

# Table of Contents

## Summarization

Modeling and Evaluating Summaries Using Complex Networks <i>Thiago Alexandre Salgueiro Pardo, Lucas Antikeira, Maria das Graças Volpe Nunes, Osvaldo N. Oliveira Jr., Luciano da Fontoura Costa</i> .....	1
SABIO: An Automatic Portuguese Text Summarizer Through Artificial Neural Networks in a More Biologically Plausible Model <i>Télvio Orrú, João Luís Garcia Rosa, Márcio Luiz de Andrade Netto</i> .....	11

## Resources

Building a Dictionary of Anthroponyms <i>Jorge Baptista, Fernando Batista, Nuno Mamede</i> .....	21
REPENTINO - A Wide-Scope Gazetteer for Entity Recognition in Portuguese <i>Luís Sarmento, Ana Sofia Pinto, Luís Cabral</i> .....	31

## Translation

Using Natural Alignment to Extract Translation Equivalents <i>Pablo Gamallo Otero</i> .....	41
Open-Source Portuguese-Spanish Machine Translation <i>Carme Armentano-Oller, Rafael C. Carrasco, Antonio M. Corbí- Bellot, Mikel L. Forcada, Mireia Ginestí-Rosell, Sergio Ortiz- Rojas, Juan Antonio Pérez-Ortiz, Gema Ramírez-Sánchez, Felipe Sánchez-Martínez, Miriam A. Scalco</i> .....	50
Weighted Finite-State Transducer Inference for Limited-Domain Speech-to-Speech Translation <i>Diamantino Caseiro, Isabel Trancoso</i> .....	60

## Named Entity Recognition

A Golden Resource for Named Entity Recognition in Portuguese <i>Diana Santos, Nuno Cardoso</i> .....	69
---------------------------------------------------------------------------------------------------------	----

Functional Aspects in Portuguese NER  
*Eckhard Bick* ..... 80

SIEMÊS - A Named-Entity Recognizer for Portuguese Relying on  
 Similarity Rules  
*Luís Sarmento* ..... 90

**Tools and Frameworks**

Tools for Nominalization: An Alternative for Lexical Normalization  
*Marco Antonio Insaurregiaga Gonzalez, Vera Lúcia Strube de Lima,  
 José Valdeni de Lima* ..... 100

A Framework for Integrating Natural Language Tools  
*João Graça, Nuno J. Mamede, João D. Pereira* ..... 110

Methods and Tools for Encoding the WordNet.Br Sentences, Concept  
 Glosses, and Conceptual-Semantic Relations  
*Bento C. Dias-da-Silva, Ariani Di Felippo, Ricardo Hasegawa* ..... 120

A Multi-agent Approach to Question Answering  
*Cássia Trojahn dos Santos, Paulo Quaresma, Irene Rodrigues,  
 Renata Vieira* ..... 131

**Systems and Models**

Adaptation of Data and Models for Probabilistic Parsing of Portuguese  
*Benjamin Wing, Jason Baldrige* ..... 140

A Set of NP-Extraction Rules for Portuguese: Defining, Learning and  
 Pruning  
*Claudia Oliveira, Maria Claudia Freitas, Violeta Quental,  
 Cícero Nogueira dos Santos, Renato Paes Leme, Lucas Souza* ..... 150

Resolving Portuguese Nominal Anaphora  
*Jorge C.B. Coelho, Vinicius M. Muller, Sandra Collovini,  
 Renata Vieira, Lucia H.M. Rino* ..... 160

Design of a Multimodal Input Interface for a Dialogue System  
*João P. Neto, Renato Cassaca, Márcio Viveiros, Márcio Mourão* .... 170

Review and Evaluation of DiZer -- An Automatic Discourse Analyzer  
 for Brazilian Portuguese  
*Thiago Alexandre Salgueiro Pardo, Maria das Graças Volpe Nunes* ... 180

Classroom Lecture Recognition <i>Isabel Trancoso, Ricardo Nunes, Luís Neves</i> .....	190
------------------------------------------------------------------------------------------	-----

## Information Extraction

Semi-supervised Learning for Portuguese Noun Phrase Extraction <i>Ruy Milidiú, Cicero Santos, Julio Duarte, Raúl Rentería</i> .....	200
Automatic Extraction of Keywords for the Portuguese Language <i>Maria Abadia Lacerda Dias, Marcelo de Gomensoro Malheiros</i> .....	204
Semi-automatically Building Ontological Structures from Portuguese Written Texts <i>Túlio Lima Baségio, Vera Lúcia Strube de Lima</i> .....	208

## Speech Processing

On the Use of Machine Learning and Syllable Information in European Portuguese Grapheme-Phone Conversion <i>António Teixeira, Catarina Oliveira, Lurdes Moutinho</i> .....	212
A Model to Computational Speech Understanding <i>Daniel Nehme Müller, Mozart Lemos de Siqueira, Philippe O.A. Navaux</i> .....	216
Phonetic Sequence to Graphemes Conversion Based on DTW and One-Stage Algorithms <i>Rafael Teruszkin, Fernando Gil Vianna Resende Jr.</i> .....	220

## Lexicon

Very Strict Selectional Restrictions: A Comparison Between Portuguese and French <i>Éric Laporte, Christian Leclère, Maria Carmelita Dias</i> .....	225
Towards a Formalization of Tense and Aspect for the Generation of Portuguese Sentences <i>Michel Gagnon, Eliana de Mattos Pinto Coelho, Roger Antonio Finger</i> .....	229

The Need for Application-Dependent WSD Strategies: A Case Study in MT

*Lucia Specia, Gabriela Castelo Branco Ribeiro,  
 Maria das Graças Volpe Nunes, Mark Stevenson* . . . . . 233

A Lexical Database of Portuguese Multiword Expressions

*Sandra Antunes, Maria Fernanda Bacelar do Nascimento,  
 João Miguel Casteleiro, Amália Mendes, Luísa Pereira, Tiago Sá* . . . . 238

**Morpho-syntactic Studies**

Dedicated Nominal Featurization of Portuguese

*António Branco, João Ricardo Silva* . . . . . 244

Part-of-Speech Tagging of Portuguese Based on Variable Length Markov Chains

*Fábio Natanael Kepler, Marcelo Finger* . . . . . 248

From Syntactical Analysis to Textual Segmentation

*Ana Luisa Leal, Paulo Quaresma, Rove Chishman* . . . . . 252

Syntactical Annotation of COMPARA: Workflow and First Results

*Susana Inácio, Diana Santos* . . . . . 256

**Web, Corpus and Evaluation**

A Complex Evaluation Architecture for HAREM

*Nuno Seco, Diana Santos, Nuno Cardoso, Rui Vilela* . . . . . 260

What Kinds of Geographical Information Are There in the Portuguese Web?

*Marcirio Silveira Chaves, Diana Santos* . . . . . 264

Corpus-Based Compositionality

*Milena Garrão, Claudia Oliveira, Maria Claudia de Freitas,  
 Maria Carmelita Dias* . . . . . 268

**Author Index** . . . . . 273