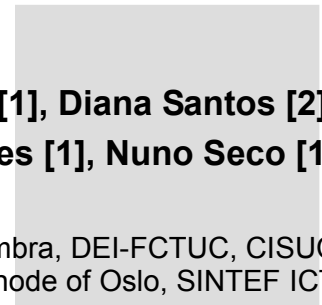


PAPEL: a lexical ontology for Portuguese



CMS, Cognitive and MediaSystems Group, University of Coimbra



**Hugo Gonalo Oliveira [1], Diana Santos [2],
Paulo Gomes [1], Nuno Seco [1]**

1 - Linateca, node of Coimbra, DEI-FCTUC, CISUC
2 - Linateca, node of Oslo, SINTEF ICT

Contents



- **Linguatoteca – node of Coimbra**
- **Objectives**
- **MRDs (as a source of semantic knowledge)**
- **Similar resources**
- **Building PAPEL**
- **Detailed example**
- **Further work/possible directions**

Liguatoteca – node of Coimbra

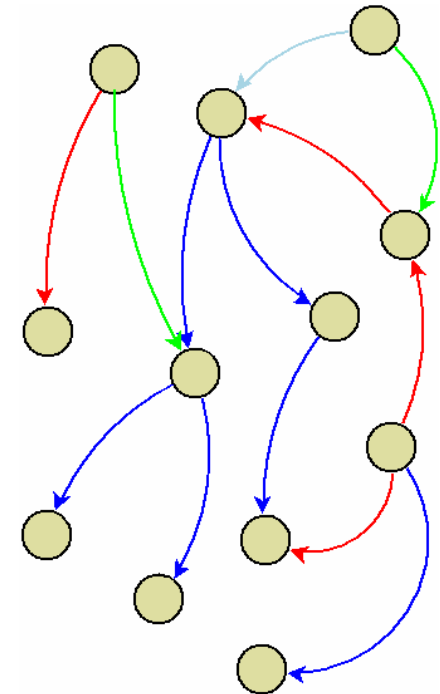
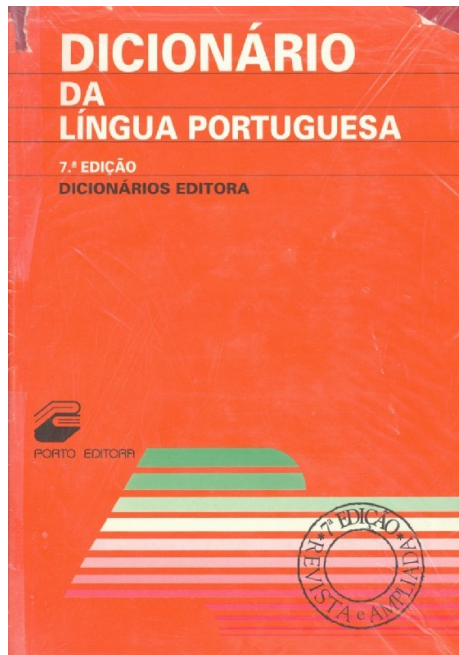


- PAPEL: a lexical ontology for Portuguese
- Floresta Sintá(c)tica: a treebank for Portuguese
- HAREM co-organization
 - named entity recognition
 - event in March 2008
 - new track ReReLEM
 - more info: http://acdc.liguatoteca.pt/aval_conjunta/HAREM/

Objectives



■ PAPEL stands for Palavras Associadas Porto Editora Linguateca



MRDs as a source of semantic knowledge

- Since 1970's (Calzolari 1977, Amsler 1981, ...)
- Typical definition structure: *genus + differentia*
 - *genus*: the superordinate concept
 - *differentia*: properties for distinction between instances of the same superordinate concept.
- Restricted and predictable vocabulary

Similar resources



- Princeton WordNet (Miller 1990)
 - freely available
 - widely used by NLP researchers
 - created “from the scratch”
 - synsets + relations
 - different kinds of relations for words of different grammatical categories
 - Hypernymy/Hyponymy, Meronymy/Holonymy.

Similar resources



- Microsoft's MindNet (Richardson et al. 1998)
 - automatically created with tools to extract relations from MRDs
 - a large amount of relations: Attributive modification, Causation, Classifier, Goal, Hypernymy, Intensifier, Location, Manner, Means...
 - a web interface: MNEX

Similar resources



- FrameNet (Baker et al. 1998)
 - description of semantic frames
- WordNet.BR (Dias da Silva 2004)
 - as far as we know only contains the Synonymy and Antonymy relations.
- WordNet.PT (Marrafa et al. 2006)
 - few information shared
 - website/search interface has not been working for the past year...

Building PAPEL



Intended relations

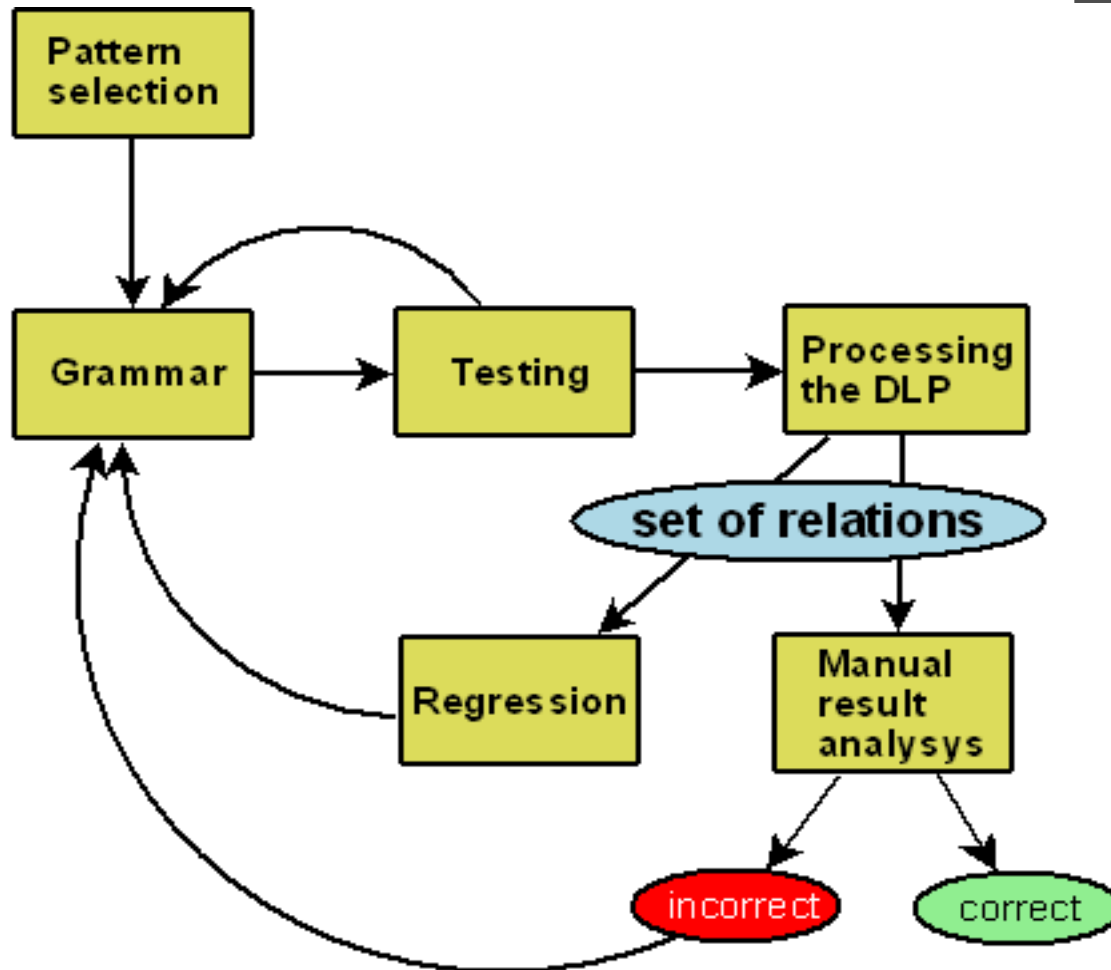
Relation	Name	Description
←Hyper→	Hypernymy	Is-a relation
←Part→	Meronymy	Part of
←Cause→	Causation	Caused by
←Attrib→	Attributive modification	Modification through a property or quality
←Loc→	Location-of	Place where something is/occurs
←Means→	Means	Used to accomplish some task
←Manner→	Manner	Way of doing something
←Goal→	Goal	Objective, aim or purpose

Building PAPEL



- Quantitative studies about the patterns used in the definitions.
- Using the chart parser PEN with specific grammars to extract relations.
- Relations between words defined and entities inside their definition.

Building PAPEL



Detailed example



- **Causation:** relation between an agent (the causer) and a result (the caused)
- The patterns used are based on verbs: ***causar, originar, provocar, producir*** and ***resultar***.

■ ■ ■ Detailed example



■ Direct relation: **Cause-of**

■ Examples of patterns:

– *que causa/original/provoca CAUSADOR_DE*

- tentação (s.f.) – coisa ou pessoa **que provoca** desejo
- tentação **Causador-de** desejo

– *causar/originar/provocar CAUSADOR_DE*

- penalizar (v. tr.) – **causar** pena, dor, aflição
- penalizar **Causador-de** {pena, dor, aflição}

■ ■ ■ Detailed example



■ Inverse relation: **Caused-by/Result-of**

■ Examples of patterns:

- *causado/a/os/as* **FREQ*** **PREP** **EFEITO_DE**
- *originado/a/os/as* **FREQ*** **PREP** **EFEITO_DE**
- *provocado/a/os/as* **FREQ*** **PREP** **EFEITO_DE**
- *produzido/a/os/as* **FREQ*** **PREP** **EFEITO_DE**
 - concussão (s.f.) – choque violento
originado por uma explosão
 - concussão **Causado-por** explosão

■ ■ ■ Detailed example



■ Examples of patterns (cont.):

– *devido a/ao/à/aos/às EFEITO_DE*

- olheiras (s. f) - círculos arroxeados que aparecem por vezes à volta dos olhos, geralmente **devido a** cansaço, insónia ou doença
- olheiras **Causado-por** {cansaço, insónia, doença}

– *efeito PREP EFEITO_DE*

- cálculo (s.m.) - acto ou **efeito de** calcular
- cálculo **Causado-por** calcular

■ ■ ■ Detailed example



■ 7253 hits

■ 4.5% errors

■ Examples of current problems:

– Specific relations inside the definition:

- estetoscópio (s.m) – instrumento para auscultar a respiração, as batidas do coração e outros sons produzidos pelo corpo.

– Negation of patterns:

- respeitar (v. tr.) – não **causar** dano

Further work/possible directions



- Improvements to PEN
- Using a broad-coverage parser like PALAVRAS (Bick 2000)
- Construction of the network:
 - grouping the words into synsets
 - word sense desambiguation

The end...



Acknowledgements

- *This work was done in the scope of the Linguateca project, promoted by FCCN, jointly funded by the Portuguese Government and the European Union (FÉDER and FSE) under contract ref. POSC/339/1.3/C/NAC*

Questions/suggestions?

References



Nicoletta Calzolari. An empirical approach to circularity in dictionary definitions. In *Cahiers de Lexicologie*, pages 118-128, 1977.

Robert A. Amsler. A taxonomy for english nouns and verbs. In *Proceedings of the 19th annual meeting on Association for Computational Linguistics*, pages 133-138, Morristown, NJ, USA, 1981. Association for Computational Linguistics.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. Introduction to wordnet: An on-line lexical database*. *Int J Lexicography*, 3(4):235244, January 1990.

Stephen D. Richardson, William B. Dolan, and Lucy Vanderwende. Mindnet: acquiring and structuring semantic information from text. In *Proceedings of the 17th international conference on Computational linguistics*, pages 10981102, Morristown, NJ, USA, 1998. Association for Computational Linguistics.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics*, pages 8690, Morristown, NJ, USA, 1998. Association for Computational Linguistics.

Bento C. Dias da Silva. Wordnet.br: An exercise of human language technology research. *Revista PaLavra*, no. 12, 2004. Série Linguagem, Processamento Automático do Português:1524, 2004.

P. Marrafa, R. Amaro, R. P. Chaves, S. Lourosa, C. Martins, and S. Mendes. Wordnet.pt new directions. In Petr Sojka, Key-Sun Choi, Christiane Fellbaum, Piek Vossen (eds.), *Proceedings of GWC'06: 3rd International Wordnet Conference*, Jeju Island, Korea, pages 319320, 2006.

Eckhard Bick. *The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. PhD thesis, Arhus University, Arhus, 2000.