

Semiautomatic Generation of a General Language Ontology

Leonardo Zilio (PPG-Letras/UFRGS)

Rodrigo Wilkens (PPG-Computação/UFRGS)

Kassius Vargas Prestes (PPG-Computação/UFRGS)

Aline Villavicencio (PPG-Computação/UFRGS)

Topics

PAPEL

Tool developed for the validation process

Human validation

PAPEL

Previous Work

- PAPEL – Palavras Associadas Porto Editora Linguateca (Oliveira et al., 2008)
 - Development of a general language comprehensive ontology
 - Series of semantic relations semiautomatically extracted from a general language dictionary
 - Use of language patterns for the extraction
 - Not completely validated
 - Around 50% of synonyms are potentially wrong (Oliveira et al., 2009)

Lists of Relations

- Each type of relation was stored in a TXT file
 - Example: Synonyms

hesitante SINONIMO_DE reticente

fugir SINONIMO_DE escapular

insensível SINONIMO_DE neve

patife SINONIMO_DE dúzia

impedir SINONIMO_DE interdizer

manchar SINONIMO_DE enodoar

Semantic Relations in PAPEL

- 60,000+ hypernyms
- 70,000+ synonyms

repartir	SYNONYM OF	partilhar
vasqueiro	QUALITY OF SOMETHING THAT CAUSES	vasca
vazar	ACTION THAT CAUSES	vazão
cabo	PART OF	vassoura
navio	HYPERNYM OF	veleiro

Selected Subset of Synonym Relations

- Nouns
- 36,504 relations

Examples of Problematic Relations

- Hypernym:
 - distribuição HIPERONIMO_DE detalhe
distribution HYPERNYM_OF detail
- Synonym:
 - patife SINONIMO_DE dúzia
rascal SYNONYM_OF dozen

**Tool developed for the
validation process**

Computational Processing of PAPEL's List of Synonyms

- 1 - Each line of the TXT files is processed
- 2 - Several empty sets are created to allocate the synonyms
- 3 - For each line:
 - If one of the nouns in a line is already present in only one previously existent set, the other noun is inserted in this set
 - If none is present in existent sets, a new one is created with both nouns
 - If both are present, then the ambiguous relation is presented to a human validator

Validation

- Development of a user-friendly interface

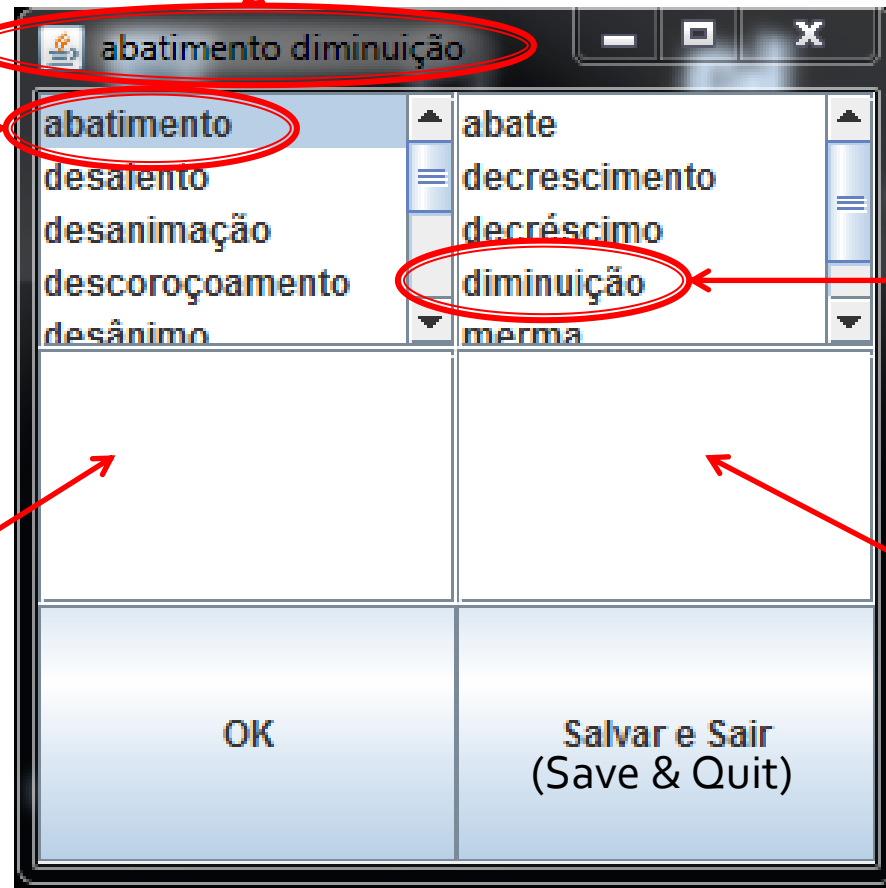


Trying to Explain How it Works

Both synonyms

First set of synonyms in which one of them is present

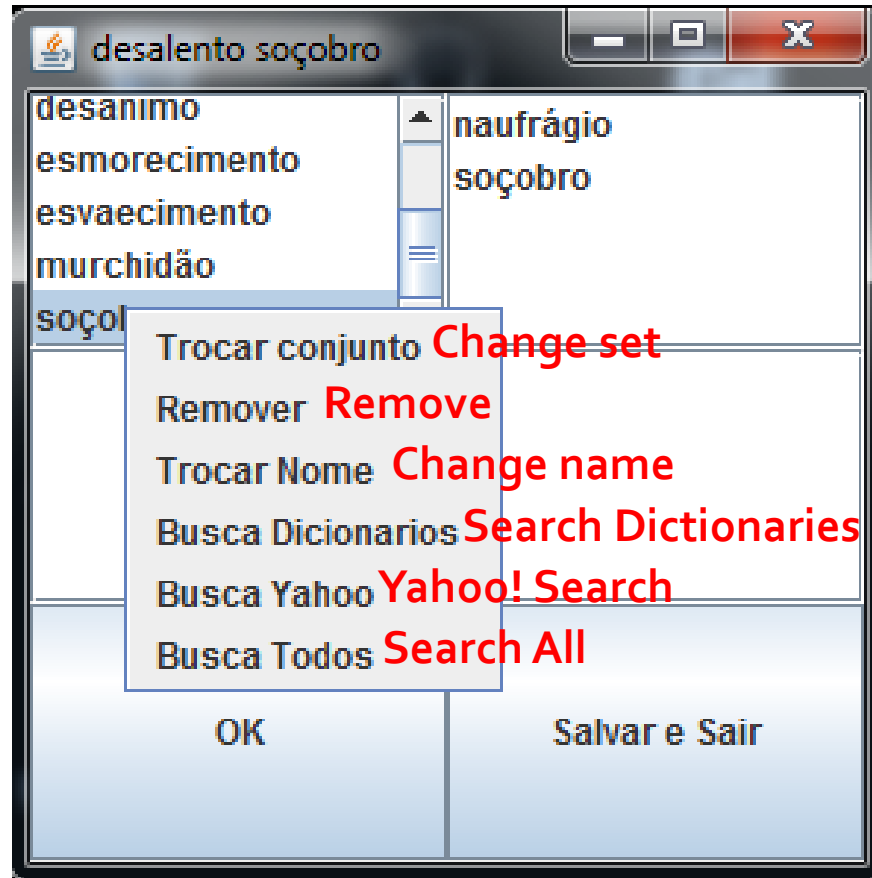
Second set of synonyms in which one of them is present



Empty set, in case a bigger reorganization is necessary

Empty set, in case a bigger reorganization is necessary

Options



Human Validation

Tasks

- Identify the senses of the pair of synonyms
- Identify which one of the senses is more relevant
- Evaluate which one of the sets of synonyms is more adequate (if any) for the subject pair
- Correction of existing sets or creation of new ones

Houaiss Dictionary

Dicionário eletrônico Houaiss da língua portuguesa 1.0

Arquivo Editar Exibir Ferramentas Ajuda

abatemento

abatemento

Acepções

- **substantivo masculino**
- ação, processo ou efeito de abater(-se)
- 1 ato ou efeito de descer, arriar, baixar
- 2 lançamento por terra; derrubada, abate
- 3 processo de matança de animais destinados ao consumo
- 4 redução, diminuição ou desconto em valor, preço, intensidade, altura etc.
- 5 diminuição ou perda de força física; enfraquecimento, definhamento
- 6 prostração moral ou acabrunhamento; entristecimento
- 6.1 sentimento de humilhação ou de vexação
- 7 **Rubrica: armamento.**
queda progressiva do projétil pela ação da gravidade
- 8 **Rubrica: construção.**
queda ou desmoronamento; demolição de um edifício
- 9 **Rubrica: engenharia.**
relação entre a flecha de um arco ou abóbada e a sua abertura; abaulamento
- 10 **Rubrica: engenharia.**
redução de altura sofrida pelo cone de concreto fresco no ensaio de consistência
- 11 **Rubrica: termo de marinha.**
ângulo formado pela quilha do navio e sua esteira, quando o vento e o mar ferem obliquamente as velas ou o costado, da perpendicular para a vante
- 12 **Rubrica: termo de marinha.**
diferença entre o rumo de governo e a direção real de avanço do navio; afastamento
- 13 **Rubrica: termo de marinha.**

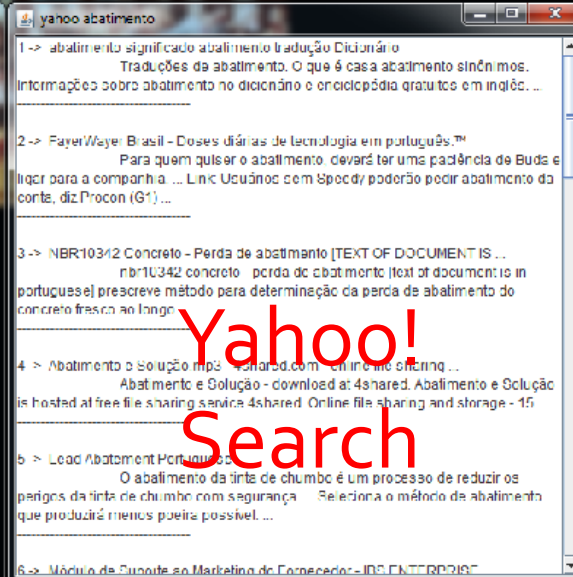
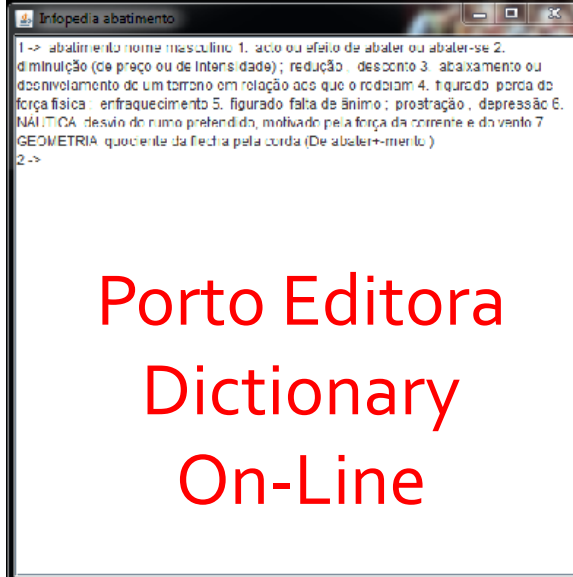
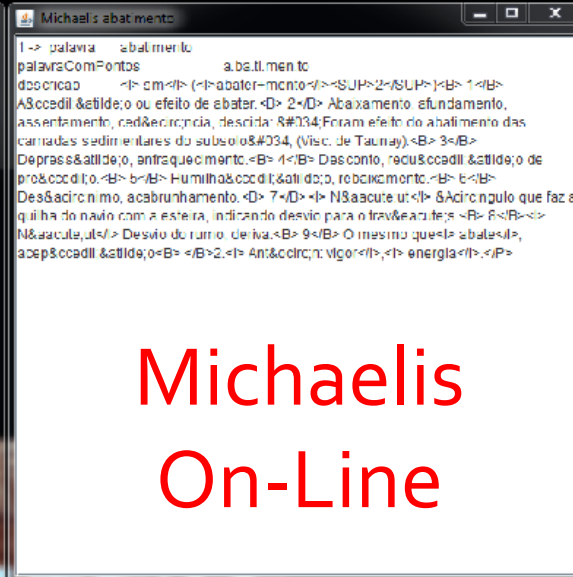
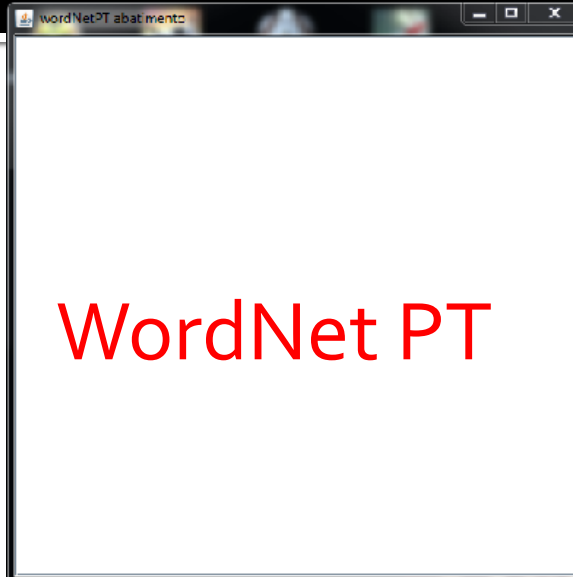
Search Options

WordNet PT

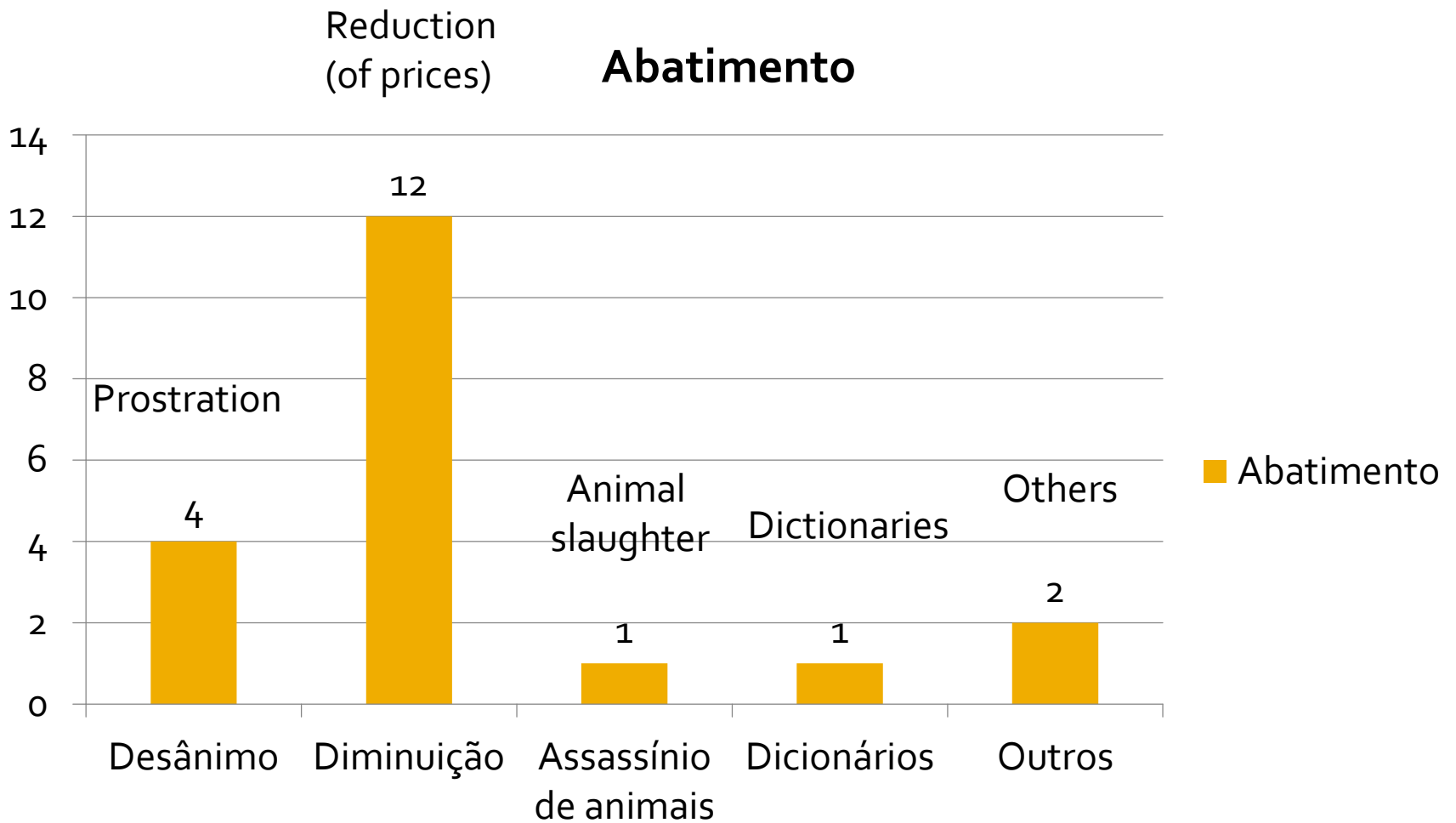
Michaelis
On-Line

Porto Editora
Dictionary
On-Line

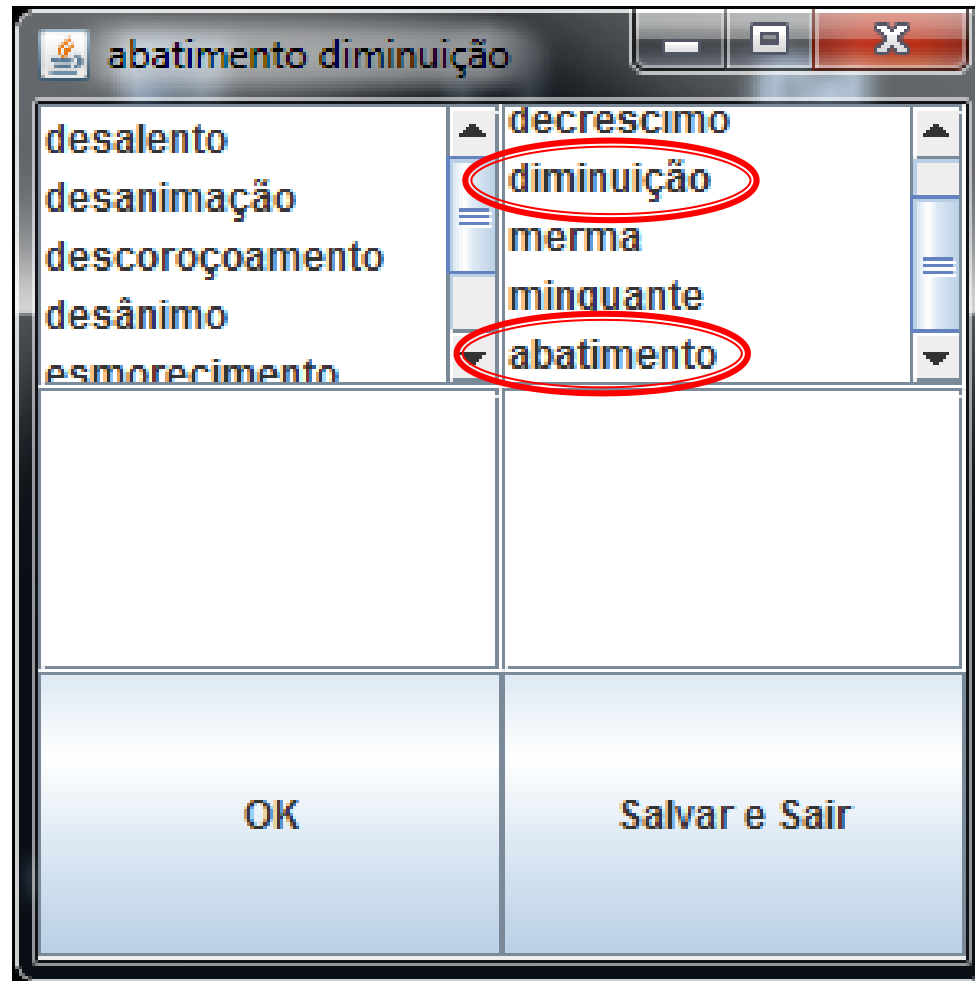
Yahoo!
Search




Yahoo!



Decision



Results

- 36,504 original relations
 - 20,096 semiautomatically validated relations
- 

Final Considerations

- Comprehensive validation of the lexical resource, in spite of its size
- Good base for a general language ontology in the steps of WordNet (Fellbaum, 1998)
- Validation of hypernyms as next step
- Congregation of both validated resources into one

References

- FELLBAUM, C. (Ed.) (1998) *WordNet: an electronic lexical database*. MIT Press.
- OLIVEIRA, H. G.; SANTOS, D.; GOMES, P.; SECO, N. (2008). Papel: A dictionary-based lexical ontology for portuguese. In: TEIXEIRA, A.; DE LIMA, V. L. S.; DE OLIVEIRA, L. C.; QUARESMA, P. (Eds.) *Proceedings of Computational Processing of the Portuguese Language (PROPOR)*, volume 5190 of LNAI, p. 31–40. Springer.
- OLIVEIRA, H. G.; SANTOS, D.; GOMES, P. (2009). Extracção de relações semânticas entre palavras a partir de um dicionário: o papel e sua avaliação. STIL 2009, *Linguamática*, p. 77–93.

Acknowledgements

- This research was partly funded by the following projects:
 - CAPES-COFECUB 707/11
 - CNPq 551964/2011-1
 - CNPq 309569/2009-5.

Muito Obrigado!

Leonardo Zilio

lzilio@ig.com.br