Looking for Hyponyms in Vector Space

Marek Rei, SwiftKey
Ted Briscoe, University of Cambridge

SwiftKey® UNIVERSITY OF CAMBRIDGE

Hyponymy

Hyponymy is a ‘*type-of*’ relation

hyponym → *hypernym*

car, ship, train → vehicle

scarlet, crimson, vermilion → red

therapy, medication, rehabilitation → treatment

Applications in NLP:

- Information Retrieval
- Summarisation
- Paraphrasing
- etc.

Tasks

Hyponym detection

Kotlerman et al. (2010), Baroni & Lenci (2011)

Italian → language ?

Hyponym acquisition

Hearst (1992), Caraballo (1999), Snow et al. (2005)

… our international program offers courses in several different languages such as Italian and Spanish, and the student is able to choose ...

Hyponym generation

? → language

Output: Italian, Spanish, Chinese, Estonian, English, ...

Evaluation Dataset

Training (1230 hypernyms), development (922 hypernyms) and test (922 hypernyms) sets.

- Contains all hyponyms for each hypernym
- Extracted from WordNet
- Includes indirect hyponyms and synonyms
- Excludes low-frequency hypernyms

On average, each hypernym in the dataset has 233 hyponyms, but the distribution is roughly exponential, and the median is 36.

Vector similarity

Method: Scoring a large pool of candidates using vector similarity.

Candidates: words in BNC with 10+ frequency (86,496 words)

Candidate	Score for (? → language)	Correct
Italian	0.35	TRUE
Spanish	0.22	TRUE
culture	0.21	FALSE
English	0.18	TRUE
Spain	0.15	FALSE

Vector spaces

1. Window

Word co-occurrences in a context window of 3 on either side, PMI weighting.

2. Collobert & Weston (2008)

Neural network for predicting the next word in the sequence. Learns dense vector representations for each word.

3. Mnih & Hinton (2007)

Hierarchical log-bilinear (HLBL) neural network. Learns to predict the vector representation for the next word in the sequence.

Vector spaces

4. Word2vec

Feedforward neural network for efficient learning of word representations. Predicts surrounding words based on the current word. (Mikolov et al., 2013)

5. Dependencies

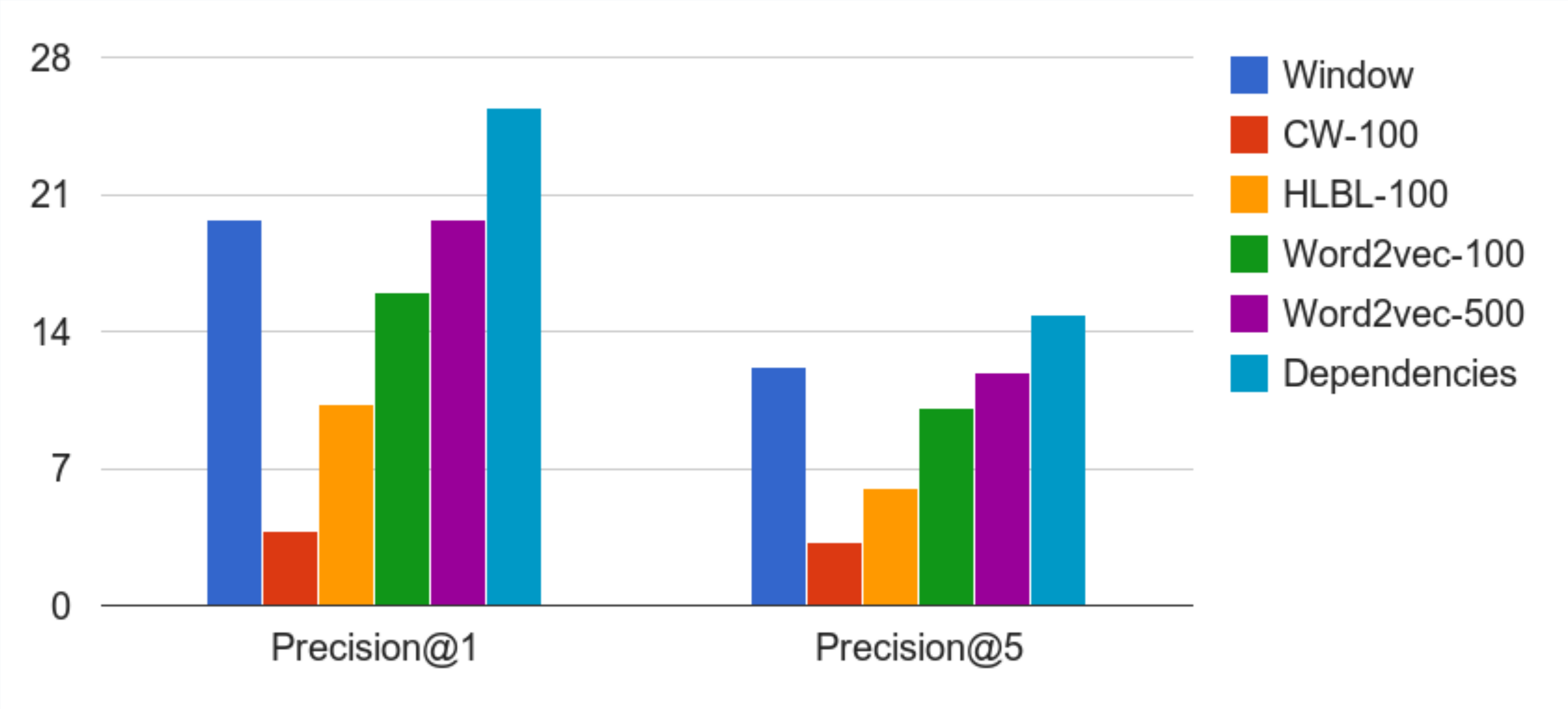
The text was parsed with RASP (Briscoe et al., 2006) and features extracted from dependency relations, weighted with PMI.

CW and HLBL were trained on RCV1, others on BNC.

Download: www.marekrei.com/projects/vectorsets/

Experiments with vector spaces

Using cosine as a scoring function



Vector offset method

Modelling semantic relations by vector offset (Mikolov et al., 2013)

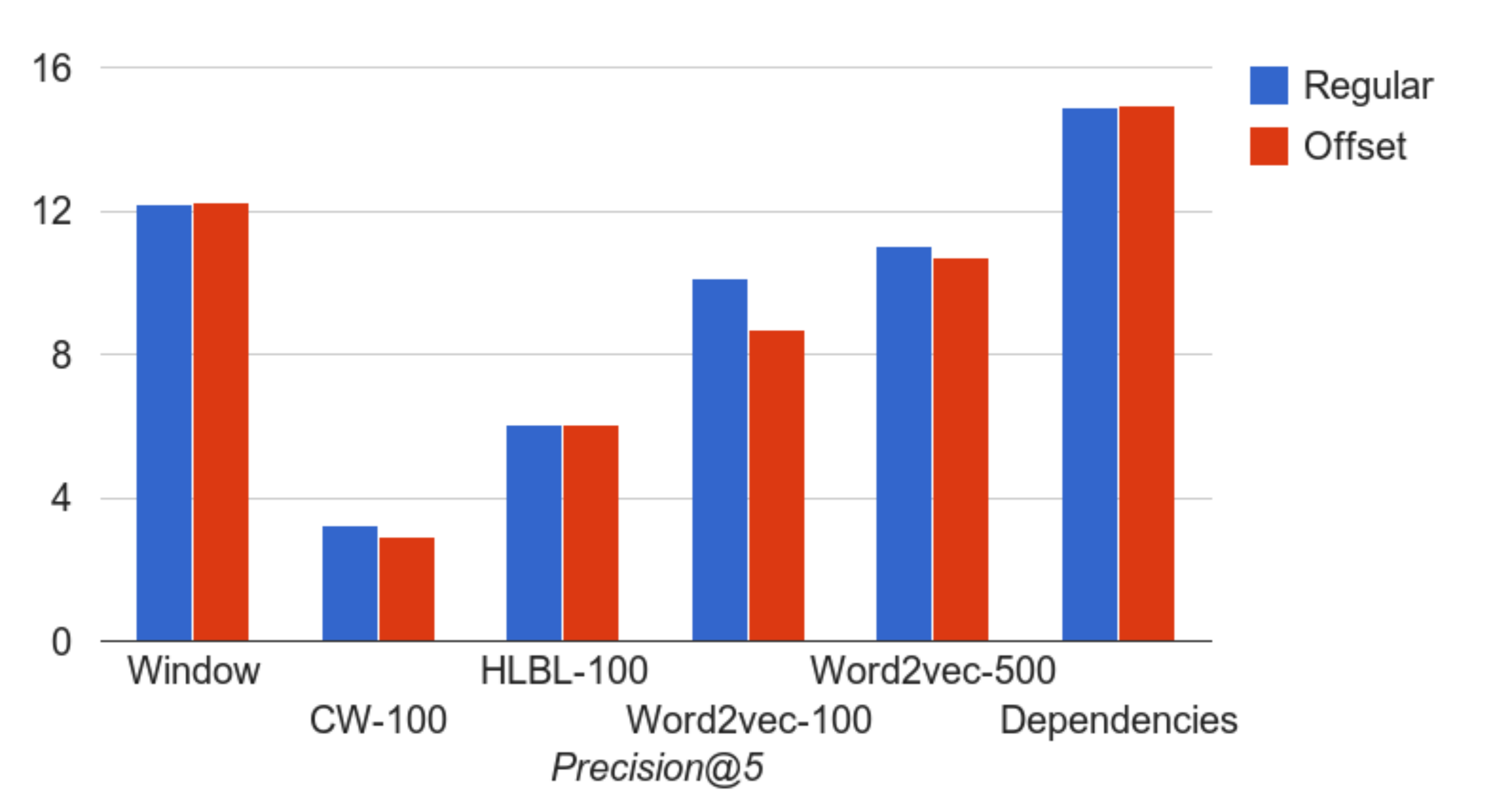
$$king - man + woman = queen$$

Can we apply this to hyponym generation?

$$bird - fish + salmon = eagle?$$

$$hypernym_A + \frac{1}{N} \sum_{i}^{N} (hyponym_i - hypernym_i)$$

Vector offset method



Vector offset method

bird	bird - fish + salmon	bird - red + crimson	bird - treatment + therapy
bird	salmon	bird	bird
mammal	bird	crimson	therapy
same-sized	**goose**	long-winged	mammal
reptile	**tern**	flightless	hedgehog
butterfly	**pheasant**	**moorhen**	sambar
wader	**plover**	reptile	reptile
lizard	**pipit**	lizard	lizard
insect	**gull**	**sea-bird**	**moorhen**
long-winged	**warbler**	babirusa	butterfly
tern	smoked	butterfly	frugivorous

Weighted cosine

We propose properties for a directional measure:

1. The shared features are more important to the directional score calculation, compared to the non-shared features.

2. Highly weighted features of the broader term are more important to the score calculation, compared to features of the narrower term.

$$z(f) = \begin{cases} (1 - \frac{r_b(f)}{|F_b|+1}) \times (1 - C) + C & \text{if } f \in F_a \cap F_b \\ C & \text{otherwise} \end{cases}$$

Similarity measures

Method	Precision@1	Precision@5
Pattern-based	8.14	4.45

Similarity measures

Method	Precision@1	Precision@5
Pattern-based	8.14	4.45
Cosine*	25.41	14.90
Lin*	21.17	12.23
DiceGen2*	21.82	14.55

Similarity measures

Method	Precision@1	Precision@5
Pattern-based	8.14	4.45
Cosine*	25.41	14.90
Lin*	21.17	12.23
DiceGen2*	21.82	14.55
WeedsPrec	0.11	0.04
WeedsRec	0.54	2.41
BalPrec	17.48	11.34
BalAPInc	15.85	9.66

Similarity measures

Method	Precision@1	Precision@5
Pattern-based	8.14	4.45
Cosine*	25.41	14.90
Lin*	21.17	12.23
DiceGen2*	21.82	14.55
WeedsPrec	0.11	0.04
WeedsRec	0.54	2.41
BalPrec	17.48	11.34
BalAPInc	15.85	9.66
WeightedCosine	25.84	15.46

Similarity measures

Method	Precision@1	Precision@5
Pattern-based	8.14	4.45
Cosine*	25.41	14.90
Lin*	21.17	12.23
DiceGen2*	21.82	14.55
WeedsPrec	0.11	0.04
WeedsRec	0.54	2.41
BalPrec	17.48	11.34
BalAPInc	15.85	9.66
WeightedCosine	25.84	15.46
Combined	**27.69**	**18.02**

Examples

scientist	sport	treatment
researcher	**football**	**therapy**
biologist	**golf**	**medication**
psychologist	club	patient
economist	**tennis**	procedure
observer	**athletics**	surgery
physicist	**rugby**	remedy
sociologist	**cricket**	regimen

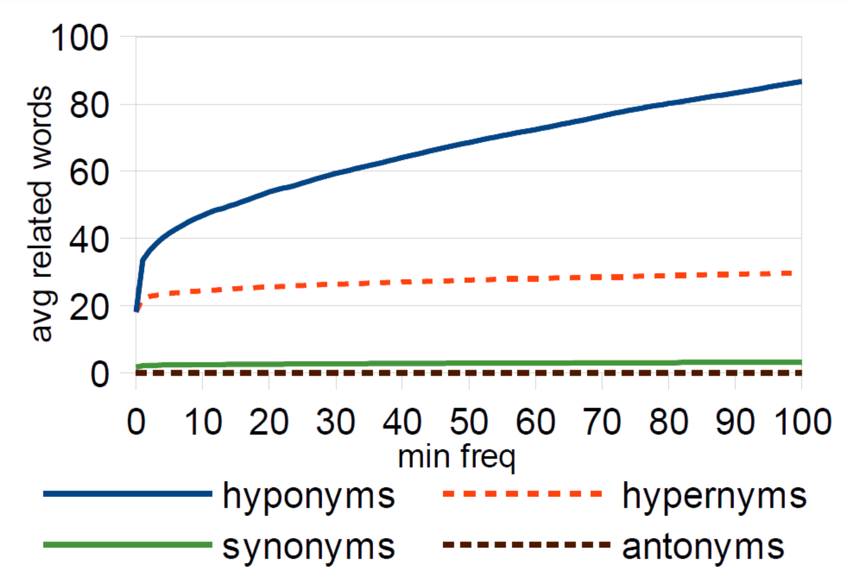
Investigating cosine

Why does symmetrical cosine perform so well?

1. There are many hyponyms, compared to other relations.

2. Directional measures focus only on the narrower term.

3. Much research on hyponym detection, but not generation.



Conclusion

- Performed a systematic evaluation of different methods for hyponym generation.
- It is important to choose the correct vector space and similarity measure for a specific task.
- Symmetric similarity measures (like cosine) perform surprisingly well.
- We constructed a new measure that outperformed others on hyponym generation.
- We release three vector sets, trained on the BNC with different methods.

Conclusion

Thank you!

Experiments with vector spaces

Candidates: BNC vocabulary with 10+ frequency (86,496 words)

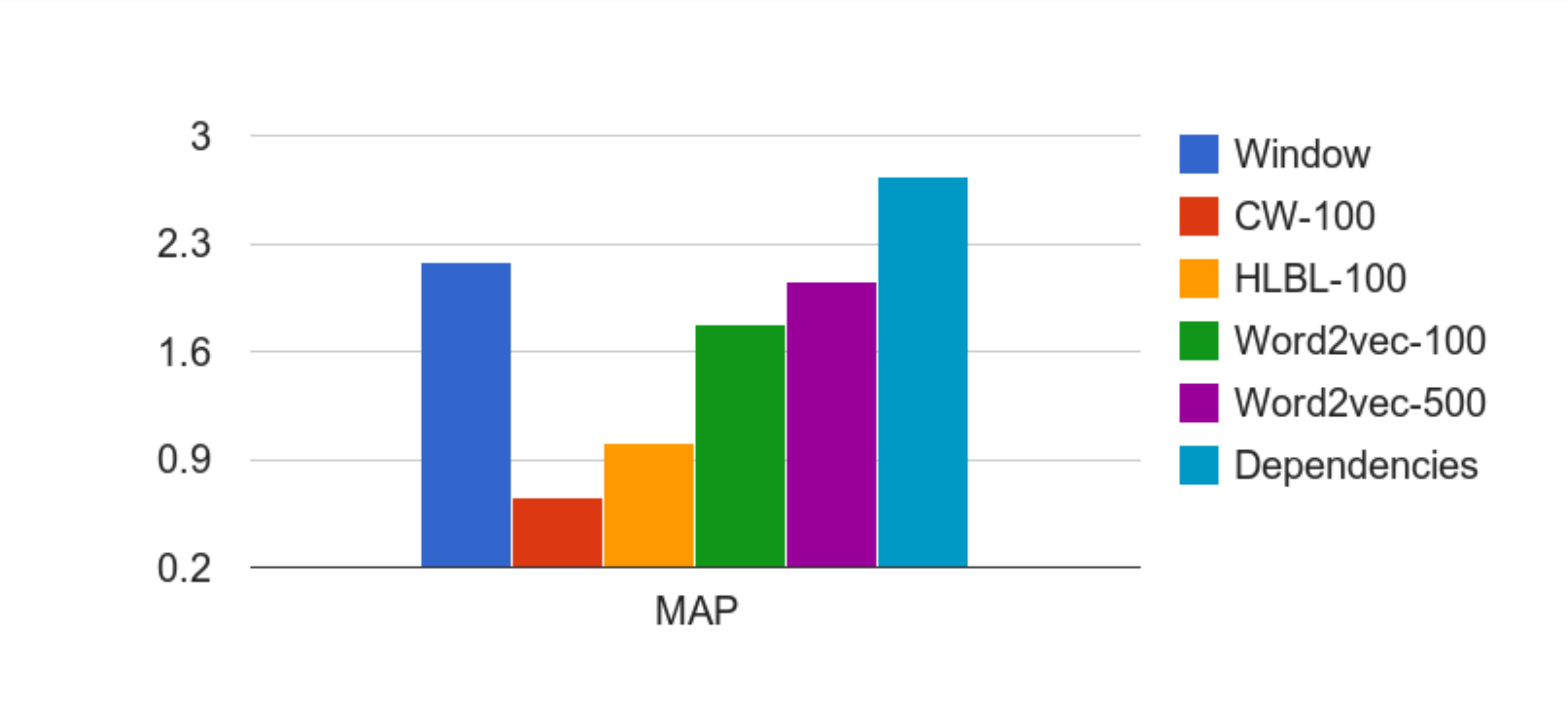
Scoring function: cosine

Vector space	MAP	Precision@1	Precision@5
Window	2.18	19.76	12.20
CW-100	0.66	3.80	3.21
HLBL-100	1.01	10.31	6.04
Word2vec-100	1.78	15.96	10.12
Word2vec-500	2.06	19.76	11.92
Dependencies	**2.73**	**25.41**	**14.90**

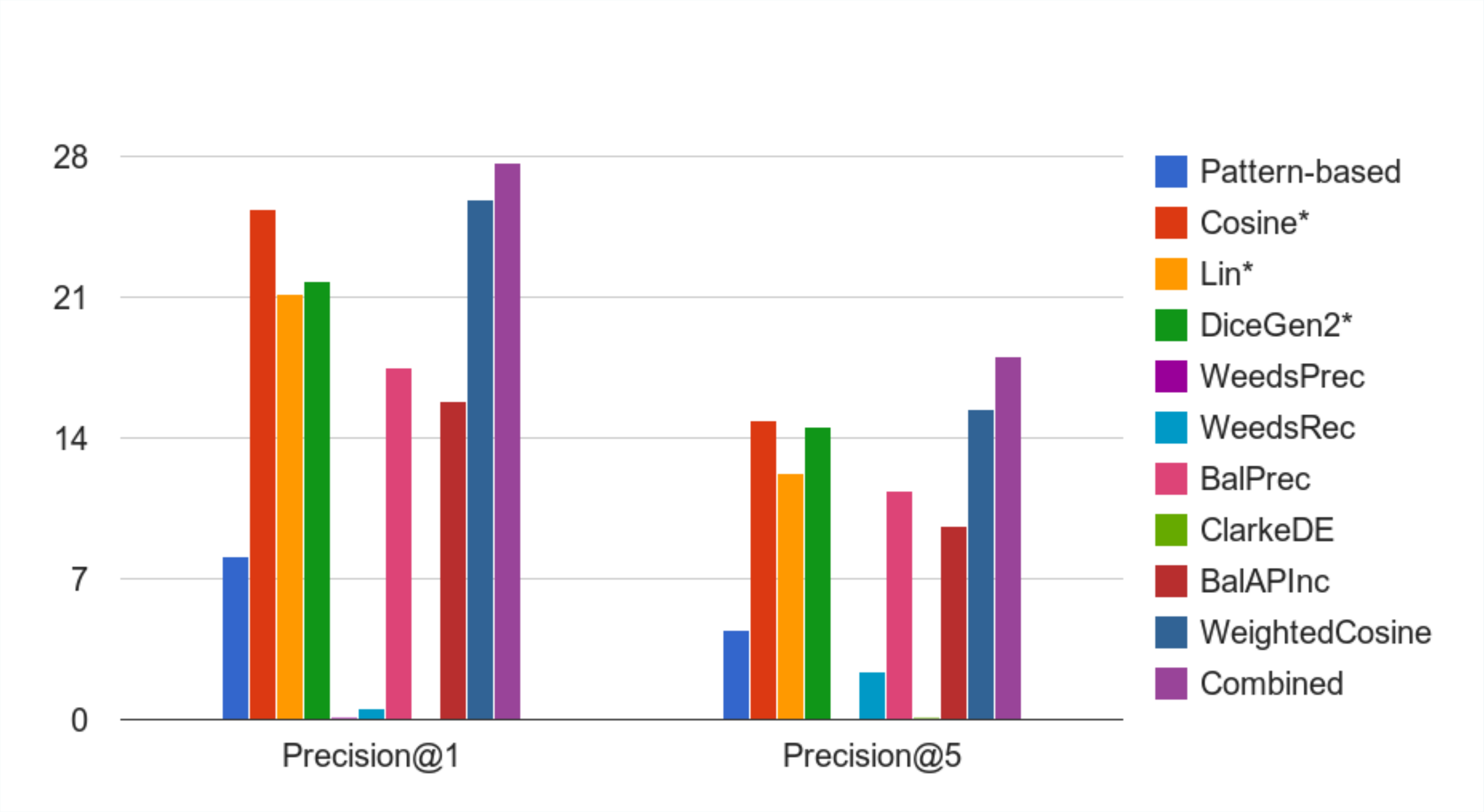
Experiments with vector spaces

Candidates: BNC vocabulary with 10+ frequency (86,496 words)

Scoring function: cosine



Similarity measures


Precision@1
Precision@5
Pattern-based
Cosine*
Lin*
DiceGen2*
WeedsPrec
WeedsRec
BalPrec
ClarkeDE
BalAPInc
WeightedCosine
Combined