Identificação de Expressões Anafóricas e Não Anafóricas com Base na Estrutura do Sintagma

Sandra Collovini, Rodrigo Goulart, Renata Vieira

Programa Interdisciplinar de Pós-Graduação em Computação Aplicada – PIPCA

Universidade do Vale do Rio dos Sinos (UNISINOS)

Av. Unisinos, 950 – 93.022-000 – São Leopoldo, RS – Brasil

{sandrac, rodrigo, renata}@exatas.unisinos

Abstract. *One of the problems in anaphora resolution is to identify which expressions are anaphoric and which are non anaphoric. In this work a group of heuristic to identify the expressions as non anaphoric, the implementation of these heuristic in an environment for anaphora resolution (ART - Anaphor Resolution Tool) and an evaluation of the obtained results is presented.*

Resumo. *Um dos problemas da resolução de anáforas é identificar quais expressões são anafóricas e quais são não anafóricas. Neste trabalho um conjunto de heurísticas para identificar as expressões não anafóricas, a implementação destas heurísticas em um ambiente para a resolução de anáforas (ART - Anaphor Resolution Tool) e uma avaliação dos resultados obtidos são apresentados.*

1. Introdução

Este trabalho trata da resolução de expressões anafóricas em textos da Língua Portuguesa, mais especificamente descrições definidas. Chamamos descrições definidas os sintagmas nominais iniciados por artigo definido (o, a, os, as). Uma expressão é considerada anafórica quando se refere a uma entidade previamente referenciada no texto por meio de outra expressão. A resolução de anáforas consiste em encontrar um antecedente para os sintagmas nominais.

A identificação de expressões anafóricas, que se referem à mesma entidade, é importante em diversas aplicações de Processamento de Linguagem Natural, por exemplo, em sumarização automática, extração de informação, recuperação de informação, tradução automática, classificação de textos, entre outros.

Trabalha-se com descrições definidas, pelo fato de ocorrerem em grande quantidade nos textos do tipo de corpus estudado [Vieira et al., 2002]. Além disso, existem vários trabalhos sobre resolução de anáforas pronominais, mas não existem muitos trabalhos que tratem especialmente as relações de co-referência entre as descrições definidas.

Estudos recentes mostram que as descrições definidas além de ocorrerem em grande número, somente em 50% dos casos são consideradas expressões anafóricas. Por isso, consideramos importante o desenvolvimento de heurísticas para identificação de descrições definidas não anafóricas no processo de resolução dessas expressões.

Neste trabalho apresenta-se a implementação e a avaliação de heurísticas no ambiente ART (*Anaphor Resolution Tool*) [Gasperin et al., 2003a] para a tarefa de identificar descrições definidas não correferentes[1], ou seja, expressões não anafóricas, com base na estrutura sintática do sintagma nominal, de acordo com estudos prévios feitos para a Língua Inglesa [Vieira, 1998].

O trabalho encontra-se assim organizado: na seção 2, são apresentadas algumas considerações sobre resolução de anáforas e um estudo do sintagma nominal. Na seção 3, é mostrada a análise de corpus. Na seção 4, uma visão geral do Ambiente de Desenvolvimento ART é dada e as heurísticas para a classificação automática são apresentadas em detalhadamente. Por fim, na seção 5 são avaliados os resultados juntamente com as considerações finais.

2. Resolução de Anáforas

A tarefa de resolução de anáforas consiste na identificação de um antecedente textual importante na interpretação de uma expressão, como é ilustrado no exemplo a seguir: *"O Eurocenter oferece cursos de Japonês na bela cidade de Kanazawa, tanto para iniciantes quanto para aqueles com conhecimento avançado da língua. Os cursos têm quatro semanas de duração".*

Devido à complexidade da tarefa de encontrar um antecedente, somado ao fato de nem todas as expressões serem anafóricas (principalmente as descrições definidas), a comunidade vem propondo que parte do processo de resolução consiste em diferenciar sintagmas nominais entre anafóricos e não anafóricos [McCarthy and Lehnert, 1995; Bean and Riloff, 1999; Cardie and Wagstaff,1999; Vieira and Poesio, 2000; Soon et al., 2001; Muller et al., 2002; Ng and Cardie, 2002a; Ng and Cardie, 2002b; Uryupina, 2003]. Para isso, uma análise do sintagma nominal do português foi realizada para adaptar heurísticas do inglês na identificação de sintagmas nominais não anafóricos.

2.1. Estudo do Sintagma Nominal

Os sintagmas são formados por vários elementos que constituem uma unidade significativa dentro da sentença, além de manterem entre si relações de dependência e de ordem [Silva and Koch, 1989]. Estes elementos podem ser uma única palavra ou um conjunto de palavras. Os sintagmas desempenham uma função na sentença e combinam-se em torno de um núcleo. A classificação do sintagma é dependente do seu núcleo, por exemplo, quando o núcleo for um nome o sintagma é classificado como sintagma nominal. Conforme Perini (2003), o sintagma nominal possui uma estrutura bastante complexa, pois é possível distinguir dentro do sintagma nominal várias funções sintáticas. O núcleo do sintagma nominal pode ser um nome (comum ou próprio) ou um pronome (pessoal, demonstrativo, indefinido, interrogativo ou possessivo). O sintagma nominal pode também ser constituído por determinantes e/ou modificadores, sendo que os modificadores antecedem ou sucedem o núcleo, enquanto os determinantes apenas o antecedem [Miorelli, 2001].

Um sintagma nominal pode ser classificado como uma expressão anafórica ou não anafórica dependendo da sua relação de co-referência no discurso. As expressões

[1] Expressões correferentes são diferentes expressões invocando o mesmo referente.

são ditas anafóricas quando fazem referência a uma entidade introduzida no texto. As anáforas podem ser pronominais, definidas, indefinidas ou demonstrativas.

Um sintagma nominal não anafórico, introduz uma nova entidade no modelo discursivo. Geralmente ocorre no início do texto com descrições indefinidas, por exemplo, *"Uma instituição social"* ou com descrições definidas complexas, por exemplo, *"O quilômetro 430 da rodovia Assis Chateau Briand"*.

Nesse trabalho, o foco dos estudos são as descrições definidas. Estudam-se as descrições definidas segundo a classificação apresentada em Vieira (1998):

1. Anafóricas Diretas: são antecedidas por uma expressão que possui o mesmo nome-núcleo e refere-se à mesma entidade no discurso, por exemplo, "*Comprei um sapato. O sapato é confortável"*.
2. Anafóricas Indiretas: são antecedidas por uma expressão que não têm o mesmo nome-núcleo do seu antecedente. Assim, o núcleo pode ser um sinônimo do antecedente ou mesmo uma elipse, referindo-se à mesma entidade já introduzida no discurso, por exemplo, "*Comprei um apartamento. A moradia fica perto daqui"*.
3. Anafóricas Associativas: introduzem um referente novo no discurso, mas que tem uma relação semântica com algum antecedente já introduzido. Assim, a descrição definida tem seu significado ancorado em um referente, por exemplo, "*Ganhei uma rifa. O número sorteado foi o 100"*.
4. Não Anafóricas: são aquelas que introduzem um novo referente no texto que não se relaciona com nenhum antecedente no discurso. Assim, não possui uma âncora para se apoiar semanticamente, por exemplo, "*O radialista da Rádio Globo Washington Rodrigues"*.

3. Análise de Corpus

O corpus utilizado nesse estudo constitui-se de um extrato do corpus NILC[2], formado por 10 textos jornalísticos retirados da Folha de São Paulo, escritos em português do Brasil. Cada documento é um arquivo texto (formato ASCII) com tamanho entre 1 Kbyte e 6 Kbytes, com um mínimo de 41 termos e um máximo de 895 termos.

O corpus estudado foi anotado sintaticamente. Para obter a análise das sentenças do corpus, utilizou-se o analisador sintático PALAVRAS[3] descrito em Bick (2000), uma ferramenta robusta para a análise sintática do português. A partir da saída do analisador sintático a ferramenta Xtractor[4] descrita em Gasperin et al. (2003b) gera três arquivos XML. O primeiro arquivo é o *arquivo de Words,* Figura 1; o segundo é o arquivo com as categorias morfossintáticas (*POS – Part of Speech*), Figura 2; e por fim, o terceiro é o arquivo com as estruturas sintáticas das sentenças representadas por *chunks*. Um *chunk*

---

[2]Núcleo Interinstitucional de Lingüística Computacional. Disponível em http://www.nilc.icmp.usp.br/nilc

[3] O analisador PALAVRAS faz parte de um grupo de analisadores sintáticos do projeto VISL (Visual Interactive Syntax Learning), do Institute of Language and Communication da University of Southern Denmark Disponível em: http://visl.sdu.dk/visl/pt/parsing/automatic/

[4] A Ferramenta Xtractor engloba a análise do corpus a partir do analisador sintático PALAVRAS, o tratamento da saída desse analisador, com a geração de três arquivos XML.

pode possuir sub-elementos *chunks* com informações das sub-estruturas das sentenças, Figura 3.

```
<words>
.....
<word id="word_69">o</word>
<word id="word_70">radialista</word>
<word id="word_71">de</word>
<word id="word_72">a</word>
<word id="word_73">Rádio_Globo_Washington_Rodrigues</word>
.....
</words>
```

Figura 1. Arquivo de Words

```
<words>
.....
<word id="word_73">
<prop canon= "Rádio_Globo_Washington_Rodrigues"
 gender="M" number="S"/>
</word>
.....
</words>
```

Figura 2. Arquivo das Categorias Morfossintáticas

```
<text>
<paragraph id= "paragraph_1">
.....
<sentence id="sentence_7" span="word_69..word_96">
<chunk id="chunk_95" ext="sta" form="fcl"
       span="word_69..word_95">
  <chunk id="chunk_96" ext="subj" form="np"
         span="word_69..word_70">
  <chunk id="chunk_97" ext="n" form="art" span="word_69">
  </chunk>
.....
```

Figura 3. Arquivo de Chunks

Nesse estudo, os atributos dos *chunks* serão utilizados para a implementação das heurísticas no Ambiente ART (seção 4). As informações de interesse dos *chunks* são:

- Atributo *ext*: representa a função do *chunk,* por exemplo, sentença ou enunciado (*ext=sta*); sujeito (*ext=subj*); núcleo (*ext=h*).
- Atributo *form*: representa a forma do *chunk,* tais como: cláusula finita (*form=fcl)*; sintagma nominal (*form=np)*; substantivo (*form=n*).

Depois da anotação sintática automática, o corpus foi analisado manualmente em relação a co-referência. A anotação manual consiste em duas etapas. Em um primeiro momento, são anotadas as descrições definidas, considerando-se que uma

descrição definida pode conter outras descrições definidas, por exemplo, *"A lista do banqueiro do jogo do bicho", "o banqueiro do jogo do bicho", "o jogo do bicho"*. Em um segundo momento, as descrições definidas são classificadas como anafóricas e não anafóricas.

Para a anotação manual do corpus, utilizou-se a ferramenta MMAX (*Multi-Modal Annotation in XML*) [Müller and Strube, 2000], específica para anotação de corpus. Essa ferramenta utiliza o *arquivo de Words*, gerado pela ferramenta Xtractor que contém todas as palavras do corpus associadas a um identificador (atributos *id* da Figura 1). Ela também utiliza um segundo arquivo que contém a estrutura do corpus (parágrafos, sentenças, cabeçalhos, etc), ilustrado na Figura 4.

```
.....
<paragraph>
      <sentence id="sentence_1" span="word_1..word_8"/>
      <sentence id="sentence_2" span="word_9..word_23"/>
</paragraph>
.....
```

Figura 4 . Arquivo da Estrutura

O resultado do processo de anotação no MMAX é um arquivo que contém a anotação de co-referência. As marcações são codificadas como elementos *markable*, cujo atributo *span* indica as palavras que formam a expressão, o atributo *pointe*r indica o identificador do antecedente. Além destes, outros atributos podem ser especificados pelo pesquisador. Para esse estudo, acrescentou-se o atributo *classification* que corresponde à classificação anafórica da expressão (Figura 5).

```
.....
<markable>
      <markable id="markable_1"
      pointer=" "
      span="word_3..word_4"
      classification="non_anaphoric"/>
</markable>
.....
```

Figura 5. Arquivo de Marcações

4. Heurísticas para identificação de descrições definidas não anafóricas

ART é uma ferramenta para resolução de expressões anafóricas, onde o processo de resolução das anáforas é baseado em heurísticas. A ferramenta é desenvolvida em Java e os dados de entrada e saída utilizam a linguagem de marcação XML.

A arquitetura da ferramenta é baseada em "*pipes & filters*", constituindo-se de um conjunto de três passos (baseados na anotação manual) com uma ou mais tarefas codificadas através de folhas de estilo XSL[5] (*eXtensible Stylesheet Language*). As heurísticas utilizam informações dos textos analisados e são implementadas com folhas de estilos XSL.

[5] Linguagem Desenvolvida pelo W3C (world Wide Web Consortium) disponível em: http://www.w3.org/Style/XSL/

Nesse estudo, testamos algumas heurísticas para identificar as descrições definidas não anafóricas com base na estrutura do sintagma. Entre as heurísticas que serão apresentadas, a heurística 1, 2 e 3 foram elaboradas com base nos estudos da Língua Inglesa detalhado em Vieira (1998) e adaptadas para a Língua Portuguesa. Já as heurísticas 4, 5, e 6 foram construídas a partir da análise das características morfossintáticas das descrições definidas do corpus anotado estudado.

Heurística 1: expressão acompanhada de um sintagma preposicional, pós-modificador (restritivo), por exemplo, *"A tarde de ontem"*. Um pós-modificador restritivo sucede o núcleo restringindo-o. Um modificador restritivo permite que o referente seja identificado através da informação do modificador que especifica a informação do núcleo. Procura-se a existência de um sintagma preposicional no *chunk* da descrição definida, ou seja, um filho desse *chunk* com o atributo *form* igual a "*pp*". A Figura 6 ilustra o *span* "word_200..word_203" que corresponde a *"a tarde de ontem"*.

```
.....
<chunk id="chunk_277" ext="p" form="np"
       span= "word_200..word_203">
.....
  <chunk id="chunk_280" ext="n" form="pp"
         span="word_202..word_203">
.....
</chunk>
.....
```

Figura 6. Trecho do Arquivo de Chunks

Heurística 2: expressão constituída de construções de apostos, por exemplo, *"O prefeito de Gravataí, Daniel Luiz Bordignom"*. O aposto é um sintagma composto, com uma expressão adjacente que o explica ou especifica. O aposto pode vir separado por vírgulas ou depois de dois pontos. No corpus estudado, há construções de apostos como no exemplo acima em que o aposto *"Daniel Luiz Bordignom"* é uma explicação sobre "*o prefeito de Gravataí"*. Nessa heurística analisa-se a estrutura sintática do *chunk,* buscando-se por uma construção de aposto, ou seja, um filho com o atributo *ext* igual a "*app*". A Figura 7 ilustra o *span* "word_49..word_54" que corresponde a *"o prefeito de Gravataí, Daniel_Luiz_Bordignom"*.

```
.....
<chunk id="chunk_71" ext="subj" form="np"
       span=" word_49..word_54">
.....
  <chunk id="chunk_78" ext="app" form="prop" span=" word_54">
.....
</chunk>
.....
```

Figura 7. Trecho do Arquivo de Chunks

Heurística 3: expressão acompanhada de uma cláusula relativa, por exemplo, *"O texto que deve ser assinado pelos jornalistas"*. Nessa heurística, procura-se a existência de uma cláusula relativa, isto é, um irmão desse *chunk* que possua o atributo *form* igual a "*pron_indp*". A Figura 8 ilustra o *span* "word_100..word_108" que corresponde a *"o texto que deve ser assinado por os jornalistas"*.

```
.....
<chunk id="chunk_152" ext="subj" form="np"
       span="word_100..word_108">
.....
  <chunk id="chunk_161" ext="subj" form="pron_indp"
         span="word_105">
.....
</chunk>
```

Figura 8. Trecho do Arquivo de Chunks

Como neste trabalho os antecedentes não estão sendo considerados, apenas a estrutura do sintagma, adicionamos algumas restrições às heurísticas relacionadas a nomes próprios utilizadas anteriormente para o inglês (4 e 5).

Heurística 4: expressão com o núcleo sendo um nome próprio composto, por exemplo, "*A Rádio Globo Washington Rodrigues*". No corpus estudado, por tratar-se de textos jornalísticos, são relatadas informações sobre locais, eventos, pessoas, empresas importantes da atualidade, sendo que uma característica observada nesses textos é a presença de nomes próprios compostos, ou seja, nomes próprios formados por dois ou mais elementos que geralmente introduzem um novo referente no discurso. Para tratar desses casos, busca-se o núcleo dessa estrutura, ou seja, o filho desse *chunk* que possua o atributo *ext* igual a "*h*" e a forma de nome próprio, isto é, o atributo *form* igual a "*prop*". A Figura 9 ilustra o span "word_72..word_73" correspondente a *"a Rádio_Globo_Washington_Rodrigues"*.

```
.....
<chunk id="chunk_100" ext="p" form="np"
       span="word_72..word_73">
.....
  <chunk id="chunk_102" ext="h" form="prop" span="word_73">
</chunk>
.....
```

Figura 9. Trecho do Arquivo de Chunks

Heurística 5: expressão acompanhada de um nome próprio, por exemplo, *"O delegado Elson Campelo"*. No corpus estudado, uma característica observada nos textos é a construção de descrições definidas com núcleo sendo um nome comum (substantivo comum), seguido de um nome próprio especificando esse núcleo e geralmente tratando-se de um novo referente no discurso. Para resolver esses casos, analisa-se a estrutura do *chunk* , localizando o seu núcleo, ou seja, o filho desse nodo que possua o atributo *ext* igual a "*h*" e a forma de nome comum (substantivo comum), isto é, o atributo *form* igual a "*n*". Em seguida, verifica-se a presença de um nome próprio, isto é, um irmão desse *chunk* que possua o atributo *form* igual a "*prop*". A Figura 10 ilustra o *span* word_186..word_188 correspondente a *"o delegado Elson_Campelo"*.

```
.....
<chunk id="chunk_258" ext="acc" form="np"
       span="word_186..word_188">
.....
  <chunk id="chunk_260" ext="h" form="n" span="word_187">
  <chunk id="chunk_261" ext="n" form="prop" span="word_188">
</chunk>
.....
```

Figura 10. Trecho do Arquivo de Chunks

Identificamos na análise do corpus um outro tipo de pós-modificador restritivo freqüente, o sintagma adjetival.

Heurística 6: expressão acompanhada de sintagma adjetival, pós-modificador (restritivo), por exemplo, *"Os momentos mais difíceis de minha carreira"*. Um pós-modificador pode se configurar como um sintagma adjetival, que possui como núcleo um adjetivo. Para essa heurística, verifica-se a presença de um sintagma adjetival nessa estrutura, ou seja, o filho desse *chunk* que possua o atributo *form* igual a "*ap*". A Figura 11 ilustra o *span* "word_22..word_28" que corresponde a *"os momentos mais difíceis de minha carreira"*.

```
.....
<chunk id= "chunk_31" ext="p" form="np"
       span= "word_22..word_28">
.....
  <chunk id="chunk_34" ext="n" form="ap"
         span= "word_24..word_28">
.....
</chunk>
.....
```

Figura 11. Trecho do Arquivo de Chunks

De posse das heurísticas desenvolvidas, é possível automatização do processo de resolução de anáforas.

5. Avaliação

Na seção anterior foi apresentado um conjunto de heurísticas para identificar as descrições definidas não anafóricas e a implementação dessas heurísticas no Ambiente ART. Para analisar os resultados, é necessário comparar os resultados da aplicação das regras da ferramenta ART e os dados da anotação manual do corpus realizada no MMAX. O corpus analisado apresenta um total de 279 descrições definidas, sendo que 131 dessas expressões são classificadas como não anafóricas pela classificação manual, e 94 pela classificação automática, conforme Tabela 1. Para avaliar os ganhos obtidos com as heurísticas propostas, comparamos as medidas de abrangência e precisão das heurísticas com o *baseline* sendo um algoritmo que considera todas as expressões definidas como não anafóricas. A comparação é apresentada na Tabela 2. Com essas heurísticas obtemos 52,6% de abrangência e 73,4% de precisão, o que representa um ganho em relação à precisão obtida com o *baseline*. Considerando-se que apresentamos apenas cinco heurísticas de análise do sintagma pode-se dizer que a abrangência é bastante significativa. Durante o processo de análise dos resultados, erros na classificação foram observados, tais como: algumas descrições definidas sem

complementos (como *"as acusações"*) são não anafóricas, pois fazem parte do título do artigo ("*Citados negam as acusações*"), ou seja, estão na primeira sentença do texto; as descrições definidas constituídas por cláusulas relativas*"*, não estão sendo tratadas pela heurística 3, com por exemplo o pronome relativo "*onde*" em *"o hotel onde se hospeda, em Brasília*, isto se deve ao fato do analisador PALAVRAS considerar o pronome relativo "*onde*" como um advérbio.

Tabela 1. Classificação Manual e Automática

	Não anafóricas	**Anafóricas**	**Total**
Classificação manual	131	148	279
Classificação automática	94	185	279

Tabela 2. Abrangência e Precisão %

	Abrangência	**Precisão**
Baseline	100	46.9
ART + Heurísticas	52.6	73.4

Como trabalhos futuros, pretende-se aumentar o número de características para a identificação das descrições definidas não anafóricas. Essas novas características levariam em conta a posição das descrições definidas no texto, para tratar, por exemplo, as descrições definidas na primeira sentença. Também se consideraria as construções copulares, por exemplo: *"O maior representante do Eurocentres no Brasil é o Stbstudent Travel Bureau"*.

Com base nas heurísticas desenvolvidas, pretende-se além de aumentar o número de características para a identificação das descrições definidas não anafóricas, também utilizá-las em experimentos de resolução de anáforas com uma abordagem de Aprendizado de Máquina Supervisionado com árvores de decisão.

6. Bibliografia

Bean, D. L. and Riloff, E. (1999) "Corpus-based Identification of Non-Anaphoric Noun Phrases". In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, p. 373–380.

Bick, E. (2000) "The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework". PhD thesis, Arhus University, Arhus.

Cardie, C. and Wagstaff, K. (1999) "Noun phrase coreference as clustering". In: Proceedings of the 1999 SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, College Park, p. 82–89.

Gasperin, C.; Vieira, R.; Goulart, R.; Quaresma, P. (2003a) "Extrating XML Syntactic Chunks from Portuguese Corpora". In: Traitement Automatique Dês Langues Minoritaires- TALN, Btaz-sur-mer, France.

Gasperin, C., Goulart, R.; Vieira, R. (2003b) "Uma Ferramenta para Resolução Automática de Co-referência". Anais do Encontro Nacional de Inteligência Artificial (ENIA 2003), Campinas, SP.

McCarthy, J. F. and Lehnert, G. (1995) “Using decision trees for coreference resolution”. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, Canada, p. 1050–1055.

Miorelli, S. (2001) “Extração do Sintagma Nominal em Sentenças em Português”. Dissertação de Mestrado, PUC, Porto Alegre.

Müller, C. and Strube, M. (2000) “MMAX: A tool for the annotation of multi-modal corpora”. In: Proceedings of the IJCAI 2001, Seattle, p. 45–50.

Muller, C.; Stefan, R.; Strube, M. (2002) “Applying Co-training to reference resolution”. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL- 2002), Philadelphia, Penn., p. 352-359.

Ng, V. and Cardie, C. (2002a) “Improving machine learning approaches to coreference resolution”. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.

Ng, V. and Cardie, C. (2002b) “Identifying Anaphoric and Non-Anaphoric Noun Phrases to Improve Coreference Resolution”. In: Proceedings of the Nineteenth International Conference on Computational Linguistics (COLING-2002), p. 730–736.

Perini, M. (2003) Gramática descritiva do português. São Paulo: Editora Ática, 380 p.

Silva, M. and Koch, I. (1989). Lingüística Aplicada ao Português: Sintaxe. São Paulo: Editora Cortez, 160 p.

Soon, W. M.; Ng, H.wee T.; Lim, D. C. Y. (2001) “A machine learning approach to coreference resolution of noun phrases”. In: Computational Linguistics, p. 521–544.

Uryupina, O. (2003) “High-precision Identification of Discourse New and Unique Noun Phrases”. In: Proceedings of the ACL Student Workshop, Sapporo.

Vieira, R. (1998) “Definite description processing in unrestricted text”. PhD thesis, University of Edinburgh, Edinburgh.

Vieira, V. and Poesio, M. (2000) “An empirically-based system for processing definite descriptions”. In: Computational Linguistics.

Vieira, R.; Salmon-Alt, S.; Schang, E. (2002) “Multilingual corpora annotation for processing definite descriptions”. In: Proceedings of the PorTAL 2002, Faro.

Vieira, R.; Gasperin, C.; Goulart, R.; Salmon-Alt, S. (2003) “From concrete to virtual annotation mark-up language: the case of COMMOn-REFs”. In Proceedings of the (ACL 2003) Workshop on Linguistic Annotation: Getting the Model Right, Sapporo.