Susana Afonso
25 Março de 2004

# Avaliação do grau de concordância entre anotadores: análise e discussão dos resultados do processo de re-revisão

## 1. Introdução

Quando se produz um recurso a ser utilizado por diversos investigadores com diversos propósitos, como é, por exemplo, a criação de um treebank (um conjunto de árvores morfossintacticamente anotadas e humanamente revistas), é fundamental proceder à sua avaliação. É complexo definir sucintamente o conceito de avaliação (para uma extensa discussão sobre o conceito de avaliação, ver Crouch, Gaizauskas e Netter 1995). É, no entanto, relevante para o presente estudo referir que avaliar um recurso significa não só medir quantitiva e qualitativamente mas também melhorar o recurso com base nos resultados da avaliação (Hirshman 1998) através do (re)planeamento de estratégias (por exemplo, medir o progresso interno de um projecto (Santos 2000)).

O presente estudo representa a análise e discussão dos resultados obtidos pela contagem das diferenças entre dois revisores humanos e entre estes e o parser. O ficheiro utilizado para a contagem e posterior análise correspondeu a 229 extractos do corpus CETENFolha no formato de Constraint Grammar (CG) e a análise das diferenças incidiu sobre a informação morfossintáctica e outras como anotações secundárias.

O CETENFolha (**C**orpus de **E**xtractos de **T**extos **E**lectrónicos **N**ILC/**Folha** de S. Paulo), www.linguateca.pt/CETENFolha/, é um corpus de natureza jornalística de aproximadamente 24 milhões de palavras em português do Brasil criado pela Linguateca com base num subconjunto do corpus NILC/São Carlos compilado pelo Núcleo Interinstitucional de Lingüística Computacional (NILC), contendo exclusivamente textos da Folha de São Paulo de 1994. O CETENFolha foi anotado pelo parser PALAVRAS (Bick 2000).

A importância deste estudo prende-se com a necessidade de avaliar o grau de concordância de análise morfossintáctica dos revisores relativamente ao parser e o grau de concordância entre os revisores. Esta avaliação torna-se relevante num contexto de criação de um treebank para o português- Floresta Sintá(c)tica (Afonso at al. 2002)- que se deseja o mais consistente possível, isto é, que represente um recurso que possa ser usado pelos potenciais utilizadores de forma consistente, e desta forma, é desejável que o corpus seja mais homogéneo do que represente variações de análise de acordo com os diferentes revisores humanos, que parece ser inevitável (Kilgarriff 1998) mas, talvez, controlável. Relativamente à (não)concordância entre os anotadores e o parser é importante avaliar o grau de robustez da análise automática. Outro objectivo é também avaliar as opções de revisão: revisão paralela ou revisão consecutiva face ao número de diferenças observadas.

O presente estudo está estruturado da seguinte forma: em primeiro lugar apresenta-se os processo seguido para a análise e discussão dos resultados da contagem, remetendo para as categorias formadas para a avaliação. Segue-se a apresentação dos resultados e a discussão dos mesmos. Finalmente, apresentam-se as conclusões do estudo.

## 2. Contabilização e processo de análise dos resultados

A contabilização das diferenças foi feita automaticamente através de um programa criado para o efeito. O ficheiro onde as diferenças foram contabilizadas foi um único ficheiro em que as alterações ao parser pelo revisor 1 bem como as alterações pelo revisor 2 ao parser e ao revisor 1 foram marcadas.

Desta forma, foi feita a contagem das diferenças relativamente a:

- a) revisor 1 e parser (alterações introduzidas à análise original do parser pelo revisor 1);
- b) revisor 2 e parser (alterações introduzidas à análise original do parser pelo revisor 2);
- c) revisor 2 e revisor 1 (alterações introduzidas pelo revisor 2 às alterações pelo revisor 1 à análise original do parser)

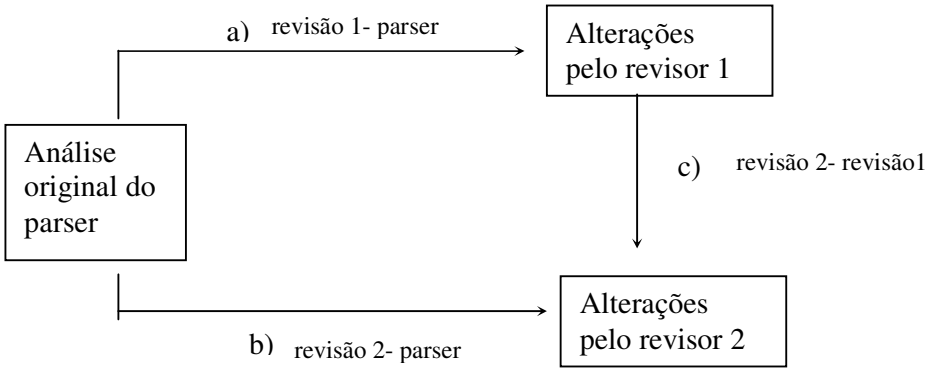


Diagrama 1: Dimensões de contagem de diferenças entre revisores e entre estes e o parser

As caixas no diagrama acima representam tão só as alterações e não os ficheiros correspondentes à análise original do parser e o modificado.

Uma vez que houve apenas um ficheiro com as alterações do revisor 1 e 2 marcadas, o revisor 2 teve acess tanto à análise original do parser como às alterações do revisor 1, estando a análise original do parser. O seguinte exemplo ilustra de que forma o revisor 2 teve acesso à análise original e às alterações do revisor 1:

**Palmeiras** **[Palmeiras] PROP M S @SUBJ>** **# [palmeira] N F P**
e [e] <co-subj> KC @CO
Corinthians [Corinthians] PROP M S @SUBJ>
treinam [treinar] <fmc> V PR 3P IND VFIN @FMV
hoje [hoje] ADV @<ADVL

As alterações à análise do parser estão marcadas por #. O revisor 2 saberia então que '[Palmeiras] PROP M S' foi o conjunto de alterações do revisor 1 ao parser cuja análise original tinha sido '[palmeira] N F P @SUBJ>'. Como a função sintáctica (@SUBJ>) não foi alterada, não está indicada por #.

Por outro lado, as alterações do revisor 2 ao parser estão marcadas por ! em caso de adição de elementos novos à análise do parser, por !R2 <alterado> em caso de alterações ao parser e #R1 em caso de alterações ao revisor 1:

...
deveriam [dever] <fmc> V COND 3P VFIN @FAUX
entrar [entrar] V INF @IMV @#ICL-AUX<
de [de] PRP @<ADVL
sola [sola] N F S @P<
em [em] <sam-> PRP @<ADVS **#R1 @<ADVL**; #@N<
esse [esse] <-sam> <dem> DET M S @>N
caso [caso] N M S @P<
e [e] <co-advs> KC @CO **!R2 <co-postnom>**

em [em] PRP @<ADVS **#R1 @<ADVL**; #@N<
todas [todo] <quant> DET F P @>N
as [o] <artd> DET F P @>N
outras [outro] <diff> <KOMP> DET F P @>N
federações [federação] N F P @P<
$.

Havelange [Havelange] PROP M S @SUBJ>
aceitaria [aceitar] <fmc> V COND 3S VFIN @FMV
engolir [engolir] V INF @IMV @#ICL-<ACC
o [o] <artd> DET M S @>N
italiano [italiano] N M S @<ACC
Antonio=Matarrese [Antonio=Matarrese] PROP M S @N<
como [como] PRP @PRD **!@#AS-<OC**
o [o] <artd> DET M S @>N
seu [seu] <poss 3S> DET M S @>N
vice-executivo [vice-executivo] N M S @AS<

Isto significa que as alterações pelo revisor 2 ao parser (representado em b) no digrama 1) são independentes de quaisquer alterações introduzidas à análise do parser pelo revisor 1.

## 3. Processo de análise e discussão

Para efeitos de análise dos resultados da contabilização das diferenças entre revisores e entre estes e o parser, formaram-se categorias e formularam-se hipóteses que a seguir se expõem. Algumas definições são também relevantes. Assim, por alterações introduzidas entende-se as alterações à análise original do parser bem como as adições à análise original. Por exemplo:

*alterações*: parser @SUBJ>
revisor @ACC>

*adição*: parser V PCP @IMV
revisor V PCP @IMV @#ICL-N<

De forma a investigar que tipo de alterações foram mais frequentemente introduzidas, as seguintes categorias foram formadas:

- Função sintáctica. A função sintáctica original, interna e/ou externa, fornecida pelo parser é alterada ;
- PoS (Part of Speech), correspondente a alterações a nível da classe de palavras;
- Morfologia, correspondente a alterações exclusivamente (isto é, sem combinação de alterações (ver em baixo)) a nível de género e número, pessoa, número, tempo verbal e modo;
- Alteração global, correspondente a alterações que envolvem todos os níveis gramaticais, por exemplo, PoS, morfologia e função, ou alterações de atomização, como os seguintes exemplos ilustram:

  a)
  ...
  deveria [dever] V COND 3S VFIN @FAUX
  significar [significar] V INF @IMV @#ICL-AUX<
  muito [muito] <quant> ADV @>A
  mais [mais] <quant> <KOMP> ADV @<ACC

```
do=que  [do=que] KS @COM @#AS-KOMP<
isso    [isso] <dem> SPEC M S @AS<
$,
mais    [mais] <quant> <KOMP> ADV @<ACC
do=que  [do=que] KS @COM @#AS-KOMP<
o       [o] <artd> DET N S @>N                          #[ele] PERS M 3S ACC @ACC>
fio     [fio] N M S @AS<                                #[fiar] V PR 1S IND VFIN @FMV
```

b)

```
Zanini  [Zanini] PROP M S @SUBJ>
diz     [dizer] <fmc> V PR 3S IND VFIN @FMV
que     [que] KS @SUB @#FS-<ACC
está    [estar] V PR 3S IND VFIN @FAUX
preparando      [preparar] V GER @IMV @#ICL-AUX<
a       [o] <artd> DET F S @>N
fundação        [fundação] N F S @<ACC
de      [de] PRP @N<
um      [um] <arti> DET M S @>N
novo    [novo] ADJ M S @>N
partido [partido] N M S @P<
$,
que     [que] <rel> SPEC M S @SUBJ> @#FS-N<             #KS @SUB @#FS-<ACC
trará   [trazer] V FUT 3S IND VFIN @FMV
explícito [explícito] ADV @<OC
em      [em] PRP @<ADVL
seu     [seu] <poss 3S> <si> DET M S @>N
programa        [programa] N M S @P<
```

c)

```
O       [o] <artd> DET M S @>N
projeto [projeto] N M S @SUBJ>
original [original] ADJ M S @N<
de      [de] <sam-> PRP @N<
o       [o] <-sam> <artd> DET M S @>N
governo [governo] N M S @P<
destinava       [destinar] <fmc> V IMPF 3S IND VFIN @FMV
a       [a] <sam-> PRP @<PIV
o       [o] <-sam> <artd> DET M S @>N
TSE     [TSE] PROP M S @P<                      #TSE=R$         [TSE=R$] PROP M S @P<
R$      [R$] N M P @<ACC
334,9   [334,9] <card> NUM M P @>N
milhões [milhão] N M P @N<
$.
```

No exemplo a) observa-se a total substituição da análise de "o", inicialmente analisado como pronome pessoal (forma acusativa) (PERS) e revisto como artigo definido/determinante (<artd> DET). É interessante observar a razão da análise do parser. Repare-se que "fio" foi analisado como verbo, daí a análise de "o" como pronome.

No exemplo b) por outro lado, a análise de "que" é globalmente alterada: de conjunção subordinativa a pronome relativo. Mais uma vez de notar a origem da análise automática: a presença de um verbo ilocutório ("dizer") seguido por uma conjunção subordinativa.

Finalmente, em c) há uma diferença de atomização introduzida uma vez que o parser considerou "TSE R$" uma unidade (ou polilexical).

- Alteração combinada, correspondente a alterações que envolvem apenas dois níveis gramaticais, por exemplo, PoS e função:

```
a           [o] <-sam> <artd> DET F S @>N
escravatura         [escravatura] N F S @SUBJ>
ter         [ter] V INF 3S @IAUX @#ICL-P<
desvirtuado         [desvirtuar] V PCP @IMV @#ICL-AUX<
e           [e] <co-pcv> KC @CO
tolhido     [tolher] V PCP @IMV @#ICL-AUX<
a           [o] <artd> DET F S @>N
evolução            [evolução] N F S @<ACC
de          [de] <sam-> PRP @N<
o           [o] <-sam> <artd> DET M S @>N
algoritmo           [algoritmo] N M S @P<
etnocêntrico        [etnocêntrico] ADJ M S @N<
que         [que] <rel> SPEC M S @ACC> @#FS-N<
os          [o] <artd> DET M P @>N
negros      [negro] N M P @SUBJ>                    #ADJ @>N
americanos          [americano] ADJ M P @N<         # N @SUBJ>
teriam      [ter] V COND 3P VFIN @FAUX
desenvolvido        [desenvolver] V PCP @IMV @#ICL-AUX<
```

Apesar de nesta categoria estarem incluídas outras como, por exemplo, morfologia e função, estas não foram consideradas nas categorias só morfologia e/ou só função. Isto significa que os valores de morfologia, função sintáctica e PoS referem-se às alterações exclusivamente nestes campos.

- Erro e irrelevante, correspondente a diferenças observadas entre o ficheiro original e o ficheiro revisto mas que não correspondem a alterações relevantes para a presente discussão, como por exemplo, contagem de diferenças de espaçamento ou sinalização de alteração a introduzir mas, de facto, não introduzida;
- Escolha de análises de função sintáctica alternativas constantes na análise inicial do parser, correspondente à escolha pelos revisores de uma das análises alternativas, e no fundo, correspondente à desambiguação de função sintáctica.
- Etiquetas secundárias, correspondente às alterações produzidas exclusivamente a este nível;
- Lema, correspondente às alterações produzidas exclusivamente a este nível.

Estabeleceram-se também algumas hipóteses relativamente aos níveis gramaticais em que as alterações nas diversas dimensões foram observadas. Assim, supôs-se que na revisão 1 da análise original do parser as alterações observadas seriam mais frequentemente da ordem da desambiguação de função, morfologia (@PRED> @ADVL>; M/F S/P) e alterações globais.

Relativamente às alterações do revisor 2 à análise original do analisador automático, a hipótese foi de que se tratariam de alterações mais finas, por exemplo de ordem semântica, como alteração nas etiquetas secundárias, e menos alterações globais bem como menos alterações por desambiguação.

O processo de análise das diferenças contabilizadas focou-se, desta forma, no seguinte:

- relativamente a revisor 1-parser e revisor 2-parser, investigou-se a que nível(eis) gramatical (função, PoS, morfologia, etc) as alterações introduzidas são mais notórias e dentro de cada nível quais são as alterações mais frequentes;

- relativamente a revisor 1-revisor 2 a questão essencial esteve em saber o número de alterações do do revisor 2 que retornaram à análise inicial do parser e a que nível(eis) gramaticais esta situação ocorreu mais frequentemente. Investigou-se igualmente quantas alterações do revisor 2 ao revisor 1 foram independentes da análise inicial do analisador automático, ou seja, quantas alterações do revisor 2 são uma nova proposta de análise;

## 4. Resultados

O número de diferenças observadas em cada uma das fases representadas no diagrama 1 foi o seguinte:

a) revisor 1 - parser: 2110 diferenças[1]
b) revisor 2 - parser: 1488 diferenças
c) revisor 2 – revisor 1: 363 diferenças

Como seria de esperar, o maior número de diferenças registou-se na revisão 1 da análise original do parser. No entanto, é também de notar o elevado número de alterações introduzidas ao parser pelo revisor 2.

### 4. 1. Revisor 1 – parser: alterações

Na tabela seguinte encontram-se os resultados mais significativos relativos ao tipo de alteração introduzida pelo revisor 1 à análise original do parser:

| Tipo de alteração | Função sintáctica | PoS | Morfologia | Alteração global | Alteração combinada | Escolha de análises alternativas originais | Etiquetas secundárias | Lema |
|---|---|---|---|---|---|---|---|---|
| Frequência | 617 | 16 | 432 | 430 | 113 | 85 | 29 | 0 |

Tabela 1: Número de alterações introduzidas pelo revisor 1 (ao parser) por categorias mais relevantes

Tal como foi inicialmente previsto, na primeira revisão da análise original a maior parte das alterações introduzidas foram a nível de função sintáctica (617 ocorrências), morfologia (432 ocorrências) e a nível global (430 ocorrências), enquanto que distinções mais finas, normalmente representadas por etiquetas secundárias, não foram tão frequentes na primeira fase de revisão (apenas 29 ocorrências).

Não se registaram alterações ao lema. De forma geral, o lema não oferece problemas em termos de representação, por um lado porque muitos dos casos foram já discutidos (como por exemplo, lemas de pronomes pessoais em forma não nominativa) e controlados (e a solução automaticamente implementada) e os casos mais complexos que requerem intervenção humana são reduzidos. A título de exemplo refira-se casos de nomes comuns ou próprios, cuja alteração condiciona alterações ao lema (por exemplo, rosa como analisado como nome comum e que deverá

[1] Esta contabilização não é precisa, porque no ficheiro correspondente às alterações do revisor 1 ao parser estão também incluídas algumas alterações do revisor 2 ao revisor 1, que deviam constar do número de alterações corrigidas.

ser revisto como nome próprio determina que o lema seja também alterado: de rosa para Rosa) ou lemas que requeiram forma plural (por exemplo, o lema de Finanças deverá ser Finanças e não Finança). É também possível que o lema sofra alterações que se incluem em alterações globais.

É também de salientar o facto de alterações a nível de PoS terem sido muito poucas, o que sugere que a este nível o parser oferece um elevado grau de confiança. Este facto pode estar relacionado talvez por a identificação/análise da classe de palavras ser essencialmente baseada por um léxico previamente elaborado, com informação, por palavra, da sua classe gramatical, género e número, tempo verbal, pessoa e modo. Desta forma, a percentagem de erro do parser poderá ter a ver possivelmente com a semelhança formal, como é o caso de sentido, substantivo e verbo sentir:

Tem [ter] <fmv> V PR 3S IND VFIN @FMV
**sentido [sentido] N M S @<ACC**

Tem [ter] V PR 3S IND VFIN @FAUX
**sentido [sentir] V PCP @IMV @#ICL-AUX<**

Analisou-se também, ao nível de função sintáctica, quais as funções mais frequentemente alteradas e quais as novas funções introduzidas.
Os resultados estão representados na tabela seguinte:

| Par função alterada/função introduzida[a)] | @N/@A | @A/@ADVL | @N/@N<PRED | @N/@ADVL | @PIV/@ADVL | @SUBJ/@SC | @SUBJ/@ACC |
|---|---|---|---|---|---|---|---|
| Frequência | 7 | 7 | 21 | 149 | 42 | 18 | 22[b)] |

Tabela 2: Frequência de par função alterada e função introduzida (revisor 1)

A tabela acima representa apenas as funções que foram alteradas e as análises introduzidas. No entanto, é importante referir que houve ocorrências não só de alteração de função mas também de valência verbal, isto é, alterações na relação verbo - constituintes. Em formato C(onstraint) G(rammar) isto equivaleria a dizer que a direcção marcada nos constituintes (por **>** , verbo à direita e **<**, verbo à esquerda) para o seu núcleo (verbo) é alterada, apesar da função ser mantida. 31 destes casos foram observados.

Relativamente aos valores da tabela 2, verifica-se que o par permutável (ver nota de rodapé a)) @N/@ADVL é o mais frequentemente modificado (149 ocorrências) do que qualquer outro par de funções.

No caso da morfologia (432 alterações pelo revisor 1 ao parser), é fundamental saber se as alterações produzidas estão relacionadas com desambiguação de formas dadas como ambíguas pelo parser (por exemplo, M(asculino)/F(eminino), S(ingular)/P(lural), etc.), conforme a hipótese de que se partiu, ou seja, que a revisão 1 iria contemplar elevada frequência de alterações por desambiguação. Os resultados encontram-se na seguinte tabela:

[a)] O par função alterada/função introduzida é permutável. Assim, por exemplo, na terceira coluna, as 149 ocorrências dizem respeito @N substituída por @ADVL e vice-versa. O mesmo se aplica aos outros pares.
[b)] Destas ocorrências, 12 são a nível de função externa, isto, é @#ICL-SUBJ / @#ICL-ACC ; @#FS-SUBJ / @#ICL-ACC ou @#AS-SUBJ / @#AS-ACC.

| Tipo de alteração | Desambiguação | | | | |
|---|---|---|---|---|---|
| | Número | Género | Tempo verbal (PS/MQP) | Pessoa | |
| | | | | 1/3S | 3S/P |
| **Frequência** | 4 | 388 | 1 | 28 | 4 |

Tabela 3: Frequência de alterações a nível morfológico pelo revisor 1 distribuídas pelas categorias mais relevantes

Como se pode observar, a maior parte das alterações ao parser em morfologia produzidas pelo revisor 1 estão relacionadas com desambiguação de formas.

### 4.2. Revisor 2-parser: alterações

Recapitulando, as alterações à análise original do parser pelo revisor 2 foram introduzidas via ficheiro alterado pelo revisor 1. Desta forma, o revisor 2 teve acesso às alterações que o revisor 1 produziu e o seu processo de revisão foi a alteração não da revisão do revisor 1 mas da análise original do parser, que o revisor 1 por sua vez não alterou (não marcadas por #).

A hipótese mais intuitiva seria que, ao contrário da revisão 1, o revisor 2 por um lado introduzisse menos alterações globais e combinadas, menos desambiguações e, por outro, introduzisse alterações relacionadas com parâmetros de análise mais finos, como por exemplo a inclusão/exclusão/modificação de etiquetas secundárias. O número de alterações em função sintáctica, PoS e morfologia descresceria relativamente às alterações do revisor 1 nos mesmos níveis.

Tal como para o revisor 1, investigou-se quais os níveis gramaticais em que ocorreram as alterações. Os resultados encontram-se na tabela seguinte:

| Tipo de alteração | Função sintáctica | PoS | Morfologia | Alteração global | Alteração combinada | Escolha de análises alternativas originais | Etiquetas secundárias | Lema |
|---|---|---|---|---|---|---|---|---|
| Frequência | 468 | 15 | 311 | 83 | 48 | 9 | 126 | 8 |

Tabela 4: Número de alterações introduzidas pelo revisor 2 ao parser por categorias mais relevantes

O seguinte diagrama apresenta os valores comparativos correspondentes às alterações do revisor 1 2 ao parser (constantes nas tabelas 1 e 4):

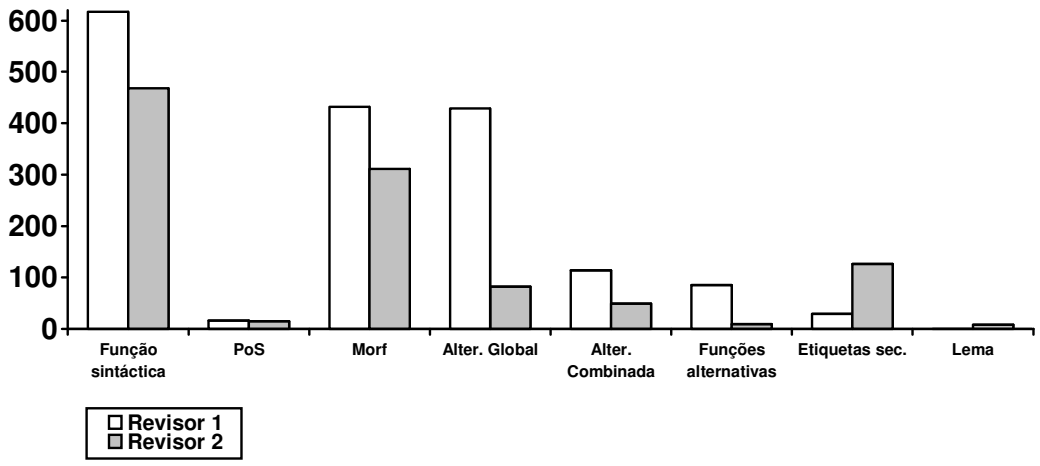


Diagrama 2: Gráfico de barras comparativo das categorias mais frequentemente alteradas pelo revisor 1 e pelo revisor 2 ao parser.

Confirmando a hipótese avançada estão os números relativos a alterações globais (82, revisor 2 contra 429, revisor 1), alterações combinadas (49 contra 114), desambiguação entre funções originalmente fornecidas pelo parser (9 contra 85) bem como os números referentes a modificação em termos de etiquetas secundárias (126, revisor 2, contra 29, revisor 1).

No entanto, observou-se 8 alterações apenas ao lema/forma base, na revisão 2 ao parser, sem que nenhuma alteração deste tipo se tenha verificado na revisão 1. Este facto pode estar relacionado, como já foi referido, com o facto de o lema ser, na revisão 1, alterado maioritariamente por alteração combinada ou global. Além disso, neste processo pode ser por erro que apenas o lema, não fosse alterado pelo revisor 1, falha identificada pelo revisor 2, como o seguinte exemplo ilustra:

*parser*:

os [o] <artd> DET M P @>N
atletas [atleta] N M P @SUBJ>
que [que] <rel> SPEC M P @SUBJ> @#FS-N<
não [não] ADV @ADVL>
atuaram [atuar] V PS 3P IND VFIN @FMV
contra [contra] PRP @<ADVL
o [o] <artd> DET M S @>N
**Vitória [vitória] <prop> N F S @P<**

*revisor 1*: Vitória [vitória] PROP M S @P<

*revisor 2*: Vitória [Vitória] PROP M S @P<

O revisor 1 introduz uma alteração combinada, alterando PoS e morfologia (género); no entanto, o lema, que também devia ter sido alterado, permanece segundo a análise original do parser. O revisor 2 procede à alteração da análise original do parser em relação ao lema.

Comparando os números relativos às alterações de função e morfologia do revisor 2 com os valores das mesmas categorias no revisor 1, verifica-se um decréscimo das alterações do revisor 2

relativamente ao revisor 1 (465 contra 617, em função, e 311 contra 432, em morfologia), que não é tão significativo, em números, isto é, a frequência de alterações em termos de função e morfologia pelo revisor 2 continuam a ser muito elevados. Relativamente a PoS, os baixos valores também se mantêm (15 alterações, revisor 2, contra 16 alterações, revisor 1).

Com base nestes resultados, investigou-se em primeiro lugar, que funções foram mais frequentemente alteradas e em termos de morfologia quais as categorias em termos de desambiguação em que o revisor 2 mais se focou.

Em termos de alteração da função sintáctica pelo revisor 2, os resultados encontram-se na tabela seguinte:

| **Par função alterada/função introduzida** | @N/@A | @A/@ADVL | @N/@N<PRED | @N/@ADVL | | @PIV/@ADV()[c)] | @SUBJ/@SC | @SUBJ/@ACC |
|---|---|---|---|---|---|---|---|---|
| **Frequência** | 11 | 23 | 2 | 147 | 135 função interna | 30 | 2 | 9 |
| | | | | | 12 função externa | | | |

Tabela 5: Frequência de par função alterada e função introduzida (revisor 2)

Comparando os números da tabela 5 com os da tabela 2, e olhando para o gráfico em baixo, em quase todos os pares função alterada/função introduzida verifica-se um descréscimo na revisão 2, como se previa. No entanto, ao contrário do que se previu, os valores de @N/@ADVL e @PIV/@ADVL (ou espécies de ADVL, isto é ADVS e ADVO), apresentaram um descréscimo foi pouco relevante. E @N/@A e @A/@ADVL sofreram um aumento do número de alterações na revisão 2.

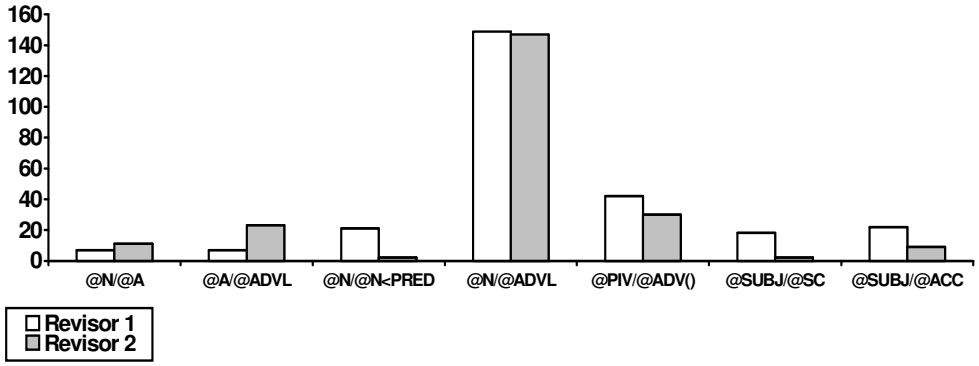


Diagrama 3: Gráfico de barras comparativo dos pares de funções mais frequentemente alterados pelo revisor 1 e pelo revisor 2 ao parser.

Ainda de salientar 87 ocorrências de alterações em termos de função externa, o que é muito significativo, uma vez que se relaciona directamente com resultados obtidos em morfologia, como veremos.

Ao nível da morfologia, investigou-se qual a frequência de alterações por tipos de desambiguação. Os resultados encontram-se na tabela seguinte:

[c)] ADV() corresponde às funções ADVL, ADVS e ADVO.

| **Tipo de alteração** | Desambiguação de pessoa | | Desambiguação tempo verbal (PS/MQP) | Desambiguação de número | Desambiguação de Género | Desambiguação de caso (NOM/PIV) |
|---|---|---|---|---|---|---|
| | 1/3S | 3S/P | | | | |
| **Frequência** | 9 | 1 | 74 | 1 | 10 | 18 |

Tabela 6: Frequência de desambiguação a nível morfológico distribuídas pelas categorias mais relevantes

É muito importante também referir que a introdução de sinalização de formação de oração por um particípio passado (ICLPCP) foi muito significativa no revisor 2, com 88 ocorrências. Isto é, o parser não assinalou formação de oração por particípio passado em 88 casos. Este valor é praticamente o mesmo que o valor obtido nas diferenças a nível de introdução de função externa (87 ocorrências), e os dois estão de facto relacionados, uma vez que a introdução de função externa irá implicar a introdução da sinalização de oração (a nível morfológico:

*parser*:

```
um          [um] <arti> DET M S @>N
estudo      [estudo] N M S @NPHR
apresentado     [apresentar] V PCP M S @N<
em          [em] <sam-> PRP @N<
a           [o] <-sam> <artd> DET F S @>N
5ª=Conferência=internacional=sobre=Transtornos=Alimentares
            [5ª=Conferência=internacional=sobre=Transtornos=Alimentares]  PROP F S @P<
```

*revisor 2*:

```
...
apresentado     [apresentar] V ICLPCP M S @IMV @#ICL-N<
em          [em] <sam-> PRP @<ADVL
...
```

Como se pode observar pelo exemplo acima, é frequente que a introdução de marcação de oração por particípio passado seja acompanhada de introdução de função externa, de modo que a frequência de introdução de marcação de oração participial (ICLPCP) e de função externa seja muito próxima quantitativamente.

O seguinte gráfico apresenta os valores comparativos entre as alterações introduzidas pelo revisor 1 e 2, em termos de morfologia:

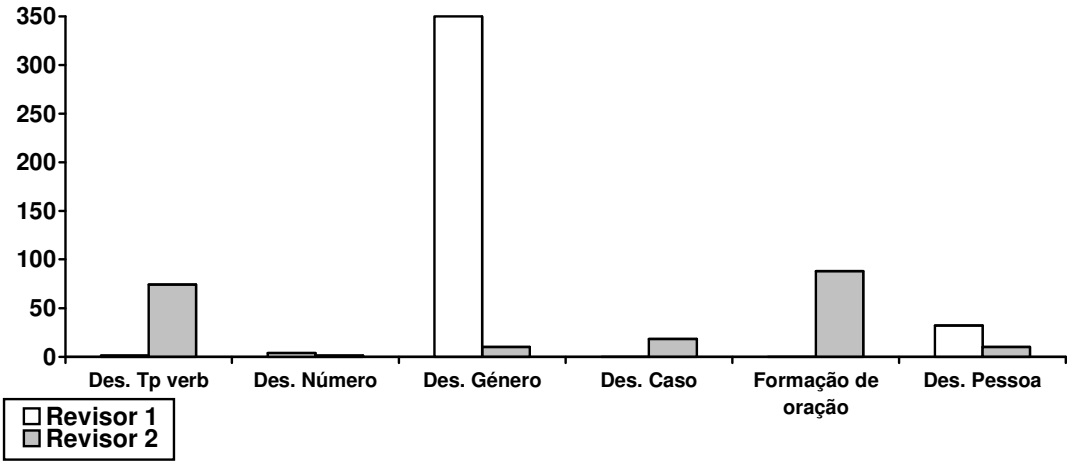


Diagrama 4: Gráfico de barras comparativo entre o revisor 1 e revisor 2 (relativamente ao parser) das alterações mais frequentes em morfologia

Conformando as expectativas, verificaram-se baixos níveis de desambiguação a nível de género e número pelo revisor 2, se comparados com os valores respeitantes às alterações introduzidas pelo revisor 1. No entanto, a desambiguação do tempo verbal (Pretérito mais que perfeito e perfeito) não teve expressão na revisão1.

### 4.3. Alterações combinadas

Como foi acima referido para as categorias formadas, por "alterações combinadas" entende-se as alterações introduzidas em mais do que um nível gramatical (por exemplo PoS e função). O objectivo de se formar esta categoria relacionou-se com a análise quantitativa de quantas alterações se produziram para cada entrada. Assim, previu-se que as análises globais e combinadas, porque envolvem uma revisão a mais do que um nível, seriam mais frequentes na revisão1. Mas de forma a não redobrar a contagem, optou-se por considerar alterações de PoS ou função aquelas alterações que exclusivamente ocorreram a estes níveis. Logo, as alterações de, por exemplo, PoS ou função nas alterações combinadas não foram contabilizadas como combinadas, PoS e função, mas tão só combinadas.

No entanto, esta categoria torna opacas as alterações a diversos níveis gramaticais cuja análise pode revelar-se importante. Desta forma, considera-se fundamental examinar mais a fundo as alterações combinadas, isto é, investigar a que níveis gramaticais as alterações combinadas ocorreram.

#### 4.3.1. Níveis de alterações combinadas: revisor 1

Os tipos mais frequentes de alterações combinadas são:

- morfologia e função (78 ocorrências);
- PoS e função (30 ocorrências);

As alterações combinadas mais frequentes envolvem alterações de função. As funções mais frequentes foram, indistintamente, nas duas dimensões acima, as seguintes:

- @N/@A (13 ocorrências);
- @SUBJ/@ACC (11 ocorrências);
- @ACC/@ACC-PASS (10 ocorrências).

Investigou-se também até que ponto as alterações combinadas acima seriam interdependentes, isto é, até que ponto a alteração de morfologia estaria dependente da alteração da função e/ou vice-versa e a alteração de PoS estaria dependente da alteração de função e/ou vice-versa.

Verificou-se que a alteração "morfologia e função" incidiu sobre a análise do pronome reflexo "se" em 18 casos:

O [o] <artd> DET M S @>N
time [time] N M S @SUBJ>
pode [poder] <fmc> V PR 3S IND VFIN @FAUX
empatar [empatar] V INF @IMV @#ICL-AUX<
a [o] <artd> DET F S @>N
segunda [segundo] <NUM-ord> ADJ F S @>N
partida [partida] N F S @<ACC
para [para] PRP @<ADVL
**se [se] PERS M 3S ACC @ACC> #M/F 3S/P @SUBJ>**
classificar [classificar] V INF 3S @IMV @#ICL-P<

Estes casos são exemplos de interdependência na revisão e diferentes níveis. Os pares de funções alteradas nestes casos foram precisamente @SUBJ/@ACC (6 ocorrências) e @ACC/@ACC-PASS (10 ocorrências). Relativamente às funções @SUBJ/@ACC, o pronome reflexo "se" terá muito claramente uma forma morfológica definida caso tenha a função de sujeito, isto é, M/F S/P. Assim, a análise de sujeito irá implicar uma alteração a nível da morfologia.

No caso de @ACC/@ACC-PASS, a revisão da função no sentido de partícula apassivante (@ACC-PASS) poderá também implicar uma alteração a nível da morfologia, uma vez que terá de haver uma concordância em género e número em relação ao sujeito.

Finalmente, as funções @N/@A. Este é um caso muito peculiar porque as 13 ocorrências referem-se ao mesmo contexto: ocorrência de unidades monetárias (US$; CR$, etc) antes dos seus modificadores e envolvendo numerais e o substantivo "milhão" ou "bilhão" em que o numeral tem a função de @>A em vez de @>N como habitualmente. Os seguinte exemplo ilustra a questão:

o [o] <artd> DET M S @>N
jornal [jornal] N M S @APP
de [de] <sam-> PRP @N<
o [o] <-sam> <artd> DET M S @>N
Vaticano [Vaticano] PROP M S @P<
$,
qualificou [qualificar] <fmc> V PS 3S IND VFIN @FMV
a [o] <artd> DET F S @>N
campanha [campanha] N F S @<ACC
de [de] PRP @N<
US$ [US$] N M P @P<
**15 [15] <card> NUM M P @>A**
milhões [milhão] N M P @N<
como [como] <rel> ADV @COM @#AS-N<
$"«
terrorismo [terrorismo] N M S @AS<
de [de] PRP @N<
imagem [imagem] N F S @P<
$"»

Note-se que em o mesmo extracto do CETENFolha usado para a contagem das diferenças, qualquer modificador de um núcleo de natureza nominal, como "milhão" possui a função @>N:

```
Apenas   [apenas] ADV @>A
5        [5] <card> NUM M P @>N
milhões  [milhão] N M P @SUBJ>
de       [de] PRP @N<
contribuintes    [contribuinte] N M P @P<
haviam   [haver] <fmc> V IMPF 3P IND VFIN @FAUX
entregue [entregar] V PCP M S @IMV @#ICL-AUX<
sua      [seu] <poss 3S> <si> DET F S @>N
declaração       [declaração] N F S @<ACC
de       [de] PRP @N<
Imposto=de=Renda         [Imposto=de=Renda] PROP M S @P<
...
```

A razão para este facto tem origem bem definida. Uma vez que o contexto com unidades monetárias representadas como acima surge pela primeira vez no corpus do CETENFolha, uma nova regra teve de ser especialmente criada para lidar com este contexto. E houve uma falha na escrita da regra, ou seja, a escolha da função @A para o numeral neste contexto em vez que @N.

### 4.3.2. Níveis de alterações combinadas: revisor 2

As alterações combinadas foram, como se referiu, menos frequentes na revisão 2. O tipo mais frequentes de alterações combinadas foi: morfologia e função (42 ocorrências).

A maior parte destes casos relaciona-se com a formação de orações participiais:

```
Os       [o] <artd> DET M P @>N
jogadores        [jogador] N M P @SUBJ>
se       [se] PERS M 3P ACC @ACC>
dividem  [dividir] <fmc> V PR 3P IND VFIN @FMV
por      [por] <sam-> PRP @<PIV
os       [o] <-sam> <artd> DET M P @>N
dez      [dez] <card> NUM M P @>N
quartos  [quarto] N M P @P<
de       [de] <sam-> PRP @N<
o        [o] <-sam> <artd> DET M S @>N
alojamento       [alojamento] N M S @P<
$,
equipados        [equipar] V ICLPCP M P @IMV !@#ICL-N<          !R2 PCP @N<PRED
com      [com] PRP @<ADVL
frigobar [frigobar] N M S @P<
...

Preocupado       [preocupar] V ICLPCP M S @IMV !@ICL-PRED>      !Su PCP @PRED>
com      [com] PRP @<ADVL
a        [o] <artd> DET F S @>N
facilidade       [facilidade] N F S @P<
de       [de] PRP @N<
comunicação      [comunicação] N F S @P<
de       [de] PRP @N<
seu      [seu] <poss 3S> DET M S @>N
adversário       [adversário] N M S @P<
...
Mário=Covas      [Mário=Covas] PROP M S @SUBJ>
acatou   [acatar] <fmc> V PS 3S IND VFIN @FMV
```

as [o] <artd> DET F P @>N
recomendações [recomendação] N F P @<ACC
...

Como se pode observar, a formação de uma oração participial implica a) a revisão da função interna, que passa a ser @IMV; 2) a adição de oração externa (@#ICL-N<PRED e @#ICL-PRED nos cexemplos acima) e 3) (possível) adição de informação morfológica sobre formação de oração-ICLPCP.

### 4.4. Revisor 2 – revisor 1: alterações

Nas alterações do revisor 2 às alterações introduzidas pelo revisor 1, investigou-se se as alterações do revisor 2 foram coincidentes com a análise original do parser, que por sua vez foi alterada pelo revisor 1, ou se as alterações do revisor 2 às alterações do revisor 1 foram distintas da análise original do parser, isto é, as alterações introduzidas pelo revisor 2 produziram uma nova análise independente da análise original do parser e, obviamente, da análise do revisor 1. Isto significa que a análise original do parser seria alterada pelo revisor 2 caso este tivesse acesso ao ficheiro original (não revisto pelo revisor 1), ou ainda que ambos os revisores teriam alterado a análise original do parser mas de formas diferentes.

#### 4.4.1 Alterações do revisor 2 coincidentes com a análise inicial do parser

O total de alterações ao revisor 1 pelo revisor 2 coincidentes com a análise original proposta automaticamente pelo parser foi 208.

Na seguinte tabela encontram-se estas alterações distribuídas por nível gramatical:

| Tipo de alteração | Função sintáctica | PoS | Morfologia | Alteração combinada | Alteração global | Escolha de análises alternativas originais | Etiquetas secundárias | Lema |
|---|---|---|---|---|---|---|---|---|
| Frequência | 167 | 2 | 22 | 1 | 0 | 16 | 0 | 0 |

Tabela 7: Número de alterações introduzidas pelo revisor 2 coincidentes com a análise original do parser por categorias mais relevantes

Relativamente à função sintáctica, é importante referir que tal como aconteceu com os resultados obtidos nas dimensões revisor 1-parser e revisor 2-parser, as funções sintácticas que são mais frequentemente modificadas continuam a ser @ADVL e @N (76 ocorrências das 167 alterações de função sintáctica).

#### 4.4.2. Alterações do revisor 2 à análise do revisor 1 distintas da análise inicial do parser

O total de alterações introduzidas pelo revisor 2 que representam uma análise independente da do parser (e obviamente do revisor 1) foi 154.

Na seguinte tabela encontram-se estas alterações distribuídas por nível gramatical:

| Tipo de alteração | Função sintáctica | PoS | Morfologia | Alteração combinada | Alteração global | Escolha de análises alternativas originais | Etiquetas secundárias | Lema |
|---|---|---|---|---|---|---|---|---|
| Frequência | 109 | 3 | 7 | 1 | 20 | 0 | 3 | 0 |

Tabela 8: Número de alterações introduzidas pelo revisor 2 não coincidentes com a análise original do parser por categorias mais relevantes[1]

Relativamente à categoria de função sintáctica, as mais recorrentes nas alterações referem-se a @N e @ADVL com 17 ocorrências (no total de 109 alterações em função sintáctica). Apesar de este valor não ser muito significativo, alerta mais uma vez para o facto do uso das funções @N e @ADVL parecer ser, de facto, uma questão que dificilmente reúne consenso.

Alguns exemplos de casos em que uma análise correcta de cada uma destas funções podem ser enumerados: impossibilidade de @ADVL face à inexistência de um elemento verbal[2]:

```
JFK       [JFK] PROP M S @NPHR
$--
A=Pergunta=Que=Não=Quer=Calar [A=Pergunta=Que=Não=Quer=Calar] PROP M S @N<PRED
Telecine  [Telecine] PROP F S @N<
$,
20h30     [20h30] N F P @N<PRED                                  !R2 @NPHR
$.
```

Outro exemplo comum ocorre em contexto de orações relativas, ou seja, com a análise da função dos elementos interna vs. externa à oração, como o exemplo ilustra:

```
O         [o] <artd> DET M S @>N
destaque  [destaque] N M S @SUBJ>
...
é         [ser] <fmc> V PR 3S IND VFIN @FMV
o         [o] <artd> DET M S @>N
inédito   [inédito] ADJ M S @>N
$"
Exterminador=do=Futuro=2=-=O=Julgamento=Final  [Exterminador=do=Futuro=2=-=O=Julgamento=Final] PROP M
S @<SC
$"
$,
em        [em] PRP @ADVL>                                        #R1 @<ADVL; #@ADVL>
que       [que] <rel> SPEC M S @P< @#FS-N<
Schwarzenegger  [Schwarzenegger] PROP M S @SUBJ>
é         [ser] V PR 3S IND VFIN @FMV
um        [um] <arti> DET M S @>N
robô      [robô] N M S @<SC
com       [com] PRP @N<
aparência         [aparência] N F S @P<
humana    [humano] ADJ F S @N<
```

[1] Outros números que fazem parte da contabilização, mas que não são relevantes, são 7 ocorrências em que a análise de revisor 2 se manteve igual à do revisor 1 e 4 ocorrências de contabilização de alterações irrelevantes.

[2] Poder-se-ia especular qual o motivo por que a análise de @ADVL é muitas vezes aplicada a estes casos. Do meu ponto de vista, este facto pode estar associado com a natureza do dependente nominal. A expressão de tempo ou lugar é prototipicamente associada à função @ADVL. Daí que um dependente nominal com estas características possa induzir o revisor a considerar a função mais central @ADVL.

A análise de "em que" em relação à oração relativa (em cima a negrito) ou em relação à oração principal não parece ser totalmente transparente. A análise original do parser é tornar "em que" um dependente de um possível núcleo nominal antecedente. No entanto, "em que" tem uma função interna (@ADVL) à oração relativa e não à oração externa (@N< desempenhada pela oração relativa como todo).

Quer dizer que o uso destas funções é por si só complexo, e não só a permutação entre as mesmas, porque @N e @ADVL são também substituídas por uma outra função (em princípio por @PIV, segundo os dados das tabelas 2 e 5).

Como se pode ver pela comparação das tabelas 7 e 8, as alterações ao revisor 1 que convergem e as que divergem da análise original do parser estão relativamente equilibradas (167 ocorrências convergentes e 108 divergentes). Isto significa que parece haver de facto uma diferença considerável no processo de revisão de cada um dos revisores. As 154 ocorrências de alteração do revisor 2 ao revisor 1 não coincidente com a análise original do parser assim o parece indicar.

Conclusões a retirar das 208 ocorrências de alteração ao revisor 1 pelo revisor 2 coincidentes com a análise original do parser são mais reservadas. Isto é, imediatamente se poderia supor que este número corresponderia à percentagem de erro humano introduzida à análise inicialmente proposta pelo parser. No entanto, esta seria segundo a perspectiva de um revisor. Por isso, de forma a se investigar a potencial percentagem de introdução de erro humano, um terceiro revisor teria de confirmar ou contrariar os valores correspondentes às alterações do revisor 2 ao revisor 1 coincidentes com o parser.

Além disso, este estudo teria outra dimensão se o ficheiro em ad correspondente ao cg que foi usado neste estudo, tivesse sido também revisto pelo revisor 1 previamente à revisão do cg pelo revisor 2. Isto porque, tendo em conta a experiência do processo de revisão da Floresta Sintá(c)tica, há sempre alterações que são introduzidas ao ficheiro em formato cg durante a revisão do ficheiro ad correspondente. Ou seja, o revisor 2 em vez de rever o ficheiro cg que foi revisto uma única só vez pelo revisor 1, reveria uma versão cg mais trabalhada. Neste caso, seria interessante analisar se o número de diferenças entre os dois revisores diminuiria.

## 5. Conclusão

Como conclusões gerais, apontam-se as seguintes:

a) relativamente às alterações de cada um dos revisores ao parser:

1. o revisor 1, como esperado, introduziu alterações na análise original do parser que estão primeiramente relacionadas com desambiguação ou alteração de formas e funções ou alterações a nível global ou alterações combinadas.
2. o revisor 2, como esperado, produziu menos alterações à análise original do parser no que diz respeito a alterações globais e desambiguação de formas (principalmente no que respeita à desambiguação de género) e funções e por outro lado introduziu alterações mais finas, como, por exemplo, a adição de etiquetas secundárias.
3. o revisor 2 produziu menos alterações em termos de alteração de função, morfologia e PoS originalmente proposta pelo parser, embora os valores respeitantes às alterações pelo revisor 2 ao parser estejam muito próximos dos valores respeitantes às alterações pelo revisor 1 ao parser.

b) relativamente às alterações do revisor 2 às alterações do revisor 1:

1. O número de alterações do revisor 2 que coincidiram com a análise inicalmente proposta pelo parser foi mais elevado do que o número de alterações pelo revisor 2 que foram distintas da análise inicial do parser. Estes resultados, no entanto, não representam directamente a introdução de erro humano, cuja avaliação teria de ser fundamentada num outro tipo de estudo que involva um terceiro revisor.

Em termos mais específicos, e relativamente às funções sintácticas mais frequentemente alteradas, as conclusões mais importantes são as seguintes:

a) @N e @ADV() foram em todas as dimensões observadas as funções que mais alterações sofreram, quer por substituição de @N por @ADVL e vice-versa, como a substituição de @ADV() por outras funções.
b) A alternância das funções @N e @ADVL levanta algumas questões importantes. Em primeiro lugar, é amplamente reconhecida a existência de reais ambiguidades linguísticas, como o clássico exemplo: "Vi o homem da bicicleta com binóculos". No entanto, se nestes casos, as duas funções terão de estar presentes como alternativas, em outros casos a desambiguação é possível.

   Assim, das 76 ocorrências referentes à substituição da função @N por @ADVL e vice-versa, contabilizou-se quantos destes casos porderiam ser de facto desambiguados sem margem de erro. Verificou-se que 61 destas ocorrências são controláveis, isto é, não se baseiam na interpretação de cada revisor mas na sintaxe.

c) Os casos cuja análise depende da interpretação de cada revisor, e com vista a se chegar a uma maior consistência entre anotadores, teriam de ser analisados em conjunto.

## Glossário[3]:

| | |
|---|---|
| **@SUBJ** | função sintáctica de sujeito |
| **@SC** | função sintáctica de predicativo do sujeito |
| **@ACC** | função sintáctica de objecto directo |
| **@ADVL** | função sintáctica de adjunto adverbial |
| **@ADVS** | função sintáctica de argumento adverbial (em relação ao sujeito) |
| **@ADVO** | função sintáctica de argumento adverbial (em relação ao objecto directo) |
| **@PIV** | função sintáctica de objecto preposicional |
| **@ACC-PASS** | partícula apassivante |
| **@N** | dependente em sintagma nominal |
| **@N<PRED** | adjecto predicativo |
| **@A** | dependente em sintagma adjectival ou sintagma adverbial |
| **@IMV** | verbo principal não finito |

[3] Para uma listagem completa das etiquetas usadas na Floresta Sintá(c)tica consultar http://visl.sdu.dk/visl/pt/symbolset-floresta.html

## Referências:

Afonso, S., Eckhard Bick, Renato Haber & Diana Santos. (2001). "Floresta sintá(c)tica: um treebank para o português", *Actas do XVII Encontro da Associação Portuguesa de Linguística* (Lisboa, Outubro de 2001), APL, pp. 533-45.

Afonso, S., Eckhard Bick, Renato Haber & Diana Santos. (2002). ""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, pp. 1698-1703.

Bick, E. (2000). *The Parsing System Palavras, Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press, 2000.

Hirshman, L. (1998). "The Evolution of Evaluation:Lessoan from the Message Understanding Conferences". In *Computer Speech and Language*, 12(4): 281-305.

Kilgarriff, A. (1998). "Gold standard datasets for evaluating word sense disambiguation programs". In *Computer Speech and Language*, 12 (4): 453-172.

Santos, D. (2000). *Tutorial on Evaluation of Natural Language Processing systems*, apresentado na Joint International Conference IBERAMIA/SBIA 2000 (Atibaia, São Paulo, Brazil), 19 November 2000.