The CORDIAL Parsing

Ernestina Carrilho
Catarina Magro
(University of Lisbon)

Overview of presentation

- short description of CORDIAL-SIN (corpus properties and current state)
- detailed account of CORDIAL-SIN syntactic annotation
  - the annotation system – goals, principles and guidelines
  - the annotation process – tools and procedures
  - the adjustments on the system required by a Portuguese corpus of dialectal speech

CORDIAL-SIN

- Syntax-oriented Corpus of Portuguese Dialects
- a corpus of spoken dialectal European Portuguese
- a geographically representative body of excerpts of spontaneous and semi-directed speech
- excerpts selected from the oral interviews gathered by the Research Group on Linguistic Variation at CLUL within the scope of several linguistic atlases
  - Informants: aged, low instruction, rural, born and raised in place of interview
- cc. 600 000 words
- 42 locations

available online:
- verbatim orthographic transcripts
- ‘normalized’ orthographic transcripts
- morphologically tagged texts


The CORDIAL-SIN project

- Building up the tagged corpus
  - from 1999 to 2007
  - national funding by Fundação para a Ciência e Tecnologia (FCT)
    - CORDIAL-SIN (PRAXIS XXI/P/PLP/13046/1998)
    - CORDIAL-SIN2 (POSI/1999/PLP/33275)
    - Dialect Syntax (POCTIL/46880/2002)
- Corpus syntactic annotation
  - since January 2008
    - DUPLEX (PTDC/LIN/71559/2006)

CORDIAL-SIN Syntactic annotation

- Based on the processes and tools used by the Penn Parsed Corpora of Historical English
  - the Penn-Helsinki Parsed Corpus of Middle English, second edition [PPCME2] (Kroch & Taylor 2000)
  - the Penn-Helsinki Parsed Corpus of Early Modern English [PPCEME] (Kroch, Santorini & Delfs 2004)
  - the Penn Parsed Corpus of Modern British English [PPCMBE] (Kroch & Santorini, under construction)
CORDIAL-SIN Syntactic annotation
Corpora network

- Other corpora using the Penn Corpora annotation scheme
  - The Tycho Brahe Corpus (a parsed corpus of historical Portuguese)
    (http://www.tycho.iel.unicamp.br/~tycho/)
    Charlotte Galves (University of Campinas, Brazil)
  - The Canadian Parsed Corpus of Historical French
    France Martineau (University of Ottawa, Canada)

CORDIAL-SIN Syntactic annotation
The Penn Corpora annotation system
Goals and Principles

- to facilitate automated searches for various constructions of interest
- (!) not to associate every sentence with a correct structural description

Dealing with uncertainty and ambiguity
Avoid controversial decisions
- Omitting undecidable information
  - VP boundaries
  - subtle distinctions (adjectival vs. verbal passives, argument vs. adjunct PPs)
- Using default (and, sometimes, linguistically unmotivated) rules
  - location of wh-traces
  - PP attachment (“when in doubt, attach high”)

CORDIAL-SIN Syntactic annotation
The Penn Corpora annotation system
Goals and Principles

passive vs impersonal se – a default rule option
- unambiguous se

passive
(IP-MAT (VB-P-3P @P@em)
  NP-SE-1 (CL @sei))
  NP-SBJ-1 (D-UM-P uns)
  (N-P panos))
  [. .])
  [SRP24]

impersonal se
(IP-MAT (CONJ mas)
  (NP-SE-1 "exp")
  (FP já)
  (NP-SBJ-1 (CL sei))
  (VP-3S vivia)
  (ADVP (ADV-R melhor))]
  [VPA15]

CORDIAL-SIN Syntactic annotation
The Penn Corpora annotation system
Goals and Principles

passive vs impersonal se – a default rule option
- ambiguous se

passive
(IP-MAT (PP (P @A)
  (NP (D-F-P @as)
    (N-P vezes)))
  (NP-SE-1 (CL @si))
  (NP-SBJ-1 (D-UM-F uma)
    (N quantidade)
    (PP (P @do)
      (NP (DEM @aquilo)))))
  [. .])
  [VPA10]

impersonal se
(IP-MAT (PP (P @A)
  (NP (D-F-P @as)
    (N-P vezes)))
  (NP-SE-1 (CL @si))
  (NP-SBJ-1 (D-UM-F uma)
    (N quantidade)
    (PP (P @do)
      (NP (DEM @aquilo)))))
  [. .])
  [VPA10]

CORDIAL-SIN Syntactic annotation
The Penn Corpora annotation system
Goals and Principles

Quite ‘flat’ structures
- multiple branching nodes
- some word-level nodes (VB, NEG, FP, a.o.)

Although...
- Rich annotation system – marking up of:
  - constituent boundaries
  - phrase and clause dependencies
  - categorial information (e.g. NP, PP, ADVP)
  - grammatical functions (e.g. SBJ, ACC)
  - sentence and clause type (e.g. EXL, QUE)
  - some null constituents
  - some transformational relations
  - other relevant information (LFD, pragmatic markers)
CORDIAL-SIN Syntactic annotation

Example

- syntactic annotation is represented as labeled bracketing
  (depth of indenting corresponds to depth of structural embedding)

- from morphologically tagged texts:
  ```
  e/[CONJ] andávamos/VB-D-1P com/P as/D-F-P redes/N-P @de/P @soldado/N , /que/WPRO são/SR-P-3P mais/ADV-R baixas/ADJ-F-P ... . /[VPA07]
  ```

- to syntactically annotated texts...

CORDIAL-SIN Syntactic annotation

Labels and extended labels

Phrase labels

<table>
<thead>
<tr>
<th>Label</th>
<th>Category (&amp; Function)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>Noun Phrase</td>
</tr>
<tr>
<td>NP-SBJ</td>
<td>Noun Phrase (subject)</td>
</tr>
<tr>
<td>NP-ACC</td>
<td>Noun Phrase (DO, nominal predicate)</td>
</tr>
<tr>
<td>NP-ADV</td>
<td>Noun Phrase (adverbial)</td>
</tr>
<tr>
<td>NP-VOC</td>
<td>Noun Phrase (vocative)</td>
</tr>
<tr>
<td>NP-DAT</td>
<td>Noun Phrase (dative)</td>
</tr>
<tr>
<td>NP-GEN</td>
<td>Noun Phrase (dative of possession)</td>
</tr>
<tr>
<td>PP</td>
<td>Prepositional Phrase</td>
</tr>
<tr>
<td>PP-ACC</td>
<td>Prepositional Phrase (partitive object)</td>
</tr>
</tbody>
</table>

CORDIAL-SIN Syntactic annotation

Labels and extended labels

Phrase labels

<table>
<thead>
<tr>
<th>Label</th>
<th>Category (&amp; Function)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP-MAT</td>
<td>Independent or conjoined declarative IP</td>
</tr>
<tr>
<td>IP-IND</td>
<td>Independent, non-declarative IP</td>
</tr>
<tr>
<td>IP-SUB</td>
<td>Subordinate IP (under CP)</td>
</tr>
<tr>
<td>IP-ADV</td>
<td>Adverbial IP</td>
</tr>
<tr>
<td>IP-INF</td>
<td>Infinitival clause</td>
</tr>
<tr>
<td>IP-GER</td>
<td>Gerund clause</td>
</tr>
<tr>
<td>IP-PPL</td>
<td>Participial clause</td>
</tr>
<tr>
<td>IP-SMC</td>
<td>Small clause</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

CORDIAL-SIN Syntactic annotation

Labels and extended labels

Phrase labels

<table>
<thead>
<tr>
<th>Label</th>
<th>Category (&amp; Function)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP-THT</td>
<td>That clause</td>
</tr>
<tr>
<td>CP-REL</td>
<td>Relative</td>
</tr>
<tr>
<td>CP-FRL</td>
<td>Free Relative</td>
</tr>
<tr>
<td>CP-CLF</td>
<td>Cleft</td>
</tr>
<tr>
<td>CP-ADV</td>
<td>Adverbial clause</td>
</tr>
<tr>
<td>CP-DEG</td>
<td>Degree clause</td>
</tr>
<tr>
<td>CP-CMP</td>
<td>Comparative clause</td>
</tr>
<tr>
<td>CP-EXL</td>
<td>Exclamative</td>
</tr>
<tr>
<td>CP-IMP</td>
<td>Imperative</td>
</tr>
<tr>
<td>CP-QUE</td>
<td>Question</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
CORDIAL-SIN Syntactic annotation
The syntactic annotation process

- automatic parsing of the POS tagged corpus
  version of the Collins/Bikel statistical parser
  (Collins 1999, Bikel 2004), modified for treebank construction by Seth Kulick

- hand correction of the parser output
  CorpusDraw – an editing annotation tool

- make available a parsed version of the corpus that allows to
  retrieve syntactic configurations of interest
  CorpusSearch – a query language for parsed corpora

CORDIAL-SIN Syntactic annotation
The syntactic annotation tools

- CorpusSearch2 (Randall 2005-2007) – a java program that
  • supports research in corpus linguistics
  • is useful both for the construction of syntactically parsed corpora and for searching them
  • runs under any Java-supported operating system (Linux, Mac, Unix, Windows) – requires Java 2, version 1.5 or later
  • expects labelled bracketing (Penn Treebank style)
  • is freely available from http://corpussearch.sourceforge.net/
  • has two main components
    • CorpusDraw
    • CorpusSearch

CORDIAL-SIN Syntactic annotation
The syntactic annotation tools

- CorpusSearch – a dedicated search engine for parsed corpora
Key features:
  • basic search functions are linguistically intuitive
    • (immediately) precedes
    • (immediately) dominates
    • exists
    • hassister
    • isroot
    • logical operators: AND OR NOT
  • end user can custom-define further linguistically relevant search expressions
  • searches can disregard material as necessary
  • the output of CorpusSearch is itself searchable

CORDIAL-SIN Syntactic annotation
The syntactic annotation tools

- CorpusDraw
  Human editing of the parser output

  Editing operations include:
  • changing syntactic tags
  • breaking up run-on sentences or consolidating fragments
  • adding subcategory information
    - NP → NP-ADV
    - CP → CP-CMP
  • changing attachment level
  • adding empty categories
  • adding indices to
    - an antecedent and its trace
    - a constituent in a noncanonical position and the position in which it is interpreted
CORDIAL-SIN Syntactic annotation

The syntactic annotation tools
CorpusSearch – query example

- A simple sample query

  node: IP*
  query: (IP* (Doms NEG))

  - asterisk is a wildcard (IP* matches IP-MAT, IP-SUB, IP-INF, etc.)
  - CorpusSearch searches the corpus for constituents with the label(s) specified in node
  - whenever it finds such a constituent, it checks whether the material in the constituent matches the condition(s) in query
  - matching tokens are recorded in an output file

CORDIAL-SIN Syntactic annotation

At present

- training the Penn parser with European Portuguese data (from July to December)
  - parse a sample file of 10,000 words
  - train the parser with the hand-corrected output
  - repeat the process with new 10,000 words files up to a maximum of 50,000 words
  - run the trained parser on the remaining 550,000 words
- defining the annotation system for European Portuguese (in collaboration with Tycho Brahe and Penn Corpora teams)
- developing the annotator’s manual

CORDIAL-SIN Syntactic annotation

Prospects

- How long does it take to produce a parsed corpus of 600,000 words?
  - mean editing speed: 500 words/hour
    (in language well-known to annotator)
  - 1 annotator
  - annotators can work approx. 4 hours/day or 20 hours/week
  
\[
\begin{align*}
600,000 \text{ words} & \div 500 \text{ words/hour} = 1,200 \text{ hours} \\
1,200 \text{ hours} & \div 20 \text{ hours/week} = 60 \text{ weeks}
\end{align*}
\]

CORDIAL-SIN Syntactic annotation

Adapting the Penn system

- Annotation schemes
  - preserved wherever possible
  - adapted
  - expanded
  - (partly) omitted in specific domains
- Label set
  - preserved
  - new extended labels added

Annotation scheme preserved

that clause complements

(IP-MAT (CONJ e)
  (NP-SBJ (PRO ele))
  (VB-D-3S disse)
  (...)
  (ADV também)
  (...)
  (IP-SUB)
  (CP-THT (C que)
    (NP-SUB (NP-SBJ "pro")
      (VB-D-3P usavam)
      (NP-ACC "*"))
    (NEG não)
    (SR-P-3S é)
    (PP (P por)
      (ADVP (ADV mal))))
  (CP-THT (C que)
    (DEM aquilo))
  (C que)
  (IP-SUB)
  (VPA17))

CP recursion

I (IP-MAT (NP-SBJ (PRO Eu))
  (VB-P-1S sei)
  (CP-THT (C que)
    (DEM aquilo))
  (C que)
  (IP-SUB)
  (VPA15))

know\_sg

that

not

for
evè

[...]

(VPA17)
<table>
<thead>
<tr>
<th>Annotation scheme adapted</th>
<th>Annotation scheme expanded</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>clitic climbing</strong></td>
<td>relative clauses</td>
</tr>
<tr>
<td>(IP-MAT (NP-SBJ &quot;pro&quot;)</td>
<td>preserved</td>
</tr>
<tr>
<td>(ADV também)</td>
<td>woman</td>
</tr>
<tr>
<td>(NP-25 (CL o))</td>
<td>(N senhora)</td>
</tr>
<tr>
<td>(VB-P-1S vou)</td>
<td>who</td>
</tr>
<tr>
<td>(IP-INF (NP-ACC &quot;ICH*-25)</td>
<td>(WPRO que)</td>
</tr>
<tr>
<td>(VB levar))</td>
<td>IP-SUB</td>
</tr>
<tr>
<td>(. .))</td>
<td>(TR-D-3S tinha)</td>
</tr>
<tr>
<td>(NP-25 (CL o))</td>
<td>seven</td>
</tr>
<tr>
<td>(IP-INF (NP-ACC &quot;ICH*-25)</td>
<td>(N-P filhas))</td>
</tr>
<tr>
<td>(VB levar))</td>
<td>[. .])</td>
</tr>
<tr>
<td>(. .))</td>
<td>(PFT25)</td>
</tr>
<tr>
<td>(IP-MAT (NP-SBJ &quot;pro&quot;)</td>
<td>preserved</td>
</tr>
<tr>
<td>(ADV também)</td>
<td>after</td>
</tr>
<tr>
<td>(NP-25 (CL o))</td>
<td>there</td>
</tr>
<tr>
<td>(IP-INF (NP-ACC &quot;ICH*-25)</td>
<td>(VB-D-3S foi)</td>
</tr>
<tr>
<td>(VB levar))</td>
<td>(NP-SBJ (D o))</td>
</tr>
<tr>
<td>(. .))</td>
<td>local</td>
</tr>
<tr>
<td>(IP-MAT (NP-SBJ &quot;pro&quot;)</td>
<td>(WPP-167 (P [CODE {em}]))</td>
</tr>
<tr>
<td>(NP-25 (CL o))</td>
<td>(NP-SBJ (WPRO que))</td>
</tr>
<tr>
<td>(IP-INF (NP-ACC &quot;ICH*-25)</td>
<td>(IP-SUB (PP <em>T</em>-167))</td>
</tr>
<tr>
<td>(VB levar))</td>
<td>(NP-SBJ (PRO eu))</td>
</tr>
<tr>
<td>(. .))</td>
<td>(VB-D-1S andava))</td>
</tr>
<tr>
<td>(IP-MAT (NP-SBJ &quot;pro&quot;)</td>
<td>preserved</td>
</tr>
<tr>
<td>(NP-25 (CL o))</td>
<td>after</td>
</tr>
<tr>
<td>(VB-P-1S vou)</td>
<td>there</td>
</tr>
<tr>
<td>(IP-INF (NP-ACC &quot;ICH*-25)</td>
<td>(VB-D-3S foi)</td>
</tr>
<tr>
<td>(VB levar))</td>
<td>(NP-SBJ (D o))</td>
</tr>
<tr>
<td>(. .))</td>
<td>local</td>
</tr>
<tr>
<td>(IP-MAT (NP-SBJ &quot;pro&quot;)</td>
<td>(WPP-167 (P [CODE {em}]))</td>
</tr>
<tr>
<td>(NP-25 (CL o))</td>
<td>(NP-SBJ (WPRO que))</td>
</tr>
<tr>
<td>(IP-INF (NP-ACC &quot;ICH*-25)</td>
<td>(IP-SUB (PP <em>T</em>-167))</td>
</tr>
<tr>
<td>(VB levar))</td>
<td>(NP-SBJ (PRO eu))</td>
</tr>
<tr>
<td>(. .))</td>
<td>(VB-D-1S andava))</td>
</tr>
<tr>
<td>(IP-MAT (NP-SBJ &quot;pro&quot;)</td>
<td>preserved</td>
</tr>
<tr>
<td>(NP-25 (CL o))</td>
<td>after</td>
</tr>
<tr>
<td>(VB-P-1S vou)</td>
<td>there</td>
</tr>
<tr>
<td>(IP-INF (NP-ACC &quot;ICH*-25)</td>
<td>(VB-D-3S foi)</td>
</tr>
<tr>
<td>(VB levar))</td>
<td>(NP-SBJ (D o))</td>
</tr>
<tr>
<td>(. .))</td>
<td>local</td>
</tr>
<tr>
<td>(IP-MAT (NP-SBJ &quot;pro&quot;)</td>
<td>(WPP-167 (P [CODE {em}]))</td>
</tr>
<tr>
<td>(NP-25 (CL o))</td>
<td>(NP-SBJ (WPRO que))</td>
</tr>
<tr>
<td>(IP-INF (NP-ACC &quot;ICH*-25)</td>
<td>(IP-SUB (PP <em>T</em>-167))</td>
</tr>
<tr>
<td>(VB levar))</td>
<td>(NP-SBJ (PRO eu))</td>
</tr>
<tr>
<td>(. .))</td>
<td>(VB-D-1S andava))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation scheme expanded</th>
<th>topic constructions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>resumptive relative</strong></td>
<td></td>
</tr>
<tr>
<td>(NP (D-F Essa)</td>
<td>Penn system: left</td>
</tr>
<tr>
<td>(ADJ-G tal)</td>
<td>dislocated noun-</td>
</tr>
<tr>
<td>(N feiticeira)</td>
<td>phrases (NP*-L-F-U)</td>
</tr>
<tr>
<td>(WNP (WPRO que))</td>
<td>(IP-MAT (NP-LFD</td>
</tr>
<tr>
<td>(IP-SUB (NP-SBJ &quot;pro&quot;)</td>
<td>(NPR Mary))</td>
</tr>
<tr>
<td>(NP-DAT-RSP-28 (CL lhe))</td>
<td>(NP-SBJ-RSP (PRO</td>
</tr>
<tr>
<td>(VB-P-3P chamam)</td>
<td>she (ADVP-TMP (ADV</td>
</tr>
<tr>
<td>(IP-SMC (PP-SBJ &quot;ICH*-28)</td>
<td>(ADVP-TMP (ADV</td>
</tr>
<tr>
<td>(NP-ACC (NP(D-F a)</td>
<td>always)) (VBP wins)</td>
</tr>
<tr>
<td>(N pata-roxa)</td>
<td>(. .))</td>
</tr>
<tr>
<td>[. .])</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation scheme expanded</th>
<th>topic constructions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>free relative</strong></td>
<td></td>
</tr>
<tr>
<td>preserved</td>
<td>(IP-MAT (CONJ e)</td>
</tr>
<tr>
<td></td>
<td>(NP-SBJ &quot;exp&quot;)</td>
</tr>
<tr>
<td></td>
<td>(HV-P-3S há)</td>
</tr>
<tr>
<td>who</td>
<td>(NP-ACC (CP-FRL</td>
</tr>
<tr>
<td>throw the fishing net</td>
<td>(WNP-1 (WPRO quen)</td>
</tr>
<tr>
<td>from the stem</td>
<td>(IP-SUB (NP-SBJ</td>
</tr>
<tr>
<td></td>
<td>&quot;T*-1)</td>
</tr>
<tr>
<td></td>
<td>(VB-SP-3S largue)</td>
</tr>
<tr>
<td></td>
<td>(NP-ACC (D-F a)</td>
</tr>
<tr>
<td></td>
<td>(N rede))</td>
</tr>
<tr>
<td></td>
<td>(PP (P por)</td>
</tr>
<tr>
<td></td>
<td>(NP (D-F a)</td>
</tr>
<tr>
<td></td>
<td>(N popa))))</td>
</tr>
<tr>
<td>[. .])</td>
<td>(VPA05)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation scheme expanded</th>
<th>topic constructions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>‘chopping’ relative</strong></td>
<td></td>
</tr>
<tr>
<td>preserved</td>
<td>(IP-MAT (CONJ E)</td>
</tr>
<tr>
<td></td>
<td>after (ADVP (ADV</td>
</tr>
<tr>
<td></td>
<td>depois))</td>
</tr>
<tr>
<td></td>
<td>there (ADVP (ADV</td>
</tr>
<tr>
<td></td>
<td>lá))</td>
</tr>
<tr>
<td></td>
<td>went (VB-D-3S foi)</td>
</tr>
<tr>
<td>the</td>
<td>(NP-SBJ (D o))</td>
</tr>
<tr>
<td>in</td>
<td>(N barco)</td>
</tr>
<tr>
<td>which</td>
<td>(WPP-167 (P [CODE</td>
</tr>
<tr>
<td></td>
<td>(PP <em>T</em>-167))</td>
</tr>
<tr>
<td>f</td>
<td>(IP-SUB (NP-SBJ</td>
</tr>
<tr>
<td>was</td>
<td>(PRO eu))</td>
</tr>
<tr>
<td>(. .))</td>
<td>(VB-D-1S andava))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation scheme expanded</th>
<th>topic constructions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>topic constructions</strong></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>(IP-MAT (NP-LFD</td>
</tr>
<tr>
<td>that</td>
<td>(NPR Mary))</td>
</tr>
<tr>
<td>such</td>
<td>(NP-SBJ-RSP (PRO</td>
</tr>
<tr>
<td>that</td>
<td>she (ADVP-TMP (ADV</td>
</tr>
<tr>
<td>fish name</td>
<td>always)) (VBP wins)</td>
</tr>
<tr>
<td>that</td>
<td>(. .))</td>
</tr>
</tbody>
</table>
(1) Mas esse peixe, já uma pessoa às vezes não o conhece.
(2) Nós tocavamos para aí uns quinhentos mil réis, ou setecentos ou oitocentos.
(3) E o lavagante já não há, também.
(4) A pesca, olhe, larga-se a rede por a borda.
Adapting the label set
New extended labels

- **IP-ANS** answers to YN and wh-questions
- **IP-POL** reinforcement of assertion
- **CP-QUE-TAG** question-tag
- **XP/IP/CP-PRG** pragmatic (generic)

In connection with pragmatics

---

**IP-ANS**

(Code INQ1 E trazia-as já feitas?)

I: And did you bring them already done

Inf: brought

(Code INF)

(Code INQ2 E tinha umas coisas para respirar?)

I: And did it have any thing to breathe through?

Inf: no [<no madam>]

(Code INF)

IP-POL

(IP-MAT (CONJ Porque)
(NP-SBJ "pro")
(SR-P-3S é)
(ADJP (ADJ branco))
(.
.)

(IP-ANS (ADVP (ADV-NEG Não_senhora))
(.
.))

(CODE INQ1 E trazia-as já feitas?)

I: And did you bring them already done

Inf: brought

(Code INF)

IP-POL

(IP-MAT (NP-SBJ "pro")
(NEG não)
(SR-P-3S é)
(PP (P @de)
(NP (D @este)
(N género)))
(. ,)

(IP-POL (NEG não)
(. .) [PFT40])

IP-POL

(IP-MAT (NP-SBJ "pro")
(NEG não)
(SR-P-3S é)
(PP (P @de)
(NP (D @este)
(N género)))
(, ,)

(IP-POL (NEG não)
(. .) [PFT40])

CP-QUE-TAG

(IP-MAT (NP-SBJ "pro")
(TR-P-3P têm)
(NP ACC (D-UM um)
(N bocadinho)
(PP (P de)
(NP (N ferrugem)))
(. ,)

(IP-POL (NEG não)
(TR-P-3P têm)
.
.) [PFT40]

XP/IP/IP(…) - PRG

(IP-MAT (NP-SBJ "pro")
(ADVP-PRG (ADV bem))
(.
.)

(IP-POL (VB-P-3P fazem)
(ADV (ADV assim))
(.
.)) [PFT40]

(IP-POL (NP-ACC)
(NEG não)
(D-UM um)
(N bocadinho)
(PP (P de)
(NP (N ferrugem)))
(, ,)

(IP-POL (NEG não)
(TRA-P-3P têm)
.
.) [PFT40]
in the past

dried

the mackerel

wants to say
the "sorelo"

look

the potatoes

grew

because
that man
is of the
religion of
the truth

know

(you know)