Tracing Research Paradigm Change Using Terminological Methods
A Pilot Study on “Machine Translation” in the ACL Anthology Reference Corpus

Anne-Kathrin Schumann[†] **and Behrang QasemiZadeh**
[†]Applied Linguistics, Translation and Interpreting
Universität des Saarlandes
Campus A2.2, 66123 Saarbrücken, Germany
anne.schumann@mx.uni-saarland.de
behrang.qasemizadeh@insight-centre.org

Abstract

This paper explores the use of terminology extraction methods for detecting paradigmatic changes in scientific articles. We use a statistical method for identifying salient nouns and adjectives that signal these paradigmatic changes. We then employ the extracted lexical units for discovering terms that are assumed to be central in characterising paradigm shifts. To assess the method’s performance, in this pilot study, we work on “machine translation” (MT) research articles sampled from the ACL anthology reference corpus. We analyse this corpus to check whether the proposed approach can trace the dramatic changes that machine translation research has experienced in the last decades: from transformational rule-based methods to statistical machine learning-based techniques.

1 Introduction

Research in computational terminology traditionally focuses on static models of knowledge acquisition and representation. Corpus-based approaches have led to an increased interest in the automatic extraction and semantic categorisation of terms with many successful applications. However, progress in the empirical description and computational modelling of *terminological dynamics* has been rather slow.

This paper suggests that terminological methods and principles can be employed in empirical investigations of *diachronic knowledge evolution*. In particular, terminological methods can provide new insights into problems of diachrony since they can be used to trace (a) how terminologies come into being, and (b) how they develop over time as the scientific field itself evolves. Empirical work on the creation and development of terminologies is especially relevant for investigations into the history of science. Furthermore, studies of this kind are also likely to benefit terminology as a discipline, since they might provide insights into the driving forces of terminological development and knowledge organization.

The method proposed here identifies lexical units the importance of which increases or decreases upon the transition from an earlier period to a more recent one. In other words, we approach history of science in the form of a trend analysis task. Formally, this task consists of two sub-tasks, namely:

(a) the detection of those periods in time when a paradigm change is taking place (e.g., as signalled by terminological dynamics in a domain);

(b) the extraction of terms that are indicative of a declining or rising paradigm.

The pilot study described in this paper relates only to the extraction of terms signalling paradigm shift (i.e., sub-task (b)). The material for our analysis consists of research articles dealing with “machine translation”. These articles are sampled from the ACL Anthology Reference Corpus (ACL ARC)—introduced in Bird et al. (2008).

Linguistically, the proposed method is inspired by studies on *register*.[1] Register linguistics approaches linguistic variation as the description of

[1]See Cabré (1998) for an elaboration of terminological aspects of register. Also, see Teich et al. (2015) for an applied perspective.

changing configurations of linguistic features on the textual level. One of the relevant dimensions for this type of study certainly is the lexicon. Accordingly, we hypothesise that paradigmatic changes in a field of knowledge are the cause of terminological dynamics. These dynamics are expressed in the form of the rise or decline of not just isolated terms but whole groups of terms.

We conclude that terms extracted by our method are salient if they are able to depict the paradigmatic change that the MT field has undergone in the last decades—that is, the advent of statistical methods in contrast to symbolic approaches that were in use earlier. The remainder of this paper is structured as follows. Section 2 briefly summarises relevant previous work. Section 3 outlines our extraction method. Section 4 reports the results of our pilot study, followed by an evaluation in Section 5. Section 6 discusses obtained results and concludes this paper.

2 Related Work

The term "paradigm" in the sense intended here goes back to Kuhn (1962). According to Kuhn, a paradigm emerges from a generally acknowledged scientific contribution to a research field. The significance of the paradigm consists in its ability to propose research problems and solutions to these problems to the relevant community. Some of Kuhn's arguments can be traced back to Fleck (1935). Fleck describes scientific communities as communities of thought ("Denkkollektive") who share habits in their way of perceiving and solving scientific problems ("Denkstil", literally "style of thought"). What is important here for our research question is that paradigms are coupled not only with specific types of problems and research methods, but also with terminologies: they constitute the inventory of lexical units used to refer to concepts that are central for a given paradigm. Consequently, they are subject to change whenever the conceptual outline of the discipline changes.

Terminological dynamics have been approached by terminology proper from various perspectives. Relevant to our study are the articles by Kristiansen (2011) and Picton (2011). Kristiansen (2011) provides a detailed account of external motivating factors of conceptual and, eventually, terminological dynamics. Picton (2011) elaborates a typology for the description of *short-term* term evolution patterns such as *neology–necrology* (i.e., appearance–disappearance of terms), term migration, and topic centrality–disappearance. Both papers, unfortunately, do not provide any methodology for the automatic detection of these dynamics.

In computational linguistics, trend analysis is usually approached by computing topic centrality and/or community influence measures and plotting them on a timeline. An example is the work by Hall et al. (2008) who try to trace the "development of research ideas over time". They employ the standard Latent Dirichlet Allocation (LDA) algorithm (Blei et al., 2003)—a term-by-document model—for identifying "topic clusters". The method involves manual selection of relevant topics and seed words in multiple runs of the LDA algorithm. Probabilities derived from the LDA model are then used for the identification of rising and declining topics. Similar to our work, the authors report experiments over the ACL ARC, using publications from 1978–2006.

A term-based approach to topic and trend analysis is proposed by Mariani et al. (2014). The analysis is conducted on the *ELRA Anthology of LREC publications* starting in 1998. A term extraction method, namely TermoStat (Drouin, 2004), is employed to extract "topic keywords". For each year, terms and their variants are grouped into synsets and the most frequent terms are found. Finally, the authors study the rank development for the 50 most frequent terms in order to extract information on whether topics designated by these terms have risen, declined, or stayed stable over the period under analysis. Relevant co-occurrences of terms are also listed.

Gupta and Manning (2011) stress that for the purpose of detailed investigations into the history of science "... an understanding of more than just the 'topics' of discussion ... " is necessary. They extract semantic information for the categories FOCUS (i.e., the main contribution of an article), TECHNIQUE, and DOMAIN from the title and abstract sentences of research papers using a set of bootstrapped patterns. They then identify *communities* using the LDA algorithm. An influence measure is defined and calculated for communities based on the number of times their FOCUS, DOMAIN, or TECHNIQUE have been adopted by other communities. Finally, results obtained from the ACL ARC are projected onto a timeline.

The work listed above has a number of shortcomings, amongst them are:

- Approaches based on topic modeling do not always provide readily interpretable topics. While many of the induced topics are convincing in terms of their lexical outline, we believe that the use of terminology, as proposed by Mariani et al. (2014), can provide more targeted information.
- For any detailed understanding of the history of a given discipline, it is insufficient to measure how "central" or "popular" certain topics were at different periods in time. Instead, the internal, fine-grained dynamics of the field such as paradigms and paradigm shifts need to be understood. To our knowledge, the work by Mariani et al. (2014) is the only one that includes a study of the lexical context of terminological units; however, this analysis is not carried out systematically. We believe that a systematic study of how groups of terms change over time can provide rich information for users that are interested in the history of a given scientific discipline (e.g., see Figure 2).

3 Detection of Lexical Rank Shifts: The Method

Our work differs from previous studies in that we exploit the notion of rank shifts for detecting fine-grained shifts rather than measuring topic centrality or popularity. The comparison of rank shifts between two lists of sorted lexical items is an established research method in the field of quantitative historical linguistics (e.g., c.f. Arapov and Cherc (1974)) and we believe that it can be adapted to our purposes.

In essence, our approach to the detection of terminological dynamics revealing a paradigm change is two-fold. *Firstly*, we extract lemmas that experience a change in their ranks upon the transition from older publications to more recent ones. We believe that these lemmas are either paradigmatic terms themselves or can be used to extract paradigmatic terms. We restrict word classes to nouns and adjectives since we believe that they are the most characteristic units for a given research paradigm. *Secondly*, we use extracted lemmas for identifying paradigmatic terms.

The first step (i.e., extraction of lemmas) consists of three sub-processes:

1. extraction of frequency per document information for all nouns and adjectives in the two sub-corpora under analysis and removal of strings containing non-alpha-numeric characters;
2. ranking of lexemes obtained for the two time periods using the method explained below;
3. comparison of the two ranked lists in order to identify those lexemes that have undergone relevant rank-shifts.

Frequency and document-related information is extracted using the IMS Open Corpus Workbench (CWB) loaded with our data (Evert and Hardie, 2011). For ranking, we employ the measure for calculating *domain consensus* proposed by Sclano and Velardi (2007). This measure—$DC_{D_i}(t)$—is defined as follows:

$$DC_{D_i}(t) = -\sum_{d_k \in D_i} nf(t, d_k) \log(nf(t, d_k)), \quad (1)$$

where d_k denotes the kth document in domain D_i, and nf is the normalised frequency of term t in $d_k \in D_i$. $DC_{D_i}(t)$ goes beyond the use of raw frequencies (e.g., as used by Mariani et al. (2014)). Instead, $DC_{D_i}(t)$ favors lexemes that are evenly distributed over all the texts in the two sub-corpora as opposed to candidates that are frequent just in a small number of texts. The process results in ranked lists of lexemes for the two time periods that we want to compare. Each lexeme either occurs in only one of the two lists or in both of them. To detect major rank shifts RS for a lexeme t that occurs in both lists, we use the following formula:

$$RS(t) = \frac{1}{R_{New}(t)} - \frac{1}{R_{Old}(t)}, \quad (2)$$

where $R(t)$ denotes the rank of t in the two ranked lists New (recent publications) and Old (early publications).

In the next step, the lemmas with highest rank shifts are employed to build partly lexicalised term extraction patterns for identifying paradigmatic terms. PoS sequence patterns are taken from the

Pattern	CWB query
adjective + noun	[pos=”JJ.*”] [lemma=”lexicon”]
past participle + noun	[pos=”VVN”] [lemma=”lexicon”]
noun + noun	[pos=”N.*”] [lemma=”lexicon”]
noun + noun + noun	[pos=”N.*”] [pos=”N.*”] [lemma=”lexicon”]
noun + preposition + noun	[pos=”N.*”] [pos=”IN”] [lemma=”lexicon”]
adjective + adjective + noun	[pos=”JJ.*”] [pos=”JJ.*”] [lemma=”lexicon”]

Table 1: Examples of partly lexicalised term extraction patterns.

ONLY_NEW	ONLY_OLD
alignment	periphrasing
tag	cannonical
annotation	transcodage
database	transcoded
baseline	pidgin
ontology	sjstem
threshold	descri
monolingual	ption
multilingual	versinn
learning	periphrasin
architecture	paragrapher
engine	subroutine
n-gram	Noninclusive
decoder	inclusiveness
tagger	quelques

(a)

UP	DOWN
word	language
translation	sentence
corpus	structure
model	analysis
result	rule
text	form
method	problem
information	semantic
feature	grammar
system	computer
approach	program
set	theory
training	way
pair	possible
source	dictionary

(b)

Table 2: The result obtained from processing and comparing the *Old* and *New* sub-corpora. Note that dues to the presence of noise in pre-processes (e.g., OCR), the extracted lists of lexemes also contain invalid lexical units such as in Table 2a.

multilingual term extraction tool *TTC TermSuite* (Daille and Blancafort, 2013)[2]. Table 1 provides examples of these patterns.

4 Experiment

As stated earlier, we used the ACL ARC as a dataset. The corpus contains research articles on the topic of human language technology dating back as far as 1965. In our experiments, we use the preprocessed segmented version of the ACL ARC (i.e., the ACL RD-TEC) provided by QasemiZadeh and Handschuh (2014). Our pilot study is limited to the research publications in the domain of MT. Given our knowledge that MT research has undergone a major paradigm shift since the late 1980s, we want to examine whether our method is able to capture and characterise this paradigm shift.

[2] `http://code.google.com/p/ttc-project/`

Up terms	Down terms
machine translation	natural language
language model	deep structure
translation system	phrase structure
word sense	transformational rule
training datum	syntactic analysis
test set	surface structure
mt system	sentence structure
translation model	physics problem
sentence pair	semantic theory
statistical machine translation	transformational grammar
machine translation system	phrase structure grammar
bleu score	average number
parallel corpus	linguistic theory
training set	conversion rule
english word	source language

Table 3: Most frequent paradigmatic term candidates extracted using the proposed lexicalised PoS sequence patterns. We consider *Up terms* and *Down terms* as indicators of topics that are *trending* and *un-trending*, respectively.

To prepare the data for experiments, we extract nouns and adjectives from papers containing either the string “machine translation” or “automatic translation”. We divide the corpus into two sets of articles: *Old* (1960s–70s) and *New* (1980s onwards). Since *New* is substantially larger than *Old*, we randomly reduce the size of the *New* set in order to make it more comparable to *Old*. Despite this effort, the two sub-corpora still have a different size and structure—*New* contains 290,337 nouns and adjectives whereas *Old* contains only 79,247.

The extracted lemmas are weighted using Equations 1 and 2. Consequently, four sets of words are generated:

- words that occur only in *New* (ONLY_NEW);
- words that occur only in *Old* (ONLY_OLD);
- words whose rank increases upon the transition from *Old* to *New* (UP);
- words whose rank decreases upon the transition from *Old* to *New* (DOWN).

The first set—items that occur only in *New*—is comparatively large and contains 14,347 adjectives and nouns. *Old*, on the other hand, has

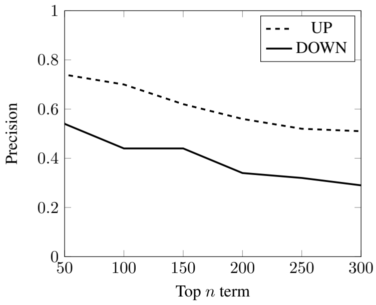

Figure 1: Precision at n for the extracted list of terms using the lexicalised patterns for the Up and Down lemmas.

Up	Down
training	transformational
corpus	routine
score	force
probability	picture
target	location
pair	numeral
evaluation	title
task	reverse
statistical	geometric
source	physics
performance	decimal
bilingual	personal
feature	intension
error	Russian
sense	storage

Table 4: The baseline lemma list: top 15 lemmas sorted by frequency and rank shifts.

7,094 unique adjectives and nouns. 1,023 lemmas have an increased rank over time, and 2,880 words are subject to rank decrease. Table 2 details the results by showing the top 15 items in each set of generated words. Table 2a shows words that occur only in *New* or only in *Old*. Table 2b, however, shows common words with the largest rank shifts. Note that ONLY_NEW and ONLY_OLD have been ranked by their assigned DC score (Equation 1), whereas Up and Down are sorted according to the score computed using Equation 2.

In the *second step*, we select the top 30 plausible noun lemmas from the UP list (shown in Table 2b) and use them for building term extraction patterns (as exemplified in Table 1). This process is also repeated for the top 30 nouns from the DOWN list. The two obtained sets of patterns are employed to extract terms from the *New* and the *Old* sub-corpora, respectively. Table 3 provides an overview over the 15 most frequent candidate terms extracted by this method. Figure 1 reports the precision for the first 300 Up and Down paradigmatic term candidates obtained by automatically comparing them to terms annotated in the ACL RD-TEC by QasemiZadeh and Handschuh (2014).

5 Evaluation

The 15 lemmas listed in Table 2b (i.e., $DC_{D_i}(t)$-ranked lemmas) are presented to 5 researchers in the area of machine translation. The evaluators are asked whether

(a) the individual lemmas in Table 2b are salient for the period they are supposed to represent (*New* and *Old*); and,

(b) the lists as a whole contain words that are typical for the mainstream research paradigms in the respective periods.

To investigate (a), participants make binary distinctions (i.e., in each of the Up and Down lists, a lemma is marked either as relevant or irrelevant). To investigate (b), participants are asked to provide a grade indicating the relevance of the lists of terms on a scale from 1 (“list is irrelevant”) to 5 (“relevant”).

In order to assess whether the $DC_{D_i}(t)$ ranking mechanism proposed in this paper (i.e., Equations 1 and 2) outperforms simpler ranking methods, we also construct a baseline data-set: nouns and adjectives in New and Old are sorted by their frequency and then evaluated by the differences in their ranks. The resulting baseline data is given in Table 4. Evaluators are asked to repeat the above-mentioned assessment also for this baseline without being aware of how both data-sets were produced. Table 5 summarizss the results of this evaluation.

Each row of the Sub-Tables 5 summarises the input from each of the expert evaluators. The first and the second column in each sub-table show the sum of positively marked Up and Down items—that is, the sum of those lemmas (out of 15) that were found salient for either the 1960s–1970s or the 1980s–2000s (sub-task (a)). The third column presents the overall evaluation of the lists (i.e., sub-task (b)). Table 5a provides the results for the list of lexical items that are ranked using the $DC_{D_i}(t)$ score (i.e., listed in Table 2b). Table 5b provides the assessments for the

Up	Down	Overall
12	10	4:5
12	10	4:5
13	12	4:5
10	10	3:5
3	4	2:5

(a)

UP	DOWN	Overall
15	5	3:5
11	2	3:5
14	11	3:5
11	6	4:5
6	2	3:5

(b)

Table 5: Each row of these tables summarises the assessment of each of the evaluators. Table 5a shows the results for the sets of lexical items ranked by $DC_{D_i}(t)$ (listed in Tables 2b). Table 5b, in contrast, provides the result for the sets of lexical items that are sorted by their raw frequencies (listed in Tables 4).

baseline list (i.e., listed in Table 4).

As can be observed in Table 5, the evaluators tend to prefer the $DC_{D_i}(t)$-ranked lexical items over the baseline data-set. Except for one of the annotators who suggests that the baseline method provides more informative output (i.e., the last row of Tables 5a and 5b), the evaluators consistently prefer the ranking mechanism proposed in this paper, assigning an overall grade of 3–4 (out of 5) points to the output. However, the difference remains but slight.

Table 6 shows the 15 most frequent terms in the *Old* and the *New* corpus, respectively. These terms were collected using the manual annotations in the ACL RD-TEC by QasemiZadeh and Handschuh (2014). By comparing these terms to the output of our method (Table 3), we observe considerable differences. Evidently, for the detection of paradigm shifts, terms extracted using semi-lexicalised part-of-speech (PoS) patterns based on our $DC_{D_i}(t)$ method are better indicators of the paradigm shift than terms ranked by their raw frequencies.

Figure 2 exemplifies some of the dynamics detected by our method. For each year, the plot shows the frequencies of terms normalised by the sum of all term frequencies extracted from the publications in that year. All plotted terms were among the top items in our Up and Down lists. Up paradigmatic terms are given in blue whereas Down paradigmatic terms are plotted in black.

Figure 2 illustrates what types of information can be drawn from the analysis conducted here. For example, we observe that “automatic evaluation” rises synchronously with “Bleu score”

Sub-Corpus Old	Sub-Corpus New
natural language	machine translation
machine translation	natural language
computational linguistics	language processing
data base	translation system
artificial intelligence	target language
language processing	computational linguistics
phrase structure	natural language processing
syntactic analysis	training data
translation system	source language
automatic translation	test set
natural languages	information retrieval
information retrieval	machine translation system
noun phrase	language model
language understanding	training corpus
noun phrases	noun phrase

Table 6: 15 most frequent terms (two tokens or longer) in the Old and the New sub-corpora. This list was collected using the manual annotations in the ACL RD-TEC and from the documents in the two Old and New sub-corpora.

and is only slightly preceded by “statistical machine translation” itself. We also find that, during the 1980s, references to “linguistic‘theory” were rather frequent, but they have largely vanished since 1990. Themes such as generative grammar or phrase structure grammar were not dominant even in the earlier decades, but they exhibit a constant decline at least since the 1990s. Evidently, the plot confirms that our attribution of terms to the categories Up and Down is justified. Moreover, this plot supports our hypothesis that paradigm shifts are lexically expressed by dynamics of whole groups of related terms.

6 Discussion and future work

For a detailed understanding of the dynamics of science, it is insufficient to measure how “central” or “popular” certain topics are at different periods of time. Instead, those groups of terms that signal paradigm changes must be detected—this is the key idea that motivates the research presented in this paper. The pilot study described here, therefore, aims at showing that terminological methods can be employed to serve this purpose, and to provide information for understanding what is going on in a scientific field at a given moment in time.

An inspection of our method’s output indicates that the renewal of vocabulary (happening by some words falling from use and others being in-

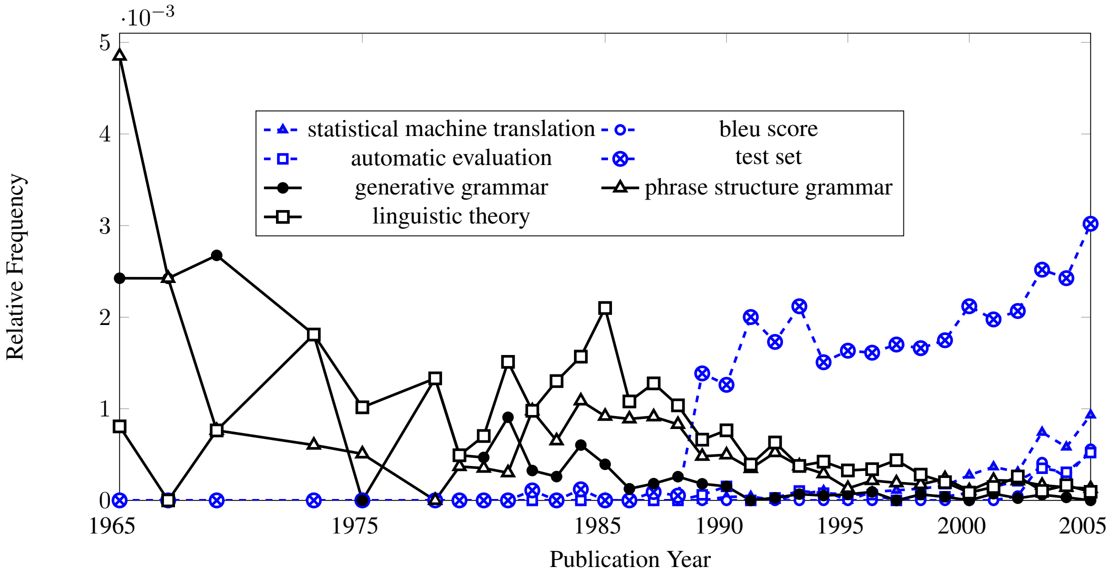

Figure 2: Terms mapped onto a timeline: For each year, the y-axis shows the frequencies of terms normalised by the sum of the frequencies of all the terms extracted in that year.

troduced) is considerable given the relatively short time span under analysis in our experiments. We observe that the content words shared by the two data sets are, in fact, a minority. However, we also observe that *Only_New* (Table 2a) clearly contains items that are indicative of more recent MT research such as "alignment", "n-gram" or "decoder". The items that are specific to *Only_Old*, on the other hand, seem to be rather spurious and low-frequent. These lexical units, rather unsurprisingly, disappear upon the transition from *Old* to *New*.

Our evaluation also indicates that the lemmas extracted by our method (Table 2b) are indicative of the respective time periods, at least as far as the top ranks are concerned. MT experts prefer the output of our proposed method over the output of the baseline method, perhaps due to the improved coverage of the relevant Down lemmas.

Moreover, the terminological evaluation of the extracted paradigmatic terms (Figure 1) shows that Up lemmas indeed help to extract valid computational linguistics terms. Performance for Down lemmas, however, is consistently worse. This difference in performance, in our opinion, is related to the higher productivity of the Up lemmas from Table 2b: Up lemmas are used in a growing number of more specific and more frequent terms, whereas Down lemmas do not experience a similar increase in frequency and specificity. That is, it is harder to distinguish irrelevant collocations containing Down terms from collocations with terminological value. Hence, term extraction performance for Down terms is worse. We believe that, if this property can be shown to hold in general, it is highly relevant as it can be used for the extraction of emergent and semantically related terms. Term extraction performance itself can be further improved by integrating standard practices such as stop-word filtering.

Last not but not least, a timeline plot of Up and Down paradigmatic terms indicates that Down terms, as expected, do not exhibit the same exponential growth as Up paradigmatic terms. However, what we also observe is that many relevant terms do not simply fall from use (e.g., the term "linguistic theory"). They may even increase their absolute frequency or become salient again in new or unforeseen contexts.

The local context of terms therefore remains an unexplored factor in trend analysis research. If we look more closely into our data, we find unexpected formulations such as "the language model in the human" or "translation model based on semantic interpretation". Future work will need to address these kinds of dynamics in superficially identical terms that are even more fine-grained than the rank shifts observed in this pilot study.

Several measures can be taken into consideration for improving our current evaluation method. Future work will also strive for a comparison of multiple sub-corpora that represent time slices of different granularity, perhaps of more similar size and structure. The detection of time periods in which paradigm shifts occur and a more precise modelling of their interplay with terminological dynamics are also important topics for future research.

Finally, we would like to mention that an important observation about the dependence of lexical dynamics on frequency has already been made by Arapov and Cherc (1974) who explicitly refer to Zipf:

> The speed of decay . . . can, in a way, be understood as the probability of decay. The higher the ordinal number (rank) of a [word] group . . . , the lower the frequency of the words belonging to that group, the higher is the speed of decay of this group.[3]

It is no surprise that term frequency does play a role in term necrology. However, the formula that we currently use for rank comparison (i.e., Equation 2) does not account for this aspect. Furthermore, the question how to compare terms the frequencies of which differ by sizes of magnitude is also yet unresolved. Future work will address these shortcomings.

Acknowledgements

We thank Mihael Arcan, Iacer Calixto, Peyman Passban, Liling Tan and colleagues for evaluating our data. We would also like to thank Prof. Elke Teich for her comments and advice. This research has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the Cluster of Excellence 'Multimodal Computing and Interaction'.

References

M. V. Arapov and M. M. Cherc. 1974. *Matematičeskie metody v istoričeskoj lingvistike*. Nauka.

Steven Bird, Robert Dale, Bonnie Dorr, Bryan Gibson, Mark Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. 2008. The acl anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. In *Proceedings of LREC'08*, Marrakech, Morocco, may. ELRA.

David Blei, Andrew Ng, and Michael Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, (3).

Maria Teresa Cabré. 1998. Do we need an autonomous theory of terms? *Terminology*, 2(5).

Béatrice Daille and Helena Blancafort. 2013. Knowledge-poor and Knowledge-rich Approaches for Multilingual Terminology Extraction. In *Cicling*.

Patrick Drouin. 2004. Detection of Domain Specific Terminology Using Corpora Comparison. In *LREC*.

Stefan Evert and Andrew Hardie. 2011. Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. In *Corpus Linguistics*.

Ludwik Fleck. 1935. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Schwabe.

Sonal Gupta and Christopher D. Manning. 2011. Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers. In *IJCNLP*.

David Hall, Daniel Jurafsky, and Christopher D. Manning. 2008. Studying the History of Ideas Using Topic Models. In *EMNLP*.

Marita Kristiansen. 2011. Domain dynamics in scholarly areas: How external pressure may cause concept and term changes. *Terminology*, 17(1).

Thomas S. Kuhn. 1962. *The Structure of Scientific Revolutions*. University of Chicago.

Joseph Mariani, Patrick Paroubek, Gil Francopoulo, and Olivier Hamon. 2014. Rediscovering 15 Years of Discoveries in Language Resources and Evaluation: The LREC Anthology Analysis. In *LREC*.

Aurelie Picton. 2011. Picturing short-period diachronic phenomena in specialised corpora: A textual terminology description of the dynamics of knowledge in space technologies. *Terminology*, 17(1).

Behrang QasemiZadeh and Siegfried Handschuh. 2014. The ACL RD-TEC: A Dataset for Benchmarking Terminology Extraction and Classification in Computational Linguistics. In *Computerm*.

Francesco Sclano and Paola Velardi. 2007. Termextractor: A Web Application to Learn the Shared Terminology of Emergent Web Communities. In *Enterprise Interoperability II: New Challenges and Approaches*. Springer.

Elke Teich, Stefania Degaetano-Ortlieb, Peter Fankhauser, Hannah Kermes, and Ekaterina Lapshinova-Koltunski. 2015. The Linguistic Construal of Disciplinarity: A Data-Mining Approach Using Register Features. *J. Assoc. Inf. Sci. Technol.*

[3] Translated from Russian.