

Volume 10, Número 2 – 2018

LinguaMÁTICA

ISSN: 1647-0818

Editores

Alberto Simões

José João Almeida

Xavier Gómez Guinovart

Conteúdo

POP: Por Outras Palavras

Alinhamentos Parafrásticos PE–PB de Predicados Verbais com o Pronome Clítico <i>lhe</i> <i>Ida Rebelo-Arnold, Anabela Barreiro, Paulo Quaresma & Cristina Mota</i>	3
Construções Conversas do Português do Brasil: Descrição e Classificação Iniciais <i>Nathália Perussi Calcia & Oto Araujo Vale</i>	13
Paráfrase de Advérbios terminados em <i>–mente</i> em Português <i>Jorge Baptista</i>	21
Deteção de Paráfrases na Língua Portuguesa usando Sentence Embeddings <i>Marlo Souza & Leandro M. P. Sanches</i>	31
Identificação de Paráfrase em Ferramentas de Resolução de Coreferência <i>Bernardo S. Consoli, Joaquim F. dos Santos Neto, Sandra C. de Abreu & Renata Vieira</i>	45
Parafraseamento Automático de Registo Informal em Formal na Língua Portuguesa <i>Anabela Barreiro, Ida Rebelo-Arnold, Jorge Baptista, Cristina Mota & Isabel Garcez</i>	53
Explorando Métodos Non-Supervisados para Calcular a Similitude Semântica Textual <i>Pablo Gamallo & Martín Pereira-Fariña</i>	63

POP – Por Outras Palavras

Este volume contém os trabalhos apresentados no POP – Por Outras Palavras, o 1º seminário sobre Ferramentas e Recursos Linguísticos para Parafraseamento em Português, realizado a 24 de Setembro de 2018 em Canela (RS), Brasil. O seminário teve como objetivo reunir investigadores linguistas e que trabalham na área do Processamento de Linguagem Natural interessados em discutir novas ideias sobre o desenvolvimento e uso de recursos linguísticos orientados para parafraseamento em português com aplicações do mundo real.

As paráfrases são extremamente importantes na comunicação humana, tanto na produção como na compreensão da linguagem, e assumem um papel cada vez mais importante em atividades e projetos de investigação. Diversas experiências linguísticas mostraram a viabilidade de usar recursos parafrásticos numa ampla variedade de aplicações de software, pois permitem reconhecer e gerar formas equivalentes de expressar o mesmo conteúdo, permitindo que os sistemas forneçam ao utilizador sugestões para dizer e escrever a mesma coisa / ideia por outras palavras, aumentar a fluência, a criatividade e a diversidade estilística. No atual estágio de desenvolvimento, os sistemas de parafraseamento exigem conhecimento linguístico e “inteligência” sensível ao contexto para “compreender” e reconhecer uma ampla variedade de expressões. Para o português, a utilidade dos recursos parafrásticos já foi explorada em cenários aplicativos, como um sistema de diálogo, para aumentar o conhecimento linguístico de um agente virtual inteligente, em ferramentas de sumarização e simplificação e também em ferramentas que visam obter tradução automática de qualidade superior. No entanto, é necessária mais investigação para a viabilidade e sucesso de um sistema de parafraseamento a longo prazo nas áreas de produção e revisão de texto, nomeadamente no desenvolvimento e melhoria de plataformas de autoria online, desenvolvendo programas interativos para ajudar os estudantes de português como língua estrangeira a produzir frases diferentes mas equivalentes ou até para estudantes nativos, para os auxiliar nas tarefas de produção e revisão dos seus textos.

Ao propor o seminário POP, queríamos (i) reunir investigadores com interesse no campo das paráfrases, e com especial enfoque no português, para aprender e partilhar informação sobre o tema; (ii) reunir um conjunto de artigos de boa qualidade que discutam as últimas tendências na área e contribuam para melhorar o estado da arte das paráfrases em português; (iii) trocar ideias e disseminar as melhores práticas para ajudar a fomentar a investigação nesta área; (iv) fomentar uma convergência de esforços de investigação para uma definição consensual dos métodos científicos, e incentivar a cooperação internacional, a fim de alcançar estratégias comuns que respondam às necessidades tecnológicas atuais; (v) discutir novas metodologias, como redes neuronais, etc., e aprender a combinar essas metodologias com esforços linguísticos; (vi) discutir desafios futuros e trocar informação sobre aspetos científicos e tecnológicos; (vii) incentivar e reforçar a criação de corpora paralelos de paráfrases para o português como conjuntos de dados para a coleta de recursos de alinhamento

parafrástico para treino e teste de sistemas de parafraseamento; e (viii) localizar fontes de financiamento para impulsionar ainda mais a investigação, apoiar a inovação e desenvolver esta tecnologia capacitante essencial.

O Comité do Programa era composto por 22 membros de Portugal (8), Brasil (7), Espanha (4), França (2) e Noruega (1), e todos os membros são especialistas de renome em Processamento de Linguagem Natural, Linguística Computacional, Engenharia da Linguagem, e áreas afins, com ampla experiência no processamento da língua portuguesa e especificamente em tópicos relacionados à paráfrase.

Os organizadores do seminário POP gostariam de reconhecer publicamente várias instituições e pessoas cuja ajuda foi imprescindível para o sucesso do seminário: a Organização do PROPOR'2018, por aceitar a proposta de integrar o POP nos eventos satélite da principal conferência internacional sobre Processamento da Língua Portuguesa, bem como pelo seu apoio constante e colaboração; todos os membros do Comité de Programa, cuja colaboração inestimável foi fundamental para o sucesso do seminário e para a sua qualidade científica; as diferentes instituições que apoiaram, de diferentes formas, a participação de autores e organizadores na conferência.

*Anabela Barreiro
Jorge Baptista
Renata Vieira
Paulo Quaresma*

Comissão de Programa POP@PROPOR2018

Sandra Aluísio, Universidade de São Paulo/ICMC/NILC (Brasil)
Jorge Baptista, Universidade do Algarve L2F/INESC-ID, Lisboa (Portugal)
Anabela Barreiro L2F/INESC-ID Lisboa (Portugal)
Lucília Chacoto, Universidade do Algarve (Portugal)
Luísa Coheur, IST/INESC-ID, Lisboa (Portugal)
Ariani Di Felippo, Universidade Federal de São Carlos (Brasil)
Claudia Freitas, PUC-Rio (Brasil)
Pablo Gamallo, Universidade de Santiago de Compostela (Espanha)
Hugo Gonçalo Oliveira, Universidade de Coimbra (Portugal)
Éric Laporte, Universidade de Paris-Est Marnela-Vallée (França)
Belinda Maia, Universidade do Porto (Portugal)
Thiago Pardo, Universidade de São Paulo (Brasil)
Paulo Quaresma, Universidade de Évora (Portugal)
Ama manda Rassi, Lionbridge (Brasil)
Ricardo Ribeiro, INESC ID Lisboa/ISCTE-IUL (Portugal)
Paolo Rosso, Universidade Politécnica de Valência (Espanha)
Diana Santos, Universidade de Oslo (Noruega)
Max Silberztein, Universidade de Franche-Comté (França)
Alberto Simões, 2Ai Lab - IPCA Braga (Portugal)
Oto Vale, Universidade Federal de São Carlos (Brasil)
Renata Vieira, Pontifícia Universidade Católica - Rio Grande do Sul, PA (Brasil)

Comissão Científica

Alberto Álvarez Lugrís,
Universidade de Vigo

Alberto Simões,
Universidade do Minho

Aline Villavicencio,
Universidade Federal do Rio Grande do Sul

Álvaro Iriarte Sanroman,
Universidade do Minho

Ana Frankenberg-Garcia,
University of Surrey

Anselmo Peñas,
Univers. Nac. de Educación a Distancia

Antón Santamarina,
Universidade de Santiago de Compostela

Antoni Oliver González,
Universitat Oberta de Catalunya,

Antonio Moreno Sandoval,
Universidad Autónoma de Madrid

António Teixeira,
Universidade de Aveiro

Arantza Díaz de Ilarraza,
Euskal Herriko Unibertsitatea

Arkaitz Zubiaga,
Dublin Institute of Technology

Belinda Maia,
Universidade do Porto

Carmen García Mateo,
Universidade de Vigo

Diana Santos,
Linguatca/Universidade de Oslo

Ferran Pla,
Universitat Politècnica de València

Gael Harry Dias,
Université de Caen Basse-Normandie

Gerardo Sierra,
Univers. Nacional Autónoma de México

German Rigau,
Euskal Herriko Unibertsitatea

Helena de Medeiros Caseli,
Universidade Federal de São Carlos

Horacio Saggion,
University of Sheffield

Hugo Gonçalo Oliveira,
Universidade de Coimbra

Iñaki Alegria,
Euskal Herriko Unibertsitatea

Irene Castellón Masalles,
Universitat de Barcelona

Joaquim Llisterri,
Universitat Autònoma de Barcelona

José João Almeida,
Universidade do Minho

José Paulo Leal,
Universidade do Porto

Joseba Abaitua,
Universidad de Deusto

Juan-Manuel Torres-Moreno,
Lab. Informatique d'Avignon - UAPV

Kepa Sarasola,
Euskal Herriko Unibertsitatea

Laura Plaza,
Complutense University of Madrid

Lluís Padró,
Universitat Politècnica de Catalunya

Marcos Garcia,
Universidade da Corunha

María Inés Torres,
Euskal Herriko Unibertsitatea

Maria das Graças Volpe Nunes,
Universidade de São Paulo

Mercè Lorente Casafont,
Universitat Pompeu Fabra

Miguel Solla Portela,
Universidade de Vigo

Mikel Forcada,
Universitat d'Alacant

Pablo Gamallo Otero,
Universidade de Santiago de Compostela

Patrícia Cunha França,
Universidade do Minho

Rui Pedro Marques,
Universidade de Lisboa

Susana Afonso Cavadas,
University of Sheffield

Tony Berber Sardinha,
Pontifícia Univ. Católica de São Paulo

Xavier Gómez Guinovart,
Universidade de Vigo

