

# Instruções para anotação de relações anafóricas e referência dêitica

Versão 2.6 – 16 de novembro de 2006

## Laboratório de Engenharia da Linguagem – LEL

Universidade do Vale do Rio dos Sinos – UNISINOS  
Av. Unisinos, 950 – 93.022-000 – São Leopoldo, RS – Brasil

[www.inf.unisinos.br/~renata/laboratorio/labor.htm](http://www.inf.unisinos.br/~renata/laboratorio/labor.htm)

## Introdução

Este documento contém instruções para anotação de informações anafóricas e dêiticas, designadamente, elaboradas para o discurso escrito do português. O modelo de anotação, aqui, apresentado é baseado em estudos realizados pelos projetos ANACORT (Vieira, 2001), TeXto (Vieira, 2002) e VENEX (Poesio, 2005) e conta com o uso de uma ferramenta de anotação, a *MMAX* (Muller and Strube, 2000). Para efeito de síntese, este documento foi organizado sem discriminar os aspectos teóricos dos operacionais, sendo que as seções 1, 2, 3, 4, 5 e 6 correspondem a etapas de anotação.

## 1. Seleção das unidades de interesse (markables)

Nós estamos preocupados, aqui, em anotar, as relações anafóricas e as referências dêiticas realizadas por sintagmas nominais, que podem ter como núcleo um nome (exemplo 1) ou um pronome (exemplo 2). Locuções articuladoras, tais como, “nesse sentido”, “por essa razão”, “como resultado”, “além disso”, “isto é”, devem ser desconsideradas.

(1) *Quarta-feira, o Brasil inicia a nova campanha de turismo. Segundo o ministro, Olívio Dutra, isso impulsionará significativamente o setor. Ele, contudo, espera resultados a médio prazo.*

(2) *Ela estava responsável com alguém pelo esquema de viagens.*

É importante observar que sintagmas nominais podem empregar uma palavra de outra classe gramatical (adjetivos, verbos etc.) como nomes – por exemplo, “o pensar” em (3) e “o maior” em (4).

(3) *O pensar é uma atividade psíquica consciente e pode ser organizado ou não.*

(4) *Maria, ao comprar dois sorvetes, sabia que o maior era o meu.*

Além disso, sintagmas nominais podem apresentar outros sintagmas nominais internos – exemplo 5 (todos devem ser selecionados).

(5) *As regiões no interior do estado incluem diversos pontos turísticos.*

*As regiões no interior do estado*

*o interior do estado*

*o estado*

Outras observações:

- os apostos não são segmentados. Por exemplo, no sintagma nominal “o ministro, Olívio Dutra,” (exemplo 1), não deve ser marcado o sintagma nominal interno, em função apositiva, “*Olívio Dutra*”.

[o ministro, *Olívio Dutra*,]<sub>mark1</sub>

Da mesma forma, os apostos sem marca explícita também não devem ser segmentados, por exemplo, em “a química Nilva Ré-poppi”, não deve ser marcado o sintagma nominal interno, “*Nilva Ré-poppi*”.

[a química *Nilva Ré-poppi*]<sub>mark1</sub>

- Quando ocorrerem apostos formados por sintagmas nominais internos que representam novas entidades, como “a empresa” no exemplo, “Antonio Calmon, diretor da empresa”, deve-se anotar uma unidade completa:

[Antonio Calmon, diretor da empresa]<sub>mark1</sub>

e outra unidade representando o sintagma nominal interno, “a empresa”.

[Antonio Calmon, diretor d[*a empresa*]<sub>mark2</sub>]<sub>mark1</sub>

No exemplo, “A Motorola, gigante internacional do setor de telecomunicações,” também devem ser marcados os sintagmas nominais internos que representam outras entidades.

[A Motorola, gigante internacional d[*o setor de [telecomunicações]*]<sub>mark3</sub>]<sub>mark2</sub>]<sub>mark1</sub>

- Quando ocorrem sintagmas nominais seguidos de parênteses em função de aposto, como em “feronômios (substâncias produzidas pelos insetos para se comunicar ou servir de atrativo sexual)”, deve-se anotar uma unidade completa:

[feronômios (substâncias produzidas pelos insetos para se comunicar ou servir de atrativo sexual)]<sub>mark1</sub>

e outra(s) unidade(s) representando o(s) sintagma(s) interno(s) que representam outra(s) entidade(s):

[feronômios (substâncias produzidas pel[os insetos]<sub>mark2</sub> para se comunicar ou servir de [atrativo sexual]<sub>mark3</sub>)]<sub>mark1</sub>

**Observação:** os casos que apresentam URL não devem ser considerados apostos - não representam a mesma entidade que a expressão anterior - como em “a Energy Resource 2000 (www.energy.org)”, a URL “www.energy.org” não deve ser anotada – não é considerada um markable na anotação:

[a Energy Resource 2000]<sub>mark1</sub>

- listas, como, “mamíferos, anfíbios, répteis e aves”, formam uma unidade completa

[mamíferos, anfíbios, répteis e aves]<sub>mark1</sub>

e uma unidade para cada nome

[[mamíferos]<sub>mark2</sub>, [anfíbios]<sub>mark3</sub>, [répteis]<sub>mark4</sub> e [aves]<sub>mark5</sub> ]<sub>mark1</sub>

- Nomes: construções em que dois nomes estão coordenados, tais como “vinho e queijo”, “inverno **ou** verão”, formam três unidades, uma completa:

[vinho e queijo]<sub>mark1</sub>

[inverno ou verão]<sub>mark1</sub>

e outras duas para cada nome:

[[vinho]<sub>mark2</sub> e [queijo]<sub>mark3</sub> ]<sub>mark1</sub>

[[inverno]<sub>mark2</sub> ou [verão]<sub>mark3</sub> ]<sub>mark1</sub>

- Modificadores: um sintagma nominal formado por modificadores coordenados, como por exemplo, “*amarelas e vermelhas*” em: “bolsas *amarelas e vermelhas*”, não devem ser separados. Exemplos:

[aspecto *ecológico e social*]<sub>mark1</sub>

[a diversidade genética animal e vegetal]<sub>mark1</sub>

[plantas silvestres e medicinais]<sub>mark1</sub>

Os exemplos citados não precisam ser separados, mas devem ser separados no caso de uma posterior referência explícita a **uma de suas partes** no decorrer do texto (por exemplo, no caso de ocorrer uma referência ao aspecto ecológico, à diversidade vegetal ou plantas silvestres nos exemplos descritos).

- Cláusulas relativas restritivas: quando um sintagma nominal apresentar na sua estrutura uma cláusula relativa restritiva, o pronome relativo (que, onde, em que, na qual etc.) não deve ser considerado, como por exemplo, o pronome relativo “que” em: “a NASA está estudando *os mamíferos que se adaptaram ao clima polar*, a fim de desenvolver novas tecnologias espaciais”:

[os mamíferos que se adaptaram ao clima polar]<sub>mark1</sub>

[o clima polar]<sub>mark2</sub>

Outro exemplo:

[uma falha geológica ativa *que pode explicar aquele terremoto*]<sub>mark1</sub>

[aquele terremoto]<sub>mark2</sub>

- Cláusulas relativas não restritivas: esses casos devem ser desconsiderados. Por exemplo, no enunciado “Um dos autores do estudo, que está publicado na edição de hoje da revista britânica Nature, deixou claro o novo modelo implementado”, o primeiro argumento (Um dos autores do estudo, que está publicado na edição de hoje da revista britânica Nature,) deve ser anotado da seguinte forma:

[Um dos autores do estudo]<sub>mark1</sub>

[a edição de hoje da revista britânica Nature]<sub>mark2</sub>

[a revista britânica Nature]<sub>mark3</sub>

Outro exemplo de como devem ser anotadas cláusulas relativas não restritivas pode ser visto na marcação, a seguir, do sintagma nominal “a chamada falha normal, em que forças opostas levam blocos rompidos do solo a modificar de posição”:

[a chamada falha normal]<sub>mark1</sub>

[forças opostas]<sub>mark1</sub>

[blocos rompidos de o solo]<sub>mark3</sub>

[o solo]<sub>mark4</sub>

[posições]<sub>mark5</sub>

Neste exercício de anotação, os sintagmas nominais são nossas unidades de interesse, denominadas a partir de agora, também, de *markables*.

A princípio, os markables são gerados automaticamente e, posteriormente, revisados manualmente.

(COMENTAR SOBRE ERROS COMUNS DE IDENTIFICAÇÃO AUTOMÁTICA DE SNS).

## 2. Indicação de correferência (member) e anáfora (pointer)

Considera-se, freqüentemente, a referência (**referenciação**) como a propriedade da linguagem de evocar entidades, tais como, pessoas, animais, lugares, fatos etc. Essas instruções lingüísticas usadas para referir são denominadas **expressões referenciais**.

A organização dessas expressões referenciais é fundamental à continuidade e à estabilidade do texto. Basicamente, nesse processo de organização utilizamos termos/expressões que retomam outros termos do próprio texto, constituindo, assim, **cadeias correferenciais**. Esse fenômeno de retomada de referentes, é denominado **correferência**. No exemplo:

(6) *O Eurocenter oferece cursos de Japonês na bela de Kanazawa. Os cursos têm quatro semanas de duração. As aulas do nível avançado incluem refeições típicas e passeios a pontos turísticos da cidade.*

observamos que na sentença inicial são introduzidas três referentes ("O Eurocenter", "cursos de Japonês" e "a bela de Kanazawa"). Na sentença seguinte, a expressão "Os cursos" retoma "cursos de Japonês". Nessa perspectiva, "cursos de Japonês" é antecedente de "Os cursos", ou seja, duas expressões referenciais que fazem menção à mesma entidade, portanto duas **expressões correferenciais**.

Quando um markable apresentar essa relação, assegure-se de adicioná-lo ao set apropriado, selecionando um antecedente com o mecanismo **member** da MMAX (com exceção de encapsulation – ver abaixo).

No contexto da referenciação encontramos a noção de anáfora, a qual está intimamente relacionado ao conceito de correferência. Em sentido amplo, a anáfora se define como toda retomada de um elemento anterior em um texto, mantendo-se a identidade referencial, como, por exemplo: "cursos de japonês" – "os cursos", "a bela de Kanazawa" – "a cidade" (em 6). Já, em

sentido restrito, o fenômeno anafórico pode não ser correferencial, o referente de uma expressão anafórica pode não ser explicitamente denotado por um mesmo referente anterior. Em (6), observamos que a expressão “As aulas do nível avançado” não é correferente a nenhum termo anterior, entretanto, apresenta parte do seu significado ancorado na expressão “cursos de Japonês”, tanto que o sentido pleno dessa expressão (As aulas do nível avançado do curso de japonês) é alcançado pela soma da expressão anafórica (“As aulas do nível avançado”) a sua âncora referencial (“cursos de Japonês”). Assim, a anáfora pode ser um fenômeno semântico de natureza inferencial, e pode ser um fenômeno de correferência. Em síntese, uma expressão anafórica pode retomar uma referência anterior mantendo uma relação de identidade (i.e., anáfora correferencial), mas também pode ativar um novo referente cuja interpretação é dependente de outras expressões referenciais anteriormente presentes do texto (i.e., anáfora associativa).

Quando um markable apresentar essa relação, certifique-se de apontar a expressão que serve de âncora textual, usando o mecanismo **pointer** da MMAX.

### 3. Atributos dos markables

Os markables são formados por atributos que representam informações anafóricas e dêiticas de sintagmas nominais da Língua Portuguesa.

Nomes dos Atributos	Descrição	Forma de Anotação
Comment	usado para inserir comentários da anotação.	Manual
np_form	os valores deste atributo são os tipos de sintagmas nominais, os quais foram baseados na lista usada no GNOME.	automática (revisão manual)
pro_form	os valores deste atributo são os tipos de pronomes, os quais foram baseados na lista usada no GNOME.	automática (revisão manual)
member	Usado no MMAX para indicar cadeias de correferência constituídas por expressões que retomam a mesma entidade no discurso.	mecanismo de anotação do MMAX

pointer	usado no MMAX pra indicar uma referência associativa.	mecanismo de anotação do MMAX
status	representa as relações possíveis entre as entidades do discurso. Os valores deste atributo podem ser <i>new</i> , <i>old</i> , <i>associative</i> e <i>deictic</i> .	automática (revisão manual)
is_bridging	Quando "status=associative", o atributo is_bridging indica o tipo de relação associativa expressa por um <i>pointer</i> .	manual
is_anaphoric	Quando "status=old", o atributo is_anaphoric especifica o tipo de relação entre a entidade do discurso e o seu antecedente ( <i>direct</i> , <i>indirect</i> , <i>pronominal</i> e <i>encapsulation</i> ).	automática (revisão manual)

#### 4. Identificação de forma dos markables

Entendendo que as configurações morfosintáticas oferecem pistas valiosas para processamento da língua, nós definimos os atributos `np_form` e `pro_form`, em que o anotador distinguirá os sintagmas nominais com núcleo nome (**np-n** = yes) dos pronomes (**np-n** = no).

➤ **np\_form** (np-n = yes)

- Sintagmas nominais com núcleo nome:

- **def-np**: sintagma nominal formado por núcleo nome comum e determinante artigo definido (o, a, os, as), exemplo: "o setor" (em 1), "As regiões no interior do estado" (em 5).
- **def-pn**: sintagma nominal composto por núcleo nome próprio e determinante artigo definido, exemplo: "o *Brasil*" (em 1). Datas iniciadas por determinante artigo definido podem ser classificadas como **def-pn**, exemplo: "os anos 60", "o século 19".
- **indef-np**: sintagma nominal constituído por núcleo nome e artigo indefinido (um, umas, uns, umas) exemplo: "uma casa".

- **dem-np**: sintagma nominal com núcleo nome e determinante pronome demonstrativo (esse, essas, aquele, isso etc.), exemplo: “essa última disputa”, “aquele terremoto”.
- **poss-np**: sintagma nominal constituído por determinante pronome possessivo (seu, suas, nossa etc.) e núcleo nome, exemplo: “sua casa”, “nossa pesquisa”.
- **int-np**: sintagma nominal constituído por determinante pronome interrogativo (quando, quem, qual etc.) e núcleo nome, exemplo: “que horas termina a aula?”
- **num-np**: sintagma nominal iniciado por um numeral (um, primeiro etc.) e núcleo nome, exemplo: “95 empresas”, “10 mil trabalhadores”.
- **quant-np**: sintagma nominal constituído por núcleo nome e seguido de quantificadores (todos os, todas as, a maioria, ambos etc.). Os casos de determinante pronome indefinido (algum, outra, certa, várias, etc.) devem ser classificados como **quant-np**.
- **coord-np**: sintagmas nominais coordenados, exemplos: “ ceras, resinas e gomas”, “Maria e João”, “café ou chá”.
- **bare-np**: sintagma nominal sem determinantes<sup>1</sup> composto por núcleo nome comum, exemplo: “quarta-feira” (em 1), “viagens” (em 2).
- **pn**: sintagma nominal sem determinantes constituído por núcleo nome próprio representando nomes de pessoas, instituições, lugares geográficos e produtos, exemplo: “Maria” (em 4), “Porto Alegre”. Datas também podem ser classificadas como **pn**, exemplo: “1732”.

- **pro\_form (np-n = no)**

- Pronomes:

- **indef-pro**: sintagma nominal formado exclusivamente por um pronome indefinido, exemplo: “alguém” (em 2).
- **dem-pro**: sintagma nominal constituído somente por um pronome demonstrativo, exemplo: “isso” (em 1).

---

<sup>1</sup> Para nós, os numerais não são considerados determinantes.



- **pes-pro:** sintagma nominal formado exclusivamente por um pronome pessoal, exemplo: "Ele" (em 1), "Ela" (em 2).
- **poss-pro:** sintagma nominal formado exclusivamente por pronome possessivo, exemplo: "meu", "teu".
- **int-pro:** sintagma nominal formado exclusivamente por um pronome interrogativo, exemplo: "quando", "quem".
- **num-ana:** sintagma nominal formado exclusivamente por determinante numeral ou cardinal, exemplo: "eu quero um", "eu preciso de dois".

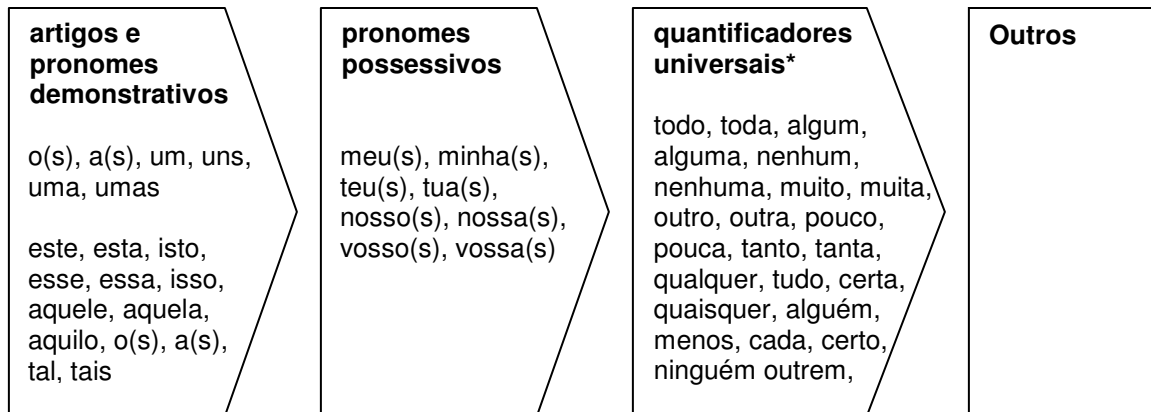
- Observações:

Nos casos de sintagmas nominais com determinantes complexos (exemplos 7 e 8) deve-se seguir a hierarquia para classificação da forma proposta neste manual. Por exemplo: em (7) e (8), as classificações adequadas são **def-np** e **dem-np** respectivamente.

(7) *Todos **os** alunos passaram na disciplina de Cálculo.*

(8) ***Esses** meus amigos participaram da festa.*

- Hierarquia para classificação da forma:



\*quantificadores universais: pronomes indefinidos

Além disso, ao contrário de algumas análises morfossintáticas que consideram os numerais como candidatos a ocupar a função *determinante*, nós limitamos os numerais à função

*modificador*. Por exemplo: “**as primeiras** experiências com inseticidas”, a classificação adequada é **def-np**.

## 5. Classificação das informações referenciais dos markables

Há ainda, no contexto da referenciação, outra possibilidade de referência: a informação requerida para interpretação de uma expressão não é encontrada no texto, mas na situação comunicativa. Quando uma expressão funciona assim é denominada **dêitica**. A referência dêitica não é coesiva, pois não liga uma expressão à outra dentro do texto. Imagine encontrar um bilhete, na rua, com a seguinte sentença:

(9) *Eu encontrarei você em trinta minutos.*

Nesse exemplo, é possível reconhecer cinco dêiticos nominais (“eu”, “você”, “trinta minutos”), uma vez que, os referentes adequados dessas expressões, somente, serão definidos de posse de informações (quem escreveu o bilhete, para quem, quando) provenientes da situação comunicativa – informações essas não presentes no bilhete (no texto), mas fora dele.

Resumidamente, pode-se dizer que a dêixis é a característica da linguagem humana que consiste em fazer um enunciado referir-se a uma situação definida, real ou imaginária, que pode ser:

- i) quanto aos participantes do ato de enunciação ou todo assunto da comunicação, exemplo: “nós” em (10);
- ii) quanto ao momento da enunciação, exemplo: “semana passada” (em 10);
- iii) quanto ao lugar onde ocorre a ação, estado ou processo, exemplo: “nesta edição” em (10).

(10) *Nesta edição, nós apresentamos o novo conversível XJ520 da Jaguar lançado na semana passada. Em primeira mão, explicaremos como funciona o motor de 620cv de potência a 7.600rpm.*

Em (10), as expressões “Nesta edição” e “nós” são ditas dêiticas, pois seus referentes só podem ser recuperados em relação à situação comunicativa. Com nosso conhecimento de mundo, nós sabemos que a primeira faz referência ao próprio objeto de leitura (a revista) e a segunda ao

corpo editorial. Já o referente da expressão dêitica “a semana passada” somente pode ser recuperado de posse da data da revista.

Como dito anteriormente, nós estamos interessados aqui com as relações anafóricas nominais. Um tipo importante de relacionamento anafórico que nós não estudaremos será a elipse, como em:

(11) *Meu livro não está aqui, [ele] sumiu.*

A par do exposto, num primeiro momento, o anotador indica o status dos markables:

➤ **status**

- **new:** o sintagma nominal introduz um novo referente no discurso sem apresentar parte de seu sentido ancorado em uma expressão anterior. Exemplo: “o novo conversível XJ520 da Jaguar” (em 10).
- **old:** o sintagma nominal é uma anáfora correferencial, ou seja, retoma um referente já introduzido por uma expressão anterior. Exemplo: “cursos de japonês” – “os cursos”, “a bela de Kanazawa” – “a cidade” (em 10).

Se um markable estiver ligado a outro como um **member** selecionar esse status (ou no caso de encapsulation – ver abaixo).

- **associative:** o sintagma nominal é uma anáfora associativa, ou seja, introduz um novo referente no discurso cujo significado está ancorado em uma expressão anterior. Exemplo: “cursos de japonês” – “as aulas do nível avançado” (em 8), “o novo conversível XJ520 da Jaguar” – “o motor de 620cv de potência a 7.600rpm” (em 10).

Se um markable estiver ligado a outro por um **pointer** marcar esse status.

- **deictic:** o sintagma nominal é uma referência dêitica. Exemplo: “Nesta edição” (10).

Atenção! Excetuando alguns poucos gêneros discursivos, como, por exemplo, o gênero narrativo, predominantemente, os pronomes pessoais de primeira (eu, me, mim, nós, conosco etc.) e segunda (tu, ti, vós, convosco etc.) pessoa funcionam como uma referência dêitica.

## 6. Classificação dos relacionamentos anafóricos correferenciais

Essa etapa de anotação é feita de modo automático após todas as etapas manuais. Agora, os casos, anteriormente, anafóricos correferenciais (status="old"), são classificados no atributo `is_anaphoric` em:

### ➤ **is\_anaphoric**

- **direct:** a expressão anafórica e seu antecedente apresentam núcleos idênticos. Exemplo: "cursos de japonês" – "os cursos" (em 8).
- **indirect:** a expressão anafórica e seu antecedente apresentam núcleos diferentes. Exemplo: "a bela de Kanazawa" – "a cidade" (em 8).
- **encapsulation:** a expressão anafórica (inclusive pronomes) retoma um trecho de texto maior que um sintagma, tais como sentenças ou mesmo parágrafos – por exemplo, em (12), "a operação" retoma "O Banco Central interveio ontem para segurar a cotação do dólar".

Nesse caso, não é necessário adicioná-lo a um **set**, entretanto, é preciso "comentar" qual trecho de texto foi retomado. Para isso, temos o atributo **comment** para adicionar esses comentários à anotação.

(12) *O Banco Central interveio ontem para segurar a cotação do dólar. A operação ocorreu quando a moeda havia alcançado R\$ 2,08.*

## 7. Classificação dos relacionamentos anafóricos associativos

Nesta etapa, é especificado, no atributo `is_bridging`, o tipo de relacionamento anafórico associativo. Assim, assegure-se de que os markables anafóricos associativos estão marcados com uma das relações seguintes:

### ➤ **is\_bridging**

- **element-of:** essa é a relação aplicada quando uma entidade do discurso evocada por uma expressão anafórica é um elemento de um grupo previamente introduzido – exemplos (13) e (14). Quando o elemento ocorre primeiro, deve-se usar a relação inversa: **element-of-inv** – exemplo (15). Atenção! Cada relação associativa tem uma relação inversa, entretanto, deve-se sempre apontar da expressão anafórica para seu antecedente, nunca de outra maneira!!!

- (13) *As turbinas do F-996 possuem sistemas distintos de segurança. A turbina traseira recebeu um sistema contra-incêndio.*
- (14) *Hoje, o Museu Nacional inaugura uma nova galeria expondo dois valiosos vasos chineses. Segundo o curador, o vaso maior, que mede 90 cm, está avaliado em mais de cinquenta milhões de dólares.*
- (15) *Por exemplo, o gás argônio é um elemento encontrado em diminuta proporção na atmosfera terrestre. Contudo, a partir de janeiro, a UNICAMP iniciará o rastreamento de todos os nobres incolores na estratosfera.*

- **subset-of:** essa relação é empregada sempre que uma expressão anafórica retoma um subconjunto de uma entidade introduzida anteriormente no texto. Essa relação também é utilizada para relacionamentos entre um tipo e seus subtipos. É possível perceber isso em (16) e (17).

- (16) *Os símios não precisam de muito espaço e se alimentam de quase tudo o que existe na floresta: folhas, brotos, frutinhas. O inverno, porém, é a estação de fartura para os bugios pela fartura de pinhões.*
- (17) *Para orientar as manobras das asas-deltas, o Pico-da-águia recebeu aparelhos do aeródromo Olavo Bilac. Perpendicularmente à extremidade sul, birutas foram instaladas.*

- **part-of:** essa relação deve ser marcada quando a expressão anafórica retoma uma parte de uma entidade já mencionado – exemplos (10) e (18). Quando a parte ocorre antes do todo, deve-se usar a relação inversa: **part-of-inv** – exemplo (19).

- (18) *O jato F-996 recebeu várias melhorias de segurança. As asas, agora, são equipadas com poderosos isolantes térmicos.*
- (19) *Uma nova torre chama a atenção dos cariocas. A Igreja Nossa Senhora Aparecida foi totalmente restaurada, sendo que a estrutura do sino foi aumentada dez metros.*

- **entity-attribute:** essa relação é empregada sempre quando a expressão anafórica retoma um atributo de uma entidade previamente mencionado – exemplos (20) e (21).

(20) *Devido aos novos padrões, as modelos estão se submetendo a severos programas de dieta. Conforme a agência Bellezza, o peso tem de ser menor do que é normalmente exigido.*

(21) *Devido ao titânio, o F-996 alcança 452,4 metros por segundo, o seu peso é quatro vezes menor do que um F-16.*

- **possessor-thing:** essa relação é marcada quando o antecedente possui a entidade evocada pela expressão anafórica associativa – exemplos (22) e (23). Frequentemente, nesse relacionamento, a expressão anafórica apresenta pronome possessivo.

(22) *Ontem, o ministro dos transportes foi convocado pela polícia para prestar esclarecimentos. Semana passada, foram encontradas, no seu apartamento, várias mercadorias contrabandeadas.*

(23) *O jogador Romário embarca essa mês para Madri. A sua viagem foi marcada para manhã de sexta-feira, após o Natal.*

- **other-bridging:** essa relação deve ser usada para outro tipo de relação não definido pelos anteriores. Por exemplo, os epítetos prestam-se particularmente a refletir a apreciação subjetiva do enunciador com relação ao discurso. Muitas vezes, as anáforas desse tipo constituem fatos de polifonia, quando a recategorização do referente é atribuída a uma outra voz que não a do enunciador, como se vê em:

(24) *Ao analisar os resultados do Sistema Nacional de Avaliação Básica do MEC, o ministro Paulo Renato Souza (Educação) afirmou que "a escola está cada vez menos interessante e motivadora, e o aluno cada vez mais dispersivo e indisciplinado". A pesquisa deste ano mostrou queda de aproveitamento nas escolas particulares. Para Paulo Renato, esse "efeito chatice" é provocado por duas razões centrais: a falta de reciclagem das escolas e a grande oferta de conhecimento fora da sala de aula, principalmente na Internet.*

Nem sempre é fácil distinguir um relacionamento part-of de um entity-attribute. Os testes lingüísticos baseados em construções lingüísticas gerais tais como "X de Y" ou "X tem Y" não são de muita ajuda: pode-se dizer que "casas têm janelas/portas/cômodos" e "casas tenha uma altura/largura/preço". Um critério razoavelmente útil é o status ontológico: as partes (part-of) tendem a ser objetos concretos, tais como portas, janelas, rodas, motores etc., e os atributos (entity-attribute) tendem a ser objetos mais abstratos, tais como altura/peso/velocidade. Os testes lingüísticos que usam verbos mais específicos podem também ser úteis: pode-se dizer dos atributos tais como a altura que a "altura é um atributo dos objetos", mas não se pode dizer que as partes dos objetos como: "as janelas são um atributo das casas". Inversamente, se pode dizer que as "janelas são partes das casas", mas não que "o comprimento é parte da casa". Também, o relacionamento possessor-thing não são aqueles entre um objeto e um de seus atributos ou suas partes. Se você não puder se decidir, marque estes casos como is\_bridging= "other-bridging" e use o atributo **comment**.

## **Síntese do procedimento de anotação do Projeto ProCaCoSA**

Nesta seção, sintetizamos o procedimento de anotação. Conforme dito anteriormente, as etapas 2, 3 e 6 são anotadas automaticamente, sendo que as demais são realizadas manualmente.

Cabe salientar que, o objetivo do projeto ProCaCoSA – Processamento de Cadeias de Correferência para a Sumarização Automática de Textos em Português – é analisar e diagnosticar alguns problemas causados à Sumarização Automática pela ocorrência de cadeias de correferência não resolvidas durante a seleção e estruturação do conteúdo de sumários. Dessa forma, o foco da anotação no âmbito do projeto são as cadeias correferenciais formadas por descrições definidas. Portanto serão anotadas apenas as informações anafóricas e dêiticas relacionadas a essas expressões (nas etapas 5, 6 e 7).

1. Leia o texto inteiro primeiramente, e certifique-se que o compreendeu. Tente fazer a marcação de um texto do começo ao fim, sem interrupções, pois, dessa maneira, será mais fácil de perceber o esquema coesivo do texto.
2. Selecione os sintagmas nominais, observando as observações da seção 1.
3. Identifique para cada markable uma dessas formas:
  - Sintagmas Nominais com núcleo nome - Figura 1:
    - a. **def-np**
    - b. **def-pn**

- c. **indef-np**
- d. **dem-np**
- e. **poss-np**
- f. **int-np**
- g. **num-np**
- h. **quant-np**
- i. **coord-np**
- j. **bare-np**
- k. **pn**

- Pronomes - Figura 2:

- l. **indef-pro**
- m. **dem-pro**
- n. **pes-pro**
- o. **poss-pro**
- p. **int-pro**
- q. **num-ana**

4. Indique se os markables são relacionados com outros por:

- a. **member:** selecione os markables correferentes usando o mecanismo member da MMAX – Figura 1.
- b. **Pointer:** apontar à expressão que serve de âncora, usando o mecanismo pointer da MMAX – Figura 2 e Figura 3.

Realize as etapas 5, 6 e 7 da anotação apenas para os markables que foram identificados como sendo da forma **def-np** ou **def-pn** (etapa 3).

5. Classifique os markables como:

- a. **new**
- b. **old** (com relação member) – Figura 1



caso um sintagma nominal retome trechos de textos maiores que um sintagma, classifique-o como old, comente qual trecho foi retomado e, no atributo **is\_anaphoric**, marque a relação **encapsulation**.

- c. **associative** (com relação pointer) – Figura 3
  - d. **deitic**
6. Classifique os relacionamentos dos markables anafóricos correferenciais (status= "old") – Figura 1:
- a. **direct**
  - b. **indirect**
  - c. **encapsulation**
7. Classifique os relacionamentos dos markables anafóricos associativos (status= "associative") – Figura 3:
- a. **element-of** (ou **element-of-inv**)
  - b. **subset-of**
  - c. **part-of** (ou **part-of-inv**)
  - d. **entity-attribute**
  - e. **possessor-thing**

a serpentes (markable\_4)

comment

member set\_1

pointer

np  yes  no

np\_form  indef-np  def-np  def-pn  dem-np  poss-np  int-np  num-np  quant-np  coord-np  bare-np  pn

status  new  old  associative  deictic

is\_anaphoric  direct  indirect  pronominal  encapsulation

Apply Undo changes

to front  suppress check  warn on extra attributes

**AutoApply is ON!**

Figura 1 – Esquema de anotação – Etapas 3, 4, 5 e 6

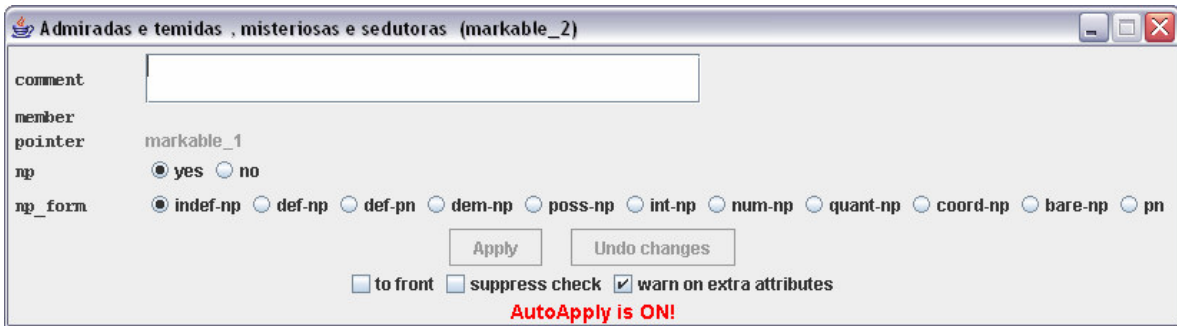


Figura 2 – Esquema de anotação – Etapas 3 e 4

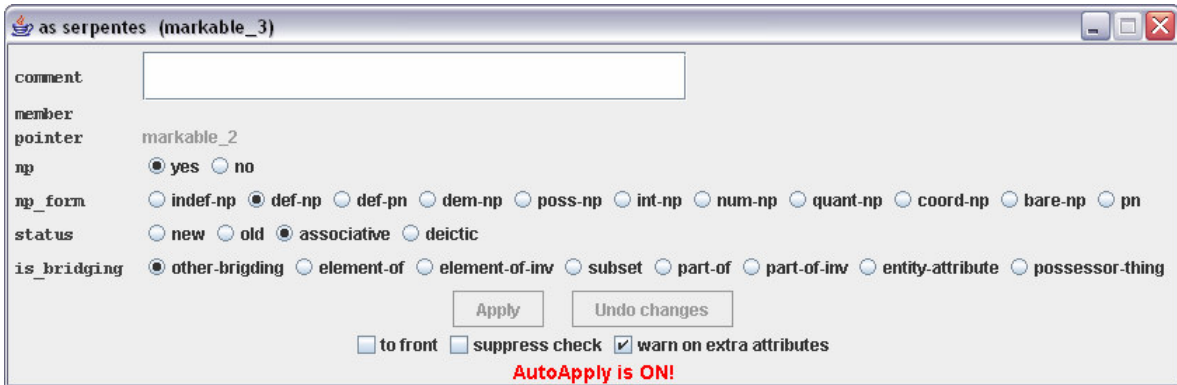


Figura 3 – Esquema de anotação – Etapas 3, 4, 5 e 7

## Apêndice

As referenciais e o esquema de anotação, bem como bibliografia para consulta estão disponíveis em: [www.inf.unisinos.br/~renata/laboratorio/labor.htm](http://www.inf.unisinos.br/~renata/laboratorio/labor.htm)

Incluimos também um arquivo exemplo em:

[www.inf.unisinos.br/~renata/laboratorio/corpora.htm/example\\_file.zip](http://www.inf.unisinos.br/~renata/laboratorio/corpora.htm/example_file.zip)