

LALIC

Laboratório de Linguística e
Inteligência Computacional

AIMOVEST

Analysis and Integration of MultiWord Expressions in Speech and Translation

UFSCar

Helena Caseli

helenacaseli@dc.ufscar.br

www.lalic.dc.ufscar.br

www.nilc.icmc.usp.br



Núcleo Interinstitucional de
Linguística Computacional

Research in Brazil

- Scenario
 - Mainly performed inside the Public Universities, in university laboratories
 - Mainly carried out by students (**undergraduate**, graduate)
 - Supervised by professors, so all the "permanent" are in fact full-time professors (we have to teach 8h/week, supervise and do research)

UFSCar's group

▪ LALIC



Laboratório de Linguística e Inteligência Computacional

- Profa. Helena Caseli (DC/UFSCar)
- Profa. Eloize Seno (IFSP/São Carlos)
- Currently: 12 students

→ www.lalic.dc.ufscar.br



Núcleo Interinstitucional de Linguística Computacional

- ~20 researchers and ~30 students

→ www.nilc.icmc.usp.br

▪ MaLL – Machine Learning Lab

- Prof. Estevam Hruschka (DC/UFSCar)

UFSCar's group

- Main topics
 - Machine translation
 - Phrase-based MT
 - Automatic post-editing
 - Automatic paraphrase identification
 - Automatic semantic relation extraction
 - Multilingual parsing
 - Machine learning
 - Never-ending machine learning
 - Deep learning

Ongoing research projects/LALIC



Learning from the web how to translate and paraphrase texts

FAPESP 2013/11811-0

09/2013-08/2015

Helena

1. Thiago – Master (2012-2014) – NEBEL

2. Cristiano – IC (2013-2014) – NESReL

3. Alexandre – TCC (2014) – NEMWEL

Eloize

4. Paulo – Master (2013-2015) – NEPaL

5. Gustavo – IC (2014) – Parap. + CS

6. Lucas – IC (2014-2015) – Syntactic parap.

IC and TCC = undergraduate students

Ongoing research projects/LALIC



AIM-WEST



FAPESP 2013/50757-0
03/2014-02/2017

Carlos Ramisch

Eloize Helena

1. Alexandre - TCC (2014) - NEMWEL
2. Thiago - Master (2012-2014) - NEBEL
3. Cristiano - IC (2013-2014) - NESReL
4. Paulo - Master (2013-2015) - NEPaL
5. Lucas - IC (2014-2015) - Syntactic parap.

Ongoing research projects/LALIC



RITA



CAPES 047/14

03/2014-02/2016

Fábio Kepler

1. Pablo – Master (2014-2016) –
Multilingual parsing

Carlos Ramisch

2. Alexandre – TCC (2014) – NEMWEL

3. Paulo – Master (2013-2015) – NEPaL

Eloize Helena

NEW research interests/LALIC

▪ *Deep Learning*

Helena

1. Fernando – TCC (2015) – Scene understanding (Socher et al., 2011) Jander Moreira
2. Pablo – Master (2014-2016) – Multilingual parsing (Legrand e Collobert, 2014) Fábio Kepler
3. Débora – PhD (2015-2018) – Automatic post-editing
4. Márcio – IC (2015-2016) – ?
5. Lucas Silva – IC (2015-2016) – ?
6. Leonardo – IC (2015-2016) – ?

AIM-WEST

- Planning
 - FAPESP budget = R\$ 48.855,07
 - 2 master or PhD internship of 3 months (2014, 2015)
 - 2015 - Pablo?
 - 2016 - Débora?
 - 1 visit (10 days) of a Brazilian researcher to the French group (2016)
 - 2016 - Eloize? Helena?

AIM-WEST

- Tasks X Students

1. Corpora construction

- Thiago (NEBEL) - parallel corpora (subtitles and lyrics)

- Alexandre (NEMWEL) - monolingual pt corpora (news from G1)

2. Protocols for translation evaluation

- Débora - automatic post-editing

3. Pre-processing of corpora

- Pablo - multilingual parsing

AIM-WEST

- Tasks X Students

- 4. Lexical and ontological resource construction for French, Portuguese and English

- Thiago (NEBEL) – bilingual lexicons

- Cristiano (NESReL) – semantic relations (useful for building ontologies)

- 5. Cognitive and data intensive models of MWE learning

- Alexandre (NEMWEL) – ML model for MWE learning

AIM-WEST

- Tasks X Students

- 6. Lexical models of MWEs

- Paulo (NEPaL) - never-ending paraphrases learning

- 7. Syntactic models of MWEs

- Pablo - multilingual parsing

- 8. Semantic models of MWEs

- Cristiano (NESReL)

- Paulo (NEPaL)

- 9. Automatic speech recognition systems

AIM-WEST

- Tasks X Students

- 10. Machine translation

- Thiago (NEBEL)

- Débora - automatic post-editing

- 11. Reports, papers and articles

- ALL

LALIC

Laboratório de Linguística e
Inteligência Computacional

AIMOVEST

Analysis and Integration of MultiWord Expressions in Speech and Translation

UFSCar

Helena Caseli

helenacaseli@dc.ufscar.br

www.lalic.dc.ufscar.br

www.nilc.icmc.usp.br



Núcleo Interinstitucional de
Linguística Computacional