ARL-TR-9050 ● Sep 2020

A User- and Text-Oriented Approach to Annotation

by Erin Zaroukian, Sue E Kase, and Michelle Vanni

NOTICES

Disclaimers

ARL-TR-9050 ● SEP 2020

A User- and Text-Oriented Approach to Annotation

Erin Zaroukian, Sue E Kase, and Michelle Vanni

Computational and Information Sciences Directorate, CCDC Army Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
September 2020	Technical Report	1 January 2018–2 December 2019

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
A User- and Text-Oriented Approach to Annotation	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Erin Zaroukian, Sue E Kase, and Michelle Vanni	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
CCDC Army Research Laboratory ATTN: FCDD-RLC-IT Aberdeen Proving Ground, MD 21005	ARL-TR-9050
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ORCID ID(s): Erin Zaroukian, 0000-0002-1381-085X; Sue Kase, 0000-0003-2732-629X; Michelle Vanni, 0000-0002-1023-5407

14. ABSTRACT
In the decades-long enterprise of information extraction (IE) system development, it has generally been presumed that systems for automatically identifying various categories of information in text are basically helpful to analysts as consumers of content leading to decision making (DM). Recent studies have shed light on the complexity of that account. Our research suggests that analysts probably benefit from different kinds of decision aids, depending on the genres of texts presented to them (e.g., tweets, narratives) and the types of questions they must, based on text content, answer (e.g., inferential, multiple choice). Here, we summarize education community research exploring this very space and present novel experiments involving different types of text (e.g., scientific expository texts, collections of military intelligence reports), asking different kinds of questions about the text (e.g., simple information queries, queries requiring multiple steps of deduction or inference), and providing different styles of annotation (e.g., user/computer generated, rich/sparse), as DM aids, to point up the relevance of these dimensions on the usefulness of annotation for improved text comprehension and problem solving by analysts and learners.

15. SUBJECT TERMS
annotation, information extraction, decision support, reading comprehension, technology use

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Erin Zaroukian
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	22	**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified			(410) 278-3203

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

Acknowledgments

Many thanks are owed to Jon Bakdash, Stephen Tratz, Claire Bonial, Clare Voss, Jeffrey Micher, Lucia Donatelli, and Justine Caylor.

1. Introduction

Annotation has the potential to greatly improve human ability to comprehend text by serving as a cognitive artifact, facilitating comprehension, or even enhancing automated reasoning systems when applied to raw text input. Reading and understanding involve many complex and interrelated processes, however, which make studying the efficacy of annotation to the human user difficult but crucially important.

In this report, we overview a relevant area of education research, focusing on the roles played by the *text type* and the *type of question being asked about the text* in developing annotation strategies that are effective for text comprehension. We begin by discussing automation schemes and variety therein. We then review existing work on variation in text types and how amenable they are to annotation for improved comprehension. Finally, we present original research across a range of annotation schemes for intelligence analysis.

2. Text Annotation and Comprehension

2.1 Annotation Schemes

Annotation schemes typically describe a hierarchical breakdown of a semantic space enabling the decomposition of a sentence into meaningful categories of information. Annotation of text is intended to aid comprehension, either directly (e.g., highlighting important passages helps a reader recognize and understand key ideas) or downstream for purposes of information extraction (IE). For example, named entity recognition (NER) may be employed to populate a database to help synthesize multiple documents. Annotation schemes can vary widely in the type and density of information they extract (e.g., one key idea, all named entities). We illustrate this variation by applying annotation styles, based on three distinct annotation schemes, varying in the complexity of their semantic integration, to a task-specific context. These annotation schemes are based on standard markup guideline definitions: Message Understanding Conference (MUC; Grishman and Sundheim 1996), Low Density Languages (LODL; Simpson et al. 2008), and Automatic Content Extraction (ACE; Doddington et al. 2004). Examples of the styles, developed on the basis of these schemes, are provided in later sections of the report.

The style developed on the basis of MUC annotation is relatively sparse, recognizing only two categories of information: numbers and named entities. Although rarely encountered in current technology, it can be used for filling templates found in legacy systems or enhancing static descriptions, inventories, and generically single-out numeric features without subcategory tags. The LODL annotation scheme, loosely based on MUC-7, was originally designed to be used on text in low-resource languages as an aid in system training for quick ramp-up of automatic content processors. The scheme expanded on the earlier work to pull out named entities in several categories, such as PERson, ORGanization, and LOCation, and distinguish NUMeric EXpression subcategories of Date, Time, money, and percentage amounts. The ACE annotation scheme is the most sophisticated of the three, recognizing events and co-references and categorizing a variety of named entities and numeric expressions, very useful for training systems for populating knowledge bases from new material.

It is noteworthy that each of the schemes from which the experiments' styles were derived originated as data designs serving the development of computational processes. It was humans who developed the guidelines, applied the annotation, and evaluated the systems trained on the annotated data. Because the tagged categories are thus meaningful to humans, the moment has arrived for researchers to explore strategies for enhancing human use and exploitation of the output of these and similar systems.

2.2 Annotation for Comprehension

Research on annotation for text comprehension acknowledges that, despite the goal of improving human comprehension of text, annotation runs the risk of increasing cognitive load to the detriment of comprehension. Factors investigated in reading comprehension research (i.e., text type, question type, context) could provide some insight on the limitations of annotation in human interpretation and understanding of text-based information.

For example, Eason et al. (2012) investigated potential differences in ability to comprehend specific types of text (i.e., narrative, expository, functional) and answering specific types of questions (i.e., initial understanding, interpretation, critical analysis). Significant differences in task performance were found among subjects assigned the three types of text. Functional texts were found to be easier to understand than both narrative and expository text. Functional text is designed to replicate "real world" situations, such as reading signs or following directions. Since the structure of functional text is commonly procedural in nature (i.e., presented in a list format), it may be easier to comprehend when the goal is to answer questions following the text. Significant differences on performance by experimental subjects assigned the three question types were observed as well. Not surprisingly, critical analysis questions proved to be the most difficult for readers overall. This question type requires readers not only to comprehend the text but also to analyze and evaluate the text while using reading strategies to recognize text structures. Higher-level cognitive skills (i.e., inferencing, planning, organizing) are required to answer these types of questions. Eason et al. (2012) is only one example in a significant body of reading comprehension research accumulated over several decades.

A limited number of reading comprehension studies actually incorporate annotation as a factor in the understanding and processing of textual information. Jan et al. (2016) conducted a study on collaborative annotation that showed improved reading comprehension when annotation was filtered such that participants viewed less, higher-quality annotation. In this study, high school chemistry students participated in a digital reading and annotating activity where they were ultimately asked to answer comprehension questions on annotated text with and without annotation filtering. The text was an expository chemistry article, and comprehension was measured through multiple-choice reading comprehension questions designed by an instructor to address recall, main idea, inference, and application. Participants annotated text through highlighting, and they labeled highlighted text as one of five categories: importance, quizzing, query, example, and summary. Participants showed better comprehension when using a high-quality annotation filter, leading the authors to recommend frequency-based filtering of user-submitted annotation to increase annotation quality and reduce annotation quantity and cognitive load. This demonstrates that more annotation is not necessarily better and that quality matters.

Additionally, research has delved into the importance of the type of task that annotation is intended to support. Wallen et al. (2005) conducted a study using expert annotation of digital texts, finding that different types of annotation facilitated different types of learning and that using multiple types of annotation at once increased cognitive load and lowered performance. They used three types of annotation that were hypothesized to support different cognitive processes: selection (select relevant information to enable high-level processes), organization (organize information into meaningful representations by building connections to support comprehension), and integration (integrate representations between visual and verbal information and prior knowledge to support processes like the creation of new ideas). Three tests were used to measure these different levels of learning: a recognition test (recall of facts), a comprehension test (conceptual understanding), and a transfer test (higher-level understanding). Participants in their study read digital 650-word expository scientific texts in one of seven conditions: plain, with one of the three annotation types, or with two annotation types. They found that the different annotation types facilitated different types of learning, and participants had lower cognitive load and performed better when using only one type of annotation compared to two types presented together.

Ben-Yehudah and Eshet-Alkalai (2014) compared print and digital texts and found that annotation's effect on comprehension varied across question types. Participants read a four-page passage about fossils from a textbook, and half of participants were asked to highlight text and take notes directly on the page in a way that they thought would help them retain information for a later test. After reading and annotating the passage, the text annotation was removed and participants were asked factual and inference questions. Participants performed similarly on factual questions with or without annotating, using digital or print materials. For inference questions, however, annotation improved performance, but only for print text.

Reading comprehension models and experimental results may provide theoretical grounding for additional studies on the cognitive processes associated with reading annotated texts. As studies mentioned show, the usefulness of annotation cannot be taken for granted because it varies depending on annotation quality, quantity, and complexity, as well as the medium. Moreover, different types of annotation facilitate different types of learning. While these studies vary with respect to annotation scheme and the types of questions asked, they are similar in that they each focus on the humans' cognizing of annotated expository texts. In the next section, we present experiments using several types of annotation schemes applied to field report style texts containing information useful for intelligence purposes.

3. Experiments

Three experiments investigating the impact of text annotation schemes on problem solving in a task-specific context were performed using crowdsourcing technologies enabling the recruitment of large numbers of participants from Amazon Mechanical Turk (AMT).

In each of the experiments, participants played the role of intelligence analysts required to solve a problem. Participants read text formatted as field reports one to three sentences in length. The reports had characteristics similar to those of expository texts. Expository texts provide information about a particular topic possibly including scientific or technical vocabulary, and exhibit distinct structural qualities (Medina and Pilonieta 2006; Eason et al. 2012; McNamara et al. 2012). Depending on the experiment and condition, an annotation scheme was applied to the text in the intelligence reports. The participant's level of reading comprehension was an important factor in solving the intelligence analysis problem.

Section 3.1 overviews two experiments comparing comprehension of intelligence reports with and without an ACE-style annotation. Accuracy and task-relevance of the annotation were manipulated across the two experiments. Section 3.2 looks at a third experiment, also using ACE annotation, along with two additional annotation schemes (MUC and LODL), which have a lower level of semantic integration. In that experiment, text was marked up with the annotation schemes and question complexity was manipulated. Section 3.3 discusses trends in question answering accuracy across the three experiments.

3.1 Experimental Laboratory for the Investigation of Collaboration, Information Sharing, and Trust (ELICIT) Experiments

In a pair of experiments, participants read lists of sentences that together describe a hypothetical adversarial attack. These lists of sentences were presented either as plain text or with an ACE-style annotation. The participants, acting in the role of intelligence analysts, were tasked with identifying the attacker, target, time, and location of the adversarial attack. Performance with and without annotation was compared to determine whether the annotation assisted in identifying the adversarial attack.

3.1.1 Text and Questions

The text used in both experiments was drawn from ELICIT (Ruddy 2007). ELICIT provides capabilities to simulate an intelligence analysis task containing a number of hypothetical adversary attack scenarios.

Each scenario is a list of 68 simple sentences that together allow a reader to deduce the attacker, target, attack time, and attack location (*Who*, *What*, *Where*, and *When*) of an anticipated adversary attack. While some answers may be straightforwardly identified in the text (e.g., “The attack will be at 3 pm”), other answers require multiple logical steps through two or more sentences to deduce the correct answer. Half of the sentences are noise (i.e., they do not help the participant deduce the correct responses), but the entire set of sentences is logically consistent. The sentences are not ordered, and, while they are loosely interconnected, do not build a narrative. Instead the sentences resemble facts, simulating streaming data collected via multiple sources of intelligence.

To solve an ELICIT scenario, participants must answer the *Who*/*What*/*Where*/*When* questions using seven dropdown menus (*When* is broken down into separate menus for month, date, time of day, and am/pm) before a 20-min timer forces them to move on. Figure 1 shows example sentences from an ELICIT scenario.

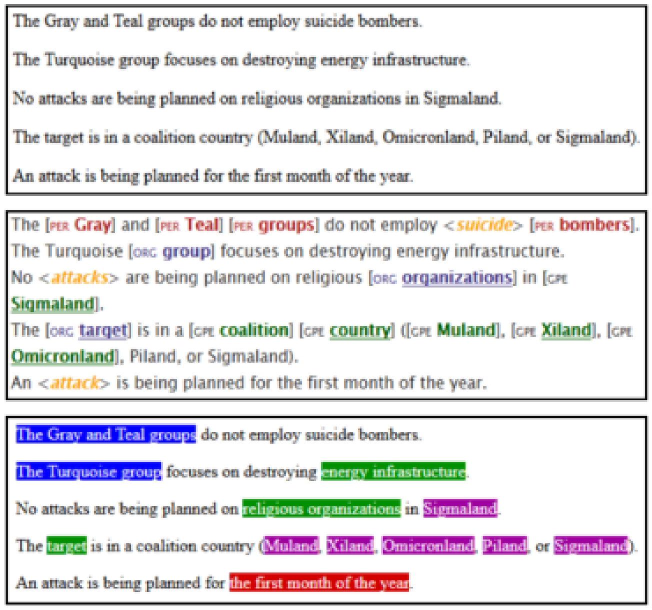

Fig. 1 Example excerpts of an ELICIT scenario: plain text (top) with ACE annotation (middle) and hand-generated annotation (bottom)

3.1.2 Experiment 1: ACE annotation

In the first experiment, the ELICIT scenarios were presented either plain text, see the top of Fig. 1, or with ACE markup from an IE pipeline (Li et al. 2013, Li and Ji 2014), see the middle of Fig. 1. This markup was generated using named entity recognition and event detection to indicate entities (e.g., person, vehicle, geo-political entity) and events (e.g., attack), shown via bracketing and subscripts, with mouse-over revealing additional information (e.g., an event's arguments, the class an entity belongs to).

The participants' accuracy and response times were compared across plain text and ACE annotation conditions. Results showed participants' performance was significantly worse (lower accuracy, slower response time) in the ACE condition compared to the plain text condition. Participants also supplied workload and preference ratings comparing these two conditions, and again the plain text condition was shown to be superior.

3.1.3 Experiment 2: Hand-generated annotation

Because the ACE annotation used in this first experiment was generated by an IE pipeline not specifically designed for the ELICIT intelligence task, its accuracy and task-relevance were not perfect. Therefore, to create a best-case scenario for annotation, the experiment was repeated using hand-generated markup (see the bottom of Fig. 1). The automatically generated ACE annotation in Experiment 1 indiscriminately draws participants' attention to all detected entities and events, whereas the hand-generated markup highlighted all and only potential answers to the *Who*/*What*/*Where*/*When* questions. In this way, the hand-generated markup is highly relevant to the task, and by carefully creating it by hand and verifying its correctness with other researchers, the markup is likely to be highly accurate. Furthermore, the hand-generated markup provided less information overall, marking up fewer words and doing so with simple highlighting instead of including information about an entity's classification. This has the advantage of reducing the risk of information overload (Marusich 2016).

Results from the experiment indicated participants still failed to show a performance advantage in the hand-generated annotation condition, although no significant performance advantage was shown for the plain text condition as was the case in Experiment 1. Further, ratings of workload and preference showed that participants preferred to work with the hand-generated annotation and felt that they performed better with it. For further details on these studies and their results, see Zaroukian (2018, 2019) and Zaroukian et al. (2019a, 2019b).

3.2 Human-Assisted Machine Information Exploitation (HAMIE) Experiment

Similar to the ELICIT experiments, participants in the HAMIE experiment acted in the role of intelligence analysts tasked with solving a problem. The task required participants to read and understand a series of sentences formatted as brief intelligence reports. Participants were randomly assigned to answer a question concerning a factious network of individuals and events in the Kandahar region. Text in the intelligence reports was marked up using one of three annotation styles schemes (MUC, LODL, ACE). Two subtasks (relation identification and link diagram formation) were designed to assist the participant in answering the question. Participants' performance was compared across the three annotation styles and three types of questions to determine the impact of annotation on problem solving in an intelligence analysis context.

3.2.1 Text and Questions

The text used in HAMIE appeared as five types of short simulated reports one to three sentences in length representing: open-source information, blog posts, biographies, records of criminal history, and intelligence information reports. Similar to the ELICIT experiment, the text was expository in nature and contained noise, which was not relevant in answering the questions. The HAMIE experiment contained three annotation conditions and three question conditions.

HAMIE's three annotation conditions (MUC, LODL, ACE) varied in the types and, hence, number of entities marked up. In addition to the ACE-based markup used in the ELICIT experiments, the two less complex markups were also compared. The MUC annotation highlighted named and numeric entities. The LODL annotation highlighted named, information, and event entities. Figure 2 shows an example intelligence report marked up with MUC (top), LODL (middle), and ACE (bottom).

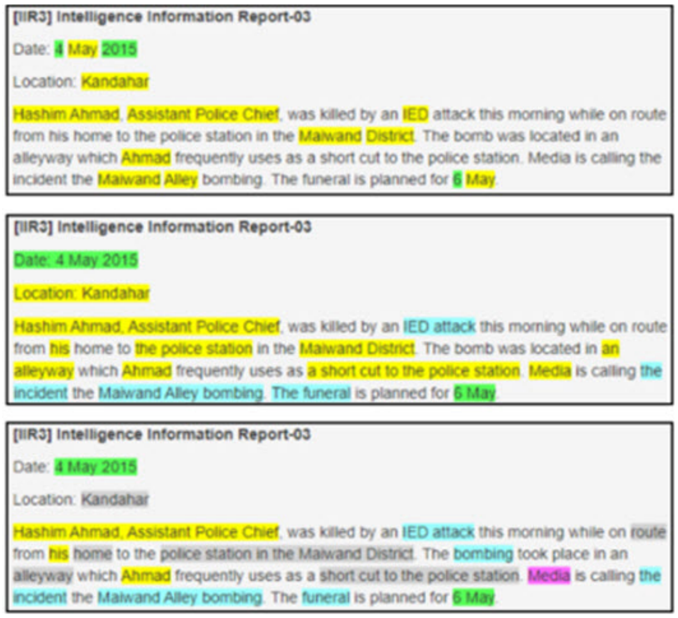

Fig. 2 Example intelligence report from HAMIE with MUC annotation (top), LODL annotation (middle), and ACE annotation (bottom)

The three question conditions in HAMIE varied by level of problem-solving difficulty. The first question required participants to determine whether critical intelligence information was true or false. For this question, the relevant facts where clearly stated in the text. The second question queried participants about potential corruptness of government leaders. This was a multiple-choice question with one correct answer, which required inference as all the facts were not clearly stated in the text. The third question asked participants to identify the individuals responsible for carrying out a bombing that killed the assistant police chief. This was also a multiple-choice question, but required the selection of three correct answers out of the seven provided. This question involved judgmental input as the text in the reports contained conflicting facts pertaining to the question solution.

3.2.2 Question Answering Performance

A number of performance recording mechanisms were built into the HAMIE task as a whole (i.e., question answering accuracy and total response time) and per individual subtask (i.e., relation identification and link diagram formation accuracy and response times). A point scheme relative to question difficulty was developed (i.e., first question 300 points, second question 1,100 points, third question 2,400 points).

Results across tasks indicated that annotation scheme significantly affected question answering performance. Participants assigned the MUC annotation scheme outperformed participants using the LODL annotation scheme and ACE annotation scheme. Performance between the ACE and LODL annotation schemes was not significantly different. This indicates that the MUC annotation scheme may constitute an ideal level of annotation for this type of task.

Additionally, question complexity affected performance. The lowest degree of performance occurred in the most difficult question scenario across the three annotation schemes. It appears that assistance provided by annotation in understanding the text is only beneficial when answering easier questions. For additional details on the HAMIE experimental design and analysis results, see Kase et al. (2017a, 2017b) and Neigel et al. (2018).

3.3 Discussion

The extent to which annotation—whether it be automatic or human-generated—helps humans to extract information from text is not well understood, especially when applied in task-relevant contexts. The three experiments described in this report represented a simplified intelligence analysis task requiring participants to answer one or more questions across a range of difficulty levels. The expository text type, which remained consistent across experiments, had a distinct structure, but without following a timeline and without the high cohesion of the narrative text type (Best et al. 2005). Several annotation schemes were compared to each other, with the most complex also compared to plain text.

The two ELICIT experiments demonstrated a failure of annotation (both ACE and hand-generated) to produce a performance advantage over plain text. While the hand-generated annotation was more accurate and supposedly task-relevant, it is possible that annotation of this sort, or annotation in general, is not particularly useful for answering deductive questions from a list of sentences. Perhaps a different type of text and a different type of problem solving would show different results.

The HAMIE experiment showed that the least complex annotation scheme (MUC), highlighting only named entities and numeric values—when compared with the more complex annotation schemes (LODL and ACE)—was associated with better performance by subjects, not only in number of points scored across subtasks but also in answering the question. Surprisingly, LODL and ACE offering a greater number of highlighted entity categories intended to aid comprehension and question answering proved to be rather ineffective. Less surprisingly, the least difficult question, the answer to which was found in facts clearly stated in the text, benefited most from applying the simple MUC annotation scheme.

The three experiments show we need to learn more about how annotated text assists readers as decision makers, with both comprehending textual content and potentially answering critical questions in a timely manner. The related research discussed in Section 2.2 offers key concepts, which can be incorporated, as factors, into experimental new directions in annotation research. These include understanding the cognitive skills and processes relevant to reading comprehension, cognition, and information overload, which also—in tandem with text type and question type—require consideration. Development of fine-grained performance metrics, which may someday inform algorithms predicting correct question answering by humans in a specified timeframe, would bridge another research gap. Lastly, investigations focused on customizing the annotation scheme for the task at hand might provide important direction for understanding a decision maker's interaction with annotated text.

4. Conclusion

With the limited research to date on the use of text annotation for comprehension, it appears that less is more. More complex texts and more complex questions will still challenge human analysts without sophisticated decision aids, the development and testing of which will have incorporated these crucial factors. However, a principled means for pairing task to tool for reduced cognitive load are, as yet, lacking. Our hope is that insights achieved in the bringing together of research in natural language processing and reading comprehension will lead to novel approaches designed to address this issue.

5. References

Ben-Yehudah G, Eshet-Alkalai Y. The influence of text annotation tools on print and digital reading comprehension. Proceedings of the 9th Chais Conference for the Study of Innovation and Learning Technologies: Learning in the Technological Era; 2014. p. 28–35.

Best RM, Rowe M, Ozuru Y, McNamara DS. Deep-level comprehension of science texts: the role of the reader and the text. Top Lang Disord. 2005;25:6583.

Doddington GR, Mitchell A, Przybocki M, Ramshaw L, Strassel S, Weischedel R. The Automated Content Extraction (ACE) program – tasks, data, and evaluation. LREC. 2004(2).

Eason SH, Goldberg LR, Young KM, Geist MC, Cutting LE. Reader-text interactions: how differential text and question types influence cognitive skills needed for reading comprehension. J Edu Psychol. 2012;104:515–528.

Grishman R, Sundheim B. Message understanding conference-6: a brief history. COLING. 1996(96).

Jan J-C, Chen C-M, Huang P-H. Enhancement of digital reading performance by using a novel web-based collaborative reading annotation system with two quality annotation filtering mechanisms. Int J Hum Comp Stud. 2016;86:81–93.

Kase SE, Vanni M, Caylor J, Hoye J. Human-assisted machine information exploitation: a crowdsource investigation of information-based problem solving. Proceedings of SPIE Defense and Security: Next Generation Analyst V; 2017a; Bellingham, WA. SPIE. p. 1–17.

Kase SE, Vanni M, Caylor J, Hoye J. A crowdsourced investigation of decision guidance approaches to information processing. Proceedings of the 22nd International Command and Control Research and Technology Symposium; 2017b.

Li Q, Ji H. Incremental joint extraction of entity mentions and relations. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics; 2014; New York, NY. ACL. p. 402–412.

Li Q, Ji H, Huang L. Joint event extraction via structured prediction with global features. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics; 2013; New York, NY. ACL. p. 73–82.

Marusich LR, Bakdash JZ, Onal E, Yu MS, O'Donovan J, Höllerer T, Buchler N, Gonzalez C. Effect of information availability on command-and-control decision making: performance, trust, and situation awareness. Hum Fact. 2018;58(2):301–321.

McNamara DS, Graesser AC, Louwerse MM. Sources of text difficulty: Across the ages and genres. In: Sabatini JP, Albo E, editors. Assessing reading in the 21st century: aligning and applying advances in the reading and measurement sciences. Lanham (MD): R&L Education; 2012.

Medina AL, Pilonieta P. Once upon a time: comprehending narrative text. In: Schumm JS, editor. Reading assessment and instruction for all learners. New York (NY): Guilford Press; 2006. p. 222–261.

Neigel AR, Caylor JP, Kase SE, Vanni MT, Hoye J. The role of trust and automation in an intelligence analyst decisional guidance paradigm. J Cogn Eng Decis Mak. 2018;12(3):239–247.

Ruddy M. ELICIT – The Experimental Laboratory for the Investigation of Collaboration, Information Sharing, and Trust. Proceedings of the 12th International Command and Control Symposium; 2007.

Simpson H, Cieri C, Maeda K, Baker K, Onyshkevych B. Human language technology resources for less commonly taught languages: Lessons learned toward creation of basic language resources. Proceedings of SALTMIL Collaboration: Interoperability between People in the Creation of Language Resources for Less-Resourced Languages; 2008; Marrakech, Morocco. LREC.

Wallen E, Plass JL, Brünken R. The function of annotations in the comprehension of scientific texts: cognitive load effects and the impact of verbal ability. EduTech Res Dev. 2005;53(3):59–71.

Zaroukian E. Information extraction for optimized human understanding and decision making. Proceedings of the 23rd International Command and Control Research and Technology Symposium; 2018.

Zaroukian E. Automated information extraction to facilitate comprehension across text difficulty levels. Proceedings of the 24th International Command and Control Research and Technology Symposium; 2019.

Zaroukian E, Caylor J, Vanni M, Kase S. Human interaction with the output of information extraction systems. Proceedings of Advances in Artificial Intelligence, Software and Systems Engineering. Applied Human Factors and Engineering. Advances in Intelligent Systems and Computing; 2019a; Washington, DC. Cham (Switzerland): Springer. p. 456–466.

Zaroukian E, Caylor J, Vanni M, Kase S. Evaluating improvement in situation awareness and decision-making through automation. Proceedings of the IEEE Conference on Cognitive and Computational Aspects of Situation Management; 2019b.

List of Symbols, Abbreviations, and Acronyms

ACE	Automatic Content Extraction
AMT	Amazon Mechanical Turk
ARL	Army Research Laboratory
CCDC	US Army Combat Capabilities Development Command
DM	decision making
ELICIT	Experimental Laboratory for the Investigation of Collaboration, Information Sharing, and Trust
HAMIE	Human-Assisted Machine Information Exploitation
IE	information extraction
LODL	Low Density Languages
MUC	Message Understanding Conference
MUC-7	Seventh Message Understanding Conference
NER	named entity recognition

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	CCDC ARL FCDD RLD DCI TECH LIB
3 (PDF)	CCDC ARL FCDD RLC IT E ZAROUKIAN S E KASE M VANNI