High Precision Analysis of NPs with a Deep Processing Grammar

STEP 2008

António Branco[1] Francisco Costa[1]

[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

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] , . . .

1 Introduction

2 Foundations

3 Design Features

4 Coverage

5 Noun Phrases

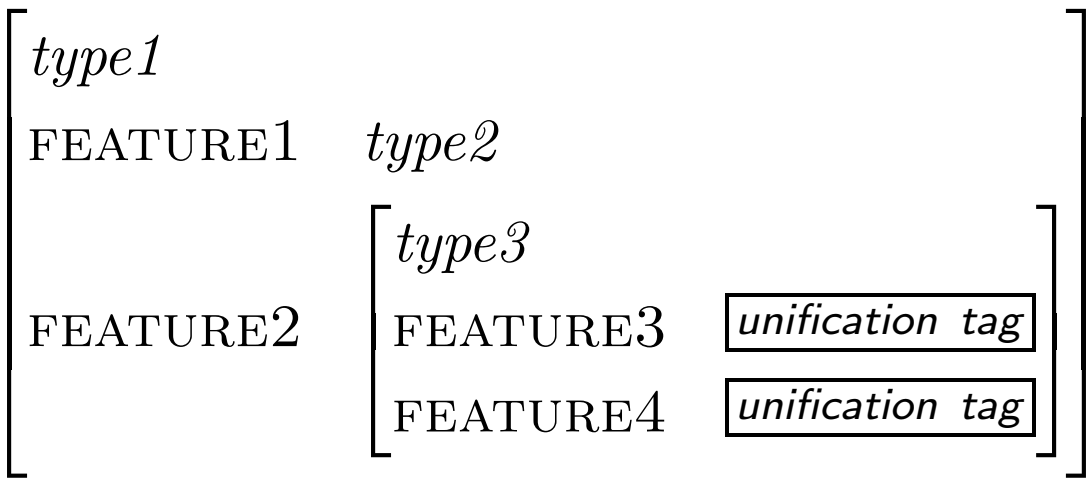
6 Evaluation

7 Applications and Further Work

8 Final Remarks

Head-Driven Phrase Structure Grammar

- Declarative formalism
- Feature structures
- Type system
- Unification



Minimal Recursion Semantics

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

$$\begin{bmatrix} \textit{mrs} \\ \text{LTOP} & \boxed{h1}\ h \\ \text{INDEX} & \boxed{e2}\ e \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{_todo_q_rel} \\ \text{LBL} & \boxed{h3}\ h \\ \text{ARG0} & \boxed{x4}\ x \\ \text{RSTR} & \boxed{h6}\ h \\ \text{BODY} & \boxed{h5}\ h \end{bmatrix}, \begin{bmatrix} \textit{_equipa_n_rel} \\ \text{LBL} & \boxed{h7}\ h \\ \text{ARG0} & \boxed{x4} \end{bmatrix}, \begin{bmatrix} \textit{_poder_v_rel} \\ \text{LBL} & \boxed{h8} \\ \text{ARG0} & \boxed{e2} \\ \text{ARG1} & \boxed{h9}\ h \end{bmatrix}, \begin{bmatrix} \textit{_vencer_v_rel} \\ \text{LBL} & \boxed{h10} \\ \text{ARG0} & \boxed{e11} \\ \text{ARG1} & \boxed{x4} \end{bmatrix} \right\rangle \\ \text{HCONS} & \left\langle \begin{bmatrix} \textit{qeq} \\ \text{HARG} & \boxed{h1} \\ \text{LARG} & \boxed{h8} \end{bmatrix}, \begin{bmatrix} \textit{qeq} \\ \text{HARG} & \boxed{h6} \\ \text{LARG} & \boxed{h7} \end{bmatrix}, \begin{bmatrix} \textit{qeq} \\ \text{HARG} & \boxed{h9} \\ \text{LARG} & \boxed{h10} \end{bmatrix} \right\rangle \end{bmatrix}$$

$$\textit{all}(x, \textit{team}(x), \textit{possible}(\textit{win}(x)))$$

$$\textit{possible}(\textit{all}(x, \textit{team}(x), \textit{win}(x)))$$

1 Introduction

2 Foundations

3 Design Features

4 Coverage

5 Noun Phrases

6 Evaluation

7 Applications and Further Work

8 Final Remarks

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`

1 Introduction

2 Foundations

3 Design Features

4 Coverage

5 Noun Phrases

6 Evaluation

7 Applications and Further Work

8 Final Remarks

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

Pt	os	problemas	
gloss	os	problems	
En	the problems		

Pt	todos	*(os)	problemas
gloss	all	the	problems
En	all (the) problems		

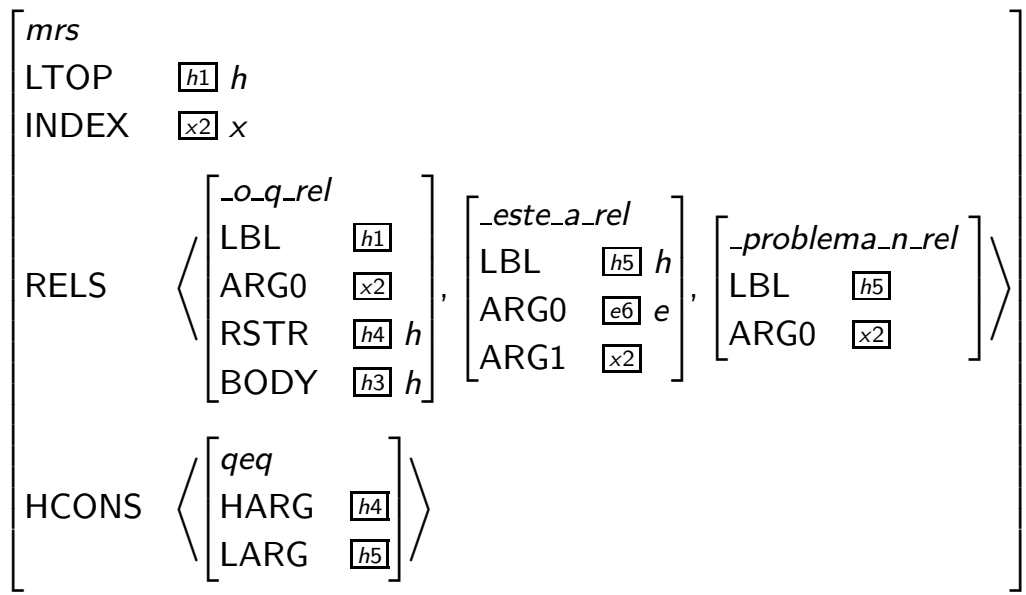
Pt	todos	estes	problemas
gloss	all	these	problems
En	all these problems		

Pt	estes	(*os)	problemas
gloss	these	the	problems
En	these problems		

Pt	os	problemas	estes
gloss	the	problems	these
En	these problems		

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		

$$
\begin{bmatrix}
\textit{mrs} \\
\text{LTOP} & \boxed{h1}\ h \\
\text{INDEX} & \boxed{x2}\ x \\
\text{RELS} & \left\langle
\begin{bmatrix} \textit{_o_q_rel} \\ \text{LBL} & \boxed{h1} \\ \text{ARG0} & \boxed{x2} \\ \text{RSTR} & \boxed{h4}\ h \\ \text{BODY} & \boxed{h3}\ h \end{bmatrix},
\begin{bmatrix} \textit{possessive_a_rel} \\ \text{LBL} & \boxed{h5}\ h \\ \text{ARG0} & \boxed{e6}\ e \\ \text{ARG1} & \boxed{x2} \\ \text{ARG2} & \boxed{x7}\ x \end{bmatrix},
\begin{bmatrix} \textit{pronoun_q_rel} \\ \text{LBL} & \boxed{h8}\ h \\ \text{ARG0} & \boxed{x7} \\ \text{RSTR} & \boxed{h9}\ h \\ \text{BODY} & \boxed{h10}\ h \end{bmatrix},
\begin{bmatrix} \textit{pronoun_n_rel} \\ \text{LBL} & \boxed{h11}\ h \\ \text{ARG0} & \boxed{x7} \\ \text{ARG1} & \boxed{r12}\ r \end{bmatrix},
\begin{bmatrix} \textit{_guitarra_n_rel} \\ \text{LBL} & \boxed{h5} \\ \text{ARG0} & \boxed{x2} \end{bmatrix}
\right\rangle \\
\text{HCONS} & \left\langle
\begin{bmatrix} \textit{qeq} \\ \text{HARG} & \boxed{h4} \\ \text{LARG} & \boxed{h5} \end{bmatrix},
\begin{bmatrix} \textit{qeq} \\ \text{HARG} & \boxed{h9} \\ \text{LARG} & \boxed{h11} \end{bmatrix}
\right\rangle
\end{bmatrix}
$$

Noun Phrases — Examples

- Determiners and predeterminers
- Possessives

Pt	a	sua	filha
gloss	the	his	daughter
En	his daughter		

$$
\left[\begin{array}{ll}
\textit{mrs} & \\
\text{LTOP} & \boxed{h1}\ h \\
\text{INDEX} & \boxed{x2}\ x \\
\text{RELS} & \left\langle \left[\begin{array}{ll} \textit{_o_q_rel} & \\ \text{LBL} & \boxed{h1} \\ \text{ARG0} & \boxed{x2} \\ \text{RSTR} & \boxed{h4}\ h \\ \text{BODY} & \boxed{h3}\ h \end{array}\right], \left[\begin{array}{ll} \textit{pronoun_q_rel} & \\ \text{LBL} & \boxed{h5}\ h \\ \text{ARG0} & \boxed{x6}\ x \\ \text{RSTR} & \boxed{h7}\ h \\ \text{BODY} & \boxed{h8}\ h \end{array}\right], \left[\begin{array}{ll} \textit{pronoun_n_rel} & \\ \text{LBL} & \boxed{h9}\ h \\ \text{ARG0} & \boxed{x6} \\ \text{ARG1} & \boxed{r10}\ r \end{array}\right], \left[\begin{array}{ll} \textit{_filho_n_-de-_rel} & \\ \text{LBL} & \boxed{h11}\ h \\ \text{ARG0} & \boxed{x2} \\ \text{ARG1} & \boxed{x6} \end{array}\right] \right\rangle \\
\text{HCONS} & \left\langle \left[\begin{array}{ll} \textit{qeq} & \\ \text{HARG} & \boxed{h4} \\ \text{LARG} & \boxed{h11} \end{array}\right], \left[\begin{array}{ll} \textit{qeq} & \\ \text{HARG} & \boxed{h7} \\ \text{LARG} & \boxed{h9} \end{array}\right] \right\rangle
\end{array}\right]
$$

Noun Phrases — Examples

- Missing nouns

Pt	um	com	água
gloss	one	with	water
En	one with water		

$$
\begin{bmatrix}
\textit{mrs} \\
\text{LTOP} & \boxed{h1}\ h \\
\text{INDEX} & \boxed{x2}\ x \\
\text{RELS} & \left\langle
\begin{bmatrix}
\textit{_um_q_rel} \\
\text{LBL} & \boxed{h1} \\
\text{ARG0} & \boxed{x2} \\
\text{RSTR} & \boxed{h4}\ h \\
\text{BODY} & \boxed{h3}\ h
\end{bmatrix},
\begin{bmatrix}
\textit{ellipsis-or-generic_n_1_rel} \\
\text{LBL} & \boxed{h5}\ h \\
\text{ARG0} & \boxed{x2}
\end{bmatrix},
\begin{bmatrix}
\textit{_com_p_rel} \\
\text{LBL} & \boxed{h5} \\
\text{ARG0} & \boxed{e6}\ e \\
\text{ARG1} & \boxed{x2} \\
\text{ARG2} & \boxed{x7}\ x
\end{bmatrix},
\begin{bmatrix}
\textit{udef_q_rel} \\
\text{LBL} & \boxed{h8}\ h \\
\text{ARG0} & \boxed{x7} \\
\text{RSTR} & \boxed{h10}\ h \\
\text{BODY} & \boxed{h9}\ h
\end{bmatrix},
\begin{bmatrix}
\textit{_água_n_rel} \\
\text{LBL} & \boxed{h11}\ h \\
\text{ARG0} & \boxed{x7}
\end{bmatrix}
\right\rangle \\
\text{HCONS} & \left\langle
\begin{bmatrix}
\textit{qeq} \\
\text{HARG} & \boxed{h4} \\
\text{LARG} & \boxed{h5}
\end{bmatrix},
\begin{bmatrix}
\textit{qeq} \\
\text{HARG} & \boxed{h10} \\
\text{LARG} & \boxed{h11}
\end{bmatrix}
\right\rangle
\end{bmatrix}
$$

Noun Phrases — Examples

- Missing nouns
- Intersective and non-intersective adjectives

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

$$
\begin{bmatrix}
\textit{mrs} \\
\text{LTOP} & \boxed{h1}\ h \\
\text{INDEX} & \boxed{x2}\ x \\
\text{RELS} & \left\langle
\begin{bmatrix}
\textit{_um_q_rel} \\
\text{LBL} & \boxed{h1} \\
\text{ARG0} & \boxed{x2} \\
\text{RSTR} & \boxed{h4}\ h \\
\text{BODY} & \boxed{h3}\ h
\end{bmatrix},
\begin{bmatrix}
\textit{_carro_n_rel} \\
\text{LBL} & \boxed{h5}\ h \\
\text{ARG0} & \boxed{x2}
\end{bmatrix},
\begin{bmatrix}
\textit{_francês_a_rel} \\
\text{LBL} & \boxed{h5} \\
\text{ARG0} & \boxed{e6}\ e \\
\text{ARG1} & \boxed{x2}
\end{bmatrix}
\right\rangle \\
\text{HCONS} & \left\langle
\begin{bmatrix}
\textit{qeq} \\
\text{HARG} & \boxed{h4} \\
\text{LARG} & \boxed{h5}
\end{bmatrix}
\right\rangle
\end{bmatrix}
$$

$\lambda P.um(x, carro(x) \wedge frances(x), P(x))$

Noun Phrases — Examples

- Missing nouns
- Intersective and non-intersective adjectives

Pt	um	antigo	presidente
gloss	a	former	president
En	a former president		

$$\begin{bmatrix} mrs \\ \text{LTOP} & \boxed{h1}\ h \\ \text{INDEX} & \boxed{x2}\ x \\ \text{RELS} & \left\langle \begin{bmatrix} _um_q_rel \\ \text{LBL} & \boxed{h1} \\ \text{ARG0} & \boxed{x2} \\ \text{RSTR} & \boxed{h4}\ h \\ \text{BODY} & \boxed{h3}\ h \end{bmatrix}, \begin{bmatrix} _antigo_a_rel \\ \text{LBL} & \boxed{h5}\ h \\ \text{ARG0} & \boxed{e6}\ e \\ \text{ARG1} & \boxed{h7}\ h \end{bmatrix}, \begin{bmatrix} _presidente_n_-de-_rel \\ \text{LBL} & \boxed{h8}\ h \\ \text{ARG0} & \boxed{x2} \\ \text{ARG1} & \boxed{r9}\ r \end{bmatrix} \right\rangle \\ \text{HCONS} & \left\langle \begin{bmatrix} qeq \\ \text{HARG} & \boxed{h4} \\ \text{LARG} & \boxed{h5} \end{bmatrix}, \begin{bmatrix} qeq \\ \text{HARG} & \boxed{h7} \\ \text{LARG} & \boxed{h8} \end{bmatrix} \right\rangle \end{bmatrix}$$

$\lambda P.um(x, antigo(presidente(x)), P(x))$

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

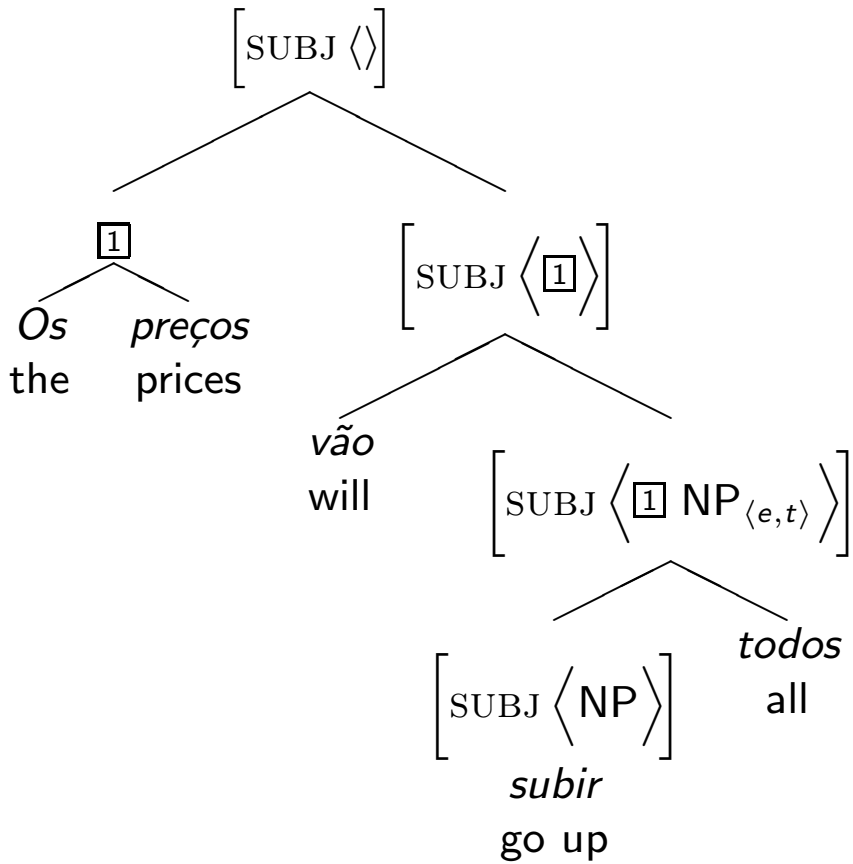
Pt	Os	preços	vão	todos	subir.
gloss	the	prices	will	all	go up

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

$all(x, price(x), will(go_up(x)))$

Noun Phrases — Examples

- Floated quantifiers



Noun Phrases — Examples

- Floated quantifiers
- Scope of adjectives and relative clauses

Pt	um	possível	médico	chinês
gloss	a	possible	doctor	Chinese
En	a possible Chinese doctor			

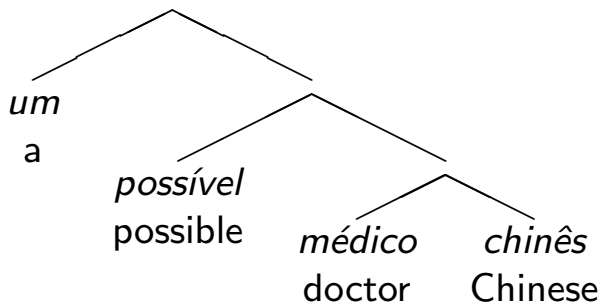
$\lambda P.a(x, possible(doctor(x) \wedge chinese(x)), P(x))$

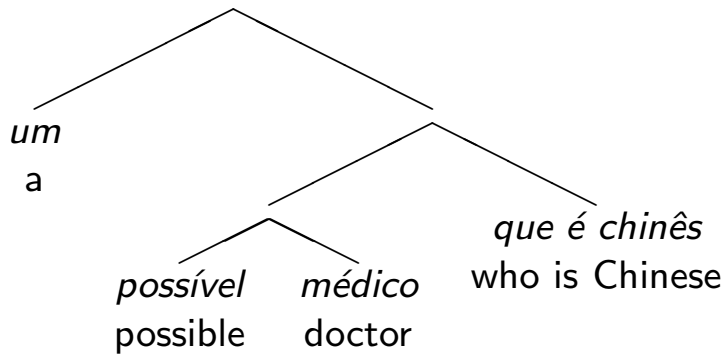
Pt	um	possível	médico	que	é	chinês
gloss	a	possible	doctor	who	is	Chinese
En	a possible doctor who is Chinese					

$\lambda P.a(x, possible(doctor(x)) \wedge chinese(x), P(x))$

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

1 Introduction

2 Foundations

3 Design Features

4 Coverage

5 Noun Phrases

6 Evaluation

7 Applications and Further Work

8 Final Remarks

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

Baldwin, T., Beavers, J., Bender, E. M., Flickinger, D., Kim, A., and Oepen, S., 2005. Beauty and the beast: What running a broad-coverage precision grammar over the bnc taught us about the grammar — and the corpus. In Kepser, S. and Reis, M., editors, *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, pages 49–70, Berlin. Mouton de Gruyter.

Branco, A. and Costa, F., 2007. Identification and handling of dialectal variation with a single grammar. In Dirix, P., Schuurman, I., Vandeghinste, V., and Eynde, F. V., editors, *Proceedings of the 17th Meeting of Computational Linguistics in the Netherlands (CLIN17)*, pages 5–19, Utrecht. LOT.

Callmeier, U., 2000. PET — A platform for experimentation with efficient HPSG processing techniques. *Natural Language Engineering*, 6(1):99–108. (Special Issue on Efficient Processing with HPSG).

Copestake, A. and Flickinger, D., 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece.

Copestake, A., Flickinger, D., Sag, I. A., and Pollard, C., 2005. Minimal Recursion Semantics: An introduction. *Journal of Research on Language and Computation*, 3(2–3):281–332.

Copestake, A., 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford.

Müller, S. and Kasper, W., 2000. HPSG analysis of German. In Wahlster, W., editor, *Verbmobil: Foundations of Speech-to-Speech Translation*, pages 238–253. Springer-Verlag, Berlin Heidelberg New York, Artificial Intelligence edition.

Pollard, C. and Sag, I., 1994. *Head-Driven Phrase Structure Grammar*. Chicago University Press and CSLI Publications, Stanford.

Siegel, M. and Bender, E. M., 2002. Efficient deep processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization. Coling 2002 Post-Conference Workshop*, pages 31–38, Taipei, Taiwan.