

Anáforas nominais definidas: balanceamento de corpus e classificação

Sandra Collovini, Renata Vieira

Programa Interdisciplinar de Pós-Graduação em Computação Aplicada
Universidade do Vale do Rio dos Sinos*
CP 275, CEP 93022-000, São Leopoldo, RS, Brazil
sandrac@exatas.unisinos.br, renatav@unisinos.br

Resumo Este trabalho apresenta a avaliação de um classificador de anaforicidade das descrições definidas que considera quatro classes distintas. Como a classificação é influenciada pela distribuição de exemplos entre as classes, e o número de exemplos é naturalmente não uniforme nos textos, experimentos com classes balanceadas foram realizados para melhor estudar as condições de aprendizado para o problema.

1 Introdução

A identificação do status do discurso tem sido reconhecida como uma tarefa relevante no processamento da língua natural. Muitos sistemas foram propostos para classificar expressões referenciais [1,2,3,4,5,6,7], a fim de reconhecer se elas representam informações velhas ou novas no discurso (anafóricas ou não anafóricas), juntamente com o problema da resolução de anáfora que busca encontrar um antecedente para as expressões que identificam uma informação dada. Na literatura, encontramos outras classificações mais detalhadas para as expressões referenciais que apresentam um número maior de classes, por exemplo, as 4 classes (novas no discurso ou não anafóricas; associativas; diretas e indiretas) apresentadas em [8]. Entretanto, a maioria dos sistemas não consideram uma classificação mais detalhada.

Neste trabalho, nós construímos e avaliamos um sistema para classificar descrições definidas (DDs) em 4 classes em textos da Língua Portuguesa. O problema de considerar 4 classes na classificação é que algumas classes possuem poucos exemplos. Por isso, nesse trabalho, realizamos experimentos onde as classes das DDs são balanceadas, para verificar se com um número equilibrado de exemplos é possível construir um classificador automático para as 4 classes.

O trabalho encontra-se assim organizado: na Seção 2, nós discutimos os trabalhos relacionados. Na Seção 3, as classes de DDs consideradas neste trabalho, o estudo de corpus e as características utilizadas para a construção do nosso classificador são apresentados. Na Seção 4, as classificações com base no corpus

* Este artigo foi realizado com apoio do CNPQ (Processos n° 310488/2005-2 e n° 381063/2005-4).

original e no corpus balanceado são apresentadas. Por fim, na Seção 5 apresentamos as considerações finais.

2 Trabalhos Relacionados

Na literatura, encontramos várias propostas de classificadores de expressões referenciais. Em [1] foi desenvolvido um sistema para a classificação de DDs em duas classes: DDs referenciais que possuem um referente anterior no texto, e DDs existenciais independentes que não possuem um referente anterior no texto e que são compreendidas isoladamente pelo leitor, sem a necessidade de um contexto. Um sistema de classificação de DDs baseado em heurísticas foi desenvolvido em [2]. Os autores estudaram os diferentes usos das DDs considerando três classes: anáforas diretas (possuem um referente anterior com mesmo nome-núcleo), novas no discurso (não possuem um antecedente) e associativas (possuem uma relação semântica com a expressão anterior), e desenvolveram três diferentes conjuntos de heurísticas para tratar esses casos. Em [5] um classificador para determinar a anaforicidade de sintagmas nominais (pronomes, DDs etc.) foi desenvolvido. Os autores classificaram as expressões em duas classes: anafóricas e não anafóricas, onde as anafóricas foram definidas como aquelas que compõem a cadeia de correferência, mas não são o seu primeiro elemento, e todos os demais elementos foram considerados como não anafóricos. Em [6] foi apresentado um sistema de classificação de DDs novas no discurso e únicas. A autora definiu as expressões novas no discurso como a primeira menção da entidade no discurso e expressões únicas quando especificam completamente o seu referente, e assim são interpretadas sem qualquer contexto. Primeiramente, foi realizada a distinção entre as DDs novas no discurso e velhas, após aplicou-se a classificação entre as DDs únicas e não únicas. Um grupo de características usuais em trabalhos anteriores para outros classificadores de DDs novas no discurso foi reexaminado e aplicado em [7].

Todos os trabalhos citados anteriormente se referem à Língua Inglesa. Algumas outras línguas foram também estudadas, mas não extensivamente [9,10,4]. Existem alguns estudos de corpora sobre correferência para a Língua Portuguesa [11], mas não implementam um sistema de resolução ou classificação de DDs para o Português.

Este trabalho estende nossos estudos para a construção de um classificador de DDs da Língua Portuguesa [12,13,14,15,16], sendo o primeiro a apresentar um classificador de DDs que abrange 4 classes. Na próxima seção, nós apresentamos as classes consideradas.

3 Corpus e Características para o Aprendizado

Inicialmente apresentamos a classificação das DDs consideradas nesse trabalho, baseadas principalmente em [17] e em alguns dos estudos discutidos na Seção 2. Nós classificamos as DDs em *Novas* (introduzem novas entidades no discurso) e *Velhas* (referem-se a entidades mencionadas previamente no discurso). As DDs *Novas* foram organizadas em dois tipos: *novas-no-discurso* (por exemplo, **A**

Folha de São Paulo) e *associativas* (por exemplo, *a operação contra o crime organizado ... o delegado encarregado*). As *associativas* introduzem uma nova entidade no discurso, mas possuem uma conexão semântica com uma expressão antecedente. Os dois tipos de DDs *Velhas* possuem uma relação de identidade com seus antecedentes com uma diferença: as *diretas* compartilham o mesmo nome-núcleo (por exemplo, *as listas apreendidas ... as listas*) e as *indiretas* apresentam diferentes nomes-núcleo (por exemplo, *A Folha de São Paulo ... O jornal*).

Os experimentos utilizaram um corpus formado por 24 artigos jornalísticos da Folha de São Paulo, escritos em Português do Brasil, correspondendo a uma parte do corpus do NILC¹. O corpus foi anotado automaticamente com informações sintáticas utilizando o analisador sintático do Português PALAVRAS² [18]. O corpus também foi anotado manualmente com correferência utilizando a ferramenta MMAX [19]. A primeira tarefa de anotação foi a distinção entre as DDs *Novas* e *Velhas*. A segunda tarefa foi a identificação dos antecedentes para as DDs *Velhas*. O corpus foi anotado por três anotadores. O consenso entre os anotadores para a primeira tarefa foi em torno de 90%. O corpus foi analisado posteriormente, dividindo as *Novas* e *Velhas* nas subclasses apresentadas anteriormente. Os valores utilizados nos experimentos são mostrados na Tabela 1. Uma grande quantidade de DDs *novas-no-discurso* foi confirmada (53%).

Classes	Corpus (%)		Sub-Classes	Corpus (%)	
Novas	644	62%	novas-no-discurso	550	53%
			associativas	94	9%
Velhas	401	38%	diretas	285	27%
			indiretas	116	11%
Total				1045	100%

Tabela 1. Subclasses das DDs Novas e Velhas

Com um estudo da literatura e uma análise de corpus, detalhados em [13], foi possível identificar um total de 16 características. Uma descrição e exemplos do corpus que ilustram cada uma das características são apresentados a seguir.

- SP: DD contém sintagma preposicional - *Os membros da classe jurídica.*
- APO: DD com construção de aposto - *O Prefeito de Gravataí, Daniel Luiz Bordignon.*
- APO_NP: DD contém nome próprio com aposto sem marca explícita - *O delegado Elson Campelo.*
- REL: DD contém cláusula relativa - *O texto que deve ser assinado pelos jornalistas.*
- NP.COM: DD com nome próprio composto - *O Othon Palace Hotel.*
- SA: DD contém sintagma adjetival - *As conversas mais antigas.*
- PRE_ADJ: DD com adjetivo anteposto ao núcleo - *O primeiro grau.*
- PRE_NUM: DD com numeral anteposto o núcleo - *Os 65 anos.*
- NUM: DD com numeral após o núcleo - *Os anos 60.*

¹ <http://www.nilc.icmp.usp.br/nilc>

² <http://visl.sdu.dk/visl/pt/parsing/automatic>

- PRON_DET: DD com outros determinantes além do artigo definido - *Os nossos arqueólogos.*
- PRE_SUP: DD contém superlativo precedendo o núcleo - *Os melhores alunos.*
- SUP: DD composta por superlativo - *O melhor.*
- TAM: DD formada por cinco ou mais termos - *O quilômetro 430 da rodovia Assis Chateau Briand.*
- COP: DD em uma construção copular - *O coreano seria a língua dos anjos.*
- S1: DD na primeira sentença do texto.
- SEM_ANT: O núcleo da DD é uma palavra que não ocorre anteriormente no texto, ou seja, não possui antecedente.

4 Classificação

Nesta seção nós apresentamos as classificações com base no corpus original e no corpus balanceado. O algoritmo de aprendizado utilizado foi o *j48* (reimplementação em Java do algoritmo de árvores de decisão *C4.5*) do pacote Weka³, com *10-fold cross-validation*⁴. As características apresentadas foram utilizadas na geração das árvores de decisão.

4.1 Classificação com o Corpus Original

O primeiro experimento de classificação com base no corpus original considerou as classes *Novas* e *Velhas*. Os resultados, na Tabela 2, são dados em termos de precisão (P), abrangência (A), F-measure (F) e taxa de acertos (C).

A segunda classificação (Tabela 3) utilizou as 4 classes - *novas-no-discurso*, *associativas*, *diretas* e *indiretas*. Os número de casos corretos baixou de 79% para 68%. Nenhuma DD foi classificada como *associativa* ou *indireta*, dado o reduzido número de exemplos existentes para essas classes (9% e 11%, respectivamente).

Uma outra classificação considerou 3 classes: *novas-no-discurso*, *diretas* e *outra*. Essa última incluindo as classes *associativas* e *indiretas*. Os resultados apresentados na Tabela 4 foram similares aos da classificação com 4 classes. Os exemplos da classe *outra* novamente não foram classificados.

Classes	P	A	F	C
Novas	79%	89%	84%	79%
Velhas	78%	63%	70%	

Tabela 2. Resultados da Classificação das classes Novas e Velhas

³ <http://www.cs.waikato.ac.nz/ml/weka>

⁴ Este método divide a base de dados em 10 partes, sendo 9 partes usadas para o treinamento e a parte remanescente para teste do algoritmo de classificação. Esse processo é repetido 10 vezes, cada vez considerando uma parte diferente para teste.

Classes	P	A	F	C
novas-no-discurso	67%	89%	77%	68%
associativas	0	0	0	
diretas	71%	76%	74%	
indiretas	0	0	0	

Tabela 3. Resultados da classificação com 4 classes

Classes	P	A	F	C
novas-no-discurso	67%	90%	77%	68%
diretas	71%	75%	73%	
outra	0	0	0	

Tabela 4. Resultados da classificação com 3 classes

O último experimento de classificação combinou as classes em pares, totalizando 6 classificações. Os resultados são apresentados na Tabela 5. Primeiramente foi realizada a combinação da classe *diretas* com as demais (*novas-no-discurso*, *associativas* e *indiretas*) - que apresentaram, respectivamente, 86% , 84% e 80% de acertos. Os resultados da classe *diretas* foram satisfatórios nas 3 classificações (F-measure média de 84%), a classe *novas-no-discurso* alcançou a melhor F-measure (89%), a classe *associativas* apresentou 75% de F-measure e a classe *indiretas* obteve 67% de F-measure. Os resultados podem estar relacionados a diferença de número de casos de cada classe (veja Tabela 1).

Classes	P	A	F	C
diretas	81%	77%	79%	86%
novas-no-discurso	88%	91%	89%	
diretas	95%	83%	89%	84%
associativas	64%	89%	75%	
diretas	89%	83%	86%	80%
indiretas	62%	73%	67%	
indiretas	0	0	0	83%
novas-no-discurso	83%	100%	91%	
indiretas	58%	45%	51%	54%
associativas	51%	63%	56%	
novas-no-discurso	85%	100%	92%	85%
associativas	0	0	0	

Tabela 5. Resultados da classificação combinando as classes

A classe *indiretas* foi combinada com as classes *novas-no-discurso* e *associativas*. A classificação de *indiretas* e *novas-no-discurso* não distinguiu os exemplos, sendo todos classificados como *novas-no-discurso*, porém a taxa de acertos geral alcançou 83%, dado que as *novas-no-discurso* somam 83% do total de exemplos das duas classes. Já na segunda classificação (classes *indiretas* e *associativas*) obteve-se, respectivamente, 51% e 56% de F-measure, indicando uma dificuldade de diferenciar as classes envolvidas (54% de acertos).

A última combinação considerou as classes *novas-no-discurso* e *associativas*, alcançando 85% de acertos, mas também não conseguiu distinguir os exemplos.

As características consideradas nas árvores resultantes são: SEM_ANT, NUM, PRE_SUP, APO_NP, S1, PRE_NUM, SA, PRON_DET (veja Tabela 6). A característica SEM_ANT mostrou-se a mais relevante para as classificações, estando no topo em todas as árvores resultantes. O principal fator de distinção é, simplesmente, a presença ou não de uma palavra idêntica no texto anterior à DD.

Classes	Características Relevantes
Novas e Velhas	SEM_ANT, NUM, PRE_SUP
4 classes	SEM_ANT, NUM, PRE_SUP, SA
3 classes	SEM_ANT, NUM, PRE_SUP, SA, S1
diretas e novas-no-discurso	SEM_ANT, NUM, PRE_SUP, SA
diretas e associativas	SEM_ANT
diretas e indiretas	SEM_ANT
indiretas e associativas	SEM_ANT, PRON_DET, PRE_NUM, APO_NP

Tabela 6. Características utilizadas nos experimentos de classificação

4.2 Classificação com o Corpus Balanceado

Devido à diferença de número de casos entre as classes (veja Tabela 1), novos experimentos foram realizados a partir de uma modificação nos exemplos que tornaram o corpus balanceado. Para construção do corpus balanceado, considerou-se o número de casos da classe *novas-no-discurso* (550 casos) como indicador do número de casos que cada classe deveria conter. Assim, os exemplos das demais classes (*associativas*, *diretas* e *indiretas*) foram duplicados até que cada classe completasse 550 casos. O corpus balanceado resultou em 2200 DDs. As classificações realizadas com o corpus original foram refeitas para o corpus balanceado.

A primeira classificação considerou as classes *Novas* e *Velhas* (Tabela 7), resultando em uma taxa de acertos de 71%, valor inferior ao do corpus original, a principal queda foi observada na precisão da classificação de *novas-no-discurso*.

A segunda classificação considerou as 4 classes (Tabela 8). Com o balanceamento das classes, os exemplos foram distribuídos entre as 4 classes, no entanto a taxa de acertos foi menor (51%). Observou-se uma maior dificuldade de distinção

das classes *novas-no-discurso* e *indiretas* que apresentaram, respectivamente, F-measure de 48% e 7%. A classe *diretas* apresentou os melhores resultados, valores similares aos encontrados utilizando o corpus original.

Classes	P	A	F	C
Novas	66%	88%	75%	71%
Velhas	82%	55%	66%	

Tabela 7. Resultados da Classificação das classes Novas e Velhas

Classes	P	A	F	C
novas-no-discurso	49%	47%	48%	51%
associativas	43%	77%	55%	
diretas	68%	78%	73%	
indiretas	46%	4%	7%	

Tabela 8. Resultados da classificação com 4 classes

Os resultados da classificação em 3 classes (Tabela 9) mostram valores mais altos em relação à classificação em 4 classes. A classe *diretas* segue apresentando bons resultados. Em geral, obteve-se 65% de acerto.

Por fim, as classificações considerando as classes em pares foram realizadas. Os resultados são apresentados na Tabela 10. A combinação da classe *diretas* com as demais, apresentou bons resultados - 85%, 87% e 79% de taxa de acertos, respectivamente.

Quando combinadas de forma balanceada, as classes *indiretas* e *novas-no-discurso* resultaram em uma distinção entre as classes, que não ocorria antes. Os resultados gerais diminuíram para 66% de exemplos corretamente classificados. Porém a classificação da classe *indiretas* passou de 0% de F-measure para 69%.

Na classificação entre as classes *indiretas* e *associativas*, observou-se 54% de taxa de acertos. Esta combinação permaneceu com a maior dificuldade em distinguir os exemplos.

A classificação entre as classes *novas-no-discurso* e *associativas* resultou em 69% de exemplos classificados corretamente. A classe *associativas* passou de 0% para 73% com o corpus balanceado.

As características presentes nas árvores são: SEM_ANT, NUM, APO_NP, S1, PRE_NUM, SA, NP_COM, SP, PRE_SUP, PRON_DET, COP, PRE_ADJ (veja Tabela 11). Em geral, o conjunto de características consideradas nas classificações balanceadas foi maior do que o apresentado na Tabela 6. A característica SEM_ANT aparece em todas as árvores resultantes, porém não necessariamente no topo. Tendo um conjunto de exemplos bem distribuído, as outras características se mostraram mais presentes no processo de classificação. Nos últimos 3 casos na Tabela 11, essa característica não foi a mais saliente (não ocorreu no topo da árvore).

Classes	P	A	F	C
novas-no-discurso	62%	53%	57%	65%
diretas	77%	80%	78%	
outra	55%	61%	58%	

Tabela 9. Resultados da classificação com 3 classes

Classes	P	A	F	C
diretas	88%	82%	85%	85%
novas-no-discurso	83%	88%	86%	
diretas	93%	80%	86%	87%
associativas	82%	94%	88%	
diretas	78%	82%	80%	79%
indiretas	81%	77%	79%	
indiretas	64%	74%	69%	66%
novas-no-discurso	69%	58%	63%	
indiretas	57%	41%	47%	54%
associativas	53%	68%	60%	
novas-no-discurso	75%	57%	65%	69%
associativas	66%	81%	73%	

Tabela 10. Resultados da classificação combinando as classes

Classes	Características Relevantes
Novas e Velhas	SEM_ANT, NUM, PRE_SUP
4 classes	SEM_ANT, PRON_DET, PRE_ADJ, APO_NP, TAM, NP_COM, SA, SP, PRE_NUM, NUM, PRE_SUP, COP, S1
3 classes	SEM_ANT, S1, SP, PRE_NUM, NP_COM, TAM, SA, PRE_ADJ, PRON_DET, NUM, PRE_SUP, COP, APO_NP
diretas e novas-no-discurso	SEM_ANT, NUM, PRE_SUP, COP
diretas e associativas	SEM_ANT, NUM, PRON_DET, APO_NP, NP_COM
diretas e indiretas	SEM_ANT, S1
indiretas e novas-no-discurso	PRE_SUP, S1, SEM_ANT, TAM, SA, PRE_ADJ, APO_NP, NP_COM, PRE_NUM, COP, SP
indiretas e associativas	PRE_NUM, APO_NP, SEM_ANT, SP, SA
associativas e novas-no-discurso	NP_COM, PRE_NUM, TAM, SA, APO_NP, NUM, PRE_ADJ, PRON_DET, SEM_ANT, S1

Tabela 11. Características utilizadas nas classificações

5 Considerações Finais

Este trabalho apresentou um sistema de classificação de DDs para a Língua Portuguesa, considerando 4 classes distintas. Com base em um estudo de corpus foram identificadas 16 características que serviram de entradas para o aprendizado do classificador. Devido a dificuldade da classificação de DDs em 4 classes, novos experimentos foram realizados com base em um corpus balanceado.

Como resultados, a classificação entre as DDs *Novas* e *Velhas* com base no corpus balanceado apresentou resultados mais baixos na taxa de acertos em relação ao corpus original (71% e 79%, respectivamente). A principal perda foi na precisão da classificação de *novas-no-discurso* (de 79% para 66% com o balanceamento). O maior número de *associativas* e *indiretas* teve influência nessa perda de precisão. A classificação considerando 4 classes com o corpus balanceado distribuiu melhor os exemplos entre as classes, mas apresentou a taxa de acertos inferior (51%) à do corpus original (68%). Observou-se a dificuldade de distinção na classe *indiretas* que apresentou 7% de F-measure. Os resultados da classificação em 3 classes com base no corpus balanceado mostrou que a classe *outra* passou de 0% de F-measure para 58%. Com o balanceamento, nós observamos que a combinação da classe *diretas* com as demais apresentou bons resultados, inclusive aumento na taxa de acertos entre as classes *diretas* e *associativas* (de 84% para 87%). Quando combinadas duas a duas e de forma balanceada, as classes *indiretas* e *novas-no-discurso* resultaram em uma distinção maior entre as classes, a classe *indiretas* passou de 0% de F-measure para 69%. Embora os resultados gerais tenham diminuído para 66% de taxa de acertos (83% com base no corpus original). A identificação de duas classes distintas ficou mais evidente. Na classificação entre as classes *indiretas* e *associativas*, manteve-se 54% de taxa de acertos. Esta combinação permaneceu como a que apresenta a maior dificuldade em distinguir os exemplos. A última classificação entre as classes *novas-no-discurso* e *associativas* apresentou perda na taxa de acertos com o balanceamento (de 85% com o corpus original para 69%). Porém, como anteriormente, a classe *associativas* passou de 0% para 73% com o corpus balanceado.

Os resultados apresentados indicam que o diferente número de exemplos entre as classes influenciam na classificação. Nos experimentos com corpus balanceado, as características estudadas serviram melhor para a identificação das classes. Nesses experimentos, um conjunto maior e mais diversificado de características foram utilizadas no aprendizado. Como trabalhos futuros, pretendemos pesquisar outras línguas românicas e integrar as tarefas de classificação com as de resolução de anáforas, isto é, determinar o antecedente das *diretas*, *indiretas* e *associativas*.

Referências

1. Bean, D.L., Riloff, E.: Corpus-based identification of non-anaphoric noun phrases. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, College Park, Maryland, USA (1999) 373–380

2. Vieira, R., Poesio, M.: An empirically-based system for processing definite descriptions. *Computational Linguistics* **26**(4) (2000) 525–579
3. Soon, W.M., Ng, H.T., Lim, D.C.Y.: A machine learning approach to coreference resolution of noun phrases. In: *Computational Linguistics*. Volume 27. (2001) 521–544
4. Müller, C., Rapp, S., Strube, M.: Applying co-training to reference resolution. In: *Proceedings of the 40th Annual Meeting of the ACL*, Philadelphia, PA (2002) 352–359
5. Ng, V., Cardie, C.: Identifying anaphoric and non-anaphoric noun phrases. In: *Proceedings of the Nineteenth International Conference on Computational Linguistics (COLING)*, Taipei, Taiwan (2002) 1–7
6. Uryupina, O.: High-precision identification of discourse new and unique noun phrases. In: *Proceedings of the 41st Annual Meeting on ACL*, Sapporo, Japan (2003) 80–86
7. Poesio, M., Alexandrov-Ksbadjov, M., Vieira, R., Goulart, R., Uryupina, O.: Does discourse-new detection help definite description resolution? In: *Proceedings of the 6th International Workshop on Computational Semantics*, Tiburg (2005) 236–246
8. Vieira, R.: *Definite Description Processing in Unrestricted Text*. PhD thesis, University of Edinburgh, Edinburgh (1998)
9. Aone, C., Bennett, S.: Evaluating automated and manual acquisition of anaphora resolution strategies. In: *Proceedings of the 33rd Annual Meeting of the ACL*, Cambridge, Massachusetts, USA (1995) 122–129
10. Guillena, R.M., Palomar, M., Ferrández, A.: Processing of spanish definite descriptions. In: *Proceedings of the Mexican International Conference on Artificial Intelligence*, Springer-Verlag (2000) 526–537
11. Salmon-Alt, S., Vieira, R.: Nominal expressions in multilingual corpora: Definites and demonstratives. In: *Proceedings of the LREC*, Las Palmas (2002) 1627–1634
12. Collovini, S., Goulart, R., Vieira, R.: Identificação de expressões anafóricas e não anafóricas com base na estrutura do sintagma. In: *II Workshop em Tecnologia da Informação e da Linguagem Humana - TIL*, Salvador (2004)
13. de Abreu, S.C.: *Análise de expressões referenciais em corpus anotado da língua portuguesa*. Master's thesis, Programa Interdisciplinar de Pós-Graduação em Computação Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo (2005)
14. Collovini, S., Coelho, J.C.B., Vieira, R.: Classificação automática de expressões anafóricas em textos da língua portuguesa. In: *V Encontro Nacional de Inteligência Artificial - ENIA*, São Leopoldo (2005)
15. Coelho, J.C.B., Collovini, S., Vieira, R.: Estudo de corpus para classificação de expressões anafóricas da língua portuguesa. In: *III Workshop em Tecnologia da Informação e da Linguagem Humana - TIL*, São Leopoldo (2005)
16. Collovini, S., Vieira, R.: Learning discourse new references in portuguese texts. In: *IFIP Conference on Artificial Intelligence - IFIP AI 2006*. IFIP World Computer Congress (WCC2006), Santiago, Chile (2006)
17. Prince, E.F.: Tooward taxonomy of given-new information. In: P. Cole, editor *Radical Gramatics*, New York, Academic Press (1981) 223–256
18. Bick, E.: *The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. PhD thesis, Arhus University, Arhus (2000)
19. Müller, C., Strube, M.: Mmax: A tool for the annotation of multi-modal corpora. In: *Proceedings of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Seattle, Washington (2001) 45–50