

# ACQUIRING PREDICATE PARAPHRASES FROM NEWS TWEETS

Vered Shwartz, Gabriel Stanovsky and Ido Dagan

Computer Science Department, Bar-Ilan University [vered1986@gmail.com](mailto:vered1986@gmail.com)

## 1. AN EVER-GROWING PREDICATE PARAPHRASE RESOURCE

[a] <sub>0</sub> introduce [a] <sub>1</sub>	[a] <sub>0</sub> welcome [a] <sub>1</sub>
[a] <sub>0</sub> appoint [a] <sub>1</sub>	[a] <sub>0</sub> to become [a] <sub>1</sub>
[a] <sub>0</sub> die at [a] <sub>1</sub>	[a] <sub>0</sub> pass away at [a] <sub>1</sub>
[a] <sub>0</sub> hit [a] <sub>1</sub>	[a] <sub>0</sub> sink to [a] <sub>1</sub>
[a] <sub>0</sub> be investigate [a] <sub>1</sub>	[a] <sub>0</sub> be probe [a] <sub>1</sub>
[a] <sub>0</sub> eliminate [a] <sub>1</sub>	[a] <sub>0</sub> slash [a] <sub>1</sub>
[a] <sub>0</sub> announce [a] <sub>1</sub>	[a] <sub>0</sub> unveil [a] <sub>1</sub>
[a] <sub>0</sub> quit after [a] <sub>1</sub>	[a] <sub>0</sub> resign after [a] <sub>1</sub>
[a] <sub>0</sub> announce as [a] <sub>1</sub>	[a] <sub>0</sub> to become [a] <sub>1</sub>
[a] <sub>0</sub> threaten [a] <sub>1</sub>	[a] <sub>0</sub> warn [a] <sub>1</sub>
[a] <sub>0</sub> die at [a] <sub>1</sub>	[a] <sub>0</sub> live until [a] <sub>1</sub>
[a] <sub>0</sub> double down on [a] <sub>1</sub>	[a] <sub>0</sub> stand by [a] <sub>1</sub>
[a] <sub>0</sub> kill [a] <sub>1</sub>	[a] <sub>0</sub> shoot [a] <sub>1</sub>
[a] <sub>0</sub> approve [a] <sub>1</sub>	[a] <sub>0</sub> pass [a] <sub>1</sub>
seize [a] <sub>0</sub> at [a] <sub>1</sub>	to grab [a] <sub>0</sub> at [a] <sub>1</sub>

» Binary verbal predicate paraphrases

» Extracted from Twitter

» Up to 86% accuracy

» Ever-growing resource:  
» Currently ~0.8M paraphrases

» Complementary to other resources

Get the resource!



<https://github.com/vered1986/Chirps>

## 2. ASSUMPTIONS

- **Main assumption:** redundant news headlines of the same event are likely to describe it with different words [1].
- **This work:** propositions extracted from tweets discussing news events, published on the same day, that agree on their arguments, are predicate paraphrases.



[Amazon] **to buy is buying to acquire** [Whole Foods]

## 3. RESOURCE CREATION

Using Twitter Search API:

1. Collecting News Tweets

`get_tweets(lang=en, filter=news) →` `→ clean_tweets()`

Amazon is buying Whole Foods in \$13.7B  
 Amazon to acquire Whole Foods Market in deal valued at nearly \$14 billion  
 ...

2. Proposition Extraction

- Extract propositions from the tweets using PropS [2]: <https://github.com/gabrielStanovsky/props>
- Get binary verbal predicate templates, and apply argument reduction [3].

[Amazon] **buy** [Whole Foods]  
 [Amazon] **acquire** [Whole Foods Market]  
 ...

3. Generating Paraphrase Instances

- We consider two predicates as paraphrases if: (1) They appear on the same day. (2) Each of their arguments aligns with a unique argument in the other predicate.
- Two levels of argument matching: **strict** (exact match / short edit distance) and **loose** (partial token matching / WordNet synonyms)

[a]<sub>0</sub> **buy** [a]<sub>1</sub> [a]<sub>0</sub> **acquire** [a]<sub>1</sub> Amazon Whole Foods  
 [a]<sub>0</sub> **buy** [a]<sub>1</sub> [a]<sub>0</sub> **acquire** [a]<sub>1</sub> Intel Mobileye  
 ...

4. Generating Types

Heuristic score for a predicate paraphrase type:

$$p_1 = [a]_0 \text{ buy } [a]_1, \quad p_2 = [a]_0 \text{ acquire } [a]_1$$

$$s(p_1, p_2) = \text{count}(p_1, p_2) \cdot \left(1 + \frac{\text{days}(p_1, p_2)}{N}\right)$$

- $\text{count}(p_1, p_2)$  assigns high scores for frequent paraphrases
- $N$  - number of days since the resource collection begun
- $\frac{\text{days}(p_1, p_2)}{N}$  eliminates noise from two arguments participating in different events on the same day, e.g.: 1) *Last year when Chuck Berry turned 90*; 2) *Chuck Berry dies at 90*

5. Resource Release

- We release our resource daily, with two files:
  - **Instances:** predicates, arguments and tweet IDs.
  - **Types:** predicate paraphrase pair types ranked in a descending order according to the heuristic accuracy score.

## REFERENCES

- [1] Yusuke Shinyama, Satoshi Sekine, and Kiyoshi Sudo. Automatic paraphrase acquisition from news articles. In *HLT*, pages 313–318. Morgan Kaufmann Publishers Inc., 2002.
- [2] Gabriel Stanovsky, Jessica Ficler, Ido Dagan, and Yoav Goldberg. Getting more out of syntax with props. *arXiv*, 2016.
- [3] Gabriel Stanovsky and Ido Dagan. Annotating and predicting non-restrictive noun phrase modifications. In *ACL*, 2016.