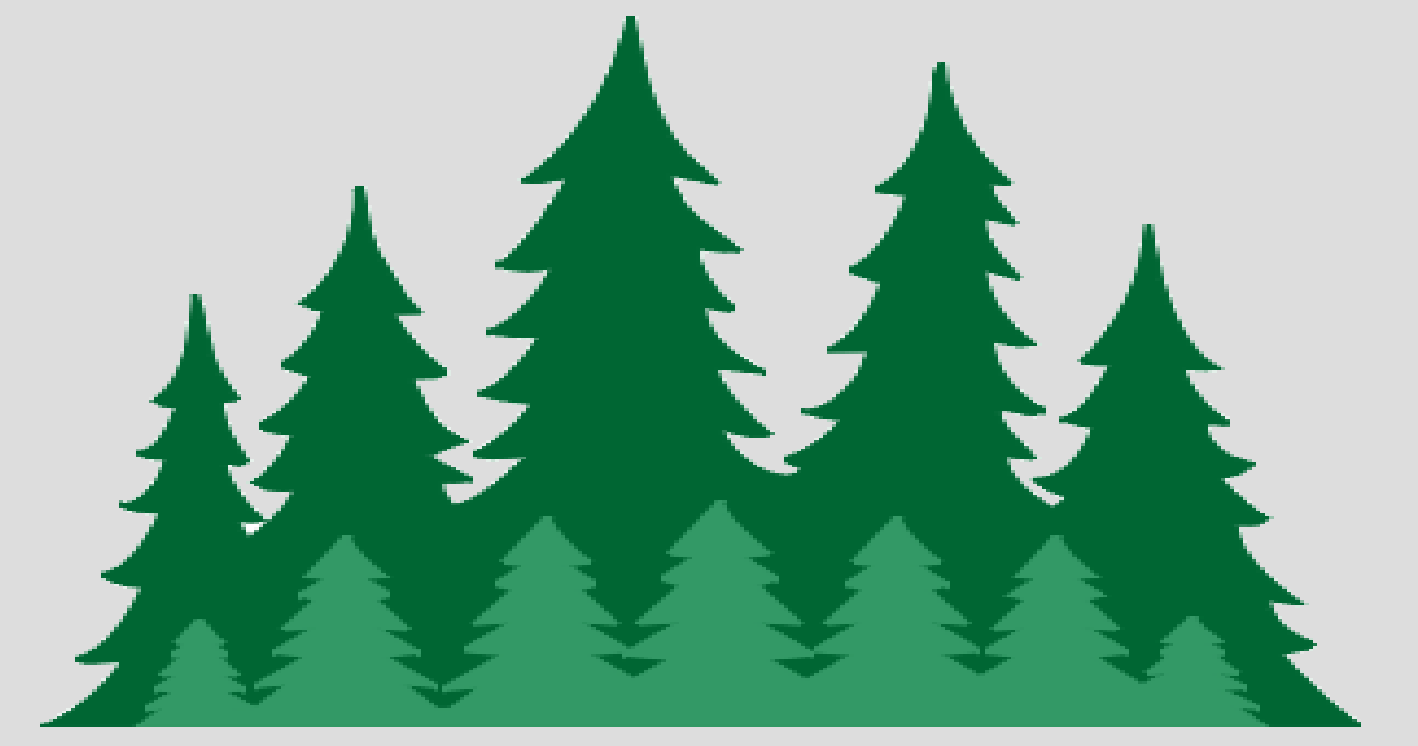


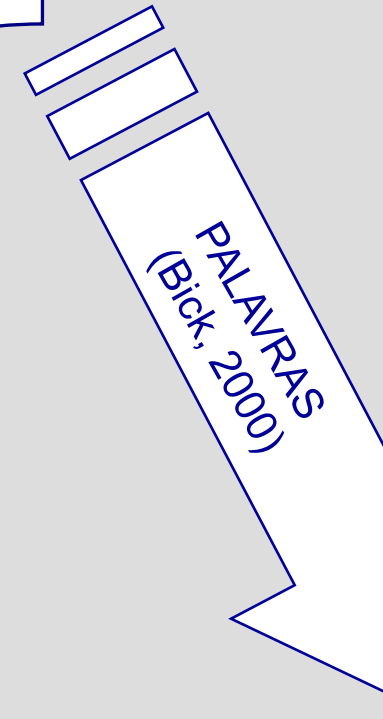
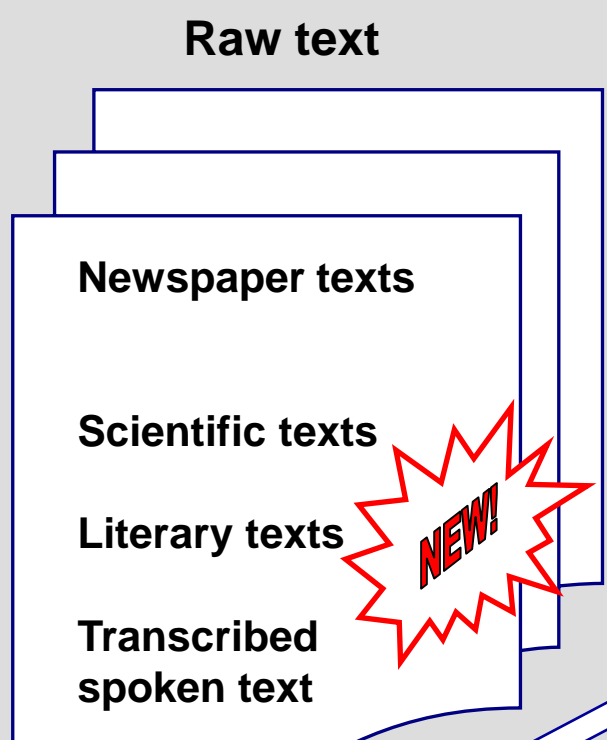
Floresta Sintá(c)tica



Cláudia Freitas (Linguatca)
Paulo Rocha (Linguatca)
Eckhard Bick (VISL)

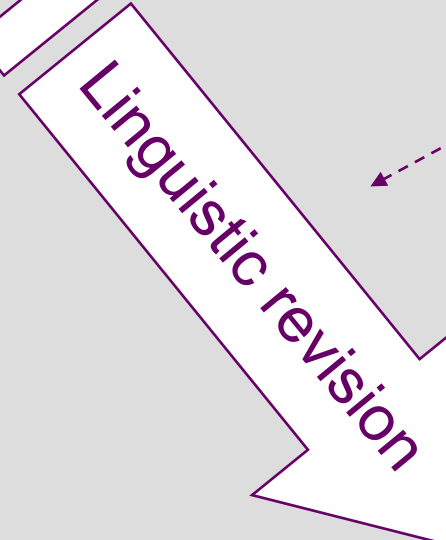
2000-2008

<http://www.linguatca.pt/Floresta/>



```

Formato árvores deitadas
CP460-12 A viatura em que seguia era igualmente roubada.
A1
STA:fcl
=SUBJ:np
=>N:art('o' <artd> F S) A
==H:n('viatura' <np-def> F S) viatura
==N<:fcl
===ADVL:pp
====H:ppp('em') em
====P<:np
=====H:pron-indp('que' <rel> F S) que
====P:v
=====MV:v-fin('seguir' <fs-rel> <no-subj> IMPF 3P IND) seguia
==P:v
====MV:v-fin('ser' IMPF 3S IND) era
==SC:adjp
==>A:adv('igualmente' <quant>) igualmente
==H:v-ppp('roubar' F S) roubada
=.
```



| Noun Modifiers | Noun Arguments |
|---|--|
| N< | N<ARGS |
| A <i>decisão</i> de ontem foi elogiada... | A <i>decisão</i> do Grupo ficaria difícil... |
| | N<ARGO |
| | (...) para que veja a sua <i>decisão</i> de boicote aos bancos |
| | N<ARG |
| (...) a comer tangerinas e a macular, com as cascas, os <i>montes</i> de neve branca ao longo das estradas. | "Um <i>monte</i> de artistas vai querer ver os concertos", diz Ohtake. |
| | ..bebia <i>copos</i> de água gaseificada. |

NEW!
Thicker

Partitives – secondary tags
"Um *monte* de artistas vai querer ver os concertos", diz Ohtake. <np-partitive1>
...bebia *copos* de água gaseificada <np-partitive1>
...metade dos *leitores* <np-partitive2>

NEW!

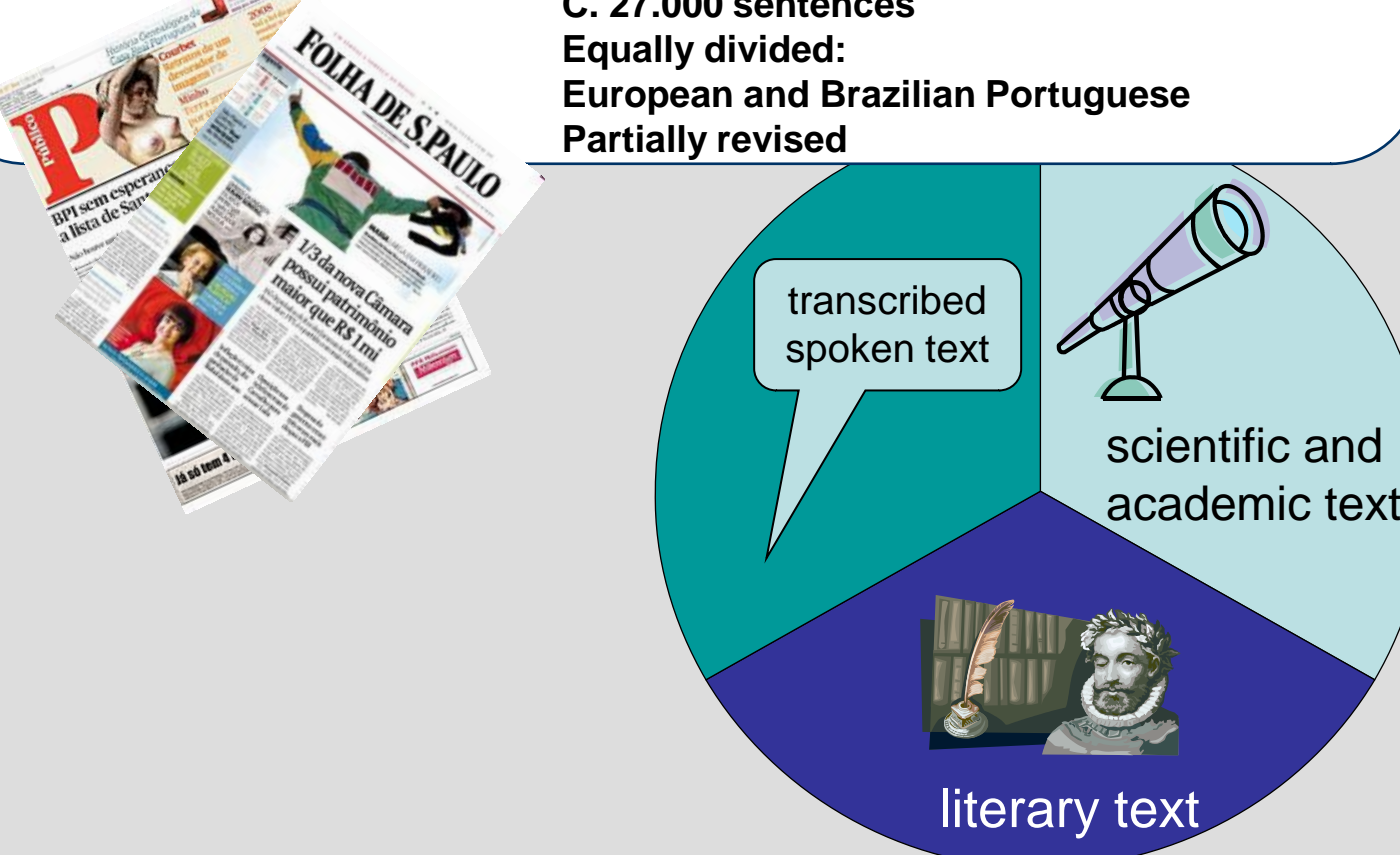
SEARCHABLES

- np-dequant estruturas com quantificadores
- np-partitive1 partitivos com "não contáveis"
- np-partitive2 partitivos com "contáveis"
- fs-conc verbos de orações concessivas
- fs-cause verbos de orações causais
- fs-conf verbos de orações conformativas
- fs-cond verbos de orações condicionais
- fs-cons verbos de orações consecutivas
- fs-temp verbos de orações temporais
- fs-fin verbos de orações finais
- fs-rel verbos de orações relativas
- fs-subst verbos de orações substantivas
- passive verbos de orações passivas
- se-passive verbos de orações passivas com se
- rel pronomes relativos
- no-subj orações sem sujeito explícito

Floresta Virgem
Unrevised newspaper texts
1.600.000 words

Bosque
Newspaper text
c. 190.000 word
9.368 sentences

Selva
Literary, scientific and transcribed spoken text
C. 100.000 words by section
C. 27.000 sentences
Equally divided:
European and Brazilian Portuguese
Partially revised



Bigger

```

Formato árvores deitadas
CP460-12 A viatura em que seguia era igualmente roubada.
A1
STA:fcl
=SUBJ:np
=>N:art('o' <artd> F S) A
==H:n('viatura' <np-def> F S) viatura
==N<:fcl
===ADVL:pp
====H:ppp('em') em
====P<:np
=====H:pron-indp('que' <rel> F S) que
====P:v
=====MV:v-fin('seguir' <fs-rel> <no-subj> IMPF 3P IND) seguia
==P:v
====MV:v-fin('ser' IMPF 3S IND) era
==SC:adjp
==>A:adv('igualmente' <quant>) igualmente
==H:v-ppp('roubar' F S) roubada
=.
```

```

Formato PennTreebank
(STA:fcl (SUBJ:np (>N:art:o:F_S:<artd> A)
(H:n:viatura:F_S: viatura)
(N<:fcl (ADVL:pp (H:ppp:em::: em)
(P<:np
(H:pron-indp:que:F_S:::rel: que)))
(P:v (MV:v-fin:seguir:IMPF_3P_IND::fs-rel: seguia)))
(P:v (MV:v-fin:ser:IMPF_3S_IND::fs-rel: era))
(SC:adjp (>A:adv:igualmente::quant: igualmente)
(S:v-ppp:roubar:F_S:::roubada))
(.
```

TGrep2
-both compiles and searches the corpus
-allows searches for the relationship between nodes (and not only their position)
-uses a standard format (PennTreebank)



Compiled Corpus

Milhafre

QUERY: "Informação" is the subject of ...

Gêneros: LIT FAL CIE

| | |
|------------|----|
| Resultados | 25 |
| confirmar | 4 |
| estar | 3 |
| faltar | 2 |
| permanecer | 2 |
| armazenar | 1 |
| caber | 1 |
| chegar | 1 |
| compor | 1 |
| contestar | 1 |
| corroborar | 1 |
| dar | 1 |
| encantar | 1 |

Variante: BR_PT

| | |
|------------|----|
| Resultados | 60 |
| ser | 12 |
| constar | 4 |
| chegar | 3 |
| estar | 3 |
| comunicar | 2 |
| confirmar | 2 |
| ficar | 2 |
| indicar | 2 |
| obter | 2 |
| surpre | 2 |
| aparecer | 1 |
| apontar | 1 |
| assegurar | 1 |
| atualizar | 1 |
| avanzar | 1 |
| averiguar | 1 |
| certificar | 1 |

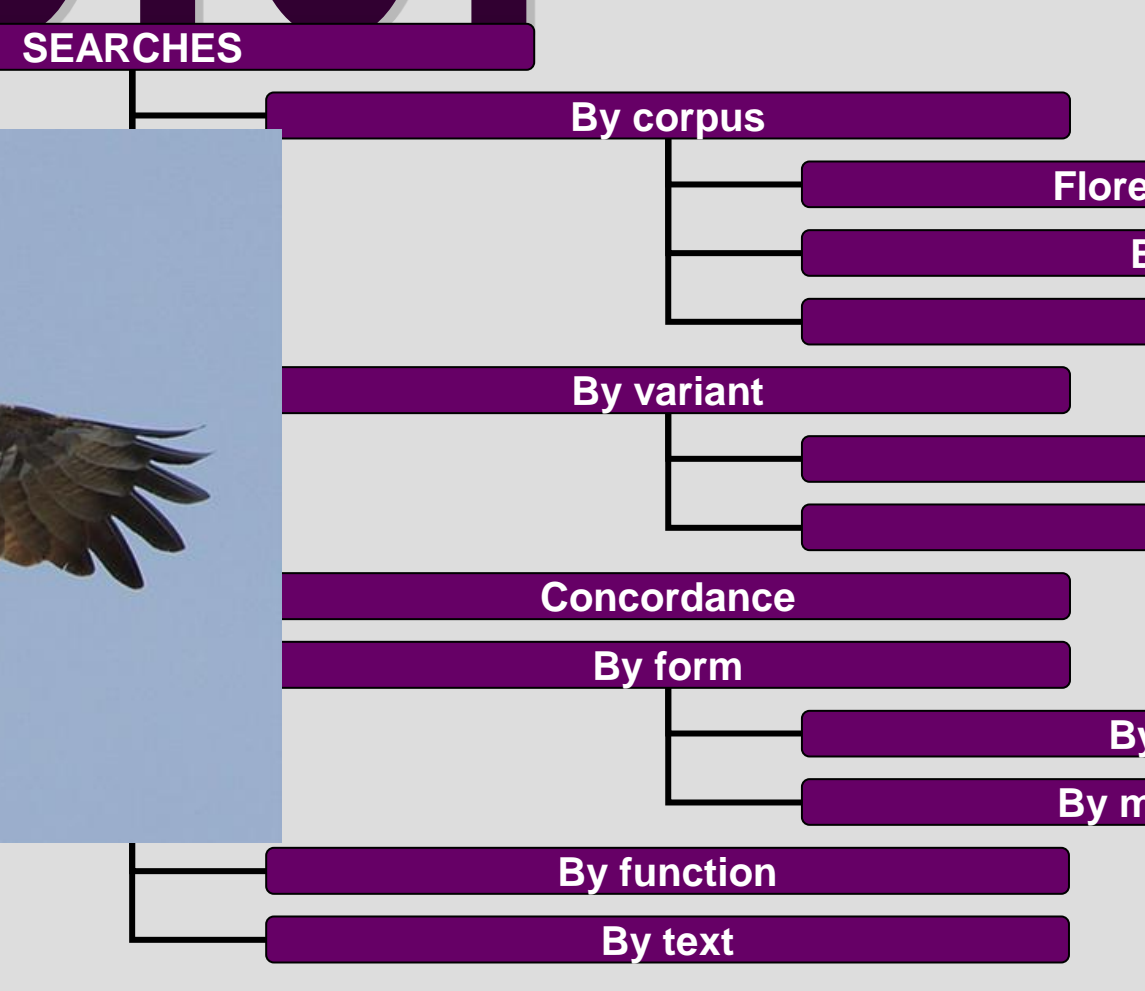
resultados completos em txt

Variante: BR_PT

| | |
|------------|---|
| Resultados | 7 |
| dar | 2 |
| chegar | 1 |
| confirmar | 1 |
| existir | 1 |
| fornecer | 1 |
| ser | 1 |

fornecer
• CP497-2 Não foram fornecidas informações sobre o balanço destes confrontos. !

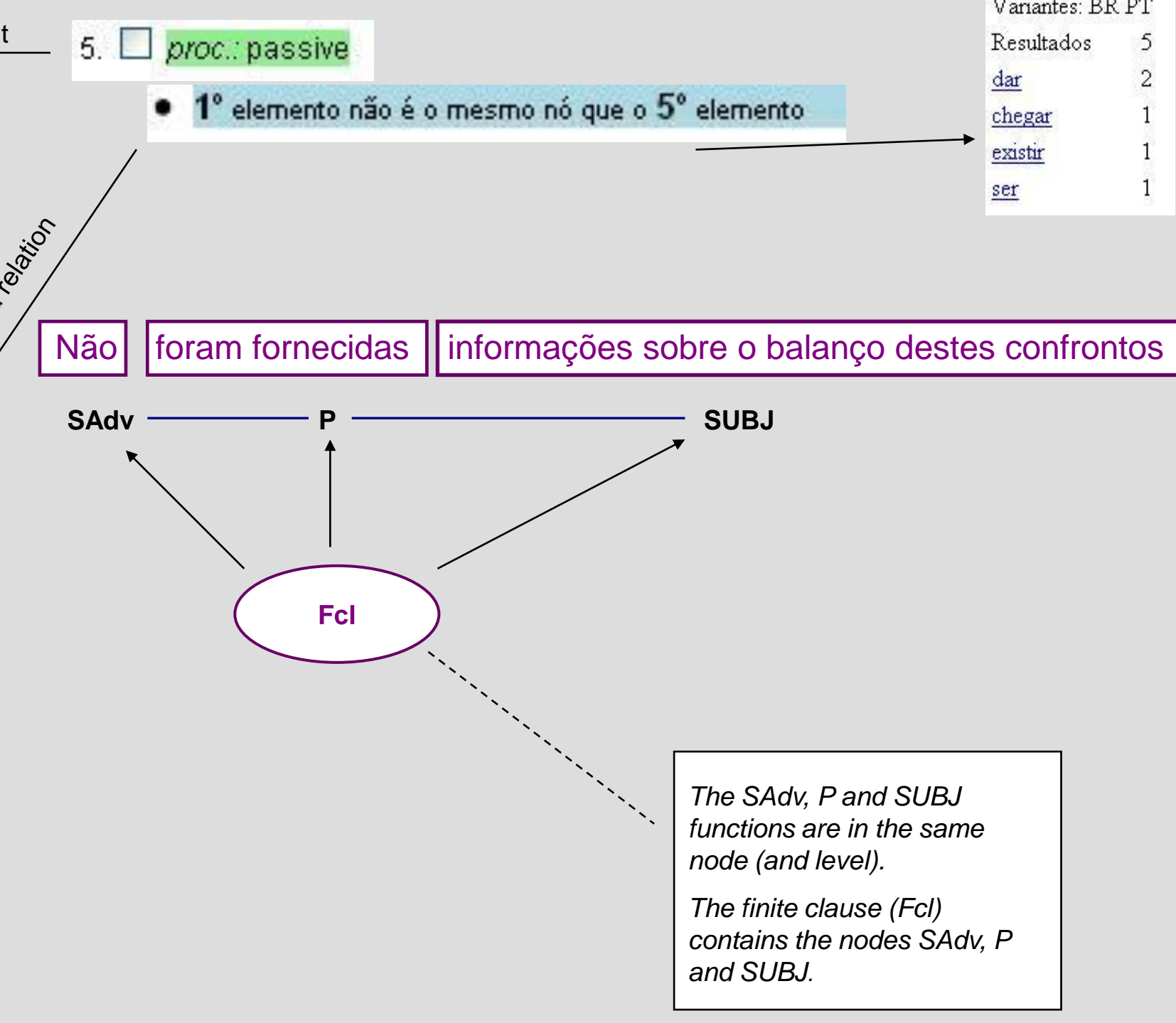
Easier



Milhafre (versão experimental)

Aplicação dos resultados:
 frases completas
 apenas extracto
 limpar resultados por
 função forma lema morfologia texto

| forma | função | morfologia |
|---|--|---|
| <input type="checkbox"/> formas oracionais | <input type="checkbox"/> enunciado | <input type="checkbox"/> elementos essenciais |
| <input type="checkbox"/> finita (Ec1) | <input type="checkbox"/> frase (FT) | <input type="checkbox"/> sujeito (SUJ) |
| <input type="checkbox"/> não-finita (Ec1) | <input type="checkbox"/> declarativa (DTA) | <input type="checkbox"/> objeto direto (OD) |
| <input type="checkbox"/> locativa (Ec1) | <input type="checkbox"/> interrogativa (QTE) | <input type="checkbox"/> objeto indireto (OI) |
| <input type="checkbox"/> Satzgruppen | <input type="checkbox"/> imperativa (IME) | <input type="checkbox"/> agente da passiva (PASS) |
| <input type="checkbox"/> nominal (np) | <input type="checkbox"/> exclamativa (EXC) | <input type="checkbox"/> particípio (PTC) |
| <input type="checkbox"/> adjetival (adjp) | <input type="checkbox"/> sujeito (SUJ1) | <input type="checkbox"/> verbo auxiliar (AV) |
| <input type="checkbox"/> adverbial (advp) | <input type="checkbox"/> objeto (OBJ) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> verbal (vp) | <input type="checkbox"/> objeto direto (OD) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> preposicional (pp) | <input type="checkbox"/> part. apassivante (ACC-PASS) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> evidenciador coord. (co) | <input type="checkbox"/> obj. ind. pronominal (OAT) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> frase discursiva (fp) | <input type="checkbox"/> obj. ind. preposicional (PIP) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> Rein lexical | <input type="checkbox"/> agente da passiva (PASS) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> substantivo (n) | <input type="checkbox"/> adjetivo adverbial (ADV) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> substantivo (n) | <input type="checkbox"/> adjetivo (ADJ) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> substantivo (n) | <input type="checkbox"/> nome próprio (ncop) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> adjetivo (adj) | <input type="checkbox"/> verbo (v) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> nome próprio (ncop) | <input type="checkbox"/> finito (v-f1a) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> verbo (v) | <input type="checkbox"/> gerúndio (v-ger) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> finito (v-f1a) | <input type="checkbox"/> particípio (v-pp) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> gerúndio (v-ger) | <input type="checkbox"/> infinitivo (v-inf) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> particípio (v-pp) | <input type="checkbox"/> artigo (art) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> infinitivo (v-inf) | <input type="checkbox"/> pronome (pcos) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> artigo (art) | <input type="checkbox"/> adjetivo (adv) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> pronome (pcos) | <input type="checkbox"/> numeral (num) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> adjetivo (adv) | <input type="checkbox"/> preposição (pp) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> numeral (num) | <input type="checkbox"/> interjeição (intj) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> preposição (pp) | <input type="checkbox"/> conjunção (conj) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> interjeição (intj) | <input type="checkbox"/> preterito (ec) | <input type="checkbox"/> verbo auxiliar (AVO) |
| <input type="checkbox"/> conjunção (conj) | | |
| <input type="checkbox"/> preterito (ec) | | |



References

- Eckhard Bick. The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. 2000
- Douglas Rohde. "TGrep2 User Manual: version 1.15". 2005
- Susana Afonso, Eckhard Bick, Renato Haber & Diana Santos. "Floresta sintá(c)tica: a treebank for Portuguese" (LREC 2002)

Linguatca is financed by the Portuguese government and the European Union through grant n° 339/1.3/C/NAC