

## **Estudo dos argumentos verbais e ambiguidade dos sintagmas preposicionais através do Águia**

### **1. Introdução**

O estudo de algumas questões através do uso do sistema de busca do Águia para melhoramento da Floresta Sintá(c)tica (Afonso et al., 2001; 2002) e do próprio Águia (ref) foi desenvolvido. Ou seja, através do Águia, puderam detectar-se problemas de consistência nas árvores do Bosque correspondentes ao CETEMPúblico<sup>1</sup> (Rocha e Santos, 2000) já revistas e assim proceder às devidas alterações. Além disso, o uso do sistema de procura permitiu avaliar o funcionamento do mesmo.

As áreas investigadas foram as seguintes: argumentos dos verbos e ambiguidade dos sintagmas preposicionais. Com base nos resultados, as alterações necessárias ao Bosque foram efectivamente produzidas.

O presente relatório está organizado da seguinte forma: em primeiro lugar descreve-se o processo de estudo dos argumentos verbais, definindo-se cada um dos argumentos estudados segundo o sistema de anotação usado, apresentando-se a estratégia e os objectivos das buscas efectuadas e as conclusões relativas a esta parte. De seguida apresenta-se o processo de estudo da ambiguidade dos sintagmas preposicionais, reflectindo-se sobre a ambiguidade sintáctica e sintáctico-semântica e suas consequências. A estratégia para a busca no Águia será apresentada bem como os tipos de ambiguidade de sintagmas preposicionais encontrados. Seguem-se as conclusões relativas a esta parte. Finalmente, apresenta-se uma proposta de melhoramento das funcionalidades do Águia.

### **2. Estudo dos argumentos dos verbos: objecto directo, indirecto e regido de preposição**

O estudo dos argumentos dos verbos (objecto directo, indirecto e objecto regido de preposição teve como objectivo geral o melhoramento do Bosque, como foi referido na Introdução. No entanto, os objectivos mais específicos foram: 1) descrever a estrutura formal de cada um dos argumentos, 2) descrever os subtipos de objecto regido de preposição e 3) tentar encontrar critérios sintácticos que possam contextualizar o uso de cada um dos argumentos.

#### **2.1. Definições segundo o PALAVRAS**

Antes da descrição do processo de busca e das conclusões, é importante definir cada um dos argumentos segundo o parser PALAVRAS (Bick, 2000) que é usado para a anotação da Foresta.

Em termos terminológicos, os argumentos dos verbos objecto directo, indirecto e objecto regido de preposição têm associadas as seguintes etiquetas de função: ACC(usative) (objecto directo acusativo), DAT (ive) (objecto indirecto pronominal) e P(reposit)IV(e) (objecto preposicional), respectivamente. Relativamente à etiqueta DAT, e como a definição terminológica o sugere, esta etiqueta é usada apenas nos

---

<sup>1</sup> Embora o Bosque contenha árvores dos corpora CETEMPúblico e CETENFolha, o estudo das questões foi apenas produzido nas árvores do CETEMPúblico, porque este estudo fora iniciado antes das árvores do CETENFolha terem sido disponibilizadas.

casos em que um pronome tem função de objecto indirecto. Isto significa, conseqüentemente, que os objectos indirectos que tenham forma de sintagma (estrutura interna complexa, composta por mais do que um elemento), normalmente sintagmas preposicionais, não serão etiquetados por DAT, mas por PIV. Por exemplo: ‘a DeMille’ (em “Jesse Lasky, a conselho de Sam Goldfish (...) dá a primeira oportunidade a DeMille para fazer um filme”) é analisado como PIV, mas ‘lhes’ (em “ficou provado que os docentes não cumprem os novos programas porque não lhes foi dada a necessária reciclagem”) é analisado como DAT.

No entanto, PIV é atribuída também outro tipo de funções como é o caso de complementos verbais iniciados por preposição, a qual é ditada ou regida pelo verbo principal. Por exemplo, ‘falar de/sobre’, ‘aderir a’. Ou seja, PIV é uma etiqueta que abrange vários tipos diferentes de funções, o que pode criar o problema de um elevado grau de subespecificação, o que dependendo dos objectivos de investigação de cada um, pode ou não ser desejável.

## 2.2. Buscas no *Águia* dos objectos directos (ACC): estratégia

A estratégia de busca no *Águia* relativamente aos argumentos dos verbos teve como objectivo o seguinte. Relativamente a ACC, os objectivos iniciais de procura foram os seguintes:

- a) quais as possíveis formas que ACC pode exibir;
- b) qual a posição que ocupa em relação ao verbo.
- c) quais os contextos sintácticos que poderão determinar as relativas posições do argumento.

Relativamente a a), apenas se pôde pesquisar através do *Águia* as formas complexas (sintagmas) e não as formas atómicas, como pronomes, por exemplo (ver ponto 4.). Os resultados descrevem-se a seguir.

### 2.2.1. Resultados: objecto directo

Quanto às possíveis formas que ACC pode exibir, observou-se que ACC é bastante flexível, admitindo tanto formas sintagmáticas como oracionais. A seguinte tabela apresenta as frequências em matéria de forma dos ACC extraídas do *Águia* por concordância (e eliminando os casos que serão necessários corrigir):

Forma	np	ap	pp	advp	icl	fcl	acl
Frequência	2792	14	4	3	351	521	1

Tabela 1: Formas exibidas por ACC e número de ocorrências (através da obtenção de concordâncias)

Como se pode ver, o objecto directo é mais frequentemente um sintagma nominal (2792 ocorrências), seguido das formas oracionais icl (oração não finita) e fcl (oração finita). Os seguintes exemplos ilustram estes casos:

**np:** “Há , no ar , *uma certa ideia de invasão*”

**icl:** “A fatura de pensamento pode dar maus resultados e nós não queremos *ter um enfarte*” .

**fcl:** “*O resto é connosco* , garante a Lipor”.

Por outro lado, as formas raras associadas ao objecto directo (pp, advp, acl) relacionam-se com casos que são periféricos, i.e., objectos de verbos ilocutórios que são apenas sintagmas preposicionais, construções que parecem não ser prototípicas de funções como objecto directo:

**pp:** “*Para as poupar*”, justifica”.

“Os apostadores teriam de abrir uma conta e depositar aí *entre 1000 e 5000 dólares ( 185 a 925 contos )*”.

**advp:** “A polícia efectuou , na manhã de ontem , *nada menos do que oito detenções* no Bairro de S.Tomé,…”

**acl:** “Há , no seu modo de ser , *como que uma espécie de timidez...*”

Note-se que extraindo estes valores através da distribuição de funções que cada uma das formas presentes na tabela pode ter, os resultados são diferentes:

Forma	np	Ap	pp	advp	icl	fcl	acl
Frequência	3376	17	9	3	417	863	8

Tabela 1: Formas exibidas por ACC e número de ocorrências (através da distribuição de funções por cada uma das formas)

Além das formas presentes na tabela, há a considerar formas pronominais e nominais (nome comum ou próprio constituindo um nó sem dependentes) que não puderam ser extraídas através do Águia, por o sistema não reconhecer expressões de procura do tipo /fun\_n(‘ACC’) ou /fun\_prop(‘ACC’) e ainda /fun\_pron-pers(‘ACC’). Assim, valores importantes referentes por exemplo aos casos que envolvem o pronome reflexo ‘se’ não puderam ser estudados.

Além do estudo das formas que a função de objecto directo pode exibir, investigou-se também se há posições preferidas em relação ao verbo que estejam relacionadas com as respectivas formas, ou se haveria contextos específicos que proporcionariam determinadas posições, já que a hipótese seria de que a posição mais frequente do objecto directo seria posterior ao verbo, dado o português ser uma língua **S**(ujeito)**V**(erbo)**O**(bjecto).

Examinou-se então a função de objecto directo em posição pré- e pós-verbal e estudou-se para cada uma destas quais os contextos que indicariam essas posições. Os resultados revelaram-se interessantes. Em primeiro lugar, a posição pós-verbal, como se previa, é a mais frequente<sup>2</sup>. Pelo facto de o português ser uma língua SVO, como referido, e a evidência dos dados empíricos, a posição pré-verbal está

<sup>2</sup> Optei por não incluir valores porque não seriam reais uma vez que os valores diferem pelo facto de haver ou não espaçamento entre ‘ ’ e os argumentos e entre a expressão regular .\* e os argumentos. Concretizando, a expressão de busca /ass\_fun\_icl(‘ACC P’) produz resultados diferentes de /ass\_fun\_icl(‘ ACC P ’) e ainda as expressões de busca /ass\_fun\_icl(‘.\*ACC P’) e ass\_fun\_icl(‘ .\* ACC P’) produzem também resultados diferentes. Daí que os resultados apresentados serão apenas qualitativos.

relacionada com contextos alguns dos quais foram identificados, através das extracções do Águia.

Desta forma, em orações finitas e não finitas, os seguintes contextos sintácticos foram identificados como propiciadores de a posição pré-verbal do objecto directo. Estes contextos são os seguintes:

- **quotação com verbos ilocutórios em fim de oração** (Ex: “*E o dinheiro não falta só às câmaras, lembra o secretário de Estado*”);
- **oração relativa** (em que o pronome relativo que inicia a oração tem função de objecto directo) (Ex: “Os dois criminosos tinham libertado os três reféns *que detinham*”);
- **presença de conjunção** (Ex: “*nem o felicitei , nem fui carinhosa*”);
- **presença de adjunto adverbial** (Ex: “Facto que foi repetidamente denunciado, tanto pelos partidos da oposição (*onde se* destacaram o PCP e o PS)...”; “...diz- lhe ofegante , mas *logo se* recusa a acreditar na realidade”);
- **construção de foco** (Ex: “*Foi* então *que o* vi tirar a boina e umas barbas pretas...”);
- **presença de pronomes interrogativos** (Ex: “*Quem o* decidiu foi o juiz William\_Osteen ,...”).

Note-se ainda que os contextos acima estão também relacionados com a forma que o objecto directo exhibe. A presença de conjunção, adjunto adverbial, presença de pronomes interrogativos e construções de foco estão relacionados com a forma pronominal (pronome pessoal) do objecto. Enquanto, por exemplo, o contexto “quotação com verbos ilocutórios em fim de oração” relaciona-se com o objecto directo preposto ao verbo mas com forma oracional (finita ou não finita). Em orações relativas, a forma do objecto é, por outro lado, pronominal (pronome relativo).

Outros contextos ainda, especialmente construções específicas como a seguir se demonstrará, parecem influenciar a pré-posição do objecto directo face ao verbo. Quanto à forma, esta parece incluir sintagmas nominais, mas não formas pronominais (pelo menos as pessoais):

- “*Quanto mais* os EUA forem capazes de deixar claro que é fútil competir com o poder americano » , *menos chances* haverá de que « outros alimentem ideias de perturbar a actual ordem mundial”
- “O Carvalhal é a mais conhecida , mas *outras há* com menos gente e , por isso mesmo , mais escolhidas por os nudistas”.
- “*Uma reacção semelhante teve* o PCP , por intermédio de Vitor\_Dias...”

Como se pode ver pelos exemplos acima, a pré-posição do objecto face ao verbo (de que é argumento) ocorre em construções específicas (chamemos à construção acima construção de causa-efeito em grau: quanto mais/menos....mais/menos)<sup>3</sup>. Pode também ocorrer em contextos em que a escolha da posição do argumento tem que ver com a estrutura de informação e sua força pragmática (intencionalidade do

---

<sup>3</sup> Note-se que uma breve extracção destas construções a partir do CETEMPúblico permite constatar a fixidez da sua forma, a qual envolve a pré-posição dos argumentos ou adjuntos face ao verbo.

interlocutor/falante). Estes casos não foram estudados, uma vez que envolvem outros parâmetros de análise que não são do âmbito da Floresta em termos de anotação automática ou manual.

### 2.3. Buscas no Águia dos objectos indirectos (DAT): estratégia

Como foi referido, os objectos directos estão associados a duas etiquetas de função cujo uso é dependente da forma do objecto. Caso o objecto indirecto tenha forma pronominal, a etiqueta de função será DAT; se tiver forma de sintagma preposicional, a etiqueta de função será PIV.

Esta secção tratará apenas dos objectos indirectos pronominais. A secção seguinte tratará dos objectos indirectos preposicionais.

A estratégia seguida para o estudo dos objectos directos pronominais foi a seguinte:

- investigar a posição de DAT face ao verbo de que é argumento;

Não se investigou diferentes formas, em primeiro lugar porque o Águia não permite ainda a procura de formas atómicas e parece não reconhecer a função DAT (ver 4.). Por outro lado, pela própria definição da etiqueta DAT, esperar-se-ia que as formas encontradas fossem apenas pronomes pessoais. Desta forma, a busca de formas que não pronomes pessoais significaria apenas a identificação de problemas e sua consequente correcção.

A investigação das posições relativas de DAT face ao verbo é importante em termos de comparação com o argumento PIV (ver 2.4.) nomeadamente aquele com função semântica idêntica a DAT.

#### 2.3.1. Resultados: objecto indirecto

Identificaram-se alguns contextos em que a posição pré-verbal ocorre:

- **presença de conjunção subordinativa** (Ex: “Mas, *se me* pergunta se o parque de feiras de Lisboa será mais rentável que o do Porto, eu pessoalmente digolhe ...que sim”);
- **oração relativa** (Ex: “...o passeio marítimo , que me dizem ser de os mais apazíveis –“ );
- **presença de adjunto adverbial** (Ex: “Ele sabe que *não lhe* dei o meu voto”; “*Sempre me* pareceu estranho nunca ter lido um artigo sobre a «ilha de Santos»”).

A posição pós-verbal é ainda a mais frequente.

### 2.4. Buscas no Águia dos objectos preposicionais (PIV): estratégia

A etiqueta PIV é usada pelo PALAVRAS cobrindo diferentes funções sintácticas que têm em comum apenas a forma de sintagma preposicional. Há, no entanto, determinados casos em que a preposição é omitida sem que haja alterações à função, mas afectando a forma, como é exemplo o seguinte caso:

“...uma vez que conseguiram *convencer* muitos senadores antiaborto *que* defender a utilização de restos fetais não é a mesma coisa que defender a prática que lhes dá origem”

Convencer é um verbo que rege um argumento com preposição (convencer de que....). No entanto, no exemplo acima, a preposição está ausente. Em casos como este, o objecto do verbo, que normalmente seria preposicional, mantém a mesma função (PIV) mas a sua forma (com presença de preposição seria sintagma preposicional) passa a ter a forma consoante a sua “nova” estrutura interna. No exemplo acima, a forma de PIV é fcl (oração finita).

Assim, as funções que PIV abrange são:

- **objecto: recipiente** (objecto indirecto na terminologia da gramática tradicional) (Ex: “...referiu *ao Público* o presidente da Comissão Nacional do RMG , Paulo Pedrosa”);
- **objecto: origem** (Ex: “E que efeitos se podem *esperar da aplicação da pena de prisão àqueles que cometeram delitos pela exigência da toxicodependência* ?”)
- **objecto regido de preposição** (determinados verbos requerem um objecto com forma preposicional) (Ex: “Enquanto até aqui o HDL tinha em atenção o parecer dos pais no tratamento de crianças a *necessitar de sangue*,...”);

Ainda uma distinção importante entre PIV e ADVS e/ou ADVO (complemento adverbial reativo a sujeito/objecto, isto é, obrigatório ao contrário de ADVL, adjunto adverbial, não obrigatório). PIV e ADVS/O são categorias muito próximas, porque exibem a mesma forma (sintagma preposicional) e ambas pertencem à estrutura argumental do verbo. No entanto, ADVS/O refere-se a complementos na sua maior parte locativos ou temporais como é o caso de “chegar às redacções”, “expulsar da ordem”; “nascer e viver em Vila Real até aos 17 anos” que são obrigatórios com determinado tipo de verbos, especialmente verbos incoativos (que expressam mudança de estado).

Não havendo etiquetas para diferenciar estas subcategorias de objectos regidos de preposição, tentou investigar-se se haveria contextos sintácticos que pudessem ser sensíveis a cada um destes casos. Por exemplo, se PIV enquanto objecto regido de preposição estaria sempre em posição pós-verbal ou se haveria contextos que evidenciassem as subcategorias objecto/recipiente ou objecto/origem.

Tentou também investigar-se no caso de PIV ser objecto/recipiente se a preposição é fixa (a preposição ‘a’).

Finalmente, contrastou-se os campos semânticos semelhantes entre PIV (objecto/recipiente) e DAT em termos de comportamento sintáctico.

#### 2.4.1. Resultados: objecto regido de preposição

Relativamente à posição de PIV, objecto regido de preposição (tipo: necessitar de...), observou-se que em orações principais o argumento situa-se quase sempre em posição pós-verbal, salvo alguns casos/contextos que são os seguintes e raros:

- **oração relativa:** “de que se fala”, “a que se aderiu”;
- **construções particulares** relacionadas com estrutura de informação:
  - o “À sua poluição deve-se ainda o desaparecimento dos famosos bivalves negros da «Mexelhoeira»...”
  - o “Só da Serenata Monumental e do Sarau não se falou:”
  - o “Não é de literatura do que aqui se trata, embora a história, densa e trágica, nos envolva desde a primeira página.”
- **construções de foco** (Ex: “Foi com eles que o Público falou”)

Investigou-se qual a frequência das subcategorias de PIV em posição pré-verbal. O subtipo mais comum de PIV em posição pré-verbal é objecto regido de preposição seguido de objecto/recipiente. Note-se, no entanto, que os casos mais frequentes de objecto regido de preposição em posição pré-verbal está maioritariamente relacionado com contextos de oração relativa. Por outro lado, a ocorrência de objecto/recipiente e objecto/origem é pouco frequente neste mesmo contexto.

A posição pós-verbal é também nos subtipos objecto/recipiente e origem a posição mais frequente.

Quanto à fixidez da preposição no subtipo objecto/recipiente, todas as ocorrências exibem a preposição ‘a’.

Relativamente aos resultados comparativos face às posições relativas dos argumentos DAT e PIV subtipo objecto/recipiente, constatou-se que este subtipo de PIV é mais flexível em termos de posição do que DAT, isto é, a sua posição relativamente ao verbo de que é argumento não está somente dependente de determinados contextos sintácticos que obrigam à posição pré-verbal. Este facto poderá estar relacionado em parte com a forma que ambos os argumentos exibem. Assim, enquanto DAT parece ter exclusivamente forma pronominal, PIV tem por forma maioritariamente sintagmas preposicionais, o que parece ser por si só mais flexível em termos de posição. Veja-se por exemplo o caso de ADVL quando tem forma preposicional ou atómica (advérbio):

- “Essa poderá vir a ser uma hipótese , até porque , *no terreno* , a capacidade dos GAT está *cada\_vez\_mais* enfraquecida” .
- “Um líder *sempre* tem os seus fiéis.”

No primeiro exemplo, a alteração de posição do ADVL (‘em o terreno’) face ao verbo nem parece estar circunscrita a nenhum tipo de restrição nem parece alterar o significado global da proposição. Por outro lado, no segundo exemplo o advérbio ‘sempre’ com função de ADVL parece ter restrições sintácticas (\*Sempre um líder tem os seus fiéis).

## 2.5. Conclusões

Como conclusões aos objectos directos, aponta-se o facto de em termos de forma este argumento ser bastante flexível, abrangendo várias formas. A mais comum é por vários motivos (como características semânticas dos objectos directos entre outros que aqui não cabe discutir) o sintagma nominal. Infelizmente não se pode comparar a frequência de ocorrência de objectos directos com forma de sintagma nominal (isto é, um núcleo nominal e dependentes) com a frequência de ocorrência de objectos directos com forma pronominal ou nominal (isto é, um só nó, logo sem dependentes), que se suspeita ser também muito frequente.

Quanto à posição do argumento face ao seu verbo, observou-se que foram identificados contextos que promovem determinadas posições do argumento face ao seu verbo. A ordem mais frequente é a pós-verbal, por isso, estes contextos estão relacionados com a posição pré-verbal do argumento. Além disso, verificou-se haver uma relação entre as posições pré-verbais e as formas que o objecto pode exibir.

Sobre o funcionamento do *Águia*, de referir o facto da procura de elementos atómicos com função de objecto tanto por concordância como por distribuição de função não foi possível, mas seria uma funcionalidade importante, uma vez que, neste caso em particular, não se pôde extrair todas as formas pronominais com função de objecto directo, cuja frequência se supõe que seja significativa.

Relativamente aos objectos indirectos, a forma, por definição da etiqueta de função (DAT), supõe-se ser sempre pronominal (pronomes pessoais, forma dativa). A posição pós-verbal deste argumento é a mais frequente, mas a posição pré-verbal foi observada em determinados contextos que foram identificados.

Finalmente, quanto aos objectos regidos de preposição, constatou-se que a etiqueta usada pelo PALAVRAS, PIV, engloba diferentes subtipos de objectos preposicionais que apenas partilham forma idêntica. Uma vez que não há etiqueta que diferencie estes tipos (diferenciação esta de ordem semântica), retiraram-se algumas conclusões sobre o comportamento sintáctico destes subtipos e sobre o comportamento de um subtipo semelhante a objecto indirecto relativamente ao comportamento de DAT (objecto indirecto pronominal).

Estes dados parecem ser pois importantes para identificar quais os contextos em que determinados argumentos ocorrem ou não, que podem ser úteis, por exemplo, para mais facilmente serem colocadas manual ou automaticamente distinções mais finas correspondentes aos diversos subtipos de PIV.

### **3. Estudo da ambiguidade dos sintagmas preposicionais (pp)**

A outra área de estudo foi a ambiguidade dos sintagmas preposicionais. O estudo incidiu essencialmente na procura de tipos de ambiguidade e em especial no estudo da ambiguidade bem conhecida que é o sintagma ser um adjunto verbal (ADV<sub>L</sub>) ou um dependente nominal (N<sub><</sub>; >N).

#### **3.1. Ambiguidade sintáctico-semântica (real) vs. ambiguidade sintáctica**

Antes de descrever o tipo de buscas e os resultados, será talvez importante ter em conta o conceito de ambiguidade linguística. Ambiguidade sintáctico-semântica relaciona-se com o facto de duas ou mais alterantivas sintácticas serem possíveis num



mesmo contexto linguístico ou com o facto de um mesmo constituinte oracional poder ter o seu núcleo (verbal) em diferentes níveis de constituintes. O seguinte exemplo é um caso de ambiguidade sintáctica:

“A lei de imprensa, *na sua actual versão*, nestes como em muitos outros aspectos , fere gravemente o direito de ser informado dos portugueses ...”

A ambiguidade surge pelo facto de a parte em itálico poder ser tanto um dependente de o sintagma nominal anterior (“A lei da imprensa na sua actual versão”) como um adjunto adverbial do verbo ‘ferir’.

No entanto, há casos em que a ambiguidade é apenas sintáctica. A partir da análise dos resultados de comparação entre as anotações de dois revisores humanos e entre estes e a análise original do parser (Afonso 2004), concluiu-se que na maior parte dos casos em que a análise original do parser foi considerada ambígua, a ambiguidade poderia ser controlada por questões ou sintácticas (ex: presença de um pp de tempo ou lugar num sintagma nominal, função de dependente nominal) ou semânticas (pelo entendimento global da proposição, contexto linguístico ou por conhecimento extra-linguístico). Ou seja, não são muitos os casos de ambiguidade sintáctica que não possam ser desambiguizados intelectualmente.

### **3.2. Buscas no Águia de sintagmas preposicionais (pp)- estratégia**

O objectivo das buscas foi essencialmente investigar se existem contextos sintácticos que são mais propícios à existência ou não existência de ambiguidade (real, não desambiguável ou sintáctica e humanamente desambiguável). A partir das buscas tentou-se também identificar possíveis tipos de ambiguidade.

Desta forma, e partindo do conhecimento da ambiguidade entre função de adjunto (ADV<sub>L</sub>) e de dependente nominal (N<, pós-nominal e >N, pré-nominal) examinaram-se as ocorrências que contemplavam estas funções com forma de sintagma preposicional. Algumas hipóteses foram elaboradas, nomeadamente quanto à possibilidade de existirem contextos que possam ser mais ou menos susceptíveis de ocorrência de ambiguidade.

Outra questão que naturalmente decorreu da análise dos dados extraídos foi a possibilidade de existência de outro tipo de ambiguidades que não entre adjunto adverbial e dependente nominal.

### **3.3. Resultados: ambiguidade pps**

Relativamente aos contextos sintácticos que possam determinar se um pp é um dependente nominal ou um adjunto adverbial, observou-se em primeiro lugar que os dependentes de forma preposicional encontram-se na sua maioria em posição posterior ao núcleo nominal que modifica<sup>4</sup>. Quanto à mesmas posições mas

---

<sup>4</sup> Embora os valores por concordância e por distribuição de funções não coincidam, em ambos os tipos de pesquisa os valores de um pp dependente pré-nominal (31 e 39 ocorrências) são claramente inferiores aos valores de dependentes pós-nominais (9246 e 6239 ocorrências).

relativamente ao adjunto adverbial, não se pôde contabilizar<sup>5</sup> qual a posição mais frequente (pré- ou pós-verbal), mas sendo o adjunto adverbial um constituinte de relativa flexibilidade em termos de posição relativa ao verbo, é de crer que não se verificasse grandes diferenças em termos de ocorrência de adjuntos adverbiais antes ou após o verbo a que se refere. Assim, segundo os dados quantitativos parece mais provável que um sintagma preposicional que esteja antes de um constituinte de forma nominal seja um adjunto adverbial e não um dependente nominal (e que então faria parte desse sintagma nominal).

Um outro contexto sintático que parece, na maioria dos casos, determinar que um constituinte é adjunto adverbial é a ocorrência deste constituinte em início de oração ou quase início de oração<sup>6</sup>). Dificilmente, nestes casos, haverá ambiguidade entre as funções adjunto adverbial e dependente nominal, uma vez que, como foi observado, a maioria dos dependentes nominais é posterior ao núcleo que modifica. Só em casos específicos como descontinuidade do sintagma nominal ou em casos em que a proposição é um np, isto é, não verbal, é que um constituinte em início de oração será um dependente nominal:

- o Adjunto adverbial em início de oração:

*Atrás de Kerrigan* na lista das grandes favoritas está a japonesa Yuka\_Stao  
Foi este homem (...) que *em Janeiro* decidiu partir a loiça

- o Dependente nominal em início de oração (descontinuidade):

*Entre os concertos previstos para Novembro*, destaca-se o de dia 3, no  
Palácio Nacional de Queluz

Quando morre alguém , e depois nós ficamos sempre a pensar que *de essa pessoa* não sabíamos quase nada

Pelo contrário, há também contextos que, por outro lado, parecem beneficiar uma potencial análise sintática ambígua. Em primeiro lugar, se os constituintes pré constituintes de forma nominal indicam uma fraca propensão para a leitura de dependente nominal parecendo beneficiar a análise de adjunto adverbial, a posição dos constituintes pós constituinte com forma nominal já parece ser portadora de possíveis ambiguidades:

“Leonhardt tornou o «prélude non-mesuré» de Anglebert numa verdadeira dissertação filosófica, cujo final, *no seu despojamento elíptico*, surgiu como um ponto de interrogação metafísico e perturbante.”

Adicionalmente, colocou-se a hipótese de a adjacência ou não adjacência entre um constituinte ADVL:pp e P poder ter influência na análise de ADVL que poderia

---

<sup>5</sup> Porque sendo a pesquisa através do Águia feita a partir do formato de árvores, a etiqueta ADVL não inclui neste formato a sua posição face ao verbo, por contraste com o formato CG em que a informação sobre a posição do ADVL está contida nas setas (@ADVL>, verbo à direita, ou @<ADVL, verbo à esquerda).

<sup>6</sup> Por exemplo sequências como [Conjunção PP V].

revelar-se mais ou menos ambígua<sup>7</sup>. Além deste contexto investigou-se ainda a hipótese de as sequências (P SC ADVL), (SUBJ ADVL P), (P ACC ADVL)<sup>8</sup> poderem ser potencialmente ambíguas relativamente à análise de ADVL uma vez que SC, SUBJ, ACC e OC são mais frequentemente um sintagma nominal (703 ocorrências por distribuição de sintagmas).

Os resultados não se revelaram muito significativos devido em parte às falhas na extração de casos e extensiva filtragem manual dos casos relevantes. No entanto, algumas conclusões, ainda que não definitivas, poderão ser adiantadas. Assim, dos 25 casos de adjacência ADVL:pp P em que ADVL não está em início de oração (depois de manualmente filtrados), 12 casos são sintacticamente ambíguos (48%), ambiguidade essa que pode ser resolvida recorrendo-se à semântica e/ou ao conhecimento extra-linguístico, e 13 (52%) casos são realmente ambíguos. Relativamente à não adjacência, dos 18 casos extraídos e filtrados, o número de casos considerados ambíguos de ordem sintático-semântica e sintacticamente ambíguos é o mesmo- 9 ocorrências (50%). Desta forma, conclui-se que apenas a adjacência não é um critério que se relacione com maior ou menor grau de ambiguidade, dado a proximidade das percentagens em cada um dos contextos. Este facto pode ter também que ver com o facto de a não adjacência incluir advérbios de negação entre outros elementos que não são muito relevantes para o conceito de não-adjacência, isto é, aqueles que seriam relevantes seriam constituintes de forma nominal. No entanto, vejamos alguns casos de ambiguidade sintáctica e sintático-semântica independentemente de ADVL e P serem ou não adjacentes:

o Ambiguidade sintáctica:

“Ronaldo, *à beira do intervalo*, inaugurou o marcador para o Inter...”

“A toxicod dependência, *por o desespero que provoca*, é um terreno propício a todo o tipo de manipulações”

“Mas se tivermos em conta a nossa dimensão, ninguém *em tão pouco tempo* cobriu o país como nós”

o Ambiguidade sintático-semântica:

“O fundamental, *para mim*, é que a arte seja livre de constrangimentos.”

“Entretanto, *durante toda a manhã de ontem*, uma testemunha tentou, na esquadra da PSP, dizer o que tinha visto e ouvido”

“Antes, *ao fim da tarde*, os manequins e crianças passaram modelos da Mango”

Os casos acima classificados de ambiguidade sintáctica são casos em que a ambiguidade é apenas de ordem sintáctica, uma vez que os sintagmas preposicionais poderiam ter função de dependentes pós-nominais já que seguem um sintagma nominal o que pode ser gerador de ambiguidade sintáctica. No entanto, as potenciais ambiguidades são semanticamente resolucionáveis. O primeiro caso, porque o pp expressa o tempo em que o marcador foi inaugurado e não a posição no tempo de Ronaldo, o segundo caso, porque o pp expressa a causa da a toxicod dependência ser

---

<sup>7</sup> Houve vários problemas relacionados com a busca de constituintes adjacentes e não adjacentes que serão descritos em 4.

<sup>8</sup> Glossário: P-> predicado; SC-> predicativo do sujeito; SUBJ-> Sujeito

um terreno propício a manipulações e o terceiro caso, muito semelhante ao primeiro, porque o pp expressa o tempo em que o país foi coberto.

No entanto, ambiguidades de ordem sintáctico-semântica foram também observadas. As ambiguidades sintáctico-semânticas acima têm que ver com a possibilidade de o mesmo constituinte ter mais do que uma função sintáctica. No primeiro caso, o pp segue um sintagma nominal e, neste caso, ao contrário dos casos classificados como ambiguidade sintáctica que também seguem um sintagma nominal), a ambiguidade não é resolucionável através da semântica ([ (O fundamental para mim) é...] ou [(O fundamental)(para mim) (é)...] no sentido de “Para mim, o fundamental é...”). Enquanto este é um caso de ambiguidade adjunto adverbial/dependente nominal, os outros casos de ambiguidade sintáctico-semântica remetem para uma outra realidade, outros tipos de ambiguidade. Estes tipos serão tratados no ponto a seguir.

Em termos de sequências e para a ambiguidade adjunto adverbial/dependente nominal, a partir destes casos de ambiguidade sintáctico-semântica e sintáctica pôde-se concluir (ainda que muito superficialmente) que nas sequências com constituintes nominais seguidos de pp (SUBJ ADVL P) ou (P SC ADVL) o adjunto adverbial tem mais possibilidades de ter uma análise mais ambígua do que sequências que não impliquem a adjacência de um constituinte nominal e outro preposicional. Consoante os outros tipos de ambiguidade encontradas (descritas em 3.4.), as sequências mais propícias à ambiguidade serão outras.

### 3.4. Tipos de ambiguidade identificados

Como se referiu no ponto anterior, através das buscas por adjacência e não adjacência de ADVL:pp e P, encontrou-se o tipo de ambiguidade já esperado (adjunto adverbial/dependente nominal) e outros tipos de ambiguidade que são descritos a seguir.

Observou-se dois tipos de ambiguidade distintos da ambiguidade adjunto adverbial/dependente nominal:

- adjunto adverbial/dependente adverbial:  
“Antes, **ao fim da tarde**, os manequins e crianças passaram modelos da Mango”
- adjunto adverbial com núcleo em dois níveis de constituintes distintos:  
“Entretanto, **durante toda a manhã de ontem**, uma testemunha tentou, na esquadra da PSP, dizer o que tinha visto e ouvido”

No primeiro caso, o sintagma preposicional pode ser considerado como uma elaboração ou uma especificação do advérbio que o precede (“Antes”). Desta forma, a função seria de dependente adverbial e “Antes, ao fim da tarde,” tornar-se-ia em um só sintagma adverbial. Por outro lado, pode também ser apenas um adjunto adverbial do verbo “passar”.

O último tipo é uma ambiguidade diferente em termos estruturais, uma vez que não se trata de um problema de um mesmo constituinte poder ter funções diferentes, mas por outro lado, é uma ambiguidade de um mesmo constituinte em termos de dependência verbal, i.e, de que verbos nas diferentes orações que compõem a proposição é o adjunto dependente. No exemplo correspondente, o sintagma

preposicional tanto pode ser um adjunto adverbial do verbo “tentar” como do verbo “dizer” a um nível mais abaixo.

### **3.5. Conclusões**

Uma vez definidos os conceitos de ambiguidade sintáctica e de ambiguidade sintáctico-semântica e com base na observação de dados, os seguintes tipos de ambiguidade foram identificados: adjunto adverbial/dependente nominal; adjunto adverbial/dependente adverbial e adjunto adverbial com núcleos (verbais) em diferentes níveis de constituintes. Com certeza existirão outros tipos de ambiguidade envolvendo o sintagma preposicional, mas estes foram os tipos passíveis de identificação através de uma extracção conseguida nem sempre nas melhores condições. De igual forma, os contextos identificados que possam potencializar possíveis casos de ambiguidade são talvez limitados. De aí que estes resultados devam ser encarados como preliminares e não finais.

## **4. Proposta de melhoramento das funcionalidades do Águia**

Ao elaborar o estudo que aqui foi descrito, várias questões relativas ao funcionamento do sistema de busca Águia surgiram. As seguintes funcionalidades que não estão aparentemente disponíveis mas que seria importante implementar listam-se seguidamente, enquanto se ilustra a sua relevância.

- a) Procura de funções de elementos atómicos, como pronomes, nomes, etc, já que na Floresta não há nós com um constituinte, logo, é à própria forma que é atribuída função. Por exemplo, tornar possível a seguinte condição /fun\_n('ACC'). Neste estudo em particular, esta funcionalidade teria tido especial interesse porque as funções objecto directo e objecto directo pronominal (entre outras, como Sujeito) podem estar associadas a elementos atómicos como os acima referidos.
- b) Procura de funções em que a forma não esteja necessariamente especificada, por exemplo /fun\_x('SUBJ'). Tentou-se usar a expressão regular .\* no lugar da forma, mas sem sucesso. Com esta busca obter-se-ia de imediato, por concordância) todas as formas possíveis que o Sujeito (ou outra qualquer função) poderia exhibir, evitando que o utilizador, ao extrair as ocorrências com base nas formas, uma por uma, se esquecesse de investigar alguma forma em particular. Poder-se-ia argumentar que as formas que uma função pode exhibir pode ser obtida através da distribuição de funções por sintagmas. No entanto, apenas os valores quantitativos e não os contextos em que ocorrem seriam conhecidos/extraídos.
- c) Procura da distribuição de funções de um sintagma preposicional relativamente a um lexema (verbo, adjectivo, etc.). Por exemplo no caso “obrigar a”, extrair automaticamente as funções da preposição. Esta funcionalidade permitiria investigar todos estes casos de forma a aumentar a consistência dos mesmos. Além disso permitiria investigar quais os verbos ou adjectivos ou outros lexemas que são regidos de preposição.

- d) Procura de funções e formas combinadas na mesma funcionalidade, isto é, procurar sequências de funções com formas específicas. Por exemplo /ass\_fun\_fcl('P ADVL:pp). Esta funcionalidade teria sido de extrema importância para extrair os contextos exactos descritos em 3.3., principalmente extrair de imediato apenas os casos em que o ADVL nessas sequências tem apenas a forma preposicional. No caso presente, ao extrair /ass\_fun\_fcl('P ADVL), todas as formas que o ADVL exhibe foram extraídas e tiveram de ser filtradas manualmente de forma a se reter apenas os casos de adjuntos adverbiais de forma preposicional.
- e) Procura de constituintes que não em início de oração. Por exemplo, neste estudo, verifiquei que a procura da sequência ('.\*ADVL P') não elimina, ao contrário do que supunha) os casos em que ADVL surge em início de oração, o que conseqüentemente promove trabalho de filtragem manual.

Quanto a dificuldades encontradas no sistema, constatou-se o seguinte. Parece haver discrepância nos valores obtidos nas diferentes opções de extracção, essencialmente por concordância e por sintagmas. Além disso, as extracções por concordância, e em especial na parte de extracção de sequências de funções (/ass\_fun\_f('F1 F2')) parecem ser sensíveis ao número de espaços entre as funções e entre estas e ' ' que as envolvem, sem que se tenha percebido a razão de tal diferença. Finalmente, a investigação dos sintagmas associados à etiqueta de função DAT produz zero resultados (ao contrário das outras etiquetas de função), o que parece indicar a possibilidade de haver algum problema de configuração ou outros.

## 5. Referências

- Afonso, Susana (2004). *Avaliação do grau de concordância entre anotadores: análise e discussão dos resultados do processo de re-revisão*. Relatório interno.
- Afonso, Susana, Eckhard Bick, Renato Haber & Diana Santos. "Floresta sintá(c)tica: um treebank para o português", Anabela Gonçalves & Clara Nunes Correia (orgs.), *Actas do XVII Encontro da Associação Portuguesa de Linguística* (Lisboa, 2-4 de Outubro de 2001), Lisboa: APL, 2002, pp.533-45.
- Afonso, Susana, Eckhard Bick, Renato Haber & Diana Santos. "Floresta sintá(c)tica: a treebank for Portuguese", in Manuel González Rodríguez & Carmen Paz Suárez Araujo (eds.), *Proceedings of LREC 2002, the Third International Conference on Language Resources and Evaluation* (Las Palmas de Gran Canaria, Spain, 29-31 May 2002), ELRA, 2002, pp. 1698-1703.
- Bick, Eckhard (2000). *The Parsing System Palavras, Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*, Aarhus University Press.
- Rocha, Paulo Alexandre & Diana Santos. "CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa", in *Actas do V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR'2000)* (Atibaia, São Paulo, Brasil, 19 a 22 de Novembro de 2000), pp.131-140.