PropBank from HPSG Treebank

António Branco

University of Lisbon

CINTIL data banks

❑ HPSG bank

  ▪ kernel linguistic data bank

  ▪ source of linguistic info to...

❑ ... extract "views"

  ▪ CINTIL treebank

  ▪ CINTIL dependency bank

  ▪ CINTIL logical form bank

  ▪ ...

Online browsing

❑ CINTIL Treebank Searcher

- ▪ http://cintiltreebank.di.fc.ul.pt

- ▪ in development: intermitent service in the near future

António Branco
Delph-in'2009, Barcelona, July 21

How about PropBanks?

- ❑ Penn tagset (Palmer et al)
  - ▪ ARG1, ..., ARG*n*, M
  - ▪ LOC, EXT, ADV, CAU, TMP, PNC, MNR, DIR, POV, PRD, NULL

- ❑ **A semantic role view?**
  - ▪ expand grammar with feature for SRole
  - ▪ run the grammar
  - ▪ select the correct parses
  - ▪ extract "view" with SRole labeled trees

- ❑ **Drawback**
  - ▪ parse explosion
  - ▪ semantic categorization task
    - ▪ WSD: for words
    - ▪ SRL: for phrases

A 2-step alternative

❑ Step 1:   Grammar

- ▪ tag with certainty
  - ▪ arguments: ARG1, ..., ARG*n*
- ▪ insert place-holders for the rest
  - ▪ modifiers: M

❑ Step 2:   Human annotators

- ▪ specify tags at the place holders
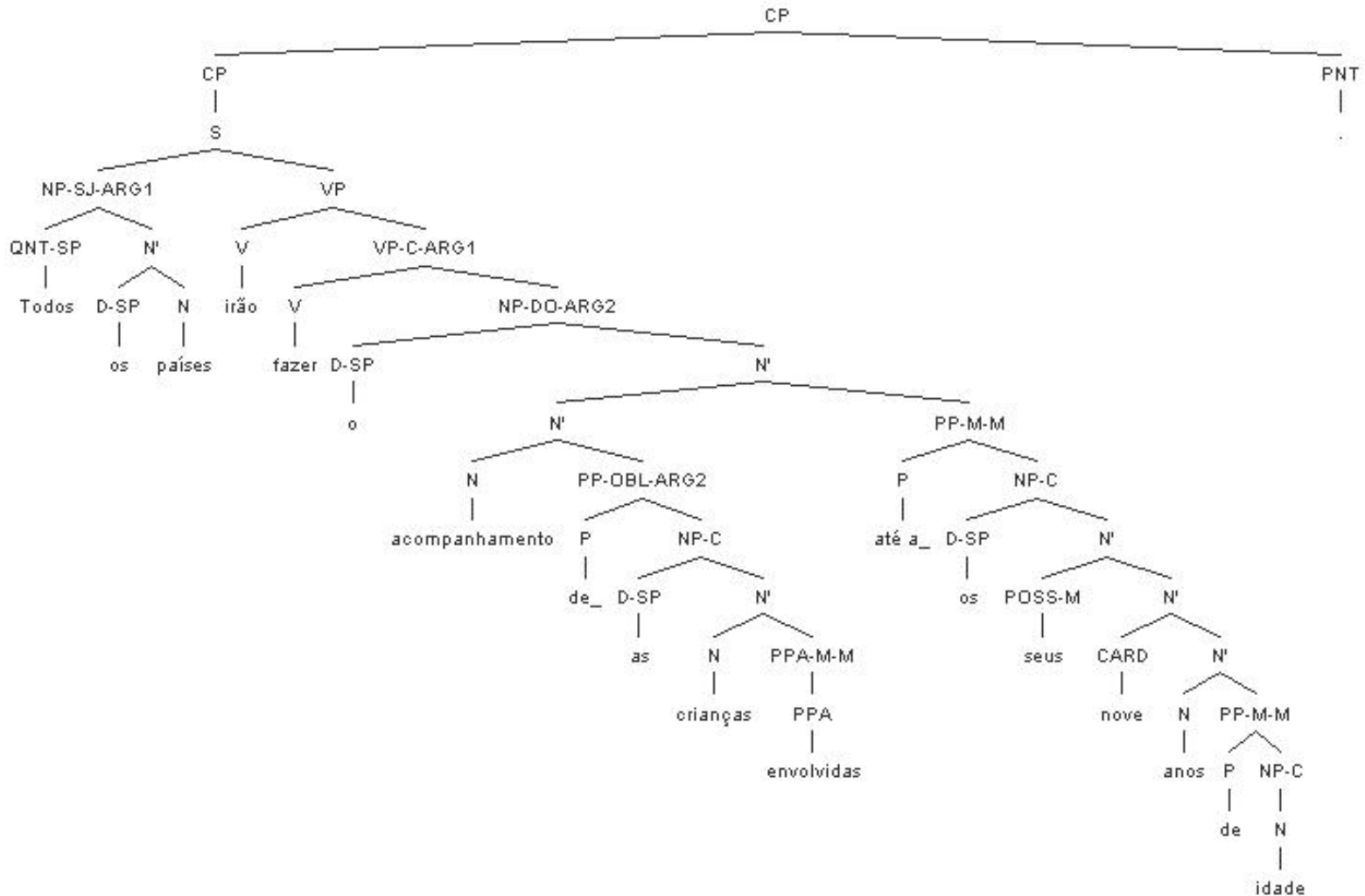- ▪ focused on what requires their decision

PropBanking workflow

1. ## Semi-labelled treebanking
   grammar + [incr tsdb()]

2. ## Conversion
   from derivation tree to annotation interface format

3. ## Labelling
   annotators + adjudicator

4. ## Retroversion
   back to treebank format

António Branco
Delph-in'2009, Barcelona, July 21

Semi-labelled tree

Incremental Multi-layer SR anotation

	A	B	C	D	E	F	G
1	Syntactic Function	Level 1 Semantic Role	Level 2 Semantic Role	Covered String	Observations	First Position	Last Position
2	PP-M	M	PNC	para a delegação evitar um conflito armado		0	7
3	NP-SJ	ARG1		a delegação		1	3
4	NP-DO	ARG2		um conflito armado		4	7
5	A-M	M	PRED	armado		6	7
6	PP-M	M	TMP	em Maio		7	8
7	NP-SJ	ARG1		a ONU		9	11
8	ADV-M	M	MNR	rapidamente		12	13
9	NP-DO	ARG2		tropas		13	14
10	PP-OBL	ARG3		para a fronteira		14	16
11							

2178 / 2179 / 2180 / 2181 / 2182 / 2183 / 2184 / 2185 / 2186 / 2187 \ **2188** / 2189 / 21

Sheet 11 / 50   PageStyle__2188__   100%   STD   Sum=0

❏ Multi-layer

- 2 levels of SR labelling used now

- Incremental
  - e.g. ARGi may be expanded to Agent, Patient, etc.
  - e.g. expand for more than 2 layers
  - ...

Dynamic PropBanking Kit

- ❏ Tools...
  - ▪ **Converter**
    - ▪ derivation tree to .xls interface format
  - ▪ **Annotations comparator**
    - ▪ input for adjudication
    - ▪ competing tags at the top of drop down menu
  - ▪ **Reverter**
    - ▪ to tree format

- ❏ ... also provision for dynamic treebanking
  - ▪ **Versions comparator**
    - ▪ previous version is kept below, in the .xls sheet

Evaluation

- ❑ First pilot experiment
  - ▪ 807sentences
  - ▪ two annotators + adjudicator
  - ▪ no pre-training

- ❑ Inter-annotator agreement
  - ▪ 334 assigned tags
  - ▪ k=0.77

- ❑ Estimate
  - ▪ Observed: 5 500 tokens / 10 hours
  - ▪ Projection: 250 Ktoken / 30 pm

Thank you!

Annotation interface

	A	B	C	D	E	F	G
1	Syntactic Function	Level 1 Semantic Role	Level 2 Semantic Role	Covered String	Observations	First Position	Last Position
2	PP-M	M	PNC	para a delegação evitar um conflito armado		0	7
3	NP-SJ	ARG1		a delegação		1	3
4	NP-DO	ARG2		um conflito armado		4	7
5	A-M	M	PRED	armado		6	7
6	PP-M	M	TMP	em Maio		7	8
7	NP-SJ	ARG1		a ONU		9	11
8	ADV-M	M	MNR	rapidamente		12	13
9	NP-DO	ARG2		tropas		13	14
10	PP-OBL	ARG3		para a fronteira		14	16
11							

2178 / 2179 / 2180 / 2181 / 2182 / 2183 / 2184 / 2185 / 2186 / 2187 \ **2188** / 2189 / 21

Sheet 11 / 50　　　PageStyle__2188__　　　100%　　　STD　　　Sum=0

António Branco
Delph-in'2009, Barcelona, July 21