Studying the Acquisition of WordNet Relations in Portuguese from Pretrained MLMs

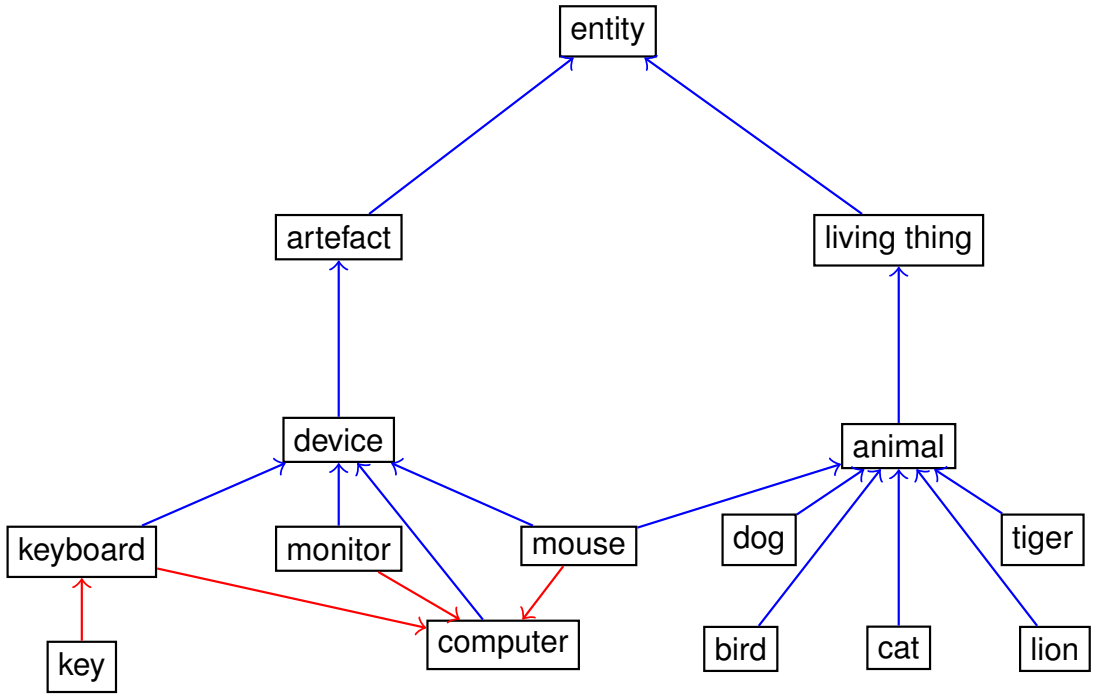
Hugo Gonçalo Oliveira

CISUC, Department of Informatics Engineering
University of Coimbra

January, 2023

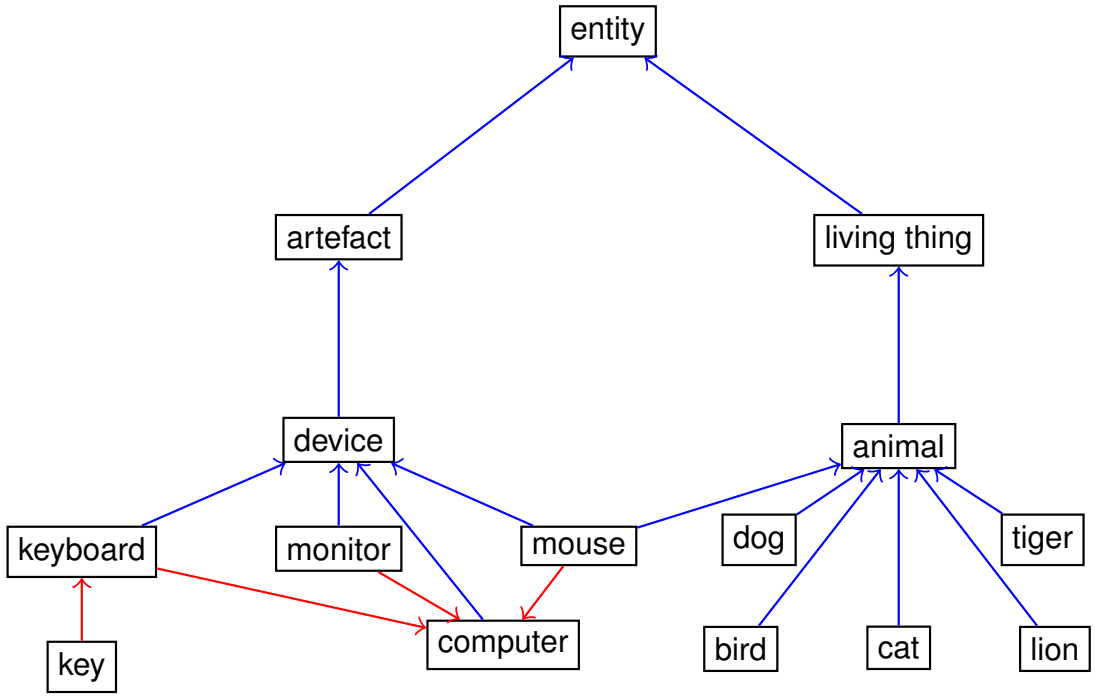
Motivation

- A **wordnet** can be seen as a graph...
  - **Nodes** = concepts, possibly denoted by words
  - **Edges** = (**explicit**) relations between nodes



Motivation

- A **wordnet** can be seen as a graph...
  - **Nodes** = concepts, possibly denoted by words
  - **Edges** = (**explicit**) relations between nodes



- In many cases, creation is **manual**!

Relation Acquisition from Corpora

- In order to automatise / accelerate the creation of wordnets...

Relation Acquisition from Corpora

- In order to automatise / accelerate the creation of wordnets...

Hearst (1992) patterns

NP , NP* , (and|or) other NP_H
NP_H such as , NP* , (or|and) NP
such NP_H as , NP* , (or|and) NP
NP_H , including NP,* (or|and) NP
NP_H , especially NP,* (or|and) NP

$\Rightarrow$ hyponymOf(NP, NP_H)

Relation Acquisition from Corpora

- In order to automatise / accelerate the creation of wordnets...

Hearst (1992) patterns

```
NP , NP* , (and|or) other NP_H
NP_H such as , NP* , (or|and) NP
such NP_H as , NP* , (or|and) NP
NP_H , including NP,* (or|and) NP
NP_H , especially NP,* (or|and) NP
```

$\Rightarrow$ hyponymOf(NP, NP_H)

* *... works by such authors as Herrick, Goldsmith, and Shakespeare.*

* *... Bruises, wounds, broken bones or other injuries ...*

Relation Acquisition from Corpora

- In order to automatise / accelerate the creation of wordnets...

Hearst (1992) patterns

```
NP , NP* , (and|or) other NP_H
NP_H such as , NP* , (or|and) NP
such NP_H as , NP* , (or|and) NP
NP_H , including NP,* (or|and) NP
NP_H , especially NP,* (or|and) NP
```

$\Rightarrow$ hyponymOf(NP, NP_H)

* *... works by such authors as Herrick, Goldsmith, and Shakespeare.*

$\rightarrow$ hyponym(Herrick, author), hyponym(Goldsmith, author), hyponym(Shakespeare, author)

* *... Bruises, wounds, broken bones or other injuries ...*

$\rightarrow$ hyponym(bruise, injury), hyponym(wound, injury), hyponym(broken bone, injury)

Relation Acquisition from Corpora

- In order to automatise / accelerate the creation of wordnets...

Hearst (1992) patterns

```
NP , NP* , (and|or) other NP_H
NP_H such as , NP* , (or|and) NP
such NP_H as , NP* , (or|and) NP
NP_H , including NP,* (or|and) NP
NP_H , especially NP,* (or|and) NP
```

$\Rightarrow$ hyponymOf(NP, NP_H)

* *... works by such authors as Herrick, Goldsmith, and Shakespeare.*

$\rightarrow$ hyponym(Herrick, author), hyponym(Goldsmith, author), hyponym(Shakespeare, author)

* *... Bruises, wounds, broken bones or other injuries ...*

$\rightarrow$ hyponym(bruise, injury), hyponym(wound, injury), hyponym(broken bone, injury)

Learn more patterns automatically

- Using WordNet relations as seeds...
  - distant supervision (Snow et al., 2005)
  - weak supervision (Pantel and Pennacchiotti, 2006)

Masked Language Model

- BERT (Devlin et al., 2019) is a Transformer-based model pretrained in the **Masked Language Modelling** task...

Masked Language Model

- BERT (Devlin et al., 2019) is a Transformer-based model pretrained in the **Masked Language Modelling** task...
  - `Global [MASK] Conference 2023.`

Masked Language Model

- BERT (Devlin et al., 2019) is a Transformer-based model pretrained in the **Masked Language Modelling** task...
  - `Global [MASK] Conference 2023.`
    - development (0.077)
    - security (0.056)
    - leadership (0.039)

Masked Language Model

- BERT (Devlin et al., 2019) is a Transformer-based model pretrained in the **Masked Language Modelling** task...
  - `Global [MASK] Conference 2023.`
    - development (0.077)
    - security (0.056)
    - leadership (0.039)
- Given text may instantiate a **lexical pattern that denotes a semantic relation**...

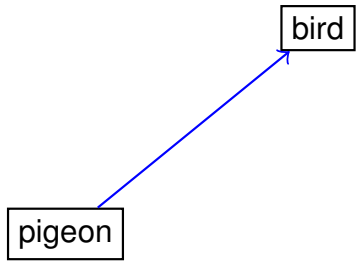
Masked Language Model

- BERT (Devlin et al., 2019) is a Transformer-based model pretrained in the **Masked Language Modelling** task...
  - `Global [MASK] Conference 2023.`
    - development (0.077)
    - security (0.056)
    - leadership (0.039)
- Given text may instantiate a **lexical pattern that denotes a semantic relation**...
  - `A [MASK] is a type of bird.`

bird

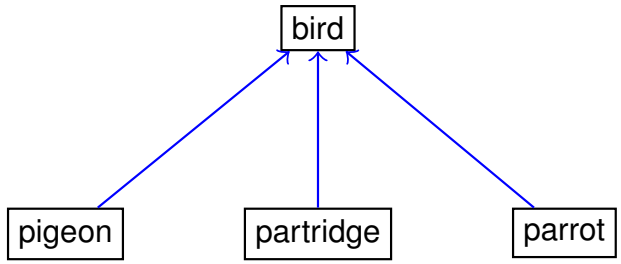
Masked Language Model

- BERT (Devlin et al., 2019) is a Transformer-based model pretrained in the **Masked Language Modelling** task...
  - `Global [MASK] Conference 2023.`
    - development (0.077)
    - security (0.056)
    - leadership (0.039)
- Given text may instantiate a **lexical pattern that denotes a semantic relation**...
  - `A [MASK] is a type of bird.`
    - pigeon (0.177)



Masked Language Model

- BERT (Devlin et al., 2019) is a Transformer-based model pretrained in the **Masked Language Modelling** task...
  - `Global [MASK] Conference 2023.`
    - development (0.077)
    - security (0.056)
    - leadership (0.039)
- Given text may instantiate a **lexical pattern that denotes a semantic relation**...
  - `A [MASK] is a type of bird.`
    - pigeon (0.177)
    - parrot (0.048)
    - partridge (0.043)
    - raven (0.037)
    - heron (0.035)



Goal

- BERT as a knowledge base? (Petroni et al., 2019)

Goal

- BERT as a knowledge base? (Petroni et al., 2019)

  ... as a Lexical Knowledge Base?

Goal

- BERT as a knowledge base? (Petroni et al., 2019)

  ... as a Lexical Knowledge Base?

  ... as a source of lexico-semantic knowledge?

  ... as a means for enriching wordnets?

Goal

- BERT as a knowledge base? (Petroni et al., 2019)

  ... as a Lexical Knowledge Base?

  ... as a source of lexico-semantic knowledge?

  ... as a means for enriching wordnets?

Possible Approach

1. Compile a set of lexical patterns for WordNet relations.
   - One fixed argument (X_1), to be replaced before prediction
   - Another to be predicted, `[MASK]`
   * e.g., for hypernymy: X_1 `is a type of [MASK].`

Goal

- BERT as a knowledge base? (Petroni et al., 2019)

  ... as a Lexical Knowledge Base?

  ... as a source of lexico-semantic knowledge?

  ... as a means for enriching wordnets?

Possible Approach

1. Compile a set of lexical patterns for WordNet relations.
   - One fixed argument (X_1), to be replaced before prediction
   - Another to be predicted, `[MASK]`
   - e.g., for hypernymy: X_1 `is a type of [MASK].`
2. Use a pretrained BERT for predicting the missing arguments

Goal

- BERT as a knowledge base? (Petroni et al., 2019)

  ... as a Lexical Knowledge Base?

  ... as a source of lexico-semantic knowledge?

  ... as a means for enriching wordnets?

Possible Approach

1. Compile a set of lexical patterns for WordNet relations.
   - One fixed argument (X_1), to be replaced before prediction
   - Another to be predicted, `[MASK]`
   * e.g., for hypernymy: X_1 `is a type of [MASK].`
2. Use a pretrained BERT for predicting the missing arguments
   - e.g., hypernyms of '*pigeon*'? (X_1 = pigeon)
   * Prompt BERT with: `pigeon is a type of [MASK].`

Goal

- BERT as a knowledge base? (Petroni et al., 2019)

  ... as a Lexical Knowledge Base?

  ... as a source of lexico-semantic knowledge?

  ... as a means for enriching wordnets?

Possible Approach

1. Compile a set of lexical patterns for WordNet relations.
   - One fixed argument (X_1), to be replaced before prediction
   - Another to be predicted, `[MASK]`
   * e.g., for hypernymy: X_1 `is a type of [MASK].`
2. Use a pretrained BERT for predicting the missing arguments
   - e.g., hypernyms of '*pigeon*'? (X_1 = pigeon)
   * Prompt BERT with: `pigeon is a type of [MASK].`
3. Assess predictions in available datasets of WordNet relations.

Goal

Instantiated to Portuguese

BERT as ...

... a Lexical Knowledge Base for Portuguese?

Goal

Instantiated to Portuguese

BERT as ...

... a Lexical Knowledge Base for Portuguese?

... a source of lexico-semantic knowlege in Portuguese?

Goal

Instantiated to Portuguese

BERT as ...

... a Lexical Knowledge Base for Portuguese?

... a source of lexico-semantic knowlege in Portuguese?

Approach

1. Compile a set of lexical patterns for WordNet relations
   - Synonymy, Antonymy, Hypernymy
   - Explore the relation-validation service VARRA (Freitas et al., 2015)

Goal

Instantiated to Portuguese

BERT as ...

... a Lexical Knowledge Base for Portuguese?

... a source of lexico-semantic knowlege in Portuguese?

Approach

1. Compile a set of lexical patterns for WordNet relations
   - Synonymy, Antonymy, Hypernymy
   - Explore the relation-validation service VARRA (Freitas et al., 2015)
2. Use a pretrained BERT for predicting the missing arguments
   - Multilingual BERT
   - Portuguese BERT (Souza et al. (2020))

Goal

Instantiated to Portuguese

BERT as ...

... a Lexical Knowledge Base for Portuguese?

... a source of lexico-semantic knowlege in Portuguese?

Approach

1. Compile a set of lexical patterns for WordNet relations
   - Synonymy, Antonymy, Hypernymy
   - Explore the relation-validation service VARRA (Freitas et al., 2015)
2. Use a pretrained BERT for predicting the missing arguments
   - Multilingual BERT
   - Portuguese BERT (Souza et al. (2020))
3. Assess predictions in available datasets.
   - B^2SG: TOEFL-like test
   - TALES: test of lexico-semantic analogies

Goal

Instantiated to Portuguese

BERT as ...

... a Lexical Knowledge Base for Portuguese?

... a source of lexico-semantic knowlege in Portuguese?

Approach

1. Compile a set of lexical patterns for WordNet relations
   - Synonymy, Antonymy, Hypernymy
   - Explore the relation-validation service VARRA (Freitas et al., 2015)
2. Use a pretrained BERT for predicting the missing arguments
   - Multilingual BERT
   - Portuguese BERT (Souza et al. (2020))
3. Assess predictions in available datasets.
   - B^2SG: TOEFL-like test
   - TALES: test of lexico-semantic analogies
4. Compare with performance of alternative methods

Prompting BERT MLM

- Manually-compiled patterns that indicate target relations in Portuguese text (Gonçalo Oliveira, 2022)
  - Synonym-of
  - Antonym-of
  - Hypernym-of, Hyponymy-of

Prompting BERT MLM

- Manually-compiled patterns that indicate target relations in Portuguese text (Gonçalo Oliveira, 2022)
  - Synonym-of
  - Antonym-of
  - Hypernym-of, Hyponymy-of
- Additional patterns adapted from VARRA (Freitas et al., 2015), a service for validating Portuguese semantic relations in corpora
  - regex $\rightarrow$ masked prompt

Relation	VARRA	Masked Prompt
Synonym-of	`[lema="PALAVRA1"] "," "isto" "é" "," [lema="PALAVRA2"]`	X_1 `, isto é, [MASK]`

Antonym-of	`[word="nem|seja|quer"] [lema="PALAVRA1"] [lema=","]* [word="nem|seja|quer"] [lema="PALAVRA2"]`	`nem` X_1 `, nem [MASK]`
		`seja` X_1 `, seja [MASK]`
		`quer` X_1 `, quer [MASK]`

Hypernym-of	`[lema="PALAVRA1"] [pos="ADJ.*"]* [lema=","]* [lema="tal"]* "como" [pos="DET.*"]* [pos="ADJ.*"]* [lema="PALAVRA2"]`	X_1 `, tal como [MASK]`
Hypernym-of	`[lema="PALAVRA2" & pos="N.*"] "e" [lema="outro"] [lema="PALAVRA1" & pos="N.*"]`	X_1 `e outro [MASK]`

B²SG[1] (Wilkens et al., 2016)

- TOEFL-like test for Portuguese
  - **Multiple-choice**: four options, only one word related to the source
  - Based on the Portuguese part of BabelNet

[1] http://www.inf.ufrgs.br/pln/resource/B2SG.zip

B²SG[1] (Wilkens et al., 2016)

- TOEFL-like test for Portuguese
  - **Multiple-choice**: four options, only one word related to the source
  - Based on the Portuguese part of BabelNet
  - Similar to the WordNet-Based Synonymy Test, covering 6 relations
    - Synonyms, Antonyms, Hypernyms, each between nouns and between verbs

Synonyms (v)				
`trancar` (lock	`barrar` block	`aviar` serve	`alienar` alienate	`progredir` progress)
`estremecer` (shudder	`tremer` tremble	`avermelhar` redden	`esconder` hide	`suspirar` sigh)
Antonyms (n)				
`esquerda` (left	`direita` right	`repressão` repression	`sétimo` seventh	`diácono` deacon)
`perfeição` (perfection	`defeito` defect	`junção` junction	`mutilação` mutilation	`prémio` premium)
Hyponyms (n)				
`matemática` (mathematics	`ciência` science	`célula` cell	`pulseira` bracelet	`libertação` release)
`chapéu` (hat	`sombreiro` sombrero	`cavalaria` cavalry	`gavião` hawk	`dissidente` dissident)

[1] `http://www.inf.ufrgs.br/pln/resource/B2SG.zip`

B²SG[1] (Wilkens et al., 2016)

- TOEFL-like test for Portuguese
  - **Multiple-choice**: four options, only one word related to the source
  - Based on the Portuguese part of BabelNet
  - Similar to the WordNet-Based Synonymy Test, covering 6 relations
    - Synonyms, Antonyms, Hypernyms, each between nouns and between verbs

Synonyms (v)				
`trancar` (lock	`barrar` block	`aviar` serve	`alienar` alienate	`progredir` progress)
`estremecer` (shudder	`tremer` tremble	`avermelhar` redden	`esconder` hide	`suspirar` sigh)
Antonyms (n)				
`esquerda` (left	`direita` right	`repressão` repression	`sétimo` seventh	`diácono` deacon)
`perfeição` (perfection	`defeito` defect	`junção` junction	`mutilação` mutilation	`prémio` premium)
Hyponyms (n)				
`matemática` (mathematics	`ciência` science	`célula` cell	`pulseira` bracelet	`libertação` release)
`chapéu` (hat	`sombreiro` sombrero	`cavalaria` cavalry	`gavião` hawk	`dissidente` dissident)

* **Accuracy:** proportion of source words for which the correctly related target is selected

[1] `http://www.inf.ufrgs.br/pln/resource/B2SG.zip`

TALES[2] (Gonçalo Oliveira et al., 2020)

- Analogy-like test of Portuguese lexico-semantic relations
  - Contents acquired from open Portuguese LKBs
  - 50 entries for each of 14 relation types, including:
    - Synonymy (nouns, verbs, adjectives), Antonymy (adjectives)
    - Hypernymy and Hyponymy (concrete nouns, abstracts nouns, verbs)

[2] https://github.com/NLP-CISUC/PT-LexicalSemantics

TALES[2] (Gonçalo Oliveira et al., 2020)

- Analogy-like test of Portuguese lexico-semantic relations
  - Contents acquired from open Portuguese LKBs
  - 50 entries for each of 14 relation types, including:
    - Synonymy (nouns, verbs, adjectives), Antonymy (adjectives)
    - Hypernymy and Hyponymy (concrete nouns, abstracts nouns, verbs)
  - Same format as BATS (Drozd et al. (2016))...

Synonym-of (n)	
`proprietário` (owner)	`dono/senhor/titular/amo` (lord/holder/master)
`sabor` (flavour)	`gosto/paladar` (taste/palate)
Antonym-of (adj)	
`novo` (young)	`velho/idoso/entradote` (old/aged/oldish)
`certo` (certain/right)	`incerto/incorreto` (uncertain/incorrect)
Hypernym-of (concrete)	
`casa` (house)	`vila/residência/cabana/estalagem/...` (vila/residence/hut/inn/...)
`livro` (book)	`volume/álbum/catecismo/missal/...` (volume/album/catechism/prayer book/...)

[2] https://github.com/NLP-CISUC/PT-LexicalSemantics

TALES[2] (Gonçalo Oliveira et al., 2020)

- Analogy-like test of Portuguese lexico-semantic relations
  - Contents acquired from open Portuguese LKBs
  - 50 entries for each of 14 relation types, including:
    - Synonymy (nouns, verbs, adjectives), Antonymy (adjectives)
    - Hypernymy and Hyponymy (concrete nouns, abstracts nouns, verbs)
  - Same format as BATS (Drozd et al. (2016))...

Synonym-of (n)	
`proprietário` (owner)	`dono/senhor/titular/amo` (lord/holder/master)
`sabor` (flavour)	`gosto/paladar` (taste/palate)
Antonym-of (adj)	
`novo` (young)	`velho/idoso/entradote` (old/aged/oldish)
`certo` (certain/right)	`incerto/incorreto` (uncertain/incorrect)
Hypernym-of (concrete)	
`casa` (house)	`vila/residência/cabana/estalagem/...` (vila/residence/hut/inn/...)
`livro` (book)	`volume/álbum/catecismo/missal/...` (volume/album/catechism/prayer book/...)

* **Accuracy:** proportion of source words for which the top prediction is a target word

* **Accuracy@10:** proportion of source words for which a target word is in the top-10 predictions

[2] `https://github.com/NLP-CISUC/PT-LexicalSemantics`

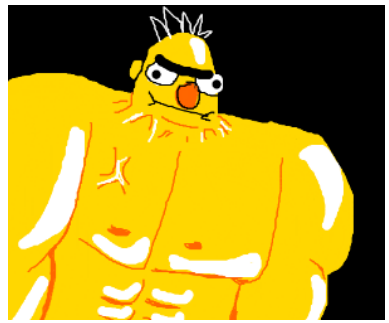
Models

- Experiments focused on three pretrained BERT models, all available from the Huggingface Hub:
  - BERTimbau (Souza et al., 2020), BERT pretrained for Portuguese
    - BERTimbau-base[3] (BERT-base), 12 layers, 110M parameters;
    - BERTimbau-large[4] (BERT-base), 24 layers, 335M parameters.
  - Multilingual BERT[5] (BERT-ML), pretrained on Wikipedia for 104 languages, 12 layers, 110M parameters.

[3] neuralmind/bert-base-portuguese-cased
[4] neuralmind/bert-large-portuguese-cased
[5] bert-base-multilingual-cased
[6] https://github.com/Qordobacode/fitbert

Models

- Experiments focused on three pretrained BERT models, all available from the Huggingface Hub:
  - BERTimbau (Souza et al., 2020), BERT pretrained for Portuguese
    - BERTimbau-base[3] (BERT-base), 12 layers, 110M parameters;
    - BERTimbau-large[4] (BERT-base), 24 layers, 335M parameters.
  - Multilingual BERT[5] (BERT-ML), pretrained on Wikipedia for 104 languages, 12 layers, 110M parameters.
- For each entry of the dataset + masked pattern for its relation...
  - B^{2}SG: FitBERT[6] instantiated with each model used for selecting the most suitable option.
  - TALES: `fill-mask` pipeline in each model, for predicting suitable tokens for the mask.



[3] `neuralmind/bert-base-portuguese-cased`
[4] `neuralmind/bert-large-portuguese-cased`
[5] `bert-base-multilingual-cased`
[6] `https://github.com/Qordobacode/fitbert`

COGNITIVE & MEDIA SYSTEMS CISUC 1 2 9 0

Results in B^{2}SG

Accuracy

Relation	PoS	Pattern		BERT-ML	BERT-base	BERT-large
Synonym-of	N	X_1 é o mesmo que [MASK] (X_1 is the same as [MASK])		0.35	0.57	**0.64**
Synonym-of	N	X_1, isto é, [MASK] (X_1, this is, [MASK])	*V*	0.33	0.58	0.62
Synonym-of	V	X_1, isto é, [MASK] (X_1, this is, [MASK])	*V*	0.32	0.50	**0.56**
Synonym-of	V	X_1, ou seja, [MASK] (X_1, i.e., [MASK])	*V*	0.49	0.54	0.37
Antonym-of	N	nem [MASK], nem X_1 (not X_1, nor [MASK])	*V*	0.44	0.76	**0.77**
Antonym-of	N	X_1 é o contrário de [MASK] (X_1 is the opposite of [MASK])		0.46	0.72	0.77
Antonym-of	V	se está a X_1 não está a [MASK] (if it is X_1, it is not [MASK])		0.46	0.60	0.62
Antonym-of	V	nem [MASK], nem X_1 (not X_1, nor [MASK])	*V*	0.29	**0.63**	0.61
Hypernym-of	N	X_1, isto é, um tipo de [MASK] (X_1, this is, a type of [MASK])	*V*	0.44	0.68	**0.71**
Hypernym-of	N	X_1, isto é, uma espécie de [MASK] (X_1, this is, a kind of [MASK])	*V*	0.41	0.63	0.70
Hypernym-of	V	a X_1 ou outras formas de [MASK] (X_1 or other forms of [MASK])	*V*	0.36	0.61	**0.66**
Hypernym-of	V	a X_1 ou outros modos de [MASK] (X_1 or other modes of [MASK])		0.37	0.57	0.61

Results in B^{2}SG

Accuracy

Relation	PoS	Pattern		BERT-ML	BERT-base	BERT-large
Synonym-of	N	X_1 é o mesmo que [MASK] (X_1 is the same as [MASK])		0.35	0.57	**0.64**
Synonym-of	N	X_1, isto é, [MASK] (X_1, this is, [MASK])	*V*	0.33	0.58	0.62
Synonym-of	V	X_1, isto é, [MASK] (X_1, this is, [MASK])	*V*	0.32	0.50	**0.56**
Synonym-of	V	X_1, ou seja, [MASK] (X_1, i.e., [MASK])	*V*	0.49	0.54	0.37
Antonym-of	N	nem [MASK], nem X_1 (not X_1, nor [MASK])	*V*	0.44	0.76	**0.77**
Antonym-of	N	X_1 é o contrário de [MASK] (X_1 is the opposite of [MASK])		0.46	0.72	0.77
Antonym-of	V	se está a X_1 não está a [MASK] (if it is X_1, it is not [MASK])		0.46	0.60	0.62
Antonym-of	V	nem [MASK], nem X_1 (not X_1, nor [MASK])	*V*	0.29	**0.63**	0.61
Hypernym-of	N	X_1, isto é, um tipo de [MASK] (X_1, this is, a type of [MASK])	*V*	0.44	0.68	**0.71**
Hypernym-of	N	X_1, isto é, uma espécie de [MASK] (X_1, this is, a kind of [MASK])	*V*	0.41	0.63	0.70
Hypernym-of	V	a X_1 ou outras formas de [MASK] (X_1 or other forms of [MASK])	*V*	0.36	0.61	**0.66**
Hypernym-of	V	a X_1 ou outros modos de [MASK] (X_1 or other modes of [MASK])		0.37	0.57	0.61

* Always higher than the random chance (25%)
  - Variable performance depending on relation, pattern, model, ...
  - BERTimbau-large generally the best option

Results in B²SG

Other methods

- Other BERT-based approaches: most similar according to...
  - `[CLS]` embeddings
  - Mean pooling of token embeddings
  - BERTimbau fine-tuned for Semantic Textual Similarity

Relation	PoS	BERT-b (patterns)	BERT-l (patterns)	BERT-b (CLS)	BERT-l (CLS)	BERT-b (tokens)	BERT-l (tokens)	BERT-STS
Synonym-of	N	0.58	0.64	0.60	0.67	0.59	0.66	0.80
Synonym-of	V	0.54	0.56	0.55	0.51	0.54	0.54	**0.75**
Antonym-of	N	0.76	0.77	0.72	0.63	0.69	0.64	0.78
Antonym-of	V	0.63	0.62	0.51	0.51	0.49	0.57	0.68
Hypernym-of	N	0.68	0.71	0.59	0.61	0.59	0.62	0.76
Hypernym-of	V	0.61	0.66	0.52	0.51	0.54	0.54	**0.71**

Results in B^{2}SG

Other methods

- Other BERT-based approaches: most similar according to...
  - `[CLS]` embeddings
  - Mean pooling of token embeddings
  - BERTimbau fine-tuned for Semantic Textual Similarity

Relation	PoS	BERT-b (patterns)	BERT-l (patterns)	BERT-b (CLS)	BERT-l (CLS)	BERT-b (tokens)	BERT-l (tokens)	BERT-STS
Synonym-of	N	0.58	0.64	0.60	0.67	0.59	0.66	0.80
Synonym-of	V	0.54	0.56	0.55	0.51	0.54	0.54	**0.75**
Antonym-of	N	0.76	0.77	0.72	0.63	0.69	0.64	0.78
Antonym-of	V	0.63	0.62	0.51	0.51	0.49	0.57	0.68
Hypernym-of	N	0.68	0.71	0.59	0.61	0.59	0.62	0.76
Hypernym-of	V	0.61	0.66	0.52	0.51	0.54	0.54	**0.71**

- Most similar according to...
  - 300-sized word embeddings, pretrained for Portuguese (Hartmann et al., 2017)

Relation	PoS	CBOW	Skip	GloVe
Synonym-of	N	0.71	**0.83**	0.81
Synonym-of	V	0.66	0.68	0.70
Antonym-of	N	0.70	0.81	**0.83**
Antonym-of	V	0.67	0.69	**0.71**
Hypernym-of	N	0.65	0.76	**0.80**
Hypernym-of	V	0.64	0.66	0.70

Results in TALES

Accuracy of best patterns

Relation	PoS	Pattern		BERT-ML	BERT-base	BERT-large
Synonym-of	N	X_1 é sinónimo de [MASK] (X_1 is a synonym of [MASK])		0.02	**0.28**	0.20
Synonym-of	V	X_1 é o mesmo que [MASK] (X_1 is the same as[MASK])		0.12	0.12	**0.34**
Synonym-of	ADJ	estar X_1 é o mesmo que estar [MASK]. (being X_1 is the same as being [MASK])		0.14	0.06	**0.24**
Antonym-of	ADJ	ser [MASK] é o contrário de ser X_1 (being X_1 is the opposite of being [MASK])		0.08	0.26	**0.38**
Hypernym-of	Abstract	a [MASK] é um tipo de X_1 (the [MASK] is a type of X_1)		0.08	0.22	**0.38**
Hypernym-of	Concrete	o [MASK], que é um tipo de X_1 (the [MASK], which is a type of X_1)	*V*	0.08	0.20	**0.24**
Hypernym-of	V	como [MASK] e outros modos de X_1 (like [MASK] and other modes of X_1)		0.00	0.08	**0.20**
Hyponym-of	Abstract	um X_1, isto é, um tipo de [MASK] (a X_1, this is, a type of [MASK])	*V*	0.02	0.24	**0.40**
Hyponym-of	Concrete	uma X_1, isto é, um tipo de [MASK] (a X_1, this is, a type of [MASK])	*V*	0.10	**0.60**	0.56
Hyponym-of	V	como X_1 ou outras maneiras de [MASK] (like X_1 and other manners of [MASK])		0.18	**0.24**	0.18

Results in TALES

Accuracy of best patterns

Relation	PoS	Pattern		BERT-ML	BERT-base	BERT-large
Synonym-of	N	X_1 é sinónimo de [MASK] (X_1 is a synonym of [MASK])		0.02	**0.28**	0.20
Synonym-of	V	X_1 é o mesmo que [MASK] (X_1 is the same as[MASK])		0.12	0.12	**0.34**
Synonym-of	ADJ	estar X_1 é o mesmo que estar [MASK]. (being X_1 is the same as being [MASK])		0.14	0.06	**0.24**
Antonym-of	ADJ	ser [MASK] é o contrário de ser X_1 (being X_1 is the opposite of being [MASK])		0.08	0.26	**0.38**
Hypernym-of	Abstract	a [MASK] é um tipo de X_1 (the [MASK] is a type of X_1)		0.08	0.22	**0.38**
Hypernym-of	Concrete	o [MASK], que é um tipo de X_1 (the [MASK], which is a type of X_1)	*V*	0.08	0.20	**0.24**
Hypernym-of	V	como [MASK] e outros modos de X_1 (like [MASK] and other modes of X_1)		0.00	0.08	**0.20**
Hyponym-of	Abstract	um X_1, isto é, um tipo de [MASK] (a X_1, this is, a type of [MASK])	*V*	0.02	0.24	**0.40**
Hyponym-of	Concrete	uma X_1, isto é, um tipo de [MASK] (a X_1, this is, a type of [MASK])	*V*	0.10	**0.60**	0.56
Hyponym-of	V	como X_1 ou outras maneiras de [MASK] (like X_1 and other manners of [MASK])		0.18	**0.24**	0.18

* Not constrained by the options → much more challenging!

- BERT-large still the best choice

Results in TALES

Accuracy@10

Relation	PoS	Pattern		BERT-ML	BERT-base	BERT-large
Synonym-of	N	X_1 é sinónimo de [MASK] (X_1 is a synonym of [MASK])		0.20	0.64	**0.70**
Synonym-of	V	X_1 é o mesmo que [MASK] (X_1 is the same as[MASK])		0.24	0.80	**0.90**
Synonym-of	ADJ	ser X_1 é o mesmo que ser [MASK]. (being X_1 is the same as being [MASK])		0.24	0.54	**0.64**
Antonym-of	ADJ	ser [MASK] é o contrário de ser X_1 (being X_1 is the opposite of being [MASK])		0.22	0.40	**0.48**
Hypernym-of	Abstract	uma [MASK], isto é, um tipo de X_1 (a [MASK], this is, a type of X_1)	*V*	0.32	**0.70**	0.62
Hypernym-of	Concrete	o [MASK], que é um tipo de X_1 (the [MASK], which is a type of X_1)	*V*	0.20	0.54	**0.56**
Hypernym-of	V	como [MASK] e outros modos de X_1 (like [MASK] and other modes of X_1)		0.04	0.54	**0.58**
Hyponym-of	Abstract	uma X_1, isto é, uma espécie de [MASK] (a X_1, this is, a kind of [MASK])	*V*	0.38	**0.66**	0.64
Hyponym-of	Concrete	uma X_1, isto é, um tipo de [MASK] (a X_1, this is, a type of [MASK])	*V*	0.40	**0.88**	0.80
Hyponym-of	V	como X_1 ou outras maneiras de [MASK] (like X_1 and other manners of [MASK])		0.54	0.64	**0.70**

* Several with correct predictions in the top-10!
  - Suggestions for manual enrichment?

Results in TALES

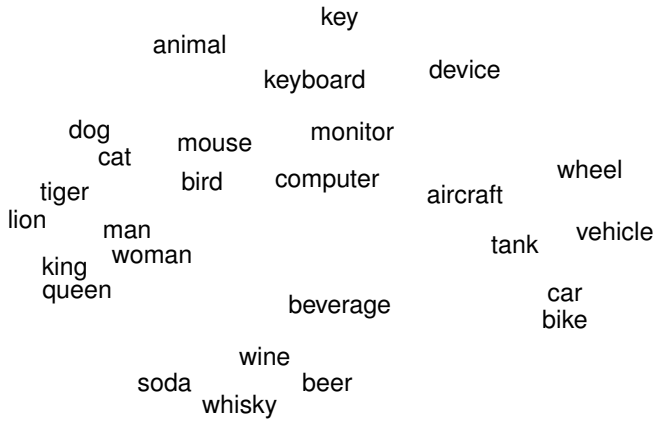
Other methods

- For discovering related words in static word embeddings (Drozd et al., 2016)

Results in TALES

Other methods

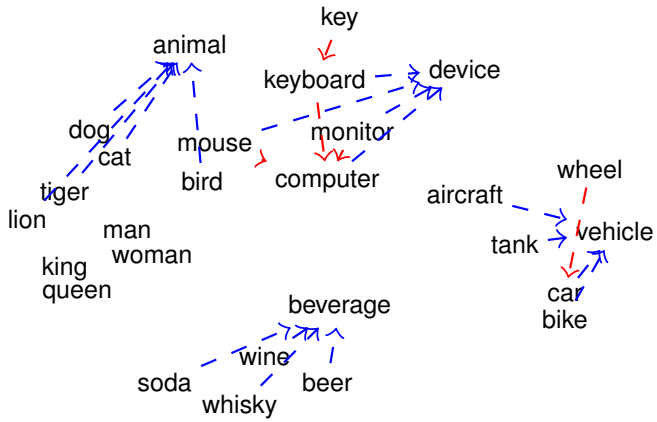
- For discovering related words in static word embeddings (Drozd et al., 2016)
  - 3CosAvg: what is to x_1 as the average of all $(\vec{x'_2} - \vec{x'_1})$?
  - LRCos: what is similar to x_1 and belongs to the class of all x'_2?



Results in TALES

Other methods

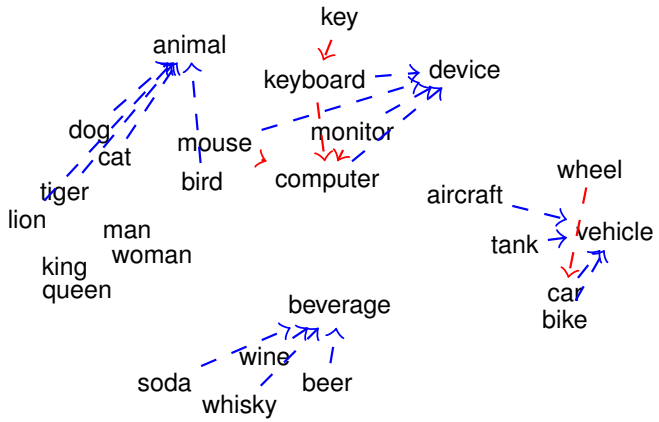
- For discovering related words in static word embeddings (Drozd et al., 2016)
  - 3CosAvg: what is to x_1 as the average of all $(\vec{x_2'} - \vec{x_1'})$?
  - LRCos: what is similar to x_1 and belongs to the class of all x_2'?



Results in TALES

Other methods

- For discovering related words in static word embeddings (Drozd et al., 2016)
  - 3CosAvg: what is to x_1 as the average of all $(\vec{x_2'} - \vec{x_1'})$?
  - LRCos: what is similar to x_1 and belongs to the class of all x_2'?



Relation	PoS	BERT-base	BERT-large	Sim	3CosAdd	3CosAvg	LRCos
Synonym-of	N	0.28	0.20	0.28*	0.18*	0.32$^{\times}$	**0.38**$^{+}$
Synonym-of	V	0.12	**0.34**	0.20^{+}	0.12^{+}	0.24^{+}	0.30^{+}
Synonym-of	ADJ	0.06	0.24	0.26*	0.10*	**0.28**$^{+}$	0.26^{+}
Antonym-of	ADJ	0.26	**0.38**	0.20*	0.14*	0.24^{+}	0.28*
Hypernym-of	Abstract	0.22	**0.38**	0.20^{+}	0.06$^{\times +}$	0.20^{+}	0.16^{*+}
Hypernym-of	Concrete	0.20	**0.24**	0.18^{+}	0.10$^{\times}$	0.20*	0.20^{+}
Hypernym-of	V	0.08	0.20	0.14*	0.08$^{\times}$	0.12^{+}	**0.22***
Hyponym-of	Abstract	0.24	**0.40**	0.08*	0.08*	0.10*	0.12*
Hyponym-of	Concrete	**0.60**	0.56	0.10^{+}	0.04$^{\times *+}$	0.14^{+}	0.28$^{\times}$
Hyponym-of	V	**0.24**	0.18	0.14^{+}	0.16*	0.16$^{\times +}$	0.22^{+}

$^{\times}$ GloVe; * word2vec-skip; $^{+}$ word2vec-cbow

COGNITIVE & MEDIA SYSTEMS CISUC 1 2 9 0

Conclusions & Future

- A pretrained BERT is a source of lexico-semantic knowledge

Conclusions & Future

- A pretrained BERT is a source of lexico-semantic knowledge
  - Better with **monolingual** and **large** models
  - Better for **noun relations**

Conclusions & Future

- A pretrained BERT is a source of lexico-semantic knowledge
  - Better with **monolingual** and **large** models
  - Better for **noun relations**
- Still ...
  - Not as good as static word embeddings for selecting related word from a restricted set
  - Not ready for being used as a LKB, useful for suggesting related words
    - May boost the **enrichment** of wordnets!
    - e.g., any Portuguese WordNet (de Paiva et al., 2016)
    - Best performance for hypernyms of nouns

Conclusions & Future

- A pretrained BERT is a source of lexico-semantic knowledge
  - Better with **monolingual** and **large** models
  - Better for **noun relations**
- Still ...
  - Not as good as static word embeddings for selecting related word from a restricted set
  - Not ready for being used as a LKB, useful for suggesting related words
    - May boost the **enrichment** of wordnets!
    - e.g., any Portuguese WordNet (de Paiva et al., 2016)
    - Best performance for hypernyms of nouns
- Handcrafted lexical patterns, based on previous work + VARRA
  - Available from: `https://github.com/NLP-CISUC/PT-LexicalSemantics/blob/master/Patterns/BERT_patterns_for_TALES_v2.txt`
  - Enough coverage?
  - Learn from seeds + corpora? (Bouraoui et al., 2020; Jiang et al., 2020)

Conclusions & Future

- A pretrained BERT is a source of lexico-semantic knowledge
  - Better with **monolingual** and **large** models
  - Better for **noun relations**
- Still ...
  - Not as good as static word embeddings for selecting related word from a restricted set
  - Not ready for being used as a LKB, useful for suggesting related words
    - May boost the **enrichment** of wordnets!
    - e.g., any Portuguese WordNet (de Paiva et al., 2016)
    - Best performance for hypernyms of nouns
- Handcrafted lexical patterns, based on previous work + VARRA
  - Available from: `https://github.com/NLP-CISUC/PT-LexicalSemantics/blob/master/Patterns/BERT_patterns_for_TALES_v2.txt`
  - Enough coverage?
  - Learn from seeds + corpora? (Bouraoui et al., 2020; Jiang et al., 2020)
- Test other models?
  - e.g., GPT3 (left-to-right instead of bidirectional / masked)

Meanwhile, ChatGPT...

Thank you for you attention.

Questions?

Thank you for you attention.

Questions?

✉ **hroliv@dei.uc.pt** | **@hgoliv**

References I

Bouraoui, Z., Camacho-Collados, J., and Schockaert, S. (2020). Inducing relational knowledge from BERT. In *Proc of AAAI Conference on Artificial Intelligence*, pages 7456–7463. AAAI Press.

de Paiva, V., Real, L., Gonçalo Oliveira, H., Rademaker, A., Freitas, C., and Simões, A. (2016). An overview of Portuguese wordnets. In *Proc 8th Global WordNet Conference*, GWC'16, pages 74–81, Bucharest, Romania.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc 2019 Conf of North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186. ACL.

Drozd, A., Gladkova, A., and Matsuoka, S. (2016). Word embeddings, analogies, and machine learning: Beyond king - man + woman = queen. In *Proc 26th International Conference on Computational Linguistics: Technical papers COLING 2016*, COLING 2016, pages 3519–3530.

Freitas, C., Santos, D., Gonçalo Oliveira, H., and Quental, V. (2015). VARRA: Validação, Avaliação e Revisão de Relações semânticas no AC/DC. In *Pesquisas e perspetivas em linguistica de corpus (Livro do IX Encontro de Linguística de Corpus, 2010)*, ELC 2010, pages 199–232. Mercado de Letras, Rio Grande do Sul, Brasil.

Gonçalo Oliveira, H. (2022). Drilling lexico-semantic knowledge in Portuguese from BERT. In *Computational Processing of the Portuguese Language – 14th International Conf, PROPOR 2022*, volume 12037 of *LNCS*, pages 387—397. Springer.

Gonçalo Oliveira, H., Sousa, T., and Alves, A. (2020). TALES: Test set of Portuguese lexical-semantic relations for assessing word embeddings. In *Procs of ECAI 2020 Workshop on Hybrid Intelligence for Natural Language Processing Tasks (HI4NLP 2020)*, volume 2693 of *CEUR Workshop Proceedings*, pages 41–47. CEUR-WS.org.

Hartmann, N. S., Fonseca, E. R., Shulby, C. D., Treviso, M. V., Rodrigues, J. S., and Aluísio, S. M. (2017). Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In *Proc of 11th Brazilian Symposium in Information and Human Language Technology (STIL 2017)*.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proc 14th Conference on Computational Linguistics*, COLING 92, pages 539–545. ACL.

Jiang, Z., Xu, F. F., Araki, J., and Neubig, G. (2020). How can we know what language models know? *Transactions of the Association for Computational Linguistics*, 8:423–438.

References II

Pantel, P. and Pennacchiotti, M. (2006). Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Procs of 21st International Conference on Computational Linguistics and 44th annual meeting of the Association for Computational Linguistics*, pages 113–120, Sydney, Australia. ACL Press.

Petroni, F., Rocktäschel, T., Riedel, S., Lewis, P., Bakhtin, A., Wu, Y., and Miller, A. (2019). Language models as knowledge bases? In *Proc 2019 Conf on Empirical Methods in Natural Language Processing and 9th Intl Joint Conf on Natural Language Processing (EMNLP-IJCNLP)*, pages 2463–2473. ACL.

Snow, R., Jurafsky, D., and Ng, A. (2005). Learning syntactic patterns for automatic hypernym discovery. *Advances in neural information processing systems*, 17:1297–1304.

Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: Pretrained BERT models for Brazilian Portuguese. In *Proc of Brazilian Conf on Intelligent Systems (BRACIS 2020)*, volume 12319 of *LNCS*, pages 403–417. Springer.

Wilkens, R., Zilio, L., Ferreira, E., and Villavicencio, A. (2016). B2SG: a TOEFL-like task for Portuguese. In *Procs 10th International Conference on Language Resources and Evaluation (LREC 2016)*. ELRA.