perfis; Sistemas Multi-agente; Processamento de Linguagem Natural, MorFEu.

## **ABSTRACT**

The proposed starting point is the modeling and implementation of a tool to acquire characteristics of users from the processing of their productions. The approach is to be given the use of techniques for natural language processing to identify progressive personal characteristics for the composition profile. The atmosphere consists of a set of rules which are applied on Natural Language Processing (PLN) to construct the profile. The procedures for the PLN delimit the type of feature is of interest to the composition and the composition profile of a knowledge base modeled by first-order logic, created with the purpose of supporting the personal identification features that are implicit in the production individual. This knowledge base is built and queried by a multi-agent system (SMA). The SMA uses this basis to achieve clarification of doubts, these doubts have arisen during the processing of individual productions that could not be remedied with language processing. The SMA requires the basis to identify personal characteristics implicit in individual productions. The purpose of the tool is to identify personal characteristics of users who use the environment Morpheus (MENEZES, 2008), and the notion of flexibility in what they propose Morpheus, allowing it to be used for many did, making it necessary to obtain personal characteristics as well as flexible. This flexibility is restricted specifications of the rules of PLN and the logic of first order.

**Keywords:** Construction Profiles, Multiagent systems, Natural Language Processing, MOrFEu.



## LISTAS DE FIGURAS

Figura 2.1 Agente Inteligente de Software. Fonte: (RUSSELL, 1995).....	22
Figura 2.2 Agente Inteligente de Software. Fonte: (WOOLDRIDGE, 1999). .....	23
Figura 2.3 Arquiteturas de Comunicação de Agentes (a) Direta e (b) Assistida. Fonte: (HUHNS and STEPHENS, 1999) .....	24
Figura 2.4 Aprendizagem em SMA. Fonte: (SEN and WEISS, 1999) .....	28
Figura 2.5 - Níveis de PLN .....	33
Figura 2.6 - Árvore de Derivação Fonte: (PAGANI, 2004) .....	34
Figura 4.1 - Perfil do Usuário (Fonte: SERT, 2012).....	44
Figura 5.1 - Solução Sistema Multi-Agente para reconhecer características .....	53
Figura 5.2 - Processamento de Texto .....	59
Figura 6.1 - Interação entre os Módulos.....	67
Figura 6.2 - Processamento de Texto Implementado no Protótipo. ....	69
Figura 6.3 - Módulo Editor de UPI .....	70
Figura 6.4 - SMA .....	72
Figura 6.5 - Exemplo de Etiquetas .....	75
Figura 6.6 - Etiquetagem Com 2 Vínculos De Dependência Gramatical .....	80
Figura 6.7 - Algoritmo Seção de Habilidade.....	80
Figura 6.8 - Análise Etiqueta Comida.....	82
Figura 6.9 - Análise Etiqueta Relação Pessoal .....	84
Figura 6.10 - Algoritmo Seção Relação Pessoal .....	84
Figura 6.11 - Análise da Etiqueta <Hprof> Profissão .....	85
Figura 6.12 - Análise Etiqueta <sport> Esporte.....	86
Figura 6.13 - Análise Hipótese .....	86
Figura 6.14 - Imagem Tela de Cadastro.....	88

Figura 6.15 - Imagem Tela de Cadastro de Verbo. ....	89
Figura 6.16 - Imagem Tela de Cadastro de Máscaras. ....	90
Figura 6.17 - Imagem Tela De Consultar/Visualizar Perfil Individual.....	91
Figura 6.18 - Imagem Tela de Consultar/Realizar Agrupamento. ....	94
Figura 6.19 - Base de Dados do SysCAP .....	96
Figura 6.20 - UPIs de Arnaldo .....	97
Figura 6.21 - Representação Do Perfil De Arnaldo Pelo Protótipo.....	98
Figura 6.22 - Representação Gráfica do Perfil do Usuário Arnaldo.....	99
Figura 6.23 - UPIs de Rômulo .....	100
Figura 6.24 - Representação do Perfil de Rômulo Pelo Protótipo .....	100
Figura 6.25 - Representação Gráfica do Perfil do Usuário Rômulo.....	101
Figura 6.26 - UPIs de Juliana .....	102
Figura 6.27 - Representação Gráfica do Perfil do Usuário Juliana.....	103
Figura 6.28 - Processamento Equivocado da Fala de Juliana .....	104
Figura 6.29 - Análise Comparativa das Etiquetas .....	105
Figura 6.30 - Máscaras da Seção de Habilidades.....	106

## **LISTA DE TABELAS**

Tabela 1.1 - Questões Norteadoras .....	17
Tabela 3.1 - Perfil no SEDHI. FONTE (SILVA, 2005).....	40
Tabela 3.2 - Especificação das Categorias IMS-LIP .....	42
Tabela 6.1 - Análise do Vínculo de Dependência Gramatical, Sequência Texto. ....	76
Tabela 6.2 - Análise do Vínculo de Dependência Gramatical, Sequência Gramatical. ...	76
Tabela 6.3 - Análise do Primeiro Vínculo de Dependência Gramatical .....	81
Tabela 6.4 - Análise do Segundo Vínculo de Dependência Gramatical. ....	81
Tabela 6.5 - Configurações do Protótipo .....	106

## **SUMÁRIO**

<b>1. INTRODUÇÃO .....</b>	<b>14</b>
1.1. MOTIVAÇÃO.....	15
1.2. OBJETIVOS.....	16
1.2.1. Objetivo Geral .....	16
1.2.2. Objetivo Específico.....	16
1.3. METODOLOGIA.....	17
1.3.1. Referencial Metodológico.....	19
1.4. ESTRUTURA DO TRABALHO.....	19
<b>2. REFERENCIAL TEÓRICO.....</b>	<b>21</b>
2.1. AGENTES INTELIGENTES .....	21
2.2. SISTEMAS MULTI-AGENTES – SMA .....	22
2.2.1. Comunicação em SMA.....	23
2.2.2. Aprendizagem em SMA .....	27
2.2.3. Coordenação em SMA.....	29
2.3. PERFIL.....	31
2.3.1. Árvores De Conhecimentos (PIERRE LÉVY, 1997).....	31
2.4. PROCESSAMENTO DE LINGUAGEM NATURAL – PLN .....	32
2.4.1. Análise Morfológica.....	33
2.4.2. Análise Sintática.....	33
2.4.3. Análise Discurso.....	36
2.4.4. Análise Pragmática .....	36
2.4.5. Etiquetagem .....	37
2.4.6. Outros Problemas .....	37
<b>3. TRABALHOS CORRELATOS.....</b>	<b>39</b>
3.1. SEDHI (SILVA, 2005).....	39
3.2. OCTOPUS .....	40
3.3. PELEP.....	41
3.4. IMS-LIP/PAPI.....	41

3.5.	ONTOLEARNER.....	43
3.6.	CONCLUSÃO.....	43
<b>4.</b>	<b>CARACTERIZAÇÃO DO PROBLEMA.....</b>	<b>44</b>
4.1.	PERSPECTIVAS DO TRABALHO.....	46
4.2.	MORFEU.....	49
4.3.	APLICAÇÕES DA PESQUISA.....	49
4.4.	CONCLUSÃO.....	50
<b>5.</b>	<b>PROPOSTA DE SOLUÇÃO.....</b>	<b>51</b>
5.1.	ESPECIFICAÇÃO DO SMA.....	52
5.2.	ESPECIFICAÇÃO DO PERFIL.....	57
5.3.	PROCESSAMENTO LINGUAGEM NATURAL.....	58
<b>6.</b>	<b>UMA IMPLEMENTAÇÃO PARA A PROPOSTA.....</b>	<b>66</b>
6.1.	ESPECIFICAÇÃO DO PROTÓTIPO.....	67
6.1.1.	Observações entre Proposta de Solução e Implementação do Protótipo.....	68
6.2.	MÓDULO EDITOR DE UPI.....	69
6.3.	MÓDULO DO SMA.....	70
6.3.1.	Agente Recolhedor de UPI - ARUPI.....	72
6.3.2.	Agente Etiquetador - AE.....	73
6.3.3.	Agente Recuperador de Características Pessoais ARCP.....	73
6.4.	MÓDULO DA INTERFACE ADMINISTRADORA DE PERFIS.....	87
6.5.	BASE DE DADOS DO SYSCAP.....	96
6.6.	EXPERIMENTO DO PROTÓTIPO.....	96
6.7.	CONFIGURAÇÃO DO PROTÓTIPO.....	106
6.8.	CONCLUSÃO.....	107
<b>7.</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>109</b>
7.1.	TRABALHOS FUTUROS.....	111
	<b>REFERÊNCIAS.....</b>	<b>113</b>
	<b>APÊNDICE A.....</b>	<b>119</b>
	<b>APÊNDICE B.....</b>	<b>126</b>

## 1. INTRODUÇÃO

Atualmente, a utilização de sistemas com propósitos web tem tomado espaço em grande escala e a proliferação de dispositivos eletrônicos portáteis (tais como celulares, tablets e computadores portáteis), que possibilitam a interação com esses ambientes, têm tornado o crescimento ainda maior. A difusão da internet contribui para que os agentes computacionais tornem-se capazes de compreender contextos com identificação de detalhes de precisão. Além disso, pesquisas relacionadas à adaptação (AUGUSTIN, 2002) trouxeram um novo modelo de interação humano-computador, provocando ainda mais a necessidade de aperfeiçoar recurso que possibilite a identificação das características pessoais de seus utilizadores. Estudos voltados para a computação ubíqua estão constantemente esbarrando na necessidade de identificação das características de seus usuários (SILVA et al., 2008) (WEISER, 1991) (SATYANARAYANAN, 2001).

Assim, torna-se desejável o desenvolvimento de aplicações capazes de extrair informações do contexto em que elas estão localizadas. A noção de contexto é definida de forma distinta por diferentes autores: Schilit et al. (SCHILIT, 1994) define contexto como "quem você é, com quem você está, e que recursos estão próximos". Já Dey e Abowd (DEY and ABOWD, 1999) definem contexto como "qualquer informação que possa ser usada para caracterizar a situação de uma entidade (onde entidade é uma pessoa, local ou objeto considerado relevante à interação entre um usuário e uma aplicação, incluindo o próprio usuário ou aplicação)". Aplicações capazes de utilizar o contexto para fornecer informações e/ou serviços relevantes ao usuário são classificadas como sensíveis a contexto (DEY and ABOWD, 1999).

Almeja-se conhecer as pessoas de forma ampla, construindo o perfil do usuário com informações que vão além das coisas que a pessoa faz. Este perfil seria construído a partir de suas produções textuais.

BRUSILOVSKY (1998) define SHA como sistemas que possuem acesso às informações sobre o usuário (dados pessoais, características, preferências, etc.) reunidas em um modelo de usuário, e as aplicam para adaptar vários aspectos visíveis

do sistema ao usuário. Em um ambiente educacional, o modelo de usuário é um modelo de aluno, que armazena os dados referentes a cada um, oferecendo dele uma descrição adequada e permitindo, assim, a adaptabilidade. Informações sobre o usuário são críticas quando se deseja oferecer algum tipo de adaptabilidade em um sistema, pois funcionam como referência para o sistema que busca adaptar seu ambiente às expectativas particulares de seus usuários. Porém, essas informações não são obtidas facilmente.

Neste mesmo contexto, as interações se proliferam para atender diferentes atividades do ser humano, sejam elas relacionadas com o trabalho, com o lazer ou com a aprendizagem. Muitas vezes, para tornarmos a experiência de uso desses sistemas mais conveniente aos nossos interesses, torna-se necessário selecionar o interlocutor mais apropriado, o que nem sempre é possível a partir do recolhimento direto de informações dos membros de uma rede de interações. Sendo assim, a proposta deste trabalho é representar características além do que normalmente está disponível em diplomas e certificados de cursos. É identificar características que podem ser o resultado da junção de pequenas informações, constituindo uma característica de interesse, revelada nos textos de seu autor e identificada através da análise dos padrões linguísticos (ou padrões identificáveis pela inteligência artificial e mineração de dados), o que permite ressaltar tais características, possibilitando delinear um perfil para o usuário. Perfil este que será construído automática ou semiautomaticamente.

Para o que se apresenta neste trabalho, os textos a serem processados e analisados estão registrados nas Unidades de Produção Intelectual (UPIs), que é a unidade básica de registro de produções no contexto do projeto MOrFEu (MENEZES, 2008), do qual este trabalho, a ser apresentado ao longo desta dissertação, é parte integrante,.

### **1.1. Motivação**

A motivação para o desenvolvimento do presente trabalho surgiu a partir da necessidade de se conhecerem características dos usuários do ambiente MOrFEu, somada à intenção de contribuir com um grande número de aplicações que poderiam

fazer uso dos recursos e ideias apresentadas por este estudo, o qual caminha com o objetivo de constituir características pessoais de forma automática ou semiautomática.

As características que são consideradas para a construção do perfil permitem identificar um sujeito hábil para a realização de atividades que necessitam de especificações minuciosas, o que vai além dos pré-conceitos ou características atribuídas às pessoas que possuem curso de qualificação profissional por exemplo.

Ao trabalharmos com pessoas, em geral, é relevante conhecê-las para que, com base neste conhecimento, possamos realizar, de forma mais adequada, a organização do trabalho.

O conhecimento prévio das características pessoais de outros indivíduos permitiria a disposição personalizada de conteúdo para cada usuário. Como exemplo, a disposição de propagandas, textos para leitura, indicações de vagas de emprego permitiriam identificar pessoas para a realização de atividades conforme a definição de seu perfil. Ou, ainda, identificar características comuns em determinadas regiões ou comportamento em determinado período do ano. Desse modo, o perfil do usuário é enriquecido de informações a respeito de seu dono conforme for sendo revelado e escrito em suas produções textuais (UPI).

## **1.2. Objetivos**

### **1.2.1. Objetivo Geral**

O propósito deste trabalho é identificar características pessoais dos autores, cujo perfil será constituído por meio do processamento e análise dos textos, as características serão obtidas através do processamento de linguagem natural.

### **1.2.2. Objetivo Específico**

Os objetivos específicos que guiam este trabalho para alcançarmos o objetivo geral é prover forma que permita identificar características pessoais por meio de textos, formas



estas que estão vinculadas à obtenção de informações peculiares, que podem ser extraídas por regras e padrões de textos, identificando informações neles contidas por meio de análise da estrutura das frases e/ou palavras-chave.

É estudada ainda uma forma de representar as regras a serem monitoradas para identificação das características pessoais.

Para obtermos sucesso na extração e representação do perfil dos usuários, serão estudados os seguintes itens:

- i. Que padrões linguísticos podem ser reveladores de características pessoais?
- ii. É possível construir perfil de usuários a partir de textos em linguagem natural?
- iii. Que ferramentas podem fornecer apoio no processamento linguístico?

### 1.3. Metodologia

Este trabalho, de cunho experimental, ganhou forma e proporções a partir do desejo de representar as características pessoais dos autores/usuários extraídas dos textos.

Para tanto, propomos um conjunto de classes, definidas a partir de diferentes perspectivas de análise das abordagens. Cada uma destas perspectivas de análise pretende responder a um conjunto de questões norteadoras, conforme descritas na Tabela 1.1.

**Tabela 1.1 - Questões Norteadoras<sup>2</sup>**

<b>Perspectivas</b>	<b>Questões Norteadoras</b>
Objetivos da Pesquisa	<ul style="list-style-type: none"> <li>• Quais são os objetivos da pesquisa?</li> <li>• Quem é o público-alvo dos perfis produzidos pela abordagem?</li> </ul>
Fonte de Dados Referenciada	<ul style="list-style-type: none"> <li>• Quais tipos de fontes de dados são utilizados pela abordagem?</li> <li>• Quais são os métodos e técnicas empregados</li> </ul>

	<p>para manipulá-los?</p> <ul style="list-style-type: none"> <li>• A fonte de dados pertence a um domínio específico?</li> </ul>
Perfil	<ul style="list-style-type: none"> <li>• Quais as ferramentas utilizadas na construção dos perfis?</li> <li>• Quais métodos linguísticos permitem a identificação de características?</li> </ul>

Vários ambientes estão extraindo informações de seus usuários para proporcionar satisfação e maior eficiência em suas atividades. Porém a extração das informações para a composição do perfil a partir de textos ainda não é abrangentemente estudada.

Desta forma, dado um conjunto de sujeitos e dado um conjunto de textos de cada sujeito, fornecidos progressivamente, desejamos obter perfis parciais desses sujeitos.

A forma como foi abordado e investigado:

1 – Estudo das aplicações a serem dadas para os perfis dos usuários dentro do ambiente MOrFEu.

2 – Levantamento bibliográfico - Nessa etapa foram buscados, na literatura, recursos que poderiam subsidiar a concepção de uma solução para o problema;

3 – Esboço de uma proposta inicial para solucionar o problema de construção de perfil.

Nessa etapa foram agrupados recursos e ideias obtidos anteriormente, investigando-se de que forma eles poderiam interagir para a solução do problema.

4 – Validação da proposta: Testes das hipóteses validaram e afirmaram a continuidade da pesquisa com as ideias e ferramentas de apoio.

Nessa etapa foi validada a hipótese selecionada, através de um meta algoritmo, identificando a viabilidade da hipótese.

5 – Desenvolvimento e implementação do protótipo.

Nessa etapa foi implementado um protótipo, como prova das afirmações do que se propõe, como recursos passíveis de implementação para a obtenção automática de características pessoais por meio de análise das UPIs do usuário.

6 – Conclusão e obtenção dos resultados.

Nessa etapa foi realizada uma análise dos resultados obtidos pela proposta e pelo resultado dos testes realizados com o protótipo.

### **1.3.1. Referencial Metodológico**

Esta pesquisa, segundo seus objetivos e pontos de vista, apresenta dupla designação: pesquisa descritiva e pesquisa exploratória.

O resultado esperado a partir da pesquisa descritiva é a solução do problema definido para a construção de perfis segundo as premissas propostas. A solução, portanto, demonstra os procedimentos usados para determinar as relações entre a linguagem natural e os elementos que constituem as características de um perfil.

Sobre o outro ponto de vista, o exploratório foi concebido pelo empenho realizado na pesquisa para a definição dos elementos que aproximaram o que se tem disponível na bibliografia com a solução proposta por este trabalho.

## **1.4. Estrutura do Trabalho**

Os próximos capítulos estão estruturados da seguinte forma:

No capítulo 2 é feito um referencial teórico sobre Agentes Inteligentes, perfil de usuários e processamento de linguagem natural.

No capítulo 3 é feita uma análise de trabalhos correlatos, identificando diferenças e características das abordagens.

No capítulo 4 é apresentada a caracterização do problema e identificação do contexto em que está sendo estudado.

No capítulo 5 expomos a nossa proposta de solução para os problemas apresentados no capítulo 4, destacando as tecnologias utilizadas e explicando a arquitetura proposta.

No capítulo 6 é apresentado o protótipo desenvolvido, com o propósito de validar nossa proposta de solução, apresentando também as ferramentas utilizadas, as formas de aprendizado e todo o fluxo do sistema.

O capítulo 7 apresenta a análise dos dados, os resultados obtidos com o protótipo e as observações pertinentes às análises dos dados submetidos.

O capítulo 8 apresenta as considerações finais e trabalhos futuros. E, finalmente, no capítulo 9, são apresentadas as referências bibliográficas.

## 2. REFERENCIAL TEÓRICO

O objetivo deste capítulo é pontuar as considerações teóricas necessárias para a “extração de perfis”, tema central para o desenvolvimento da presente dissertação.

Iniciamos pela apresentação dos conceitos de agentes de softwares inteligentes. Delineamos brevemente, na Seção 2.1, as características dos agentes de softwares e, na Seção 2.2, abordamos a teoria dos sistemas multiagentes SMA, ressaltando como a teoria estabeleceu os principais elementos e a dinâmica de construção dos SMA.

Em seguida, apresentamos estudos realizados para a construção de modelos de perfil e a extração das informações que o compõem na Seção 2.3. Em uma subseção 2.3.1, realizamos a ideologia para a representação das características pessoais. E, no final, na Seção 2.4, realizamos uma apresentação dos vários pontos a serem considerados para se trabalhar com processamento de linguagem natural.

### 2.1. Agentes Inteligentes

Agentes inteligentes são frequentemente propostos como entidades apropriadas para permitir a análise de projeto de sistemas complexos. Formados por várias entidades que se comportam de forma autônoma, permitem interagir uns com os outros, a fim de alcançar um objetivo comum (ou seja, a funcionalidade geral do sistema) (JENNINGS et al., 1998).

Qualidades de um agente inteligente (WOOLDRIDGE and JENNINGS, 1995):

- Reatividade: capacidade de perceber seu ambiente e responder na hora certa às mudanças que ocorrem, para satisfazer seus objetivos de projeto.
- Proatividade: capacidade de exibir comportamento direcionado a objetivos, tomando iniciativa para satisfazer seus objetivos de projeto.
- Habilidade social: capacidade de interagir com outros agentes (e possivelmente com humanos) para satisfazer seus objetivos de projeto.

Veja na figura 2.1 a representação gráfica do agente e seus atuadores.

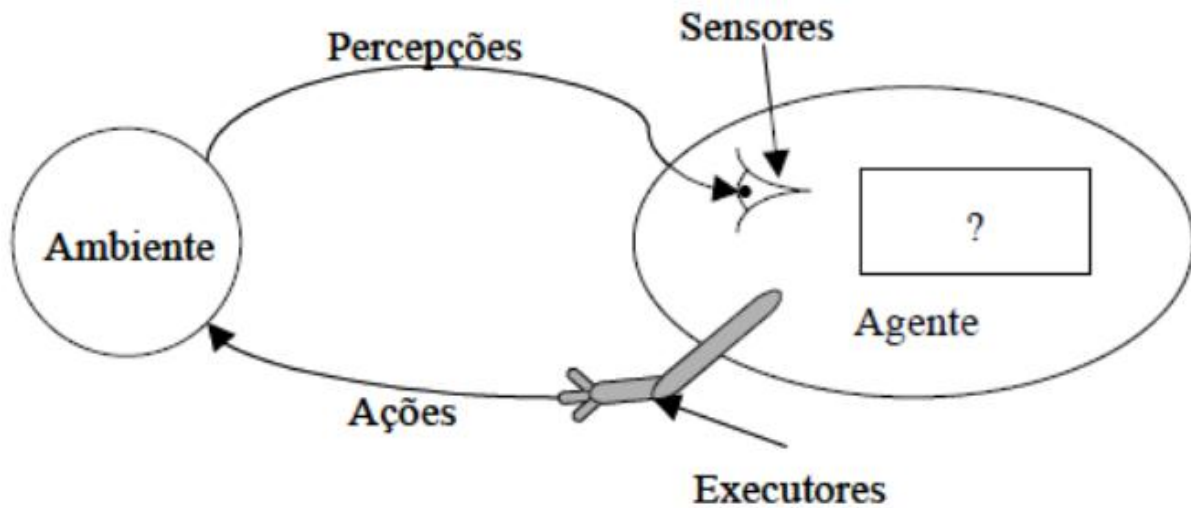


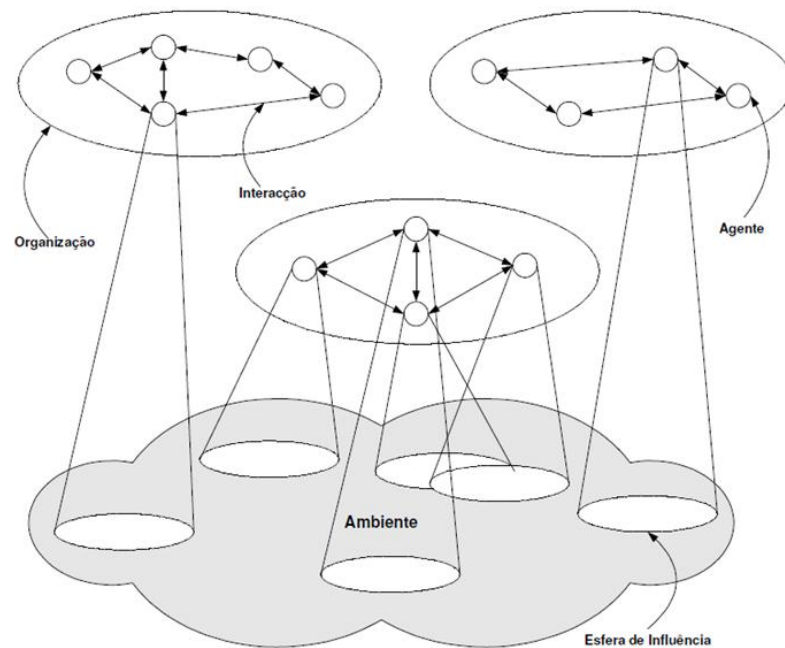
Figura 2.1 Agente Inteligente de Software. Fonte: (RUSSELL, 1995).

Na Figura 2.1 temos dois elementos: o ambiente e o agente. O agente contém os sensores representados na Figura 2.1 como os olhos, e contém os executores representados na Figura 2.1 por um braço. Permite que o agente realize manipulações no ambiente ao qual ele interage.

As características sociais para os agentes, em alguns casos, nos remete a outras abordagens dos agentes, que são os SMAs, os quais veremos na seção 2.2.

## 2.2. Sistemas Multi-Agentes – SMA

Um Sistema Multi-Agente é um sistema computacional em que dois ou mais agentes interagem ou trabalham em conjunto, de forma a desempenhar determinadas tarefas ou satisfazer um conjunto de objetivos. A investigação científica e a implementação prática de Sistemas Multi-Agente estão focalizadas na construção de padrões, princípios e modelos que permitam a criação de pequenas e grandes sociedades de agentes automáticas ou semiautomáticas, capazes de interagir convenientemente de forma a atingirem os seus objetivos (LESSER, 1999).



**Figura 2.2 Agente Inteligente de Software. Fonte: (WOOLDRIDGE, 1999).**

Na Figura 2.2 é ilustrado um Sistema Multi-Agente, contendo múltiplos agentes, cada qual com diferentes capacidades de percepção e ação no mundo. Cada agente terá uma esfera de influência distinta sobre o ambiente, ou seja, será capaz de influenciar diferentes partes do ambiente (JENNINGS, 2000). Estas esferas de influência podem coincidir dependendo das relações existentes entre os agentes.

Num SMA, os agentes que o compõem, podem estar a trabalhar em conjunto para atingir um só objetivo geral, ou rumo a objetivos individuais separados que podem, no entanto, estar relacionados, sendo desta forma necessária a interação entre os diferentes agentes no sentido de atingirem os seus próprios objetivos.

### **2.2.1. Comunicação em SMA**

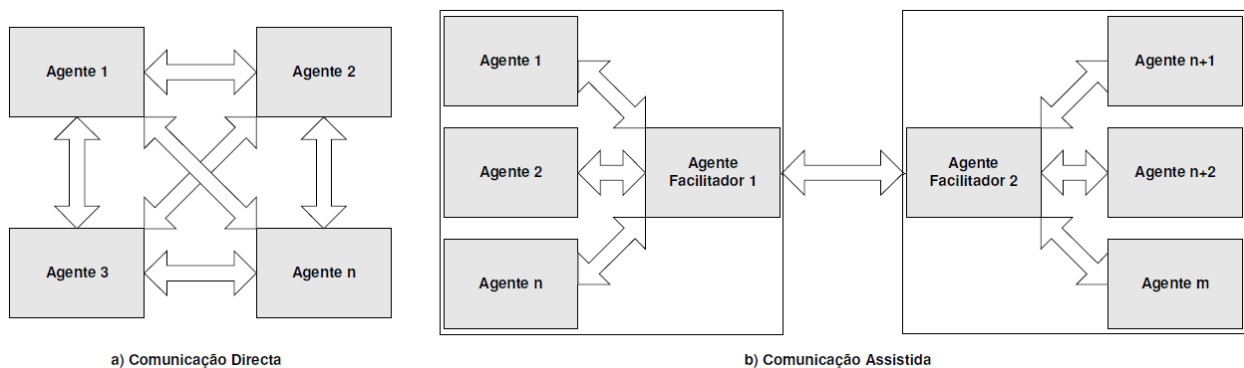
A comunicação é essencial em sistemas multiagentes, pois os agentes em uma sociedade ou em grupo precisam de se comunicar, a fim de estabelecer um diálogo na resolução de seus objetivos. Em geral, a maioria das línguas de comunicação de agente (ACLs) e a teoria de comunicação são baseadas na Teoria de atos de fala desenvolvida por (AUSTIN, 1962) e, posteriormente, prorrogada por (SEARLE, 1969).

#### **2.2.1.1. Arquiteturas de Comunicação**

Huhns e Stephens denotam que a comunicação entre os agentes pode assumir uma de duas arquiteturas básicas (HUHNS and STEPHENS, 1999), representadas esquematicamente na Figura 2.3:

**Comunicação Direta:** Os agentes tratam da sua própria comunicação sem intervenção de qualquer outro agente. Neste tipo de arquitetura, cada agente comunica diretamente com qualquer outro agente, sem qualquer intermediário, como pode ser observado na Figura 2.3 a).

**Comunicação Assistida:** Os agentes apoiam-se em agentes especiais designados “Agente Facilitador”, de forma a efetuarem a comunicação com os outros agentes. Nesta arquitetura, a organização de agentes é do tipo sistema federado. Nestes casos, se um dado agente  $i$  desejar enviar uma mensagem a outro agente  $j$ , terá primeiro de enviá-la para o “Agente Facilitador”, que se encarregará de a reencaminhar ao seu destinatário, como pode ser observado na Figura 2.3 b).



**Figura 2.3 Arquiteturas de Comunicação de Agentes (a) Direta e (b) Assistida. Fonte: (HUHNS and STEPHENS, 1999)**

Genesereth apresenta uma organização de sistema de agentes federados, vista na Figura 2.3 (b), que consiste de agentes e facilitadores como um meio de atingir a interoperabilidade entre estes. Uma organização federada sugere que agentes não se comunicam diretamente com outros, mas cada agente comunica-se com seu facilitador local que, por sua vez, comunica-se com outros facilitadores. Desta forma, cada grupo de agentes ligado a um único facilitador é uma federação, onde um agente rende parte de sua autonomia ao facilitador (GENESERETH, 1994).



Esta abordagem pode ser utilizada para agrupar os agentes nos diversos processadores de um ambiente distribuído. Desta maneira, cada federação corresponderia a um processador onde o respectivo facilitador teria controle sobre a conexão de rede.

A comunicação entre agentes está diretamente relacionada com o nível de implementação das comunicações. Esta comunicação pode ser implementada de duas maneiras:

- Memória partilhada (quadro-negro);
- Passagem de mensagens entre agentes.

De uma forma geral, a passagem de mensagens entre agentes é o modo de comunicação mais utilizado. As soluções de memória partilhada são em geral mais difíceis de distribuir e obrigam a metodologias adicionais para a sincronização dos agentes, enquanto que a passagem de mensagens pode facilmente ser usada para sincronização.

#### **2.2.1.2. Características da Comunicação**

A comunicação tem dois fins principais: partilha de conhecimento, informação, crenças ou planos com outros agentes; e coordenação de atividades entre agentes. No entanto, a realização de comunicação que permita atingir estas duas metas requer a definição de uma linguagem de comunicação comum ou partilhada, caracterizada por: sintaxe, semântica, vocabulário, pragmática, modelo do domínio do discurso.

#### **2.2.1.3. Linguagens de Comunicação**

No início dos anos 90, foi fundado nos Estados Unidos da América, o Knowledge Sharing Effort (KSE) financiado pelo DARPA, com o objetivo de desenvolver protocolos para a troca e representação de informação entre sistemas de informação autônomos. O KSE gerou dois produtos finais principais (FININ et al., 1993):

**A Linguagem KQML (Knowledge and Query Manipulation Language).** Segundo (WOOLDRIDGE and JENNINGS, 1995) KQML é uma linguagem externa para comunicações entre agentes. Define um invólucro para formatar mensagens que

determina o significado locutório da mensagem. O KQML não está preocupado com o conteúdo da mensagem, mas sim com a caracterização da informação necessária à compreensão desse conteúdo.

Assim, ela define um formato comum para mensagens. Uma mensagem KQML pode rudemente ser pensada como um objeto (POO): cada mensagem tem uma performativa (que pode ser pensada como uma classe de mensagem) e um número de parâmetros.

Exemplo:

```
(ask-one
:content (PRICE IBM ?price)
:receiver stock-server
:language LPROLOG
:ontology NYSE-TICKS
)
```

O atributo content é a mensagem a ser passada. O receiver indica para quem se deseja enviar a mensagem. O language indica a linguagem usada, onde é assumido que o destinatário compreende essa linguagem. O ontology identifica a terminologia usada na mensagem.

**Formato KIF (Knowledge Interchange Format).** A KIF é uma linguagem que se destina explicitamente a representar o conhecimento sobre um domínio de discurso específico. Foi desenvolvido primariamente como forma de definir o conteúdo de mensagens expressas em KQML.

Para além do KQML e KIF, existem várias linguagens definidas no âmbito da comunicação em Sistemas Multiagente. De entre as mais utilizadas destaca-se o FIPA ACL (Agent Communication Language).

### **FIPA ACL**

Em (WOOLDRIDGE and JENNINGS, 1995) relata que 1995 a Foundation for Intelligent Physical Agents, FIPA, iniciou o desenvolvimento de padrões para sistemas de agentes, definindo a FIPA ACL (FIPA, 1999) – Agente Communication Language superficialmente similar à KQML: definida como uma linguagem “externa” para mensagens, com 20 performativas para definição de interpretação desejada de

mensagens, e não obriga qualquer linguagem específica para o conteúdo das mensagens. Veja exemplo da linguagem FIPA ACL.

```
(inform
  :sender agent1
  :receiver agente2
  :content (price good2 150)
  :ontology hpl-auction
)
```

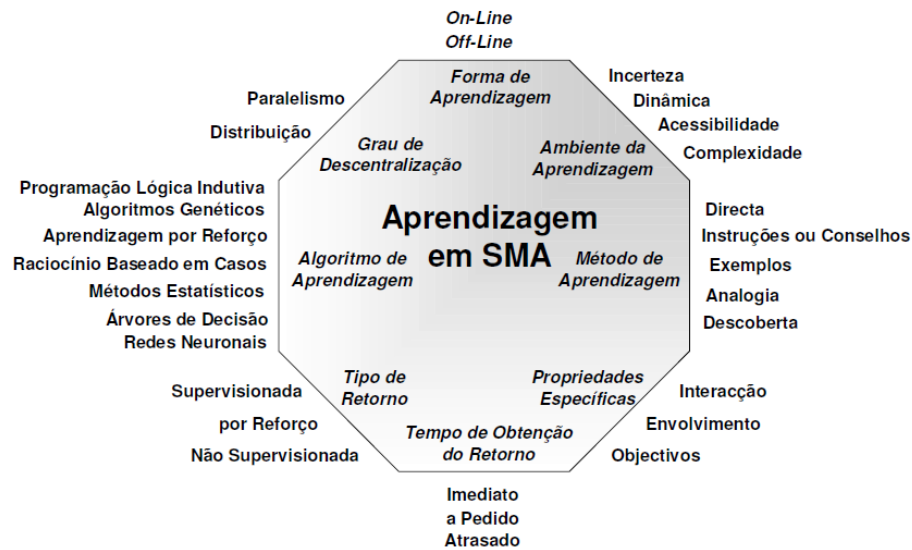
### 2.2.2. Aprendizagem em SMA

As duas categorias de aprendizagem descritas (Ex.: aprendizagem interativa e individual) cobrem um largo espectro de tipos de aprendizagem mais específicos que podem ocorrer no contexto de um SMA.

**Aprendizagem Interativa.** Situações em que os agentes, de forma coletiva, procuram atingir os seus objetivos de aprendizagem comuns.

**Aprendizagem Individual.** Situações em que cada agente procura atingir os seus próprios objetivos de aprendizagem, mas em que a sua aprendizagem é afetada por outros agentes: os seus conhecimentos, crenças, intenções, etc.

Na Figura 2.4, é apresentado um diagrama que resume os principais tópicos de investigação no âmbito da aprendizagem em SMA.



**Figura 2.4 Aprendizagem em SMA. Fonte: (SEN and WEISS, 1999)**

As abordagens à aprendizagem em SMA podem ser classificadas com maior precisão utilizando um conjunto de características auxiliares (SEN and WEISS, 1999):

- Grau de Descentralização.
- Forma de Aprendizagem.
- Ambiente de Aprendizagem.
- Propriedades Específicas da Interação.
- Propriedades Específicas do Envolvimento.
- Propriedades Específicas dos Objectivos.
- Método de Aprendizagem.
  - Aprendizagem Direta.
  - Aprendizagem por Instruções ou Conselhos.
  - Aprendizagem por Exemplos.
  - Aprendizagem por Analogia.
  - Aprendizagem por Descoberta.
- Tipo de Retorno da Aprendizagem.
  - Aprendizagem Supervisionada.
  - Aprendizagem por Reforço.
  - Aprendizagem Não Supervisionada.

- Tempo de Obtenção do Retorno.
  - Imediata.
  - A Pedido.
  - Atrasada.

Utilizando estas características, é possível caracterizar o processo de aprendizagem a diversos níveis. Tendo em conta o elevado número de combinações possível, é fácil concluir que a determinação da metodologia de aprendizagem mais adequada a cada caso específico é, em si, também um problema significativo neste contexto.

### **2.2.3. Coordenação em SMA**

A coordenação é o ponto central no SMA, pois é o processo no qual os agentes se engajam para garantir que o conjunto deles que compõe o sistema irá atuar de uma maneira coerente (NWANA et al., 1996). Quando os agentes trabalham para cumprir determinado objetivo, eles devem agir como uma unidade, coordenando suas ações, minimizando esforços redundantes, compartilhando recursos, dentre outros. A coordenação em SMA pode ser definida como a forma harmoniosa e coerente de trabalhar em conjunto para atingir um objetivo comum. Uma coordenação efetiva é essencial para que os agentes alcancem seus objetivos em um SMA (EXCELENTE-TOLEDO; JENNINGS, 2004). Desta coordenação se espera o controle das várias formas de dependência que aparecem naturalmente quando os agentes possuem objetivos relacionados, quando habitam o mesmo ambiente ou quando compartilham recursos.

O conceito de coordenação é intuitivo e é muito simples possuir uma noção do que significa a palavra coordenação. Por exemplo, quando se leciona numa faculdade bem gerida, ou se assiste a um jogo de futebol entre duas equipes capazes de realizar excelentes jogadas coletivas, é fácil perceber que existem ações de grupo bem coordenadas. Exemplos de ausência de coordenação verificam-se quando: se chega a um hotel onde se tinha uma reserva e o hotel está cheio, se chega a uma sala para lecionar uma aula e esta se encontra ocupada.

A coordenação pode ser definida como: “o ato de trabalhar em grupo de forma harmoniosa” (MALONE E CROWSTON., 1991).

Desta forma, diversos autores propuseram definições distintas, mas inter-relacionadas do termo coordenação, destacando-se:

“Processo pelo qual um agente raciocina acerca das suas ações locais e das ações (antecipadas) dos outros para tentar assegurar que a comunidade atue de maneira coerente” (JENNINGS, 1996).

Na maioria destas definições, destaca-se a noção de grupo, trabalho conjunto, harmonia e objetivo comum.

Sichman e Demazeau (SICHMAN, 1994) (SICHMAN E DEMAZEAU, 1995) afirmam que a necessidade de coordenação resulta da existência de relações de dependência que estão relacionadas com o fato de um agente necessitar do outro de forma a poder satisfazer os seus objetivos.

Jennings (JENNINGS, 1996) afirma que existem três razões principais para a necessidade de coordenar as ações de múltiplos agentes:

- Existem dependências entre as tarefas dos agentes, as quais ocorrem quando a realização de uma tarefa de um determinado agente influencia na execução de uma tarefa executada por outro agente;
- Agentes operam sobre restrições globais, as quais se tornam presentes quando os recursos estiverem escassos. A coordenação é então necessária para lidar com esta limitação, a qual não é respeitada quando os agentes apresentam um comportamento individual; e
- Agentes individuais não possuem informação, recursos ou capacidade para resolver problemas, eles possuem competência para realizar tarefas diferentes ou possuem informação sensorial complementar devido à, por exemplo, localização distribuída no ambiente ou capacidade limitada de percepção.

Para (NWANA et al.,1996), os mecanismos para coordenação multiagente podem ser classificados em quatro grupos: coordenação por estrutura organizacional, coordenação por criação de contratos, coordenação por planejamento multiagente e coordenação por negociação.

### **2.3. Perfil**

Em (ZIMMER, 2001) apresenta a relevância de conhecer o perfil dos profissionais com um refinamento e detalhes de características que modela um profissional. O propósito de compor grupos para a realização de trabalhos colaborativos com melhor grau de produtividade requer o conhecimento de várias características pessoais.

WARNER (1987) cita HARRIS (1984 apud WARNER, 1987) que ressalta não ser suficiente representar o conhecimento somente com estruturas estáticas, uma vez que a representação de conhecimento requer que uma estrutura interna de significados seja construída. Em WARNER (1987), apresenta uma visão mais ampla, classificando o conhecimento por um sistema com componentes: morfológico e fonético, sintático, semântico e pragmático.

Ainda segundo WARNER (1987), os mecanismos formais e as estruturas para representação de conhecimento são cobertos em inúmeros trabalhos. Dentre eles, CULLINGFORD (1986 apud WARNER, 1987) que considera a estrutura de conhecimento como sendo um aglomerado de relações conceituais, algoritmos de senso-comum e preferências semânticas, e finalmente GRISHMAN (1986 apud WARNER, 1987) que apresenta uma visão das principais estruturas e formalismos para representação sintática e semântica de conhecimento (WARNER, 1987, p. 82).

(MONTGOMERY, 1969) reconheceu a importância da semântica na teoria da gramática, o que, segundo a autora, era o caminho para a automação, com a especificação de uma lógica formal para representação semântica.

#### **2.3.1. Árvores De Conhecimentos (PIERRE LÉVY, 1997)**

Pierre Lévy escritor, filósofo e sociólogo francês, nos faz refletir sobre a riqueza e a beleza do nosso processo de aprendizagem, processo que acontece desde que

nascemos até o fim de nossas vidas, em todos os espaços em que vivemos e convivemos.

Pierre faz uma metáfora do processo de aprendizagem humano com o evoluir de uma árvore. A árvore, sendo um ser vivo que se renova constantemente, trocando suas folhas, crescendo ou secando galhos, raízes, flores e frutos, compara-se ao processo de aprendizagem humano.

A árvore captura nutrientes do solo e os transforma em galhos, folhas, raízes, flores e frutos. Assim é o processo do conhecer, que transforma as informações, conhecimentos e experiências em sabedoria e habilidades.

Saberes ou experiências como: saber cozinhar, contar histórias, cuidar de crianças, costurar e outros são patentes que permitem que o indivíduo seja valorizado por aquilo que ele sabe, por suas competências, e não por aquilo que ele não sabe (diferente dos preconceitos de habilidades adquiridos por diplomas).

A ideia de Lévy pode tornar-se ainda mais ampla se as informações que constituem a árvore de conhecimento contiverem outras, de preferências pessoais, como: alimentos, bebidas, esportes, parentes, amigos e regiões conhecidas. Assim, teremos uma abrangência de aplicações enorme.

O conjunto das patentes inseridas na Árvore de Conhecimento forma um brasão, que é uma representação gráfica dos saberes de cada indivíduo, ou seja, a sua identidade cognitiva. Realizamos uma analogia de Árvore de conhecimento com Perfil e Brasão, com estereótipo de uma pessoa.

#### **2.4. Processamento de Linguagem Natural – PLN**

O processamento da linguagem natural (PLN) é uma subárea de pesquisa que trata computacionalmente os diversos aspectos da comunicação humana, como sons e textos. De forma geral, o PLN visa fazer o computador entender a linguagem humana. E, para isso, considera o entendimento a partir de algumas fases do processamento de texto que são: morfológico, sintático, semântico, discursivo e pragmático (GONZALEZ,



2003). Cada uma dessas fases será apresentada nas subseções a seguir, observe a sequência em que são processadas na Figura 2.5.



**Figura 2.5 - Níveis de PLN**

#### **2.4.1. Análise Morfológica**

Na análise morfológica, cada palavra é analisada e classificada isoladamente, levando em consideração a sua própria estrutura. Estas palavras são classificadas de acordo com sua categoria gramatical específica. Analogamente equivaleria a organizar cada palavra na sua classificação morfológica, ou seja, verbos, artigos, adjetivos, substantivos, advérbios, conjunções, pronomes, interjeições, preposições e numerais.

Em uma sentença permanece válida, mesmo quando substituída por outra do mesmo tipo, por exemplo, os verbos, adjetivo, substantivos, etc. Há vários grupos que caracterizam o comportamento dos vocábulos das linguagens. A morfologia trata as palavras de conforme a sua estrutura gramatical, flexão e classificação (OLIVEIRA, 2004).

#### **2.4.2. Análise Sintática**

Por meio da gramática da linguagem que se pretende analisar e das informações do analisador morfológico, o analisador sintático procura construir árvores de derivação para cada frase, mostrando como as palavras estão relacionadas entre si. A análise sintática pode ser definida como o estudo minucioso da montagem e do modo de colocar as várias partes de um todo em posições adequadas (ARARIBÓIA, 1988).

Toda linguagem é definida por uma gramática e um léxico. A gramática descreve como a frase pode ser decomposta em sintagmas. Um sintagma é uma sequência de palavras que formam uma unidade significativa. Cada sintagma tem uma palavra principal, que é chamada de núcleo e outras palavras dependentes deste núcleo. Recursivamente, as palavras que acompanham o núcleo podem formar outros sintagmas.

Vejamos exemplo de análise sintática: Primeiro, precisamos de uma gramática para representar o fato em que uma sentença (S) pode ser formada por um nome próprio (NP) seguido de um sintagma verbal (SV). Recursivamente, a gramática deve conter uma regra que permite combinar um verbo e um nome próprio para formar um sintagma verbal. Eis as duas regras que essa gramática contém:

$S \Rightarrow NP, SV$

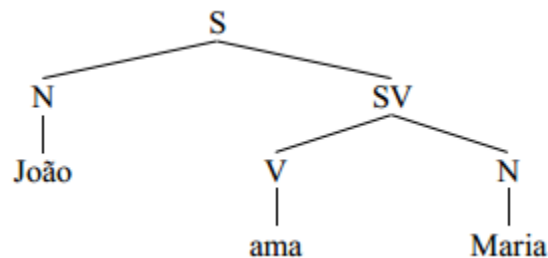
$SV \Rightarrow V, NP$

Para determinar a estrutura, é preciso um léxico que relacione as palavras com as categorias que podem lhe ser atribuídas:

V: ama

NP: Maria, João

Com essa gramática e esse léxico, podemos identificar uma estrutura sintática para a sentença “João viu Maria” na Figura 2.6.



**Figura 2.6 - Árvore de Derivação Fonte: (PAGANI, 2004)**

A árvore de derivação tem como objetivo organizar a lista de palavras que formam uma sentença em uma estrutura hierárquica.

Nos sistemas de processamento de linguagem natural, o maior problema é a transformação de uma frase potencialmente ambígua numa não ambígua, a qual será utilizada pelo sistema. Esta transformação é conhecida como parsing.

A análise semântica (AS) tem como objetivo descobrir o significado das palavras no contexto de uma frase ou sentença.

Mapeia-se o significado das palavras isoladas para os objetos apropriados na base de dados e ainda se criam estruturas corretas que correspondem ao modo como os significados das palavras isoladas combinam entre si.

Segundo (OLIVEIRA, 2004), a AS é feita para dar sentido às estruturas das palavras que foram reagrupadas pelo analisador sintático, uma vez que o AM já identificou cada uma delas individualmente.

A semântica é dividida em duas formas: a semântica léxica e a semântica gramatical.

**A semântica léxica:** busca uma representação conceitual para modelar o sentido. Para elaborar esta representação pode-se decompor a semântica em unidades léxicas. Para isso, compreender o relacionamento entre todas as palavras de uma frase é tão importante quanto saber o significado isolado de cada uma delas.

**A semântica gramatical:** tenta descrever o sentido lógico de uma frase. Como não há uma correspondência instantânea entre a sintaxe e a semântica, a mesma estrutura sintática pode gerar várias representações de semântica, ou seja, sentidos diferentes na mesma frase.

Por exemplo, “uma professora de capoeira pernambucana” pode referir-se a uma pessoa nascida em Pernambuco, a qual ensina capoeira, ou a uma pessoa que ensina capoeira no estilo em que é praticada em Pernambuco (OLIVEIRA, 2004).

O processo de uma ou mais representações semânticas em árvores de derivação é chamado de análise semântica, podendo ser feita posteriormente ou durante a geração das árvores.

Um sistema com capacidade de interpretação semântica avançada conseguiria solucionar a maioria dos casos de ambiguidade encontrados nestas sentenças, visto que, não há regra sintática que solucionaria estes problemas. Um sistema com tal capacidade teria um grande problema, que é o alto processamento de informações, para que pudesse obter a melhor aceção na tradução. Para isso, a consulta em outras

bases de dados seria indispensável, com a finalidade de encontrar a resposta satisfatória (OLIVEIRA, 2004).

#### **2.4.3. Análise Discurso**

A análise de discurso tem o objetivo de obter o significado de uma frase ou sentença mais próximo do contexto, considerando as sentenças antecessoras e sucessoras. O significado de cada sentença pode influenciar no significado da sentença sucessora. Com isto, vê-se a necessidade de que todo o contexto de um conjunto de sentenças seja analisado, para que se possa compreender o sentido da sentença, principalmente quando se trata de textos ou diálogos.

Vamos considerar a seguinte frase: “Eu quero imprimir o arquivo .int do João”. Já se sabe do que se trata nesta sentença. Agora se faz necessária a integração do discurso, para especificar que indivíduos estão sendo citados na sentença. Para que isso possa ser esclarecido, precisa-se de um modelo de contexto que permita descobrir qual usuário fez o pedido (aquele que digitou “Eu”) e quem é o usuário “João”. Sabendo-se, quem é João, pode-se definir qual é o arquivo (ou arquivos) com extensão .init, cujo dono é João. Este tipo de reconhecimento só será possível se o programa for capaz de compreender diversas sentenças, percorrendo uma grande base de conhecimento ou fortes restrições de domínio do discurso, de maneira a diminuir a base de conhecimento (SPECIA, 2000c).

#### **2.4.4. Análise Pragmática**

A pragmática é o estudo de como a linguagem é utilizada para atingir objetivos. É a análise da informação que se deseja transmitir além do que está escrito, devendo-se levar em consideração o contexto da informação. Segundo a Pragmática, o contexto dentro do qual a comunicação foi efetivada influi na compreensão do enunciado emitido, assim se uma pessoa diz à outra: “*Como está frio aqui*” pode-se entender pelo contexto um pedido para que a janela seja fechada.

A pragmática possui dois pontos importantes: o relacionamento entre as frases (no qual a construção de uma nova representação de um texto depende da representação da nova frase apoiada na frase precedente), e o contexto (que depende da situação e da condição em que ocorre o enunciado). À medida que novas sentenças vão sendo enunciadas, vão aumentando o universo a suas referências, unindo as já existentes à base de dados (VIEIRA and STRUBE, [2002 e 2004]).

Essa análise é importante porque uma sentença, após ser reconstruída, pode não ter na sua estrutura nenhum dado de interesse de interpretação. Por exemplo, nesta sentença “Bela camisa Fernandinho” podem ser identificados dois sentidos distintos, que vão depender do modo de interpretação.

A pragmática tem uma abrangência muito maior do que simplesmente analisar uma sentença isolada, ela passa a analisar todo o discurso, relacionando a língua a seu uso. Pode-se entender como discurso o texto ou a fala, que são divididos em unidades menores denominadas sentenças. Cada uma dessas sentenças podem ter significados para gerar sentidos diferentes que possam comprometer totalmente o significado do texto analisado.

#### **2.4.5. Etiquetação**

O processo de etiquetação é a adição de informação para cada palavra do texto, identificando uma descrição sintática, que inclui a classe morfológica.

Uma das dificuldades existentes na tarefa da classificação morfológica encontra-se em sua susceptibilidade à ambiguidade. Um etiquetador morfológico robusto deve levar em conta não apenas as informações lexicais da palavra a ser anotada, mas também informações a respeito do contexto em que esta palavra se encontra.

A etiquetação sintática e morfológica é um recurso muito utilizado pelas ferramentas de PLN.

#### **2.4.6. Outros Problemas**

Outros problemas pertinentes ao processamento da língua natural e o tratamento de ambiguidades e de estruturas complexas como anáforas, elipses e outras são temas que têm estimulado várias pesquisas para a solução de tais situações.

A aplicação de PLN neste trabalho é para a identificação de informações específicas nos textos dos usuários do sistema MOrFEu, fazendo-se necessário o conhecimento geral das fases em que os textos percorrem até que sejam identificadas e reveladas as características linguísticas de cada palavra no seu específico contexto.

Segundo (SAVADOVSKY, 1988) um dos problemas mais difíceis de tratar computacionalmente será o da ambiguidade de linguagens naturais, ou seja, a existência de várias formas de entender uma mesma frase.

a) *Ambiguidade léxica*: ocorre quando uma palavra pode ser interpretada de diversas maneiras. Por exemplo, a sentença: “João procurou um banco”. A palavra “banco” pode se referir à procura de um banco financeiro ou de um lugar para se sentar. Normalmente o sentido é percebido no contexto da frase;

b) *Ambiguidade sintática*: ocorre quando uma mesma sentença pode ser mapeada em mais de uma estrutura sintática válida. Por exemplo: “O menino viu o homem de binóculo”. Esta frase possui dois tipos de interpretação, como se o menino estivesse de binóculo ou, então, o homem estivesse de binóculo;

c) *Ambiguidade semântica*: ocorre quando temos mais de um significado para a mesma frase e acompanha a ambiguidade sintática, quando as diversas árvores sintáticas produzem análises semânticas válidas, como em: “Pedro viu Maria passeando” não se sabe se Pedro estava passeando e viu Maria, ou Pedro viu Maria passeando, na loja;

d) *Ambiguidade anafórica*: ocorre quando uma anáfora pronominal pode ser relacionada a duas ou mais palavras antecedentes e distintas. Por exemplo, “o ladrão entrou na casa do prefeito e tirou toda a sua roupa”, nesta frase a palavra “sua” pode estar relacionada com ladrão ou prefeito.

### **3. TRABALHOS CORRELATOS**

Em (IANNINI, 2012) apresenta o trabalho com o propósito de identificar o perfil de profissionais da área de Tecnologia da informação, tendo como pontos de interesses para a composição do perfil algumas características como nível de estudo, renda familiar, influência familiar na escolha da profissão, etc. Durante sua pesquisa deseja descobrir fatores comportamentais característicos dos profissionais, além de identificar a preferência por idiomas e o nível de conhecimento na língua estrangeira.

(ARELLANO, 2005) realizou uma pesquisa para obter perfil de usuários que utilizam lista de discussão da área de ciência da computação, querendo obter características como o comportamento e o que os levaram a utilizar a lista de discussão, perfil dos usuários que postaram suas dúvidas como sexo, idade, escolaridade, identificação de categorias de dúvidas comuns. A pesquisa foi realizada por meio de preenchimento de formulários.

Em (OLIVEIRA, 2004) foi realizado um estudo para identificar o perfil profissional e pessoal de estudantes do Curso de Secretariado Executivo, com o propósito de obter especificamente informações como: faixa etária, gênero, atividade profissional, responsabilidade familiar, remuneração, etc. Pretendem analisar quais são as representações sociais que contribuem como fatores intervenientes na escolha do curso. Fazem um relato da dificuldade inerente à obtenção das informações e as informações para compor o perfil dos alunos foram obtidas por meio de respostas aos questionários.

#### **3.1. SEDHI (SILVA, 2005)**

O SEDHI possui como características principais: classificar os estudantes em três tipos de perfis pré-definidos, baseados nos dados individuais e os dados de interação destes estudantes com o sistema; inicializar e atualizar o Modelo de Estudante a cada interação do estudante com o sistema; e adaptar a navegação no conteúdo conforme o perfil do estudante contido no Modelo de Estudante.

O curso é organizado em cinco módulos divididos em capítulos, tópicos e subtópicos. Em cada um destes módulos o estudante pode acessar um conjunto de recursos de apoio, a saber: “Glossário”, “Dica do Professor” e “Mergulhe Fundo”.

O SEDHI considera um conjunto de atributo-valor que se refere aos registros do sistema sobre as ações do estudante em relação ao (1) tempo gasto no conteúdo, (2) dúvidas enviadas para o tutor, (3) interação na lista de discussão, (4) reforço no acesso ao FAQ (Frequently Asked Question), (5) desempenho nos exercícios e (6) experiência em outros cursos Web.

O sistema faz a classificação dos usuários em três perfis: Básico, Intermediário e Avançado. De acordo com a classificação do perfil do estudante, o sistema SEDHI adapta a navegação no conteúdo por meio das técnicas de ocultação e anotação de links.

**Tabela 3.1 - Perfil no SEDHI. FONTE (SILVA, 2005)3**

ATRIBUTO-VALOR						PERFIS
Tempo	Dúvidas	Interação	Reforço	Desempenho	Experiência	
3	3	3	3	3	3	Básico
2	2	2	2	2	2	Intermediário
1	1	1	1	1	1	Avançado

Na Tabela 3.1 podemos perceber os valores usados pelo sistema para monitorar a interação do aluno com o sistema e retirar informações que irão ser utilizadas para classificar o aluno em um dos três perfis.

### **3.2. OCTOPUS**

No OCtoPUS os estereótipos são definidos em lógica de descrição, permitindo uma melhor definição dos perfis de acordo com o domínio. A definição dos perfis é idealizada de forma genérica e aprimorada conforme a interação do usuário.

O OCtoPUS utiliza análises semânticas e ontologias com o objetivo de inferir as informações de forma personalizada para compor o perfil do usuário.



### **3.3. PELEP**

O modelo PeLeP (LEVIS 2008) foi idealizado para estar conectado a um sistema ubíquo e utiliza o histórico de interação dos aprendizes para aperfeiçoar o perfil. Porém, por estar baseado no PAPI/LIP também dificulta o processamento das informações de perfil por dispositivos computacionais. Esse diferencial, o modelo OCtoPUS oferece, ao usar as ontologias baseadas em lógica de descrição para a representação das informações do usuário.

### **3.4. IMS-LIP/PAPI**

Os IMS-LIP (2001) e PAPI (2002) são as principais especificações de perfis que incluem informações detalhadas para personalização: histórico do aprendiz, atividades atuais e objetivos. Contudo, eles provêm somente um framework sintático para armazenar tais informações, fornecendo uma estrutura para descrições textuais. Dessa forma, o processamento textual do perfil é limitado, pois essas seções dos perfis são projetadas para serem lidas por pessoas ao invés de dispositivos computacionais. Sem qualquer formalismo ou padrão para expressar as informações dos aprendizes, qualquer tentativa dos dispositivos entenderem essa informação será baseada em técnicas de similaridade de textos não estruturados. O uso das tecnologias da web semântica é uma forma de definir tais vocabulários de forma que a informação seja processada tanto por humanos quanto por máquinas.

A especificação PAPI (PAPI, 2001) originou-se no grupo IEEE Learning Technology Standards Committee (LTSC). PAPI especifica a sintaxe dos dados sobre o aluno, caracterizando suas informações pessoais, seus conhecimentos, preferências, etc. As informações são divididas em seis categorias principais, que visam organizar os diversos tipos de dados do aluno.

A especificação LIP (Learner Information Package) (LIP 2001) foi definida pelo Global Learning Consortium da IMS, que é um consórcio de instituições educacionais, companhias de software e editoras. O objetivo do consórcio é promover a disseminação de especificações que permitam ambientes de aprendizagem distribuídos (LIP 2001). No LIP, as informações do aluno são separadas em onze categorias.

O LIP possui a vantagem de ter sido desenvolvido com base no PAPI. Além disso, o padrão LIP tem maior expressividade na área educacional e, pelo fato de ser desenvolvido por um consórcio que envolve universidade e empresas, ele está em constante atualização.

O IMS-LIP é um modelo de dados que descreve características de um estudante. Os elementos de dados que descrevem as características de um estudante são agrupados em 11 (onze) categorias, como pode ser observado na Tabela 3.2.

**Tabela 3.2 - Especificação das Categorias IMS-LIP4**

<b>Categoria</b>	<b>Descrição</b>
Identification	Agrupar informações sobre dados biográficos e demográficos relevantes para o aprendizado
Goal	Descreve os objetivos e aspirações pessoais do estudante. Pode ainda conter informações para monitorar o progresso na conquista desses objetivos.
QCL	Agrupar informações sobre as qualificações, certificações e licenças (licença para exercer medicina, advocacia, etc.) conquistadas pelo estudante.
Activity	Qualquer atividade relacionada ao aprendizado em andamento, por exemplo, treinamentos, estágios e serviço militar.
Transcript	Agrupar informações sobre o desempenho acadêmico de um estudante em uma instituição.
Interest	Agrupar informações sobre hobbies e outras atividades recreativas realizadas pelo estudante.
Competency	Agrupar informações sobre as habilidades adquiridas pelo estudante. Essas habilidades podem estar ou não associadas a algum treinamento formal ou trabalho realizado pelo estudante.
Affiliation	Armazena informações sobre as organizações profissionais das quais o estudante faz parte.
Accessibility	Armazena informações de acessibilidade genéricas sobre o estudante, tais como, conhecimento de línguas, deficiências e preferências cognitivas.
Securitykey	Armazena o conjunto de senhas e chaves de segurança do estudante para transações com sistemas de informação de estudantes e serviços.
Relationship	Descreve as possíveis relações existentes entre dados das outras categorias.

### **3.5. ONTOLEARNER**

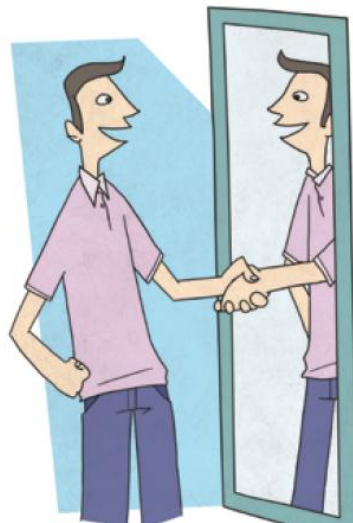
A OntoLearner é uma ontologia para a representação de aluno que foi definida com base nos padrões PAPI e LIP. Segundo (MUSA – 2007) ambos os padrões lidam com muitas categorias relacionadas com informações sobre o aluno e apresentam deficiências na sua representação conceitual e em algumas características. Foi realizada uma análise em cada elemento desses padrões, sendo escolhidos aqueles que foram considerados os mais importantes para um sistema de EaD adaptativo. Por essa razão, a Ontologia OntoLearner utiliza mutuamente os principais elementos de cada um dos dois padrões, gerando o modelo mínimo necessário.

### **3.6. Conclusão**

Foi percebido que existe um grande número de aplicações que se utilizam de características pessoais, com o propósito de entender comportamento, conhecimentos, preferências, facilidade, etc. Identificamos ainda que existe uma variação de métodos para se coletar dados a fim de compor o perfil do usuário, sendo a maioria deles por meio de preenchimento de formulários. Não foi localizado, na literatura, nenhum trabalho que utilize métodos linguísticos para coletar informação de texto com o objetivo de compor características pessoais ou representação de perfil com foco no propósito d que este trabalho se propõe.

## 4. CARACTERIZAÇÃO DO PROBLEMA

No desenvolvimento deste trabalho, pretendemos criar recursos para a identificação de perfis contendo suas características que representem informações que sejam de interesse de ser representada no perfil permitindo que outras pessoas te conheçam por meio da análise destas características que compõem o perfil assim como até mesmo o próprio usuário possa se identificar, identificar suas características ilustradas na Figura 4.1.



**Figura 4.1 - Perfil do Usuário (Fonte: SERT, 2012) 7**

Ao pensar em características pessoais podemos pensar em uma grande quantidade de informação. Vamos exemplificar algumas, levando em conta que você tem uma bagagem hereditária, como: altura, cor da pele, dos olhos, temperamento, talentos etc., em combinação com o meio em que você vive: sua família, comunidade, cidade, momento histórico etc.

Em (SERT, 2012) diz ser importante ir mais longe, acrescentando detalhes e listando as várias tarefas envolvidas nas atividades realizadas no dia a dia, no trabalho, na escola, na comunidade ou em casa... Enfim, deve ser valorizado tudo o que sabemos fazer.

Neste trabalho, partimos do pressuposto de que os participantes de um ambiente em rede, ao se manifestarem nos diversos espaços, muito provavelmente falarão de si mesmos, de suas atividades cotidianas, de suas preferências, das pessoas com as quais se relacionam, do que gostam, aonde vão, aonde deixam de ir e etc., dando assim possibilidades para que os demais usuários venham a conhecê-los. Propomos a implementação de um ambiente capaz de identificar as principais características de um usuário do sistema de forma progressiva. Quanto mais o usuário interagir com o ambiente, realizando suas tarefas e se comunicando, mais completo vai se tornando o perfil do usuário. Entendemos que hoje em dia e, cada vez mais, os usuários têm deixado registrado em texto as ações realizadas em seu dia a dia.

O ambiente proposto tem como ponto de partida a modelagem e implementação de uma ferramenta para aquisição de características dos usuários a partir do processamento de suas produções. A abordagem a ser dada é o uso de técnicas de processamento de linguagem natural para identificação progressiva de características pessoais para a composição de perfil.

O ambiente constitui-se de um conjunto de regras que são aplicadas sobre o Processamento de Linguagem Natural (PLN) para a construção do perfil. As regras de PLN delimitam o tipo de característica que é de interesse para a composição do perfil, assim como a composição de uma base de conhecimento, modelada por lógica de primeira ordem e criada com o propósito de apoiar a identificação de características pessoais que estão implícitas nas produções individuais. Esta base de conhecimento é construída e consultada por um Sistema multiagente (SMA). O SMA utiliza esta base para realizar esclarecimentos de dúvida surgidas durante o processamento das produções individuais que não puderam ser sanadas com o processamento linguístico. O SMA necessita da base para identificar características pessoais implícitas nas produções individuais.

O objetivo da ferramenta é identificar características pessoais dos usuários que utilizam o ambiente MOrFEu (MENEZES, 2008), tendo a noção da flexibilidade à qual se propõe, permitindo que ele seja utilizado para diversos fins e aplicações e tornando

necessário à obtenção de características pessoais de forma também flexível. Esta flexibilidade está restrita às especificações das regras de PLN e lógicas de primeira ordem.

O problema está associado à descoberta de características pessoais a partir do processamento de textos produzidos por um indivíduo ou seus possíveis "parceiros".

O sistema é previamente treinado para identificar tais características que são de interesse de uma determinada aplicação. O período de treinamento do sistema é o período em que são definidos e configurados os padrões linguísticos, assim como as premissas em lógica que permitem a identificação de características explícitas ou implícitas.

Na Seção 4.1 elencamos as perspectivas do trabalho que legitimam a pesquisa na construção automática de perfil.

Na Seção 4.2 é feita uma contextualização do projeto no ambiente MOrFEu.

Na Seção 4.3 é feita uma analogia das aplicações da pesquisa.

Na Seção 4.4 é feita uma conclusão.

#### **4.1. Perspectivas do Trabalho**

O estudo das informações que irão representar o perfil dos usuários, a partir de textos escritos pelo próprio usuário, requer a escolha de um conjunto de características que serão utilizadas como base para estruturar e representar o perfil.

Portanto, com o intuito de preencher esta lacuna, propomos, nesta seção, a caracterização da pesquisa neste campo apoiados na revisão bibliográfica de publicações do meio acadêmico que, explicitamente, declaram a intenção de produzir perfil de modo automático a partir de texto.

Apresentamos o interesse em implementar técnicas e recursos que permitam a obtenção de características pessoais que constituam um perfil extraído de forma automática por meio de textos. Como exemplo de aplicações podemos citar: desambiguação de consulta em ambientes de pesquisa, personalização de páginas

web conforme perfil dos usuários, ofertas comerciais personalizadas, aplicações com fins de recrutamento de pessoa, estudos estatísticos de comportamento e características pessoais, sites de relacionamentos, ambientes educacionais, etc. Considerando a diversidade de informação que pode estar contida nos textos e a diversidade de possibilidades de interesses, pensamos em um ambiente em que as regras linguísticas para a extração das informações possam ser modeladas e ajustadas para se obter a informação conforme o interesse e objetivo do perfil do usuário.

Propusemo-nos a tratar textos em português do Brasil, em conformidade com as recomendações da UNESCO que orientam para a criação e aplicação de recursos em língua materna (UNESCO, 2009, p.14).

Assim, ao longo desta dissertação, investigamos o problema da construção de perfil a partir de textos na linguagem natural em português, considerando ineficiente e dispendiosa a atividade de atualização do perfil do usuário de forma manual (AROYO at all, 2006).

Seguindo este raciocínio, pensamos em obter o perfil dos usuários indiferente de seu domínio de atuação e sem restringir nenhum tipo de domínio. Pensamos que as pessoas podem possuir características que relatam e revelam, em sua rotina diária, sendo elas atuantes de domínios diferentes, assim como ter mais de uma profissão. E seria normal esperar que essa pessoa escreva sobre seu cotidiano, revelando tais características de seu perfil.

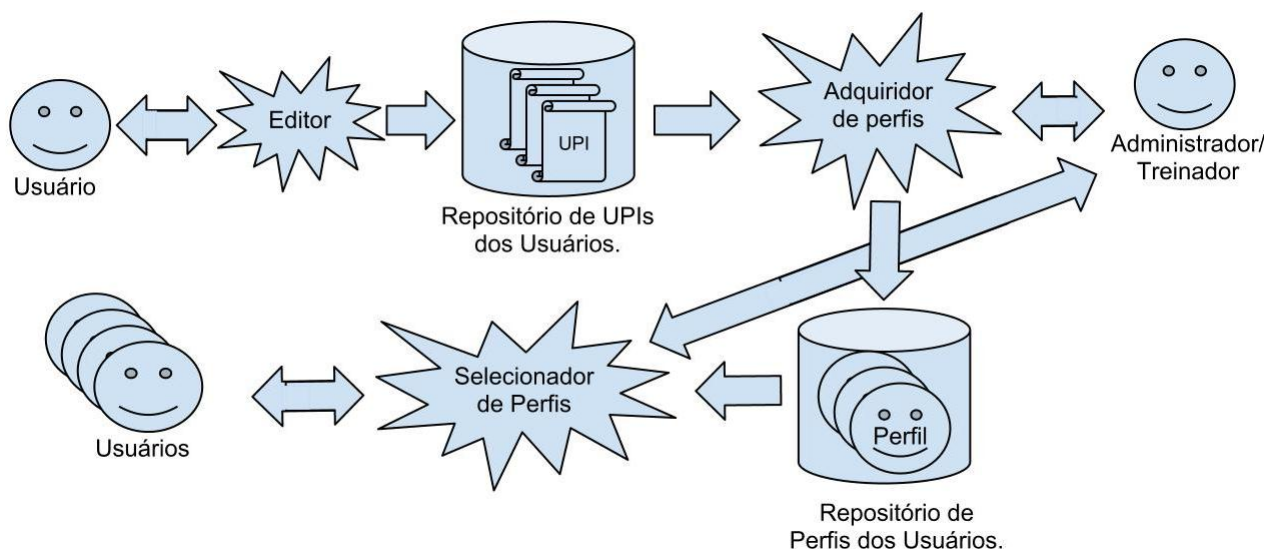
É comum pessoas, em seu dia a dia, se envolverem em vários contextos, por exemplo, contexto doméstico, contexto familiar, trabalho (podendo ter mais de um trabalho), etc.

Considerada a situação em que o usuário revele características de outra pessoa, neste caso, consideramos a necessidade de que esta informação deve ser validada pelo próprio usuário citado.

As produções de cada autor seguem o seu cotidiano de trabalho no ambiente MOrFEu e, conforme o autor escreve as suas UPIs, o sistema vai modelando e construindo o

seu perfil. Assim que surge uma UPI no MOrFEu, instantaneamente já entram em análise e atualização do perfil. O sistema fica monitorando, durante todo o tempo, tudo o que é produzido no MOrFEu.

Formalmente, o problema resume-se na necessidade de transformar uma Unidade de Produção Intelectual (UPI), com conteúdo expresso em português do Brasil e extrair as informações que compõem o perfil de um usuário. O conteúdo da UPI é constituído por um conjunto de sentenças  $s_1...s_n$ , ou seja,  $upi = \{s_1, s_2, \dots, s_n\}$ . Para cada sentença  $S_i$ , definida por meio de linguagem natural, em que  $i$  designa o índice da sentença na UPI, é necessário identificar verbos  $v_i...v_n$  a serem analisados em cada sentença. Uma característica para o perfil é constituída por um verbo e por um complemento. O complemento pode ser uma ou mais palavras que dão sentido ao verbo em um contexto. Veja na Figura 4.2 a visão geral da abordagem do problema.



**Figura 4.2 - Esquema da Abordagem**

Entendendo a Figura 4.2 o usuário interage com o ambiente, escrevendo as UPIs por meio do “Editor”. As UPIs são processadas pelo “Adquiridor de Perfil” e geram os perfis de cada usuário com base nos textos das UPIs. O professor pode realizar consultas de Perfil que lhe interesse, conforme for a sua intenção, por meio do “Selecionador de



Perfil". O administrador realiza a seleção dos perfis, informando quais as características de seu interesse e obtém o resultado.

#### **4.2. MOrFEu**

O MOrFEu é um ambiente para o trabalho cooperativo e, portanto, existem produções individuais onde as pessoas vão registrando seus escritos, tornando-se, portanto, um espaço apropriado para hospedar o sistema que será apresentado nesta dissertação.

A proposta de flexibilidade se fundamenta na possibilidade de transferir para os usuários a responsabilidade pela criação e a administração de suas produções individuais ou coletivas.

No MOrFEU, o elemento básico de autoria é a Unidade de Produção Intelectual (UPI), usada para registrar as produções dos usuários, sendo que cada UPI está ligada ao seu autor.

#### **Editor de Unidade de Produção Intelectual (UPI)**

No MOrFEU, segundo MENEZES, 2008, o elemento básico de autoria é a UPI, usado para registrar as produções dos usuários. A ferramenta básica é um editor de UPI. Cada UPI possui um autor, um título e um conteúdo (corpo). O conteúdo pode usar todo e qualquer elemento de representação digital suportado em html. Uma versão de UPI pode ser usada em diversas situações. Por exemplo, uma UPI que for enviar uma mensagem para um colega pode também ser usada para publicação em um fórum.

#### **4.3. Aplicações da Pesquisa**

Os objetivos definidos pela pesquisa para a construção de perfil restringe os textos como sendo aspectos que o usuário relata sobre a sua rotina diária. Não existe especificação de público alvo, todos os usuários que interagirem com o ambiente terão a definição de seu perfil construído conforme as suas produções.

Propomos a definição do perfil de pessoas sem restrições de foco específico. Pretendemos obter a melhor composição das informações que caracterizam uma pessoa, permitindo a utilização para a aplicação educacional, comércio ou qualquer outra aplicação específica que se possa imaginar. Desejamos que o perfil contenha

informações para ter um conhecimento prévio de uma pessoa sem ter a necessidade de conhecê-la.

#### **4.4. Conclusão**

Como pode ser observado no capítulo 3 e contextualizado no ambiente MOrFEu, são diversas as situações em que surge a necessidade de as características pessoais serem identificadas. Sendo o MOrFEu um ambiente flexível e passível de modelagem para atendimento de particularidades, concluímos que podemos ter uma relação de equivalência com muitos dos trabalhos apresentados no capítulo de trabalhos correlatos.

Apresentaremos a seguir, no capítulo 5, como as informações podem ser extraídas das UPIs para compor o perfil do usuário

## 5. PROPOSTA DE SOLUÇÃO

Para contextualizar nossas decisões, iniciamos apresentando as justificativas de cada um dos recursos adotados na proposta. O ambiente MOrFEu foi selecionado por ser um ambiente que atende às características de requisitos da proposta. Sendo ele um ambiente de controle de autoria, o que o destacou diante de outros ambientes foi a flexibilidade e, ainda, pelo fato de ser propriedade do grupo de estudo da Ufes para fins acadêmicos.

A decisão em trabalhar com PLN surgiu do propósito de identificar a contribuição que as técnicas de PLN podem apresentar para se obterem características pessoais com base em suas produções. Assim, também, a aplicação de agentes se deu pela necessidade de empregar inteligência e flexibilidade ao sistema.

A abordagem para solução aqui proposta se assenta sobre o ambiente MOrFEu, o qual contém todas as produções dos usuários. Essas produções são UPIs que ficam organizadas nos espaços individuais, vinculadas em diversos veículos de comunicação (VCom) a partir de suas postagens. Todas essas produções precisam ser organizadas para serem processadas pelo sistema. As produções de cada usuário revelam características de seu autor. Nossa abordagem propõe uma solução distribuída em agentes de softwares inteligentes.

A construção de perfil de usuários a partir de textos utilizando métodos linguísticos requer o enfrentamento dos desafios inerentes à manipulação da linguagem natural, além das particularidades da língua portuguesa. Então, a adoção de recursos que evidenciem os vínculos de dependência gramatical de cada palavra em seu contexto é fundamental para o êxito desta abordagem.

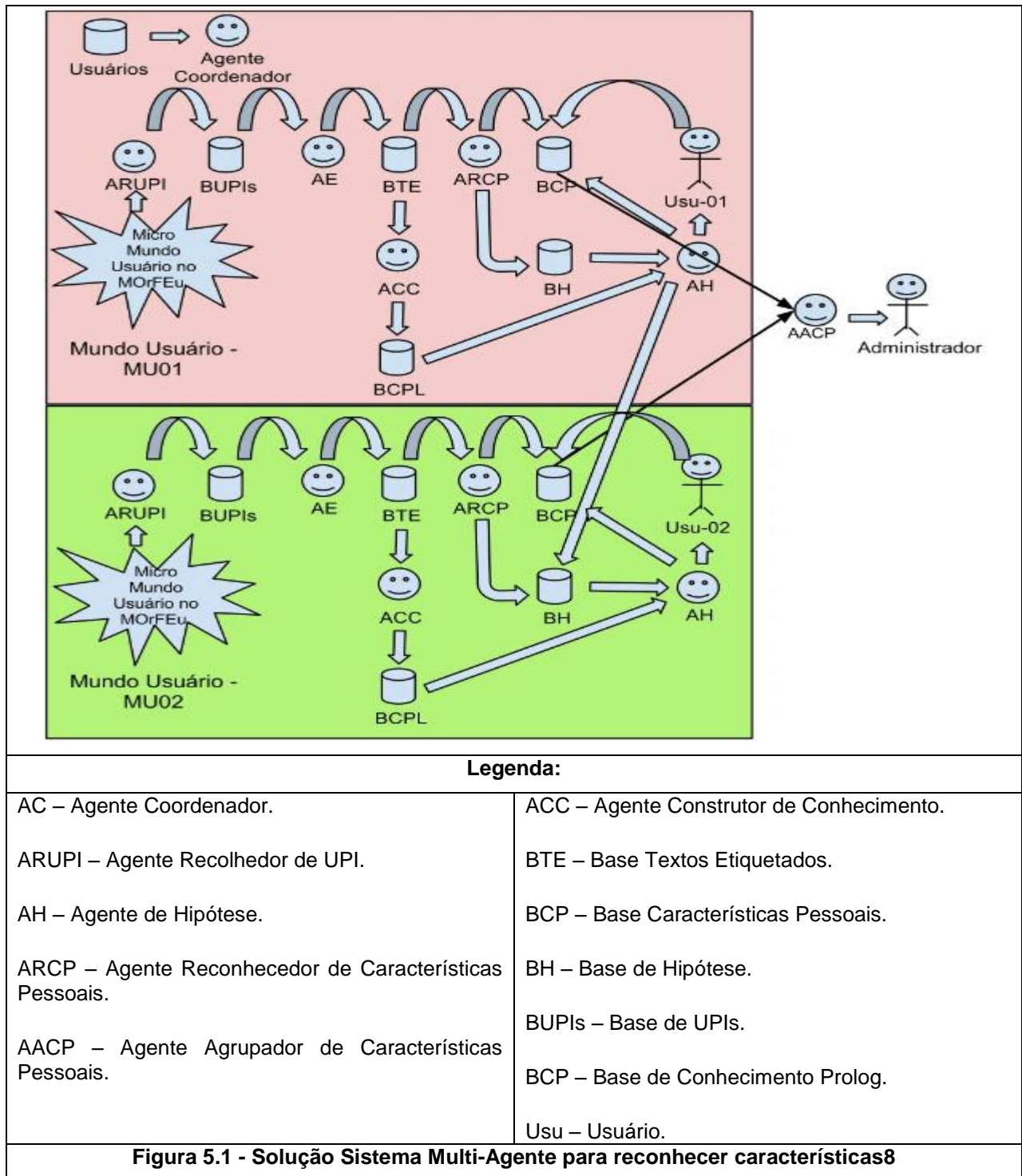
A abordagem se inicia com a obtenção das UPIs do usuário. Primeiramente, o conteúdo do documento é processado, com o objetivo de eliminar marcas de formatação para que o mesmo passe a conter somente texto puro. Em seguida, o conteúdo é separado em orações que, por sua vez, são divididas em *tokens*, ou seja, em palavras, números e sinais. A seguir, cada *token* é etiquetado com a etiqueta morfológica, sintática e semântica, deixando evidente a estrutura da oração.

A qualidade das aplicações que utilizam estes sistemas de processamento de texto é altamente dependente de uma perfeita ou completa análise com interpretação de textos. Assim, um grande desejo dos pesquisadores de texto com PLN é obter análise de texto de alta qualidade em níveis pragmáticos, obtendo as informações completas e recuperáveis.

Os textos obtidos a partir da ferramenta de processamentos de textos e etiquetagem são processados para a obtenção de características pessoais, características essas que são recuperadas por meio de padrões linguísticos que foram identificados como reveladores de características pessoais em textos português brasileiro na primeira pessoa. Os textos escritos por uma pessoa revelam características pessoais, expressas de forma explícita ou em casos em que o autor citar características pessoais de pessoa também de forma explícita.

### **5.1. Especificação do SMA**

Na Figura 5.1 apresentamos uma arquitetura multiagente elaborada como parte de nossa proposta de solução para aquisição de perfis. A proposta considera que, para cada usuário, será instanciada uma réplica da arquitetura multiagente padrão. Na Figura 5.1 podemos observar que, para cada usuário, existirá uma federação específica, denominada na Figura 5.1 de Mundo Usuário - MU.



A troca de mensagens entre os agentes acontece utilizando-se um quadro negro. As informações possuem um status e cada status define um agente que irá processar no texto. Assim que uma agente conclui o seu objetivo, ele altera o status para que outro agente processe o texto, realizando outra tarefa e passando de agente para agente, de forma sequencial, conforme demonstrado na Figura 5.1.

A seguir são descritos cada um dos componentes da arquitetura.

#### **Agente Recolhedor de UPI (ARUPI):**

Considerando os diversos veículos de comunicação, as mensagens precisam ser coletadas e agrupadas por autor, ficando na responsabilidade do agente ARUPI monitorar o ambiente e coletar as UPIs do seu usuário. As UPIs ditas neste contexto são unidades ou produções intelectuais, em formato de texto, rotuladas como UPIs pelo ambiente MOrFEu;

#### **Agente Etiquetador (AE):**

O agente etiquetador é responsável por realizar a análise lexical, morfológica e semântica, adicionando aos textos as etiquetas.

#### **Agente Reconhecedor de Características Pessoais (ARCP):**

O ARCP irá localizar os verbos e estes verbos normalmente revelam ações, de modo que essas serão possíveis características pessoais que serão validadas por uma base de fatos e uma base de padrões linguísticos que revelam características pessoais.

O ARCP fica monitorando os textos etiquetados e tentando identificar, por meio de verbos e padrões de textos, informações de características pessoais. Em algumas situações, a revelação de tais características pode não ser conclusiva, surgindo o levantamento de hipóteses. Isso ocorre quando o agente suspeita de uma característica em potencial, mas não possui todas as informações para concluir. Então, ele separa a sentença para que, em um ponto futuro, quando obtiver tal informação, seja concluída a revelação de tal característica.

O ARCP pode identificar hipótese de características de outro usuário. No caso em que um usuário comente sobre a característica de outra pessoa, esta informação é

entendida como uma hipótese, dependendo da confirmação da própria pessoa a respeito de tal característica. Porém esta confirmação só será obtida se a pessoa que foi citada como possuidora de alguma característica, for usuária do sistema. Em casos em que a pessoa ainda não é um usuário, a hipótese é registrada com a pendência ainda de que a pessoa se torne um usuário para, posteriormente, obter a validação de tal característica. Um exemplo de uma frase poderia ser “Carolina vai fazer o bolo”. Carolina pode não ser uma usuária ainda do sistema e ficaria pendente a confirmação dela.

Outra situação utilizada é a hipótese com pendência que seria o caso em que não se consegue identificar o nome do possuidor da característica. Exemplo: “Minha esposa vai fazer o bolo”. Não se sabe, por meio da frase, qual o nome da esposa para que possa ser verificado se tal pessoa é usuária do sistema e assim validar a informação realizando a pergunta para ela. Exemplo “Você sabe fazer bolo?” Para isso é registrada a hipótese da sentença e o AH fica encarregado de resolver tais pendências.

#### **Agente de Hipótese (AH):**

O AH fica constantemente avaliando as hipóteses e monitorando os fatos que são inseridos na base de conhecimento prolog (BCP). Assim que algum novo conhecimento é inserido, as hipóteses relacionadas aos usuários do novo conhecimento inserido são todas reprocessadas para verificar se dúvidas foram sanadas. Se algum fato inserido na BCP possibilitar a elucidação de uma hipótese, este fato será utilizado pelo AH.

O AH quando identifica uma hipótese referente a outro usuário do sistema, transfere esta hipótese para que seja tratada pela federação de seu usuário.

#### **Agente Construtor de Conhecimento (ACC):**

O agente construtor de conhecimento tem a função de registrar na base prolog todas as características do seu usuário para que o AH possa utilizar-se do conhecimento para sanar dúvidas.

Então, sempre que a pessoa revelar uma informação, esta informação será registrada na base de conhecimento. Imaginamos a seguinte frase:

“Minha esposa faz o almoço”.

Minha [meu] <poss 1S> **DET F S @>N #1->2**  
 esposa [esposa] <Hfam> **N F S @SUBJ> #2->3**  
 faz [fazer] <fmc> <mv> **V PR 3S IND VFIN @FS-STA #3->0**  
 o [o] <artd> **DET M S @>N #4->5**  
 almoço [almoço] <occ> **N M S @<ACC #5->3**

Pela análise semântica, é possível identificar que a palavra esposa trata de uma relação familiar do usuário identificada pela etiqueta <Hfam>, ficando registrado em lógica de primeira ordem.

Com esta frase, surge a hipótese de que uma terceira pessoa (não se sabe quem, ainda) faz almoço. Esta informação ficaria registrada como hipótese, quando for identificado qual o nome da esposa. Então o AH verificará se a esposa é uma usuária do sistema e, caso seja, irá realizar a pergunta se ela “faz almoço”, concluído assim a hipótese e atualizando no perfil da esposa que ela tem um marido com o nome de “João”. Desta forma, a federação de agentes da esposa poderá realizar aferições pendentes com relação ao seu marido caso existam.

Os agentes com relação ao seu marido ficam observando as várias postagens escritas pelo usuário, com o propósito de identificar qualquer padrão que revele uma possível informação a ser considerada a respeito do autor. Por exemplo, quando conhecemos uma pessoa, estamos sempre tirando conclusões sobre cada atitude da mesma: sua forma de vestir, andar, falar, sotaques, gírias, repetições, vocabulários, pessoas com as quais se relaciona, horários que realiza qualquer atividade, Isso tudo será alvo para o agente tentar extrair informação e gerar lebres/alertas.

### **Agente Agrupador de Características Pessoais (AACP):**

Diante das diversas características que uma pessoa pode possuir, surgem também diversas formas de obter o melhor proveito destas características em grupos. O objetivo do AAHP é permitir criar grupos de pessoas que definam um determinado número de características para atender uma necessidade. Esta é composta por diversos requisitos de características.



Por exemplo, existem pessoas com conhecimentos técnicos e específicos que precisam ser agrupados para a construção de um conhecimento mais amplo. Agrupando estas várias pessoas que possuem conhecimentos específicos e também distintos, obterá um grupo de pessoas com um potencial maior, além da possibilidade de repassarem conhecimentos uns para os outros, tornando-os acessíveis a um grupo maior de pessoas. A correta identificação das características e agrupamentos evitaria a repetição de pessoas similares em um mesmo grupo, pois esta característica pode estar sendo necessária em outro grupo.

## **5.2. Especificação do Perfil**

Para a extração das informações e constituição do perfil de cada usuário a partir dos textos por ele escrito, fez-se necessária a criação de regras linguísticas que permitem identificar cada tipo de informação que pode compor as características que se pretende para o perfil do usuário com suas particularidades linguísticas.

(...) sistemas em linguagem natural devem representar fatos sobre estruturas linguísticas, sobre a relação dessas estruturas com o mundo e sobre estruturas cognitivas de pessoas, assim como a necessidade de tratar as ações seguindo a noção de que as pessoas estão inseridas num padrão de interação (...) (WINOGRAD, 1981 apud WARNER, 1987, p. 81)

Warner (1987) cita Harris (1984 apud Warner, 1987) que ressalta não ser suficiente representar o conhecimento somente com estruturas estáticas, uma vez que a representação de conhecimento requer que uma estrutura interna de significados seja construída.

Para identificar as características das pessoas em cada categoria e assim compor o seu perfil, faz-se necessária a identificação gramatical e frasal dos textos e, posteriormente, uma análise destes textos processados. Para cada um dos conjuntos de características que compõem o perfil do usuário, existe um conjunto de regras que devem ocorrer no texto. A ocorrência das regras no texto é que as caracterizam como uma informação pertinente e de interesse à composição do perfil.

As regras que definem as características de interesse para o perfil são utilizadas pelo ARCP, sendo possíveis configurações para as regras: classificação morfológica (verbo,

substantivo, adjetivo, etc.), sintática, semântica ou junção de todas para atender a uma particularidade de interesse de característica contida no texto.

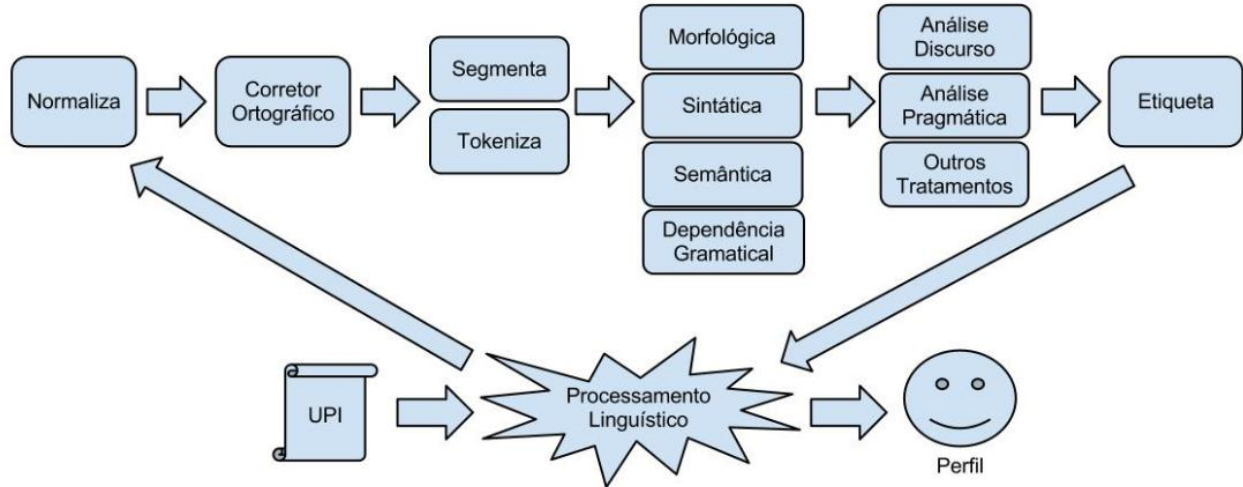
### **5.3. Processamento Linguagem Natural**

Quando algum conhecimento linguístico é considerado, a etiquetagem gramatical do texto é um dos passos iniciais. Um etiquetador gramatical (part-of-speech tagger) é um sistema que identifica, por meio da colocação de uma etiqueta (tag), a categoria gramatical de cada item lexical do texto analisado (BICK, 1998). Enquanto um etiquetador morfológico inclui informações sobre categorias morfológicas, como substantivo, adjetivo, etc., um etiquetador sintático acrescenta etiquetas indicando as funções sintáticas das palavras, como sujeito, objeto direto, etc.

Além da etiquetagem ou marcação gramatical, existe a etiquetagem semântica (VIEIRA 2000), que anexa informação relacionada ao significado, podendo indicar os papéis dos itens lexicais na sentença, como agente, processo e estado.

Propomos uma abordagem em que o conteúdo de um documento deve ser fragmentado até que suas tokens, de natureza linguística, tornem-se disponíveis. Posteriormente, as tokens são processadas, analisadas e etiquetadas revelando as características morfológicas, sintáticas e semânticas de cada token.

Equacionando o problema, projetamos um modelo de processo composto por um conjunto de atividades que se concatenam por meio de suas entradas e saídas, como ilustrado na Figura 5.2.



**Figura 5.2 - Processamento de Texto9**

A abordagem se inicia com a seleção da UPI que será processada. Primeiramente, o conteúdo do documento é processado com o objetivo de eliminar marcas de formatação para que o mesmo passe a conter somente texto puro. Em seguida, o conteúdo é separado em orações, que por sua vez, é dividido em tokens, ou seja, em palavras, números e sinais. A seguir, para cada token é identificado a sua classificação morfológica, sintática e semântica. Essas características são etiquetadas com a etiqueta morfológica (Part-of-Speech tag ou POS tag) correspondente, deixando evidente a estrutura de construção da oração.

São 14 as atividades do processamento de texto considerado para a obtenção de perfil:

- i. *Normalização de Texto;*
- ii. *Corretor Ortográfico;*
- iii. *Segmentação de Texto;*
- iv. *Tokenização;*
- v. *Análise Morfológica;*
- vi. *Análise Sintática.*
- vii. *Análise Semântica.*
- viii. *Sequência Gramatical*
- ix. *Análise de Discurso*

- x. *Análise de Pragmática*
- xi. *Etiquetação.*
- xii. *Segmentação Proposicional, Anáforas.*
- xiii. *Reconhecimento de Características Pessoais; e*
- xiv. *Construção do Perfil.*

As atividades são coordenadas de forma sequencial, do primeiro ao último, orquestradas de forma que, a cada atividade executada, seja produzido um único artefato que será tomado como insumo para a atividade seguinte, do documento original até a obtenção da informação para a composição do perfil.

### **Normalização de Texto**

A atividade é responsável por eliminar os marcadores de formatação existentes no conteúdo de um documento, tais como *tags*, estilos de fontes, metadados. A eliminação destas marcas permite a manipulação de arquivos em diferentes formatos de texto.

A existência de inúmeros formatos de documentos aliada à velocidade de surgimento de novas extensões e de atualizações de versão que não guardam compatibilidade com as versões anteriores são fatores complicadores para a atividade que depende, fundamentalmente, da definição de filtros apropriados para tratamento das diferentes codificações.

### **Corretor Ortográfico**

A correção ortográfica insere-se no problema mais geral de correção de textos escritos, que inclui, também, a correção da construção de frases e a correção do estilo. Temos, então, três níveis de manipulação de um documento. O primeiro, ortográfico, limita-se a controlar a correção das palavras isoladas, independentemente do seu uso. O nível gramatical, imediatamente acima, preocupa-se com problemas como os de concordância, pontuação e uso de preposições.

Por fim, o nível estilístico. Neste nível, trabalha-se com frases gramaticalmente bem construídas, mas que, por uma questão de estilo, deveriam ser corrigidas. Estas correções têm como objetivo tornar o texto mais legível, tornando as construções

gramaticais mais simples, ou reduzindo o tamanho das frases. A correção de estilo, pode envolver também a verificação da conformidade do texto com linhas de orientação previamente estabelecidas, como por exemplo, as elaboradas para a escrita de documentos técnicos.

### **Segmentação de Texto**

A manipulação de textos geralmente pressupõe a habilidade de dividi-lo em sentenças individuais (BIRD; KLEIN; LOPER, 2009, p. 112). A segunda atividade, denominada Segmentação de Texto, é justamente o mecanismo que provê a segmentação do conteúdo de texto puro em sentenças. Cada sentença limita um conjunto semântico mínimo para definição de uma proposição. A segmentação de textos é realizada por meio da identificação de caracteres finalizadores de sentenças, principalmente, dos sinais de pontuação. A dificuldade desta atividade está na correta associação dos sinais de pontuação ao fim de sentença, uma vez que tais sinais também podem demarcar abreviações de nomes, separação de dígitos em datas, horas, números de telefones e números ordinais (CIMIANO, 2006).

### **Tokenização**

A terceira atividade é a atividade de identificação de *tokens*, ou seja, de elementos de cada oração, inclusos neste conjunto: palavras, números e sinais de pontuação. Essa etapa, denominada de *Tokenização*, utiliza-se, basicamente, de sinais gráficos como espaços e algoritmos para o reconhecimento de entidades limítrofes de um *token*.

Segundo (CIMIANO, 2006), o uso dos espaços em branco nem sempre definem de forma adequada as fronteiras das palavras, sobretudo de palavras compostas e de nomes próprios. Por isso, é recomendada a utilização conjunta *tokenizador* juntamente com técnicas de reconhecimento de nomes de entidades.

### **Análise Morfológica**

A Morfologia é a parte da gramática que estuda as palavras de acordo com a classe gramatical a que ela pertence. Quando nos referimos às classes gramaticais, logo

sabemos que se refere àquelas dez, que são: substantivos, artigos, pronomes, verbos, adjetivos, conjunções, interjeições, preposições, advérbios e numerais.

### **Análise Sintática**

Enquanto o analisador léxico-morfológico lida com a estrutura das palavras e com a classificação das mesmas em diferentes categorias, o analisador sintático trabalha em nível de agrupamento de palavras, analisando a constituição das frases.

Segundo (GONZALEZ at LIMA, 2003), a análise sintática (parsing) é o procedimento que avalia os vários modos de como combinar regras gramaticais, com a finalidade de gerar uma estrutura de árvore que represente a estrutura sintática da sentença analisada. Se a sentença for ambígua, o analisador sintático (parser) irá obter todas as possíveis estruturas sintáticas que a representam.

O papel do processamento sintático varia em importância (NUNES [1999] apud GONZALEZ and LIMA, 2003). Ele, tradicionalmente, ocupa posição de destaque, com a semântica sendo considerada uma interpretação da sintaxe, mas, também, pode ser considerado em posição secundária, de acordo com os pesquisadores denominados semântico-gerativistas. Neste último caso, a sintaxe é uma projeção da semântica. Entretanto, qualquer que seja a visão adotada, o processamento sintático é uma etapa indispensável para viabilizar o processamento semântico que passamos a discutir a seguir.

### **Análise Semântica**

Enquanto a sintaxe corresponde ao estudo de como as palavras agrupam-se para formar estruturas em nível de sentença, a semântica está relacionada ao significado, não só de cada palavra, mas também do conjunto resultante delas. O processamento semântico é considerado um dos maiores desafios do Processamento de Linguagem Natural, pois se vincula, de um lado, com a morfologia e a estrutura sintática e, de outro lado em alguns casos, com informações da pragmática (SAINT-DIZIER [1999] apud GONZALEZ and LIMA, 2003).

Vejamos os exemplos “Me lambuzo todo chupando manga” e “Não posso sair com essa manga rasgada”. Na primeira frase a palavra “manga” será identificada como <fruta> e, na segunda frase, a palavra “manga” será identificada como <parte da roupa>.

O analisador semântico analisa o sentido das estruturas das palavras que foram reagrupadas pelo analisador sintático, uma vez que o analisador morfológico permitiu identificar estas palavras individualmente.

### **Sequência Gramatical**

Sequência Gramatical é uma teoria para estruturar sentenças "ambíguas" como “O policial viu a velha com o binóculo”. O exemplo supracitado não deixa claro quem estava de binóculo. Mostrando que a uma mesma sequência linear de termos podem corresponder diferentes estruturas sintagmáticas - ou seja, é um exemplo de que "Uma sequência gramatical é muito mais do que apenas uma sequência de elementos: é, entre outras coisas, uma hierarquia de constituintes" (PERINI, 2006 p.104). As teorias sintáticas têm a tarefa de descrever e explicar as hierarquias de constituintes que formam sequências gramaticais. Como vimos no início, a teoria gerativa entende a formação potencialmente infinita de sequências gramaticais hierarquicamente estruturadas como a característica central da linguagem humana. Essa teoria, em particular, portanto, toma para si a tarefa de descrever e explicar, não apenas um dado universo de sequências gramaticais, mas também o potencial de geração infinita de sequências gramaticais - algo como a "receita", ou o "algoritmo", da estruturação de constituintes.

### **Etiquetação.**

O etiquetador gramatical (ou pos tagger) é um sistema responsável por identificar, em uma sentença, cada um dos itens lexicais de cada uma das palavras.

A etiquetagem (JURAFSKY, 2000) é o processo de assinalamento de um marcador de classe gramatical (ou outro marcador ou ‘etiqueta’ de interesse) a cada palavra, num corpus. Esse processo corresponderia à ‘tokenização’, no processamento das linguagens de programação. A etiquetagem, como trata de linguagem natural, lida com

um número bem maior de situações de ambiguidade. A entrada para a etiquetagem é uma cadeia de itens lexicais e um conjunto específico de etiquetas; a saída é o conjunto de itens lexicais com a melhor etiqueta associada a cada item. Exemplo de etiquetagem utilizado neste trabalho pode ser visto na seção 6.3.3.1 “Especificação do Conceito Geral PLN”.

### **Segmentação Proposicional e Anáforas.**

Nesta etapa é realizada então a Segmentação Proposicional que separa as sentenças em proposições, de modo a facilitar as etapas de extração de conhecimento. Para realização desta fase, são analisadas as etiquetas provenientes da análise sintática e morfológica, de modo a identificar a existência de sentenças que possuem duas ou mais proposições, a fim de separá-las. Para separação dos textos em proposições, de forma a melhorar o desempenho do sistema ao extrair conhecimento, identificamos alguns tipos de sentenças em que pode haver mais de uma proposição. Veja o exemplo: “Vou consertar a geladeira e depois a bicicleta”. O sintagma verbal (vou consertar) está constituindo o significado para duas proposições sendo, “vou consertar a geladeira” e “vou consertar depois a bicicleta”.

Em VIEIRA (2001) é apresentado que, de acordo com a estrutura sintática de uma sentença, é possível estabelecer uma representação lógica correspondente, onde o verbo indica uma relação entre os argumentos expressos por sujeito e complemento verbal (objeto direto ou indireto).

BELTRAME (2011) apresenta uma solução para a segmentação de anáforas proposicionais. Como métodos para solução do problema, foi implementado o algoritmo proposto por KOWATA (2010) fazendo alguns ajustes e aplicando heurística proposta por FREITA (2005) que segue as premissas: o vértice sujeito anterior com o maior número de filhos terá a probabilidade maior de ser o sujeito de predicados anafóricos.

Outros processamentos poderiam ser adicionados para facilitar e melhorar a qualidade dos resultados para a obtenção do conhecimento e identificação de características



peçoais. Como exemplo de processos, podemos citar o processamento de análise de discurso e análise pragmática.

### **Reconhecimento de Características Pessoais.**

Por meio de um conjunto de padrões linguísticos baseados nas etiquetas, grupos linguísticos são reconhecidos em uma atividade que visa à identificação de elementos candidatos a características e complementos. Uma vez reconhecidas as características pessoais dos candidatos, busca-se a identificação do complemento, que é o restante da informação que contextualiza a característica.

Observou-se que as características que o autor revela possuir em seus textos ocorrem sempre vinculadas a um verbo, iniciando um contexto semântico com verbos na primeira pessoa do Singular ou Plural. Este padrão é base para a extração de características que constituem o perfil do usuário.

Há uma “lista de verbos” que são reveladores de características e de informações para a construção do perfil. Caso exista um verbo no texto, ele será analisado para verificar se tal verbo está na primeira pessoa Singular/Plural. Se essa condição básica for satisfeita, temos um verbo candidato à revelação de uma característica. A lista de verbos pode ser específica tanto quanto for a abordagem, assim como as “máscaras” e “tag semântica”, que forem especificadas para a identificação de uma característica pessoal. A composição destes conceitos (lista de verbos, máscaras e tag semântica) podem ser vistos na Tabela 6.5.

## 6. Uma Implementação para a Proposta

Ao longo do Capítulo 5, apresentamos uma abordagem computacional para a construção de perfis a partir de textos em português do Brasil. Neste Capítulo, relataremos o desenvolvimento de uma implementação realizada com a intenção de validar a factibilidade da proposta.

Na Seção 6.1, descrevemos as especificações técnicas do protótipo computacional denominado de *SysCAP (Systema Construtor Automático de Perfil)*. O *SysCAP* é composto por 3 módulos que materializam as atividades do processo de construção de perfis dos usuários.

Para assegurar o entendimento do funcionamento da ferramenta *SysCAP* no contexto da abordagem, detalhamos cada módulo em uma seção. Na seção 6.2 é apresentado o módulo editor de UPI.

Na Seção 6.3, demonstramos o módulo SMA, apresentando como a construção dos perfis é abordada pelo SMA e, em subseções, são apresentadas as especificações de cada agente que compõe o SMA, além das especificações e características consideradas sobre o processamento de linguagem natural.

Na Seção 6.4, apresentamos o módulo de interface administradora de perfil.

## 6.1. Especificação do Protótipo

Optamos por implementar uma representação do perfil dos usuários dividida em 6 subseções (habilidades, relações pessoais, profissão, comer, beber e esporte).

Escolheu-se criar as seis seções citadas acima, pela disposição de etiquetas semânticas, que identifica valores semânticos para tais ocorrências nos textos. A escolha também se deu pela impressão de ser coerente para exemplificar características pessoais, que poderiam compor parcialmente ou completamente valores para o perfil dependendo do propósito e aplicação desejados.

Dividimos a implementação do protótipo em três módulos sendo:

- 1 – Módulo Editor de UPI.
- 2 – Módulo do SMA.
- 3 – Módulo Administradora.

Na Figura 6.1 pode ser observado como ocorre a interação dos módulos para a solução implementada no protótipo.

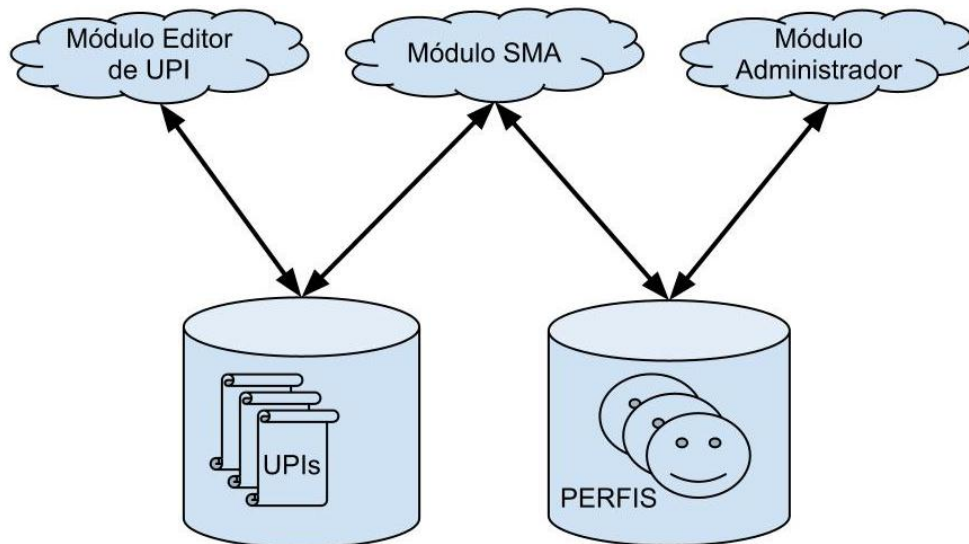


Figura 6.1 - Interação entre os Módulos10

### **6.1.1. Observações entre Proposta de Solução e Implementação do Protótipo.**

Neste tópico iremos fazer uma relação da proposta com o que foi implementado no protótipo para a validação, considerando os domínios de SMA, PLN.

#### **6.1.1.1. Implementado no Protótipo de SMA**

Considerando a proposta apresentada na Figura 5.1, foram implementados no protótipo os agentes AC, ARUPI, AH, ARCP. As bases de dados BTE, BCP, BH, BUPIs, foram implementadas, sendo cada uma das bases propostas uma tabela de um único banco de dados.

As funções desenvolvidas pelo agente AACP foram implementadas no módulo de administração para teste das possibilidades de agrupamentos e verificação das características, sendo implementado utilizando recurso de orientação e objeto e não agentes.

Os agentes e bases que não foram implementados por razões de tempo foram ACC e a BCP.

A comunicação entre os agentes foi implementada utilizando o método de quadro negro, deixando marcadas no banco de dados informações necessárias para estabelecer a coordenação entre os agentes. A leitura do ambiente é feita a cada 5 segundos, os agentes monitoram os seus ambientes, verificando se existe algo para ser processado.

O agente de hipótese infere hipóteses sobre as citações de terceiros referentes às seções de habilidades.

#### **6.1.1.2. Implementado no Protótipo de PLN**

Considerando a proposta apresentada na Figura 5.2, foram implementados no protótipo os processos: Segmentar, Tokenizar, Análise Morfológica, Análise Sintática, Análise Semântica, Dependência Gramatical e Etiquetagem. Como pode ser observado na Figura 6.2, os processos com um “X” em vermelho não foram implementados.

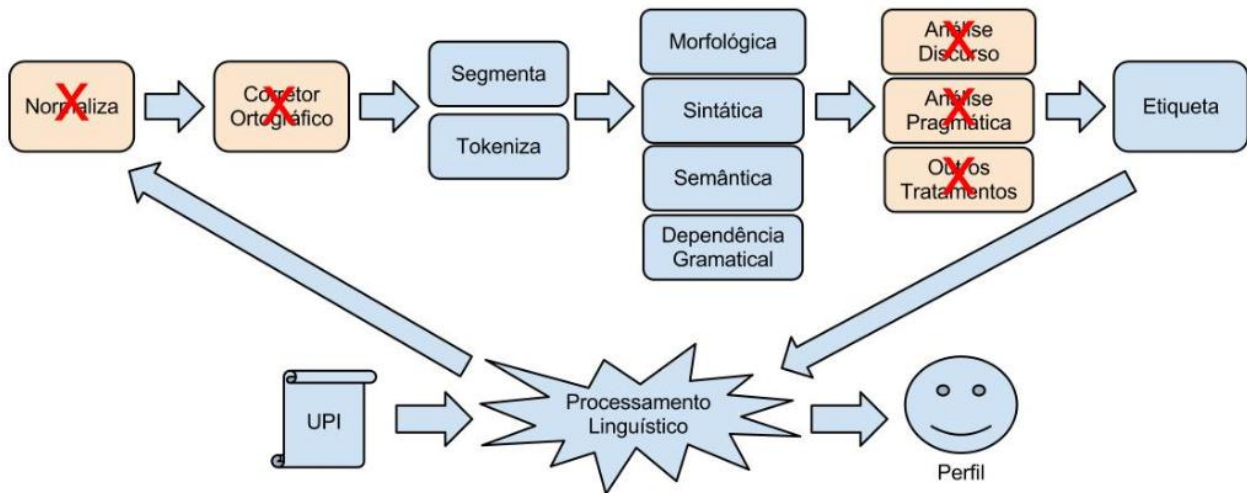


Figura 6.2 - Processamento de Texto Implementado no Protótipo. 11

Por questões de tempo, não foram implementados para os processos: análise de discurso, análise pragmática, análise de proposições e anáforas.

Sendo assim, consideramos que os textos tenham somente uma proposição e não tenham anáforas, estando normalizados e corretamente escritos.

## 6.2. Módulo Editor de UPI.

Para agilizar os testes do protótipo, optamos por trabalhar isolado da implementação atual do MOrFEu. Para isso, escrevemos um editor particular de UPIs. O módulo editor de UPI permite a construção e edição de UPIs e todas as suas produções textuais são escritas e apresentadas por meio deste.

Este módulo foi desenvolvido em PHP para prover recurso de acesso pela internet à edição de UPI. O usuário se autentica (login e senha) e é direcionado para o editor de UPI.

Veja na Figura 6.3 a interface, edição de uma coleção de UPIs por usuário.

[Sair](#) [Atualizar](#)

**Perfil**  
 Nome: Araldo  
 Login: araldo

**Confirmação de Habilidade**

Maikson escreveu que você é/sabe:  
**desenvolver**

**Postagens**

Publicar Texto:

2012-06-19 20:39:54	Araldo desenvolveu o protótipo em java.
2012-05-27 02:19:04	faço uniformes escolares de malha fria.

**Figura 6.3 - Módulo Editor de UPI12**

Na parte superior da Figura 6.3, é apresentada a “confirmação de características”. Nesta confirmação, são mostradas todas as características reveladas sobre o usuário logado, escritas nas produções de UPIs de outro usuário do sistema e identificadas pelo módulo SMA. São chamadas de hipótese, já que o sistema não conseguiu identificar com certeza se a pessoa possui realmente tal característica, dependendo da confirmação do usuário. O usuário pode confirmar ou negar as características. Caso o usuário confirme, a característica é adicionada ao seu perfil. Caso negue, ela é ignorada.

### **6.3. Módulo do SMA.**

O SMA é o principal módulo para a construção do perfil. Para o desenvolvimento deste protótipo, foi modelado o perfil, contendo um conjunto de características que chamamos de seções. As seções que compõem o perfil são: habilidades, comida, bebida, relações pessoais, profissão e esporte. Neste módulo utilizamos as especificações das informações que compõem o perfil, como sendo o norteador de nosso objetivo, utilizando como recurso, para alcançar o objetivo, as técnicas de PLN e as técnicas de modelagem de SMA.

Para facilitar o entendimento, são apresentadas as características de cada agente em subseções, permitindo organizar as manipulações necessárias que foram realizadas.

O módulo SMA do SysCAP foi implementado, utilizando a linguagem de programação java, banco de dados mysql. Os agentes foram implementados utilizando o framework JADE.

O SMA contém um agente coordenador, sendo que este agente coordenador tem como finalidade instanciar os agentes de cada usuário. Assim que o usuário é criado, o agente coordenador instancia toda a federação de agentes que estará manipulando suas UPIs e construindo o perfil de seu usuário.

Cada usuário criado no sistema recebe uma federação de agentes, que é composta por 4 agentes, sendo:

- Agente Recolhedor de UPI,
- Agente Etiquetador,
- Agente Recuperador de Características Pessoais e
- Agente Hipótese.

Observe a Figura 6.4, que descreve como é composto o protótipo SMA.

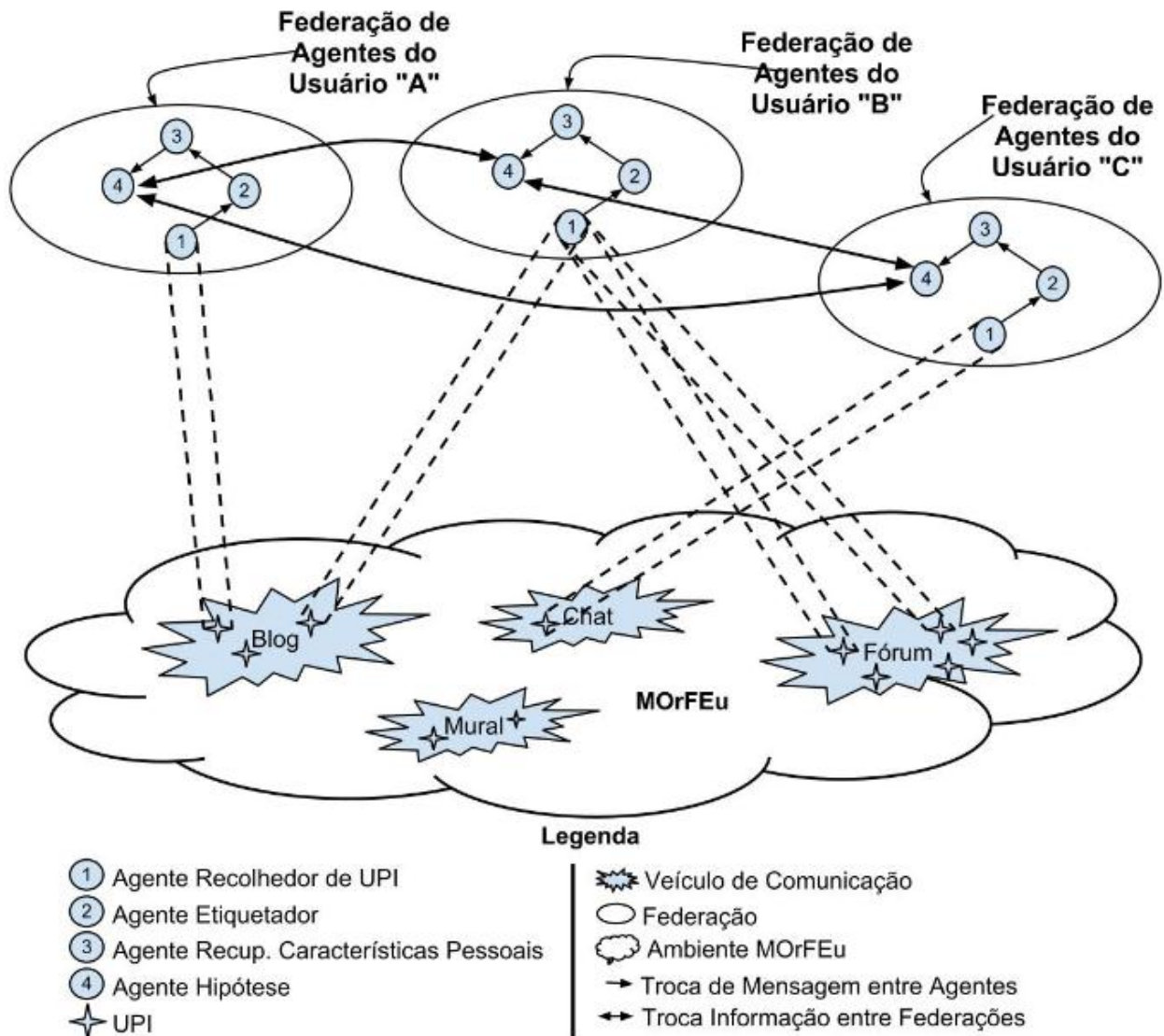


Figura 6.4 - SMA1314

Cada federação tem acesso à coleção de UPIs de seu “dono”, processando apenas as que já foram publicadas em algum veículo de comunicação.

### 6.3.1. Agente Recolhedor de UPI - ARUPI

Ao capturar a UPI, a ARUPI a marca como já capturada, para que a identifique e não a capture novamente. Desta forma, o agente fica monitorando constantemente tudo que é produzido pelo seu “dono” e, assim que o usuário publica uma nova UPI, ela é



copiada. As UPIs copiadas são processadas pelo próximo agente da federação, o agente etiquetador.

### **6.3.2. Agente Etiquetador - AE**

O AE utiliza-se da ferramenta de processamento linguístico, parser PALAVRAS VISL(BICK, 2000) para realizar todo o trabalho de etiquetação. O AE utiliza o parser PALAVRAS VISL disponível em <http://beta.visl.sdu.dk/visl/pt/parsing/automatic/dependency.php>. Esta ferramenta realiza a segmentação, tokenização, análise morfológica, análise sintática, análise semântica e a etiquetação, identificando todas as informações processadas do texto.

O AE fica monitorando a tabela de “postagemcapturada” e obtendo todas as postagens que ainda não foram identificadas como processada. O AE realiza a etiquetação e marca a postagem como processada, evitando retrabalhos. O resultado deste processamento é armazenado em tabela do banco de dados chamada de “textoetiquetado”, conforme pode ser visto na Figura 6.4, sendo esta tabela insumo para o processamento do próximo agente.

### **6.3.3. Agente Recuperador de Características Pessoais ARCP**

O ARCP é o agente com maior responsabilidade para o objetivo do trabalho. Ele atua sobre o texto etiquetado e extrai as características pessoais, identificando padrões de ocorrência na estrutura léxica frasal. Os padrões de ocorrência no texto são monitorados tanto para a identificação das características quanto dos complementos das características.

Os padrões de estruturas léxicas que são monitorados nos textos são configuráveis e chamados de máscara de complementos. Iremos apresentar o que é, e como funcionam estas máscaras de complementos: Em cada seção que compõe o perfil do usuário, existe um conjunto específico de regras linguísticas para identificar, em um texto, as características pessoais de cada seção e uma lista distinta de máscaras de complementos. A lista de máscaras de complemento é configurável pelo administrador do sistema, permitindo que ele realize inserção, alteração ou exclusão no conjunto de

máscaras de complementos. A seguir vamos entender a conceituação geral do processamento de linguagem natural a que os textos são submetidos, dando sustentação conceitual de PLN à implementação deste protótipo.

### 6.3.3.1. Especificação do Conceito Geral PLN.

Para a implementação do protótipo, optamos por utilizar o parser PALAVRAS VISL por ele disponibilizar alguns recursos que proporcionam comodidade ao propósito deste protótipo. Uma das funcionalidades que ele oferece é o tratamento “dependência gramatical (BICK, 2010)” que oferece as relações sintáticas das palavras em forma de etiquetas. Durante a geração das etiquetas de dependências gramaticais, ele realiza o tratamento de proposições. Outra funcionalidade que influenciou na escolha desta ferramenta é a identificação de papéis semânticas (BICK, 2007), oferecendo 157 classes de tags semânticas (VISL, 2012), identificando valores semânticos e considerando o contexto.

Os padrões de estrutura frasal dos textos são obtidos por meio das etiquetas adquiridas pelo parser PALAVRAS VISL. O VISL recebe uma frase e devolve uma etiquetagem da frase, com o seguinte formato:

frase = <palavra 1> <palavra 2> <palavra 3> . . . <palavra n>

onde, para cada palavra temos:

<palavra 1><etiquetagem 1>

<palavra 2><etiquetagem 2>

<palavra 3><etiquetagem 3>

...

<palavra n><etiquetagem n>

Onde, para cada etiquetagem temos:

<etiquetagem 1> [lematizado] <secondary tags> <word class tags> <inflection tags>  
<syntactic tags> <dependency grammar>

Onde:

lematizado é a palavra em sua forma básica;

<secondary tags>, contém principalmente etiquetas léxicas, classe de palavra e função sintática;

<word class tags>, contém principalmente as classificações morfológicas das palavras;

<inflection tags>, contém as etiquetas de sexo, número, caso, pessoa, modo, formas nominais;

<syntactic tags>, contém as etiquetas sintáticas.

<dependency grammar>, contém as dependências gramaticais que constitui #<origem> -> <destino>.

Veja no apêndice “A” quais as etiquetas finais para <secondary tags> <word class tags> <inflection tags> <syntactic tags>.

As etiquetas dos vínculos de dependência gramatical identificam a sequência em que as palavras estão organizadas para compor o significado semântico da frase. Vejamos um exemplo na frase “vou consertar a geladeira”, veja como fica a frase etiquetada na Figura 2.15.

```
vou [ir] <fmc> <aux> V PR 1S IND VFIN #1->0
consertar [consertar] <vH> <mv> V INF #2->1
a [o] <artd> DET F S #3->4
geladeira [geladeira] <Lh> <furn> N F S #4->2
```

**Figura 6.5 - Exemplo de Etiquetas15**

Vamos observar as etiquetas que sucedem o símbolo ‘#’. Na primeira linha temos “vou [ir] <fmc> <aux> V PR 1S IND VFIN #1->0”. O primeiro número indica o número da linha atual, e o segundo número, o que sucede o sinal de “->” indica a que palavra existe um vínculo de dependência gramatical, então podemos observar que esta primeira linha não tem vínculo de dependência gramatical com nenhuma outra palavra, porém outras palavras têm vínculo de dependência gramatical com ela. Vejamos as etiquetas da segunda palavra do texto “consertar [consertar] <vH>

<mv> **V INF #2->1**”, podemos observar, por meio da etiqueta, que a palavra da linha #2 tem um vínculo de dependência gramatical com a palavra da primeira linha. Para manter a continuidade do vínculo de dependência gramatical, vamos analisar a quarta palavra da frase e depois analisaremos a terceira palavra, “geladeira [geladeira] <Lh> <furn> **N F S #4->2**”. Percebemos assim que a linha 4 tem um vínculo de dependência gramatical com a linha 2 e, por último, podemos observar a terceira palavra do texto “a [o] <artd> **DET F S #3->4**” que continua com o vínculo por indicar que a linha 3 tem um vínculo de dependência gramatical com a linha 4.

Para facilitar o entendimento, vamos aplicar um filtro nas informações das etiquetas do texto e organizá-las em uma tabela. Veja a Tabela 6.1 .

**Tabela 6.1 - Análise do Vínculo de Dependência Gramatical, Sequência Texto. 5**

<b>Vínculo de Dependência Gramatical</b>	de	#1->0	#2->1	#3->4	#4->2
<b>Classificação Léxica</b>		V	V	DET	N
<b>Palavra</b>		vou	consertar	a	Geladeira

Na Tabela 6.1 temos, na primeira linha, os indicadores de vínculo de dependência gramatical; na segunda linha, a classificação léxica das palavras; e na última linha, as palavras da frase. A ordem em que as palavras ocorrem nas colunas da tabela estão organizadas conforme foram escritas. A Tabela 6.2 está organizada seguindo a ordem do vínculo de dependência gramatical.

**Tabela 6.2 - Análise do Vínculo de Dependência Gramatical, Sequência Gramatical. 6**

<b>Vínculo de Dependência Gramatical</b>	de	#1->0	#2->1	#4->2	#3->4
--	----	-------	-------	-------	-------

Classificação Léxica	V	V	N	DET
Palavra	vou	consertar	geladeira	A

Realizando uma análise das etiquetas ordenadas conforme o vínculo de dependência gramatical indicado nas etiquetas, tem-se a sequência da estrutura léxica da frase como sendo V V N DET. Respectivamente são verbos (V) auxiliar “vou”, seguido do verbo principal “consertar”, sucedendo o verbo(V) principal o substantivo(N) “geladeira” e por último a determinante (DET) “a”.

Pode ser identificada a questão de verbo auxiliar ou verbo principal pela etiqueta. Veja na Figura 6.5 a ocorrência da etiqueta <aux> na primeira linha “vou [ir] <fmc> <aux> **V PR 1S IND VFIN #1->0**”. Esta etiqueta caracteriza o verbo como sendo verbo auxiliar, já a etiqueta <mv> caracteriza a ocorrência de um verbo principal, o que ocorre na segunda linha da frase “consertar [consertar] <vH> <mv> **V INF #2->1**”.

A informação de verbo auxiliar e verbo principal é importante para esse trabalho. Como as características pessoais estão sempre vinculadas ao verbo principal da oração, então faz-se necessário obter o verbo principal da oração e verificar a pessoa do verbo. Se o verbo estiver na primeira do plural ou do singular, significa que está relacionado com o autor.

Para obter a informação de que o verbo principal contém um vínculo com o autor, verificamos se a pessoa do verbo é a 1º pessoa do singular ou do plural. A indicação de pessoa a que se refere a frase pode estar nas etiquetas do verbo principal ou nas etiquetas do verbo auxiliar.

No Exemplo apresentado na Figura 6.5, podemos observar a etiqueta 1S na primeira linha “vou [ir] <fmc> <aux> **V PR 1S IND VFIN #1->0**”. Esta etiqueta caracteriza o verbo como sendo primeira pessoa do singular.

A etiqueta que está entre colchetes é a palavra lematizada no exemplo apresentado na Figura 6.5. Podemos observar a etiqueta na primeira linha aplicada sobre o verbo vou, o que resultou na etiqueta do verbo lematizado, sendo [ir]. Para as palavras utilizadas

para formar as características pessoais assim como os complementos são sempre utilizadas as etiquetas lematizadas das palavras.

O complemento do verbo é uma informação que pode ser configurada. Esta configuração é feita realizando a inserção de estrutura léxica das palavras que se deseja obter como complemento do verbo principal.

Remetendo-nos ao exemplo apresentado na Figura 6.5, percebemos que o autor tem a habilidade de consertar geladeira que é organizada como sendo a habilidade: consertar e; complemento: geladeira. Para que o complemento geladeira seja identificado como uma informação de interesse para complementar a habilidade, é necessário que esteja inserida a máscara de complemento N, que indica o interesse em obter o substantivo que sucede um verbo principal que está na primeira pessoa.

Podemos imaginar a necessidade de ter um complemento que capture uma informação maior e mais significativa para o verbo que identifica uma característica para uma das seções, considerando a configuração das máscaras de complemento como sendo:

- N
- N N
- N ADJ
- N ADJ N
- N PRP N ADJ

Durante o processamento dos textos etiquetados como resultado, vamos tendo o perfil do usuário sendo definido com suas características em suas respectivas seções.

#### **6.3.3.2. Constituição do Perfil.**

Inicialmente o ARCP obtém os textos etiquetados e faz a separação dos seguimentos de sentenças dos textos. A separação das sentenças em um texto com várias sentenças é realizada por meio da etiqueta </s> que indica fim de sentença.

Cada sentença é submetida para a identificação das possíveis características pessoais que compõem o perfil, e este é composto por seções que organizam as características

pessoais do usuário. Para este protótipo, foram implementadas as seguintes seções: habilidades, comida, bebida, relações pessoais, profissão e esporte. Toda característica é composta por dois itens: característica e complemento.

Temos seis seções (habilidades, comida, bebida, relações pessoais, profissão e esporte) e a identificação de características citadas por terceiro, o que gera uma hipótese de característica para outro usuário.

Complemento pode ser uma ou mais palavras que contextualizam a característica e possivelmente complementa o sentido da característica.

A seguir, esclareceremos todos os processos realizados pelo agente para identificar as características de cada seção. Inicialmente veremos com detalhes o processo para a seção de habilidade.

Em cada seção existe a necessidade de regras específicas para a correta identificação da característica (verbo) e a configuração das máscaras de complementos.

#### **6.3.3.2.1. Seção Habilidades Pessoais.**

Para identificar as características de habilidades, o ARCP, inicia obtendo o texto etiquetado, seguido de uma busca por verbos. Quando ele encontra um verbo, realiza toda a verificação, se este verbo atende as regras de habilidades.

Para atender as regras de habilidades o verbo deve: ser um verbo principal; estar na primeira pessoa e não possuir palavra de negação no vínculo de dependência gramatical. Atendendo a necessidade básica da característica habilidade, busca-se agora o complemento da habilidade.

Para se obter o complemento da habilidade, percorrem-se os vínculos de dependência semântica aos quais o verbo de habilidade está incluso, podendo ser mais de um vínculo. Percorre-se cada vínculo montando a estrutura frasal de classificação léxica do texto etiquetado. Obtendo esta estrutura, realiza-se a verificação para ver se existe uma máscara de complemento que case perfeitamente com o início da estrutura léxica da sentença. Para isso, testam-se todas as máscaras de complemento da lista até encontrar uma. Inicia-se a comparação pelo item da lista que contém maior número de

termos no complemento. Percebemos que os complementos com maior número de termos resulta em um complemento mais contextualizado.

Vamos analisar a sentença “faço uniformes escolares de malha fria”.

```

faço [fazer] <fmc> <mv> V PR 1S IND VFIN @FS-STA #1->0
uniformes [uniforme] <cloH> N M P @<ACC #2->1
escolares [escolar] <nh> <np-close> ADJ M P @N< #3->2
de [de] <np-close> PRP @N< #4->2
malha [malha] <mat-cloth> <cloH> <tool> <activity> <ac> N F S @P< #5->4
fria [frio] <jh> <np-close> ADJ F S @N< #6->5

```

**Figura 6.6 - Etiquetação Com 2 Vínculos De Dependência Gramatical16**

Considerando o algoritmo apresentado na Figura 6.7 - Algoritmo Seção de Habilidade17 para análise e obtenção das características de habilidade.

```

Vetor token[]= sentenca.cortar("<br/>")
Enquanto( token.temProximo) faça
  Se ( token.conten("<mv>") ) então
    Se( percorreViculoProcurando(token, sentença,"1ª Pessoa")=Verdadeiro) então
      Se( percorreViculoProcurando(token , sentença,"Negação")=Falso) então
        mascaraTexto = montarMascara( token, sentença)
        Vetor mascarasAceita[] = obterMascaraHabilidade()
        Se ( comparar(mascarasAceita, mascaraTexto))então
          complemento = obterComplemento(token, sentença, mascaraTexto)
          inserirCaracteristica(token, complemento)
        fimSe
      fimSe
    fimSe
  fimSe
FimEnquanto

```

**Figura 6.7 - Algoritmo Seção de Habilidade17**

Ao analisarmos esta frase, podemos perceber que ela revela algumas características de habilidades. Ao processar esta frase pelo parser PALAVRAS, obtivemos as etiquetas com os vínculos semânticos, como pode ser observado na Figura 6.6. Tais etiquetas revelam existir dois vínculos de dependência gramatical relacionados ao verbo principal “faço”. Então iremos organizar duas tabelas de vínculo semântico, pois as etiquetas se abrem em dois caminhos de vínculos semânticos a partir do verbo de interesse “faço”. Vejamos a Tabela 6.3 e Tabela 6.4.



Tabela 6.3 - Análise do Primeiro Vínculo de Dependência Gramatical7

<b>Vínculo de Dependência Gramatical</b>	#1->0	#2->1	#3->2
<b>Classificação Léxica</b>	V	N	ADJ
<b>Palavra</b>	faço	uniformes	escolares

Na Tabela 6.3, podemos identificar a revelação da característica de habilidade de fazer uniformes escolares. Se considerarmos que temos cadastrado o interesse em obter complemento de N ADJ, teremos a captura da seguinte característica de habilidade: fazer; complemento: uniforme escolar.

Tabela 6.4 - Análise do Segundo Vínculo de Dependência Gramatical. 8

<b>Vínculo de Dependência Gramatical</b>	#1->0	#2->1	#4->2	#5->4	#6->5
<b>Classificação Léxica</b>	V	N	PRP	N	ADJ
<b>Palavra</b>	faço	uniformes	de	malha	fria

Como pode ser visto nas Tabelas 6.3 e Tabelas 6.4, as duas características podem ser capturadas pelo ARCP, pois ele realiza a análise considerando as etiquetas não tendo limite no número de vínculos que for criado para cada sentença. Para que o/os vínculos sejam capturados como complementos de características, basta que determinada máscara de complemento esteja configurada. Para capturar o complemento, ele inicializa testando o complemento com maior número de termos e passando para a próxima máscara de complemento, até encontrar uma máscara que tenha todos os elementos, ocorrendo na ordem em que foi obtida no texto. Caso exista o casamento idêntico da máscara, obtém-se o complemento da habilidade.

### 6.3.3.2.2. Seção de Comidas.

O ARCP tem como objetivo capturar as características pessoais do autor, tratando-se das coisas que ele come ou gosta de comer.

A metodologia de processo do texto se aplica da mesma forma. Alguns ajustes de regras e complementos vão identificar todas as particularidades.

Para atender as regras da seção de comida, o verbo deve estar na primeira pessoa e constar na lista de verbos de comida. Caso o verbo não possua etiqueta de pessoa, verifica-se se existe a etiqueta de pessoa em algum verbo vinculado semanticamente ao verbo que consta na lista de verbos da seção de comida.

O administrador pode adicionar ou remover verbos que sejam de interesse para revelar características de comida.

Outra regra que necessita ser atendida para ser uma característica de comida, é a ocorrência de algo comestível. Então surge a necessidade de verificar se existe alguma palavra com etiqueta de “<food>” no vínculo. Esta tag de <food> é adicionada pelo parser PALAVRAS que identifica as coisas que são comestíveis.

Atendendo estas necessidades básicas de comida, busca-se, agora, o complemento da característica de comida. O processamento do complemento para a característica de comida é o mesmo que os das características de habilidade, porém a lista de máscaras de complementos de comidas é outra.

Vejamos um exemplo na frase “gosto de comer moqueca de cação”, veja como fica a frase etiquetada na Figura 6.8.

```
gosto [gostar] <vH> <fmc> <mv> V PR 1S IND VFIN #1->0
de [de] PRP #2->1
comer [comer] <vH> <mv> V INF #3->2
moqueca [moqueca] <food-h> N F S #4->3
de [de] <np-close> PRP #5->4
cação [cação] <Aich> N M S #6->5
```

Figura 6.8 - Análise Etiqueta Comida18

Ao analisarmos esta frase,

podemos perceber que ela revela que a pessoa gosta de moqueca. Podemos imaginar que na lista de verbos de interesse para a seção de comida tenha os verbos: gostar, comer, degustar, alimentar, deliciar, etc.

Com esta configuração para a lista de verbos de comida permite obter duas formas distintas para a comida moqueca. Considerando o verbo gostar seria necessário que na lista de complementos tenha “PRP V N” ou “PRP V N PRP N” que se obteria a característica de comida: gostar; e respectivamente o complemento: “de comer moqueca” ou “de comer moqueca de cação”, caso existisse as duas máscaras na lista, ele pegaria a máscara com maior número de termos.

#### **6.3.3.2.3. Seção de Bebidas.**

O ARCP tem como objetivo capturar as características pessoais do autor se tratando das coisas em que ele bebe ou gosta de beber.

A metodologia de processo do texto se aplica da mesma forma da de comida com a única exceção de que se trata de bebida agora, então, a regra que necessita ser atendida para ser uma característica de bebida, é a ocorrência de algo bebível, então surge a necessidade de verificar se existe alguma palavra com etiqueta de “<drink>” no vínculo, esta tag de <drink> identifica as coisas que são bebíveis.

Para esta seção também existe uma lista de verbos reveladores de características de bebida, assim como a lista de máscaras de complementos de características de bebida.

#### **6.3.3.2.4. Seção de Relações Pessoais.**

O ARCP tem como objetivo capturar as relações pessoais do autor, identificando qual a característica e complemento, a característica deve conter as informações do tipo: pai, mãe, irmão, amigo, colega, etc; o complemento conter algo que identifica quem é a pessoa que tem tal característica, então no complemento deve conter os nomes próprios das pessoas.

A regra principal para levantar indício de uma relação pessoal é a identificação da tag semântica <Hfam>, a tag <Hfam> identifica todas as palavras que são do tipo de relação pessoal.

No complemento para esta seção existe uma regra semântica de validação, que consiste em identificar a tag léxica PROP que identifica ser nome próprio e a tag <hum> que significa ser humano, na Figura 6.11 podemos observar a ocorrência de tais tags.

Vejamos um exemplo na frase “tio Irineu chegou”, veja como fica a frase etiquetada na Figura 6.9.

```
meu [meu] <poss 1S> DET M S #1->2
tio [tio] <Hfam> N M S #2->4
Irineu [Irineu] <hum> <np-close> PROP M S #3->2
chegou [chegar] <fmc> <mv> V PS 3S IND VFIN #4->0
```

**Figura 6.9 - Análise Etiqueta Relação Pessoal19**

Ao analisarmos esta frase etiquetada na Figura 6.9, seguindo o algoritmo apresentado na Figura 6.10, podemos perceber que ela revela que o autor possui um tio, e o nome desse tio é Irineu, ficando a característica: tio e o Complemento: Irineu.

```
Vetor token[]= sentenca.cortar("<br/>")
Enquanto( token.temProximo) faça
  Se( token.contem ("<hfam>") ) então
    Se( percorreViculoProcurando(token,sentença,"possecivo 1 pessoa")=Verdadeiro)então
      Se( percorreViculoProcurando(token,sentença, "1ª Pessoa" )=Verdadeiro) então
        Se( percorreViculoProcurando(token,sentença,"<hfam>")=Verdadeiro)então
          Se( percorreViculoProcurando( token, sentença,"PROP")=Verdadeiro)então
            Se( percorreViculoProcurando( token, sentença,"<hum>")=Verdadeiro)então
              Se( percorreViculoProcurando( token, sentença,"Negação")=Falso)então
                mascaraTexto = montarMascara( token, sentença)
                Vetor mascarasAceita[] = obterMascara("SeçãoRelaçãoPessoal")
                Se ( comparar(mascarasAceita, mascaraTexto))então
                  complemento = obterComplemento(token, sentença, mascaraTexto)
                  inserirCaracteristica(token, complemento,"SeçãoRelaçãoPessoal")
                fimSe
              fimSe
            fimSe
          fimSe
        fimSe
      fimSe
    fimSe
  fimSe
FimEnquanto
```

**Figura 6.10 - Algoritmo Seção Relação Pessoal 20**

### 6.3.3.2.5. Seção Profissão

O ARCP tem como objetivo capturar as características profissionais do autor, identificando qual a característica e complemento. A característica deve conter as informações do tipo: ser; o complemento conter algo que identifica qual é a profissão que possui.

A metodologia de processo do texto se aplica da mesma forma que o da seção de habilidade, com a necessidade da tag de <Hprof> ocorrendo no complemento da sentença. A tag <Hprof> identifica todas as palavras que indicam uma profissão, exemplo: cantor, apresentador, pedreiro, professor, administrador, bombeiro, policial, etc.

Vejam um exemplo na frase “Eu sou professor de matemática”. Veja como fica a frase etiquetada na Figura 6.11 1.

```
eu [eu] PERS M 1S NOM #1->2
sou [ser] <vK> <fmc> <mv> V PR 1S IND VFIN #2->0
professor [professor] <Hprof> N M S #3->2
de [de] <np-close> PRP #4->3
matemática [matemática] <domain> N F S #5->4
```

Figura 6.11 - Análise da Etiqueta <Hprof> Profissão<sup>21</sup>

Ao analisarmos esta frase etiquetada na Figura 6.11 podemos perceber que ela revela que o autor é professor de matemática, ficando a característica: ser e o Complemento: professor de matemática, dependendo da configuração da máscara de complemento de profissão.

### 6.3.3.2.6. Seção Esportes

O ARCP tem como objetivo capturar as características de esportes do autor, identificando qual a característica e complemento. No campo de característica deve conter um verbo; o complemento deve conter algo que identifica qual é o esporte. No complemento teremos nomes de esportes, exemplo: nadar, correr, futebol, vôlei, tênis, basquete, etc.

A metodologia de processo do texto se aplica da mesma forma que o da seção de habilidade, com a necessidade da tag de <sport> ocorrendo no complemento da sentença. A tag <sport> identifica todas as palavras que indicam esporte. Vejamos exemplo da ocorrência da tag de esporte na Figura 6.12, pela qual foi processada a frase “eu jogo futebol de campo”.

```
eu [eu] PERS M/F 1S NOM #1->2
jogo [jogar] <vH> <fmc> <mv> V PR 1S IND VFIN #2->0
futebol [futebol] <sport> N M S #3->2
de=campo [de=campo] <pp> ADJ #4->3
```

**Figura 6.12 - Análise Etiqueta <sport> Esporte22**

Ao analisarmos esta frase etiquetada na Figura 6.11, podemos perceber que ela revela que o autor possui a característica de jogar futebol, ficando a característica: jogar e o Complemento: futebol de campo, dependendo da configuração da máscara de complemento de esporte.

#### **6.3.3.2.7. Hipóteses**

Outro objetivo deste agente é identificar as características que o autor revela a respeito de outro usuário do sistema, obtendo hipóteses. As hipóteses são encontradas quando ocorre a citação no texto de alguma característica que outro usuário do sistema possui.

Vejamos um exemplo na frase “João fez uniformes escolares de malha fria”. Veja como fica a frase etiquetada na Figura 6.13.

```
João [João] <hum> PROP M S #1->2
fez [fazer] <fmc> <mv> V PS 3S IND VFIN #2->0
uniformes [uniforme] <cloH> N M P #3->2
escolares [escolar] <nh> <np-close> ADJ M P #4->3
de [de] <np-close> PRP #5->3
malha [malha] <mat-cloth> <cloH> <tool> <activity> <ac> N F S #6->5
fria [frio] <jh> <np-close> ADJ F S #7->6
```

**Figura 6.13 - Análise Hipótese23**

Ao analisarmos a Figura 6.13, podemos perceber que as características reveladas na frase exemplo se tratam de características de outra pessoa diferente da do próprio autor.

1. Para que o ARCP perceba que as características reveladas são de outra pessoa e que esta pessoa é um usuário do sistema, existe um conjunto de regras que precisam ser atendidas. Vejamos cada uma:
2. Possuir um verbo que consta na lista de características de interesse.
3. O verbo ou algum outro verbo que consta no vínculo estar na 3ª pessoa.
4. O vínculo que contenha a classificação léxica de PROP possuir a classificação semântica de <hum>; e se esta palavra for um nome próprio de humano, este nome deve constar na base de usuários.
5. No vínculo não constar palavra de negação.
6. Possuir uma máscara de complemento de características pessoais, a mesma lista de máscara e regras de características do próprio autor.

Identificando no texto etiquetado todas estas restrições, o ARCP insere tal característica na base de hipótese do usuário citado. Esta hipótese será validada pelo agente de hipótese de sua federação de agentes.

#### **6.3.3.2.8. Agente Hipótese AH**

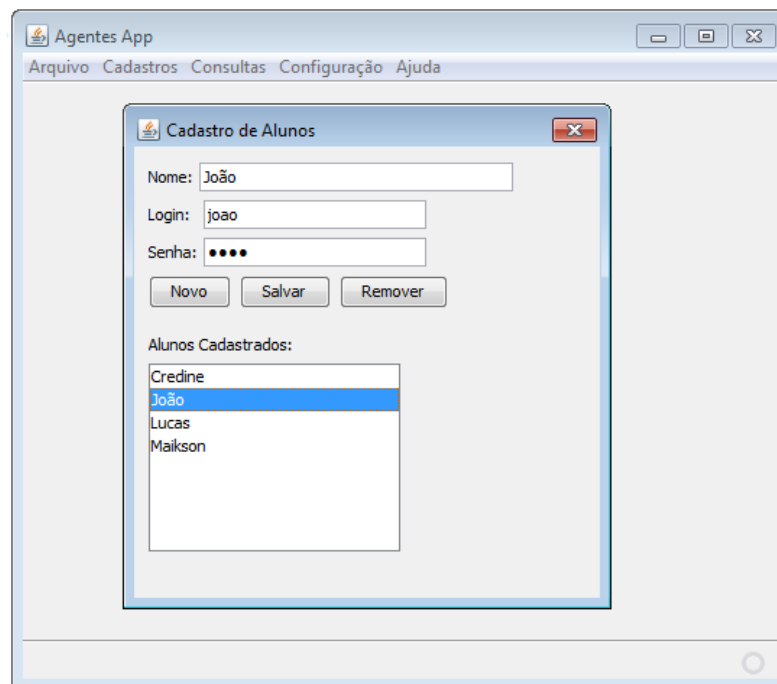
O AH tem a função de realizar a pergunta para o usuário que foi citado possuir tal habilidade. A pergunta é realizada por meio da interface de edição de UPI. Realizando a pergunta ao usuário, se tal hipótese é verdadeira ou não, se o usuário validar a hipótese, tal hipótese é inserida como habilidade. Caso seja negada a hipótese, é marcada como negada e passa a não mais ser exibida.

#### **6.4. Módulo da Interface Administradora de Perfis.**

Este módulo tem 4 funcionalidades principais: cadastro de usuários, configuração de informações para o SMA, consultar/visualizar perfil individual, consulta/realizar agrupamentos.

Vamos apresentar cada uma das funcionalidades.

**Cadastro de Usuário:** Para o propósito do protótipo, o cadastro do usuário é simples, possuindo somente os campos de nome, login e senha, como pode ser visto na Figura 6.14.



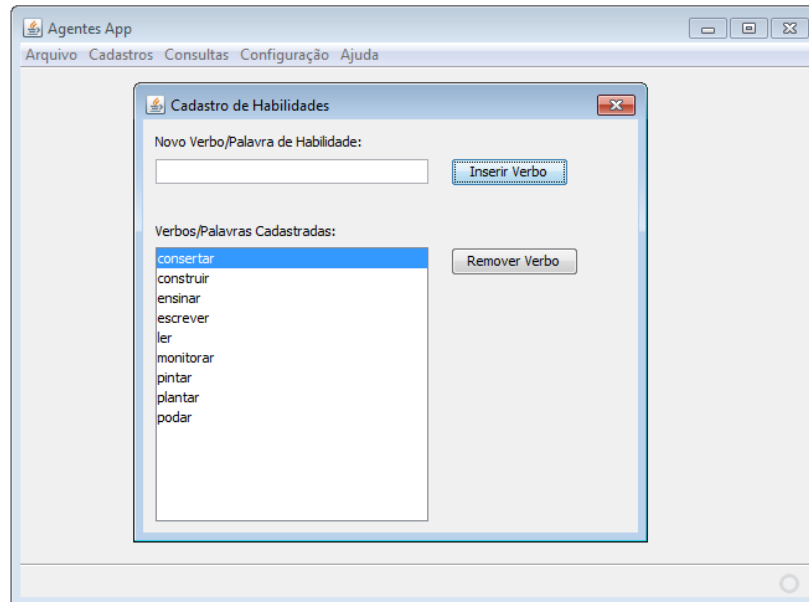
**Figura 6.14 - Imagem Tela de Cadastro. 24**

Para realizar o cadastro conforme apresentado na Figura 6.14, clica-se em novo, informam-se os valores e clica-se em salvar. Assim, o usuário recém cadastrado é exibido na lista de usuários cadastrados.

**Configuração de Informações Para SMA:** A funcionalidade de configurações para SMA envolve cadastro de verbos de interesses para monitoramento das características e cadastro das máscaras de características para todas as seções do perfil do usuário.

O cadastro dos verbos e máscaras é feito para o coletivo, ou seja, o perfil de todos os usuários é montado sobre as mesmas configurações de verbos e máscaras. Veja na Figura 6.15 a tela de cadastro de verbos para a seção de características de habilidade.



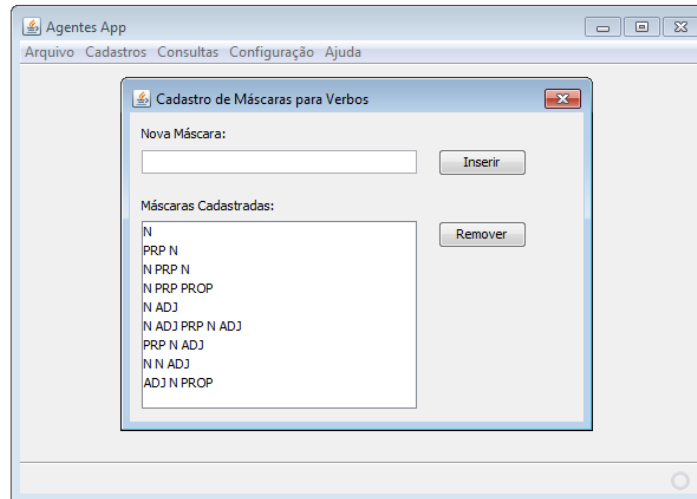


**Figura 6.15 - Imagem Tela de Cadastro de Verbo. 25**

Imagina-se a ferramenta, trabalhando com um número grande de verbos, possibilitando assim a captura de diversas características.

Para cadastrar um verbo, digita-se o verbo no campo e clica-se no botão inserir verbo. Como pode ser visto na Figura 6.15, o verbo recém cadastrado é exibido na lista de verbos.

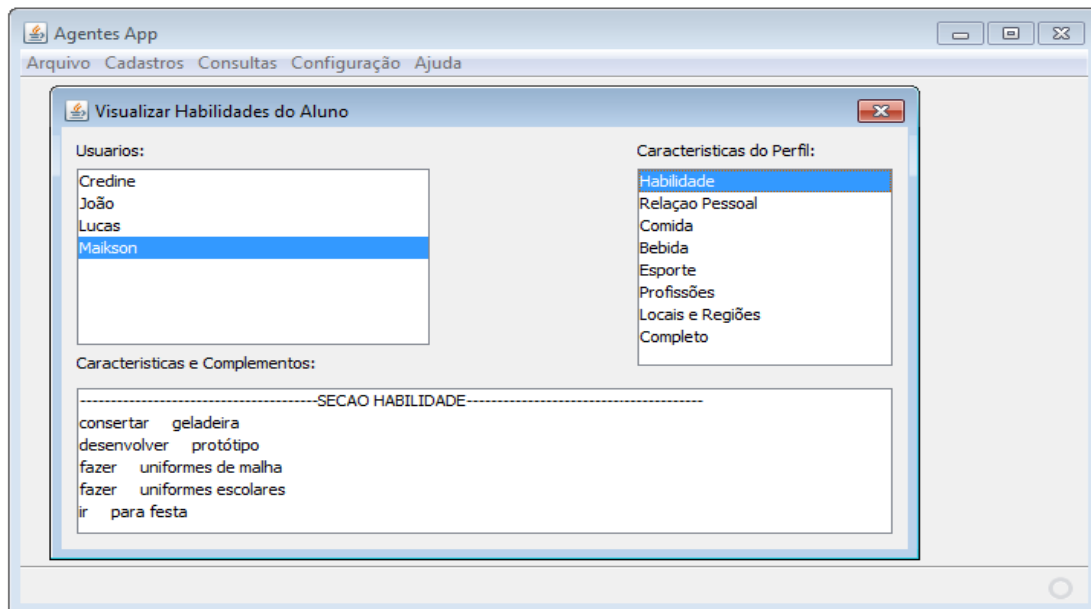
Vejamos a tela de cadastro de máscaras de complemento de características na Figura 6.16. As máscaras de complementos são usadas pelo agente ARCP para filtrar os complementos.



**Figura 6.16 - Imagem Tela de Cadastro de Máscaras. 26**

Para cadastrar uma máscara, digite-a no campo e clique no botão inserir verbo. Como pode ser observado na Figura 6.16, a máscara recém-cadastrada é exibida na lista de máscaras.

**Consultar/Visualizar Perfil Individual:** A funcionalidade de visualizar perfil permite que o administrador selecione na lista de usuários de qual usuário deseja visualizar as características do perfil. Após selecionar o usuário, é necessário que ele selecione a seção de característica que deseja visualizar, tendo na lista como última opção a visualização de todas as características. Ao selecionar a opção de característica, são listados as características e complementos conforme a opção selecionada, como pode ser visto na Figura 6.17.



**Figura 6.17 - Imagem Tela De Consultar/Visualizar Perfil Individual. 27**

A Figura 6.17 demonstra a seleção do usuário Maikson e a seleção da seção de habilidades.

**Consultar/Realizar Agrupamentos:**

A funcionalidade desenvolvida na interface apresentada a seguir é uma das possibilidades de utilização do perfil em ambiente escolar. Imaginamos uma necessidade de trabalhar com um conjunto de alunos denominado de turma. Deseja-se formar agrupamentos das pessoas conforme as suas características.

1 - Sendo assim, foram estudadas as seguintes possibilidades de agrupamentos de pessoas considerando as informações da seção de habilidade de cada uma das pessoas:

1.1 – Para que o administrador possa montar grupos de pessoas, tendo cada grupo um mesmo subconjunto de habilidades em comum. Serão especificadas as habilidades de cujos se deseja selecionar. Estas pessoas serão distribuídas uma em cada grupo, completos com as demais pessoas que constituem a turma. Esta funcionalidade será útil para o caso em que o Interessado desejar determinar certo nível de conhecimento para cada grupo. Assim, ele garantirá as especificações mínimas para o desenvolvimento de uma determinada atividade, mantendo um nível de similaridade entre os grupos, pois o usuário com habilidades específicas estará em grupos distintos.

1.2 – Para que o Interessado possa dividir a turma em dois grupos, tendo em um grupo todas as pessoas que tenham algumas determinadas habilidades e, em outro grupo, todas as pessoas que não possuem estas determinadas habilidades. Esta funcionalidade será útil para o caso em que se deseja separar as pessoas que possuem determinadas habilidades das que não possuem.

1.3 – Para que o Interessado possa visualizar o perfil de uma pessoa com todas as suas habilidades e complementos com as relações pessoais e locais de que conhece, possibilitando gerenciar as melhores formas de aproveitar-se destas informações para compor os grupos.

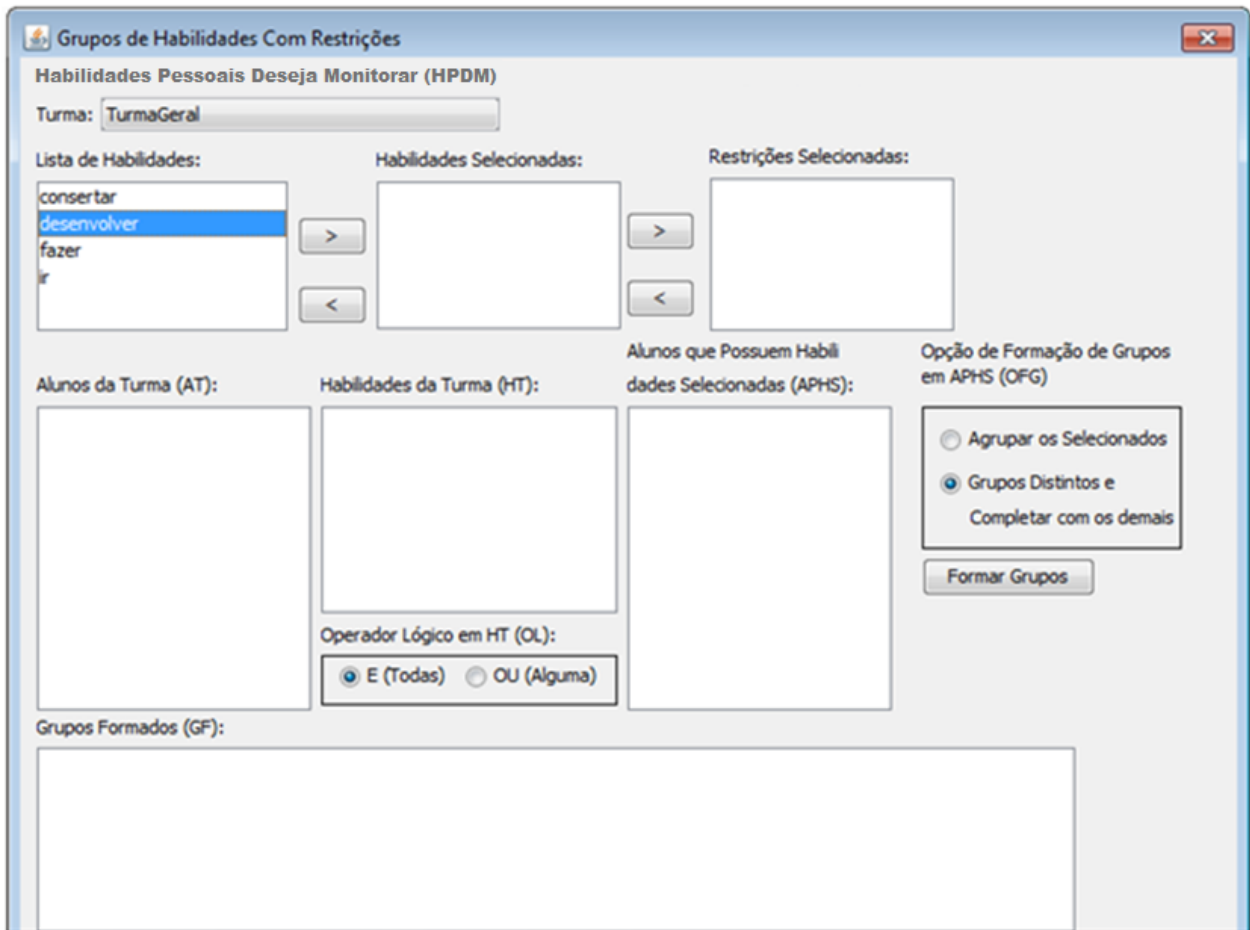
2 – Considerando outra sessão de característica (seção de relações pessoais), podemos imaginar que o Interessado tenha a opção de, ao compor os grupos, considerar as relações afetivas, tendo as seguintes opções:

2.1 – Privilegiar relações positivas: dentre várias opções de pessoas que atendam às possibilidades de agrupamentos, a que tiver relações positivas com os membros do grupo terá a preferência ao se agrupar.

2.2 – Privilegiar relações negativas: dentre várias opções de pessoas que atendam às possibilidades de agrupamentos, a que tiver relações negativas com os membros do grupo terá a preferência ao se agrupar.

2.3 – Impedir relações negativas: Será impedido de se agrupar se existirem conflitos pessoais com membros do grupo.

2.3 – Ignorar relações afetivas: Não serão consideradas as relações afetivas para a criação dos grupos.



**Figura 6.18 - Imagem Tela de Consultar/Realizar Agrupamento. 28**

A seguir, descreve-se como funciona a utilização da interface para formar os grupos, ilustrado na Figura 6.18 .

- i. Na seção HPDM, são mostradas as habilidades das pessoas que se deseja monitorar para a turma. As habilidades seleccionadas irão gerar uma lista de pessoas da turma que tenham tais habilidades e esta lista de pessoas será mostrada na seção AT.
- ii. Na seção AT é mostrada uma lista de pessoas e, por meio desta, seleccionam-se as pessoas que desejam estar visualizando as habilidades, sendo estas mostradas na seção HT.
- iii. A seção HT apresenta uma lista de habilidades. Nesta seção só são mostradas as habilidades dos grupos de pessoas seleccionados e que pertencem às

habilidades selecionadas na seção HPDM. As habilidades que forem selecionadas na seção HT indicam o interesse em saber as pessoas que (quais pessoas) possuem tais habilidades. Esta informação é mostrada na seção APHS. Para localizar as pessoas com as habilidades selecionadas, existe a opção de operação lógica de E/OU, que é selecionada na seção OL.

- iv. A seção OL permite optar por duas formas de montar os grupos E/OU. Respectivamente, na opção E, aparecem as pessoas que possuem todas as habilidades selecionadas simultaneamente na seção HT e estas pessoas serão listadas na seção APHS. A segunda opção OU irá mostrar as pessoas que possuem pelo menos uma das habilidades selecionadas na seção HT e estas pessoas serão listadas na seção APHS.
- v. Na seção APHS apresenta-se uma lista de pessoas. As pessoas selecionadas nesta seção formarão grupos que serão colocados na seção GF. As opções de formação de grupos estão definidas na seção OFG, sendo duas: “agrupar os selecionados” ou “grupos distintos e completar com os demais”.
- vi. A seção OFG permite optar por duas formas de montar os grupos: “agrupar os selecionados” ou “grupos distintos e completar com os demais”. Respectivamente, a primeira formará um grupo com as pessoas selecionadas na seção APHS e criará outros grupos com os demais membros da turma que não foram selecionados, caso existam. Já na segunda opção, as pessoas selecionadas na seção APHS formarão grupos distintos e serão completos os grupos com as demais pessoas da turma. Estes grupos serão mostrados na seção GF.
- vii. Na seção GF é mostrada uma lista de grupos formados conforme as especificações nas demais seções (AT, HT, APHS, OL e OFG).
- viii. A seção CCAC permite que seja montada uma String para Formação de Grupo (SFG), que permite descrever as opções de agrupamentos. Os grupos dos alunos serão mostrados na seção GFC. Veja na seção CCAC um exemplo de uma SFG.

ix. Na seção GFC são mostrados os grupos formados conforme as especificações na seção CCAC descritas pela SFG.

Nas seções AT, HT e APHS, os trechos em azul exemplificam informações selecionadas.

### 6.5. Base de Dados do SysCAP

A visão da base de dados do protótipo SysCAP pode ser visualizada na Figura 6.19, para um entendimento da relação de postagem pertencente ao autor no módulo de interface do MOrFEu e os demais conceitos de: postagem capturada, texto etiquetado, característica/habilidade e hipótese.

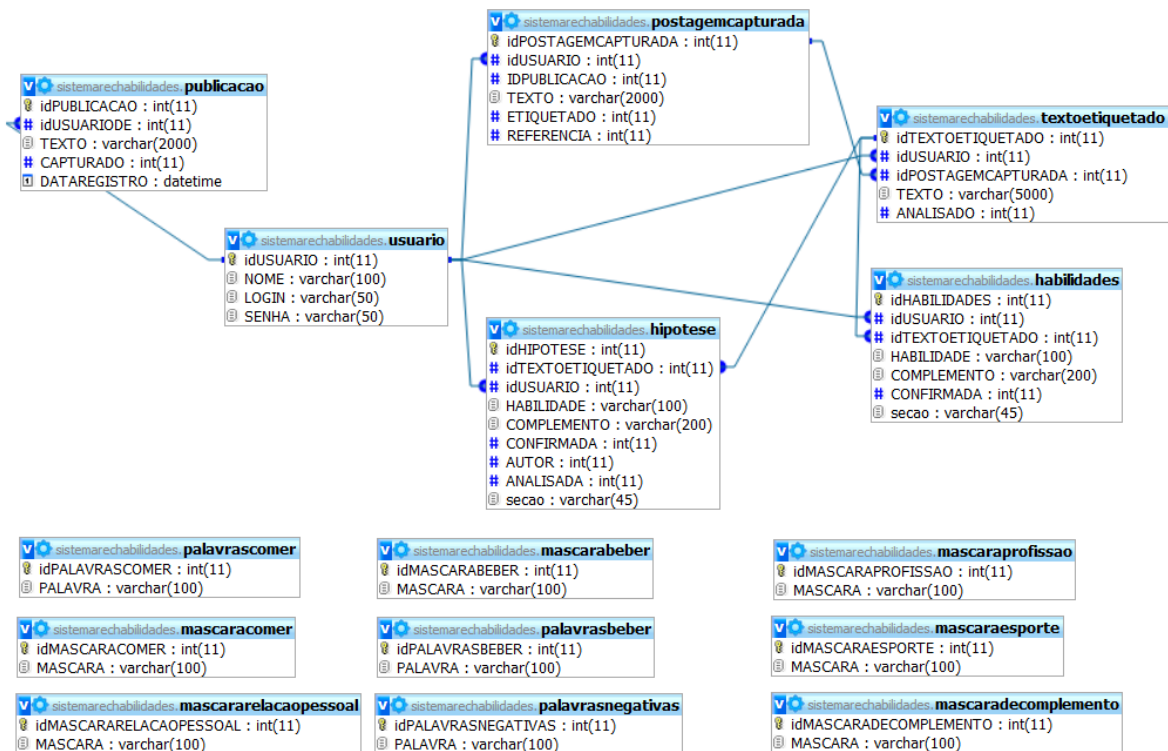


Figura 6.19 - Base de Dados do SysCAP29

### 6.6. Experimento do Protótipo

Nesta Seção, apresentamos a experimentação realizada com o protótipo SysCAP. Os experimentos foram conduzidos por meio do protótipo computacional, baseado na UPIs



produzidas no ambiente MORFEu. No MORFEu, todas as produções dos usuários são nomeadas de Unidade de Produção Intelectual (UPI) e é mantida na biblioteca de UPI que contém todas as produções do usuário. Esta biblioteca contém todas as produções do autor, indiferente do veículo de comunicação ao qual a UPI possa estar vinculada.

Apresentamos a biblioteca de UPIs de 3 usuários. Foram selecionados trechos de UPIs dos usuários que demonstram a boa utilização da proposta e as dificuldades encontradas, um fator definitivo para obter o perfil dos usuários, que está diretamente ligado à clareza e qualidade dos textos. Qualidade do texto para a aplicação dos testes do protótipo envolve algumas características textuais para se obter a coleta adequada das características pessoais para o perfil do usuário, tais como: ausência de abreviação, ausência de erros de grafia, ausência de erros de pontuação, textos escritos com a concreta especificação de estar na primeira pessoa, textos que não revelam informações pessoais, etc.

### **Análise do perfil do Arnaldo com base em suas UPI**

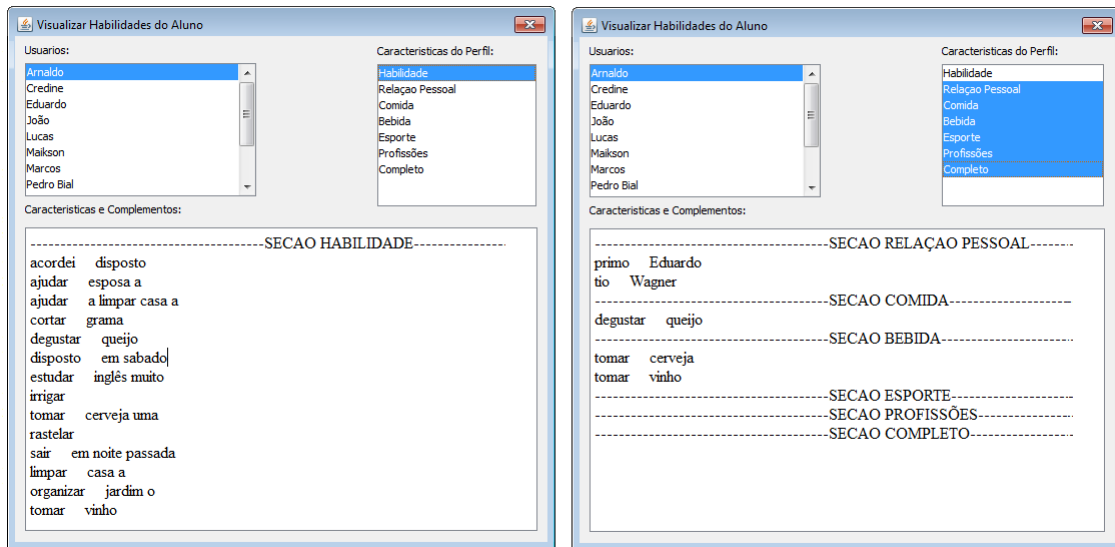
#### *Arnaldo:*

Hoje estou estudando inglês muito. Porém ontem sai com meu tio Wagner e meu primo Eduardo, tomamos umas cervejas e por fim, Eduardo comeu pizza, eu e Wagner degustamos queijo e tomamos vinho.

Saímos na noite passada e ainda assim acordei bem disposto no sábado, ajudei a minha esposa a limpar a casa, depois a esposa foi fazer o almoço e eu fiquei organizando o jardim, cortei a grama, rastelei e irriguei.

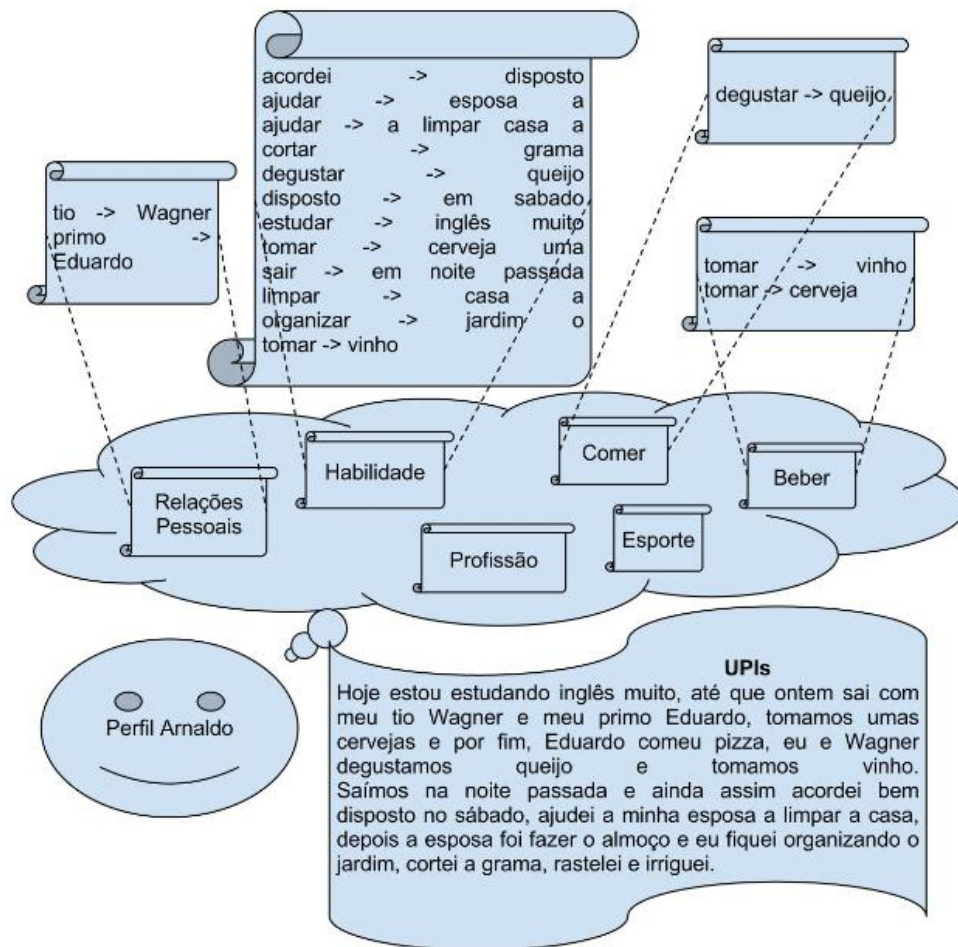
**Figura 6.20 - UPIs de Arnaldo30**

Os textos da coleção de UPIs de Arnaldo, Rômulo e Juliana são textos fictícios montados propriamente para revelação dos objetivos e funcionalidades do protótipo. Veja na Figura 6.21 o perfil do usuário Arnaldo obtido pelo protótipo SysCAP com base das UPIs que pode ser observado na Figura 6.20.



**Figura 6.21 - Representação Do Perfil De Arnaldo Pelo Protótipo31**

Para ilustrar o perfil do Arnaldo obtido pelo protótipo, foi feita uma imagem que o representa graficamente. Na Figura 6.222 é representado o perfil do usuário Arnaldo.



**Figura 6.22 - Representação Gráfica do Perfil do Usuário Arnaldo32**

No decorrer do processamento do texto produzido por Arnaldo e a obtenção de suas características para seu perfil, levantamos a hipótese de que Eduardo tem a característica de “comer” e o complemento “pizza” para a seção de habilidade. Então, considerando que Eduardo é um usuário do sistema, quando ele for entrar na Interface do MOrFEu, será solicitado que confirme ou negue a hipótese de que ele tem a característica de “comer” e o complemento de “pizza”.

### **Análise do perfil de Rômulo com base em suas UPIs**

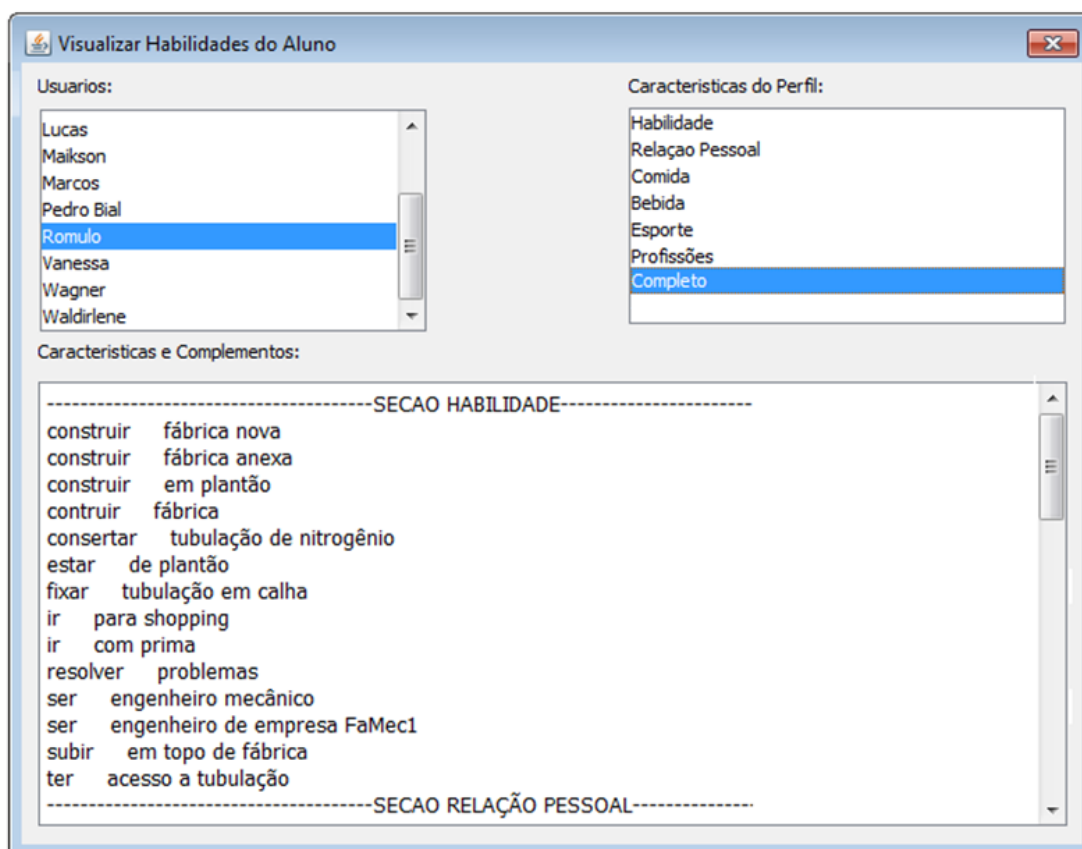
Realizando o processamento da UPI do usuário Rômulo ilustrado na Figura 6.2, foi possível obter o perfil mostrado na Figura 6.23.

**Romulo:**

Meu fim de semana foi só correria. Eu sou engenheiro mecânico da empresa FabricaMec e estava de plantão, tive que resolver vários problemas. Estamos construindo uma nova fábrica, anexa à FabricaMec1, mas neste plantão tive que consertar a tubulação de nitrogênio e subir no topo da fábrica para ter acesso a tubulação. Marcos estava soldando as conexões e eu estava fixando a tubulação na calha.

Amanhã vou jogar vôlei de areia na praia, eu gosto de praia.

**Figura 6.23 - UPIs de Rômulo33**



**Figura 6.24 - Representação do Perfil de Rômulo Pelo Protótipo34**

Para ilustrar o perfil do Rômulo obtido pelo protótipo, foi feita uma imagem que o representa graficamente. Na Figura 6.25 é demonstrado o perfil do usuário Rômulo.

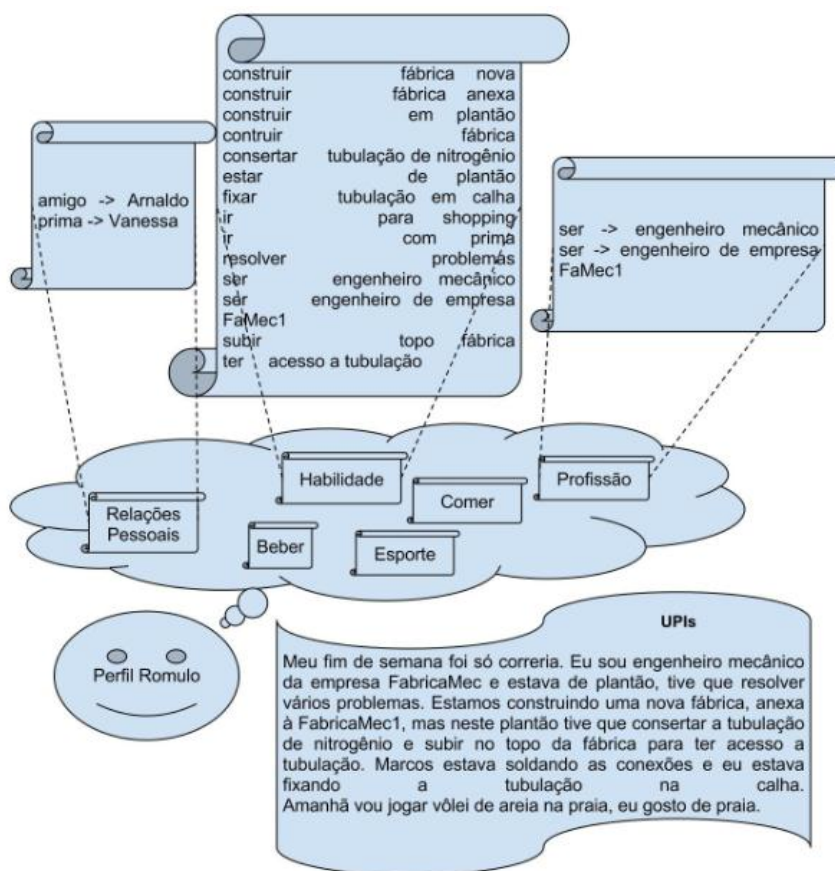


Figura 6.25 - Representação Gráfica do Perfil do Usuário Rômulo. 35

Os textos da coleção de UPIs de Rômulo com seus respectivos textos etiquetados estão disponíveis no Apêndice B.

### Análise do perfil da Juliana com base em suas UPIs

Na Figura 6.26 podem ser observadas as UPIs textuais da Juliana processadas pelo protótipo.

**Juliana:**

Meu nome é Juliana, não gosto de ficar parada. Estou sempre querendo o melhor e aprender mais, tenho paixão pelo trabalho. Entrei na MRTextil como Assistente Técnico.

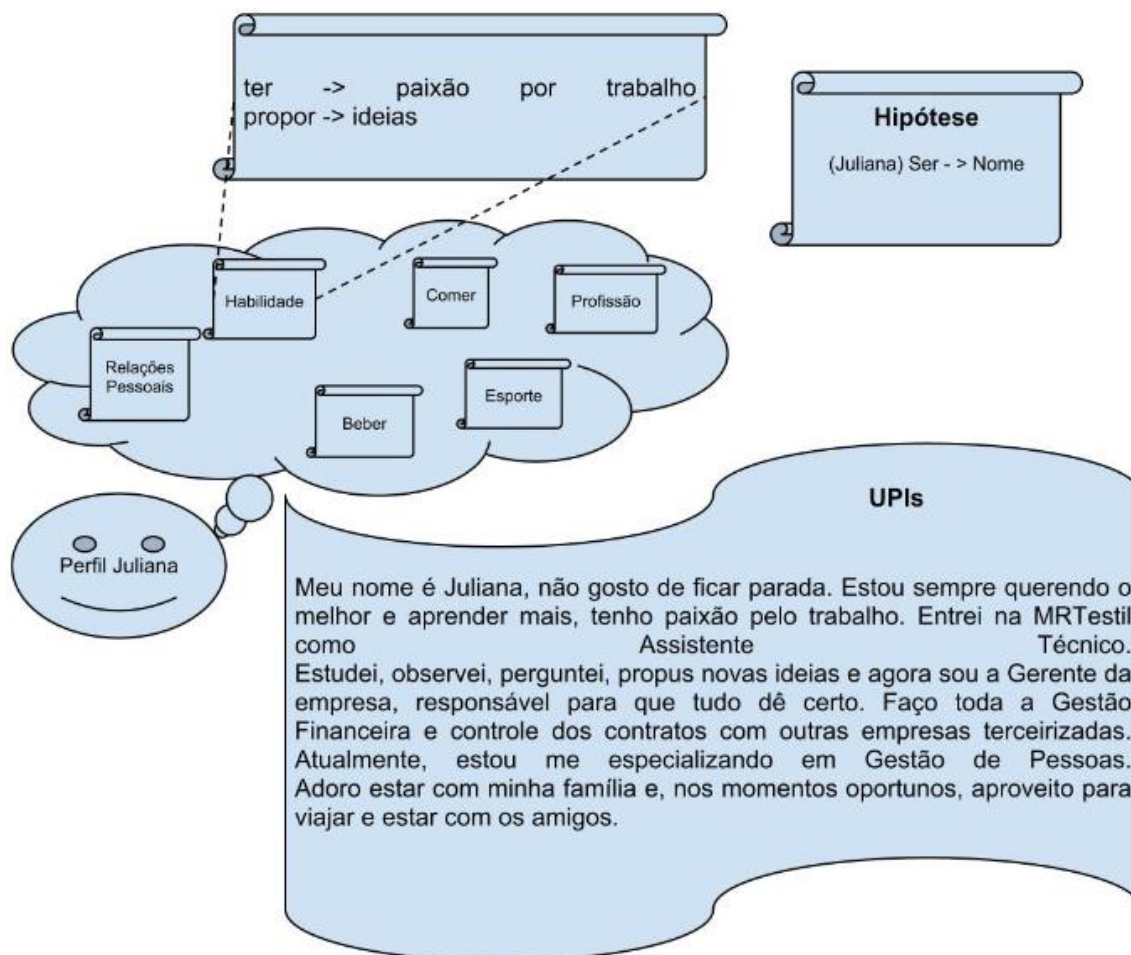
Estudei, observei, perguntei, propus novas ideias e agora sou a Gerente da empresa, responsável para que tudo dê certo. Faço toda a Gestão Financeira e controle dos contratos com outras empresas terceirizadas.

Atualmente, estou me especializando em Gestão de Pessoas.

Adoro estar com minha família e, nos momentos oportunos, aproveito para viajar e estar com os amigos.

**Figura 6.26 - UPIs de Juliana36**

Na Figura 6.27 é demonstrada a representação gráfica do perfil da usuária Juliana obtido pelo protótipo.



**Figura 6.27 - Representação Gráfica do Perfil do Usuário Juliana. 37**

Ao analisarmos a Figura 6.27, percebemos que foi obtido abaixo do esperado o resultado do processamento das UPIs de Juliana. Ao comparar com as características identificadas facilmente por humanos, percebeu-se que o algoritmo implementado no protótipo obteve poucas características pessoais.

Vamos analisar com mais detalhe os fatores que impediram a identificação das demais características.

Durante o processamento dos textos de Juliana, pode ser percebido a ocorrência de equívocos. Veja o processo de etiquetação da frase “*Meu nome é Juliana, não gosto de ficar parada.*” na Figura 6.28.

Meu [meu] <poss 1S> **DET M S** #1->2  
 nome [nome] <ac-cat> <percep-f> **N M S** #2->3  
 é [ser] <vK> <fmc> <mv> **V PR 3S IND VFIN** #3->0  
 Juliana [Juliana] <hum> **PROP M S** #4->3  
 , #5->0  
 não [não] **ADV** #6->7  
 gosto [gostar] <vH> <fmc> <mv> **V PR 1S IND VFIN** #7->3  
 de [de] **PRP** #8->7  
 ficar [ficar] <vK> <mv> **V INF** #9->8  
 parada [parar] <jh> **V PCP F S** #10->9

**Figura 6.28 - Processamento Equivocado da Fala de Juliana38**

Ao analisarmos esta frase da Figura 6.28, podemos perceber que ela revela algumas características, revelando o nome do autor e que ela não gosta de ficar parada. Ao processar esta frase pelo parser PALAVRAS obtivemos as etiquetas com os vínculos semânticos como pode ser visto na Figura 6.28. Estas etiquetas revelam existir dois vínculos de dependência gramatical relacionados ao verbo principal “ser”, a token de número 2 “nome” (sendo ela um substantivo - N) tem o vínculo de dependência gramatical com o verbo assim como a token de número 4 “Juliana” (sendo ela um nome próprio - PROP) tem o vínculo de dependência gramatical com o verbo principal “ser”, assim o verbo possui a possibilidade de dois complementos N(nome) e o PROP(Juliana).

Na Figura 6.28, ao processar a terceira token “é”, trata-se do verbo ser, sendo ele o verbo principal da oração na terceira pessoa. Ao se aplicar o algoritmo para identificar as características pessoais, o verbo “ser” encontra-se na terceira pessoa, não sendo possível identificar como revelação de característica do próprio autor. Quando foi aplicado o algoritmo de revelação de característica de terceiro, atendeu todas as condições para o levantamento de hipótese, para a validação do verbo, sendo:

- 1- Ser um verbo principal com a tag <mv>.
- 2- Não ter no vínculo de dependência gramatical a ocorrência de uma palavra de negação.
- 3- Ser verbo na terceira pessoa.
- 4- Haver um nome próprio no vínculo, identificado pela tag PROP.



- 5- Haver a constatação de que o nome próprio é de pessoa, identificado pela tag <hum>.
- 6- Existir o nome próprio cadastrado na base de usuários do sistema.
- 7- Obter complemento adequado conforme configurado.

Verifica-se que a frase oferece dois vínculos de dependência gramatical para complemento, sendo o complemento PROP ou N. Considerando a configuração atual de N como complemento, obteve-se a hipótese de que Juliana tem a característica “ser” com complemento “nome”, sendo esta uma hipótese equivocada, pois a frase revela informação sobre o próprio autor, fato que precisa ser estudado e identificada a possível solução.

Outra situação que merece um tratamento mais adequado ocorre quando é identificada uma palavra de negação no vínculo de dependência gramatical. A ocorrência de uma palavra de negação no vínculo de dependência gramatical inviabiliza a obtenção de qualquer característica do processo, fazendo com que qualquer característica revelada seja descartada. Um possível tratamento mais adequado seria a possibilidade de identificar o significado inverso.

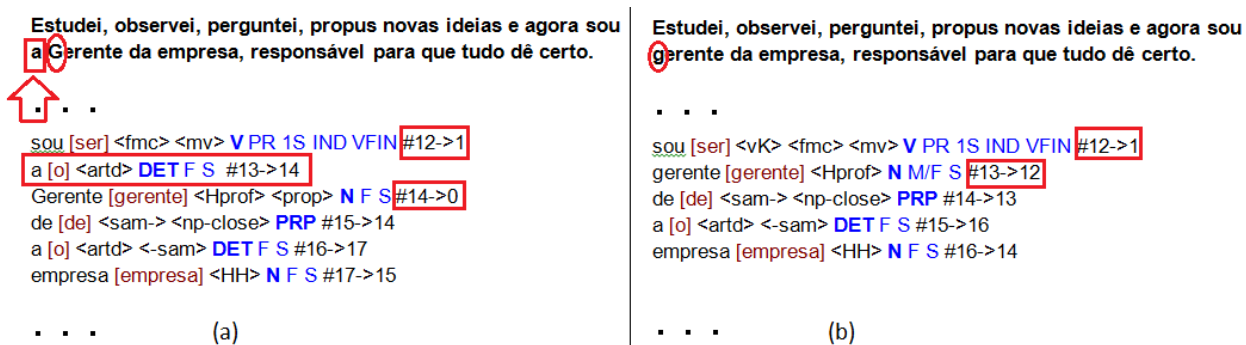


Figura 6.29 - Análise Comparativa das Etiquetas<sup>39</sup>

Analisando a frase processada na Figura 6.29 (a), percebe-se que o vínculo de dependência gramatical da token “gerente” não está ligado ao verbo “ser” e a falta deste vínculo impede que a característica profissional seja identificada. O fato de o token “a” estar entre o verbo e o sujeito quebrou o vínculo de dependência gramatical entre o substantivo “Gerente” e o verbo “Sou”.

Na Figura 6.29 (b), foi retirado o token “a” e trocado o “G” por “g”, permitindo a correta classificação semântica da palavra “gerente” como sendo uma profissão observada pela tag <Hprof> e o correto vínculo de dependência gramatical entre o verbo e o substantivo, fato que permitiria a identificação de característica para a seção de habilidade assim como para a seção de profissão.

### 6.7. Configuração do Protótipo

A seguir é apresentada a configuração das máscaras de complementos e palavras/verbos que permitiram a obtenção das características e complementos de cada uma das seções do perfil dos usuários Arnaldo, Rômulo e Juliana conforme pode ser visto respectivamente nas Figura 6.222, Figura 6.25 e Figura 6.27. Inicialmente vamos apresentar as máscaras de complementos da seção de habilidade. Observe a Figura 6..

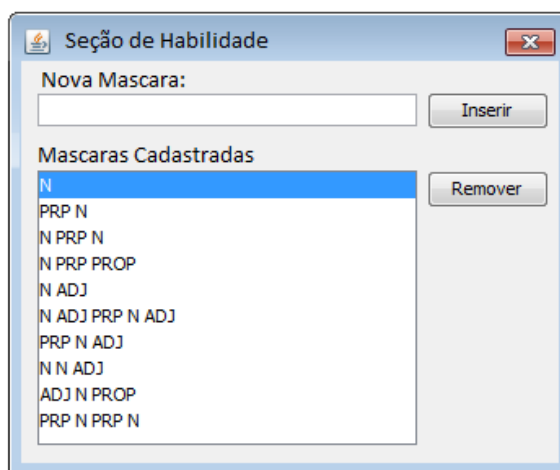


Figura 6.30 - Máscaras da Seção de Habilidades40

Observe a configuração das regras de ocorrência durante o processamento dos textos para obtenção de características e composição do perfil na Tabela 6.5.

Tabela 6.5 - Configurações do Protótipo910

	Mascara	Lista de Verbos	Tag Semântica
<b>Habilidade</b>	N; PRP N; N PRP N; N PRP PROP;	Verbo identificado com a tag semântica <mv>	

	N ADJ; N ADJ PRP N ADJ; N N ADJ; ADJ N PROP; PRP N PRP N		
<b>Comida</b>	N; N ADJ; ADJ; PRP N; N PRP N ADJ	“comer”, “alimentar”, “gostar”, “degustar”, “ingerir”, “deliciar”, “adorar”	Ocorrência da tag semântica <food>
<b>Bebida</b>	N; N ADJ; ADJ; PRP N; N PRP N ADJ	“beber”, “tomar”, “gostar”, “degustar”, “ingerir”	Ocorrência da tag semântica <drink>
<b>Profissão</b>	N; N ADJ; N PRP N PROP	Verbo identificado com a tag semântica <mv>	Ocorrência da tag semântica <Hprof>
<b>Relação Pessoal</b>	PROP	Palavras identificadas com a tag semântica <Hfam>	Ocorrência da tag semântica <hum>
<b>Esporte</b>	N; N ADJ	Verbo identificado com a tag semântica <mv>	Ocorrência da tag semântica <sport>
<b>Palavras Negativas</b>		“não”, “exceto”, “nunca”, “jamais”.	

## 6.8. Conclusão

Os testes realizados possibilitaram a conclusão de que a ideia proposta apresenta bons resultados quando submetida a textos escritos gramaticalmente corretos e que contenham apenas uma proposição. Possibilitou a identificação de alguns pontos de melhorias e a necessidade de poder especificar com mais detalhes as regras de complementos.

Revelaram a necessidade de se processar os resultados obtidos inicialmente para que os complementos fiquem de forma mais adequada. Também foi identificado que alguns complementos ficaram meio fora de contexto. Como exemplo, podemos observar a Figura 6.21 sobre o perfil de Arnaldo: na seção de habilidade obtivemos a característica: “ajudar” e o complemento: “esposa a”. O artigo “a” tem ficado fora da correta ordem, fazendo-se necessário um tratamento mais adequado para estas situações.

## 7. CONSIDERAÇÕES FINAIS

Percebemos, no decorrer do trabalho, a dificuldade de identificar conjuntos de informações que representem as características das pessoas. A dificuldade de identificar um padrão de perfil de usuários que atenda às diversas necessidades e interesses em obter determinados tipos de informações para compor o perfil das pessoas.

Foi percebido pela análise dos trabalhos correlatos que os tipos de informações são sempre diferentes para cada implementação. O perfil de usuário para cada implementação contém necessidades e aplicações diferentes, nos direcionando a refletir em como seria correto representar o perfil do usuário, considerando a necessidade de aplicação proposta por este trabalho.

O trabalho implementou um modelo de perfil e métodos para coletar dados para a composição automática do perfil dos usuários de um ambiente virtual de aprendizagem. Os dados coletados para a composição do perfil são extraídos de textos escritos pelo próprio usuário, utilizando o processamento de linguagem natural e a análise e implementação de regras e padrões que permitem identificar características do próprio autor.

A implementação do protótipo validou as expectativas, pois foi possível ter uma boa representação de perfil dos usuários com os dados que foram monitorados para a formação do perfil pessoal dos usuários.

A extração automática dos dados para o perfil dos usuários por meio de textos e técnicas de PLN foi validada pela implementação do protótipo e apresentou ser eficiente para os textos apresentados. Os resultados obtidos atenderam com êxito às expectativas para a composição de perfil do usuário nos casos em que o autor tem o hábito de escrever corretamente sobre o conceito de gramática, ortografia, concordância.

As análises dos resultados em que se obteve menor eficiência do protótipo SysCAP foram os resultados nos quais se processaram textos escritos com erros de pontuação, acentuação, abreviações ou vícios da linguagem.

A utilização da ferramenta de processamento linguístico escolhida facilitou o desenvolvimento do protótipo. A identificação dos grupos semânticos e os vínculos de dependência gramatical, além das classificações morfológicas e sintáticas disponibilizadas pela ferramenta VISL, nos possibilitaram descobrir de que forma obter as informações que desejamos extrair dos usuários.

A utilização das técnicas de SMA deu uma flexibilidade e a disponibilidade de organização e modularização do problema, possibilitando a implantação de inteligência no processo.

Foi identificada a possibilidade de maiores especificidades na representação das seções do protótipo, por meio de técnicas que complementem a caracterização semântica oferecida pelo etiquetador utilizando lógica descritiva.

Foi identificada a possível necessidade de especificações maiores para os complementos das características, como por exemplo, não identificar que o complemento seja somente uma classificação morfológica, mas que possa ter especificação sintática como número, gênero, tempo, etc.

## 7.1. Trabalhos Futuros

Lista de trabalhos futuros que complementam o objetivo da dissertação.

- 1 – O experimento foi realizado sobre textos escritos na primeira pessoa. Pensamos em ampliar a ideia sobre um domínio maior de texto e tratar situações em que se deseja obter características sobre algum objeto, pessoas, empresas, lugares, conceitos, etc. Realizando o processamento de textos escritos sem restrição de serem textos de autor conhecido, isso permitiria que várias pessoas ao se referenciar a um conceito como “política”, seria criada a característica sobre um conceito geral de “política” de forma que a atribuição de adjetivos ao conceito “política” fosse criando sua definição de forma pública (glossário de conceitos), quando possível identificar o autor do texto as características sobre o conceito seria também identificado como características pessoais do autor sobre o conceito “política”.
- 1.1 – Evoluir as regras e padrões para identificar características de coisas diversas.
- 1.2 – Implementar recursos no SMA que possibilite a captura de todos os textos dos usuários e identificassem as características ditas pelas pessoas sobre as diversas coisas.
- 1.3 – Implementar agentes que colete textos fora do ambiente MOrFEu permitindo que outros ambientes sejam analisados.
- 1.4 – Possíveis aplicações seria útil para pesquisa de satisfação outra seria a criação de um glossário regional, identificação de ontologia local.
- 2 – Implementação de uma ontologia de domínio sobre os vários domínios, que permita identificar as características dos usuários que se relaciona com cada domínio, essa ontologia de domínio permitiria a composição de uma base de conhecimento específica para cada domínio, tornando o sistema mais “inteligente”.
- 3 – Realizar teste da proposta em ambientes educacionais e profissionais, realizando a integração do protótipo com o ambiente MOrFEu.
- 4 – Realizar uma pós etiquetagem dos textos retornado pelo parser PALAVRAS VISL para adição de outras classes e talvez a identificação de subclasses semânticas.

- 5 – Implementar o agente construtor de conhecimento ACC com base de conhecimentos em PROLOG.
- 6 – Verificar necessidade de aperfeiçoamento no agente de hipótese considerando a base de conhecimento.
- 7 – Implementar análise de texto em nível de análise de discurso e análise pragmática.
- 8 – Identificar uma forma de criar marcadores das características com o tempo em que foram obtidas, possibilitando comparações de estados do perfil em períodos diferentes.
- 9 – Verificar formas para a especificação e representação flexível do perfil do usuário.
- 10 – Implementar extração de hipóteses de citação de terceiros para todas as seções de características do perfil.



## REFERÊNCIAS

- ARARIBÓIA, G. **Inteligência Artificial - um curso prático**. Rio de Janeiro: LTC, 1988. p. 149-237.
- ARELLANO, M. A. M. **Perfil de usuários de lista de discussão da área de ciência da informação**, 2005.
- AROYO, L. et al. (2006) Ontology-based personalization in user-adaptive systems. In: 2ND INTERNATIONAL WORKSHOP ON WEB PERSONALIZATION, RECOMMENDER SYSTEMS AND INTELLIGENT USER INTERFACES (WPRSIUI'06), Dublin, Ireland.
- AUGUSTIN I. et al. **ISAM: a Software Architecture for Adaptive and Distributed Mobile Applications**. In: VII IEEE Symposium on Computers and Communications, Messina, 2002, pp. 333-339.
- AUSTIN, John L. **How to do Things with Words**. Clarendon, Oxford, 1962.
- BELTRAME, W. A. R. Um Sistema de Disseminação Seletiva da Informação Baseado em Cross-Document Structure Theory. Dissertação (Mestrado em Informática). Disponível em: < [http://www.bdt.d.ufes.br/tesesimplificado/tde\\_arquivos/13/TDE-2012-02-09T120128Z-653/Publico/Dissertacao%20Walber.pdf](http://www.bdt.d.ufes.br/tesesimplificado/tde_arquivos/13/TDE-2012-02-09T120128Z-653/Publico/Dissertacao%20Walber.pdf) > Acesso em: 01 ago. 2012.
- BICK, E. (2010). A Dependency-based Approach to Anaphora Annotation. In: (eds.) Extended Activities Proceedings. **9th International Conference on Computational Processing of the Portuguese Language** Apr. 27-30. Porto Alegre, Brazil. pp. xxx. ISSN 2177-3580.
- BICK, E. (2007-9). Automatic Semantic Role Annotation for Portuguese. In: Proceedings of TIL 2007 - 5th Workshop on Information and Human Language Technology / **Anais do XXVII Congresso da SBC** (Rio de Janeiro, July 5-6, 2007). ISBN 85-7669-116-7, pp. 1713-1716.
- BICK, E. (1998). Structural Lexical Heuristics in the Automatic Analysis of Portuguese. In: **The 11th Nordic Conference on Computational Linguistics (Nodalida '98)**, Proceedings. Copenhagen: Center for Sprogteknologi (CST) and Department of General and Applied Linguistics (IAAS), University of Copenhagen.
- BICK, E. (2000-2). **The Parsing System Palavras - Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework**. Aarhus: Aarhus University Press (preprint version) -- dr.phil. thesis (cf. the Disputatio for an introduction).

BICK, Eckhard. **Gramática Constritiva na Análise Automática de Sintaxe Portuguesa**. In: Berber Sardinha, Tony (ed.), *A Língua Portuguesa no Computador [The Portuguese Language on the Computer]*. FAPESP. ISBN: 85-7591-044-2, Campinas: Mercado de Letras, São Paulo 2005.

BIRD, S; KLEIN, E.; LOPER, E. **Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit**. CA: O’Reilly Media, 2009.

BRUSILOVSKY, P. Methods and Techniques of Adaptive Hypermedia. In: ADAPTATIVE HYPERTEXT AND HYPERMEDIA. 1998, Hardcover. Proceedings... [S.I.]: Springer, 1998.

CIMIANO, P. **Ontology learning and population from text: algorithms, evaluation and applications**. New York: Springer-Verlag, 2006, 171p.

DEY A. K.; ABOWD G. D. Towards a better understanding of context and context-awareness. In: HUC '99: PROCEEDINGS OF THE 1ST INTERNATIONAL SYMPOSIUM ON HANDHELD AND UBIQUITOUS COMPUTING, 1999, pp. 304-307.

EXCELENTE-TOLEDO, C. B.; JENNINGS, N. R. The Dynamic Selection of Coordination Mechanisms. **Journal of Autonomous Agents and Multi-Agent Systems**, Amsterdam, v.9, p.55–85, 2004.

FREITAS, S. A. A. D. *Interpretação Automatizada de Textos: Processamento de Anafóras*. Vitória: Universidade Federal do Espírito Santo, 2005. Tese de Doutorado.

FININ, T. et al. Specification of the KQML agent communication language. DARPA knowledge sharing initiative external interfaces working group. 1993.

GENESERETH, M. R.; KETCHPEL, S. P. Software agents. **Communications of the ACM**, New York, v.37, n.7, p.48–53, July 1994.

GONZALEZ, M.; LIMA, V. L. S. “Recuperação de Informação e Processamento da Linguagem Natural”. In: XXIII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2003, Campinas, 2003. **Anais do III Jornada de Mini-Cursos de Inteligência Artificial**, v. 3, p.347-395.

HUHNS, M. N.; STEPHENS, L. M. Multiagent Systems and Societies of Agents, em *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, editado por Gerhard Weiss, The MIT Press, cap 2, p. 121-164, 1999.

IANNINI, O. **Pesquisa do Perfil dos Profissionais de Tecnologia da Informação**. Março de 2010. Disponível em: <[http://www.seprorj.org.br/media/pesquisa\\_perfil\\_sintese.pdf](http://www.seprorj.org.br/media/pesquisa_perfil_sintese.pdf)>. Acesso em: 01 ago. 2012.

IEEE LTSC (2002). PAPI Learner, Draft 8 Specification. Disponível em: <<http://edutool.com/papi/>>. Acesso em: 10 jun. 2012.

IMS-LIP (2001). **IMS Learner Information Packaging Information Model Specification**, Final Specification Version 1.0. Disponível em: [<http://www.msglobal.org/profiles/lipinfo01.html>]. Acesso em: 02 jun. de 2012.

JENNINGS, N. R. On agent-based software engineering. **Artificial Intelligence**, v. 117, p. 277-296, March 2000.

JENNINGS, N.; SYCARA, K.; WOOLDRIDGE, M. A Roadmap of Agent Research and Development. *Autonomous Agents and Multi-Agent Systems*. **Kluwer Academic Publishers**, Boston, n. 1, p. 7- 38, 1998.

JENNINGS, N. et al. Agent-based Business Process Management. **International Journal of Cooperative Information Systems** 5(2-3): pp. 105–130, 1996.

JURAFSKY, D., MARTIN, J. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Prentice-Hall.

KOWATA, J. H. **Uma Abordagem Computacional para Construção de Mapas Conceituais a partir de Textos em Língua Portuguesa do Brasil**. Dissertação (Mestrado em Informática) – Programa de Pós-Graduação em Informática, Universidade Federal do Espírito Santo, Vitória, Agosto de 2010.

LESSER, V. Cooperative Multi-Agent Systems: A personal View of the State of the Art. **IEEE Transactions on Knowledge and Data Engineering**, v. 11, n. 1, p. 133-142, 1999.

LEVIS, Darci et al. Aperfeiçoamento Automático do Perfil do Aprendiz em Ambientes de Educação Ubíqua. **Revista Brasileira de Informática na Educação**, v. 16, n. 1, p. 29-41, 2008.

LÉVY, P. *Lintelligence collective: pour une anthropologie du cyperspace*. Paris: Découverte/ Poche, p. 21- 24, 1997.

MALONE, T. W.; CROWSTON, K. G. To- ward an interdisciplinary theory of coordina- tion. Tech. Rep. no. 120. Massachusetts Insti- tute of Technology, Center for Coordination Science, Cambridge, Mass. *ACM Computing Surveys*, v. 26, n. 1, March, 1994.

MENEZES, C.S. et al. MOrFEu – Multi-Organizador Flexível de Espaços Virtuais para Apoiar a Inovação Pedagógica em EAD. **Anais do XIX SBIE – Simpósio Brasileiro de Informática na Educação**. Fortaleza - CE, 2008.

MONTGOMERY, C. A. Automated Language Processing. **Annual Review of Information Science and Technology**, v. 4, p. 145-174, 1969.

MUSA, D. L.; OLIVEIRA, J. P. M. de. OntoLearner: Uma Ontologia para Perfis de Alunos Baseada em Padrões. In: XVIII SIMPÓSIO BRASILEIRO DE INFORMÁTICA EDUCAÇÃO - SBIE, Rio Grande do Sul, 2007, Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS) Mackenzie, p. 412 – 421.

NWANA, H.S.; LEE, L.; JENNINGS, N.R. Coordination in Software Agent Systems. **BT Technology Journal**, 14(4): pp. 79–88, 1996.

OLIVEIRA, F. A. D. **Processamento de Linguagem Natural: princípios básicos e a implementação de um analisador sintático de sentenças da língua portuguesa**. Programa de Pós-graduação em Computação, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2004. Disponível em: <[http://www.inf.ufrgs.br/procpar/disc/cmp135/trabs/992/Parser/parser.html#\\_Toc470452821](http://www.inf.ufrgs.br/procpar/disc/cmp135/trabs/992/Parser/parser.html#_Toc470452821)> . Acesso em: 03 jun. 2012.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA EDUCAÇÃO, CIÊNCIA E CULTURA. **Padrões de Competência em TIC para professores**. Marco Político. 2009. Disponível em: <<http://unesdoc.unesco.org/images/0015/001562/156210por.pdf>>. Acesso em: 05 jul. 2012.

PAGANI, L. A. (2004). Analisador gramatical em Prolog para gramáticas de estrutura sintagmática. **Revista Virtual de Estudos da Linguagem – ReVEL**. ano 2, n. 3. [[www.revelhp.cjb.net](http://www.revelhp.cjb.net)].

PAPI (2001). “PAPI Learner, Draft 8 Specification”, Disponível em: <<http://edutool.com/papi/>>. Acesso em: 08 jun. de 2012.

PERINI, M.A. (2006).Princípios de Lingüística Descritiva. São Paulo: Parábola.  
RUSSEL, S. J.; NORVIG P. **Artificial Intelligence: A Modern Approach**, Englewood Cliffs, Nova Jersey: Prentice Hall, 1995.

SATYANARAYANAN, M. **Pervasive Computing: Vision and Challenges**. IEEE Personal Communications. 2001. 8 v. p. 10-17.

SAVADOVSKY, P. **Introdução ao Projeto de Interfaces em Linguagem Natural**. São Paulo: SID Informática, 1988.

SCHILIT, B.; ADAMS, N.; WANT R. **Context-aware computing applications**. In: PROCEEDINGS OF THE WORKSHOP ON MOBILE COMPUTING SYSTEMS AND APPLICATIONS. 1994, p. 85-90.

SEARLE J. R. **Speech Acts: an Essay in the Philosophy of Language**. Cambridge University Press, Cambridge, 1969.

SEN, S.; WEISS, G. Learning in Multiagent Systems. **Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence**. In: WEISS, G. (Ed.). Cambridge: The MIT Press, 1999. p.259-298.

SERT. Secretaria Estadual do Emprego e Relações do Trabalho Disponível em: <[http://www.timedoemprego.sp.gov.br/Apostilas/C2/TE\\_VOL\\_1\\_2.pdf](http://www.timedoemprego.sp.gov.br/Apostilas/C2/TE_VOL_1_2.pdf)> Acesso em: 14 ago. 2012.

SICHMAN, J.; Demazeau, Y. (1995). Exploiting social reasoning to deal with agency level inconsistency. In **Proceedings of the 1 st International Conference on Multi-Agent Systems (ICMAS-95)**, San Francisco, CA, pp. 352-359.

SICHMAN, J. S. et al. (1994) A social reasoning mechanism based on dependence networks. In **Proceedings of the 11 th European Conference on Artificial Intelligence** (ECAI-94), Amsterdam, pp. 188-192.

SILVA, Gisele Trentin. **Adaptação em um sistema educacional hipermídia baseada na classificação de perfis de usuários**. 2005. 124 f. Dissertação (Mestrado em Ciência da Computação) – Curso de Pós-graduação em Ciências da Computação, Universidade Federal de Santa Catarina, Florianópolis.

SILVA, D. M.; BARBOSA, J.; VIEIRA, R. **OCtoPUS: um modelo para classificação de perfil de usuários usando trilhas em sistemas ubíquos**. In COMPANION PROCEEDINGS OF THE XIV BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB. Vila Velha, Espírito Santo, Brazil, 2008, p. 185-188.

SPECIA, L. **Modelagem de um Interpretador Lexical para a Linguagem DART**. Cascavel, 2000.

VIEIRA, R. Textual Co-reference annotation: a Study on Definite Descriptions. In: VIII CONGRESSO DA SOCIEDADE ARGENTINA DE LINGÜÍSTICA, Mar del Plata, Argentina, 2000.

VIEIRA, R; STRUBE, V. L.. **Lingüística Computacional: princípios e aplicações**. Rio Grande do Sul, [entre 2002 e 2004].

VISL., Visual Interactive Syntax Learning. Disponível em: <<http://beta.visl.sdu.dk/visl/pt/info/portsymbol.html>>. Acesso em: 13 Jul. 2012.

ZIMMER, Marco Vinício. **A criação de conhecimento em equipes virtuais: um estudo de caso em empresa do setor de alta tecnologia**. 2001. 113 f. Dissertação (Mestrado em Administração) – Programa de Pós- Graduação em Administração,

Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/3368/000292086.pdf?sequence=1>>. Acesso em: 21 jul. 2012.

WARNER, A.J. Natural Language Processing. **Annual Review of Information Science and Technology**. v. 22, p. 79-108, 1987.

WEISER M. **The computer for the 21st century**: vision and challenges. Scientific America, 1991, pp. 94-104

WOOLDRIDGE, M.; JENNINGS, N. Intelligent agents: Theory and Practice. **The Knowledge Engineering Review**, v. 10, n. 2, p. 115-152, 1995.

WOOLDRIDGE, M.; JENNINGS, N. R.; KINNY, D. A methodology for agent-oriented analysis and design. In: INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS, 3., 1999. Proceedings... Disponível em:<<http://www.springerlink.com/content/t1p651134r076800/>>. Acesso em: 10 jul. 2012.

## APÊNDICE A

Neste trabalho utilizamos o Analisador léxico e morfológico *Visual Interactive Syntax Learning* (VISL), para realizar as análises Léxicas e Morfológicas dos textos a serem processados pelo sistema.

O Projeto VISL teve como ponto de partida o sistema português PALAVRAS, e construiu ferramentas e bancos de dados Linguísticos para trabalhar através da WEB (AFONSO et al., 2001). Este projeto teve como (AFONSO et al., 2001):

“A principal motivação para o projeto Processamento computacional do português participar na Floresta era a construção de um recurso que pudesse eventualmente ser usado para a avaliação de analisadores sintáticos e outras ferramentas computacionais, a partir de uma base de objetos (árvores) comum, validada por linguistas.”

Neste sentido o projeto VISL divide o processo de construção da Floresta Sintática em duas fases: Uma de pré-processamento e a outra de revisão da análise automática.

Na primeira fase ocorreu os aspectos que não são morfossintáticos nem estruturais. Também houve a parte lexicográfica, que consistiu no enriquecimento do léxico português do PALAVRAS de cerca de 8.000 a 9.000 lexemas. Já a segunda fase consistiu na criação automática e revisão manual do corpus.

Para tornar o sistema mais robusto e eficiente, como também para aumentar o Recall o VISL baseou as ferramentas de análise no formalismo *Constraint Grammar- CG* (Gramática Constritiva), que visa auxiliar nas análises automáticas para determinar qual classe gramatical escolher quando uma palavra é classificada em várias classes. A CG utiliza-se de regras que analisam o contexto onde a palavra é empregada para determinar assim qual é realmente a classe gramatical da palavra (BICK, 2005). No estilo “me diga com quem andas que te direi quem és”.

Além da classificação morfológica o VISL também trata da classificação sintática, determinando a função de cada palavra na frase. Para a representação da estrutura sintática das palavras o analisador utiliza-se de algumas etiquetas específicas, que podem ser melhores estudadas no site oficial da ferramenta (VISL, 2012).

Com a análise léxica o VISL também identifica os sintagmas que compõem as sentenças, também representado por etiquetas, na notação “chata”. Não esquecendo também que o VISL lematiza os termos analisados, apresentando o texto lematizado.

No projeto VISL as ferramentas linguísticas, utilizando a CG, realizam várias tarefas cruciais referentes ao Processamento de Linguagem Natural que nossa proposta exige. Tal como apresentado anteriormente o VISL realiza análise Morfológica e sintática como também realiza a lematização dos termos.

O projeto utiliza-se de etiquetas (*tags*) para representar o resultado de suas análises. Com a utilização do método CG o processo de escolha entre classificações ambíguas, torna-se mais preciso e eficaz. Entretanto, para realizar este método de forma a classificar corretamente grande parte dos termos, é necessário um grande número de regras. As CG mais maduras contém tipicamente milhares de regras, cerca de 2.000 regras para cada nível (BICK, 2005).

O VISL oferece etiquetas semânticas que adicionam uma semântica maior nos metadados inseridos nos textos analisados. As etiquetas semânticas estão em fase experimental, e atualmente conta com 157 etiquetas semânticas (VISL, 2012).

As etiquetas em azul e negrito são etiquetas morfológicas (*Word Class Tags*) que determinam as classes das palavras, como por exemplo: N – Substantivo, V- Verbo, DET- Determinadores.

Já as etiquetas em azul são as *Inflection Tags* que apresentam informações como gênero, número (plural ou singular), pessoa (1º pessoa do plural), entre outros.

E no final de cada linha temos as etiquetas sintáticas que indicam a função sintática de cada palavra na sentença, como por exemplo (Sujeito, Objeto Direto, etc.). As etiquetas sintáticas estão representadas na cor verde e iniciam com o símbolo @.

### **TAG parser PALAVRAS VISL**

Veja as principais tags retornadas pelo parser PALAVRAS VISL.

#### **CLASSE TAGS WORD**

**N** Substantivos



**PROP** nomes próprios  
**SPEC** especificadores  
**DET** determinantes  
**PERS** pronomes pessoais  
**ADJ** adjetivos  
**ADV** Advérbios  
**V** verbos  
**NUM** numerais  
**PRP** preposição  
**KS** conjunções subordinativas  
**KC** conjunções Coordenadas  
**IN** Interjeições

### TAGS inflexão

**Gênero:** M (masculino), F (feminino), M / F [para: N ', PROP ", SPEC", DET, PERS, ADJ, V PCP, NUM]

**Número:** S (singular), P (plural), S / P [de: N, PROP ", SPEC", DET, PERS, ADJ, V PCP, V VFIN, INF, NUM]

**Caso :** NOM (nominativo), ACC (acusativo), DAT (dativo), PIV (prepositivo), ACC / DAT, NOM / PIV [para: PERS]

**Pessoa :** 1 (primeira pessoa), 2 (segunda pessoa), 3 (terceira pessoa), 1S, 1P, 2S, 2P, 3S, 3P, 1/3S, 0/1/3S [para: pessoas, V VFIN, V INF]

**Tempo:** PR (presente), IMPF (imperfecto), PS (simples Perfeito), MQP (Mais-que-Perfeito), FUT (futuro), COND (condicional) [para: V VFIN]

**Modo :** IND (indicativo), SUBJ (subjuntivo), IMP (imperativo) [para: V VFIN]

**Conjugação :** VFIN (verbo finito), INF (infinitivo), PCP (particípio), GER (gerúndio) [para: V]

### Tags Subclasse

**<artd>** artigo definido (DET)

**<arti>** artigo indefinido (DET)

**<quant>** pronome quantificador (DET: <quant1>, <quant2>, <quant3>, SPEC: <quant0>) ou advérbio intensificador

**<dem>** pronome demonstrativo (DET: <dem>, SPEC: <dem0>)

**<poss>** pronome possessivo (DET)

**<refl>** pronome reflexivo pessoal ("se" PERS ACC, "si" PERS PIV)

**<si>** uso reflexivo da 3. pessoa possessiva

**<reci>** utilização recíproca de pronome reflexivo (= "um ao outro")

**<coll>** coletivo reflexivo ("reunir-se", "associar-se")

**<diff>** diferenciador (DET) (por exemplo, "e outros temas", "a mesma diferença")

**<ident>** identificador (DET) (por exemplo, "o próprio usuário", "a si mesmo")

**<rel>** pronome relativo (DET, SPEC)

**<interr>** pronome interrogativo (DET, SPEC)

**<post-det>** normalmente localizado como pós-determinante (DET @ N <)

**<post-attr>** tipicamente pós-posicionado adjetivo (ADJ @ N <)

<ante-attr>tipicamente pré-posicionado adjetivo (ADJ @> N)  
 <adv> pode ser usado adverbialmente (ADJ @ ADVL)  
 <ks>advérbio relativo usado como uma conjunção subordinação  
 <kc> advérbio conjuntural (POIs, EntreTanto)  
 <det> determinante uso / inflexão do advérbio ("Ela estava nua Toda").  
 <foc> concentrar advérbio marcador (também formas de "ser")  
 <prp> advérbio relativo usado como uma preposição

### Etiquetas de valência:

<vt> verbo monotransitivo com o objeto acusativo  
 <vi> (<ve>) verbo intransitivo (verbo negativo)  
 <vtd> verbo intransitivo com objetos acusativos e dativos  
 <PRP^vp> verbo monotransitivo com o objeto preposicional  
 <va> verbo transitivo com o argumento adverbial:  
 <vU> "verbos impessoais"  
 <x> verbo auxiliar com infinitivo  
 <x+PCP> verbo auxiliar com particípio  
 <x+GER> verbo auxiliar com gerúndio  
 <PRP^xp> verbo auxiliar com preposição  
 <xt> verbo "auxiliar" com o assunto infinitivo cláusula no caso acusativo  
 <PRP^xtp> verbo "auxiliar" com o objeto acusativo preposicionado e objeto que contém uma cláusula de infinitivo com seu sujeito  
 <vr> verbos reflexivos  
 <qv> verbo "impessoal"

### Tags semânticas para substantivos

#### Protótipos de origem animal:

<A> Animal, guarda-chuva tag (clone, Fêmea, fóssil, parasito, predador)  
 <AA> grupo de animais (cardume, enxame, passarada, ninhada)  
 <Adom> animal doméstico ou mamífero de grande porte (propensos a ter formas femininas, etc: terneiro, Leão / Leoa, cachorro)  
 <AAdom> grupo de animais domésticos (boiada)  
 <Aich> Água animal (tubarão, Delfim)  
 <Amyth> animal mitológico (basilisco)  
 <Azo> animal terrestre (raposa)  
 <Aorn> Pássaro (águia, bem-te-vi)  
 <Aent> inseto (borboleta)  
 <Acell> célula animal (bactérias, células do sangue: Linfócito)

#### Protótipos de plantas:

<B> Plant, guarda-chuva tag  
 <BB> Grupo de plantas, o plantio (campo, floresta, etc: mata, nabal)  
 <Btree> Tree (oliveira, palmeira)  
 <Bflo> flor (rosa, taraxaco)  
 <Bbush> Bush, arbusto (Rododendro, Tamariz)

**Protótipos humanos:**

- <H> Humano, guarda-chuva tag
- <HH> grupo de seres humanos (organizações, grupos, empresas, por exemplo, da Editora)
- <Hattr> tag guarda-chuva Atributivo humana (muitos -ista,-ante)
- <Hbio> Humano classificados por critérios biológicos (idade, raça etc, caboclo, mestiço, bebê, adulto)
- <Hfam> Humano relação com a família ou privadas (pai ,noiva)
- <Hideo> humano ideológica (comunista , implica <Hattr>), também: seguidor, discípulo (dadaísta)
- <Hmyth > Humanoid mítica (deuses, fadas humanóides conto, Curupira, duende)
- <Hnat> Nacionalidade humano (brasileiro , alemão), também: habitante (lisboeta)
- <Hprof> Profissional humano (marinheiro , implica <Hattr>), também: passatempo esporte, (alpinista)
- <Hsick> Doente humano (alguns: asmático, diabetico , cp <sick>)
- <Htit> substantivo Título (rei , minha senhora)

**Local e protótipos espaciais:**

- <L> lugar
- <Labs> lugar abstrato
- <Lciv> civilizações, cidade, país, região (igual <L> + <HH>, *Cidade, País*)
- <Lcover> Cover, tampa (*colcha, lona, tampa*)
- <Lh> lugar funcional humana, construído ou humano-utilizado (*Aeroporto*)
- <Lopening> abertura, buraco
- <Lpath> acessos (estrada, rua etc .: *rua, pista*)
- <Lstar> objeto estelar (planetas, cometas: *Planeta, quasar*)
- <Lsurf> superfície (*face, verniz .*, cp <Lcover>)
- <Ltip> lugar ponta, ponta (*pico, Pontinha* , cp <. Labs>)
- <Ltop>lugar, geográfica natural (*Promontório, pântano*)
- <Ltrap> lugar armadilha (*Armadilha, armazelo*)
- <Lwater> lugar de água (rio, lago, mar: *Fonte, foz, lagoa*) .

**Protótipos de veículos:**

- <V> veículo, marca guarda-chuva e veículo terrestre (carro, comboio: *o carro, Comboio, tanque, Teleférico*)
- <VV> grupo de veículos (frota, comboio: *Frota, esquadra*)
- <Vwater> veículo Água (navio: *Navio, submersível, Canoa*)
- <Vair> veículo aéreo (avião: *hidroplano, jatinho*)

**Protótipos Resumo:**

- <ac> Resumo contável
- <ac-sign> sinal, símbolo (*parêntese, semicolcheia*)
- <am> massa abstrata / não-contável

<ax> Resumo / conceito, nem contábil nem massa

### Protótipos de ação:

<act> Ação  
 <act-beat> batida ação (surra, *pancada*)  
 <act-d> fazer-ação (normalmente Dar / Fazer + N, *tentativa, Homenagem*)  
 <act-s> ato de fala ou ato comunicativo (*proposta, ordem*)  
 <act-trick> truque ação (fraude, *jeito*)

### Protótipos Anatômicos:

<an> substantivo anatômico, marca guarda-chuva (*carótida, clítoris, dorso*)  
 <anmov> anatomia móvel (braço, perna, *Braço, bíceps, Cotovelo*)  
 <anorg> órgãos (coração, fígado, *hipófise, Coração, Testículo*)  
 <anost> óssea (*calcâneo, fíbula, vértebra*)  
 <anzo> anatomia Animal (*rúmen, carapaça, chifres, tromba*)  
 <anorn> anatomia Bird (*bico, pluma*)  
 <anich> anatomia dos peixes (alguns: *brânquias, Siba*)  
 <anent> anatomia do inseto (alguns : *tentáculo, Olho Composto*)  
 <anbo> Anatomia Vegetal (*bulbo, caule, a Folha*)  
 <amount> substantivo quantidade (*bocada, Teor, sem-FIM*)  
 <bar> substantivo barreira (*dique, limite, muralha*)  
 <build> edifício (*casa, Cidadela, garagem*).

### Protótipos coisa:

<cc-beauty> objeto ornamental  
 <cc-fire> fogo (fogueira, centelha, *chispa, fogo, Girandola*)  
 <cc-handle> alça (*garra, alça, chupador*)  
 <cc-light> artefato de luz (*lâmpião, farol*)  
 <cc-r>objeto de leitura (*cupom, bilhete, carta*)  
 <cc-rag> pano (toalha, guardanapo, tapete, pano)

### Protótipos Coletivos:

<coll> conjunto, coletivo (conjunto aleatório ou sistemático / composto / multidão de semelhantes, porém distintos pequenas peças, Conjunto, Série)  
 <coll-cc> coisa coletiva, pilha (baralho)  
 <coll-B> planta parte coletiva (Buque, folhagem)  
 <coll-sem> coletiva semântica, a coleta (ARQUIVO, Repertório)  
 <coll-tool> ferramenta conjunto, coletivo (instrumentário, prataria)  
 <domínio> domínio (assunto, profissão)  
 <drink> bebidas (*cachaça, leite, guaraná, moca*)

### Protótipos de Alimentos:

<food> alimentos (*aveia, açúcar, carne*, até agora)

<food-c> alimentos contábil (*ovo, dente de alho*)

<food-h> alimentos preparados (*caldo verde, lasanha*)

<food-ch> comida preparados contáveis (*biscoito, pastel*)

<fight> luta, o conflito

<fruit> frutas, nozes

### Protótipos percepção:

<percep-f> o que você sente (sentidos ou sentimentos, dor, *arrepio, aversão, desgosto, cócegas*)

<percep-w> impressão visual (que você vê, *arco-íris, vislumbre*)

<pict> imagem

### OUTRAS CATEGORIAS Semanticamente

<vH> verbo com sujeito humano

<vN> verbo com sujeito inanimado

## APÊNDICE B

Segue textos que etiquetados que foram utilizados no experimento do protótipo.

### Texto de Arnaldo:

Hoje estou estudando inglês muito, até que ontem sai com meu tio Wagner e meu primo Eduardo, tomamos umas cervejas e por fim, Eduardo comeu pizza, eu e Wagner degustamos queijo e tomamos vinho.

Sáimos na noite passada e ainda assim acordei bem disposto no sábado, ajudei a minha esposa a limpar a casa, depois a esposa foi fazer o almoço e eu fiquei organizando o jardim, cortei a grama, rastelei e irriguei.

### Texto Etiquetado de Arnaldo:

Hoje [hoje] **ADV @ADVL** #1->2  
 estou [estar] <fmc> <aux> **V PR 1S IND VFIN @FS-STA** #2->0  
 estudando [estudar] <clb> <vH> <mv> **V GER @ICL-AUX** #3->20  
 inglês [inglês] <ling> <Hnat> **N M S @<ACC** #4->3  
 muito [muito] <quant> **DET M S @<ACC** #5->4  
 , #6->0  
 até [até] **PRP @<ADVL** #7->3  
 que [que] <clb> <clb-fs> <rel> **SPEC M S @SUBJ** #8->10  
 ontem [ontem] **ADV @ADVL** #9->10  
 sai [sair] <mv> **V PR 3S IND VFIN @FS-P** #10->7  
 com [com] **PRP @<PIV** #11->10  
 meu [meu] <poss 1S> **DET M S @>N** #12->13  
 tio [tio] <Hfam> **N M S @P** #13->11  
 Wagner [Wagner] <hum> <np-close> **PROP M S @N** #14->13  
 e [e] <co-prparg> **KC @CO** #15->13  
 meu [meu] <poss 1S> **DET M S @>N** #16->17  
 primo [primo] <Hfam> **N M S @P** #17->13

Eduardo [Eduardo] <hum> <np-close> **PROP M S @N**< #18->17  
 , #19->0  
 tomamos [tomar] <vH> <fmc> <mv> **V PR 1P IND VFIN @FS-STA** #20->2  
 umas [um] <quant> **DET F P @>N** #21->22  
 cervejas [cerveja] <drink> **N F P @<ACC** #22->20  
 e [e] <co-fmc> <co-fin> **KC @CO** #23->2  
 por=fim [por=fim] <pp> **ADV @ADVL**> #24->27  
 , #25->0  
 Eduardo [Eduardo] <hum> **PROP M S @SUBJ**> #26->27  
 comeu [comer] <vH> <fmc> <mv> **V PS 3S IND VFIN @FS-STA** #27->2  
 pizza [pizza] <food-c-h> **N F S @<ACC** #28->27  
 , #29->0  
 eu [eu] **PERS M/F 1S NOM @SUBJ**> #30->33  
 e [e] <co-subj> **KC @CO** #31->30  
 Wagner [Wagner] <hum> **PROP M S @SUBJ**> #32->30  
 degustamos [degustar] <vH> <fmc> <mv> **V PR 1P IND VFIN @FS-STA** #33->2  
 queijo [queijo] <food-c-h> <food-h> **N M S @<ACC** #34->33  
 e [e] <co-fmc> <co-fin> **KC @CO** #35->2  
 tomamos [tomar] <vH> <fmc> <mv> **V PR 1P IND VFIN @FS-STA** #36->2  
 vinho [vinho] <drink> **N M S @<ACC** #37->36  
 . #38->0  
 </s>

Saímos [sair] <fmc> <mv> **V PR 1P IND VFIN @FS-STA** #1->0  
 em [em] <sam-> **PRP @<ADVL** #2->1  
 a [o] <artd> <-sam> **DET F S @>N** #3->4  
 noite [noite] <temp> <per> **N F S @P<** #4->2  
 passada [passado] <np-close> **ADJ F S @N<** #5->4  
 e [e] <co-fmc> <co-fin> **KC @CO** #6->1  
 ainda=assim [ainda=assim] <kc> **ADV @ADVL**> #7->8

acordei [acordar] <fmc> <mv> **V PS 1S IND VFIN @FS-STA #8->1**  
 bem [bem] <quant> **ADV @>A #9->10**  
 disposto [dispor] <vH> <jh> <mv> **V PCP M S @ICL-<SC #10->8**  
 em [em] <sam-> **PRP @<ADVL #11->10**  
 o [o] <artd> <-sam> **DET M S @>N #12->13**  
 sábado [sábado] <temp> **N M S @P< #13->11**  
 , #14->0  
 ajudei [ajudar] <vH> <fmc> <mv> **V PS 1S IND VFIN @FS-STA #15->1**  
 a [o] <artd> **DET F S @>N #16->18**  
 minha [meu] <poss 1S> **DET F S @>N #17->18**  
 esposa [esposa] <Hfam> **N F S @<ACC #18->15**  
 a [a] **PRP @<ADVL #19->15**  
 limpar [limpar] <mv> **V INF @ICL-P< #20->19**  
 a [o] <artd> **DET F S @>N #21->22**  
 casa [casa] <build> **N F S @<ACC #22->20**  
 , #23->0  
 depois [depois] **ADV @ADVL> #24->27**  
 a [o] <artd> **DET F S @>N #25->26**  
 esposa [esposa] <Hfam> **N F S @SUBJ> #26->27**  
 foi [ser] <fmc> <mv> **V PS 3S IND VFIN @FS-STA #27->1**  
 fazer [fazer] <mv> **V INF @ICL-<SC #28->27**  
 o [o] <artd> **DET M S @>N #29->30**  
 almoço [almoço] <occ> **N M S @<ACC #30->28**  
 e [e] <co-fmc> <co-fin> **KC @CO #31->1**  
 eu [eu] **PERS M/F 1S NOM @SUBJ> #32->33**  
 fiquei [ficar] <fmc> <aux> **V PS 1S IND VFIN @FS-STA #33->1**  
 organizando [organizar] <clb> <vH> <mv> **V GER @ICL-AUX< #34->38**  
 o [o] <artd> **DET M S @>N #35->36**  
 jardim [jardim] <Lh> <BB> **N M S @<ACC #36->34**  
 , #37->0



cortei [cortar] <fmc> <mv> V PS 1S IND VFIN @FS-STA #38->1  
 a [o] <artd> DET F S @>N #39->40  
 grama [grama] <Btree> <BB> N F S @<ACC #40->38  
 , #41->0  
 rastelei [rastelar] <vH> <fmc> <mv> V PS 1S IND VFIN @FS-STA #42->1  
 e [e] <co-fin> KC @CO #43->1  
 irriguei [irrigar] <vH> <fmc> <mv> V PS 1S IND VFIN @FS-STA #44->1  
 . #45->0  
 </s>

### Texto de Rômulo:

Meu fim de semana foi só correria. Eu sou engenheiro mecânico da empresa FabricaMec e estava de plantão, tive que resolver vários problemas. Estamos construindo uma nova fábrica, anexa à FabricaMec1, mas neste plantão tive que consertar a tubulação de nitrogênio e subir no topo da fábrica para ter acesso a tubulação. Marcos estava soldando as conexões e eu estava fixando a tubulação na calha.

Amanhã vou jogar vôlei de areia na praia, eu gosto de praia.

### Texto etiquetado de Rômulo

Meu [meu] <poss 1S> DET M S @>N #1->2  
 fim=de=semana [fim=de=semana] <per> <temp> N M S @SUBJ> #2->3  
 foi [ser] <vK> <fmc> <mv> V PS 3S IND VFIN @FS-STA #3->0  
 só [só] ADV @>N #4->5  
 correria [correria] <occ> <activity> N F S @<SC #5->3  
 . #6->0  
 </s>

Eu [eu] PERS M 1S NOM @SUBJ> #1->2  
 sou [ser] <vK> <fmc> <mv> V PR 1S IND VFIN @FS-STA #2->0  
 engenheiro [engenheiro] <Hprof> N M S @<SC #3->2

mecânico [mecânico] <nh> <np-close> **ADJ M S @N<** #4->3  
 de [de] <sam-> <np-close> **PRP @N<** #5->3  
 a [o] <artd> <-sam> **DET F S @>N** #6->7  
 empresa [empresa] <HH> **N F S @P<** #7->5  
 FabricaMec [FabricaMec] <org> <np-close> **PROP F S @N<** #8->7  
 e [e] <co-fmc> <co-fin> **KC @CO** #9->2  
 estava [estar] <vK> <fmc> <mv> **V IMPF 1/3S IND VFIN @FS-STA** #10->2  
 de [de] **PRP @<SC** #11->10  
 plantão [plantão] <H> **N M S @P<** #12->11  
 , #13->0  
 tive [ter] <fmc> <aux> **V PS 1S IND VFIN @FS-STA** #14->2  
 que [que] **KS @PRT-AUX<** #15->14  
 resolver [resolver] <vH> <mv> **V INF @ICL-AUX<** #16->14  
 vários [vários] <quant> **DET M P @>N** #17->18  
 problemas [problema] <ac> **N M P @<ACC** #18->16  
 . #19->0  
 </s>

Estamos [estar] <fmc> <aux> **V PR 1P IND VFIN @FS-STA** #1->0  
 construindo [construir] <clb> <vH> <mv> **V GER @ICL-AUX<** #2->1  
 uma [um] <arti> **DET F S @>N** #3->5  
 nova [novo] **ADJ F S @>N** #4->5  
 fábrica [fábrica] <build> **N F S @<ACC** #5->2  
 , #6->0  
 anexa [anexo] <nh> <np-close> **ADJ F S @N<PRED** #7->5  
 a [a] <sam-> **PRP @<ADVL** #8->2  
 a [o] <artd> <-sam> **DET F S @>N** #9->10  
 FabricaMec1 [FabricaMec1] <org> **PROP F S @P<** #10->8  
 , #11->0  
 mas [mas] <clb> **KC @CO** #12->1

em [em] <sam-> **PRP** @<ADVL #13->2  
 este [este] <dem> <-sam> **DET M S** @>N #14->15  
 plantão [plantão] <H> **N M S** @P< #15->13  
 tive [ter] <fmc> <aux> **V PS 1S IND VFIN** @FS-STA #16->1  
 que [que] **KS** @PRT-AUX< #17->16  
 consertar [consertar] <vH> <mv> **V INF** @ICL-AUX< #18->16  
 a [o] <artd> **DET F S** @>N #19->20  
 tubulação [tubulação] <tube> **N F S** @<ACC #20->18  
 de [de] <np-close> **PRP** @N< #21->20  
 nitrogênio [nitrogênio] <cm-chem> **N M S** @P< #22->21  
 e [e] <co-inf> **KC** @CO #23->18  
 subir [subir] <mv> **V INF** @ICL-AUX< #24->18  
 em [em] <sam-> **PRP** @<ADVL #25->24  
 o [o] <artd> <-sam> **DET M S** @>N #26->27  
 topo [topo] <event> <Ltip> **N M S** @P< #27->25  
 de [de] <sam-> <np-close> **PRP** @N< #28->27  
 a [o] <artd> <-sam> **DET F S** @>N #29->30  
 fábrica [fábrica] <build> **N F S** @P< #30->28  
 para [para] **PRP** @<ADVL #31->24  
 ter [ter] <mv> **V INF** @ICL-P< #32->31  
 acesso [acesso] <Lpath> <f-right> <event> **N M S** @<ACC #33->32  
 a [o] <artd> **DET F S** @>N #34->35  
 tubulação [tubulação] <tube> **N F S** @<SUBJ #35->32  
 . #36->0  
 </s>

Marcos [Marcos] <hum> **PROP M/F S** @SUBJ> #1->2  
 estava [estar] <fmc> <aux> **V IMPF 3S IND VFIN** @FS-STA #2->0  
 soldando [soldar] <clb> <vH> <mv> **V GER** @ICL-AUX< #3->8  
 as [o] <artd> **DET F P** @>N #4->5

conexões [conexão] <cc> <ac> **N F P @<ACC #5->3**  
 e [e] <co-fmc> <co-fin> **KC @CO #6->2**  
 eu [eu] **PERS M/F 1S NOM @SUBJ> #7->8**  
 estava [estar] <fmc> <aux> **V IMPF 1S IND VFIN @FS-STA #8->2**  
 fixando [fixar] <clb> <vH> <mv> **V GER @ICL-AUX< #9->8**  
 a [o] <artd> **DET F S @>N #10->11**  
 tubulação [tubulação] <tube> **N F S @<ACC #11->9**  
 em [em] <sam-> <np-close> **PRP @N< #12->11**  
 a [o] <artd> <-sam> **DET F S @>N #13->14**  
 calha [calha] <part-build> **N F S @P< #14->12**  
 . #15->0  
 </s>

Amanhã [amanhã] **ADV @ADVL> #1->3**  
 vou [ir] <fmc> <aux> **V PR 1S IND VFIN @FS-STA #2->0**  
 jogar [jogar] <vH> <mv> **V INF @ICL-AUX< #3->2**  
 vôlei [vôlei] <sport> **N M S @<ACC #4->3**  
 de [de] <np-close> **PRP @N< #5->4**  
 areia [areia] <mat> **N F S @P< #6->5**  
 em [em] <sam-> **PRP @<ADVL #7->3**  
 a [o] <artd> <-sam> **DET F S @>N #8->9**  
 praia [praia] <Ltop> **N F S @P< #9->7**  
 , #10->0  
 eu [eu] **PERS M/F 1S NOM @SUBJ> #11->12**  
 gosto [gostar] <vH> <fmc> <mv> **V PR 1S IND VFIN @FS-STA #12->2**  
 de [de] **PRP @<PIV #13->12**  
 praia [praia] <Ltop> **N F S @P< #14->13**  
 . #15->0  
 </s>

**Texto de Juliana:**

Meu nome é Juliana, não gosto de ficar parada. Estou sempre querendo o melhor e aprender mais, tenho paixão pelo trabalho. Entrei na MRTextil como Assistente Técnico.

Estudei, observei, perguntei, propus novas ideias e agora sou a Gerente da empresa, responsável para que tudo dê certo. Faço toda a Gestão Financeira e controle dos contratos com outras empresas terceirizadas.

Atualmente, estou me especializando em Gestão de Pessoas.

Adoro estar com minha família e, nos momentos oportunos, aproveito para viajar e estar com os amigos.

**Texto Etiquetado de Juliana:**

Meu [meu] <poss 1S> **DET M S @>N #1->2**  
 nome [nome] <ac-cat> <percep-f> **N M S @SUBJ> #2->3**  
 é [ser] <vK> <fmc> <mv> **V PR 3S IND VFIN @FS-STA #3->0**  
 Juliana [Juliana] <hum> **PROP M S @<SC #4->3**  
 , #5->0  
 não [não] **ADV @ADVL> #6->7**  
 gosto [gostar] <vH> <fmc> <mv> **V PR 1S IND VFIN @FS-STA #7->3**  
 de [de] **PRP @<PIV #8->7**  
 ficar [ficar] <vK> <mv> **V INF @ICL-P< #9->8**  
 parada [parar] <jh> **V PCP F S @<SC #10->9**  
 . #11->0  
 </s>

Estou [estar] <fmc> <aux> **V PR 1S IND VFIN @FS-STA #1->0**  
 sempre [sempre] **ADV @ADVL> #2->3**  
 querendo [querer] <clb> <vH> <mv> **V GER @ICL-AUX< #3->10**  
 o [o] <artd> **DET M S @>N #4->5**  
 melhor [bom] <KOMP> <SUP> <jh> <n> **ADJ M S @<ACC #5->3**

e [e] <co-fmc> <co-fin> **KC @CO #6->1**  
 aprender [aprender] <vH> <mv> **V INF @ICL-UTT #7->0**  
 mais [mais] **ADV @<ADVL #8->7**  
 , #9->0  
 tenho [ter] <fmc> <mv> **V PR 1S IND VFIN @FS-STA #10->1**  
 paixão [paixão] <percep-f> <am> **N F S @<ACC #11->10**  
 por [por] <sam-> <np-close> **PRP @N< #12->11**  
 o [o] <artd> <-sam> **DET M S @>N #13->14**  
 trabalho [trabalho] <am> <act-d> **N M S @P< #14->12**  
 . #15->0  
 </s>

Entrei [entrar] <vH> <fmc> <mv> **V PS 1S IND VFIN @FS-STA #1->0**  
 em [em] <sam-> **PRP @<SA #2->1**  
 a [o] <artd> <-sam> **DET F S @>N #3->4**  
 MRTeXtil [MRTeXtil] <inst> **PROP F S @P< #4->2**  
 como [como] <com> <adv-rel> **PRP @<ADVL #5->1**  
 Assistente=Técnico [Assistente=Técnico] <hum> <asarg> **PROP M S @P< #6->5**  
 . #7->0  
 </s>

Estudei [estudar] <vH> <fmc> <mv> **V PS 1S IND VFIN @FS-STA #1->0**  
 , #2->0  
 observei [observar] <vH> <fmc> <mv> **V PS 1S IND VFIN @FS-STA #3->1**  
 , #4->0  
 perguntei [perguntar] <vH> <fmc> <mv> **V PS 1S IND VFIN @FS-STA #5->1**  
 , #6->0  
 propus [propor] <vH> <fmc> <mv> **V PS 1S IND VFIN @FS-STA #7->1**  
 novas [novo] <jh> **ADJ F P @>N #8->14**  
 ideias [ideia] <sem-c> **N F P @<ACC #9->7**

e [e] <co-fmc> <co-fin> **KC @CO** #10->1  
 agora [agora] <kc> **ADV @ADVL** #11->12  
 sou [ser] <fmc> <mv> **V PR 1S IND VFIN @FS-STA** #12->1  
 a [o] <artd> **DET F S @>N** #13->14  
 Gerente [gerente] <Hprof> <prop> **N F S @P<** #14->0  
 de [de] <sam-> <np-close> **PRP @N<** #15->14  
 a [o] <artd> <-sam> **DET F S @>N** #16->17  
 empresa [empresa] <HH> **N F S @P<** #17->15  
 , #18->0  
 responsável [responsável] <jh> <np-close> **ADJ M/F S @N<PRED** #19->17  
 para [para] **PRP @<ADVL** #20->12  
 que [que] <clb> <clb-fs> **KS @SUB** #21->23  
 tudo [tudo] <quant> **SPEC M S @ACC>** #22->23  
 dê [dar] <mv> **V PR 1/3S SUBJ VFIN @FS-P<** #23->20  
 certo [certo] **ADJ M S @<PRED** #24->23  
 . #25->0  
 </s>

Faço [fazer] <fmc> <mv> **V PR 1S IND VFIN @FS-STA** #1->0  
 toda=a [todo=o] <quant> **DET F S @>N** #2->3  
 Gestão=Financeira [Gestão=Financeira] <occ> **PROP F S @<ACC** #3->1  
 e [e] <co-acc> **KC @CO** #4->3  
 controle [controle] <am> **N M S @<ACC** #5->3  
 de [de] <sam-> <np-close> **PRP @N<** #6->5  
 os [o] <artd> <-sam> **DET M P @>N** #7->8  
 contratos [contrato] <sem-r> **N M P @P<** #8->6  
 com [com] <np-close> **PRP @N<** #9->8  
 outras [outro] <KOMP> <diff> **DET F P @>N** #10->11  
 empresas [empresa] <HH> **N F P @P<** #11->9  
 terceirizadas [terceirizar] <vH> <np-close> **V PCP F P @N<PRED** #12->11

. #13->0

</s>

Atualmente [atualmente] **ADV @ADVL** #1->5

, #2->0

estou [estar] <fmc> <aux> **V PR 1S IND VFIN @FS-STA** #3->0

me [eu] <refl> **PERS M/F 1S ACC @ACC** #4->5

especializando [especializar] <clb> <vH> <mv> **V GER @ICL-AUX<** #5->3

em [em] **PRP @<PIV** #6->5

Gestão=de=Pessoas [Gestão=de=Pessoas] <occ> **PROP M/F S @P<** #7->6

. #8->0

</s>

Adoro [adorar] <vH> <fmc> <mv> **V PR 1S IND VFIN @FS-STA** #1->0

estar [estar] <vK> <mv> **V INF @ICL-<ACC** #2->1

com [com] **PRP @<SC** #3->2

minha [meu] <poss 1S> **DET F S @>N** #4->5

família [família] <HH> **N F S @P<** #5->3

e [e] <co-fmc> <co-fin> **KC @CO** #6->1

, #7->0

em [em] <sam-> **PRP @ADVL** #8->13

os [o] <artd> <-sam> **DET M P @>N** #9->10

momentos [momento] <dur> <f-q> **N M P @P<** #10->8

oportunos [oportuno] <nh> <np-close> **ADJ M P @N<** #11->10

, #12->0

aproveito [aproveitar] <vH> <fmc> <mv> **V PR 1S IND VFIN @FS-STA** #13->1

para [para] **PRP @<ADVL** #14->13

viajar [viajar] <vH> <mv> **V INF @ICL-P<** #15->14

e [e] <co-inf> **KC @CO** #16->15

estar [estar] <vK> <mv> **V INF @ICL-P<** #17->15



com [com] PRP @<SC #18->17  
os [o] <artd> DET M P @>N #19->20  
amigos [amigo] <Hfam> N M P @P< #20->18  
. #21->0  
</s>