# Computing Semantic Change Using Context Volatility

A. Niekler
University Leipzig
Universität Leipzig, Institut für Informatik, ASV, P818
Augustusplatz 10, Leipzig
+49-341-97-32239
aniekler@informatik.uni-leipzig.de

C. Kahmann
University Leipzig
Universität Leipzig, Institut für Informatik, ASV, P818
Augustusplatz 10, Leipzig
+49-341-97-32239
kahmann@informatik.uni-leipzig.de

G. Heyer
University Leipzig
Universität Leipzig, Institut für Informatik, ASV, P816
Augustusplatz 10, Leipzig
+49-341-97-32231
heyer@informatik.uni-leipzig.de

## ABSTRACT

In this paper we describe problems occurring while modeling the dynamics of terminology context in diachronic corpora and how to solve them. The foundation for this work is context volatility, a measurement for contextual change [4], that we use as our main measuring unit. The computation of context volatility for a word relies on the significance-values of its co-occurrent terms and the corresponding co-occurrence ranks. In the ongoing process its necessary to evaluate these ranks precisely. After applying context volatility according to [4] it can be shown that there are many cases for which a co-occurrence of two words at a specific point of time has no joint occurrence of its terms in other time stamps. This leaves gaps in the assigned ranks for the co-occurrences which must be handled accordingly. Just setting an arbitrary rank is not accompanied by a reasonable model. In this paper we present solutions and ideas to overcome this problem. We show that the use of very sparse term-term matrices leads to undesired results. We use corpus level statistics and a recommender system to recalculate frequencies and co-occurrence statistics on the time slices. Within such a setting we can use the measure context volatility in a way more consistent manner and we propose the idea for a well-defined statistical model based on the presented results.

## Categories and Subject Descriptors

**• Mathematics of computing → Stochastic processes**
**• Information systems → Collaborative filtering**

Keywords: Semantic Change; Context Volatility; Recommender System; Gaussian Process

## 1. INTRODUCTION

Terms in diachronic text corpora may exhibit a very dynamic change in context. As a result, a method for describing patterns for the emergence of new terms but also contextual changes of existing words is needed. The identification of emerging terms or contextual changes is of relevance to applications and analysis in different fields, e.g. political science, marketing studies or technology mining. We analyze this variation not just through macro views, e.g. Dynamic Topic models [1], but more by looking at the key terms which drive the changes at the micro level. Often these hot button words fan the flames of a debate. Our focus for identifying emerging terms is related to the notion of centrality [9]. The main idea of context volatility is considering the change of a words global context to model a change of its usage. The rate of change is indicative of how much the opinion of stakeholders agree/disagree on the appropriate usage of a term. To achieve all this, we need to be able to compute each terms volatility in a reliant and proven way which isn't always possible robustly due to the properties of diachronic text corpora.

In this paper we present strategies to robustly calculate the context volatility preventing any bias. In section 1 we will discuss related work to introduce the notion of context change. The definition of the context volatility measure and involved problems and solutions is given in section 2. Those techniques will be applied in section 2 and we conclude the paper with the idea for a proper statistical model for context volatility in section 4.

## 2. RELATED WORK ON CONTEXT CHANGE

Several studies address the analysis of variation in context of terms in order to detect semantic change and the evolution of terms. Three different approaches to describe contextual variations can be distinguished: (1) methods based on the analysis of patterns and linguistic clues to explain term variations, (2) methods that explore the latent semantic space of single words, and (3) methods for the analysis of topic membership.

(1) Most studies focus on particular terms, and look for linguistic clues and different patterns of variation in their usage to better understand the dynamics of terms such as [8] or [9]. These studies take a particular term as starting point and inspect its neighboring context to classify, analyze and predict changes of usage. In contrast, our approach takes a whole corpus as starting point, and aims at detecting terms that exhibit a high rate of contextual variation for some time.

(2) Distributional properties of text have been used to study the dynamics of terms in diachronic texts. [6] use latent semantics of words in order to create representations of a term's evolution. [6] proposed a similar method, which uses multidimensional scaling to find latent semantic structures, and compare them for different periods. These approaches try to model semantic change over time by setting a certain time period as reference point and comparing the latent semantic space to that reference over time. Terms can thus be compared with respect to their semantic distance or similarity over time. Again, our approach differs from these because we do not start with a fixed set of terms to study and trace their evolution, but rather we want to detect terms in a collection of documents that may be indicative of semantic change.

(3) Assuming a Bayesian approach, topic modeling is another method to analyze the usage of terms and their embeddedness within topics over time [10,11]. These studies identify terms, which have changed in usage and context, and show that this change can be quantified by the probability of a term's membership in a topic cluster within the topic model used. Approaches like the one of [1] model the dynamics of a term's topic membership directly and allow the model to slightly change its co-occurrence structure over time. [12] modify hierarchical Dirichlet processes to measure the changing share of salient topics over time, and thus help to identify topics and terms that for are very prominent for some time. [7] has

extended this approach to identify topics that for some period of time contain rapidly changing terms, and thus can be considered to be indicative of conceptual changes. However, topic model based approaches always require an interpretation of the topics and their context. In effect, the analysis of a term's change is always relative to the interpretation of the global topic cluster, and strongly depends on it. Topic models only generate a macro view on document collections. In order to identify contextual variations, we also need to look at the key terms that drive the changes at the micro level.

In sum, while related work on the dynamics of terms usually starts with a reference (like pre-selected terms, some pre-defined latent semantics structures, or given topic structures), we aim at automatically identifying terms that exhibit a high degree of contextual variation in a diachronic corpus. The typological category of centrality as introduced by [9] tries to capture the observation that central terms simultaneously appear or disappear in a corpus when the key assumptions, or consensus, amongst the stakeholders of a domain change. The measure of context volatility is intended to support exploratory search for such central terms in diachronic corpora, in particular, if we want to identify periods of time that are characterized by substantial semantic transformation. However, we do not claim that our measure quantifies meaning change or semantic change, the measure quantifies the dynamics of a term's contextual information within a diachronic corpus.

# 3. CONTEXT VOLATILITY

The definition of context volatility is introduced in [4]. The computation of context volatility is based on term-term matrices for every time slice derived from a diachronic corpus. Those matrices hold the co-occurrence information for each time slice. One can use different significance-measures such as log-likelihood-ratio, dice or mutual information to represent the co-occurrences. At first it is necessary to compute for every word $w$ of the vocabulary $V$ and every time slice $T$ the set of co-occurrences, e.g. the term-term matrix $C_t$ with co-occurrence weights for every time slice. The matrix has dimension $V$ x $V$. In the next step we determine for every word the rank of all concurrent words for all time slices as a matrix $R_{V,T}$ where the rows represent the ranks of all co-occurent words of $w$ throughout the time slices, e.g. the rows of the matrix. This matrix has dimension $V$ x $T$ and is produced for every word in $V$. The third step is the computation of the context volatility of a word and for a given history $h$ in the time slices $T$ by computing the inter quartile range (IQR) of all ranks that the co-occurents of word $w$ take for all time slices in $h$, e.g. the IQR of a row in $R_{w,T}$, where we limit the row to $t$ elements of $h$. The result is a matrix $CV_{w,T}$, where each row contains the IQR at a time slice $t$ for a given history. The last step computes the global context volatility for a word $w$ by averaging the columns, e.g. all co-occurents in $CV_{w,T}$. This represents the mean IQR for all contextual information about a word and we get an average rate of rank changes for the co-occurrences of a word $w$. The result is a vector $S_w$ which represents the quantity of context change as defined by the context volatility given in the following formula.

$$CV_{w,T} = \frac{1}{C_{w,T}} \sum_{i} IQR\left(Rank(C_{w,i}, T)\right) \quad (1)$$

## 3.1 Limitations

If one tries to apply the IQR to the matrix $R_{V,T}$ the question arises; How to handle missing co-occurrences for a time slice which appear in other time slices and thus in the whole corpus? To apply the context volatility calculation, we therefore must assign an arbitrary rank to unseen co-occurrences. This is indeed the case if we apply the context volatility calculation to the ranks of sparse local co-occurrence information. We could set the rank of unseen co-occurrences to 0 or the maximum rank, e.g. the size of the vocabulary. But this influences the calculation and introduces a bias towards the artificially introduced ranks. This process would not be reasonable because unobservable contexts do not change our understanding of a concept. There are two more naive but not reasonable possibilities. The first option is not applying any rank to those not observable co-occurrences and just ignoring them at all when calculating the IQR. This alternative has a big deficit because there is a massive loss of information. The knowledge about two words not occurring together at all at a particular point of time is a very important information, which should not be excluded. The second option is setting the non-observable co-occurrences to the maximum rank which was applied to a not zero-value in the column $T$ of $R_{V,T}$ . But with this setting the rank is mainly dependent on the number of words which co-occur with the specific word in the specific time slice or the size of the vocabulary. One can imagine a word pair, which only jointly occurs once, having a high IQR just because of the variation of the maximum rank for every time slice. A special case is the situation in which one examines the volatility of a word, which doesn't occur before a certain time slice at all. The maximum rank for this word and the time slices in which it does not occur at all in the documents can't be applied because there is no non zero-value, which one can refer to. Those examples show that the arbitrary assigning of rank values introduces unforeseeable effects. Such being the case, a much more reliant way would be an approximation of the unseen co-occurrences by using the information of the other time slices and the co-occurrences of the same time slice as well.

## 3.2 Global co-occurrence information

The time slices could be seen as local contexts in time. In contrast, the co-occurrence statistics summarized from all documents, e.g. all time slices, form the global context without time dependence. One solution to the problem stated in the above section is the replacement of the missing local ranks by global co-occurrence information from the whole corpus. If the context is not overwritten by local information we use the information found in all documents throughout all time slices. With this procedure we have no missing ranks within the IQR calculations of the context volatility.

## 3.3 Recommender System

Our second approach is utilizing recommender systems in order to fill the matrix $R_{V,T}$. Precisely, we will use the collaborative filtering strategy. The main idea is to detect similar words in our term-term matrix via cousins-similarity of word vectors which should then behave quite analogical. We overtake the missing values mutually throughout the similar words and therefore fill the sparse local term-term-matrices. A recommended word vector can be calculated using matrix factorization or in a more naive way by applying arithmetic mean values.

$$x_{i,j} = \frac{1}{n}\sum_{k=1}^{n} x_k \text{ where } \frac{\sum_{j=1}^{m} x_i x_k}{\sqrt{\sum_{j=1}^{m} x_i^2}\sqrt{\sum_{j=1}^{m} x_k^2}} > mu \quad (2)$$

The calculation of the arithmetic mean is shown in formula 2. One can see the cosine-similarity for two word vectors $x_i, x_k$. Mu describes a threshold parameter, which limits the set of similar word vectors. The arithmetic mean is then calculated from all word vectors which exceed mu in their cosine similarity.

# 4. EXAMPLES

In this section we present examples for the approximation techniques presented in section 2 and 2. The first method uses collaborative filtering in a recommender system setting where we aim to use local (same time slice) information to approximate the unseen co-occurrence information. The second approach utilizes global (all time slices as a whole) information in order to approximate unseen word combinations for a time slice. Both attempts are shown for two independent data sets.

## 4.1 Co-occurrence approximation using a recommender system

The applied example uses a German news-corpus with documents published in 2015. The corpus is part of the Wortschatz-project and consists of 300000 sentences[1]. For simplicity we set our time slices used for volatility computation to be the whole months from January to December. In 2015 some so called "crises" were apparent in the media. Two of them were the "Flüchtlingskrise" (refugee crisis) and the "Griechenlandkrise" (Greece financial crisis). Both have been discussed very controversial in the media. Therefore, we aimed to see whether we are able to measure this controversy via using context volatility. Initially we calculated 12 term-term matrices, one for every month. The approach works for any kind of significance measure for the co-occurrence matrix. The second step was applying a rank to every word vector of the 12 term-term matrices. Having done this, we we used a recommender system approach to fill the zero-entries of at some time co-occurring words. One can imagine the words "Merkel" and "Flüchtling" (refugee) not occurring together in January, but still we know from the other months that these two words form a significant co-occurrence. As for reasons mentioned in section 2 we have disadvantages when computing volatility while keeping the zero-values unadjusted. For example, we therefore aim to replace the zero-entry $X_{Merkel,Flüchtling}$ in the term-term matrix for January. First off, we use the cosine-similarity to identify those word vectors $X_k$ in the term-term matrix, which behave quite analog to $X_{Merkel}$. A few obvious results are the word vectors: $X_{Kanzlerin}$, $X_{CDU}$ and $X_{Regierung}$. Having found the similar word vectors there are different possibilities of recalculating the zero-entry for $X_{Merkel,Flüchtling}$. For example, one can now use the "Flüchtling"-column of all found similar word vectors and compute the arithmetic mean which can be used as replacement for the zero-entry. We could reduce the sparsity of the term-term matrices using this approach from about 98% to 95%. This might not look like a lot, but in fact we have more than doubled the non-zero entries in the local term-term matrices. Subsequently we calculated the context volatility on the observed and approximated entries for the word vectors for a timespan $h$ of three months. Finally, we calculated the mean of the IQR's for one word and one time span.

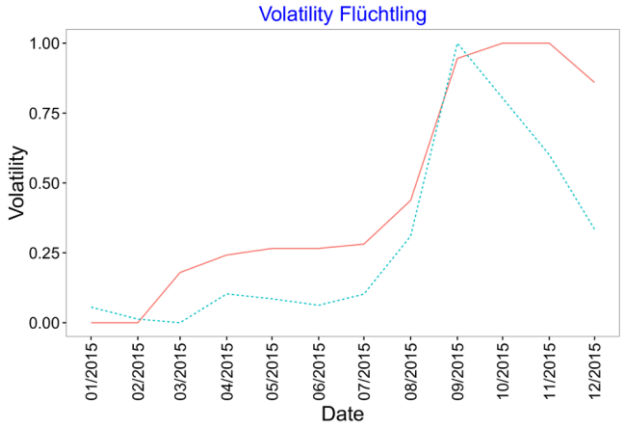


**Figure 1: Context valatility for "Flüchtling" over 12 time spans from January to December 2015**

For example, we calculated the context volatility values of "Flüchtling" and "Griechenland" as shown in fig. 1 and 2. The data points belong to the 12 time spans introduced. The last one describes the months from October to December. In the figures the context volatility is represented as red line and the term frequency is depicted as dotted blue line. According to the results the discussion about refugees heated up around July. At the same time lots of incendiary attacks on refugee accommodations took place, which made the topic very present in the media. Likewise speaking for Greece in the German news, one can see a high context volatility from May until September. In this time span the debate about the third rescue package and the resignation of Alexis Tsipras occurred. Another observation is the fact that the context change not necessarily correlates with the word frequency. For example, the word "Griechenland" is discussed and used in different contexts before June 2015 which could be useful as weak signal for the ongoing events.

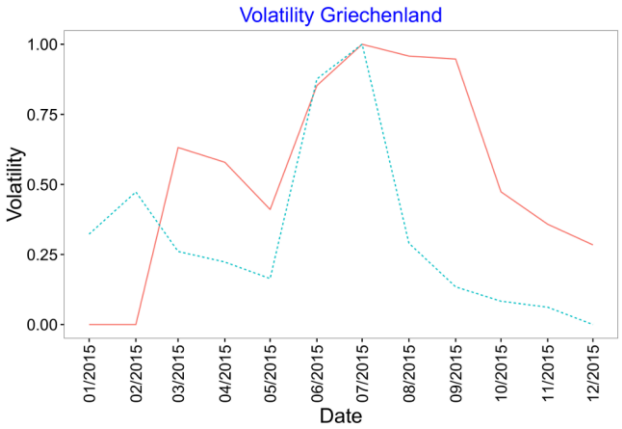


**Figure 2: Context valatility for "Griechenland" over 12 time spans from January to December 2015**

## 4.2 Co-occurrence approximation using global statistics

Our second use case is based on 397,729 articles from altogether 3,841 editions of the German weekly newspaper DIE ZEIT covering the period from 1946 – 2011. We computed in a first step 30 topics based on the Latent Dirichlet Allocation Topic Model (LDA) [2]. Thirty topical fields could be distinguished. Among them, one topic relates to "financial and economic policies" (fig. 3). We will focus on this relation as name for the topic represented by the 30 most probable terms inferred by the LDA model.

| |
|---|
| dollar, milliarde, jahr, prozent, geld, million, gewinn, zins, kredit, markt, fond, pfund, geschäft, kasse, bank, unternehmen, verlust, währung, investor, kunde, umsatz, anteil, konzern, schuld, investition, gold, verkauf, monat, versicherung, kauf |

**Figure 3: Word representation of the topic "financial and economic policies"**

In our case study, we wanted to test whether our context volatility measure is able to recognize the last financial crisis in 2007-2009. In order to do so, we applied the measure on a suitable sub-corpus of the whole data for one topic ("financial and economic policies") and the years 2005 to 2010. The term "Kredit" (loan) is a good example due to its strong context fluctuation within our exemplary

[1] The Wortschatz-project collected more than 250 languages in different corpora sizes. The project can be found under http://wortschatz.uni-leipzig.de/.

issue. The ranges of values of the term frequency and context volatility were aligned in order to overlay both longitudinal plots. We set a history *h* for the calculation of the context volatility of 6 months. The co-occurrence statistics were calculated for each month using the dice significance measure for co-occurrences, which corresponds to monthly time slices. This means that we calculated a context volatility for each word at a time *t* based on the contextual changes from the last 6 months. As one can see in fig. 4 the context volatility for the term "Kredit" (loan) in 2007 is a weak indication for the underlying goings-ons. The usage of the word is already higher than before but the actual number of the co-occurrences in the time slices is very low. Looking at the developments of the financial crisis we should see more context change for this term in 2007. Following the idea that we have a global knowledge about a term which is locally modified within the time slices we should prevent the effects that very sparse co-occurrence information influences the context volatility too much. In fig. 5 the context volatility was calculated using co-occurrence significance values from a global co-occurrence statistic as described in section 2. The context volatility data emphasizes on the year 2007 which was the starting point of the financial crisis. Since we have no artificial values for the ranks but proper global context information the local context change is better represented by the volatility calculation and no bias towards an artificially set rank is introduced. The missing local co-occurrences were simply replaced by the global values. This leads to a co-occurrence matrix for each time slice where all co-occurrences found in the corpus are present and information within the time slices alters the significance values locally.

Fig. 5 also shows that the relative word frequency does not correlate with the context volatility. Apparently, the possible change of context, the discursivity, salience, or centrality of a term, cannot fully be reflected by its frequency of usage. Further interpretations could be that the striking term is discussed from different points of view and context volatility thus reflects controversial discussion, or it can even be considered a weak signal for new adjustments within mainstream or established contexts. Of course, we can also calculate the volatility for the whole time span of the corpus highlighting terms, which appear in different contexts more often than other terms (see tab. 1). Note, that the frequency rank is different from the context volatility rank. The table shows that words like "Kredit" (loan), "Schuld" (debt) or "Risiko" (risk) are ranked higher according their context volatility value. This can also help to extract vocabulary which appears in many contexts besides its high usage as thus can be seen as hot terms for a domain or topic.

**Table 1: Rank of context volatility and frequency**

| Vol. Rank | Word | Volatility | Freq. Rank |
|---|---|---|---|
| 14 | kredit | 0.557370184 | 20 |
| 15 | fond | 0.465242881 | 24 |
| 16 | anleger | 0.451005025 | 22 |
| 17 | markt | 0.422110553 | 16 |
| 18 | investor | 0.382328308 | 26 |
| 19 | zins | 0.365159129 | 28 |
| 20 | wert | 0.164991625 | 41 |
| 21 | risiko | 0.119346734 | 32 |
| 22 | schuld | 0.090452261 | 51 |
| 23 | krise | 0.069932998 | 30 |

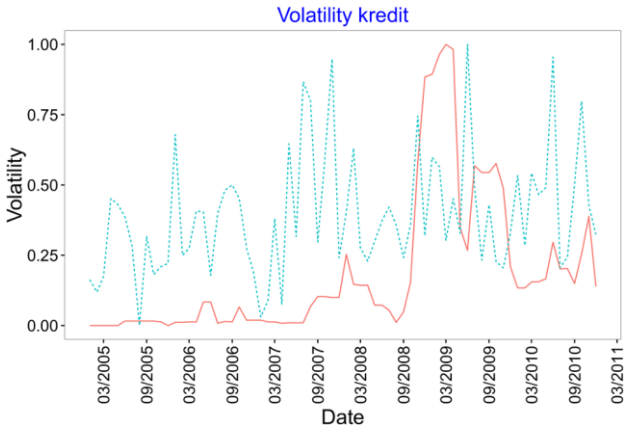


**Figure 4: Context volatility of the term "kredit" calculated with sparse co-occurrence data in the time slices.**

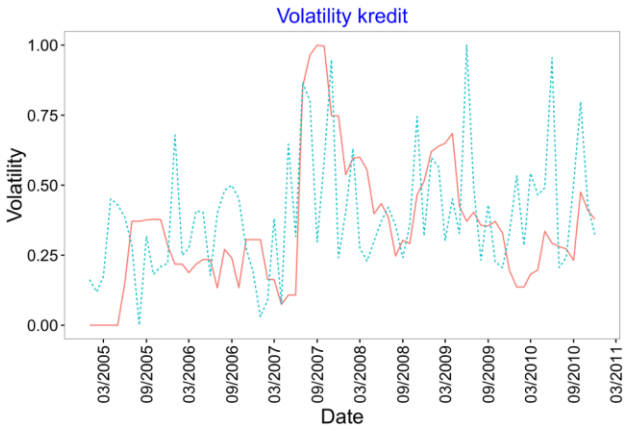


**Figure 5: Context volatility of the term "kredit" calculated with usage of global co-occurrence data in the time slices.**

## 4.3 Further usage for context volatility

Having a reasonable approximated term-term matrix for each time slice improves the context volatility method itself but also enables some even more extensive techniques of analyzing data. For example, we want to get an even deeper look into the analysis of issues and hot topics. Having identified an issue by using context volatility, we can then look for particular terms which led to the high change in the IQR of the identified issue. And as a result we hope to obtain information about what persons/institutions may exhibit a constant behavior in spreading a rumor to a big issue. Other facets, which we want to consider are not just those topics/issues/words, whose context changes over time, but rather those whose context stays almost stable over the whole time slices observed and try to understand why this is the case.

# 5. FUTURE WORK

The measurement of semantic change via computing volatility enables researchers to analyze data in a new manner. Context volatility is able to bare out controversial discussed topics (issues) and accurately describe their progression in the media. But the determination of volatility requires the usage of an uninterrupted data situation in order to produce reliable assertions. Unfortunately, this is very often not the case. Therefor we presented two approaches on how to fill the data sets using either local information (collaborative filtering) or global information (global co-occurrence statistics). Both methods work quite well in limited

situations, but lack a universal applicability. By just using local information of similar word vectors, we can't be sure whether these really behave likely relating to the one word we want to adjust for. Using global statistics, we deny information about the dynamical behavior of the word pair in the sense of assuming the word pairs behave identical over all time slices considered. This is not always the case. Both approaches have their limitations that might produce some bias in the measurement. However, the combination of both, using local as well as global information in order to reliably approximate the term-term matrix, is a promising attempt. For that reason, the development of a Bayesian model which address both solutions is a promising direction to go for. Furthermore, a Gaussian process fulfills all of those requirements. The usage of global information as prior on a Gaussian process which is adjusted by the local information as likelihood would be a reasonable model for this purpose. In more detail, we can train a Gaussian process for every word vector over the time using the word vectors of all co-occurring terms w.r.t. the sparse word vector to approximate. This will allow us to model and approximate the dynamics of the significance of word pairs. This is a huge benefit in comparison to the hypothesis that the significance of word pairs has no describable dynamics and is therefore predictable using the arithmetic mean. Ultimately, this adjusted data situation facilitates the reliable computation and interpretation of context volatility as a measure for semantic change, controversy and term dynamics in diachronic corpora.

# 6. REFERENCES


[1] Blei, D. M. and Lafferty, J. D. (2006). *Dynamic topic models.* In Proceedings of the 23rd international conference on Machine learning.

[2] Blei, D. M., Ng, A.Y., and Jordan, M.I. (2003). *Latent dirichlet allocation*. The Journal of Machine Learning Research 3, 993–1022.

[3] Downs, A. 1972. *Up and down with ecology – the "issue-attention cycle".* Public Interest 28, 38-50.

[4] Heyer, G., Kantner, C., Niekler, A., Overbeck, M. and Wiedemann, G. 2016 *Modeling the dynamics of domain specific terminology in diachronic corpora.* In Proceedings of the 12th International conference on Terminology and Knowledge Engineering (TKE 2016).

[5] Hilpert, M. (2011): *Dynamic Visualizations of Language Change: Motion Charts on the Basis of Bivariate and Multivariate Data from Diachronic Corpora*. International Journal of Corpus Linguistics 16 (4): 435–61.

[6] Jatowt, A. and Duh, K. (2014): *A framework for analyzing semantic change of words across time*. In Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries.

[7] Jähnichen, P. (2015): Topics over time – *A new approach to dynamic topic models*, Ph.D. Thesis, Leipzig University.

[8] Fernández-Silva, S., Freixa J. and Cabré, M. T. (2011): *A proposed method for analysing the dynamics of cognition through term variation*. Terminology 17(1). p. 49-73.

[9] Picton, A. 2011. *Picturing Short-Term Diachronic Phenomena in Specialised Corpora. A Textual Terminology Description of the Dynamics of Knowledge in Space Technologies*. Terminology, 17(1), 134-156.

[10] Rohrdantz, C., Hautli A., Thomas Mayer, Miriam Butt, Daniel A. Keim, and Frans Plank (2011): *Towards Tracking Semantic Change by Visual Analytics*. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2.

[11] Rohrdantz, C., Niekler, A., Hautli A., Butt M. and Keim, D. A. (2012): *Lexical Semantics and Distribution of Suffixes: A Visual Analysis.* In Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH.

[12] Zhang, J. et. al. (2010): *Evolutionary Hierarchical Dirichlet Processes for Multiple Correlated Time-varying Corpora*. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining.