High Precision Analysis of NPs with a Deep Processing Grammar
STEP 2008

António Branco\textsuperscript{1}    Francisco Costa\textsuperscript{1}

\textsuperscript{1}Universidade de Lisboa

September 23, 2008
Outline

1 Introduction
2 Foundations
3 Design Features
4 Coverage
5 Noun Phrases
6 Evaluation
7 Applications and Further Work
8 Final Remarks
Introduction

- LXGram
  - Hand-crafted precision grammar for Portuguese
  - Deep linguistic processing
- Head-Driven Phrase Structure Grammar (HPSG) [Pollard and Sag, 1994]
- Minimal Recursion Semantics (MRS) [Copestake et al., 2005]
- Linguistic Knowledge Builder (LKB) [Copestake, 2002]
- PET parser [Callmeier, 2000]
- DELPH-IN consortium
- HPSGs for other languages, based on the same formalisms and tools:
  - English [Copestake and Flickinger, 2000]
  - German [Müller and Kasper, 2000]
  - Japanese [Siegel and Bender, 2002], ...
Head-Driven Phrase Structure Grammar

- Declarative formalism
- Feature structures
- Type system
- Unification

```
\[
\begin{array}{c}
type1 \\
\text{FEATURE1} \\
type2 \\
\text{FEATURE2} \\
\end{array} \\
\begin{array}{c}
type3 \\
\text{FEATURE3} \quad \text{unification tag} \\
\text{FEATURE4} \quad \text{unification tag} \\
\end{array}
\]
```
Minimal Recursion Semantics

- Underspecification of scope
- Conjunction represented indirectly
- Abstraction from many grammar details
- Interface to applications

\[
\begin{align*}
mrs & \quad LTOP \quad h_1 \ h \\
INDEX & \quad e_2 \\
RELS & \quad \langle \langle todo_q_rel \rangle, \langle equipa_n_rel \rangle, \langle poder_v_rel \rangle, \langle vencer_v_rel \rangle \rangle \\
HCONS & \quad \langle \langle qeq \rangle, \langle qeq \rangle, \langle qeq \rangle \rangle
\end{align*}
\]

\[
\text{all}(x, \text{team}(x), \text{possible}(\text{win}(x)))
\]

\[
\text{possible}(\text{all}(x, \text{team}(x), \text{win}(x)))
\]
Design Features

- Bidirectionality
  - Parsing and generation

- Precision
  - Ungrammatical sentences are unparsable

- Broad coverage
  - Some support for linguistic variation (Portugal, Brazil)

- Efficiency
  - PET is an efficient engine
  - Grammar engineering techniques

- Robustness
  - Unknown word mechanisms
  - Robustness rules (e.g. sentences with no verb)

- Availability
  http://nlxgroup.di.fc.ul.pt/lxgram
Coverage

- Previous work on guessing information about unknown words from the output of shallow tools

- Size of the grammar:
  - 24484 lines of code (including comments)
  - 53 syntax rules
  - 40 lexical rules
  - 3154 total types
  - 414 types for lexical items
  - 2718 hand-built lexical entries
Coverage

- Implemented phenomena:
  - Declarative sentences, imperatives and yes/no questions
  - Some subcategorization frames of verbs, nouns and adjectives
  - Comparative constructions
  - Noun phrase structure
  - Verbal modification
  - Relative clauses
  - Null subjects and objects
  - Floated quantifiers
1. Introduction
2. Foundations
3. Design Features
4. Coverage
5. Noun Phrases
6. Evaluation
7. Applications and Further Work
8. Final Remarks
Noun Phrases — Examples

Determiners and predeterminers

<table>
<thead>
<tr>
<th>Pt</th>
<th>gloss</th>
<th>En</th>
</tr>
</thead>
<tbody>
<tr>
<td>os problemas</td>
<td>os problems</td>
<td>the problems</td>
</tr>
<tr>
<td>todos *(os) problemas</td>
<td>all the problems</td>
<td>all (the) problems</td>
</tr>
<tr>
<td>estes (*os) problemas</td>
<td>these the problems</td>
<td>these problems</td>
</tr>
<tr>
<td>todos estes problemas</td>
<td>all these problems</td>
<td>all these problems</td>
</tr>
<tr>
<td>os problemas estes</td>
<td>the problems these</td>
<td>these problems</td>
</tr>
</tbody>
</table>
Noun Phrases — Examples

- Determiners and predeterminers

“estes problemas”, “os problemas estes” ("these problems")
Noun Phrases — Examples

- Determiners and predeterminers
- Possessives

Pt  a sua guitarra
gloss  the his guitar
En  his guitar

\[
\begin{array}{c}
\text{mrs} \\
\text{LTOP} \ h  h_1 \\
\text{INDEX} \ x  x_2 \\
\text{RELS} \\
\text{LBL} \ h_1 \\
\text{ARG0} \ x_2 \\
\text{RSTR} \ h_4 \\
\text{BODY} \ h_3 \\
\text{arg} \\
\text{HARG} \ h_4 \\
\text{LARG} \ h_5 \\
\text{LBL} \ h_8 \\
\text{ARG0} \ h_6 \\
\text{RSTR} \ h_9 \\
\text{BODY} \ h_{10} \\
\text{arg} \\
\text{HARG} \ h_9 \\
\text{LARG} \ h_{11} \\
\text{LBL} \ h_{11} \\
\text{ARG0} \ h_7 \\
\text{RSTR} \ h_5 \\
\text{BODY} \ h_7 \\
\end{array}
\]
Noun Phrases — Examples

- Determiners and predeterminers
- Possessives

Pt
a sua filha

Gloss
the his daughter

En
his daughter

\[
\begin{align*}
\text{Pt} & \quad \text{a sua filha} \\
\text{Gloss} & \quad \text{the his daughter} \\
\text{En} & \quad \text{his daughter}
\end{align*}
\]
Noun Phrases — Examples

- Missing nouns

Pt   um com água
  gloss one with water
En   one with water
Noun Phrases — Examples

- Missing nouns
- Intersective and non-intersective adjectives

Pt | um carro francês
---|-------------------
gloss | a car French
En | a French car

\[ \lambda P. \text{um}(x, \text{carro}(x) \land \text{frances}(x), P(x)) \]
Noun Phrases — Examples

- Missing nouns
- Intersective and non-intersective adjectives

Pt | um | antigo | presidente

gloss | a former president

En | a former president

\( \lambda P. \text{um}(x, \text{antigo}(\text{presidente}(x)), P(x)) \)
Noun Phrases — Examples

- Floated quantifiers

**Pt** Todos os preços vão subir.
**gloss** all the prices will go up

**En** All prices will go up.

**Pt** Os preços todos vão subir.
**gloss** the prices all will go up

**Pt** Os preços vão todos subir.
**gloss** the prices will all go up

**Pt** Os preços vão subir todos.
**gloss** the prices will go up all

\[ all(x, \text{price}(x), \text{will}(\text{go\_up}(x))) \]
Noun Phrases — Examples

- Floated quantifiers

```
[SUBJ ⟨⟩]

[SUBJ ⟨1⟩]

Os
the

preços
prices

vão
will

[SUBJ ⟨1⟩ NP⟨e,t⟩]

todos
all

[subir
go up]
```
### Noun Phrases — Examples

- Floated quantifiers
- Scope of adjectives and relative clauses

\[
\lambda P. a(x, possible(doctor(x) \land chinese(x)), P(x))
\]

- um possível médico chinês
- possible doctor Chinese
- a possible Chinese doctor

\[
\lambda P. a(x, possible(doctor(x)) \land chinese(x), P(x))
\]

- um possível médico que é chinês
- possible doctor who is Chinese
- a possible doctor who is Chinese
Noun Phrases — Examples

- Floated quantifiers
- Scope of adjectives and relative clauses
1 Introduction
2 Foundations
3 Design Features
4 Coverage
5 Noun Phrases
6 Evaluation
7 Applications and Further Work
8 Final Remarks
Evaluation

- Preliminary evaluation
  - Newspaper text, 145 sentences (avg 22 words)
  - Unknown words approximated via a PoS tagger and a morphological analyzer
  - 13.1% of sentences parsed

- Older experiment [Branco and Costa, 2007]
  - 180K short sentences (5–9 words) from newspaper text
  - Similar approach to unknown words
  - 26% coverage

- Test suites
  - 422 positive examples, 429 negative examples (851 total)

- Comparison
  - ERG parses 57% of the sentences in the BNC for which it has full lexical span (32% of total; 10–20 words/sentence)
  [Baldwin et al., 2005]
1 Introduction
2 Foundations
3 Design Features
4 Coverage
5 Noun Phrases
6 Evaluation
7 Applications and Further Work
8 Final Remarks
Applications and Further Work

■ Previous work:
  ■ Automatically discriminate between texts written in European Portuguese and Brazilian Portuguese

■ Current work:
  ■ Integration of an external lexicon (14K nouns, 3K verbs and 3K adjectives so far)
  ■ Treebanking
  ■ Stochastic disambiguation
  ■ Grammar extension
  ■ Integration in a question answering system
Final Remarks

- LXGram: hand-crafted precision grammar for Portuguese
- Parsing and generation
- HPSG, MRS
- Heavy focus on NP phenomena until now
- More clausal level phenomena being implemented now
References


