

## Uma ferramenta para resolução automática de correferência

Rodrigo Goulart, Caroline Gasperin, Renata Vieira

Programa de Pós Graduação em Computação Aplicada  
Centro de Ciências Exatas e Tecnológicas  
Universidade do Vale do Rio dos Sinos,  
São Leopoldo, RS – Brasil  
{rodrigo,caroline,renata}@exatas.unisinos.br

### ABSTRACT

This work presents the architecture of a tool for automatic coreference resolution in texts. The tool deals with syntactic analysed corpora encoded in XML. The anaphor resolution process and the encoding formats are described in this paper.

### RESUMO

Este trabalho apresenta a arquitetura de uma ferramenta para resolução automática de correferência em textos. A ferramenta trata corpora analisados sintaticamente e codificados em XML. Tais formatos são detalhados e as etapas do processo de resolução das anáforas são descritas neste artigo.

## 1 Introdução

Estamos desenvolvendo uma ferramenta multilíngüe para resolução de expressões anafóricas como as descrições definidas (*os garotos*), demonstrativas (*esses garotos*) e pronomes (*eles*).

Uma expressão é dita anafórica quando esta se refere a uma entidade que foi previamente referenciada no texto por outra expressão. Ambas as expressões são consideradas correferentes, e a

primeira é vista como o antecedente da segunda. A informação sobre quais expressões são correferentes em um texto pode ser de grande utilidade em sistemas de recuperação e extração de informação, classificação de textos, sumarização de textos, entre outros.

Nosso processo de resolução das anáforas é baseado em heurísticas geradas através de um estudo prévio do comportamento das expressões correferentes em textos escritos (7). A ferramenta proposta tem como formato a linguagem de marcação XML para entrada, manipulação e saída dos dados.

Na seção seguinte, é detalhado o fenômeno da correferência e também os esquemas de anotação de corpus utilizados em nossos estudos. Na seção 3, são apresentados os principais aspectos da ferramenta proposta: o formato dos dados, as heurísticas que serão utilizadas e a arquitetura do sistema. Ao final, são apresentadas as conclusões deste trabalho.

## 2 Correferência

O fenômeno de correferência que ocorre na linguagem natural consiste em duas ou mais expressões de um texto se referirem a uma mesma entidade do discurso. Quando uma entidade é referenciada pela primeira vez em um texto, a expressão que a descreve é dita nova no discurso. Quando tal entidade é retomada no texto, a expressão que a descreve é dita anafórica, sendo considerado seu antecedente a expressão anterior correferente. As expressões anafóricas podem ser de diferentes tipos, como:

- pronominais: o termo anafórico é um pronome. Ex: *O menino leu o livro, mas **ele** não gostou.*
- definidas: compostas por um substantivo antecedido por um artigo definido. Ex: *Comprei uma casa. **A casa** fica longe daqui.*
- demonstrativas: compostas por um substantivo antecedido por um pronome demonstrativo. Ex: *Comprei uma casa. **Essa casa** será sempre minha.*

## 2.1 Esquemas XML para anotação de correferência

Em (11; 10) são detalhados estudos feitos em um corpus contendo textos escritos em Português e Francês, em que as expressões anafóricas foram marcadas manualmente, bem como seus antecedentes no texto. Esse processo de marcação foi feito com a utilização da ferramenta MMAX (*Multi-Modal Annotation in XML*) (5). O MMAX requer que os dados estejam representados em XML, seguindo uma estrutura determinada. O corpus deve estar de acordo com o formato apresentado na Figura 1.

```
...  
<word id="word_10">Os</word>  
<word id="word_11">acidentes</word>  
<word id="word_12">lotaram</word>  
...
```

Figura 1. Formato do corpus.

De acordo com as tendências para padronização da anotação de corpora apresentadas em (3), adotou-se o arquivo de palavras, como proposto em (4), como o repositório básico do corpus, para o qual todas as demais anotações lingüísticas apontam. Chama-se marcação *stand-off* aquela feita separadamente do corpus, isto é, em que as marcações são armazenadas em outro repositório diferente daquele onde está o corpus. Tais marcações contêm apontadores para os elementos do corpus a que se referem. Com isso, a integridade do texto não é comprometida pela anotação e pode-se ter, mais facilmente, diversos níveis de anotação sobre um mesmo corpus.

O resultado da anotação de correferência é apresentado como mostra a Figura 2, onde cada expressão referencial analisada é representada por um elemento <markable>, cujo atributo *span* indica as palavras que a formam, o atributo *pointer* indica o identificador do antecedente, e o atributo *classification* corresponde à natureza anafórica da descrição definida, como por exemplo anáfora direta ou nova no discurso.

### 3 Etapas da marcação manual

Em nossos experimentos, a marcação manual foi feita por mais de um anotador para verificar a concordância entre suas marcações ao final. No entanto, em nosso primeiro experimento, verificamos que a concordância entre os anotadores foi prejudicada por algumas dificuldades no processo de marcação, pois toda a anotação devia ser feita em uma única tarefa: identificar as expressões anafóricas, identificar seu antecedente (expressão correferente) e finalmente classificar a coreferência em cada caso. Concluimos, então, que eram muitas decisões a serem tomadas pelo anotador em um único passo.

Desta forma, para facilitar o trabalho dos anotadores e, consequentemente, melhorar sua concordância, decidimos dividir o processo de marcação em 4 passos:

1. selecionar as expressões anafóricas: esta tarefa é feita por apenas um anotador, já que não há discordância sobre quais são as expressões anafóricas do texto.
2. identificar o antecedente para cada expressão anafórica, se possível;
3. classificar os correferentes: se a expressão anafórica tem um antecedente e esse é correferente, a relação deve ser classificada.
4. classificar os não correferentes: se a expressão anafórica não tem um antecedente ou se seu antecedente não é correferente, a relação deve ser classificada.

```
...  
<markable id="markable_2"  
  pointer="markable_4"  
  span="word_10..word_11"  
  classification="direct"/>  
...
```

Figura 2. Formato da marcação.

O MMAX não estava completamente adaptado a trabalhar com diversos passos para a marcação, ou seja, a ter o processo de marcação dividido, não garantindo a consistência entre a anotação feita em passos diferentes. Para tratar esse problema, desenvolvemos uma nova maneira de especificar os esquemas de marcação do usuário no MMAX e a propusemos aos desenvolvedores do MMAX, que implementaram nossas idéias (6).

Primeiramente, os esquemas de anotação do MMAX não permitiam relacionar um atributo com outro, de forma que seus valores eram independentes. Assim, o usuário não podia restringir a marcação de um atributo dependendo do valor marcado previamente para outro atributo. Por causa disso, todas as combinações de valores de atributos eram disponibilizadas ao anotador, causando confusão.

```
<!ELEMENT schema (level+)>
<!ELEMENT level (value+)>
<!ATTLIST level
  id ID \#REQUIRED
  attribute CDATA \#REQUIRED
  read_only (yes|no) "no"
>
<!ELEMENT value EMPTY>
<!ATTLIST value
  id ID \#REQUIRED
  name CDATA \#REQUIRED
  default (yes | no) \#IMPLIED
  next CDATA \#IMPLIED
>
```

Figura 3. DTD do esquema de marcação.

O aprimoramento do MMAX consiste em descrever o esquema de marcação em XML de acordo com a DTD mostrada na Figura 3. Isso permite mostrar ao anotador apenas as opções coerentes com o que foi marcado nos passos anteriores, evitando erros. O atributo *next* restringe as opções que estarão disponíveis ao anotador de acordo com o valor associado no passo anterior ou no passo corrente. Além disso, os atributos marcados nos passos anteriores podem ser indicados como “read-only” nos próximos passos. A Figura 4 apresenta como exemplo o esquema de anotação para

o passo 3. Neste exemplo, o atributo *form* foi marcado no passo 1 e o atributo *coreference* foi marcado no passo 2. Se esse último atributo tiver o valor “coreferent”, então o anotador deve marcar o atributo *classification*.

Após dividir nossos experimentos em passos e utilizar os novos esquemas para o MMAX, a concordância entre as marcações dos anotadores aumentou, como esperado. Os resultados da anotação serão utilizados para validar os resultados de nossa ferramenta para resolução automática de correferência.

```
<annotationScheme>
  <level id="level_1" attribute="form">
    <value id="value_1" name="defNP" next="level_2"/>
    <value id="value_2" name="other_NP"/>
    <value id="value_3" name="part_of_sentence"/>
    <value id="value_4" name="sentence"/>
    <value id="value_5" name="other"/>
  </level>
  <level id="level_2" attribute="coreference">
    <value id="value_6" name="coreferent" next="level_3"/>
    <value id="value_7" name="non_coreferent"/>
    <value id="value_8" name="none"/>
  </level>
  <level id="level_3" attribute="classification">
    <value id="value_9" name="direct"/>
    <value id="value_10" name="indirect"/>
    <value id="value_11" name="other"/>
    <value id="value_12" name="none"/>
  </level>
</annotationScheme>
```

Figura 4. Exemplo de esquema.

#### 4 Ferramenta de resolução automática de correferência

Com base no estudo de corpus referido na seção anterior, estamos desenvolvendo uma ferramenta capaz de identificar automaticamente os antecedentes das expressões anafóricas.

Em (8), é apresentada uma ferramenta para resolução au-

tomática de correferência baseada nas heurísticas levantadas anteriormente. Esta ferramenta utilizava a linguagem PROLOG para armazenar os dados. Com base nesse trabalho anterior, estamos desenvolvendo uma nova ferramenta para atender as tendências atuais de anotação de corpora.

#### 4.1 Formato dos dados

Para representação dos dados de entrada e saída da ferramenta proposta, adotamos a linguagem XML e respeitamos os formatos usados no MMAX. Além disso, informações adicionais baseadas na análise sintática feita pelo analisador sintático PALAVRAS para o português (1), são utilizadas. A saída desse analisador é convertida em um conjunto de arquivos XML: o arquivo de palavras (elementos `<word>`), compatível com o MMAX; um arquivo com as categorias morfo-sintáticas (POS - Part of Speech) das palavras do corpus, como na Figura 5 (onde o elemento  $n$  indica um substantivo); e um arquivo com as estruturas sintáticas das sentenças, representadas por “chunks”. Um `<chunk>` representa a estrutura interna da sentença e pode conter sub-chunks, como mostrado na Figura 6, onde o `<chunk>` pai 1 (atributo `id="chunk_1"`) é um sintagma nominal (atributo `form="np"`) e o `<chunk>` filho 2 é um substantivo (atributo `form="n"`) e também o núcleo do sintagma nominal (atributo `function="h"`).

```
...
<word id="word_2">
  <n canon="acidente" gender="M" number="P"/>
</word>
...
```

Figura 5. Categorias morfo-sintáticas.

A quantidade de informação sintática utilizada pode variar de acordo com as heurísticas de resolução de anáforas utilizadas.

```
...  
<chunk id="chunk_1" function="subj" form="np" span="word_49..word_51">  
  <chunk id="chunk_2" function="h" form="n" span="word_51..word_51"/>  
</chunk>  
...
```

Figura 6. Chunks.

## 4.2 Heurísticas

O processo de resolução de anáforas é realizado por um conjunto de heurísticas desenvolvidas a partir de experimentos manuais de resolução. Esse processo pode ser decomposto de forma simplificada em dois passos: (i) a identificação das anáforas no texto e dos candidatos a antecedente; (ii) a identificação dos pares (anáfora, antecedente) corretos. Contudo, esses passos podem ser realizados de diferentes formas, e cada uma delas pode ser mais ou menos eficaz em função do texto. Por exemplo, a seleção de candidatos a antecedente pode ser realizada selecionando-se todos os sintagmas nominais do texto, ou apenas as descrições definidas.

Os exemplos deste artigo consideram como anáforas todas as descrições definidas do texto, como candidatos todos os sintagmas nominais, e para fins de ilustração do processo de resolução utilizamos a heurística mais simples de resolução de anáforas: anáforas diretas.

Segundo Poesio em (7), as anáforas diretas são aquelas cujo núcleo do termo anafórico é igual ao de seu antecedente. Tomemos como exemplo as sentenças “Três acidentes graves marcaram o final de semana. Os acidentes lotaram o hospital local”. O núcleo “acidentes” da descrição definida na segunda sentença é igual ao núcleo do sujeito da primeira sentença, o que caracteriza uma relação de anáfora direta.

## 4.3 Arquitetura

A ferramenta em desenvolvimento corresponde a um processo de 3 fases compostas por uma ou mais regras codificadas através de folhas de estilo (stylesheets) (Figure 7). Cada regra é conectada a

outra através de *pipes*, que filtram o fluxo de dados<sup>1</sup>.

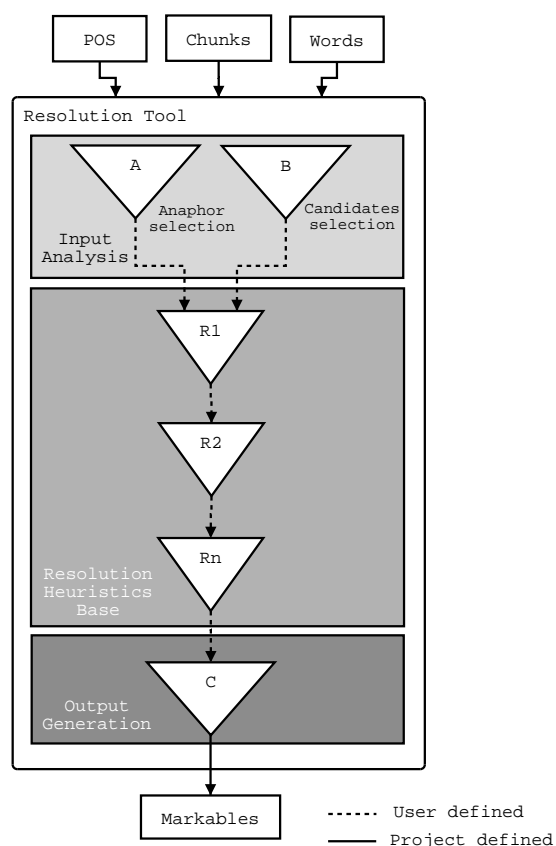


Figura 7. Arquitetura da ferramenta

A primeira fase, chamada *Input Analysis*, faz a extração de termos anafóricos e de candidatos com base nos arquivos de palavras, categorias morfo-sintáticas e chunks (Figuras 8, 9 e 10 respectivamente)<sup>2</sup>. A partir destas informações as regras A e B geram duas árvores de nodos que serão utilizadas na próxima fase.

<sup>1</sup>Estratégia baseada em Pipes & Filters Design Pattern, proposta por Gamma (2).

<sup>2</sup>Estes arquivos são baseados na frase exemplo da seção 4.2.

```

<words>
  <word id="word_1">Três</word>
  <word id="word_2">acidentes</word>
  <word id="word_3">graves</word>
  <word id="word_4">marcaram</word>
  <word id="word_5">o</word>
  <word id="word_6">final</word>
  <word id="word_7">de</word>
  <word id="word_8">semana</word>
  <word id="word_9">.</word>
  <word id="word_10">Os</word>
  <word id="word_11">acidentes</word>
  <word id="word_12">lotaram</word>
  <word id="word_13">o</word>
  <word id="word_14">hospital</word>
  <word id="word_15">local</word>
  <word id="word_16">.</word>
</words>

```

Figura 8. Words

```

...
  <word id="word_13">
    <art canon="o" gender="M" number="S">
      <secondary_art tag="artd"/>
    </art>
  </word>
  <word id="word_14">
    <n canon="hospital" gender="M" number="S">
      <secondary_n tag="inst"/>
    </n>
  </word>
  <word id="word_15">
    <adj canon="local" gender="M" number="S"/>
  </word>
</words>

```

Figura 9. POS

```

...
<sentence id="sentence_2" span="word_10..word_16">
  <chunk id="chunk_13" ext="sta" form="fcl" span="word_10..word_15">
    <chunk id="chunk_14" ext="subj" form="np" span="word_10..word_11">
      <chunk id="chunk_15" ext="n" form="art" span="word_10"/>
      <chunk id="chunk_16" ext="h" form="n" span="word_11"/>
    </chunk>
  </chunk>
</sentence>
...

```

Figura 10. Chunks

A árvore gerada pela regra A é formada por nodos <anaphor>, cujo o atributo *span* corresponde a uma descrição definida do texto (Figura 11). A árvore gerada pela regra B (Figura 12) é formada por nodos <candidate> que indicam através de seu *span* os sintagmas nominais do texto (nos experimentos preliminares utilizamos todos os sintagmas nominais como candidatos).

A próxima fase, chamada *Resolution Heuristics Base*, é composta por um conjunto de regras (R1 à Rn) que implementam as heurísticas e filtram os nodos <anaphor> em busca dos seus antecedentes. Cada uma implementa uma heurística diferente<sup>3</sup> e pode conter parâmetros que definem seu comportamento. Além disso, a ordem em que são aplicadas pode ser modificada, sendo assim, o usuário configura a ordem e os parâmetros da execução das heurísticas visando alcançar melhores resultados.

```

<anaphor-set>
  <anaphor span="word_5..word_8" pointer="">o final de semana </anaphor>
  <anaphor span="word_10..word_11" pointer="">Os acidentes </anaphor>
  <anaphor span="word_13..word_15" pointer="">o hospital local </anaphor>
</anaphor-set>

```

Figura 11. Anaphor

Considerando, por exemplo, que a regra R2 implementa a resolução de anáforas diretas, para cada nodo <anaphor>, busca-se um <candidate> com o mesmo núcleo, e *span* anterior ao do nodo

<sup>3</sup>Com a finalidade de enfatizarmos a arquitetura global da ferramenta, o conjunto completo de heurísticas não é apresentado neste trabalho.

```

<candidate-set>
  <candidate span="word_1..word_3">Três acidentes graves </candidate>
  <candidate span="word_5..word_8">o final de semana </candidate>
  <candidate span="word_10..word_11">Os acidentes </candidate>
  <candidate span="word_13..word_15">o hospital local </candidate>
</candidate-set>

```

Figura 12. Candidates

```

<markables>
  <markable id="markable_1" pointer="" span="word_5..word_8"
    classification="discourse_new"/>
  <markable id="markable_2" pointer="markable_4" span="word_10..word_11"
    classification="direct"/>
  <markable id="markable_3" pointer="" span="word_13..word_15"
    classification="discourse_new"/>
  <markable id="markable_4" pointer="" span="word_1..word_3"/>
</markables>

```

Figura 13. Markables

<anaphor> analisado. Para realizar esta tarefa, a regra R2 executa os seguintes passos:

1. Para cada nodo <anaphor> é selecionado o <chunk> correspondente com mesmo *span*;
2. Seleciona-se então o filho desse nodo que tem o atributo *ext* igual a “h” (indicando o núcleo do sintagma);
3. Então cada nodo <candidate> que antecede a anáfora é comparado com o nodo <anaphor>;
4. O núcleo do nodo <candidate> é comparado com do nodo <anaphor>:
  - (a) Se os dois núcleos são iguais: o antecedente foi encontrado, o atributo *pointer* é então preenchido com o *span* do candidato e o atributo *classification* com o valor correspondente ao da heurística em questão (neste caso *direct*).

- (b) Se os núcleos são diferentes: se retorna ao passo 3 até que os núcleos sejam iguais ou através de um critério de parada<sup>4</sup>.

No exemplo em questão a regra R2 encontrará um antecedente para o nodo <anaphor> (<anaphor span="word\_10..word\_11" pointer="word\_1..word\_3"> que corresponde ao nodo <candidate span="word\_1..word\_3">). Esta sequência de testes é realizada para cada nodo <anaphor> e a busca do antecedente se repete a cada regra.

A última fase, chamada *Output Generation*, gera os resultados em formato de saída específico para que possam ser analisados e/ou compartilhados. Neste caso apresentamos os resultados no formato do MMAX para então compará-los com a marcação manual. Para isso cada nodo <anaphor> gera um <markable> com *ID* único (por exemplo, *ID*="markable\_1"). O atributo *pointer* recebe o *ID* do markable correspondente ao *span* do atributo *pointer* do nodo <anaphor>. Se esse markable ainda não foi criado, um novo nodo <markable> é gerado utilizando este *span*. Nodos <anaphor> com *pointer* vazio geram nodos <markable> classificados como novos no discurso.

#### 4.4 Protótipo ART

Um protótipo da ferramenta (Figura 14), chamada ART (Anaphor Resolution Tool), foi construído com a finalidade de identificar as características básicas da interface da ferramenta.

A arquitetura apresentada na seção 4.3 possui características dinâmicas, como por exemplo a definição da ordem das regras e seus parâmetros, e o usuário da ferramenta deve ter acesso a elas por meio da interface. O protótipo implementa as seguintes características:

- Visualização e edição dos arquivos de entrada;

---

<sup>4</sup>O limite de parada é definido por uma janela (quantidade) de palavras, sentenças ou parágrafos por exemplo.

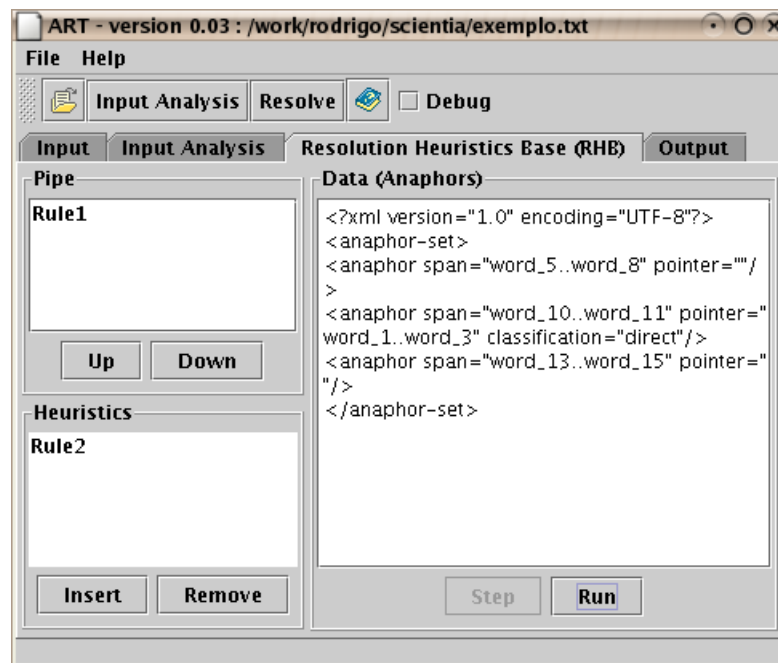


Figura 14. ART

- Interface de visualização e edição das três fases da arquitetura e suas regras;
- Construção dinâmica do *Pipe* de regras (inserção, remoção e definição da ordem das regras);
- Acesso aos parâmetros das regras (Folhas de estilo) e definição de valores para cada execução;

O objetivo é que a ferramenta se torne um ambiente para o desenvolvimento de novas heurísticas, suportando outras formas de manipulação do texto (como por exemplo a visualização gráfica dos elementos coreferentes).

## 5 Conclusão

Apresentamos a arquitetura de uma ferramenta para resolução automática de correferência. Essa ferramenta processa automaticamente corpus analisado sintaticamente e representado na linguagem XML, seguindo recomendações de padrões em desenvolvimento para anotação de corpora (XCES (3), ISO TC37 SC4). A vantagem de se ter os dados codificados em XML é a possibilidade de utilizar-se as ferramentas já existentes para manipulação dos dados nessa linguagem.

Para avaliar-se a precisão dos resultados da ferramenta, os dados de saída serão comparados com os dados obtidos na marcação manual realizada em estudos prévios do corpus (9; 11; 10).

Como esse trabalho se insere no contexto do projeto de cooperação internacional Brasil-França COMMON-REFs, a ferramenta foi concebida para tratar corpora multilíngüe, ou seja, que possa ser utilizada para a resolução de correferência em outras línguas além do português, visto que as heurísticas podem ser acopladas a ferramenta incrementalmente através de folhas de estilo independentes. Em estudos de corpus prévio sobre o francês (9; 11; 10), também foi realizada a anotação manual de correferência, o que poderá ser utilizado para avaliação da ferramenta em relação a esta língua, e de facilitar o reuso dos recursos linguísticos gerados.

## Agradecimentos

Agradecemos ao CNPq e INRIA, pelo financiamento do projeto COMMON-REFs; ao Eckhard Bick, por sua ajuda no uso do analisador PALAVRAS; ao Paulo Quaresma que desenvolveu o parser XML para o analisador PALAVRAS; ao Christoph Müller e Michael Strube, por proverem a base de nossos esquemas de anotação e acolherem nossas sugestões em relação ao MMAX; e a Susanne Salmon-Alt, pela cooperação no desenvolvimento dos estudos de corpora.

## Referências

- [1] Eckhard Bick. *The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. PhD thesis, Århus University, Århus, 2000.
- [2] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series. Addison-Wesley Publishing Company, New York, 1995.
- [3] Nancy Ide and Laurent Romary. Standards for language resources. In *Proceedings of the LREC 2002*, pages 839–844, Las Palmas de Gran Canaria, 2002.
- [4] Christoph Müller and Michael Strube. Annotating anaphoric and bridging expressions with MMAX. In *Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue*, pages 90–95, Aalborg, Denmark, 2001.
- [5] Christoph Müller and Michael Strube. MMAX: A tool for the annotation of multi-modal corpora. In *Proceedings of the IJCAI 2001*, pages 45–50, Seattle, 2001.
- [6] Christoph Müller and Michael Strube. Multi-level annotation in mmax. In *Proceedings of the SIGdial Workshop on Discourse and Dialogue*, Sapporo, 2003.
- [7] Massimo Poesio and Renata Vieira. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216, 1998.
- [8] Daniela Rossi, Clarissa Pinheiro, Nara Feier, and Renata Vieira. Resolução de correferência em textos da língua portuguesa. *Revista Eletrônica de Iniciação Científica*, 1(2), 2001.
- [9] Susanne Salmon-Alt and Renata Vieira. Nominal expressions in multilingual corpora: Definites and demonstratives. In *Proceedings of the LREC 2002*, Las Palmas de

Gran Canaria, 2002.

- [10] Renata Vieira, Susanne Salmon-Alt, Caroline Gasperin, Emmanuel Schang, and Gabriel Othero. Coreference and anaphoric relations of demonstrative noun phrases in multilingual corpus. In *Proceedings of the DAARC 2002*, Estoril, 2002.
- [11] Renata Vieira, Susanne Salmon-Alt, and Emmanuel Schang. Multilingual corpora annotation for processing definite descriptions. In *Proceedings of the PorTAL 2002*, Faro, 2002.

## Referências

- [1] Eckhard Bick. *The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. PhD thesis, Århus University, Århus, 2000.
- [2] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series. Addison-Wesley Publishing Company, New York, 1995.
- [3] Nancy Ide and Laurent Romary. Standards for language resources. In *Proceedings of the LREC 2002*, pages 839–844, Las Palmas de Gran Canaria, 2002.
- [4] Christoph Müller and Michael Strube. Annotating anaphoric and bridging expressions with MMAX. In *Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue*, pages 90–95, Aalborg, Denmark, 2001.
- [5] Christoph Müller and Michael Strube. MMAX: A tool for the annotation of multi-modal corpora. In *Proceedings of the IJCAI 2001*, pages 45–50, Seattle, 2001.
- [6] Christoph Müller and Michael Strube. Multi-level annotation in mmax. In *Proceedings of the SIGdial Workshop on*

*Discourse and Dialogue*, Sapporo, 2003.

- [7] Massimo Poesio and Renata Vieira. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216, 1998.
- [8] Daniela Rossi, Clarissa Pinheiro, Nara Feier, and Renata Vieira. Resolução de correferência em textos da língua portuguesa. *Revista Eletrônica de Iniciação Científica*, 1(2), 2001.
- [9] Susanne Salmon-Alt and Renata Vieira. Nominal expressions in multilingual corpora: Definites and demonstratives. In *Proceedings of the LREC 2002*, Las Palmas de Gran Canaria, 2002.
- [10] Renata Vieira, Susanne Salmon-Alt, Caroline Gasperin, Emmanuel Schang, and Gabriel Othero. Coreference and anaphoric relations of demonstrative noun phrases in multilingual corpus. In *Proceedings of the DAARC 2002*, Estoril, 2002.
- [11] Renata Vieira, Susanne Salmon-Alt, and Emmanuel Schang. Multilingual corpora annotation for processing definite descriptions. In *Proceedings of the PorTAL 2002*, Faro, 2002.