Measuring the Acceleration of the Social Construction of Time using the BOE (Boletin Oficial del Estado)

Elena Fernández Fernández[a], Mirco Schoenfeld[b] and Juergen Pfeffer[a]

[a] *Bavarian School of Public Policy, Technical University of Munich, 80333 Munich, Germany*
[b] *University of Bayreuth, 95447 Bayreuth, Germany*

Abstract

The Practice of Conceptual History, by Reinhart Koselleck, explores the idea that there is a direct relationship between technological advancements and an acceleration in the social construction of time. This paper will quantify this theory by measuring information density and information variety of narratives in a BOE (Boletín Oficial del Estado) dataset of thirty years (1988-2018). Using Quantitative Narrative Analysis, we will define a narrative unit as a triplet of Subject, Verb, Object (SVO), and we will define information density (ID) as the ratio of narrative units per words per year. Afterwards, we will quantify the different contexts of narratives to measure information variety (IV) by constructing a network of semantic closeness from trained word embeddings. This paper will present an increased IV and ID over the observation time, indicating more and more facts being reported. The results will show evidence of an acceleration of the social construction of time.

Keywords

Information Density, Quantitative Narrative Analysis, Word Embedding

1. Introduction

In this paper we present a quantitative computational approach of a theory developed in the fields of History, Sociology, and Media and Communication Studies, that has traditionally been explored qualitatively: the acceleration of the social construction of time.

Natural time could be defined as time rhythms given to us by nature, such as days and nights, seasons, or years. However, humans have developed different technologies that have allowed them to alter such natural time-frames, and have been able to parse time into unnatural smaller units artificially created, such as hours, minutes, or seconds. This artificial segmentation of time has been labeled by some authors [16], as the social construction of time.

Koselleck, as well as many other scholars [29, 34], have proposed the idea that the more technology advances, it is possible to parse time into increasingly smaller segments. Improvements in fields such as chronology, electricity, or transportation systems, indeed allow humans to rush natural time into unnaturally smaller parcels, often with the aim of increasing productivity.

The capacity of parsing time into progressively smaller units has created a sensation of acceleration. This means that, in the same amount of natural time, and due to the upgrading of certain technologies, it is possible to do more things. Let's remember, for example, those days when it was only possible to work with natural light, and the revolution that electricity

CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands
elena.fernandez@hfp.tum.de (E. Fernández Fernández); mirco.schoenfeld@uni-bayreuth.de (M. Schoenfeld); juergen.pfeffer@tum.de (J. Pfeffer)
0000-0001-6596-6349 (E. Fernández Fernández)

brought, as then it was possible to easily increase the number of working hours per day, and therefore, increase productivity. Another example could be improvements in transportation systems, that would allow carrying all sorts of goods faster and therefore quicken the connection of markets worldwide. Enhancements in chronology, communication systems, and broadly speaking, technology, therefore, allow humans to control time for their own benefit, and able them to accomplish higher rates of activity.

This acceleration of the social construction of time has been measured qualitatively by some scholars in Social Sciences and Humanities by witnessing a shortening in the periodization of history. From the nine centuries that the Middle Ages covered, to a twentieth century that was historically parsed in decades, it is possible to observe Koselleck's proposal.

Our main goal is to find a method that is capable of detecting a condensation of narratives over time as well as a diversification of contexts of actors of narratives. Therefore, our paper will operationalize the shortening in the periodization of history quantitatively by measuring both information density of narrative units and information variety of entities related to the narrative units in a BOE (*Boletín Oficial del Estado*) dataset of thirty years (1988-2018).

We will employ the idea that, in order to transform raw text into quantifiable narrative units, the extraction of SVO triplets (Subject, Verb, Object), is a reliable technique that is based on the linguistic properties of text [8]. We will use this definition of narrative units to calculate the ratio of different narrative units per number of words per year as the number of unique SVO triplets normalized by the sum of words. We will use the term Information Density (ID) to describe this measurement.

Our choice of method for measuring ID, Quantitative Narrative Analysis [8] as the baseline of our analysis, instead of other forms of measurement of Information Density, such as entropy, is a well informed decision. We consider that Quantitative Narrative Analysis indeed captures the essence of information density in the context of our thesis.

Afterwards, after normalizing the narrative units (SVO triplets) over the number of words per year, we would like to observe whether there is an increasing number of SVO triplets per year as we become more modern, as a first step to detect information compression. This increase in average SVO triplets over time highlights the idea that there are more narratives being reported in the press, mirroring an increase in rates of human activity. ID, therefore, is a successful way of operationalizing the shortening in the periodization of history, which, as mentioned, is the qualitative method that has been used by some authors [16] to measure an acceleration in the social construction of time.

Besides density of information, we will also measure the diversification of contexts of entities involved in the narratives as Information Variety (IV). To do this, we identify the different semantics in which the actors of the SVO triplets appear supporting the number of narrative units by the context of these narratives. Subsequently, subjects are connected in a network to contextually close words and we measure IV as the average degree of nodes in this network. The semantic contexts of the actors are obtained in the vector space of specially trained word embeddings.

This second step in our investigation stresses the necessity of using a method that is capable of identifying the semantic situation in which a narrative appears. Information Variety, therefore, provides significant additions to our baseline analysis of Information Density. By using word embeddings to observe the narrative environment in which the actors of narratives appear, and finally, building a network, it is possible to quantify the different contexts in which the involved entities appear. Observing a historical increasing number of connections of each node in the network implies that the entities appear in more contexts, which mirrors the thesis of the acceleration of the social construction of time.

The *Boletín Oficial del estado* (previously known as *La Gaceta de Madrid*, and ongoingly published since 1697), provides an excellent medium for scrutinizing information behaviour, and will be the object of study of this paper. Newspapers, whether corporate or state owned, have been accepted as a unit of recorded history in the academic community since the nineteenth century [11, 19].

This paper will therefore provide an innovative approach to the computational methods that can be used to analyze information behaviour over time. By observing a historical increase both in IV and in ID over a constant unit of time (one year), and a constant unit of information (number of words per year), our method will be able to empirically measure higher rates of human activity, or the speed of life, so to speak.

2. Related Work

Our domain knowledge theory, time as an element of study, has been extensively explored by several scholars in the fields of Social Sciences and Humanities. This paper dialogues with several well-known critics, such as the above-named historian Reinhart Koselleck, or the sociologists Judy Wacjman [34] and Hartmut Rosa [29]. Contributions from the field of Media and Communication studies are also considered, such as Robert Hassan [12], or Dylan Mulvin [22]. Other interdisciplinary works such as Jonathay Crary [5], are as well contemplated.

Additionally, Time has been explored as a theoretical field of enquiry across different disciplines, such as Psychology. Taatgen and van Rijn [23] analyze how time is processed by the human mind from a cognitive perspective. We hope that our proposed computational methodology can be used to push the current state of the art of Time across disciplines, as well as to dialogue with diverse theoretical approaches on the topic.

In terms of our methodology, we are conversing with state of the art Natural Language Processing techniques. Traditionally speaking, entropy has been used as an effective approach to measure information density in Computational Linguistics. As such, several works like the ones proposed by [4], [14], or *Redundancy and reduction: Speakers manage syntactic information density* [15], have used this procedure (to the point that Jaeger coined his theory Uniform Information Density Hypothesis in 2006).

Regarding our choice of methodology for measuring Information Density, we believe that not using entropy, but Quantitative Narrative Analysis [8] as the baseline of our analysis, is a well informed decision, as we have already explained. Franzosi's technique has been accepted in the academic community for over a decade now, and has been adapted by several researchers in the fields of Computational Social Sciences and Digital Journalism [33, 32].

Measuring the variety of information over time is closely related to capturing the semantics of text. Before, there have been topic models and, among them, dynamic topic models which were used to capture semantic change [3]. Recently, the task of semantic change detection has experienced a great boost by the use of word representation learning and other machine learning methods [17]. For example, Tang et al. propose a model to learn time-aware vector representations simultaneously solving the problem of aligning different models [35], Frerman et al. propose the detection of semantic change using Bayesian statistics [9], and Basile et al. develop a technique called Temporal Random Indexing [2]. At the same time, approaches based on word embeddings are equally popular [7, 10, 1, 30].

For the scope of our work, however, it is especially interesting how to model the semantics.

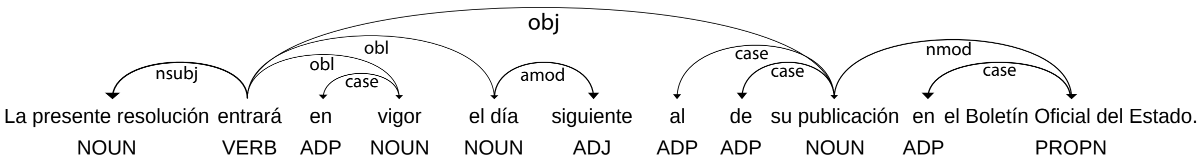


Figure 1: A parse tree of the sentence "La presente resolución entrará en vigor el día siguiente al de su publicación en el Boletín Oficial del Estado." with collapsed noun phrases.

For this task, there is a variety of ways to learn vector representations of words [21, 20, 25, 6, 26]. Generally speaking, in a vector representation model, each word is represented by a vector in a multidimensional vector space containing information about the contexts in which a word is commonly used. On the trained vectors, simple arithmetical operations of linear algebra are possible through which semantic relationships can be expressed.

Shoemark et al. proposed a framework recently which thorougly evaluated different embedding techniques as well as different experimental setups to measure semantic change [31]. We will make use of their findings for our approach of measuring information variety over time.

The use of computational research methods in Social Sciences and Humanities is a relatively recent phenomenon. There are some examples in the fields of Digital Journalism that post similar questions and related methodologies [27]. However, and to our knowledge, we are the first ones who come with a method to analyze fluctuations in information behaviour historically with a corpus of newspaper.

3. Methodology

Measuring the social construction of time consists of two parts, i.e. Information Density (ID) and Information Variety (IV). The former expresses the number of narratives in the texts by quantifying the number of subject-verb-object (SVO) triplets weighted per number of documents per year. The latter expresses the number of different semantics of the involved actors of narratives.

3.1. Information Density: Extraction of SVO Triplets

The extraction of SVO triplets first requires the resolution of grammatical dependencies in a sentence. For this dependency parsing task we used the Python library spacy. Spacy is an open-source natural language processing library that performs tokenization, part-of-speech-tagging and dependency parsing. Its dependency parser offers state-of-the art accuracy [13]. The parser creates a parse tree for a given sentence, in which the main verb represents the root node and the phrases dependent on the verb are inserted as child nodes (see Figure 1 for an example parse tree). Since clauses can depend on other clauses or on the main clause, this is a recursive process and the parse tree may consist of several sub-trees each with a verb as root node.

Accordingly, we traversed the parse trees of the sentences and identified all the verbs. Starting from each verb, we have extracted the nominal subjects and the related direct objects from the tree. In this way, triplets in the form of subject-verb-object were created. Of course, not every grammatically correct sentence contains a nominal subject, i.e. the resulting number of SVO-triplets in a text document is clearly smaller than the number of its sentences.

To describe the development of ID over time, we obtain a unique set of SVO-triplets for any year and normalize the size of that set by the number of words of that year to account for years with smaller numbers of text documents as well as years with shorter but more documents.

3.2. Information Variety: Semantic Network Analysis

The information variety is measured using a network of semantic closeness, i.e. a network in which terms are connected to terms with a semantic relation. To construct this network, the semantics of words must first become quantifiable. For this purpose, we train word embeddings using the gensim [28] implementation of the continuous bag of words (CBOW, [21]).

In order to measure the course of IV over time, a special experimental setup is necessary. In that, we make use of the recommendations of Shoemark et al. [31] which involve training one model per year, in which only the texts from one year are included. From each of the models we derive a separate network of semantic closeness expressing the semantic relationships for a single point in time. Finally, the IV is expressed through the average degree of the nodes of each individual network resulting in a time series of average degrees representing the development of IV over time.

Prior to the training of the underlying models, we perform a pre-processing of the texts. This includes first of all to transfer all letters into their lower case representation. Then all indicators of enumerations (e.g. '1)' or 'A.'), all words with up to two letters ('y', 'el', 'la', ...) and all sentences consisting of only one word are removed. We also remove a number of stop words that are irrelevant for our application. Finally, all numbers are replaced by wildcards.

From the pre-processed data we then train word vectors with 300 dimensions, a window size of 5 words and 15 iterations. We further fix the minimum number of occurrences of words to be 5 meaning each word had to appear at least 5 times in a given year to be included into the embedding. This decision was made to prevent the models to become too noisy. For the remaining hyperparameters we use the default values of gensim.

From the resulting embeddings we derive the networks of semantic closeness. For this purpose, those neighbours of a word in vector space are determined which do not fall below a certain cosine similarity, i.e. those neighbors are semantically particularly close to a given word. For the word and each of the neighboring words, nodes and connecting edges are created in the network, i.e. the node of the word is connected to each node of its neighboring words. The threshold value of cosine similarity, from which a semantic connection between two words is instantiated in the network, is fixed to be 0.9.

Measuring the course of IV over time with a fixed threshold of cosine similarity requires the word embedding vectors to be aligned to the same coordinate axes. This is due to the fact that low-dimensional embeddings might appear orthogonally transformed arbitrarily which does not affect comparisons of cosine similarity within years. But it effectively hinders a comparison of absolute cosine similarity values across years. Hence, we use orthogonal Procrustes to align the low-dimensional embeddings of different years as proposed by [10]. We align all models to the model of the last point in time employing the respective recommendation of Shoemark et al. [31].

In order to focus our measurement of information variety on the actors of the identified narratives, we construct the networks of semantic closeness considering mainly the subjects of the identified SVO triplets. Therefore, we gather a unique list of subjects that are present throughout all years. We then traverse this list for every individual model of word embeddings of one year. As soon as a term is found in vector space, we identify all semantically close terms

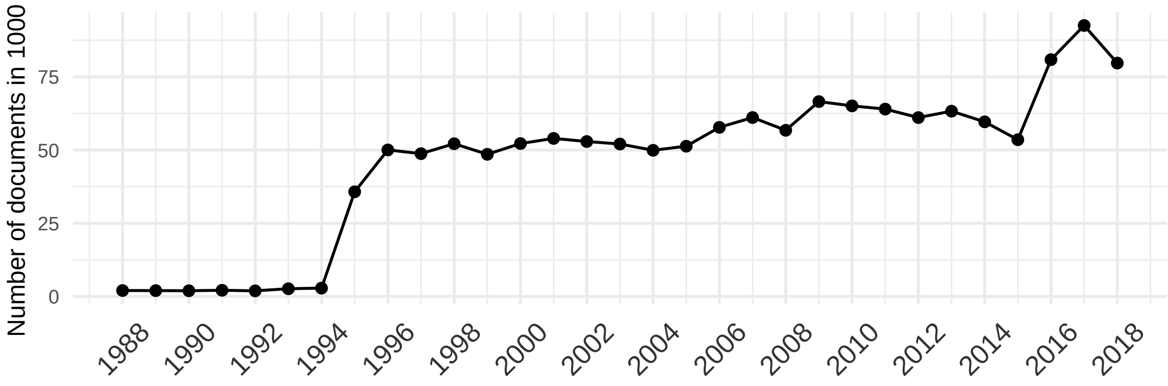


Figure 2: Number of documents over time.

which are closer than the above-mentioned cosine similarity threshold. From this resulting list of words, we add nodes to the network where necessary and instantiate edges between these nodes and the search term, namely the node of the actor.

4. The Corpus

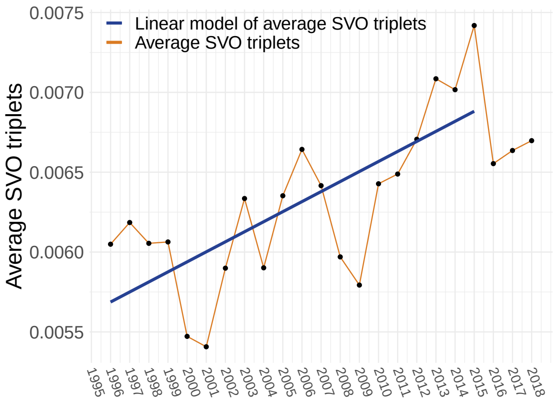
Our selected corpus, *Boletín Oficial del Estado (BOE)*, is the official gazette of the Kingdom of Spain. Dating back to 1661 (officially known as *La Gaceta de Madrid* from 1661 to 1936), the corpus is free and publicly available in a digitized form, therefore, an ideal medium of enquiry for historical analysis. Consequently, we are using a *Boletín Oficial del Estado (BOE)* dataset of thirty years (1988-2018) that we have obtained by using the API facilitated by the Spanish Government. The government maintains an archive of all articles of the BOE that is accessible from its website[1]. Every article is available as a PDF document containing a scan of the original print, and as a text file containing the textual representation of that article. We decided to download text files only and leave out those articles that were available in their PDF representation only. Our corpus has a size of 1.425.072 documents in total. Figure 2 depicts the number of documents over time. For the years 1988–1994, we were able to collect roughly 2070 documents on average per year. This was due to the fact, that early articles were mostly present as PDF only.

We have selected this dataset due to several reasons. Firstly, as indicated by Núñez de Prado [24], this government owned newspaper has been continuously published since 1697. It is therefore an excellent object of analysis where to observe information fluctuations historically. Secondly, it is easily accessible through the Spanish Government website.

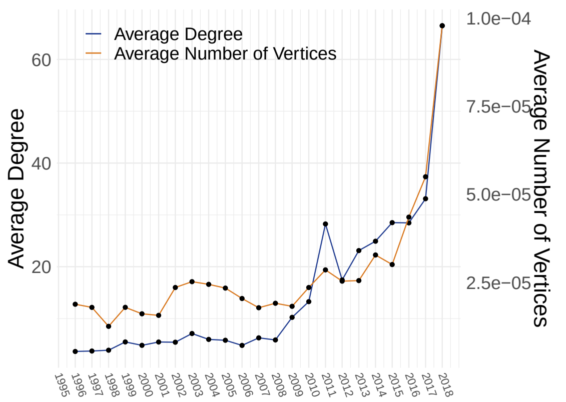
5. Results

One of the challenges that we have experienced in our analysis, has been the small amount of data in the years up to 1995 which can be seen from Figure 2. This led to a high amount of noise in our word embedding models for those years, i.e. our models contained a lot of words in close semantic proximity to each other which might have beeen due to the fact that there

[1]https://boe.es/diario_boe/



(a) Number of average SVO triplets, i.e. ID, (normalized by the number of words per year) and a fitted linear regression model.



(b) Average degree of vertices, i.e. IV, and the average number of vertices in semantic networks (normalized by the number of words per year).

Figure 3: Information Density (ID) and Information Variety (IV) over time

was too few data to separate those words better. Consequently, we decided to exclude those years from our analysis and report our findings for the time span between 1996 and 2018.

After extracting the average SVO triplets using the described techniques, a consistent increment in Information Density can be observed up to the year 2015. Figure 3a shows both the average number of SVO triplets per document and the prediction of a linear regression model fitted to the average number of SVO triplets over time.

The number of SVO triplets was normalized with the number of words per year to control for the length of the documents over time. As shown in Figure 2, the number of documents increases quite constantly. Therefore, normalizing the number of narratives with the total yearly document length better accounts for the possibilities that simply the number of documents increases or that documents become longer.

For the years 2016 – 2018, the chart shows a significant drop in the average number of SVO triplets per document. After inspection of the underlying data we found that in 2016 the number of publications increased noticeably, inter alia, due to announcements of auctions ordered by the court – short documents containing only a link to a website without any complete sentences, and, consequently, lacking any SVO triplets. Hence, we consider the corresponding drop in Figure 3a a platform effect which we will take into account in future work [18]. This is also the reason why the linear regression model ignores the observations from 2016 onwards.

Generally, using ID as a means to operationalize the shortening of the periodization of history quantitatively, is proved to be a successful solution. The overall trend of increasing numbers of weighted SVO triplets shows that there are more weighted narrative units per year as we become more modern, and, therefore, indicates that there are more things happening in the same amount of time.

Koselleck's qualitative analysis of observing a shortening in the periodization of history as a result of the upgrading of certain technologies arriving to society, assumed causality between both events (the acceleration of the social construction of time as a result of technological development). Our operationalization of this causality, ID, provides empirical data of an increase in the speed of life. ID grants the possibility of observing exact fluctuations of information over time and matching them with technological advancements arriving to society in a very concrete manner.

Yet, we acknowledge the fact that a linear increase of ID over time might not be a realistic assumption that led to fitting a linear regression model. We will address this in future work by thoroughly inspecting the causing effects of the spikes in the chart (e.g. in 2003, 2006, 2013, and 2015) and the observable drops (e.g. 2001 & 2009).

The results of the next step of the analysis, the analysis of IV (Information Variety), display as well a stable increasing trend of condensation of information in the selected time frame.

Figure 3b depicts the average degree in our networks, comparing it to the average number of vertices per year. For the average number of vertices, we normalized the number of vertices of the networks by the yearly word count. This results in the chart depicting the yearly share of words that were considered for the networks. At the same time, this puts the average degree in the corresponding networks into better relation.

As it can be seen from Figure 3b, both the relative size of the networks increase over time as well as the average degree of the nodes in the networks. This means that there are more and more actors appearing in the narratives over time which is reflected by the average number of vertices over time. At the same time, these actors are connected to an increasing number of related terms which is shown by the increasing average degree of the nodes of the network. From this Figure, we conclude that an increasing number of actors is reported about in an increasing number of contexts over time.

Our implementations of analysis of Information Density and Information Variety in a government owned newspaper data set of thirty years, have therefore shown a historical augmentation of both measurements. This rise in both ID and IV therefore proof quantitatively the theory of the Social Acceleration of Time proposed in the Social Sciences and Humanities.

Higher rates in human activity are shown by the fact that there are more facts being reported in the news in a constant unit of measurement. It is possible thus to argue that there is relation of causality between an increase in the speed of life as technology becomes more prevalent in society.

6. Conclusions and future work

We have presented a computational method that successfully quantifies fluctuations of information behaviour historically by measuring changes in rates of IV and ID. The rampant increase in the speed of life is exemplified in the increment both in IV and ID over an observational time of thirty years using the same object of study (the BOE). Our method therefore is capable of successfully quantifying Koselleck's theory about the acceleration in the social construction of time.

As we have already noted, the use of computational research methodologies in the fields of Social Sciences and Humanities is a very recent phenomenon. Consequently, we are opening a new line of research: developing computational approaches suitable for measuring information behaviour historically. We believe that the method that we have presented in this paper can have many applications in the research fields of Intellectual History, History of Ideas, and Cultural Studies broadly speaking.

Future work includes applying this method to other corpora in different languages (English, French, and German), as well as in different historical time-frames, in order to perform similar observations and validate our theory using more examples. Our final goal will be to propose how, on the one hand, there is a general trend of increasing rates of ID and IV in all the newspapers chosen (mirroring the increasing rates of human activities world wide); and, on the other hand, how different countries show different ID and IV proportions, which will open new research lines for enquiring how our method could possibly be used to test different rates of industrialization.

Acknowledgments

Thanks very much to Prof. Dylan Mulvin (London School of Economics, Department of Media and Communications), for the theoretical discussions about Time as a discipline of study.

References

[1] R. Bamler and S. Mandt. "Dynamic Word Embeddings". In: *Proc. of the 34th ICML*. Vol. 70. Australia, 2017, pp. 380–389.

[2] P. Basile and B. McGillivray. "Exploiting the Web for Semantic Change Detection". In: *Discovery Science*. Cham: Springer, 2018, pp. 194–208. ISBN: 978-3-030-01771-2.

[3] D. M. Blei and J. D. Lafferty. "Dynamic Topic Models". In: *Proc. of the 23rd ICML*. USA, 2006, pp. 113–120. ISBN: 1595933832. DOI: 10.1145/1143844.1143859.

[4] C. Coupé et al. "Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche". In: *Science Advances* 5.9 (2019).

[5] J. Crary. *24/7: Late Capitalism and the Ends of Sleep*. Verso, 2013.

[6] J. Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *arXiv:1810.04805* (2018).

[7] H. Dubossarsky, D. Weinshall, and E. Grossman. "Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models". In: *Proc. of the 2017 EMNLP*. Denmark, 2017, pp. 1136–1145. DOI: 10.18653/v1/D17-1118.

[8] R. Franzosi. *Quantitative Narrative Analysis*. Sage, 2010.

[9] L. Frermann and M. Lapata. "A Bayesian Model of Diachronic Meaning Change". In: *Transactions of the ACL* 4 (2016), pp. 31–45. DOI: 10.1162/tacl_a_00081.

[10] W. L. Hamilton, J. Leskovec, and D. Jurafsky. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change". In: *Proc. of the 54th Annual Meeting of the ACL*. Germany, 2016, pp. 1489–1501. DOI: 10.18653/v1/P16-1141.

[11] K. A. Hansen and N. Paul. *Future Proofing the News. Preserving the First Drafts of History.* Rowman & Littlefield, 2017.

[12] R. Hassan. *The Age of Distraction: Reading, Writing, and Politics in a High-Speed Networked Economy*. New Brunswick (U.S.A.) ; London (U.K.): Taylor & Francis, 2011.

[13] M. Honnibal and M. Johnson. "An Improved Non-monotonic Transition System for Dependency Parsing". In: *Proc. of the 2015 EMNLP*. Portugal, 2015, pp. 1373–1378. DOI: 10.18653/v1/D15-1162.

[14] T. F. Jaeger. "Redundancy and reduction: Speakers manage syntactic information density". In: *Cognitive Psychology* 61.1 (2010), pp. 23–62. ISSN: 0010-0285. DOI: 10.1016/j.cogpsych.2010.02.002.

[15] T. F. Jaeger. *Redundancy and Syntactic Reduction in Spontaneus Speech.* Stanford University, 2006.

[16] R. Koselleck. *The Practice of Conceptual History: Timing History, Spacing Concepts.* Stanford University Press, 2002.

[17] A. Kutuzov et al. "Diachronic word embeddings and semantic shifts: a survey". In: *Proc. of the 27th COLING*. USA, 2018, pp. 1384–1397.

[18] M. M. Malik and J. Pfeffer. "Identifying platform effects in social media data". In: *10th ICWSM*. Germany, 2016, pp. 241–249.

[19] S. E. Martin and K. A. Hansen. *Newspapers of Record in a Digital Age. From Hot Type to Hot Link*. Praeger, 1998.

[20] T. Mikolov et al. "Distributed Representations of Words and Phrases and their Compositionality". In: *arXiv:1310.4546* (2013).

[21] T. Mikolov et al. "Efficient estimation of word representations in vector space". In: *arXiv:1301.3781* (2013).

[22] D. Mulvin. "Media Prophylaxis: Night Modes and the Politics of Preventing Harm". In: *Information & Culture* 53.2 (2018), pp. 175–202.

[23] H. v. R. Niels A. Taatgen. "An Integrated Theory of Prospective Time Interval Estimation: The Role of Cognition, Attention, and Learning". In: *Pscyhological Review* 114.3 (2007), pp. 577–598.

[24] S. Núñez de Prado. "De la Gaceta de Madrid al Boletín Oficial del Estado". In: *Historia y Comunicación Social* 7 (2002), pp. 147–160.

[25] J. Pennington, R. Socher, and C. Manning. "Glove: Global Vectors for Word Representation". In: *Proc. of the 2014 EMNLP*. Qatar, 2014, pp. 1532–1543. DOI: 10.3115/v1/D14-1162.

[26] M. E. Peters et al. "Deep contextualized word representations". In: *arXiv:1802.05365* (2018).

[27] P. Razgovorov and D. Tomás. "Creación de un corpus de noticias de gran tamano en espanol para el análisis diacrónico y diatópico del uso del lenguaje". In: *Procesamiento del Lenguaje Natural* (2019).

[28] R. Řehůřek and P. Sojka. "Software Framework for Topic Modelling with Large Corpora". English. In: *Proc. of the LREC 2010 Workshop on New Challenges for NLP Frameworks.* Malta, 2010, pp. 45–50.

[29] H. Rosa. *Alienation and Acceleration: Towards a Critical Theory of Late-modern Temporality.* NSU Press, 2010.

[30] M. Rudolph and D. Blei. "Dynamic Embeddings for Language Evolution". In: *Proc. of the 2018 WWW*. France, 2018, pp. 1003–1011. ISBN: 9781450356398. DOI: 10.1145/3178876.3185999.

[31] P. Shoemark et al. "Room to Glo: A Systematic Comparison of Semantic Change Detection Approaches with Word Embeddings". In: *Proc. of the 2019 EMNLP-IJCNLP*. China, 2019, pp. 66–76. DOI: 10.18653/v1/D19-1007.

[32] S. Sudhahar, R. Franzosi, and N. Cristianini. “Automating Quantitative Narrative Analysis of News Data”. In: *Proc. of the 2nd Workshop on Applications of Pattern Analysis*. Proc. of Machine Learning Research. Spain, 2011, pp. 63–71.

[33] W. Van Atteveldt. *Semantic Network Analysis. Techniques for Extracting, Representing, and Querying Media Content*. BookSurge, 2008.

[34] J. Wacjman. *Pressed for Time: The Acceleration of Life in Digital Capitalism*. The University of Chicago Press, 2015.

[35] Z. Yao et al. “Dynamic Word Embeddings for Evolving Semantic Discovery”. In: *Proc. of the 11th WSDM*. USA, 2018, pp. 673–681. ISBN: 9781450355810. DOI: 10.1145/3159652.3159703.