DIACR-Ita at EVALITA 2020: Task Guidelines

Pierpaolo Basile
Univerisità di Bari "A. Moro"
pierpaolo.basile@uniba.it

Tommaso Caselli
Rijksuniveristeit Gronigen
t.caselli@rug.nl

Annalina Caputo
Dublin City University
annalina.caputo@dcu.iel

Rossella Varvara
Univerisità di Firenze
rossella.varvara@unifi.it

Pierluigi Cassotti
Univerisità di Bari "A. Moro"
pierluigicassotti@gmail.com

1 Background

The present task focuses on the automatic recognition of lexical semantic change, combining together computational and historical linguistics. Given contextual information from corpora, the aim is to automatically detect if a given word has changed its meaning over time.

Words meanings can evolve in different ways. They can undergo *pejoration* or *amelioration* (when meanings become respectively more negative or more positive) or they can be object of *broadening* (also referred to as *generalization* or *extension*) or *narrowing* (also known as *restriction* or *specialization*). The English word *dog* is an example of broadening, since its more general meaning came from the late Old English "dog of a powerful breed" (Traugott, 2006). On the contrary, the Old English word *deor* with the general meaning of "animal" became *deer* in present-day English. Semantic changes can be further classified based on the cognitive process they result from, i.e. either from *metonymy* or *metaphor*. Lastly, it is possible to distinguish among changes due to language-internal factors or to language-external ones (Hollmann, 2009). The latter type usually reflects a change in society, as in the case of technological advancements (e.g. *cell*, from the meaning of "prisoner cell" to "cell phone").

The problem of the automatic analysis of lexical semantic change is gaining momentum in the NLP and Computational Linguistics communities, as showed by growing number of publications on the diachronic analysis of language and the organisation of related events such as the 1st International Workshop on Computational Approaches to Historical Language Change [1] and the project "Towards Computational Lexical Semantic Change Detection" [2]. Following this trend, SemEval 2020 will host for the first time a task on automatic recognition of lexical semantic change: the SemEval 2020 Task 1 - Unsupervised Lexical Semantic Change Detection[3]. While this task targets a number of different languages, namely Swedish, Latin, and German, Italian is not present.

Many are the existing approaches, data used, and evaluation strategies to detect semantic drift. Most of the approaches rely on diachronic word embeddings, some of these created as post-processing of static word embedding, such as Hamilton et al. (2016), while others create dynamic word embeddings, where vectors share the same space for all time periods (Del Tredici et al., 2016; Yao et al., 2018; Rudolph and Blei, 2018; Dubossarsky et al., 2019). Finally, recent work exploits word sense induction algorithms to discover semantic shifts (Tahmasebi and Risse, 2017; Hu et al., 2019). A more complete state of the art is described in a critically and concise way in the latest surveys (Tahmasebi et al., 2018; Kutuzov et al., 2018; Tang, 2018).

Almost all the previously mentioned methods use English as the target language for the diachronic analysis, while other languages remain under-explored. To date, only one evaluation has been carried out on the Italian language using the Kronos-it dataset (Basile et al., 2019).

Existing approaches presented in the literature, which are not language-dependent, may be easily tested for the Italian language. Moreover, DIACR-Ita may foster the implementation of new algorithms purposely designed for the Italian language.

[1] https://languagechange.org/events/2019-acl-lcworkshop/
[2] https://languagechange.org/
[3] https://competitions.codalab.org/competitions/20948

2 Task Description

The goal of the task is to establish if a set of words (target words) change their meaning across two periods, t_1 and t_2, where t_1 precedes t_2.

Following the SemEval 2020 Task 1 setting, we rely on the comparison of two time periods. In this way we tackle two issues: 1) we reduce the number of time periods for which data has to be annotated; 2) we reduces the task complexity, allowing different model architectures to be applied to it, widening the range of possible participants.

Participants will be provided with two corpora C_1 and C_2 (for time periods t_1 and t_2, respectively), and a set of target words. For each of them, systems have to decide whether a word changed or not its meaning between t_1 and t_2 according to the occurrences of target word(s) in sentences in C_1 and C_2. For instance, the meaning of the word "imbarcata" is known to have expanded (i.e, it has acquired a new sense) from t_1 to t_2. [4] This will be reflected in different occurrences of use in sentences between C_1 and C_2.

The task is formulated as a closed task (i.e., participants must train their models on the data that are provided).

3 Development and Test Data

3.1 Data construction

We will provide a diachronic corpus, a collection of documents extracted by newspapers and books written in the Italian language labeled with temporal information. We will split the corpus into two sub-corpora each belonging to a specific time period, t_1 and t_2, where t_2 follows in time t_1. We name the corpus respectively: C_1 and C_2.

To identify the set of candidate words that are attested to have undergone a meaning change, we will rely on several resources: the Sabatini Coletti Italian dictionary [5], the Kronos-it dataset [6] and the DELI [7] dictionary.

We will semi-automatically build a gold standard that indicates for each target word if it changes its meaning from t_1 to t_2.

3.2 Data Release Policy

The corpora used in this task are strongly pre-processed and randomized versions of the original corpora and are made freely available. The original authors retain their respective rights, where applicable. The manually annotated datasets will be made available by the organisers.

3.3 Provided data

During the development stage we will provide the following data to the participants:

- trial target words for which predictions are needed;
- the gold standard of the trial target words;
- a sample submission for the trial target words;
- two trial corpora that participants can use to develop their models and test them on the trial target words;
- two training corpora;
- an evaluation script.

[4] The word originally refers to an acrobatic manoeuvre of aeroplanes. Nowadays, it is also used to refer to the state of being deeply in love with someone.

[5] Available on-line: `https://dizionari.corriere.it/dizionario_italiano/`

[6] `https://github.com/pippokill/kronos-it`

[7] DELI: Dizionario Etimologico della Lingua Italiana

Trial data do not reflect real data, but we will artificially build them in order to provide an example of data useful for developing the systems.

The training corpora will be two time-specific Italian corpora. To maximise the closed-task formulation, we will provide pre-processed corpora. In particular, we will adopt a tab separated format, with one token per line. For each token, we will provide its corresponding part-of-speech and lemma. Sentences will be separated by empty lines. Data will be pre-processed with UDPipe [8] using the ISDT-UD v2.5 model. An example of the data format is illustrated below.

```
Questa	PRON	questo
è	AUX	essere
una	DET	uno
frase	NOUN	frase
.	PUNCT	.

Questa	PRON	questo
è	AUX	essere
un'	DET	uno
altra	ADJ	altro
frase	NOUN	frase
.	PUNCT	.
```

Participants are free to combine the available information as they want. Furthermore, to facilitate the generation of word embeddings, we will make available a script for generating a one sentence per line format.

Target words will be provided as lemmas in a dedicated file (one lemma per line), as illustrated below:

```
velina
legno
plastica
atomica
oscar
distanza
```

3.4 Gold standard

We will provide a text file with, for each line, a target lemma and its class. The classes are:

- 0: the target word does not change meaning
- 1: the target word changes meaning

4 Evaluation

The task is formulated as a binary classification problem. Systems predictions will be evaluated against the change points annotated in the gold standard using accuracy.

The test set (G) will contain both positive (P) and negative (N) examples $G = P \cup N$. For example:

$$P = \{velina, oscar, plastica, atomica\}$$

$$N = \{legno, distanza\}$$

Negative words are those that did not undergo a change in their meaning. Systems' predictions involve both positive and negative classified targets $Pr = Pr_{pos} \cup Pr_{neg}$. Then true positives (positive targets classified as positive) are $TP = P \cap Pr_{pos}$, true negatives (negative targets classified as negative) are $TN = N \cap Pr_{neg}$, false negatives (positive targets classified as negative) are $FN = P \cap Pr_{neg}$ and

[8] http://lindat.mff.cuni.cz/services/udpipe/run.php

false positives (negative targets classified as positive) are $FP = N \cap Pr_{pos}$.
We can then compute the accuracy as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

4.1 Baselines

We will provide two baseline models:

- Frequencies: The absolute value of the difference between the words' frequencies is computed;
- Collocations: For each word, we build two vector representations consisting of the Bag-of-Collocations related to the two different time periods ($T0$ and $T1$). Then, we compute the cosine similarity between the two BoCs.

In both baseline models, we will use a threshold to predict if the word has changed its meaning.

5 How to submit your runs

Results should be provided in a text file with one target word with its class for each line. Compress the text file in a zip archive.

References

Pierpaolo Basile, Giovanni Semeraro, and Annalina Caputo. 2019. Kronos-it: A dataset for the Italian semantic change detection task. In *CEUR Workshop Proceedings*, volume 2481.

Marco Del Tredici, Malvina Nissim, and Andrea Zaninello. 2016. Tracing metaphors in time through self-distance in vector spaces. In *CEUR Workshop Proceedings*. 3rd Italian Conference on Computational Linguistics, CLiC-it 2016 and 5th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, EVALITA 2016 ; Conference date: 05-12-2016 Through 07-12-2016.

Haim Dubossarsky, Simon Hengchen, Nina Tahmasebi, and Dominik Schlechtweg. 2019. Time-Out: Temporal Referencing for Robust Modeling of Lexical Semantic Change. In *57th Annual Meeting of the Association for Computational Linguistics*, pages 457–470. Association for Computational Linguistics (ACL), sep.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronie word embeddings reveal statistical laws of semantic change. In *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, volume 3, pages 1489–1501, may.

Willem Hollmann. 2009. Semantic change. In *English Language: Description, Variation and Context*, pages 301–313. Basingstoke: Palgrave.

Renfen Hu, Shen Li, and Shichen Liang. 2019. Diachronic Sense Modeling with Deep Contextualized Word Embeddings: An Ecological View. In *57th Annual Meeting of the Association for Computational Linguistics*, pages 3899–3908.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. *27th International Conference on Computational Linguistics*.

Maja Rudolph and David Blei. 2018. Dynamic Embeddings for Language Evolution. In *WWW '18: Proceedings of the 2018 World Wide Web Conference*, pages 1003–1011. Association for Computing Machinery (ACM).

Nina Tahmasebi and Thomas Risse. 2017. Finding IndividualWord Sense Changes and their Delay in Appearance. In *International Conference Recent Advances in Natural Language Processing*, pages 741–749. Assoc. for Computational Linguistics Bulgaria, nov.

Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2018. Survey of Computational Approaches to Lexical Semantic Change. *1st International Workshop on Computational Approaches to Historical Language Change 2019*.

Xuri Tang. 2018. A state-of-the-art of semantic change computation. *Natural Language Engineering*, 24(5):649–676, sep.

Elizabeth Closs Traugott. 2006. Semantic change: Bleaching, strengthening, narrowing, extension. In *Encyclopedia of Language and Linguistics*. Elsevier.

Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. 2018. Dynamic word embeddings for evolving semantic discovery. In *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, volume 2018-Febua, pages 673–681.