

# Panel — Introduction



Funded by CAMELEON Project — CAPES-COFECUB project 707/11

Porto Alegre, December 5, 2011

# Partner teams I

- Grenoble, France
  - LIG-GETALP
  - LIG-EXMO



# Partner teams II

- Porto Alegre, RS, Brazil
  - PLN PUCRS
  - PLN INF-UFRGS



# Partner teams III

- São Carlos, SP, Brazil
  - NILC



# 2011 Missions

- Carlos Ramisch (Jul/2011-Sep/2011)
- Renata Vieira (Sep/2011)
- Helena Caseli (Nov/2011)
- Emmanuelle Esperança-Rodier (Nov-Dec/2011)
- Mathieu Mangeot (Dec/2011)

# Areas

- Computational Linguistics (Computer Science and Linguistics)
- Machine Translation
- Web Semantics
- Information Retrieval

# Goals I

- Creation of bi- and multi-lingual lexical resources (dictionaries, etc.) for Portuguese, French and English
- Development of new automatic and collaborative techniques for lexical and ontological resources creation, enrichment and management



# Goals II

- Focus on multiword expressions and their impact in resources and applications
- Integrate multiword expressions across lexica and ontologies
- Integrate resources across languages
- Formation goal: exchange and formation of Brazilian and French researchers (PhD and Post-doc)



# Applications

- Better dictionaries for machine translation systems
- Better tools for semantic extraction from documents
- Improved tools for cross lingual information retrieval
- Faster and cheaper dictionary and ontology creation (current bottleneck)
- Available resources for language teaching

# First year outcomes

## Work packages

- WP1 - Multiword expressions and MT
- WP2 - JeuxDeMots Portuguese
- WP3 - NLP-based ontology enhancement

# WP1 - Multiword expressions and MT

- Automatic language-independent identification: `mwetoolkit`
- Integration with SMT system
- TODO: Evaluation and error analysis

# WP2 - JeuxDeMots Portuguese

- Interface translation
- Adaptation for regional variation
- Initial dictionary
- `http://jeuxdemots.liglab.fr/por`
- TODO: Populate, analyse, compare



# WP3 - NLP-based ontology enhancement

- Choice of domain: conferences
- Choice of seed keywords from ontology labels
- Download and clean corpus from Google
- Process with parser, extract useful information
- TODO: intrinsic and extrinsic evaluation

# Future directions

- Continue work on WPs 1–3
- WP4 - MWEs for alignment of lexical and ontological resources
- WP5 - Evaluation based on cross-lingual IR

CAMELEON 2011

Workshop CAMELEON 2011

First Workshop on Collaborative and Automatic Methods for the Multilingualisation of Lexica and Ontologies (CAMELEON 2011)

2011, December 5  
Instituto de Informática- UFRGS - Porto Alegre, RS - Brazil  
Auditório Castilho (Verde)

The first CAMELEON workshop will take place on December 5, 2011. It be a whole day dedicated to presenting and discussing the work related to French and Brazilian research on Computational Linguistics, Information Retrieval and Web Semantics with special focus on lexical and multilingual applications.

Programme

Invited talks and panel

- `http://cameleon.imag.fr`

# Questions? Comments?



Funded by CAMELEON Project — CAPES-COFECUB project 707/11

Porto Alegre, December 5, 2011