

# MAKING HISTORY COUNT

## Information Retrieval Project

---

Gabriele Cerizza

Università degli Studi di Milano

## SEMANTIC SHIFTS

---

## » Objective

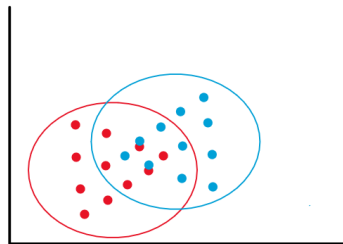
- › Measure the shift in word semantics between two time periods

## » Approach

- › Compare word embeddings of the same word obtained from models trained on corpora of different time periods

## Problem

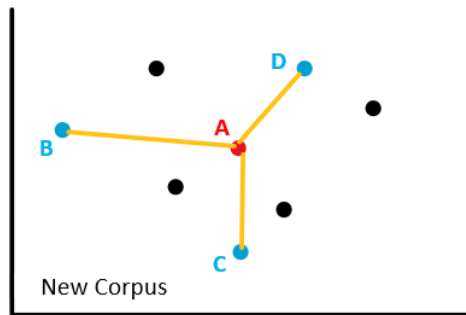
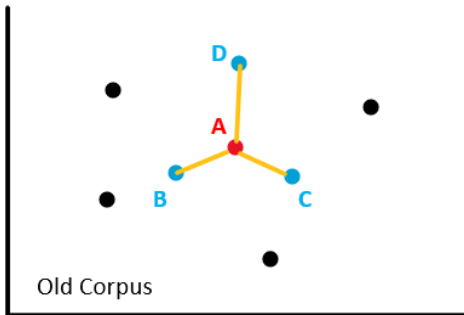
Word embeddings may be differently rotated along the axes



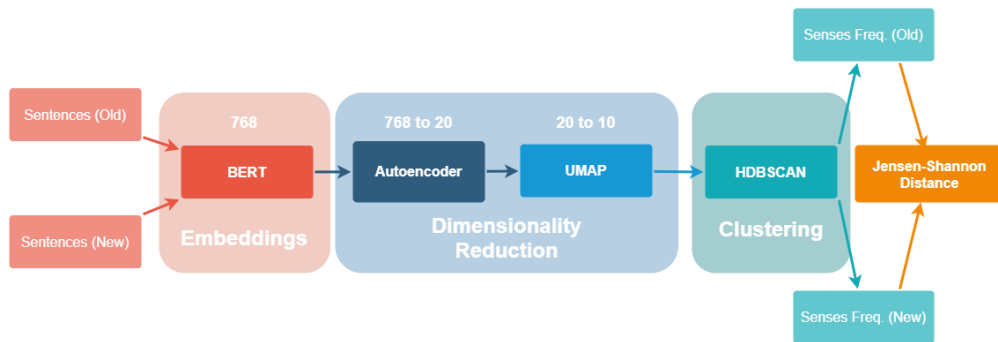
- » FastText Word Embeddings
  - › from documents dated 1860-1939 (Corpus of Historical American English)
  - › from contemporary news and Wikipedia
- » Only the 5000 most frequent words
- » Orthogonal Procrustes alignment
  - › noisy projections
- » Cosine distance between word embeddings

- » Same word embeddings as OP
- » Second-order similarity instead of alignment

$$\frac{\text{MSE}(D_{t_1}(w, nn_{t_1}), D_{t_2}(w, nn_{t_1})) + \text{MSE}(D_{t_1}(w, nn_{t_2}), D_{t_2}(w, nn_{t_2}))}{2}$$



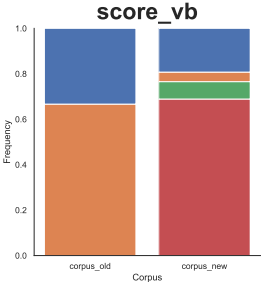
- » BERT contextualized embeddings
  - › a different embedding for each context in which a word appears
  - › no need to train multiple models
- » Corpora from SemEval-2020
  - › shuffled sentences from 1810-1860 and 1960-2010



OP	NN	JSD
cd	deletion	negro (adj.)
romney	km	people (noun)
km	gay	golf (noun)
diff	diff	shopping (noun)
deletion	outstanding	overall (adj.)
tv	implement	businessman (noun)
template	highlight	switch (verb)
isis	parameter	investor (noun)
bot	red	motor (noun)

isis (OP)	
Old similar	New similar
greek	islamic
temple	muslim
classic	arab
palestinian	syrian

gay (NN)	
Old similar	New similar
merry	homosexual
joyous	lesbian
gaiety	heterosexual
gayest	queer



Sense examples

- The Orioles scored three runs off Doc Medich right away .
- Seven Rebels scored in double figures as Vegas poured it on 132-90.
- In domestic matters the Labor Government scored very limited success .
- But it was scored in his heart . never to be forgotten , and he longed for vengeance .

## HISTORICAL EVENT EXTRACTION

---



### » Objective

- › Detect historical events in a text
- › Extract event arguments: dates, historical figures, locations
- › Evaluate on Wikipedia pages

### » Definition of historical events and entities

- › Events and entities that may be featured in a history textbook

### » Histo Corpus issues

- › Annotated events are not “historical”
- › Different language from Wikipedia
- › Event arguments are not annotated

## » RAMS Data Set (2020)

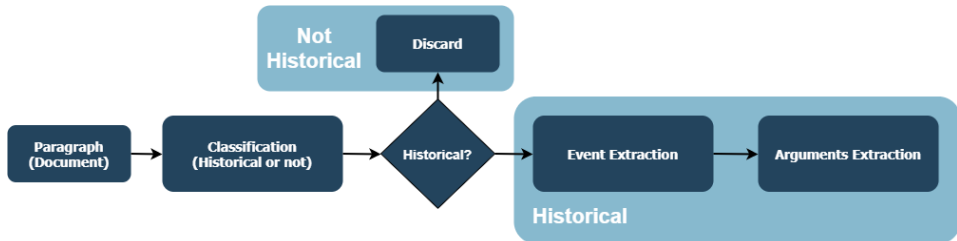
- › 139 event types
- › up to 5 arguments per event
- › 5 sentences and 1 event per document
- › arguments cross event sentence boundaries

## » Wikipedia Data Set

- › 1024 pages
- › pages, paragraphs and linked entities classified as historical or not based on Wikidata
- › 16075 historical paragraphs and 23673 non-historical paragraphs
- › linked entities annotated in the BIO format

## » Linked entities are not enough for classification

- › only the first mention is linked
- › surface form of an entity can vary greatly (“up the toe of Italy” for “Italian campaign (World War II)”)



» Paragraph classification

- › a multi-task model (MTL) to jointly learn the class of the paragraph and the tag of the entities

» Event extraction

- › enumerate spans and obtain span embeddings using BERT
- › use a FFNN to classify each span and find the event “trigger”

» Argument extraction

- › retrieve a pre-defined template for the identified event with “<arg>” placeholders
- › let BART fill the placeholders by providing the text of the paragraph (Conditional Text Generation)

Event: conflict.attack.invade

<arg> attacked <arg> using <arg> at <arg> place

Model	Acc	P	R	F1
MTL	84.2	80.6	84.2	82.4
BiLSTM	74.4	74.7	74.4	74.5
BERT	73.0	76.4	73.0	74.7

Model	Events			Arguments		
	P	R	F1	P	R	F1
Li et al.	-	-	-	-	-	48.6
Zhang et al.	-	-	-	-	-	40.1
Zhang et al. (TCD)	-	-	-	-	-	41.8
Wen et al.	-	-	-	-	-	48.6
Wei et al.	-	-	-	53.1	42.7	47.4
Lai et al.	71.9	74.7	73.2	-	-	-
Pouran Ben Veyseh et al.	55.5	78.6	65.1	-	-	-
Our BiLSTM baseline	0.1	2.3	0.2	-	-	-
Our BERT baseline	1.1	1.7	1.1	-	-	-
Our model (EventGen)	29.8	27.9	24.8	-	-	-
Our model	2.2	2.9	2.3	62.7	42.0	50.3

“In late 1941, Japan’s government, led by Prime Minister and General Hideki Tojo, decided to break the US-led embargo through force of arms. On December 7, 1941, the Imperial Japanese Navy launched a surprise attack on the American fleet at Pearl Harbor, Hawaii. This brought the US into World War II on the side of the Allies. Japan then successfully invaded the Asian colonies of the United States, the United Kingdom, and the Netherlands, including the Philippines, Malaya, Hong Kong, Singapore, Burma, and the Dutch East Indies.”

- » Predicted event: “conflict.attack.invade”
- » Predicted arguments: “Imperial Japanese Navy attacked American fleet using arms at Pearl Harbor, Hawaii place”

“In early 1944, towards the end of Tucker’s time in Italy, during the Battle of Monte Cassino, Allied commanders were engaged in a controversy regarding what action should be taken against the monastery at Monte Cassino. The Germans had declared it a military-free zone but many senior commanders were reluctant to believe that the Germans would not occupy such a strategically important position. Tucker had found a book dated 1879 in a Naples bookshop giving details of the construction of the monastery at Monte Cassino which his division was to attack. He wrote a memorandum to his Corps commander, Lieutenant-General Bernard Freyberg, concluding that it should be demolished to prevent its occupation. He also pointed out that with 150-foot (45 m) high walls made of masonry at least 10 feet (3 m) thick, there was no practical means for field engineers to deal with the place, and that bombing with blockbuster bombs would be the only solution since 1,000 pound bombs would be “next to useless”. General Sir Harold Alexander, commanding the Allied Armies in Italy, agreed to the bombing (which did not employ blockbuster bombs) and the ruins were occupied by German forces which held the position until 18 May.”

- » Predicted event: “artifactexistence.damageddestroy.destroy”
- » Predicted arguments: “Allied commanders destroyed monastery using blockbuster bombs instrument in Monte Cassino place”

Thank You



”After the Treaty of Versailles, treaties with Austria, Hungary, Bulgaria, and the Ottoman Empire were signed. However, the negotiation of the treaty with the Ottoman Empire was followed by strife, and a final peace treaty between the Allied Powers and the country that would shortly become the Republic of Turkey was not signed until 24 July 1923, at Lausanne .”

- » Predicted event: “government.agreements.unspecified”
- » Predicted arguments: “Allied Powers and Ottoman Empire signed an agreement in Lausanne place”

“In June 2005, The Times newspaper alleged that Mladić had demanded a \$5 million (£2.75 million) “compensation” to be given to his family and bodyguards if he gave himself up to the ICTY in the Hague. In January 2006, a Belgrade court indicted 10 people for aiding Mladić in hiding from 2002 to January 2006. An investigation showed Mladić spent his time in New Belgrade, a suburb of the capital.”

- » Predicted event: “justice.initiatejudicialprocess.chargeindict”
- » Predicted arguments: “court charged or indicted 10 people before <arg> court or judge for aiding Mladić crime in Belgrade place”