Rediscovering ACL Discoveries Through the Lens of ACL Anthology Network Citing Sentences

Dragomir Radev
EECS Department
University of Michigan
Ann Arbor, MI, USA
radev@umich.edu

Amjad Abu-Jbara
EECS Department
University of Michigan
Ann Arbor, MI, USA
amjbara@umich.edu

Abstract

The ACL Anthology Network (AAN)[1] is a comprehensive manually curated networked database of citations and collaborations in the field of Computational Linguistics. Each citation edge in AAN is associated with one or more citing sentences. A citing sentence is one that appears in a scientific article and contains an explicit reference to another article. In this paper, we shed the light on the usefulness of AAN citing sentences for understanding research trends and summarizing previous discoveries and contributions. We also propose and motivate several different uses and applications of citing sentences.

1 Introduction

The ACL Anthology[2] is one of the most successful initiatives of the Association for Computational Linguistics (ACL). It was initiated by Steven Bird in 2001 and is now maintained by Min-Yen Kan. It includes all papers published by ACL and related organizations as well as the Computational Linguistics journal over a period of four decades.

The ACL Anthology Network (AAN) is another successful initiative built on top of the ACL Anthology. It was started in 2007 by our group (Radev et al., 2009) at the University of Michigan. AAN provides citation and collaboration networks of the articles included in the ACL Anthology (excluding book reviews). AAN also includes rankings of papers and authors based on their centrality statistics in the citation and collaboration networks. It also includes the citing sentences associated with each citation link. These sentences were extracted automatically using pattern matching and then cleaned manually. Table 1 shows some statistics of the current release of AAN.

The text surrounding citations in scientific publications has been studied and used in previous work. Nanba and Okumura (1999) used the term *citing area* to refer to citing sentences. They define the *citing area* as the succession of sentences that appear around the location of a given reference in a scientific paper and has connection to it. They proposed a rule-based algorithm to identify the *citing area* of a given reference. In (Nanba et al., 2000) they use their citing area identification algorithm to identify the purpose of citation (i.e. the author's reason for citing a given paper.)

Nakov et al. (2004) use the term *citances* to refer to citing sentences. They explored several different uses of *citances* including the creation of training and testing data for semantic analysis, synonym set creation, database curation, summarization, and information retrieval.

Other previous studies have used citing sentences in various applications such as: scientific paper summarization (Elkiss et al., 2008; Qazvinian and Radev, 2008; Mei and Zhai, 2008; Qazvinian et al., 2010; Qazvinian and Radev, 2010; Abu-Jbara and Radev, 2011a), automatic survey generation (Nanba et al., 2000; Mohammad et al., 2009), and citation function classification (Nanba et al., 2000; Teufel et al., 2006; Siddharthan and Teufel, 2007; Teufel, 2007).

[1] http://clair.si.umich.edu/anthology/

[2] http://www.aclweb.org/anthology-new/

Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries, pages 1–12,
Jeju, Republic of Korea, 10 July 2012.

Number of papers	18,290
Number of authors	14,799
Number of venues	341
Number of paper citations	84,237
Citation network diameter	22
Collaboration network diameter	15
Number of citing sentences	77,753

Table 1: Statistics of AAN 2011 release

Comparison	Contrast/Comparison in Results, Method, or Goals
Basis	Author uses cited work as basis or starting point
Use	Author uses tools, algorithms, data, or definitions
Description	Neutral description of cited work
Weakness	Limitation or weakness of cited work

Table 2: Annotation scheme for citation purpose

In this paper, we focus on the usefulness of the citing sentences included in AAN. We propose several uses of citing sentences such as analyzing the trends of research, understanding the impact of research and how this impact changes over time, summarizing the contributions of a researcher, summarizing the discoveries in a certain research field, and providing high quality data for Natural Language Processing tasks. In the rest of this paper we present some of these ideas and provide examples from AAN to demonstrate their applicability. Some of these ideas have been explored in previous work, but we believe that they still need further exploration. However, most of the ideas are novel to our knowledge. We present our ideas in the following sections.

2 Temporal Analysis of Citations

The interest in studying citations stems from the fact that bibliometric measures are commonly used to estimate the impact of a researcher's work (Borgman and Furner, 2002; Luukkonen, 1992). Several previous studies have performed temporal analysis of citation links (Amblard et al., 2011; Mazloumian et al., 2011; Redner, 2005) to see how the impact of research and the relations between research topics evolve overtime. These studies focused on observing how the number of incoming citations to a given article or a set of related articles change over time. However, the number of incoming citations is often not the only factor that changes with time. We believe that analyzing the text of citing sentences allows researchers to observe the change in other dimensions such as the purpose of citation, the polarity of citations, and the research trends. The following subsections discuss some of these dimensions.

2.1 Temporal Analysis of Citation Purpose

Teufel et al. (2006) has shown that the purpose of citation can be determined by analyzing the text of citing sentences. We hypothesize that performing a temporal analysis of the purpose for citing a paper gives a better picture about its impact. As a proof of concept, we annotated all the citing sentences in AAN that cite the top 10 cited papers from the 1980's with *citation purpose* labels. The labels we used for annotation are based on Teufel et al.'s annotation scheme and are described in Table 2. We counted the number of times the paper was cited for each *purpose* in each year since its publication date. This analysis revealed interesting observations about the paper impacts. We will discuss these observations in Section 2.3. Figure 1 shows the change in the ratio of each purpose with time for Shieber's (1985) work on parsing.

2.2 Temporal Analysis of Citation Polarity

The bibliometric measures that are used to estimate the impact of research are often computed based on the number of citations it received. This number is taken as a proxy for the relevance and the quality of the published work. It, however, ignores the fact that citations do not necessarily always represent positive feedback. Many of the citations that a publication receives are neutral citations, and citations that represent negative criticism are not uncommon. To validate this intuition, we annotated about 2000 citing sentences from AAN for citation polarity. We found that only 30% of citations are positive, 4.3% are negative, and the rest are neutral. In another published study, Athar (2011) annotated 8736 citations from AAN with their polarity and found that only 10% of citations are positive, 3% are negative and the rest were all neutral. We believe that considering the polarity of citations when conducting temporal analysis of citations gives more insight about

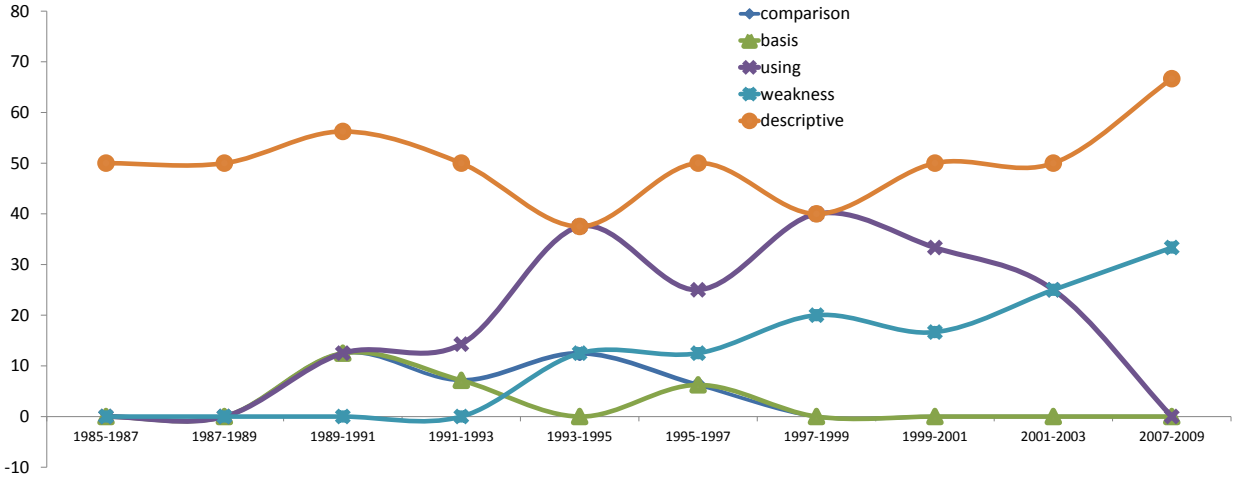

Figure 1: Change in the citation purpose of Shieber (1985) paper

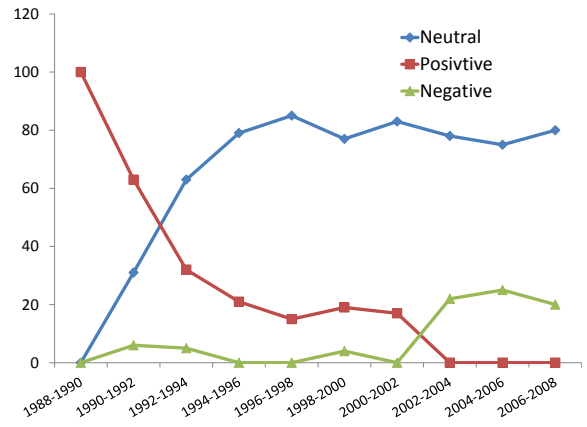

Figure 2: Change in the polarity of the sentences citing Church (1988) paper

how the way a published work is perceived by the research community over time. As a proof of concept, we annotated the polarity of citing sentences for the top 10 cited papers in AAN that were published in the 1980's. We split the year range of citations into two-year slots and counted the number of positive, negative, and neutral citations that each paper received during that time slot. We observed how the ratios of each category changed overtime. Figure 2 shows the result of this analysis when applied to the work of Kenneth Church (1988) on part-of-speech tagging.

2.3 Predict Emergence of New Techniques or Decline of Impact of Old Techniques.

The ideas discussed in Sections 2.1 and 2.2 and the results illustrated in Figures 1 and 2 suggest that studying the change in citation purpose and citation polarity allow us to predict the emergence of new techniques or the decline in impact of old techniques. For example, the analysis illustrated in Figure 2 shows that the work of Ken Church (1988) on part-of-speech tagging received significant positive feedback during the 1990s and until early 2000s before it started to receive more negative feedback. This probably can be explained by the emergence of better statistical models for part-of-speech (POS) tagging (e.g. Conditional Random Fields (Lafferty et al., 2001)) that outperformed Church's approach. However, as indicated by the neutral citation curve, Church's work continued to be cited as a classical pioneering research on the POS tagging task, but not as the state-of-the-art approach. Similar analysis can be applied to the change in citation purpose of Shieber (1985) as illustrated in Figure 1

2.4 Study the Dynamics of Research

In recent research, Gupta and Manning (2011) conducted a study that tries to understand the dynamics of research in computational linguistics (CL). They analyzed the abstracts of CL papers included in the ACL Anthology Reference Corpus. They extracted the contributions, the domain of application, and the

	apply	propose	extend	system
Abstracts	1368	2856	425	5065
Citing Sentences	**2534**	**3902**	**917**	**6633**

Table 3: Comparison of trigger word occurrences in abstracts vs citing sentences.

techniques and tools used in each paper. They combined this information with pre-calculated article-to-community assignments to study the influence of a community on others in terms of techniques borrowed and the maturing of some communities to solve problems from other domains. We hypothesize that conducting such an analysis using the citing sentences of papers instead of (or in combination with) abstracts leads to a more accurate picture of research dynamics and the interaction between different research communities. There are several intuitions that support this hypothesis.

First, previous research (Elkiss et al., 2008) has shown that the citing sentences that cite a paper are more focused and more concise than the paper abstract, and that they consistently contain additional information that does not appear in abstracts. This means that additional characteristics of a paper can be extracted from citing sentences that cannot be extracted from abstracts. To verify this, we compared abstracts vs citing sentences (within AAN) in terms of the number of occurrences of the *trigger words* that Gupta and Manning (2011) deemed to be indicative of paper characteristics (Table 3). All the abstracts and citing sentences included in the 2011 release of AAN were used to get these numbers. The numbers clearly show that the trigger words appear more frequently in the set of citing sentences of papers than they do in the paper abstracts. We also found many papers that none of the *trigger words* appeared in their abstracts, while they do appear in their citing sentences. This suggests that more paper properties (contributions, techniques used, etc.) could be extracted from citations than from abstracts.

Second, while the contributions included in an abstract are the claims of the paper author(s), the contributions highlighted in citing sentences are collectively deemed to be important by peer researchers. This means that the contributions extracted from citations are more important from the viewpoint of the community and are likely to reflect research trends more accurately.

	Rank		
word	1980s	1990s	2000s
grammar	22	71	123
model	75	72	26
rules	77	89	148
statistical	-	69	74
syntax	257	1018	683
summarization	-	880	359

Table 4: Ranks of selected keywords in citing sentences to papers published in 80s, 90s and 2000s

We performed another simple experiment that demonstrates the use of citing sentences to track the changes in the focus of research. We split the set of citing sentences in AAN into three subsets: the set of citing sentences that cite papers from 1980s, the set of citing sentences that cite papers from 1990s, and the set of citing sentences that cite papers from 2000s. We counted the frequencies of words in each of the three sets. Then, we ranked the words in each set by the decreasing order of their frequencies. We selected a number of keywords and compared their ranks in the three year ranges. Some of these keywords are listed in Table 4. This analysis shows, for example, that there was more focus on "grammar" in the computational linguistics research in the 1980s then this focus declined with time as indicated by the lower rank of the keyword "grammar" in the 1990s and 2000s. Similarly, rule based methods were popular in the 1980s and 1990s but their popularity declined significantly in the 2000s.

3 Scientific Literature Summarization Using Citing Sentences

The fact that citing sentences cover different aspects of the cited paper and highlight its most important contributions motivates the idea of using citing sentences to summarize research. The comparison that Elkiss et al. (2008) performed between abstracts and citing sentences suggests that a summary generated from citing sentences will be different and probably more concise and informative than the paper abstract or a summary generated from the full text of the paper. For example, Table 5 shows the abstract of Resnik (1999) and 5 selected sentences that cite it in AAN. We notice that citing sentences con-

tain additional facts that are not in the abstract, not only ones that summarize the paper contributions, but also those that criticize it (e.g., the last citing sentence in the Table).

Previous work has explored this research direction. Qazvinian and Radev (2008) proposed a method for summarizing scientific articles by building a similarity network of the sentences that cite it, and then applying network analysis techniques to find a set of sentences that covers as much of the paper facts as possible. Qazvinian et al. (2010) proposed another summarization method that first extracts a number of important key phrases from the set of citing sentences, and then finds the best subset of sentences that covers as many key phrases as possible.

These works focused on analyzing the citing sentences and selecting a representative subset that covers the different aspects of the summarized article. In recent work, Abu-Jbara and Radev (2011b) raised the issue of coherence and readability in summaries generated from citing sentences. They added a preprocessing and postprocessing steps to the summarization pipeline. In the preprocessing step, they use a supervised classification approach to rule out irrelevant sentences or fragments of sentences. In the postprocessing step, they improve the summary coherence and readability by reordering the sentences, removing extraneous text (e.g. redundant mentions of author names and publication year).

Mohammed et al. (2009) went beyond single paper summarization. They investigated the usefulness of directly summarizing citation texts in the automatic creation of technical surveys. They generated surveys from a set of Question Answering (QA) and Dependency Parsing (DP) papers, their abstracts, and their citation texts. The evaluation of the generated surveys shows that both citation texts and abstracts have unique survey-worthy information. It is worth noting that all the aforementioned research on citation-based summarization used the ACL Anthology Network (AAN) for evaluation.

4 Controversy Identification

Some arguments and claims made by researchers may get disputed by other researchers (Teufel, 1999). The following are examples of citing sentences that dispute previous work.

(1) Even though prior work (Teufel et al., 2006) argues that citation text is unsuitable for summarization, we show that in the framework of multi-document survey creation, citation texts can play a crucial role.

(2) Mining the Web for bilingual text (Resnik, 1999) is not likely to provide sufficient quantities of high quality data.

In many cases, it is useful to know which arguments were confirmed and accepted by the research community and which ones where disputed or even rejected. We believe that analyzing citation text helps identify these contrasting views automatically.

5 Comparison of Different Techniques

Citing sentences that compare different techniques or compare the techniques proposed by the author to previous work are common. The following sentences are examples of such comparisons.

(3) In (Zollmann et al., 2008), an interesting comparison between phrase-based, hierarchical and syntax-augmented models is carried out, concluding that hierarchical and syntax-based models slightly outperform phrase-based models under large data conditions and for sufficiently non-monotonic language pairs.

(4) Brill's results demonstrate that this approach can outperform the Hidden Markov Model approaches that are frequently used for part-of-speech tagging (Jelinek, 1985; Church, 1988; DeRose, 1988; Cutting et al., 1992; Weischedel et al., 1993), as well as showing promise for other applications.

(5) Our highest scores of 90.8% LP and 90.5% LR outperform the scores of the best previously published parser by Charniak (2000) who obtains 90.1% for both LP and LR.

Extracting such comparisons from citations can be of great benefit to researchers. It will allow them to quickly determine which technique works better for their tasks. To verify that citation text could be a good source for extracting comparisons, we created a list of words and phrases that are usually used to express comparisons and counted their frequency in AAN citing sentences. We found, for example, that the word *compare* (at its variations)

Abstract	STRAND (Resnik, 1998) is a language-independent system for automatic discovery of text in parallel translation on the World Wide Web. This paper extends the preliminary STRAND results by adding automatic language identification, scaling up by orders of magnitude, and formally evaluating performance. The most recent end-product is an automatically acquired parallel corpus comprising 2491 English-French document pairs, approximately 1.5 million words per language.
Selected Citing Sentences	Many research ideas have exploited the Web in unsupervised or weakly supervised algorithms for natural language processing (e.g. , Resnik (1999)) Resnik (1999) addressed the issue of language identification for finding Web pages in the languages of interest. In Resnik (1999), the Web is harvested in search of pages that are available in two languages, with the aim of building parallel corpora for any pair of target languages. The STRAND system of (Resnik, 1999), uses structural markup information from the pages, without looking at their content, to attempt to align them. Mining the Web for bilingual text (Resnik, 1999) is not likely to provide sufficient quantities of high quality data.

Table 5: Comparison of the abstract and a selected set of sentences that cite Resnik (1999) work

appears in about 4000 sentences, and that the words *outperform* and *contrast* each appears in about 1000 citing sentences.

6 Ontology Creation

It is useful for researchers to know which tasks and research problems are important, and what techniques and tools are usually used with them. Citation text is a good source of such information. For example, sentence (6) below shows three different techniques (underlined) that were used to extend tools and resources that were created for English so that they work for other languages. For another example, sentence (7) shows different tasks in which re-ranking has been successfully applied. These relations can be easily extracted from citing sentences and can be possibly used to build an ontology of tasks, methods, tools, and the relations between them.

(6) Another strain of research has sought to exploit resources and tools in some languages (especially English) to construct similar resources and tools for other languages, through heuristic projection (Yarowsky and Ngai, 2001; Xi and Hwa, 2005) or constraints in learning (Burkett and Klein, 2008; Smith and Eisner, 2009; Das and Petrov, 2011; McDonald et al., 2011) or inference (Smith and Smith, 2004).

(7) (Re)rankers have been successfully applied to numerous NLP tasks, such as parse selection (Osborne and Baldridge, 2004; Toutanova et al., 2004), parse reranking (Collins and Duffy, 2002; Charniak and Johnson, 2005), question-answering (Ravichandran et al., 2003).

7 Paraphrase Extraction

It is common that multiple citing sentences highlight the same facts about a cited paper. Since these sentences were written by different authors, they often use different wording to describe the cited paper facts. This motivates the idea of using citing sentences to create data sets for paraphrase extraction. For example, sentences (8) and (9) below both cite (Turney, 2002) and highlight the same aspect of Turney's work using slightly different wordings. Therefore, sentences (8) and (9) can be considered paraphrases of each other.

(8) In (Turney, 2002), an unsupervised learning algorithm was proposed to classify reviews as recommended or not recommended by averaging sentiment annotation of phrases in reviews that contain adjectives or adverbs.

(9) For example, Turney (2002) proposes a method to classify reviews as recommended/not recommended, based on the average semantic orientation of the review.

The paraphrase annotation of citing sentences consists of manually labeling which sentence consists of what facts. Then, if two citing sentences consist of the same set of facts, they are labeled as paraphrases of each other. For example, if a paper has 50 sentences citing it, this gives us a paraphrasing data set that consists of 50*49 = 2450 pairs. As a proof of concept, we annotated 25 papers from AAN using the annotation method described above. This data set consisted of 33,683 sentence pairs of which 8,704 are paraphrases.

The idea of using citing sentences to create data sets for paraphrase extraction was initially suggested

by Nakov et al. (2004) who proposed an algorithm that extracts paraphrases from citing sentences using rules based on automatic named entity annotation and the dependency paths between them.

8 Scientific Article Classification

Automatic classification of scientific articles is one of the important tasks for creating publication databases. A variety of machine learning algorithms have been proposed for this task. Many of these methods perform the classification based on the title, the abstract, or the full text of the article. Some other methods used citation links in addition to content to make classification decisions. Cao and Gao (2005) proposed a two-phase classification system. The system first applies a content-based statistical classification method which is similar to general text classification. In the second phase, the system uses an iterative method to update the labels of classified instances using citation links. A similar approach is also proposed by Zhang et al. (2006). These approaches use citation links only to improve classification decisions that were made based on content. We hypothesize that using the text of citing sentences in addition to citation structure and content leads to more accurate classification than using the content and citation links only.

9 Terminology Translation

Citing sentences can also be used to improve machine translation systems by using citing sentences from different languages to build parallel corpus of terms and their translations. This can be done by identifying articles written in different languages that cite a common target paper, then extracting the citing sentences from each paper. Word alignment techniques can then be applied to the text surrounding the reference to the common target paper. The aligned words from each source can then be extracted and used as translations of the same term. Sentences (10) and (11) below illustrate how the application of this proposed method can identify that the underlined terms in sentence 10 (Spanish) and sentence 11 (English) are translations of each other.

*(10) Spanish: Se comprobó que la agrupación por bloques ofrecía mejores resultados que, la introducción de vocabulario (Hearst, 1997) o **las cadenas léxicas** (Hearst, 1994) y, por tanto, es la que se ha utilizado en la segunda fase del algoritmo.*

*(11) English: This can be done either by analyzing the number of overlapping **lexical chains** (Hearst, 1994) or by building a short-range and long-range language model (Beeferman et al., 1999).*

10 Other Uses of Citing Sentences

Nakov et al. (2004) proposed several other uses of citing sentences. First, they suggested using them as a source for unannotated comparable corpora. Such comparable corpora can be used in several applications such as paraphrase extraction as we showed earlier. They also noticed that the scientific literature is rife with abbreviations and synonyms, and hence, citing sentences referring to the same article may allow synonyms to be identified and recorded. They also proposed using citing sentences to build a model of the different ways used to express a relationship between two entities. They hypothesized that this model can help improve both relation extraction and named entity recognition systems. Finally, they proposed improving the indexing and ranking of publications by considering, in addition to the content of the publication, the text of citing sentences that cite it and their contexts.

11 Summarizing 30 years of ACL Discoveries Using Citing Sentences

The ACL Anthology Corpus contains all the proceedings of the Annual Meeting of the Association of Computational Linguistics (ACL) since 1979. All the ACL papers and their citation links and citing sentences are included in the ACL Anthology Network (ACL). In this section, we show how citing sentences can be used to summarize the most important contributions that have been published in the ACL conference since 1979. We selected the most cited papers in each year and then manually picked a citing sentence that cites a top cited and describes it contribution. It should be noted here that the citation counts we used for ranking papers reflect the number of incoming citations the paper received *only* from the venues included in AAN. To create the summary, we used citing sentences that has the reference to the cited paper in the beginning of the sentence. This is

1979	Carbonell (1979) discusses inferring the meaning of new words.
1980	Weischedel and Black (1980) discuss techniques for interacting with the linguist/developer to identify insufficiencies in the grammar.
1981	Moore (1981) observed that determiners rarely have a direct correlation with the existential and universal quantifiers of first-order logic.
1982	Heidorn (1982) provides a good summary of early work in weight-based analysis, as well as a weight-oriented approach to attachment decisions based on syntactic considerations only.
1983	Grosz et al. (1983) proposed the centering model which is concerned with the interactions between the local coherence of discourse and the choices of referring expressions.
1984	Karttunen (1984) provides examples of feature structures in which a negation operator might be useful.
1985	Shieber (1985) proposes a more efficient approach to gaps in the PATR-II formalism, extending Earley's algorithm by using restriction to do top-down filtering.
1986	Kameyama (1986) proposed a fourth transition type, Center Establishment (EST), for utterances E.g., in Bruno was the bully of the neighborhood.
1987	Brennan et al. (1987) propose a default ordering on transitions which correlates with discourse coherence.
1988	Whittaker and Stenton (1988) proposed rules for tracking initiative based on utterance types; for example, statements, proposals, and questions show initiative, while answers and acknowledgements do not.
1989	Church and Hanks (1989) explored tile use of mutual information statistics in ranking co-occurrences within five-word windows.
1990	Hindle (1990) classified nouns on the basis of co-occurring patterns of subjectverb and verb-object pairs.
1991	Gale and Church (1991) extract pairs of anchor words, such as numbers, proper nouns (organization, person, title), dates, and monetary information.
1992	Pereira and Schabes (1992) establish that evaluation according to the bracketing accuracy and evaluation according to perplexity or crossentropy are very different.
1993	Pereira et al. (1993) proposed a soft clustering scheme, in which membership of a word in a class is probabilistic.
1994	Hearst (1994) presented two implemented segmentation algorithms based on term repetition, and compared the boundaries produced to the boundaries marked by at least 3 of 7 subjects, using information retrieval metrics.
1995	Yarowsky (1995) describes a 'semi-unsupervised' approach to the problem of sense disambiguation of words, also using a set of initial seeds, in this case a few high quality sense annotations.
1996	Collins (1996) proposed a statistical parser which is based on probabilities of dependencies between head-words in the parse tree.
1997	Collins (1997)'s parser and its re-implementation and extension by Bikel (2002) have by now been applied to a variety of languages: English (Collins, 1999), Czech (Collins et al. , 1999), German (Dubey and Keller, 2003), Spanish (Cowan and Collins, 2005), French (Arun and Keller, 2005), Chinese (Bikel, 2002) and, according to Dan Bikels web page, Arabic.
1998	Lin (1998) proposed a word similarity measure based on the distributional pattern of words which allows to construct a thesaurus using a parsed corpus.
1999	Rapp (1999) proposed that in any language there is a correlation between the cooccurrences of words which are translations of each other.
2000	Och and Ney (2000) introduce a NULL-alignment capability to HMM alignment models.
2001	Yamada and Knight (2001) used a statistical parser trained using a Treebank in the source language to produce parse trees and proposed a tree to string model for alignment.
2002	BLEU (Papineni et al., 2002) was devised to provide automatic evaluation of MT output.
2003	Och (2003) developed a training procedure that incorporates various MT evaluation criteria in the training procedure of log-linear MT models.
2004	Pang and Lee (2004) applied two different classifiers to perform sentiment annotation in two sequential steps: the first classifier separated subjective (sentiment-laden) texts from objective (neutral) ones and then they used the second classifier to classify the subjective texts into positive and negative.
2005	Chiang (2005) introduces Hiero, a hierarchical phrase-based model for statistical machine translation.
2006	Liu et al. (2006) experimented with tree-to-string translation models that utilize source side parse trees.
2007	Goldwater and Griffiths (2007) employ a Bayesian approach to POS tagging and use sparse Dirichlet priors to minimize model size.
2008	Huang (2008) improves the re-ranking work of Charniak and Johnson (2005) by re-ranking on packed forest, which could potentially incorporate exponential number of k-best list.
2009	Mintz et al. (2009) uses Freebase to provide distant supervision for relation extraction.
2010	Chiang (2010) proposes a method for learning to translate with both source and target syntax in the framework of a hierarchical phrase-based system.

Table 6: A citation-based summary of the important contributions published in ACL conference proceedings since 1979. The top cited paper in each year is found and one citation sentence is manually picked to represent it in the summary.

because such citing sentences are often high-quality, concise summaries of the cited work. Table 6 shows the summary of the ACL conference contributions that we created using citing sentences.

12 Conclusion

We motivated and discussed several different uses of citing sentences, the text surrounding citations. We showed that citing sentences can be used to analyze the dynamics of research and observe how it trends. We also gave examples on how analyzing the text of citing sentences can give a better understanding of the impact of a researcher's work and how this impact changes over time. In addition, we presented several different applications that can benefit from citing sentences such as scientific literature summarization, identifying controversial arguments, and identifying relations between techniques, tools and tasks. We also showed how citing sentences can provide high-quality for NLP tasks such as information extraction, paraphrase extraction, and machine translation. Finally, we used AAN citing sentences to create a citation-based summary of the important contributions included in the ACL conference publication in the past 30 years.

References

Amjad Abu-Jbara and Dragomir Radev. 2011a. Coherent citation-based summarization of scientific papers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 500–509, Portland, Oregon, USA, June. Association for Computational Linguistics.

Amjad Abu-Jbara and Dragomir Radev. 2011b. Coherent citation-based summarization of scientific papers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 500–509, Portland, Oregon, USA, June. Association for Computational Linguistics.

F. Amblard, A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro. 2011. On the temporal analysis of scientific network evolution. In *Computational Aspects of Social Networks (CASoN), 2011 International Conference on*, pages 169 –174, oct.

Awais Athar. 2011. Sentiment analysis of citations using sentence structure-based features. In *Proceedings of the ACL 2011 Student Session*, pages 81–87, Portland, OR, USA, June. Association for Computational Linguistics.

Christine L. Borgman and Jonathan Furner. 2002. Scholarly communication and bibliometrics. *ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY*, 36(1):2–72.

Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 155–162, Stanford, California, USA, July. Association for Computational Linguistics.

Minh Duc Cao and Xiaoying Gao. 2005. Combining contents and citations for scientific document classification. In *Proceedings of the 18th Australian Joint conference on Advances in Artificial Intelligence*, AI'05, pages 143–152, Berlin, Heidelberg. Springer-Verlag.

Jaime G. Carbonell. 1979. Towards a self-extending parser. In *Proceedings of the 17th Annual Meeting of the Association for Computational Linguistics*, pages 3–7, La Jolla, California, USA, June. Association for Computational Linguistics.

David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 263–270, Ann Arbor, Michigan, June. Association for Computational Linguistics.

David Chiang. 2010. Learning to translate with source and target syntax. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1443–1452, Uppsala, Sweden, July. Association for Computational Linguistics.

Kenneth Ward Church and Patrick Hanks. 1989. Word association norms, mutual information, and lexicography. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pages 76–83, Vancouver, British Columbia, Canada, June. Association for Computational Linguistics.

Kenneth Ward Church. 1988. A stochastic parts program and noun phrase parser for unrestricted text. In *Proceedings of the Second Conference on Applied Natural Language Processing*, pages 136–143, Austin, Texas, USA, February. Association for Computational Linguistics.

Michael John Collins. 1996. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pages 184–191, Santa Cruz, California, USA, June. Association for Computational Linguistics.

Michael Collins. 1997. Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 16–23, Madrid, Spain, July. Association for Computational Linguistics.

Aaron Elkiss, Siwei Shen, Anthony Fader, Güneş Erkan, David States, and Dragomir Radev. 2008. Blind men and elephants: What do citation summaries tell us about a research article? *J. Am. Soc. Inf. Sci. Technol.*, 59(1):51–62.

William A. Gale and Kenneth W. Church. 1991. A program for aligning sentences in bilingual corpora. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 177–184, Berkeley, California, USA, June. Association for Computational Linguistics.

Sharon Goldwater and Tom Griffiths. 2007. A fully bayesian approach to unsupervised part-of-speech tagging. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 744–751, Prague, Czech Republic, June. Association for Computational Linguistics.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 44–50, Cambridge, Massachusetts, USA, June. Association for Computational Linguistics.

Sonal Gupta and Christopher Manning. 2011. Analyzing the dynamics of research by extracting key aspects of scientific papers. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 1–9, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.

Marti A. Hearst. 1994. Multi-paragraph segmentation expository text. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 9–16, Las Cruces, New Mexico, USA, June. Association for Computational Linguistics.

George E. Heidorn. 1982. Experience with an easily computed metric for ranking alternative parses. In *Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics*, pages 82–84, Toronto, Ontario, Canada, June. Association for Computational Linguistics.

Donald Hindle. 1990. Noun classification from predicate-argument structures. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 268–275, Pittsburgh, Pennsylvania, USA, June. Association for Computational Linguistics.

Liang Huang. 2008. Forest reranking: Discriminative parsing with non-local features. In *Proceedings of ACL-08: HLT*, pages 586–594, Columbus, Ohio, June. Association for Computational Linguistics.

Megumi Kameyama. 1986. A property-sharing constraint in centering. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 200–206, New York, New York, USA, July. Association for Computational Linguistics.

Lauri Karttunen. 1984. Features and values. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics*, pages 28–33, Stanford, California, USA, July. Association for Computational Linguistics.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 2*, pages 768–774, Montreal, Quebec, Canada, August. Association for Computational Linguistics.

Yang (1) Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 609–616, Sydney, Australia, July. Association for Computational Linguistics.

Terttu Luukkonen. 1992. Is scientists' publishing behaviour rewardseeking? *Scientometrics*, 24:297–319. 10.1007/BF02017913.

Amin Mazloumian, Young-Ho Eom, Dirk Helbing, Sergi Lozano, and Santo Fortunato. 2011. How citation boosts promote scientific paradigm shifts and nobel prizes. *PLoS ONE*, 6(5):e18975, 05.

Qiaozhu Mei and ChengXiang Zhai. 2008. Generating impact-based summaries for scientific literature. In *Proceedings of ACL-08: HLT*, pages 816–824, Columbus, Ohio, June. Association for Computational Linguistics.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011, Suntec, Singapore, August. Association for Computational Linguistics.

Saif Mohammad, Bonnie Dorr, Melissa Egan, Ahmed Hassan, Pradeep Muthukrishan, Vahed Qazvinian, Dragomir Radev, and David Zajic. 2009. Using citations to generate surveys of scientific paradigms. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 584–592, Boulder, Colorado, June. Association for Computational Linguistics.

Robert C. Moore. 1981. Problems in logical form. In *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, pages 117–124, Stanford, California, USA, June. Association for Computational Linguistics.

Preslav I. Nakov, Ariel S. Schwartz, and Marti A. Hearst. 2004. Citances: Citation sentences for semantic analysis of bioscience text. In *In Proceedings of the SIGIR04 workshop on Search and Discovery in Bioinformatics*.

Hidetsugu Nanba and Manabu Okumura. 1999. Towards multi-paper summarization using reference information. In *IJCAI '99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 926–931, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Hidetsugu Nanba, Noriko Kando, Manabu Okumura, and Of Information Science. 2000. Classification of research papers using citation links and citation types: Towards automatic review article generation.

Franz Josef Och and Hermann Ney. 2000. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 440–447, Hong Kong, October. Association for Computational Linguistics.

Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 160–167, Sapporo, Japan, July. Association for Computational Linguistics.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 271–278, Barcelona, Spain, July.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July. Association for Computational Linguistics.

Fernando Pereira and Yves Schabes. 1992. Inside-outside reestimation from partially bracketed corpora. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 128–135, Newark, Delaware, USA, June. Association for Computational Linguistics.

Fernando Pereira, Naftali Tishby, and Lillian Lee. 1993. Distributional clustering of english words. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pages 183–190, Columbus, Ohio, USA, June. Association for Computational Linguistics.

Vahed Qazvinian and Dragomir R. Radev. 2008. Scientific paper summarization using citation summary networks. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 689–696, Manchester, UK, August. Coling 2008 Organizing Committee.

Vahed Qazvinian and Dragomir R. Radev. 2010. Identifying non-explicit citing sentences for citation-based summarization. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 555–564, Uppsala, Sweden, July. Association for Computational Linguistics.

Vahed Qazvinian, Dragomir R. Radev, and Arzucan Ozgur. 2010. Citation summarization through keyphrase extraction. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 895–903, Beijing, China, August. Coling 2010 Organizing Committee.

Dragomir R. Radev, Pradeep Muthukrishnan, and Vahed Qazvinian. 2009. The acl anthology network corpus. In *NLPIR4DL '09: Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*, pages 54–61, Morristown, NJ, USA. Association for Computational Linguistics.

Reinhard Rapp. 1999. Automatic identification of word translations from unrelated english and german corpora. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 519–526, College Park, Maryland, USA, June. Association for Computational Linguistics.

Sidney Redner. 2005. Citation statistics from 110 years of physical review. *Physics Today*, 58(6):49–54.

Philip Resnik. 1999. Mining the web for bilingual text. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 527–534, College Park, Maryland, USA, June. Association for Computational Linguistics.

Stuart M. Shieber. 1985. Using restriction to extend parsing algorithms for complex-feature-based formalisms. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pages 145–152, Chicago, Illinois, USA, July. Association for Computational Linguistics.

Advaith Siddharthan and Simone Teufel. 2007. Whose idea was this, and why does it matter? attributing scientific work to citations. In *In Proceedings of NAACL/HLT-07*.

Simone Teufel, Advaith Siddharthan, and Dan Tidhar. 2006. Automatic classification of citation function. In *In Proc. of EMNLP-06*.

Simone Teufel. 1999. Argumentative zoning: Information extraction from scientific text. Technical report.

Simone Teufel. 2007. Argumentative zoning for improved citation indexing. computing attitude and affect in text. In *Theory and Applications, pages 159170*.

Ralph M. Weischedel and John E. Black. 1980. If the parser fails. In *Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics*, pages 95–95, Philadelphia, Pennsylvania, USA, June. Association for Computational Linguistics.

Steve Whittaker and Phil Stenton. 1988. Cues and control in expert-client dialogues. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 123–130, Buffalo, New York, USA, June. Association for Computational Linguistics.

Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proceedings of 39th Annual Meeting of the Association for Computational Linguistics*, pages 523–530, Toulouse, France, July. Association for Computational Linguistics.

David Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 189–196, Cambridge, Massachusetts, USA, June. Association for Computational Linguistics.

M. Zhang, X. Gao, M.D. Cao, and Yuejin Ma. 2006. Neural networks for scientific paper classification. In *Innovative Computing, Information and Control, 2006. ICICIC '06. First International Conference on*, volume 2, pages 51 –54, 30 2006-sept. 1.