

# IX Brazilian Symposium in Information and Human Language Technology - STIL



October 21-23, 2013  
Fortaleza, Ceará, Brazil  
Fábrica de Negócios

PROCEEDINGS

## **Acknowledgements**

The Program Committee chairs acknowledge the financial support to the conference provided by the Brazilian Computer Society (SBC), the University of Fortaleza (UNIFOR) and the Federal University of Ceará. We thank the Program Committees of the IX Brazilian Symposium in Information and Human Language Technology and Collocated Events for all the reviews that they produced. Last but not least, we thank the local organization led by Prof. Vasco Furtado (General Local Chair) and Profa. Vlândia Pinheiro (STIL Local Chair).

Sandra Maria Aluísio and Valéria Delisandra Feltrim

(STIL 2013 Program Chairs)

**STIL 2013**

**IX Brazilian Symposium in Information and Human  
Language Technology and Collocated Events**

III Workshop on Portuguese Description

III Student Workshop on Information and Human Language Technology

IV Workshop on RST and Text Studies

**Proceedings of the Conference**

**October 21 to 23, 2013**

**Fortaleza, CE**

## **Table of Contents**

[IX Brazilian Symposium in Information and Human Language Technology](#)

[III Workshop on Portuguese Description](#)

[III Student Workshop on Information and Human Language Technology](#)

[IV Workshop on RST and Text Studies](#)

## IX Brazilian Symposium in Information and Human Language Technology

STIL is the bi-annual Language Technology event supported by the Brazilian Computer Society (SBC) and by the Brazilian Special Interest Group on Natural Language Processing. In 2013 it takes place at Fortaleza, Ceará, Brazil. The conference has a multidisciplinary nature and covers a broad spectrum of disciplines related to Human Language Technology, such as Linguistics, Computer Science, Psychology, Information Science, among others. It aims at bringing together both academic and industry participants that work on those areas.

The topics of interest centered around work in human language technology in general from perspectives as diverse as Computer Science, Linguistics and Information Science, such as text mining, written and spoken language processing, terminology, lexicology and lexicography and natural language generation. We received 65 submissions from Brazil, Great Britain, Peru, Germany, United States, and Portugal. Each paper was reviewed by at least three members of the Program Committee which had 64 members from 7 countries and 34 institutions. After a rigorous reviewing process, 15 papers were selected for oral presentation (acceptance rate of 23%), and 17 papers were selected for poster presentation, including some long papers reallocated as short papers.

Besides the technical sessions with paper and poster presentations, STIL 2013 includes two invited talks and two tutorials in its program. The invited speakers are Dr. Maite Taboada (Simon Fraser University, CA) and Dr. Diana Maynard (University of Sheffield, UK). The tutorials will also be given by Dr. Taboada and Dr. Maynard.

We want to thank all the authors, members of the Program Committee, Invited Speakers, SBC staff, and the Local and General Chairs of **STIL 2013**. We hope you enjoy this volume.

Sandra Maria Aluísio and Valéria Delisandra Feltrim

(STIL 2013 Program Chairs)

## Organizing Committee

### Program Chairs

Sandra Maria Aluísio (ICMC/USP)

Valéria Delisandra Feltrim (UEM)

### Local Chair

Vlândia Pinheiro (UNIFOR)

### Program Committee

Adriano Zanette (UFRGS, Brazil)

Aldebaro Klautau (UFPA, Brazil)

Aline Villavicencio (UFRGS, Brazil)

Andre Gustavo Adami (University of Caxias do Sul, Brazil)

António Branco (University of Lisbon, Portugal)

Carlos Augusto Prolo (UFRN, Brazil)

Carlos Ramisch (Joseph Fourier University, Grenoble, France)

Carolina Evaristo (University of Sheffield)

Caroline Hagege (Xerox Research Centre Europe, France)

Cassia Trojahn (IRIT-UTM, France)

Cícero Nogueira dos Santos (IBM Research, Brazil)

Clarissa Castella Xavier (PUC/RS, Brazil)

Daniel Nehme Muller (UFRGS, Brazil)

Ethel Schuster (Northern Essex Community College, USA)

Fábio Kepler (Universidade Federal do Pampa, Brazil)

Helena Caseli (UFSCar, Brazil)

Heliana Mello (UFMG, Brazil)

Ivandré Paraboni (EACH/USP, Brazil)

Jorge Baptista (University of Algarve, Portugal)

Laura Alonso i Alemany (FaMAF – UNC, Argentina)

Lianet Sepúlveda Torres (ICMC/USP, Brazil)

Lucelene Lopes (PUC/RS, Brazil)

Lucia Rino (UFSCar, Brazil)

Magali Sanches Duran (ICMC/USP, Brazil)

Marcelo Finger (IME/USP, Brazil)

Maria das Gracas Volpe Nunes (ICMC/USP, Brazil)

Marlo Souza (UFRGS, Brazil)

Nelson Neto (UFPA, Brazil)

Palmira Marrafa (University of Lisbon, Portugal)

Paloma Moreda (University of Alicante, Spain)

Renata Vieira (PUC/RS, Brazil)

Rodrigo Wilkens (UFRGS, Brazil)

Ronaldo Martins (Univas, Brazil)

Rove Chishman (UNISINOS, Brazil)

Sara Candeias (University of Coimbra, Portugal)

Sérgio Antônio Andrade de Freitas (UnB, Brazil)  
Stella Tagnin (FFLCH/USP, Brazil)  
Ted Pedersen (University of Minnesota, USA)  
Thiago A. S. Pardo (ICMC/USP, Brazil)  
Vera Lucia Strube de Lima (PUC/RS, Brazil)  
Vlândia Pinheiro (UNIFOR, Brazil)  
Wilker Aziz (University of Wolverhampton, UK)

### **Additional Reviewers**

Arnaldo Candido Junior (ICMC/USP, Brazil)  
Bento Dias da Silva (UNESP, Brazil)  
Carmen Dayrell (COMET/USP, Brazil)  
Christopher Shulby (ICMC/USP, Brazil)  
Daniel Lucrédio (UFSCar, Brazil)  
Erick Maziero (ICMC/USP, Brazil)  
Heloisa Camargo (UFSCar, Brazil)  
Leandro Mendonça de Oliveira (EMBRAPA, Brazil)  
Lucia Specia (University of Sheffield, UK)  
Maria Jorge (USP, Brazil)  
Marilde Santos (UFSCar, Brazil)  
Márcio Dias (UFG, Brazil)  
Norton Roman (EACH/USP, Brazil)  
Osvaldo de Oliveira Jr. (IFSC/USP, Brazil)  
Paula Cardoso (ICMC/USP, Brazil)  
Paulo Schreiner (UFRGS, Brazil)  
Pedro Balage Filho (ICMC/USP, Brazil)  
Renata Fortes (ICMC/USP, Brazil)  
Roger Granada (PUC/RS, Brazil)  
Roseli Francelin Romero (ICMC/USP, Brazil)  
Sandra Maria Aluísio (ICMC/USP, Brazil)  
Valéria Delisandra Feltrim (UEM, Brazil)

### **Natural Language Processing Steering Committee**

Thiago A. S. Pardo, Universidade de São Paulo, USP/São Carlos  
Aline Villavicencio, Universidade Federal do Rio Grande do Sul - UFRGS  
Helena M. Caseli, Universidade Federal de São Carlos – UFScar  
Sandra Maria Aluísio, Universidade de São Paulo, USP/São Carlos  
Ariani Di Felippo, Universidade Federal de São Carlos - UFSCar

## **Invited Talk**

### **Computational linguistics and discourse**

Maite Taboada

Simon Fraser University

#### **Abstract:**

Computational linguistics has never entirely forgotten about discourse, but it is safe to say that it has, until recently, neglected information above the sentence level. Much of the progress in language processing in the last couple of decades has emphasized classification methods that rely on low-level features, such as unigrams or bigrams. The tide is certainly turning, with more and more attention to the types of high-level structures that are necessary in tasks such as semantic role labelling, summarization and sentiment analysis. In this talk, I will provide a historical overview of computational treatments of discourse, but also a summary of the most interesting efforts in utilizing the rich information that discourse features can contribute to natural language processing. I will in particular concentrate on discourse parsing and sentiment analysis.

#### **Short Bio:**

Maite Taboada is Associate Professor of Linguistics at Simon Fraser University (Canada). She holds MA and PhD degrees from the Universidad Complutense de Madrid (Spain), and an MSc in Computational Linguistics from Carnegie Mellon University. Maite works in the areas of discourse analysis, systemic functional linguistics and computational linguistics. She has carried out research on coherence and cohesion, information structure and turn-taking. She has also participated in research projects in natural language generation, machine translation and software agents. Ongoing research addresses the study of opinion and sentiment in text, including a system that extracts sentiment automatically. Other current areas of research are coherence in multimodal documents, and the study of cataphoric relations. She is a frequent program committee member and reviewer in the computational linguistics community, and is currently a member of the Editorial Board of the Journal of Natural Language Engineering.



## **Invited Talk**

### **What do you really mean when you tweet? Challenges for opinion mining on social media**

Diana Maynard  
University of Sheffield

#### **Abstract:**

Social media is fast becoming a crucial part of our everyday lives, not only as a fun and practical way to share our interests and activities with geographically distributed networks of friends, but as an important part of our business lives also. On the one hand, it affords a highly effective means of promotion and advertising for companies, as well as market watch activities to keep an eye on competitors and collaborators, while on the other hand, it enables companies and institutions to acquire valuable feedback by analysing what their customers have to say. All kinds of predictions can be made based on such knowledge, from gauging current political opinions and predicting stock price movements relative to public mood, to the more frivolous but still highly lucrative predictions about Oscar winners and film revenues. There are already hundreds of tools for social media and sentiment analytics available, but they are far from perfect both in quality and coverage. Appearances can be deceptive: what you really mean when you tweet, using sarcasm, slang, or even just assumptions about context and world knowledge, may be very different from a surface interpretation. Furthermore, there is an enormous amount of multimedia material on the social web which could provide useful information if only such analysis tools could interpret the sentiments behind it. This talk will introduce the topics of opinion mining and social media analytics, in particular looking at the challenges they impose for an NLP system. It will investigate the impact of non-standard text in social media, use of sarcasm, swear words, non-words, short sentences, multiple languages and so on, which impede the success of current NLP tools to perform good analysis, and will examine tools being developed in some current cutting-edge research projects, including not only text-based research but also multimedia analysis.

#### **Short Bio:**

Diana Maynard is a research fellow at the University of Sheffield, UK. She has a PhD in automatic term recognition from Manchester Metropolitan University, and has been involved in research in natural language processing since 1994. Her main interests are in information extraction, opinion mining, social media, terminology and semantic web technologies. Since 2000 she has led the development of USFD's open-source multilingual Information Extraction tools, and has led research

teams on a number of UK and EU projects. She is chair of the annual GATE training courses and leads the GATE consultancy on IE and opinion mining. She has published extensively, organised a number of national and international conferences, workshops and tutorials, including the annual ISWC Semantic Web Challenge for the last 3 years. She has given a number of invited talks and keynote speeches, including most recently an invited tutorial at the Sentiment Analysis Symposium in 2012.

# **Tutorial 1**

## **Practical Opinion Mining for social media**

Diana Maynard

University of Sheffield

This tutorial will introduce the concepts of sentiment analysis and opinion mining from unstructured text in social media, looking at why they are useful and what tools and techniques are available. It will cover both rule-based and machine learning techniques, provide some background information on the key underlying NLP processes required, and look in detail at some of the major problems and solutions, such as detection of sarcasm, use of informal language, spam opinion detection, trustworthiness of opinion holders, and so on. The techniques will be demonstrated with real applications developed in GATE, an open-source language processing toolkit. Hands-on exercises and relevant materials will be provided for participants to try out the applications, and to experiment with building their own tools. Alternative tools to GATE will also be discussed and hands-on experiments provided for these.

The tutorial will be divided into 3 sections, as follows:

### 1. Introduction to Opinion Mining in Social Media

Explanation of the concept of opinion mining and the motivation for it.

Introduction to the main subtasks of an opinion mining system and the major challenges.

Discussion of some of the strengths and weaknesses of current systems, and how to evaluate them.

### 2. Introduction to GATE

Brief introduction to the concept of GATE (a language processing toolkit) and its major components.

### 3. Applications and Hands-On Experimentation

Machine learning for sentiment analysis of consumer reviews.

Rule-based application for opinion finding in social media.

## **Tutorial 2**

### **Annotating signals for discourse relations**

Maite Taboada

Simon Fraser University

Discourse relations are the building blocks of discourse, explaining how and why discourse is coherent and how it is effective in communicating the author's goals (which is why they are also known as coherence or rhetorical relations). They are, as a consequence, of great value to a number of computational tasks that involve processing beyond the sentence, such as summarization, information extraction or sentiment analysis.

In this tutorial, I will explain the concept of discourse relations using Rhetorical Structure Theory, but also making reference to related approaches to the study of discourse relations. Topics to be included:

- Rhetorical relations
- Existing resources: RST Discourse Treebank, Penn Discourse Treebank
- Relation signalling: discourse markers and beyond
- Annotating signals for discourse relations
- Reliability of annotation
- Annotation layering

## Long Papers

1. [AgreeCalc: Uma Ferramenta para Análise da Concordância entre Múltiplos Anotadores](#)

*Alexandre Rossi Alvares, Norton Roman*

2. [MetaAnn: Um Gerador de Ferramentas para Anotação de Textos](#)

*Tiago Emanuel Infante Missão, Norton Roman*

3. [Improving CoGrOO: the Brazilian Portuguese Grammar Checker](#)

*William Silva, Marcelo Finger*

4. [Uma Investigação sobre Algoritmos de Diferentes Abordagens de Aprendizado Supervisionado na Classificação de Papéis Retóricos em Resumos Científicos](#)

*Vinícius Mourão Alves de Souza, Valéria D. Feltrim*

5. [O Subframe Sentença no Complexo Frame Processo de Conhecimento no Direito Processual Civil](#)

*Thais Minghelli, Anderson Bertoldi, Rove Chishman*

6. [Subtopic Annotation in a Corpus of News Texts Steps Towards Automatic Subtopic Segmentation](#)

*Paula Cardoso, Thiago Pardo, Maite Taboada*

7. [O Reconhecimento de Entidades Nomeadas por meio de Conditional Random Fields para a Língua Portuguesa](#)

*Daniela do Amaral, Renata Vieira*

8. [Análise Automática de Coerência Usando o Modelo Grade de Entidades para o Português](#)

*Alison Polpeta Freitas, Valéria D. Feltrim*

9. [Aplicando Pontos de Corte para Listas de Termos Extraídos](#)

*Lucelene Lopes, Renata Vieira*

10. [Geração de Expressões de Referência usando Relações Espaciais](#)

*Diego Silva, Ivandré Paraboni*

11. [Mac-Morpho Revisited: Towards Robust Part-of-Speech Tagging](#)

*Erick Fonseca, João Luís Rosa*

12. [Implementação de um Separador Silábico Gratuito Baseado em Regras Linguísticas para o Português Brasileiro](#)

*Willian da Rocha, Nelson Neto*

13. [Text Simplification as Tree Transduction](#)

*Gustavo Paetzold, Lucia Specia*

14. [Automatic Disambiguation of Homographic Heterophone Pairs Containing Open and Closed Mid Vowels](#)

*Christopher Shulby, Gustavo de Mendonca Almeida, Vanessa Marquifável*

15. [Desambiguação Lexical de Sentido com uso de Informação Multidocumento por meio de Redes de Co-ocorrência](#)

*Fernando Antônio Nóbrega, Thiago Pardo*

## Short Papers

16. [Realização Superficial Baseada em Regras](#)  
*Douglas Silva, Eder Novais, Ivandré Paraboni*
17. [JWN-Br – Uma API Java para a WordNet.Br](#)  
*Vitor Machado Oliveira, Norton Roman*
18. [Geração de features para resolução de correferência: Pessoa, Local e Organização](#)  
*Evandro Fonseca, Renata Vieira e Aline Vanin*
19. [Modalidade no Português Brasileiro: as estruturas condicionais](#)  
*Priscila Osório Côrtes*
20. [Um repositório de verbos para a anotação de papéis semânticos disponível na web](#)  
*Magali Duran, Jhonata Pereira Martins, Sandra Aluísio*
21. [Entity-centric Sentiment Analysis on Twitter data for the Portuguese Language](#)  
*Marlo Souza, Renata Vieira*
22. [Approaches for Helping Brazilian Students Improve their Scientific Writings](#)  
*Ethel Schuster, Richard Lizotte, Sandra Aluísio, Carmen Dayrell*
23. [BrGram: uma gramática computacional de um fragmento do português brasileiro no formalismo da LFG](#)  
*Leonel Figueiredo de Alencar*
24. [Geração de instruções em mundos virtuais: primeiros passos](#)  
*Diego Sandrim, Felipe Freire, Ivandré Paraboni*
25. [Extração de Vocabulário Multilíngue para Tradução em Domínios Especializados](#)  
*Lucas Hilgert, Renata Vieira*
26. [Corpus-amostra Coelho Netto: compilação, anotação e ocorrências em textos literários dos séc. XIX e XX](#)  
*Francimary Martins*
27. [Acoustic, Phonetic and Prosodic Features of Parkinson's disease Speech](#)  
*Jorge Proença, Arlindo Veiga, Sara Candeias, Fernando Perdigão*
28. [Classificação de Prioridade de Tweets utilizando Máquinas de Vetor de Suporte](#)  
*Vinícius Pazzini, Tiago Schenkel, Mikael Poetsch, Ricardo Araújo*

29. [An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis](#)

*Pedro Balage Filho, Thiago Pardo, Sandra Aluísio*

30. [Uma investigação no apoio da avaliação da usabilidade em Sistemas Sociais usando Processamento da Linguagem Natural](#)

*Marília Mendes, Elizabeth Furtado, Miguel Franklin de Castro*

31. [Uma Ferramenta para Projeto de Sistemas de Diálogos para Call Center Baseados em Asterisk](#)

*Estillac Borges Filho, Pedro Batista, Aldebaro Klautau*



### **III Workshop on Portuguese Description - JDP**

A III Jornada de Descrição do Português (III JDP) integrou o IX Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL 2013), que ocorreu entre os dias 21 e 23 de outubro de 2013, em Fortaleza, CE.

A JDP visa aproximar as pesquisas em Linguística Descritiva da área do Processamento Automático das Línguas Naturais (PLN), integrando mais efetivamente os linguistas e os cientistas da computação que precisam atuar de forma interdisciplinar para promover avanços no processamento automático da língua portuguesa.

Diante disso, os trabalhos submetidos à JDP vinculam-se aos grandes temas da descrição linguística, a saber: Fonética e Fonologia; Lexicologia, Lexicografia e Terminologia; Sintaxe; Semântica; Texto e Discurso, nas mais diversas correntes teóricas.

Acredita-se que a iniciativa da JDP em fomentar o trabalho colaborativo de linguistas e cientistas da computação pode contribuir para colocar a língua portuguesa numa posição de destaque no cenário mundial, fazendo frente às demais línguas (inglês, francês, espanhol, etc.) que vislumbraram essa interdisciplinaridade já na década de 1960.

Em sua terceira edição, a JDP recebeu 7 submissões, as quais foram aceitas para apresentação oral. Entre elas, tivemos 3 estados brasileiros representados (São Paulo, Minas Gerais e Santa Catarina), além de um trabalho proveniente do exterior (Portugal).

Mostra-se abaixo a lista dos tópicos abordados pelos artigos aceitos:

Sintaxe	1
Texto/Sumarização automática	1
Lexicologia/Lxicografia	2
Fonética/Processamento de Fala	3

## Comitê de Organização

### Comitê de Programa

Claudia Dias de Barros (UFSCar/DL/PPGL/NILC)

Ariani Di Felippo (UFSCar/DL/NILC)

Oto Araújo Vale (UFSCar/DL/NILC)

### Comitê Científico

*Ariani Di Felippo (Universidade Federal de São Carlos, Brasil)*

*Bento Carlos Dias-da-Silva (Universidade Estadual Paulista, Brasil)*

*Cláudia Dias de Barros (Universidade Federal de São Carlos, Brasil)*

*Cláudia Zavaglia (Universidade Estadual Paulista, Brasil)*

*Éric Laporte (Université Paris Est, França)*

*Flávia Bezerra de Menezes Hirata-Vale (Universidade Federal de São Carlos, Brasil)*

*Gladis Maria de Barcellos Almeida (Universidade Federal de São Carlos, Brasil)*

*Jorge Baptista (Universidade do Algarve, Portugal)*

*Juliano Desiderato Antonio (Universidade Estadual de Maringá, Brasil)*

*Magali Sanches Duran (Universidade de São Paulo, Brasil)*

*Margarita Correia (Universidade de Lisboa, Portugal)*

*Maria José Bocorny Finatto (Universidade Federal do Rio Grande do Sul, Brasil)*

*Oto Araújo Vale (Universidade Federal de São Carlos, Brasil)*

*Rove Chishman (Universidade do Vale do Rio dos Sinos, Brasil)*

*Stella Esther Ortweiler Tagnin (Universidade de São Paulo, Brasil)*

*Tony Berber Sardinha (Pontifícia Universidade Católica, São Paulo, Brasil)*

*Violeta Quental (Pontifícia Universidade Católica, Rio de Janeiro, Brasil)*

## Artigos

1. [Análise automática de entoação emotiva \(colérica, triste e neutra\) pelo aplicativo ExProsodia](#)

*Waldemar F. Netto, Daniel O. Peres, Marcus V. M. Martins e Maressa de F. Vieira*

2. [Análise dos falsos aumentativos no Português do Brasil](#)

*Roana Rodrigues e Oto Araújo Vale*

3. [Em busca de métodos de detecção da complementaridade para a Sumarização Automática Multidocumento](#)

*Jackson Wilke da Cruz Souza e Ariani Di Felippo*

4. [Frames e Multilinguagem: contribuições da Semântica de Frames e da FrameNet para a constituição de dicionários eletrônicos multilíngues](#)

*Maucha Andrade Gamonal e Tiago Timponi Torrent*

5. [Sistema automático de pronúncia de verbos](#)

*Sara Candeias, Arlindo Veiga e Fernando Perdigão*

6. [Tipologia sintática das construções com os verbos-suporte dar, fazer e ter](#)

*Amanda P. Rassi, Cláudia D. Barros e Maria Cristina A. dos Santos-Turati*

7. [Um algoritmo prosódico para Português do Brasil](#)

*Vera Vasilévski e Márcio José Araújo*

### **III Student Workshop on Information and Human Language Technology**

O III Workshop de Iniciação Científica em Tecnologia da Informação e da Linguagem Humana (TILic 2013) integrou o IX Brazilian Symposium in Information and Human Language Technology (STIL 2013), que ocorreu entre os dias 21 e 23 de outubro de 2013, em Fortaleza, CE.

O TILic é um workshop de iniciação científica que busca (i) agregar alunos de graduação das áreas de ciência da computação e de linguística engajados em pesquisas relacionadas ao processamento automático das línguas naturais, e (ii) fomentar a interação e estimular o debate entre os alunos de graduação de diferentes áreas que tomam as línguas naturais e suas tecnologias como objetos de estudo. Essa é terceira edição do evento, realizado sempre em conjunto com o STIL, em São Carlos (2009) e em Cuiabá (2011). A cada edição, o TILic atrai jovens pesquisadores para a apresentação de seus temas de pesquisa e compartilhamento de experiências e ideias com pesquisadores mais experientes.

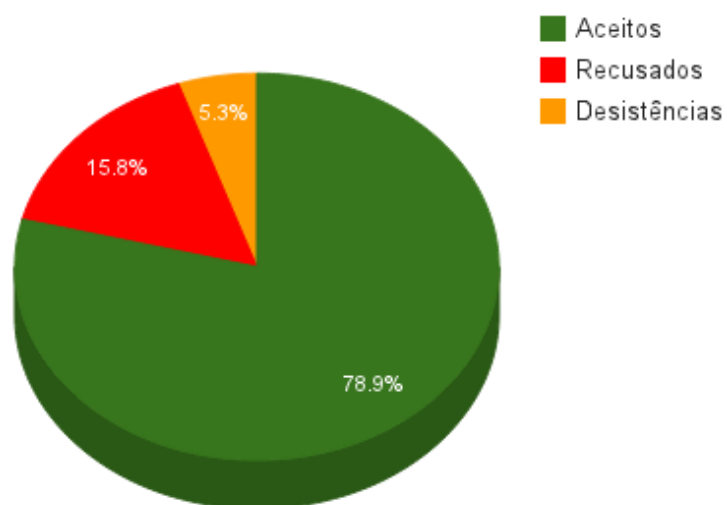
Esta edição do TILic recebeu 19 submissões de artigos. Cada artigo foi revisado por 3 membros do comitê de programa de acordo com os seguintes critérios: formato, pertinência, clareza, coerência, robustez e qualidade da revisão bibliográfica. Os artigos deveriam respeitar o limite de 3 páginas e seguir o formato da Sociedade Brasileira de Computação. Dos 19 artigos submetidos, 16 foram aceitos (84.2%) para publicação de acordo com os pareceres dos revisores. Destes, 15 estão contidos nestes anais (houve 1 desistência) e foram apresentados no formato de pôster em seção dedicada aos trabalhos dos jovens pesquisadores durante a programação principal do STIL.

Dos 15 artigos publicados, 8 foram produzidos por alunos da área de ciência da computação, e 7 por alunos da área de letras e linguística, mostrando um bom equilíbrio pluridisciplinar na comunidade. No entanto, a distribuição geográfica dos artigos é pouco homogênea, com 11 artigos (73%) provenientes do estado de São Paulo e 4 artigos da região Sul (dos quais 3 do Rio Grande do Sul e 1 do Paraná). Em edições futuras, gostaríamos de ver uma maior participação de jovens pesquisadores de outros estados do Brasil. Os tópicos dos artigos são bastante variados, com predominância de artigos sobre sumarização automática (33%), mas também 3 artigos sobre a criação e adaptação de recursos lexicais, 2 artigos sobre análise de sentimentos e emoções, 2 artigos sobre educação e ensino de línguas, 2 artigos sobre tradução automática e corpora paralelos e 1 artigo sobre análise sintática (parsing).

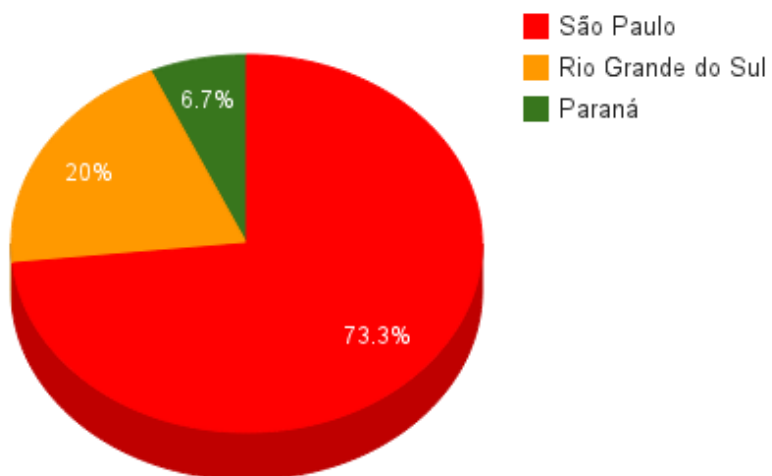
O evento contou com o patrocínio da North American Chapter of the Association for Computational Linguistics (NAACL).

Acompanhe abaixo os gráficos com os principais dados estatísticos dos artigos:

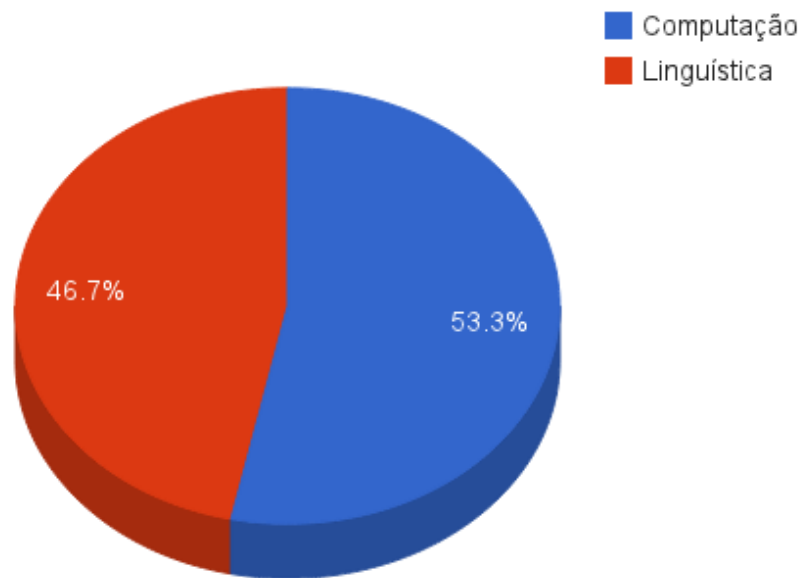
### Proporção de artigos aceitos



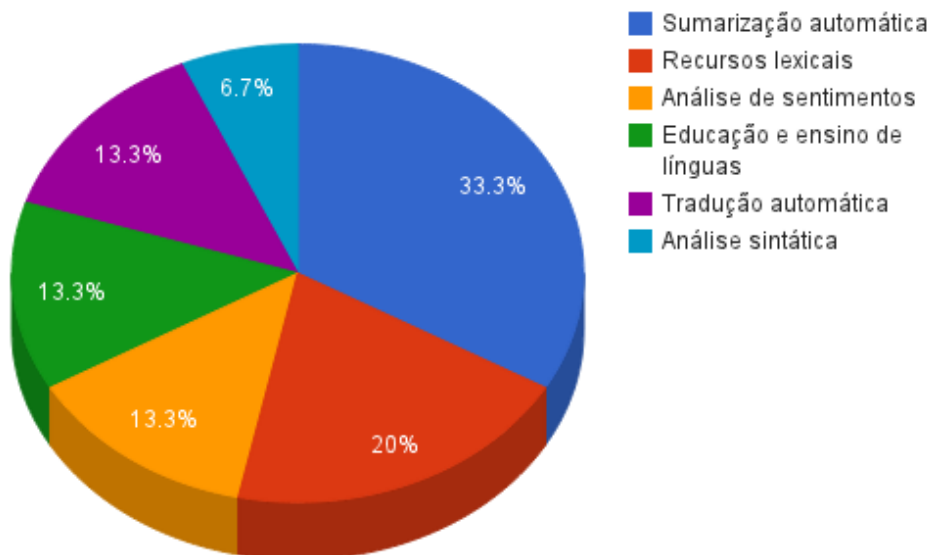
### Distribuição geográfica dos primeiros autores



### Áreas de atuação dos primeiros autores



### Tópicos abordados nos artigos



## **Comitê de Organização**

### **Comitê de Programa**

Helena de Medeiros Caseli  
Professora do Departamento de Computação (DC) da Universidade Federal de São Carlos (UFSCar)  
Pesquisadora sênior do Núcleo Interinstitucional de Linguística Computacional (NILC)

Carlos Ramisch  
Professor da Universidade Aix-Marseille (França)  
Pesquisador do grupo TALEP do Laboratório de Informática Fundamental (LIF)

Pedro Paulo Balage Filho  
Doutorando da Universidade de São Paulo (ICMC-USP)  
Pesquisador do Núcleo Interinstitucional de Linguística Computacional (NILC)

### **Comitê Científico**

Wilker Aziz, University of Wolverhampton (Reino Unido)  
Daniel Beck, University of Sheffield (Reino Unido)  
José Guilherme Camargo de Souza, University of Trento (Itália)  
Arnaldo Cândido Junior, NILC, ICMC, USP (Brasil)  
Lucía Castro, ICMC-USP (Brasil)  
Carolina Evaristo Scarton, Universidade de São Paulo (Brasil)  
Aline Evers, UFRGS (Brasil)  
Caroline Gasperin, SwiftKey (Reino Unido)  
Roger Leitzke Granada, Université Toulouse II - Le Mirail (França)  
Fábio Natanael Kepler, Universidade Federal do Pampa - UNIPAMPA (Brasil)  
Lucelene Lopes, PUCRS (Brasil)  
Leandro Henrique Mendonça de Oliveira, Empresa Informática Agropecuária (Brasil)  
Bianca Franco Pasqualini, UFRGS (Brasil)  
Eloize Rossi Marques Seno, Instituto Federal de São Paulo - campus São Carlos (Brasil)  
Magali Sanches Duran, NILC (Brasil)  
Paulo Schreiner, UFRGS (Brasil), MIT (Estados Unidos)  
Lianet Sepulveda Torres, Universidade de São Paulo (Brasil)  
Dayse Simon Landim Kamikawachi, UFSCar/São Carlos (Brasil)  
Rodrigo Souza Wilkens, UFRGS (Brasil)  
Leonardo Zilio, UFRGS (Brasil)

## Artigos

1. [Métodos de Pós-Edição Automática aplicados à Tradução Estatística Inglês-Português](#)  
*Lucas Avanço, ICMC-USP*
2. [A Visualization System for Multi-document Summarization](#)  
*Jader B. P. Lima, ICMC-USP*
3. [Adequação de Grafos de Flexão Nominal à nova Ortografia](#)  
*Nathalia Perussi Calcia, DL-UFSCar*
4. [Integrando Recursos de Processamento de Linguagem Natural e Web-Semântica em um Ambiente Educacional, para Apoio aos Alunos](#)  
*Lucas L. Martins, UNISINOS*
5. [RAA – Um Revisor de Alinhamento Sentencial Automático em Corpora Paralelos](#)  
*Davi Diório Mendes, ICMC-USP*
6. [Análise de Sentimento: Descrição de uma Ferramenta de Anotação de Textos Opinativos](#)  
*Marco Nemetz Bochernitsan, Faculdade de Informática-PUCRS*
7. [Um Método para o Pré-Processamento de Textos na Identificação Automática de Emoções para o Português do Brasil](#)  
*Lohann Paterno Coutinho Ferreira, PUCPR*
8. [Identificação da Redundância na Sumarização Automática Multidocumento: Explorando Métodos Superficiais](#)  
*Jackson W. C. Souza, DL-UFSCar*
9. [Investigação de Métodos de Segmentação e Agrupamento de Subtópicos para Sumarização Multidocumento](#)  
*Rafael Ribaldo, ICMC-USP*
10. [Tipologia das Expressões Adjetivais do Português do Brasil](#)  
*Amanda dos Santos Carneiro, UFSCar*
11. [Dificuldades Apresentadas por Hispanofalantes na Produção de Textos Acadêmicos em Língua Portuguesa: a Questão dos Heterogênicos](#)  
*Roana Rodrigues, DL-UFSCar*



12. [Estudo de Métodos Clássicos de Sumarização no Cenário Multidocumento Multilíngue](#)  
*Fabricao E. da S. Tosta, DL-UFSCar*
  
13. [Delineamento Conceitual de Corpus Via Indexação Léxico-conceitual: Primeiros Resultados](#)  
*Andressa C. I. Zacarias, DL-UFSCar*
  
14. [Gramática dos Nomes de Doença: Uma Proposta Inicial](#)  
*Maria Beatriz Escanhuela Sajovic, DL-UFSCar*
  
15. [Comparação de Desempenho da Análise Sintática Não Supervisionada do Português e do Inglês](#)  
*Pablo Botton da Costa, UNIPAMPA*

## IV Workshop on RST and Text Studies

O IV Workshop “RST and Discourse Studies” (<http://encontrorst2013.wix.com/encontro-rst-2013>) ocorreu em 23 de outubro de 2013 como evento satélite do 9º Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL), em Fortaleza, Brasil.

O workshop reúne pesquisadores, professores e estudantes de Linguística, Linguística Computacional ou Ciência da Informação que trabalham com a *Rhetorical Structure Theory* (MANN & THOMPSON, 1988) ou teorias e modelos afins. O objetivo é permitir a divulgação das pesquisas realizadas em diferentes instituições e grupos de pesquisas de vários países.

A primeira edição do workshop foi realizada no 16º Intercâmbio de Pesquisas em Linguística Aplicada (InPLA), em 2007, na PUC-SP. Em 2009 ocorreu a segunda edição, desta vez com o nome de “Encontro Brasileiro de RST” e vinculado ao STIL 2009, em São Carlos (<http://www.inf.ufrgs.br/stil09/>). O III Workshop “A RST e os Estudos do Texto” foi realizado também no âmbito do STIL (2011) em Cuiabá ([http://corpus.iingen.unam.mx/RST\\_2011/](http://corpus.iingen.unam.mx/RST_2011/)).

Esta edição do Workshop “RST and Discourse Studies” recebeu 12 submissões de trabalhos. Cada trabalho foi revisado por dois membros do Comitê Científico do evento seguindo os seguintes critérios: atendimento às normas do evento, originalidade do tema, adequação da fundamentação, clareza na apresentação das ideias, pertinência dos objetivos, adequação da metodologia, qualidade da análise e relevância da bibliografia. Os participantes optaram por um dos dois tipos de trabalhos aceitos pela organização: artigos completos de até oito páginas de conteúdo (incluindo tabelas e figuras), com duas páginas adicionais de referências ou resumos expandidos de até quatro páginas de conteúdo e uma página adicional de referências. Os trabalhos deveriam seguir o formato da Sociedade Brasileira de Computação (SBC).

Dos doze trabalhos submetidos, nove foram aceitos para publicação de acordo com os pareceres dos revisores, dentre os quais cinco artigos completos e dois resumos expandidos (não foram recebidas as versões finais dos demais trabalhos). As apresentações ocorreram na forma de comunicações orais durante o evento.

Novamente o evento se caracterizou pela participação de pesquisadores e de alunos de diversas regiões não só do Brasil, mas também de outros países (Canadá, Espanha, França, País Basco). Dessa forma, estiveram representadas universidades brasileiras – Universidade Estadual de Maringá (UEM), Universidade Federal de São Carlos (UFSCar), Universidade Federal do Ceará (UFC), Universidade Federal de Minas Gerais (UFMG), Universidade do Vale do Rio dos Sinos (Unisinos) – e universidades estrangeiras – University of the Basque Country (UPV/EHU), Université d’Avignon et des Pays de Vaucluse (UAPV), École Polytechnique de Montréal.

Apresenta-se a seguir a lista de temas abordados pelos artigos aceitos:

Sumarização discursiva	2
Língua falada	1
Relações retóricas	3
Anotação de corpus	1

### **Referências**

MANN, W. C.; THOMPSON, S. A. Rhetorical Structure Theory: toward a functional theory of text organization. *Text*, v. 8, n. 3, p. 243-281, 1988.

## **Comitê de Organização**

### **Comitê de Programa**

Juliano Desiderato Antonio (Universidade Estadual de Maringá, Brazil)

Mikel Irukieta (University of the Basque Country, Spain)

### **Comitê Científico**

Arantza Diaz de Ilarraza (University of the Basque Country, Spain)

Horacio Saggion (Universitat Pompeu Fabra, Spain)

Iria da Cunha (Universitat Pompeu Fabra, Spain)

Juan-Manuel Torres (Laboratoire Informatique d'Avignon, France)

Juliano Desiderato Antonio (Universidade Estadual de Maringá, Brazil)

Lucia Rino (Universidade Federal de São Carlos, Brazil)

Maite Taboada (Simon Fraser University, Canada)

Maria Beatriz Nascimento Decat (Universidade Federal de Minas Gerais, Brazil)

Maria Eduarda Giering (Universidade do Vale do Rio dos Sinos, Brazil)

Mikel Irukieta (University of the Basque Country, Spain)

Thiago Pardo (Universidade de São Paulo, Brazil)

## Artigos

1. [A RST em análises do português em uso: uma trajetória](#)

*Maria Beatriz Nascimento Decat*

2. [Plano de texto, conforme a Análise Textual dos Discursos \(ATD\), e organização retórica, de acordo com a RST: uma relação possível?](#)

*Maria Eduarda Giering, Elizabet Beatriz Follmann, Érica Ehlers Iracet*

3. [A Relação Retórica de Condição na fala de Mariana \(MG\): uma análise inicial](#)

*Danúbia Aline Silva*

4. [A RST e os estudos da língua falada: uma investigação em elocuições formais do português brasileiro](#)

*Juliano Desiderato Antonio*

5. [Insights for better RST segmentation of texts in Portuguese?](#)

*Lucia Helena Rino, Ariani Di Felippo, Thiago A. Salgueiro Pardo*

6. [The RST Basque TreeBank: an online search interface to check rhetorical relations](#)

*Mikel Iruskieta, Itziar Gonzalez-Dios, María Jesús Aranzabe, Mikel Lersundi, Arantza Diaz de Ilarraza, Oier Lopez de Lacalle*

7. [Segmentação discursiva automática: uma avaliação preliminar em francês](#)

*Rémy Saksik, Alejandro Molina, Andréa Carneiro Linhares, Juan-Manuel Torres-Moreno*

SPONSORED BY



ORGANIZED BY



FUNDAÇÃO EDSON QUEIROZ  
UNIVERSIDADE DE FORTALEZA  
ENSINANDO E APRENDENDO



UNIVERSIDADE  
FEDERAL DO CEARÁ

SUPPORTED BY



Prefeitura de  
**Fortaleza**



GOVERNO DO  
ESTADO DO CEARÁ



F U N C A P



Ministério da  
Educação

