

Dead parrots make bad pets: Exploring modifier effects in noun phrases

Germán Kruszewski, Marco Baroni
Center for Mind/Brain Sciences
University of Trento, Italy

Podstata

- jak přívlastky (tu shodné) ovlivňují význam celé jmenné fráze
- zelený papoušek x mrtvý papoušek
- předchozí výzkum – psychologie, teor. lingvisti
– intenzionální fráze jako *falešný, údajný, toy*
(*toy gun – dětská pistolka*)

Jak přívlastky ovlivňují prototypy

- Eleanor Rosch – jestli je slovo charakteristické pro vyšší skupinu – pták: vrabec x tučňák, páv
- narušení prototypů: příklad živorodky



Vyvozování

- Jestliže papoušek = mazlíček, pak i živý papoušek = mazlíček
- Ale proč mrtvý papoušek (obvykle) ≠ mazlíček?
- Znalost toho, jak přívlastky ovlivňují toto vyvozování, by se dala využít při **rozpoznávání anafor** (mrtvý papoušek může být spojen se zvířetem, ale ne třeba s dýcháním (dýchajícím tvorem), přestože pro zvířata je typické, že dýchají) **a parafrází**

Cíl práce

- vytvoření open-access datasetu **NBP** (Norwegian Blue Parrot)
- snaha data počítačově zpracovat s využitím **distribuční sémantiky**
- a navíc **kompozicionální** distribuční sémantiky, protože se jedná o fráze, nejen jednotlivá slova

Postup při tvorbě NBP

- 1. tzv. narušené trojice: $\langle m, h, c \rangle$, kde m dělá h **netypickým** pro kategorii c
- 2. sada trojic bez takto silného modifikátoru
celkem 489 narušených trojic a 1938 “nenarušených”
- 3. kombinace – celkem 5849 párů
- 4. anotace

typy anotací

- 1. *jestli patří fráze do stejné kategorie jako substantivum* – je mrtvý papoušek pořád papoušek?
- 2. *jestli patří do některé z nadřazených kategorií, do kterých patří substantivum* – je pořád domácím mazlíčkem?
- 3. *ad prototypy* – jak typickým papouškem je mrtvý papoušek? a jak typickým mazlíčkem?

<i>instance</i>	<i>class</i>	<i>memb.</i>	<i>typ.</i>
top membership, top typicality			
gourmet soup	food	1.00	7.00
huge tiger	predator	1.00	7.00
sugared soda	drink	1.00	7.00
live fish	animal	1.00	7.00
Thai rice	rice	1.00	7.00
silver spoon	spoon	1.00	7.00
low membership, low typicality			
fatal shooting	sport	0.20	1.40
human egg	food	0.40	1.50
perfume bottle	drinkware	0.10	1.30
explosive vest	commodity	0.30	1.90
lemon water	chemical	0.20	1.60
creamy rice	bean	0.20	1.30
top membership, (relatively) low typicality			
sick tuna	tuna	1.00	3.20
explosive vest	vest	1.00	3.50
perforated sieve	tool	1.00	4.20
bottled oxygen	substance	1.00	4.30
grilled trout	creature	1.00	4.40
educational toy	amusement	1.00	4.50

- u membership: v 86 % případů shoda 8-10 anotátorů

Automatické zpracování – výběr

- chceme model, který pozná, jak málo/moc typičtí jsou mrtví papoušci jako mazlíčci v porovnání s papoušky obecně
= rozpoznání, co je instance a co je vyšší kategorie – **správný směr relace**
- př.: **membership** – využití asymetrie
skupina $h \rightarrow c$
- mult, wadd, (dil, fulladd, lexfunc) – různé modely
– nejlepší výsledky z nejjednodušších metod

Závěr

- “ňatuknutí” problému
- pro dostupné nástroje zatím obtížný, ale s tím se trochu počítalo
- je třeba dalšího výzkumu

Zdroje

- <https://www.aclweb.org/anthology/S/S14/S14-1021.pdf>
- https://www.youtube.com/watch?v=1k-WsYPm_MQ