

Site Update Lisbon

DELPH-IN Summit, St. Wendel, Germany, July 30, 2013

António Branco
University of Lisbon



1. Data sets development
2. Basic research
3. New projects

1 Data sets

- METANET project
 - successful completion, Jan 2013
- **CINTIL treebanks**
 - new releases of DeepBank and companion vistas
 - in META-SHARE
- Parallel DeepBank
 - inaugural release & publication of **ParDeepBank**, Dec 2012
 - Dan Flickinger, Valia Kordoni, Yi Zhang, António Branco, Kiril Simov, Petya Osenova, Catarina Carvalheiro, Francisco Costa and Sérgio Castro, 2012, "ParDeepBank: Multiple Parallel Deep Treebanking", *Proceedings, TLT2012 - The 11th International Workshop on Treebanks and Linguistic Theories*, Lisbon, pp.97-108.

Monolingual data sets

- Volume
 - Over 10 000 sentences, 110 000 tokens
- Methodology
 - double blind annotation followed by adjudication
- CINTIL collection
 - CINTIL DeepBank (hpsg)
 - CINTIL TreeBank (constituency)
 - CINTIL DepBank (dependency)
 - CINTIL PropBank (semantic roles)
 - CINTIL LogicalFormBank (mrs)

Multilingual data sets

- WSJ translation into Portuguese
 - complete: 25 folders
- Parallel deep treebanking
 - double blind annotation followed by adjudication
- Volume
 - over 4 000 sentences

2 Basic research

- **Temporal processing**
 - Francisco's PhD just completed
 - “aprovado com distinção e louvor por unanimidade”
 - paper in inaugural issue of Journal of Language Modelling, last week
- **OOV robustness**
 - João's PhD very close to completion
- New students joining
 - Helena Pimentel (PhD), Luís Morgado (MA)

3 Projects

- National: DP4LT
 - just started last June
- European: QTLeap
 - negotiation of contract ongoing

DP4LT

- Deep Processing for Language Technology
- FCT, 2 fte PostDoc, 2 fte graduate
- 24 months, started May 15
- Basic research: increasing robustness with focus on OOV handling
- Applied: lexicon development

QLeap

- Quality Translation by Deep Language Engineering Approaches
- EC, 8 partners, ca. 14 RTD fte
- 36 months, start Nov 1
- Explore new research paths opened by advanced multilingual resources
 - Aligned DeepBanks
 - Linked Data
- Stimulus to further advancements in deep processing

opportunities in deep processing research and beyond

- **Advance deep processing**
 - robustness in face of oov and mwe
 - semantic representation convergence/harmonization
 - generation
 - ...
- **Advance LRTs**
 - continue ParDeepGram
 - DeepBanks enhanced with lexical and distributional semantics
 - projecting DeepBanks from “less deep” treebanks
 - ...
- **Advance quality MT**
 - MRS2MRS SMT
 - statistical transfer
 - transfer rules learning
 - ...

4 Offer

- There is a celebration to do in 2014
 - 10th edition of DELPHIN Summits
- There is a symbolic occasion
 - for retrospective assessment and outlook into the future
- There is a full circle to close
 - and another one to “restart” where the series was initiated
- There is support to profit of
 - QTLeap
- There is a nice place and weather to enjoy
 - ...

want to come to Lisboa?



Thank you.