Punalogy: a pun analogy generator

Louis Berrane, Abir Lettifi, Camille Saran, Maeva Sillaire

Project, 2nd Presentation

07/11/2022

Outline

First Dataset

Semantic links dataset :

- Choose 'funny' lexical fields[1]
- Retrieve list of nouns of a given lexical field with wikidata SPARQL
- Get their antonyms, synonyms, hyperonyms, hyponyms with WORDNET

index	word	syn	ant	hyper	hypo
11	pundit	initiate,learned_person,savant		['scholar', 'scholarly_person', 'bookman', 'student']	['polymath']
13	almoner	medical_social_worker		['social_worker', 'caseworker', 'welfare_worker']	
15	intermediary	mediator,go-between,intermediator,intercessor		['negotiator', 'negotiant', 'treater']	['conciliator', 'make-peace', 'pacifier', 'peacemaker', 'reconciler'],['diplomat'],['harmonizer', 'harmoniser'],['interpreter', 'translator'],['matchmaker', 'matcher', 'marriage_broker'],['mediatrix'],['moderator'],['second_hand']
21	saddler			['maker', 'shaper']	
26	sociologist			['social_scientist']	['demographer', 'demographist', 'population_scientist'],['psephologist']
27	marketer	seller,vender,vendor,trafficker		['merchant', 'merchandiser']	['cosmetician'],['dealer'],['flower_girl'],['fruiterer'],['huckster', 'cheap-jack'],['peddler', 'pedlar', 'packman', 'hawker', 'pitchman'],['selling_agent'],['ticket_agent', 'booking_clerk'],['underseller']
31	shunter			['locomotive', 'engine', 'locomotive_engine', 'railway_locomotive']	
34	orthoptist			['specialist', 'medical_specialist']	
35	dressmaker	modiste,needlewoman,seamstress,sempstress		['garmentmaker', 'garment-worker', 'garment_worker']	

Figure – dataset example

Sub-words dataset

- Tokenization in subwords
- Creation of a second dataset

- Example : Mango $\rightarrow ma, an, ng, go, man, ang, ngo, mang, ango$

bet is to better as count is to countter

word	syn	ant	hyper	hypo
te	NaN	NaN	NaN	NaN
bet	count	NaN	[gamble]	['ante']
be	NaN	NaN	NaN	NaN
er	NaN	NaN	NaN	NaN

Prototype

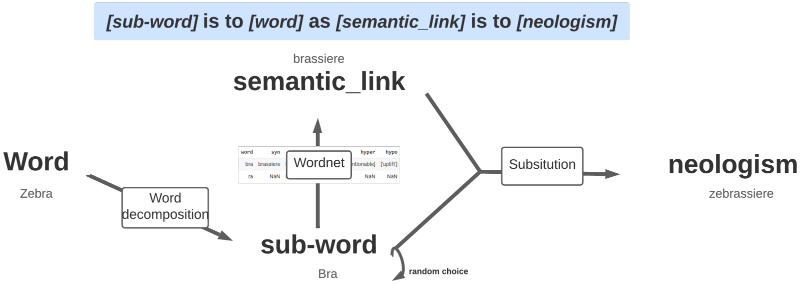
prototype

Word : *zebra*
Punalogy generated : *bra is to zebra as brassiere is to zebrassiere*

word	syn	ant	hyper	hypo
bra	brassiere	NaN	[undergarment, unmentionable]	['uplift']
ra	NaN	NaN	NaN	NaN

Figure – Zebra mini-dataset

Use of a template - Template-based approach [2]



- "best" punalogy ? *zebrassiere, zeundergarment, zeunmentionable, zeuplift*
- doable/not doable
- major issues
- define tasks : model ?
- word embedding ?

Meme generation

- Use of a template
- Use of an image generator (Dall-E)
  `https://github.com/lucidrains/DALLE-pytorch`

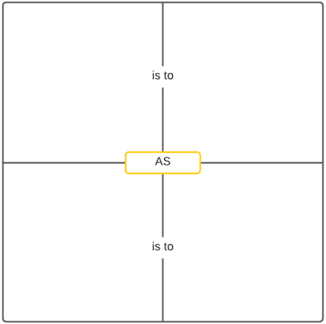


Figure – Meme template

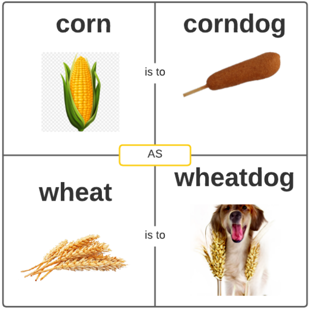


Figure – Meme example

evaluation

State-of-the-art : No optimal evaluation in NLP for humor [2]

- Soft Turing Test [3]
- Human evaluation on a Likert Scale (0 : not funny, 5 : really funny)
- Humorous Frequency (HF) [4]

Regular and systematic evaluation of our method

- Human Evaluation
- with Likert Scale (0 : not funny, 5 : really funny)
- Based on our actual results
- In order to improve and recalibrate our method (adding lexical parameters)

Issues, possible solutions ?

- Word Composition function
- Morphological blending

*bat is to batman as cream is to crea**mm**an*
*air is to dictionnaire as vent is to dictio**nnv**ente*
*ad is to avocado as promotion is to avo**cp**romotiono*

	word	syn	ant	hyper
0	use	NaN	NaN	[activity]
1	us	NaN	NaN	NaN
2	mou	NaN	NaN	NaN
3	se	NaN	NaN	NaN
4	mo	Missouri	NaN	[time]

- "rare" words (*"mou"*) - frequency ?
- Antonym/Synonym checking(Wordnet)

Timeline

Next session (18/11) :

- Enhance existing prototype
  Morphological blending ? Better semantical selection ?
- Integrate dall-e image generator
- toward a "hybrid" approach ?
  - Use of a model
    Which one ? For what tasks ? What embedding ?

Thank you !

Questions ? Suggestions ?

rat is to ratatouille as denounce is to denounceatouille
rob is to robot as overcharge is to overchargeot
perfume is to Evaneperfume as to scent is to Evanescent
bed is to bedroom as get up is to get uproom
ran is to orange as run is to orunge
design is to designet as to plan is to planet
adult is to adultgo as to man is to mango
man is to woman as adult is to woadult
lass is to sunglasses as girl is to sunggirles
pine is to philippines as conifer is to philipconifers

References I

[1] Rada Mihalcea and Stephen Pulman. Characterizing humour : An exploration of features in humorous texts. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 337–347, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[2] Miriam Amin and Manuel Burghardt. A survey on approaches to computational humor generation. In *Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 29–41, Online, December 2020. International Committee on Computational Linguistics.

[3] Kim Binsted and Graeme D. Ritchie. Computational rules for generating punning riddles. 1997.

References II

[4] Alessandro Valitutti. How many jokes are really funny ? towards a new approach to the evaluation of computational humour generators. In Bernadette Sharp, Michael Zock, Michael Carl, and Arnt Lykke Jakobsen, editors, *Proceedings of International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2011)*, pages 189–200, Denmark, 2011. Samfundslitteratur. International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2011) ; Conference date : 20-08-2011 Through 21-08-2011.