

# Comparison Pattern Matching & Creative Simile Recognition

Vlad Niculae · vlad@vene.ro · http://vene.ro  
 Université de Franche Comté, University of Bucharest

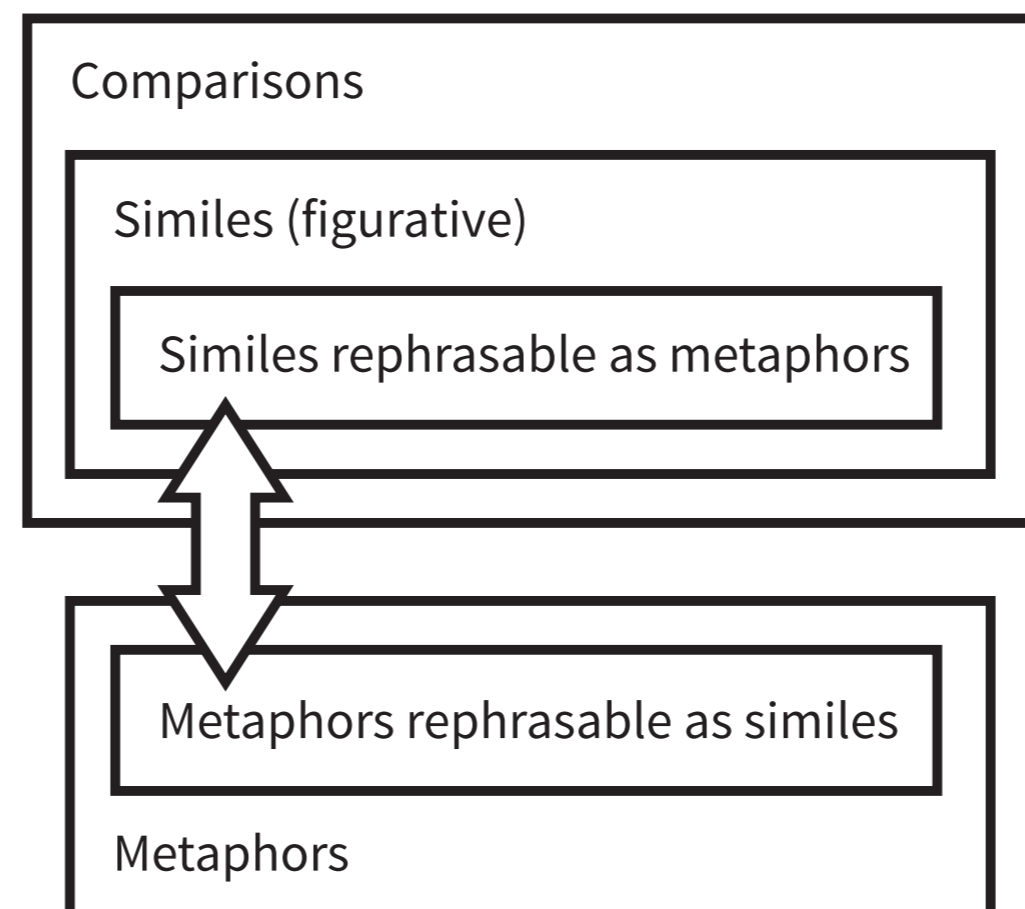
## Comparisons and Similes

**Comparison:** phrase that expresses the likeness of two entities

*He is like his father.*

**Simile:** figurative (non-literal) comparison; figure of speech with different range than metaphor

*My coffee tastes like rain-soaked cigars.*



## Constituents of a Comparison:

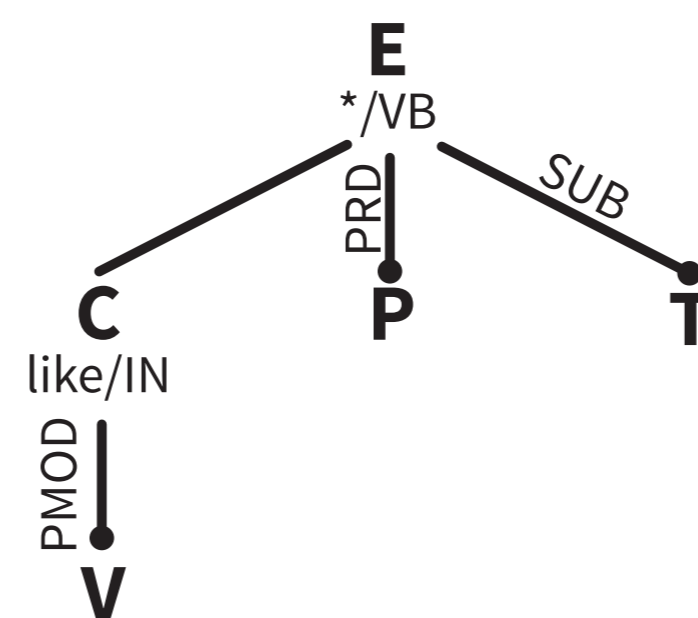
- topic
- eventuality
- predicator
- comparator
- vehicle

He looked like a broiled frog, hunched over his desk, grinning and satisfied.

His mouth tasted like the bottom of a parrot's cage.

## Pattern Encoding

**Dependency patterns** for classes of comparisons  
 Small **Python lib** to match patterns against corpora  
 Pattern format permits **automatic generation**

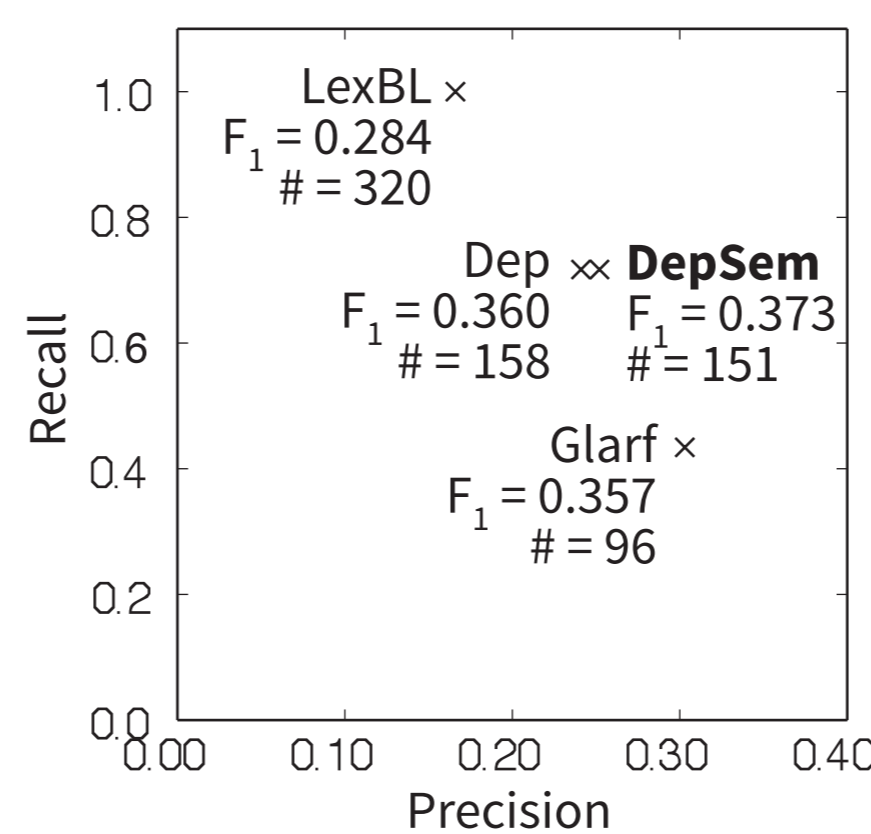


```

dict(slot='E',
    pos=lambda x: x.startswith('VB'),
    kids=[
        dict(slot='C',
            form='like',
            pos='IN',
            kids=[dict(slot='V',
                deprel='PMOD')]),
        dict(slot='T',
            deprel='SUB',
            optional=True),
        dict(slot='P',
            optional=True,
            deprel='PRD')
    ])
    
```

## Corpora for Automatic Evaluation

**VUAMC** corpus manually annotated for metaphor; includes some similes  
 mFlag annotation: metaphor flags  
*like* occurs 57 times (0.35% per sent)



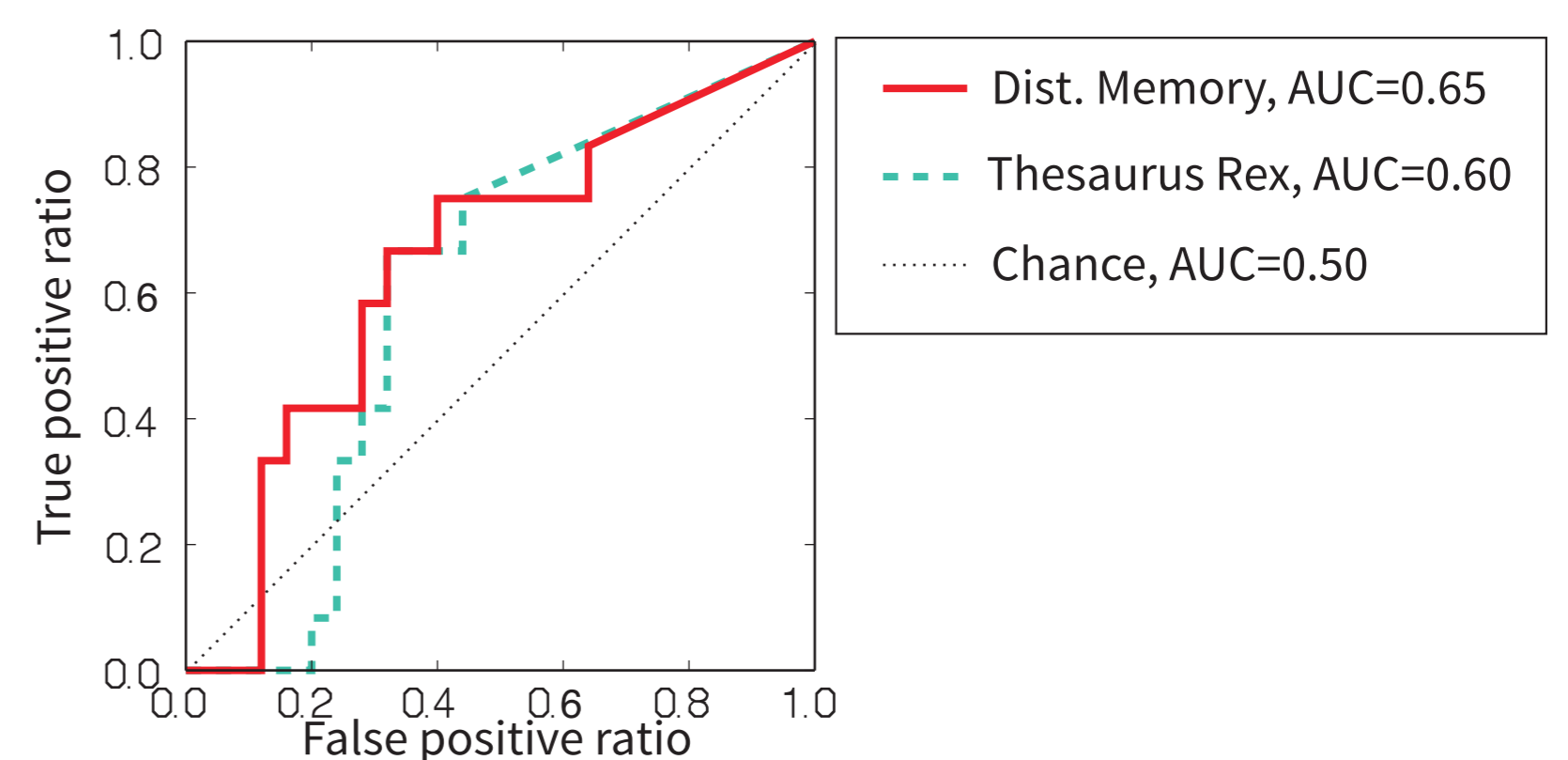
LexBL match occurrences of like/IN  
 Glarf previous work (Niculae & Yaneva, 2013)  
 Dep this work  
**DepSem** semantic heuristic dropping literal comparisons

## Measuring Figurativeness and Creativeness

Syntactical systems extract comparisons  
 The evaluation only rewards similes

Hypothesis for a **semantic approximation**:  
 figurative vs. creative: distinct but not orthogonal  
**literalness ~ conventionality ~ semantic similarity**  
 (comparisons of similar things are more likely to be literal)

Semantic resources (*canned*):  
**Distributional Memory** (Baroni & Lenci, 2010):  
 random projection of distributional *word-link-word* tensor  
**Thesaurus Rex** (Veale & Li, 2013):  
 ontology of stereotypical properties mined using  
*as X as Y* and *X such as Y* patterns



## Retrieved Examples:

The piano ripples like patent leather.  
 [DM(piano, leather) = 0.076]

This is a vivid and funny production and their expertise makes the intricate puppetry go like a dream.  
 [DM(puppetry, dream) = 0.076]

Ink, like paint, uses subtractive colour mixing while the video monitor uses the additive colours; red, green and blue, to produce the same effect.  
 [DM(ink, paint) = 0.502]