

# SemEval-2013 Task 4: Free Paraphrases of Noun Compounds

Iris Hendrickx, Zornitsa Kozareva, Preslav Nakov,  
Diarmuid Ó Séaghdha, Stan Szpakowicz, Tony Veale

Atlanta, GA, June 14, 2013

# Outline

- 1 Overview
- 2 Task Description
- 3 Evaluation
- 4 Participants, Results, Conclusion

# Overview (I)

**Noun compound (NC):** sequence of two or more nouns that act as a single noun, e.g., *colon cancer*, *suppressor protein*, *tumor suppressor protein*, *colon cancer tumor suppressor protein*, etc.

**Task:** interpret the meaning of two-word English NCs

## Applications

- Question Answering
- Machine Translation
- Information Retrieval

# Overview (II)

## Difficulties in NC interpretation (Lapata & Lascarides 2003)

- 1 the compounding process is highly productive
- 2 the semantic relation is implicit
- 3 contextual and pragmatic factors influence interpretation

# Overview (III)

## Related work

- based on semantic similarity
  - (Nastase & Szpakowicz 2003, 2006; Moldovan & al. 2004; Kim & Baldwin 2005; Girju 2007; Ó Séaghdha & Copestake 2007)
- based on paraphrasing
  - e.g., *olive oil* = '*oil that is extracted from olive(s)*'
  - (Vanderwende 1994; Kim & Baldwin 2006; Butnariu & Veale 2008; Nakov & Hearst 2008)

# Task Description (I)

- **Target:** two-word NCs, e.g. *air filter*
- **Goal:** produce an explicitly ranked list of free paraphrases, e.g.,
  - *1 filter for air*
  - *2 filter of air*
  - *3 filter that cleans the air*
  - *4 filter which makes air healthier*
  - *5 a filter that removes impurities from the air*
  - ...
- **Evaluation:** comparison to a similar list produced by human annotators

# Task Description (II)

**Data collection:** using *Amazon Mechanical Turk*.

	<b>Total</b>	<b>Min / Max / Avg</b>
<u>Trial/Train (174 NCs)</u>		
paraphrases	6,069	1 / 287 / 34.9
unique paraphrases	4,255	1 / 105 / 24.5
<u>Test (181 NCs)</u>		
paraphrases	9,706	24 / 99 / 53.6
unique paraphrases	8,216	21 / 80 / 45.4

**Statistics:** number of paraphrases with and without duplicates, minimum / maximum / average per noun compound.

# Task Description (III)

## Training Dataset

- 174 NCs from (Ó Séaghdha, 2007)
- 4,255 human paraphrases

## Test Dataset

- 181 NCs from (Ó Séaghdha, 2007)
- 8,216 human paraphrases



# Evaluation (I)

## The Scoring Strategy

The participating systems' paraphrases are matched against those in the "gold" standard: at word/stem level (fuzzy matches allowed), then at phrase level (overlapping n-grams, no determiners), then at the paraphrase level (to find the highest-ranking match for each). Scores and ranks for all of these are combined. *See the paper for all gory details.*

# Evaluation (II)

## Paraphrase Matching

- Isomorphic mode: each system paraphrase is matched with a different gold-standard paraphrase.
- Non-isomorphic mode: multiple system paraphrases may match the same gold-standard paraphrase.
- Rank multipliers reward system paraphrases which match gold-standard paraphrases highly ranked by humans.

# Evaluation (III)

$w_1$

CUTTING

$p$

CUTS

$w_2$

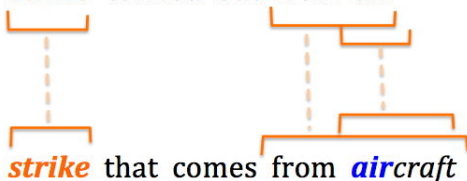
$$\text{score}(w_1, w_2) = \left[ \frac{2|p|}{|w_1| + |w_2|} \right]^2$$

*Scoring non-identical words  
with a common prefix ( $|p| \geq 3$ )*

# Evaluation (IV)

NN = “air strike”

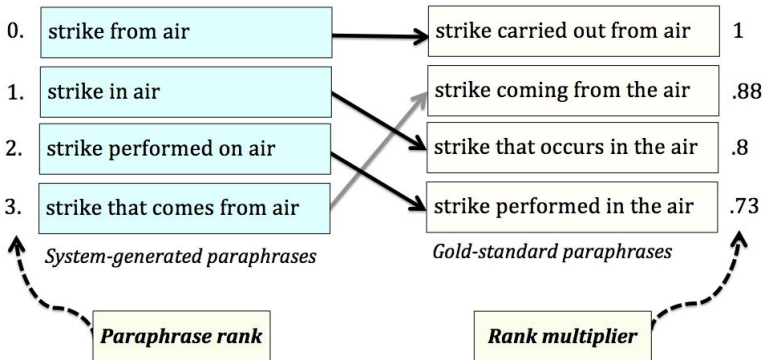
*strike* carried out from *air*



*N-gram overlap scores allow for near-identical words*  
Overall score normalized to **[0 ... 1]**

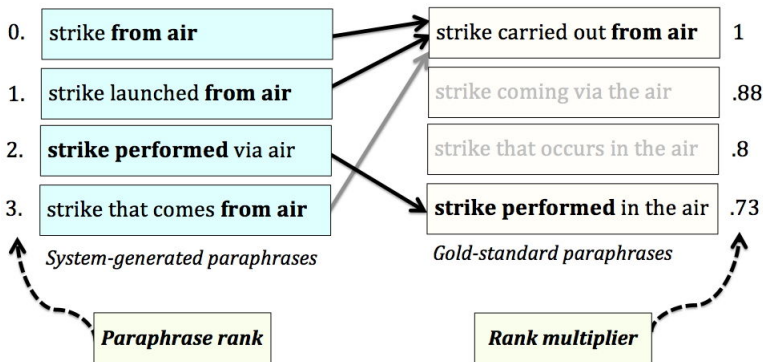
# Evaluation (V)

## Isomorphic (1 to 1) Matching



# Evaluation (VI)

## Non-isomorphic (*many to 1*) Matching



# Participants

- **MELODI**: semantic vector space model built from the UKWAC corpus; used features on the head noun to train a MaxEnt classifier.
- **IIITH**: probabilities of the preposition co-occurring with a relation to identify the class of the noun compound; uses Google n-grams, BNC and ANC.
- **SFS**: templates and fillers from training data, 4-gram language model, and a MaxEnt reranker. To find similar compounds, used Lin's WordNet similarity and statistics from the English Gigaword and the Google n-grams.

# Results

Team	isomorphic	non-isomorphic
SFS	23.1	17.9
IIITH	23.1	25.8
MELODI-Primary	13.0	54.8
MELODI-Contrast	13.6	53.6
<i>Naive Baseline</i>	<i>13.8</i>	<i>40.6</i>

## Baseline

For each test compound  $M H$ , generate the following paraphrases, in this precise order:

*H of M, H in M, H for M, H with M, H on M, H about M, H has M, H to M, H used for M, H used in M.*



# Conclusion

## Achievements

- Created a new dataset of free paraphrases for noun-noun compound interpretation; available for further research.
- Proposed two new evaluation metrics.
- Offered insights into the current approaches to the task.

This work has been partially supported by a grant from Amazon, which we used on MTurk.

We also thank our annotators: Dave Carter, Chris Fournier and Colette Joubarne.