Semantic Shift Detection in Vatican Publications: a Case Study from Leo XIII to Francis

Silvana **Castano**[1], Alfio **Ferrara**[1], Stefano **Montanelli**[1] and Francesco **Periti**[1]

[1]Università degli Studi di Milano
Department of Computer Science
Via Celoria, 18 - 20133 Milano, Italy

Abstract

In the recent years, *word embedding models* are being proposed to effectively detect language change and semantic shift in diachronic corpora. In this paper, we present a comparative analysis of different word embedding approaches by considering a case-study based on an Italian diachronic corpus of Vatican publications of Popes from Leo XIII to Francis (1898-2020). Four different approaches are considered, characterized by the adoption of different embedding models each one trained over the publications of a specific pope. The paper aims to explore whether and how word embedding techniques are successful in detecting semantic shifts over the language used by popes.

Keywords

Computational Humanities, Word Embeddings, Semantic Shift Detection

1. Introduction

In the recent years, the use of machine learning models in the field of Computational History is gaining more and more attention [1]. In particular, the application of *word embedding* techniques to the analysis of historical corpora is providing interesting and promising research results [2]. However, when historical corpora span different time periods, a number of linguistic issues can emerge. A word can evolve across the years by acquiring/losing meanings or by changing the context in which it is employed. For examples, the word *gay* shifted from meaning 'cheerful' to 'homosexual' during the 20th century, or the word *girl* having meant 'young person of either gender' in the past [3]. We refer to this process as *semantic shift*. Although in the past decades the automatic detection of semantic shift had been already investigating through data-driven approaches [4, 5], solutions based on *word embedding models* are currently being proposed and they are characterized by i) time-oriented splitting of a considered diachronic corpus into sub-corpora in which a coherent language without semantic shifts can be assumed, and ii) comparison of word embeddings derived from the sub-corpora to capture the semantic shift of words across different time periods. These approaches leverage the idea that semantically-related words are close the one to the others in the embedding space [6]. However, word embeddings from different temporal vector spaces cannot be naturally compared due to their stochastic

nature. Consequently, different approaches have been proposed to enable the embedding comparison across different models.

Motivations.   In this paper, we present a comparative analysis of different word embedding approaches by exploiting a diachronic corpus of Vatican publications from Leo XIII to Francis (1898-2020). The goal of the work is twofold. On one side, we aim at exploring whether and how word embedding techniques are successful in detecting semantic shifts over official documents and real documents that address a large audience over a long time period. Moreover, the paper aims at comparing and discussing the effectiveness of different literature approaches to capture the semantic shifts on a corpus of limited size and highly unbalanced nature like the Vatican publications corpus. On the other side, the corpus of Vatican publications represents a textual dataset of great interest, motivated not only by the exceptional historical depth of the corpus, but also by two reasons concerned with the nature of Vatican documents. The first reason is that the Catholic Church, through the writings of its popes, has always dealt with the most relevant issues in the public debate of its time, alongside the themes of faith and worship. Therefore, these writings constitute a historical source of primary importance for reconstructing an important part of the human cultural history. The second reason is related to the presence in the writings of the Holy See of terms and concepts that are characterized by a poor semantic shift over time alongside others that have instead remarkably changed both in terms of relevance and context. The former are mainly terms referring to the dogmas of faith which, albeit with some variations, essentially remained stable in the discourse of the popes. On the opposite, the latter are terms that describe well the way in which the attention of the public discourse shifted over time to different topics, such as the environment, the role of science, and many historical events of the human history. For these reasons, the corpus of Vatican documents is a perfect laboratory for experimenting with the techniques of semantic shift detection and this work constitutes a first step in the investigation of this very rich heritage of human culture.

The paper is organized as follows. In Section 2, we discuss the related work. In Section 3, we present our case-study on Vatican publications. The methods used for the case study analysis is described in Section 4. The results of the case study are presented in Section 5. In Section 6, we finally provide our concluding remarks.

2. Related work

As a general remark, word embeddings approaches to semantic shift detection are based on time-sliced corpora and separate embedding models. The comparison of different word representations over time (one per model) is enforced through a distance measure such as for example the cosine or jaccard similarity. A simple **Non-Aligned (NA)** method for semantic shift detection is proposed in [7], where the use of a word over the time is detected through the analysis of the word context in different time periods. In particular, the idea is to consider the top-n neighbors of a word in each temporal embedding model and to measure the overlap of these lists suggesting that smaller overlaps means drastic changes. However, an alternative and more typical solution is based on the idea to align word representations (i.e., embeddings)

which live in different temporal spaces before compared them. In [8], an **Incremental Update (IU)** mechanism of the embedding models is proposed. After a model is trained on a first period, it is then updated with data from the following time periods by saving its state as a new period model each time. In [9], the idea is to align embedding models to a unique vector space using heuristic local alignments per word based on the assumption that the set of nearest words in the embedding space change for words that have a shift. Then, changes between periods are detected by a distance-based distributional time series for each word in the corpus. The idea of using a similar transformation in the *temporal correspondence problem* is proposed in [10], where, given an input term (e.g., *iPod*) and a target time (e.g., *1980s*), the task is to predict the counterpart of the query that existed in the target time (e.g., *walkman*). The approach in [11] relies on the orthogonal **Procrustes (PR)** as a global alignment mechanism for temporal embedding spaces in the evaluation framework of different embedding techniques for detecting semantic shifts. Further studies attempted to combine information captured by the embedding models and the frequency of changes for capturing word shifts (e.g., [12, 13]).

In [14], the idea of creating dynamic embedding models is proposed where data across all the time periods are shared so that there is no need to align embedding spaces trained on separate sub-corpora. A Bayesian version of the skip-gram model with a latent time series as prior is proposed in [14]. Similarly, in [15], the authors propose to extend the skip-gram model by modeling time as a continuous variable. In [16], a different approach is presented in which word embeddings for each time period were not first learned, then aligned, but rather learned and aligned at the same time. As a further approach, the idea of [17] is to train embeddings on a corpus as a whole while tagging some word of interest with a special tag that indicate which corpus it comes from. As a result, an individual time-dependent embedding is created for each target word. To avoid the embedding alignment through orthogonal transformations, in [18], the authors propose to compute **Second-Order embeddings (SO)**, namely embeddings that share the same temporal space since obtained by modeling the meaning of words by means of their semantic similarity relations with all the other words in the vocabulary.

As a final remark, we note that an increasing interest is emerging about the use of contextualised pre-trained models for semantic shift detection [19, 20]. However, in this paper, such approaches are not considered since recent comparisons show that static embedding models, like Word2Vec, outperfomed the contextualised ones for semantic shift detection [21].

3. The Vatican corpus

The considered corpus of Vatican publications contains 27,831 documents extracted from the digital archive of the Vatican website[1]. The corpus consists of all the web-available documents at downloading time from Leo XIII to Francis (1878-2020) and the popes represent a natural criterion for splitting the corpus along the time, meaning that a separate sub-corpus is defined for each pope with associated publications. Furthermore, we stress that the documents have been downloaded in Italian. This choice is motivated as follows:

- The documents on the Vatican website are available in various languages, including Italian, Latin, English, Spanish, and German. We decided to work with the Italian language since

---

[1]https://www.vatican.va

a largest number of documents can be obtained in this language (consider that only 14,384 documents are available in English).

- In addition, although the official language of the Holy See is Latin, some of the available texts are not real official documents of the Catholic Church (e.g., encyclicals, apostolic constitutions, letters or exhortations), but they are about official documents of minor dogmatic importance (e.g., homilies, audiences, messages, biographies). Again, the number of available Latin documents about the publications from popes (i.e., 5,027 texts) is strongly less than the number of Italian documents.

A summary description of the considered Vatican corpus is provided in Table 1. Tokens represent the text units (i.e., words, terms) extracted from the Vatican documents through a text lowercasing step. As a further feature of the considered Vatican corpus, we note that the size of the sub-corpora from the popes varies from few documents (e.g., 19 documents from John Paul I) up to some thousands (e.g., 15,307 documents from John Paul II), meaning that the overall dataset is an example of unbalanced corpus.

Pope	Documents	Paragraphs	Tokens	Papacy beginning
Leo XIII	137	3,897	399,470	1878-02-20
Pius X	48	1,474	138,336	1903-08-04
Benedict XV	166	3,398	248,563	1914-09-03
Pius XI	95	4,371	370,728	1922-02-06
Pius XII	678	19,074	1,502,876	1939-03-02
John XXIII	377	14,014	702,157	1958-10-28
Paul VI	3,345	64,285	4,312,149	1963-06-21
John Paul I	19	370	23,862	1978-08-26
John Paul II	15,307	406,871	22,729,204	1978-10-16
Benedict XVI	3,008	68,055	4,713,339	2005-04-19
Francis	4,377	101,035	6,374,557	2013-03-13

Table 1
The considered corpus of Vatican publications.

4. Methods for the case study analysis

In this paper, we consider four different literature approaches to semantic shift detection for application to the Vatican corpus. In particular, we selected a *non-aligned* (*NA*) approach [7] and three different *aligned* solutions to make comparable the temporal vector spaces of different time periods, namely *Procrustes* (PR) [11], *Incremental Updates* (IU) [8], and *Second Order Embeddings* (SO) [18]. In our comparative analysis, the following data processing steps are executed, namely *time-oriented splitting*, *word embeddings construction* and *alignment*, and *semantic shifts detection*.

Time-oriented splitting. The Vatican corpus is split by creating a separate sub-corpus for each pontificate. Due to the short pontificate of John Paul I and the lack of documents from Pius X and Pius XI, we decide to group their documents with those of the immediately preceding popes. As a result, we merge the documents of John Paul I with those of Paul VI, the documents of Pius XI with those of Benedict XV, and the documents of Pius X with those of Leo XIII, respectively.

Word embeddings construction. For each one of the considered approaches (i.e., NA, PR, IU, SO), we train 100-dimensional word embeddings over each sub-corpus by exploiting the Gensim's implementation of Word2Vec.[2]

Word embeddings alignment. For the three aligned solutions, the alignment of embeddings belonging to separate vector spaces is executed as follows.

Procrustes (PR). We perform a cross-time alignment through the Procrustes implementation available at www.github.com/williamleif/histwords. The Procrustes assumption is that each word space has axes similar to the axes of the other word spaces, and two word spaces are different due to a rotation of the axes:

$$R^{(t)} = arg\ min_{Q^T Q=I} ||QW^t - W^{t+1}||_F$$

where W^t and W^{t+1} are matrices of word embeddings learn at year t and $t+1$ respectively, and Q is an orthogonal matrix that minimizes the Frobenius norm of the difference between W^t and W^{t+1} [11].

Incremental Updates (IU). We consider the model on the sub-corpus related to Leo XIII (the first pope in the dataset by time), and then we update the model with data of subsequent popes saving its state each time as a new pope model. Each model m_{i+1} is initialized with the word vectors from the previous model m_i [8].

Second Order Embeddings (SO). As proposed in [18], we build second order embeddings by modeling the words by means of their semantic similarity relations with all the other words in the vocabulary. Denoting an embedding of a word w_i at time period t as $w_i(t) \in \mathbb{R}^{100}$ we consider the vectors:

$$\tilde{w}_i(t) = \big(\text{sim}(w_i(t),\ w_1(t)),\ ...,\ \text{sim}(w_i(t),\ w_{|V|}(t))\big)$$

where V is a common vocabulary of all the words in all the time periods and sim is a similarity function such as the cosine similarity. For computational purposes, we define the common vocabulary V by relying on mutual information values computed between words and classes of text (e.g., encyclicals, apostolic exhortations, homilies) associated with each pope. For each class of text, we select the top-500 words by mutual information score. Similarly to the experiment performed in [18], we only keep words associated with nouns, adjectives, and verbs. Furthermore, we exclude stopwords and words shorter than 4 characters.

Semantic shifts detection. Word vectors from distinct time-sliced models cannot be directly compared due to the stochastic nature of Word2Vec. This issue does not preclude the comparison of distances between pair of words over time, which means that it is possible to compare the semantic similarities of a pair of words in distinct models. For the sake of clarity, as an example, we consider the case of temperature. Temperatures from different scales, such as Celsius and

---

[2]https://radimrehurek.com/gensim/

Kelvin, cannot be directly compared. They need to be aligned, i.e., one has to be converted to the scale of the other. However, since scales are related to an additive constant, we can directly compare deltas of temperatures computed in different scales.

Similarly, we decide to exploit:

1. *non-aligned embeddings* to analyze the relative position of word pairs (i.e., the distance between their vectors) in different vector spaces. With respect to the above example, this corresponds to compare temperature deltas in different scales;

2. *aligned embeddings* to analyze the positions of a word over time (i.e., the distance between the vector of that word and itself in distinct aligned vector spaces). With respect to the above example, this corresponds to convert a temperature from a scale $s1$ to another $s2$ before comparing it with another temperature in scale $s2$.

Pairwise word similarity. We exploit non-aligned embeddings to compute the pairwise cosine similarity between a pair of word vectors w_1 and w_2 across time in two different models m_i and m_j. In particular, as we chronologically trained the models pope by pope (they follow each other over time without overlapping) we show how cosine similarity between word vectors could highlight the strength of the relationship in the perspective of different popes.

$$cosine(w_{1_{m_i}}, w_{2_{m_j}}) = \frac{w_{1_{m_i}} \cdot w_{2_{m_j}}}{||w_{1_{m_i}}|| \, ||w_{2_{m_j}}||}$$

Word context comparison. We exploit non-aligned embeddings to explore the context of words. Given a word w, we investigate the top-n words corresponding to the n closest vectors to the vector of w (i.e. the n most similar words to w) in each embedding model. In other words, the n closest vectors to the vector of w are the top-n vectors with highest cosine similarity value from that vector. Besides learning how neighbors change over time for different popes, we estimate the context similarity of a given word w between each pair of popes by computing the jaccard similarity score between the n most similar words to w in their respective models m_i and m_j.

$$jaccard(top\text{-}n_{m_i}, top\text{-}n_{m_j}) = \frac{|top\text{-}n_{m_i} \cap top\text{-}n_{m_j}|}{|top\text{-}n_{m_i} \cup top\text{-}n_{m_j}|}$$

Self word similarity. The need of aligned embeddings rises to mutually compare words over time. By relying on the cosine similarity, we detect meaning change independent from neighboring words by considering the self similarity of a word w throughout consecutive time models m_i, m_{i+1}.

$$cosine(w_{m_i}, w_{m_{i+1}}) = \frac{w_{m_i} \cdot w_{m_{i+1}}}{||w_{m_i}|| \, ||w_{m_{i+1}}||}$$

5. Case study results

In this section, we discuss the results of the approaches presented in Section 4 for semantic shift detection applied to the Vatican corpus of Section 3 according to pairwise word similarity, word context comparison, and self word similarity. similarity.

One of the main problems in evaluating the results is that it is difficult to define a ground truth that provides information about the expected shifts in the Vatican corpus. To address this issue, we run our tests by exploiting three main categories of words.

- Words representing **long-term concepts** in the Vatican publications (e.g., `jesus`, `eucharist`, ...). These terms represent central concepts in the Church, usually related to theological issues. For those terms, then, we expect to observe a limited shift of meaning in the publications of the different popes.
- Words representing **concepts from the past** (e.g., `heresy`, `perversion`, ...). These terms are related to topics that have been central in the Vatican publications in the past, but that nowadays are less present in the popes publications in favour of new words that are more strictly related to events and social phenomena that are perceived as important at the present time. For those terms, we expect to observe a decreasing trend along the temporal dimension.
- Words representing **concepts from today** (e.g., `environment`, `science`, ...), that are the opposite of the concepts from the past, namely words representing concepts that are important nowadays for which we expect to observe a growing trend along time.

For the sake of readability, the considered words from the Vatican corpus are translated from Italian to English.

Pairwise word similarity. In Figure 1, we show examples of word pairs taken from long-term concepts (first row), concepts from the past (second row), and concepts from today (third row), respectively. For each pair of words, we compare their cosine similarity in the models trained on the different popes, exploiting both aligned and non-aligned embeddings. The relevant issue in this experiment is that the comparison between words is based exclusively on their relative position in the vector space. As a consequence, we do not have any information about the stability of the meaning of each single word per se. The only information available is about the meaning of a word with respect to the other in the same pair. As typically occurs for word embedding methods, the proximity assumption holds. Thus, if two words are similar (i.e., their are close in the vector space) we can derive that their meaning is also similar since the two words are used in a similar context.

Concerning long-term concepts (first row of Figure 1), the cosine similarity values for each word pair are stable in time. In particular, we note that the pairs are essentially composed by a word and its consolidated epithet (i.e., `Virgin Mary`, `Jesus Christ`) or alias (i.e., `Eucharist`, also called `Most Blessed Sacrament`). Such similarity values suggest that the meaning shift for these long-term concepts is limited as expected.

For the concepts from the past, the trend of the pairwise similarity between the considered words is decreasing. In particular, we note that pairs having a strong similarity in the past can be characterized by a lower similarity in the publications of recent popes, like ‹`perversion`, `novelty`›. This means that these words were originally used in the same context, but their linguistic and thematic context is changed along time, either because they are no more used together or because one of the two, or both, are now rarely used.

For the concepts from today, we observe the opposite phenomenon. The similarity of word pairs increases over different pontificates, suggesting that the cultural changes that characterize
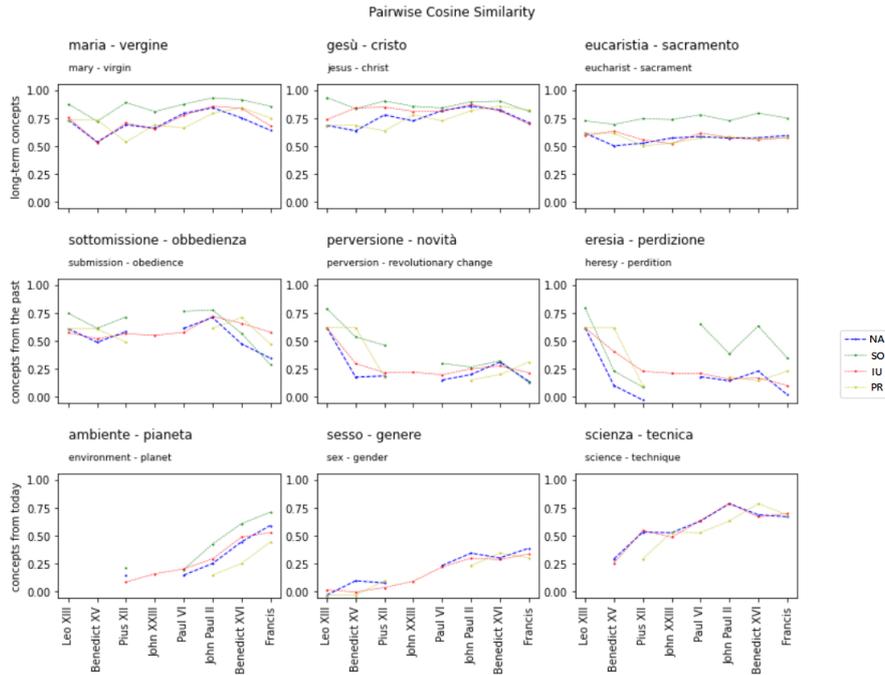
Figure 1: *Pairwise Word Similarity*. The cosine similarity values between pairs of words in different embedding models. The first, second, and third row shows pairs of words whose similarity value is i) almost stable, ii) decreasing, and iii) increasing over time (over different pontificates), respectively. Breaks in curves appear if a vector could not be found corresponding to a word w in a certain model, i.e., in a certain pope, or there was too little evidence for w. The values of the cosine similarity can be found on the y-axis of each plot.

the 21$^{\text{th}}$ century have induced popes to increasingly use the two terms in similar contexts. A pretty clear example of this behavior is given by the pair ‹science, technique› where we note that the trend of the word `technique` is to become almost a synonym of the word `science`, but only after the 70s with John Paul II. In this respect, it is also interesting to consider the new words closely related to a certain pair introduced by a pope in comparison with the dictionary of the previous one. The new words of a pope are determined as the set difference between the subset s of the vocabulary of a pope m_i and the entire dictionary of the previous pope m_{i-1}, where s is the set of the 30 words closest to the mean vector of a certain word pair in the embedding model related to m_i. For example, with respect to the pair ‹environment, planet›, Francis introduced the words `amazonia`, `biodiversity`, `deforestation`, `ecosystem`, `energetic`, and `oceans`. With respect to the pair ‹sex, gender›, Francis introduced the words `mistreatment`, and `homosexuals`; while for the pair ‹science, technique› John Paul II introduced the words `astronomy`, `biology`, `biomedical`, `branch`, `cosmology`, `computer science`, `engineering`, `molecular`, `psychiatry`, `technological`.

As a further remark, we note that the same trend result in the relative position of the considered word pairs can be detected either using aligned or non-aligned embeddings.[3]

---

[3]The breaks in the lines do not appear in IU models due to the main limitation of this approach: the recognition of

Word context comparison. In Figure 2, we consider the target words `jesus`, `environment`, `heresy` and we explore their context composed of 1, 5, and 10 most similar words in the different embedding models, namely the words corresponding to the 1, 5, 10 vectors that closest to the vector of the target. The color gradation describes the intensity of the jaccard similarity value between any pair of popes. Obviously, the diagonal always shows the darkest color since the jaccard similarity value between a pope and himself is equal to 1. About the word `jesus`, the top-1 plot of Figure 2 show that all the popes except Francis share the most similar word. This result confirms the observations of pairwise word similarity reported in Figure 1, where the pair ‹`jesus`, `christ`›is almost unchanged from Leo XIII to Benedict XVI. Also when the
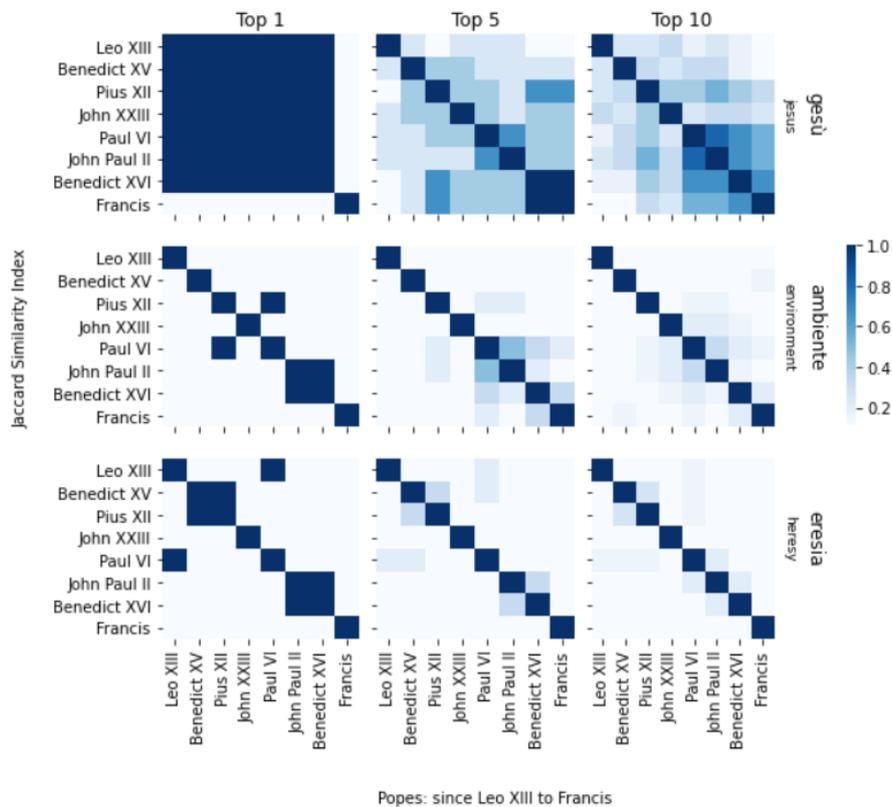

Figure 2: *Word Context Comparison.* Heatmaps of the jaccard similarity values computed between any pair of popes about `jesus` (first row), `environment` (second row), and `heresy` (third row) with respect to their top 1 (left column), 5 (middle column), and 10 (right column) similar words.

contexts of top-5 and top-10 words are considered, the stable behavior of `jesus` can be observed over different pontificates (i.e., many dark areas can be recognized on the first row). On the

---

a word change is possible only if the word has enough occurrences in the considered time period. If the occurrences of a word dramatically decrease (or completely disappear), its word vector will remain the same and hence it is not possible to observe any change [8].

opposite, the words `environment` and `heresy` are affected by semantic shifts. About the word `environment` a shift can be observed in both Paul VI and Pius XII. About the word `heresy`, the shift is less pronounced. The context of the word `heresy` is more similar in the popes of the past, rather than in those of the recent periods. As an exception, the similarity values between John Paul II and Benedict XVI denote a semantic shift in the context of the considered target words. Exploring the closest common words to `heresy` from Benedict XV and Pius XII, we find `nestorio` (i.e., the name of an Archbishop of Constantinople from which Nestorianism - a doctrine condemned as a heretic by the Council of Ephesus in 431 - takes its name), `condamnation`, `apostasy`. When John Paul II and Benedict XVI are considered, the closest common words to `heresy` are `arian` and `arianism` that are about the heresy of Ario, condemned as a heretic by the first council of Nicaea in 325.

Self word similarity. In this experiment, we consider the position of a word with respect to itself, by measuring the similarity of a word vector at time t with respect to the vector of the same word at time $t - 1$. In particular, in Figure 3, we observe the trend of the self cosine similarity for the words `environment`, `travel`, and `progress`.The similarity measures are computed by exploiting both the aligned methods (blue line) and the non-aligned one (green line). Since Leo XIII is the first Pope of our corpus, it is not possible to calculate the self cosine similarity of a word with respect to the model of the previous Pope. For this reason, the lines reported in the figure start from Benedict XV instead of Leo XIII. Since in this experiment we compare a word with itself in different models, we expect to observe high values of similarity with a limited variation. However, this expectation is confirmed only for the aligned methods SO and IU. This is due to the fact that independent models trained on different corpora of different periods can be directly compared only when models are aligned as it occurs in SO and IU. In the case of Procrustes PR, the low values of the self cosine similarity reveal that the alignment mechanism adopted by this method is not suitable for small-sized, unbalanced datasets like the considered Vatican corpus. According to the literature, low values of self similarity can be associated with a semantic change of the considered word, while high values of self similarity denote stable word meanings. As a result, we claim that successive increasing values of self similarity suggest a strengthening of the word meaning, while successive decreasing values of similarity suggest a weakening of the word meaning. About the considered target words, we note that the trends of the self cosine similarity are different for IU and SO models, but they share the increasing/decreasing direction of some shifts, such as for example between Paul VI and John Paul II for the word `environment`. This can be interpreted as a consolidation of the word meaning. Furthermore, both SO and IU models share shifts between John XXIII and Paul VI for the word `progress`, but this behavior is less evident in the SO model. This can be due to the dimensionality reduction applied when the second order embeddings are built.

6. Concluding remarks and future work

In this paper, we considered different approaches to semantic shift detection and we discussed the results obtained on a corpus of Vatican publications related to popes from Leo XIII to Francis (1878-2020). The results show that word embedding can be successfully employed in semantic
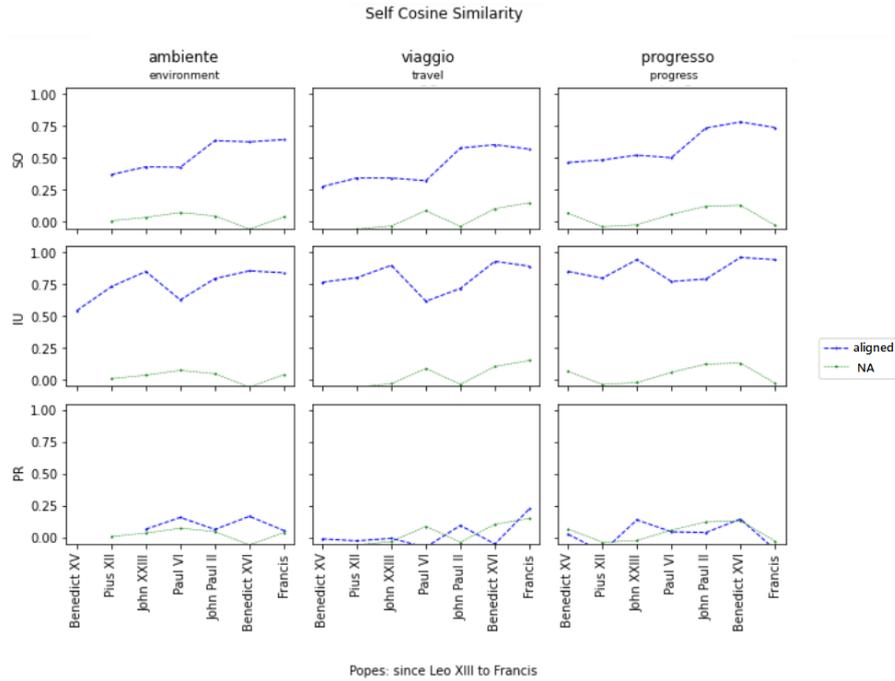
Figure 3: *Self cosine similarity.* The self cosine similarity of the target words `environment`, `travel`, and `progress` calculated over subsequent time/pope models according to Second Order embedding (SO), Incremental Updates (IU), and orthogonal Procrustes (PR) (blue lines). (green) The self cosine similarity of the same target words calculated over subsequent time/pope models according to the non-aligned models (green lines)

shift detection, even when a small-sized, unbalanced dataset is considered like the Vatican corpus. Both aligned and non-aligned approaches have been exploited in the proposed case study. The results reveal that the alignment of embedding models over different vector spaces is not required when we consider pairs of words belonging to different time periods. On the opposite, to successfully detect the meaning shift of a word along time over different vector spaces require the adoption of an alignment mechanism, so that the word vectors belonging to different periods are comparable. However, when alignment approaches are adopted, our results show that the change of a word over time can be noisy and the interpretation of the word behavior can be difficult (e.g., see the case study results of the Procrustes method when the self word similarity is considered).

Ongoing and future work are focused on exploring semantic shift detection techniques by relying on contextualized word embedding models like BERT. In this direction, BERT-like models allow to capture the *sense differentiations* of a target word, meaning that they can detect the different meanings of the considered target according to the different contexts in which the word is used throughout the whole corpus. Furthermore, contextualized embeddings can leverage the benefits of existing pre-trained models, thus avoiding the execution of a (costly) training phase over each time-sliced sub-corpus.

Acknowledgments

References

[1] C.-m. Au Yeung, A. Jatowt, Studying how the Past is Remembered: Towards Computational History Through Large Scale Text Mining, in: Proc. of the CIKM, ACM, 2011, pp. 1231–1240.

[2] J. Bjerva, R. Praet, Word Embeddings Pointing the Way for Late Antiquity, in: Proc. of the LaTeCH, ACL, 2015, pp. 53–57.

[3] N. Tahmasebi, L. Borin, A. Jatowt, Y. Xu, S. Hengchen (Eds.), Computational Approaches to Semantic Change, LSP, 2021.

[4] E. Sagi, S. Kaufmann, B. Clark, Tracing Semantic Change with Latent Semantic Analysis, Current ethods in historical semantics 73 (2011) 161–183.

[5] S. Mitra, R. Mitra, M. Riedl, C. Biemann, A. Mukherjee, P. Goyal, That's sick dude!: Automatic Identification of Word Sense Change across Different Timescales, arXiv preprint arXiv:1405.4392 (2014).

[6] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient Estimation of Word Representations in Vector Space, in: ICLR Workshop Papers, 2013.

[7] H. Gonen, G. Jawahar, D. Seddah, Y. Goldberg, Simple, Interpretable and Stable Method for Detecting Words with Usage Change across Corpora, in: Proc. of ACL, 2020, pp. 538–555.

[8] Y. Kim, Y.-I. Chiu, K. Hanaki, D. Hegde, S. Petrov, Temporal Analysis of Language through Neural Language Models, arXiv preprint arXiv:1405.3515 (2014).

[9] V. Kulkarni, R. Al-Rfou, B. Perozzi, S. Skiena, Statistically Significant Detection of Linguistic Change, in: Proc. of WWW, 2015, pp. 625–635.

[10] Y. Zhang, A. Jatowt, S. Bhowmick, K. Tanaka, Omnia Mutantur, Nihil Interit: Connecting Past with Present by Finding Corresponding Terms Across Time, in: Proc. of ACL, 2015, pp. 645–655.

[11] W. L. Hamilton, J. Leskovec, D. Jurafsky, Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change, arXiv preprint arXiv:1605.09096 (2016).

[12] I. Stewart, D. Arendt, E. Bell, S. Volkova, Measuring, Predicting and Visualizing Short-Term Change in Word Representation and Usage in VKontakte Social Network, in: Proc. of ICWSM, 2017.

[13] A. Englhardt, J. Willkomm, M. Schäler, K. Böhm, Improving Semantic Change Analysis by Combining Word Embeddings and Word Frequencies, International Journal on Digital Libraries 21 (2020) 247–264.

[14] R. Bamler, S. Mandt, Dynamic Word Embeddings, in: Proc. of the ICML, 2017, pp. 380–389.

[15] A. Rosenfeld, K. Erk, Deep Neural Models of Semantic Shift, in: Proc. of the NAACL-HLT, ACL, 2018.

[16] Z. Yao, Y. Sun, W. Ding, N. Rao, H. Xiong, Dynamic Word Embeddings for Evolving Semantic Discovery, in: Proc. of the WSDM, ACM, 2018, pp. 673–681.

[17] H. Dubossarsky, S. Hengchen, N. Tahmasebi, D. Schlechtweg, Time-Out: Temporal

Referencing for Robust Modeling of Lexical Semantic Change, in: Proc. of the ACL, ACL, 2019.

[18] S. Eger, A. Mehler, On the Linearity of Semantic Change: Investigating Meaning Variation via Dynamic Graph Models, arXiv preprint arXiv:1704.02497 (2017).

[19] R. Hu, S. Li, S. Liang, Diachronic Sense Modeling with Deep Contextualized Word Embeddings: An Ecological View, in: Proc. of the ACL, ACL, 2019.

[20] S. Montariol, M. Martinc, L. Pivovarova, Scalable and Interpretable Semantic Change Detection, in: Proc. of the NAACL, ACL, 2021.

[21] D. Schlechtweg, B. McGillivray, S. Hengchen, H. Dubossarsky, N. Tahmasebi, SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection, in: Proc. of the SemEval, ICCL, 2020.