The Case for Readability of Crisis Communications in Social Media

Irina Temnikova
itemnikova@qf.org.qa

Sarah Vieweg
svieweg@qf.org.qa

Carlos Castillo
chato@acm.org

Qatar Computing Research Institute, Doha, Qatar

ABSTRACT

The readability of text documents has been studied from a linguistic perspective long before people began to regularly communicate via Internet technologies. Typically, such studies look at books or articles containing many paragraphs and pages. However, the readability of short messages comprising a few sentences, common on today's social networking sites and microblogging services, has received less attention from researchers working on "readability".

Emergency management specialists, crisis response practitioners, and scholars have long recognized that clear communication is essential during crises. To the best of our knowledge, the work we present here is the first to study the readability of crisis communications posted on Twitter—by governments, non-governmental organizations, and mainstream media. The data we analyze is comprised of hundreds of tweets posted during 15 different crises in English-speaking countries, which happened between 2012 and 2013. We describe factors which negatively affect comprehension, and consider how understanding can be improved.

Based on our analysis and observations, we conclude with several recommendations for how to write brief crisis messages on social media that are clear and easy to understand.

1. INTRODUCTION

1.1 What is Readability?

Readability is the ease with which a written text can be read or understood by a reader [8]. Readability is different from "reading ability", which corresponds to the reading skills of the reader [22], and also differs from "legibility" [8], which is concerned with the physical characteristics of a text (font, spacing, and text position on the sheet/screen). Readability is usually expressed as a numerical score of a text, which score is based on series of "readability features", which increase or decrease the text's readability and reading comprehension.

Readability is a well-studied concept that scholars and researchers began to focus on in the beginning of the 20th century [21, 39]. Initially, levels or indexes of readability were applied by teachers regarding educational texts, with the aim of establishing whether the texts were readable for a specific grade level. Nowadays it is well-established that the readability level of texts is a crucial factor regarding their appropriateness for particular ages and/or audiences; in educational, professional, and everyday settings. For example, child mortality due to car accidents is often correlated to the improper installation and/or use of car seats. In turn, incorrect installation and/or use is often due to the poor readability of installation instructions [8, 40]. The readability features are usually defined in accordance with the requirements of a specific group of readers (e.g. students of a certain age or grade level, second language learners of a particular language, and/or dyslexic readers). Readability features span all linguistic levels of text and can include: level of vocabulary (e.g."acquiesce" vs "agree"), word length (e.g. "television" vs "tv"), word ambiguity (e.g. "take the *right* turn, instead of the wrong one" vs "take the *right* turn, instead of the left one"), figurative language (e.g. "The teacher is a dragon." vs "The teacher is scary."), sentence length (a 3-word sentence vs a 20 word sentence), syntactic complexity (a simple sentence vs a sentence with subordination), syntactic and/or modifier ambiguity ("I saw *a man with a telescope*." vs "*I saw* a man *with a telescope*."), cohesion, illogical or unclear word or phrase order, number of paragraphs, and additional factors.

Readability studies are usually divided into "classic" and "modern" studies. *Classic readability studies* [4, 10, 18, 21, 26, 39] usually combine a handful of (up to 5) features (such as sentence length in words, word length in syllables, and number of difficult words) into *"readability formulae"* – metrics which give a single numerical estimation of how difficult a text is to read by correlating its difficulty to a specific grade level or reader age. *Modern readability approaches* [11, 16, 27, 30, 35] employ statistical techniques, such as Machine Learning (ML), which allow for studying the impact of a large number of readability features.

1.2 Clarity of Official Communication

While readability is a well-studied concept, existing studies tend to concentrate on text genres such as educational materials and news articles rather than on messages published to social media sites and microblogging services by official organizations during crises. However, the clarity of communications other than social media messages by such organizations and agencies (government, banks, hospitals, etc.) has been of interest for decades. One example is the Plain English Movement. Its aim is to provide a larger population access to official (e.g. legal, medical) documents. In the UK, the Plain English Movement generated the Plain English Campaign,[1] a UK-based organization which evaluates the clarity of official documents and provides guidelines for writing in Plain English to businesses and institutions such as British Gas, British Telecom,

[1]http://www.plainenglish.co.uk/ accessed on January 21st, 2015.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2741718.

UK councils, and government departments. Examples of official communication converted into plain English (see Table 1) can be found on the Plain English Campaign's website[2], which clearly illustrate why writing in a simple style is necessary.

Table 1: "Before" and "after" messages included in the documentation of the Plain English Campaign.

Before	After
If there are any points on which you require explanation or further particulars we shall be glad to furnish such additional details as may be required by telephone.	If you have any questions, please phone.
High-quality learning environments are a necessary precondition for facilitation and enhancement of the ongoing learning process.	Children need good schools if they are to learn properly.

The need for clear communication is equally as important when it comes to formal response organizations, governments, NGOs and media outlets relaying messages in times of crisis. Communication is considered important in crisis situations and must be kept under control [23, 28, 41]. At the same time it is critical in times of crisis that messages which contain time- and safety-sensitive information are correctly understood [3, 36]. This is especially important because in stressful situations people understand differently due to the very short reaction time [19]. Though there are guidelines for how governments and other formal organizations should communicate during crisis situations (see the Australian government's "Choosing your words" guide[3]) and emergency managers have the option to employ controlled languages [33, 38], to the best of our knowledge, there are no studies which investigate how clear, readable, and comprehensible are real messages published on social media sites during times of crisis.

The aim of this article is to address this gap, and conduct a preliminary investigation on whether existing crisis messages published by government, Non-Governmental Organizations (NGOs), and mainstream media on social media are clear enough for an average person to understand. In this preliminary investigation we aim to study which readability features affect crisis messages and make them more difficult or easy to read and understand, as evaluated by human judges. We employ knowledge and techniques borrowed from readability, text simplification, and psycholinguistics, and analyze a sample of crisis messages posted by official organizations on the popular platform Twitter [4] during various crises. In addition, we turned to crowdsourcing for human assessment and annotation.

The remainder of the paper is organized as follows: Section 2 presents the related work in short messages and tweet readability, Section 3 describes the data collection and annotation, Section 4 provides a qualitative analysis of the annotated crisis tweets, Section 5 provides a quantitative analysis of the annotated tweets, and we offer conclusions in Section 6.

2. READABILITY OF SHORT MESSAGES

Though readability has been studied for close to a century, most studies focus on longer texts (schoolbooks, news articles, technical manuals, and administrative documents), and not on short messages. In this section, we examine related work which has been previously done on the clarity or readability of short messages.

[2]http://www.plainenglish.co.uk/campaigning/examples.html accessed on January 21st, 2015.

[3]http://www.em.gov.au/Emergency-Warnings/Documents accessed on January, 21st, 2015.

[4]https://twitter.com accessed on January 28th, 2015.

The readability (or clarity) of the following types of short messages has been discussed, but minimally: web summaries [17], Airbus and computer software end-user warning messages [14, 37], traffic and train stations signs [25, 33], movie subtitles [7], multiple-choice test items and test questions [12, 31], among others.

In this work we examine the readability of messages posted on Twitter, which are commonly known as "tweets". Tweets are short messages of up to 140 characters. A few studies [1, 5, 13, 42] have investigated the readability of tweets. However, most of them employed readability as a tweet selection method for another application, e.g. topic summarization [42], or a search engine [13]. Davenport and DeLine [5] have applied the Flesch Reading Ease readability formula [10] on a collection of 17.4 million tweets. They found that tweets are significantly less readable than other short messages, such as SMS, or chat. In addition, they correlated the obtained readability scores with the geographic locations of Twitter users. All of these studies applied existing readability metrics developed for other types of texts, or application-dependent heuristic features to a tweet's readability estimation, without taking into account the readability specificities of the tweets per se. As yet, no research has investigated the textual features characteristic for tweets that affect their readability, nor the readability of tweets posted during crises.

Our analysis of the related work is summarized in Table 2, which lists the readability problems we believe might affect tweets.

Table 2: Readability problems applicable to tweets.

Problem	Reference(s)
Long sentences and long messages	Manning [24], Bravo-Lillo et al. [2], Harley [15], May [25], Plain English Campaign
Text choppiness, ellipsis, and truncation	Pym [32], Rose et al. [34], Kanungo and Orr [17], May [25]
Misspellings	Kanungo and Orr [17], Plain English Campaign
Use of hashtags	Davenport and DeLine [5]
Unknown abbreviations and acronyms	Spaggiari et al. [37]
Important ideas not highlighted	Manning [24]
Pictures unrelated to the message	Manning [24]
Use of nominalizations, impersonal style (e.g. "Police calling for immediate evacuation." instead of "Police: Immediately Evacuate!")	Manning [24], Plain English Campaign
Non-standard word order	Spaggiari et al. [37]
Ambiguous words, attachment, and syntax	Spaggiari et al. [37], Harley [15]
Use of unfamiliar words (e.g. "allocate" instead of "give" or "amnesia" instead of "loss of memory")	Manning [24], Bravo-Lillo et al. [2], May [25], Harley [15], Plain English Campaign
Not using connectives (consequently, however, first, but)	Manning [24]
Use of passive voice instead of active (e.g. "The police stopped the riot." instead of "The riot was stopped by the police.")	Manning [24], Harley [15], Plain English Campaign

3. DATA ANNOTATION

3.1 Data Selection

For our experiment, we use the collection from Olteanu et al. [29], CrisisLexT26.[5] This is a freely available collection of tweets from 26 crisis events, which happened in 2012 and 2013, with about 1,000 tweets per crisis, labeled for "informativeness" (*Informative* or *Non-informative*), "information type" (*Affected individuals*, *Infrastructure and utilities*, *Donations and volunteering*, *Caution and advice*, *Sympathy and emotional support*, *Other useful information*), and "source" (*Eyewitness*, *Government*, *NGOs*, *Business*, *Media*, *Outsiders*). From this collection, we have selected a sample of tweets, following these criteria:

1. Crises occurring in countries with a large population of English speakers (see Table 3).
2. Informativeness: *Informative*.
3. Sources: *Government*, *NGOs*, and *Media*.

Table 3: Crises used in this study.

Crisis	Country
2013 Alberta floods	Canada
2013 Australia bushfires	Australia
2013 Bohol earthquake	Philippines
2013 Boston bombings	USA
2013 Colorado Floods	USA
2013 Glasgow helicopter crash	UK
2013 Los Angeles airport shooting	USA
2013 Lac Mégantic train crash	Canada
2013 Manila floods	Philippines
2013 New York train crash	USA
2013 Queensland floods	Australia
2013 Savar building collapse	Bangladesh
2013 Singapore haze	Singapore
2013 Typhoon Yolanda	Philippines
2013 West Texas explosion	USA

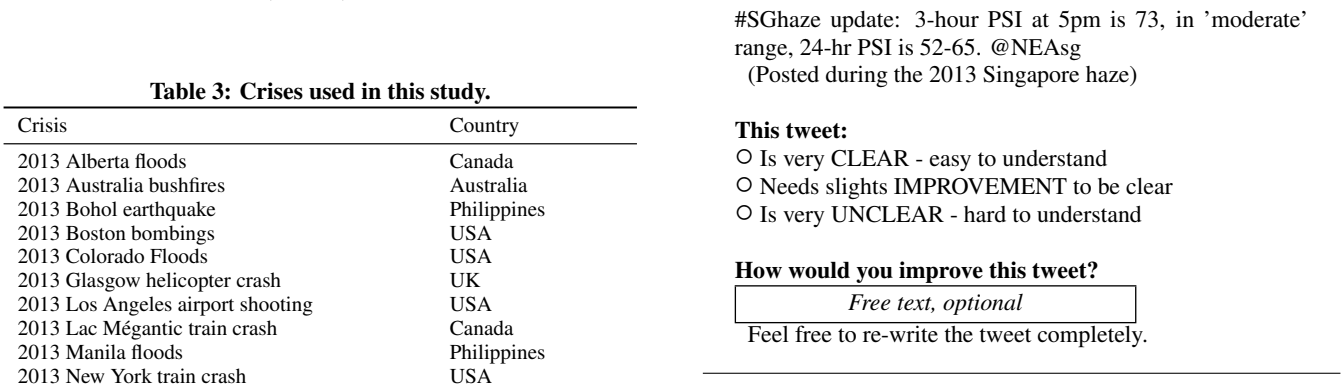
3.2 CrowdFlower Experiment

For the annotation task, we used the crowdsourcing platform CrowdFlower.[6] We employed crowdsourcing workers so that we may get the opinions of people regarding the simplicity and ease of understanding they experience when reading crisis tweets. As a pre-processing step, we randomly selected 500 tweets from the sample, mentioned in Section 3.1, and removed "RT @user:" from the tweet text, as it is usually not visible through Twitter's interface.

Using crowdsource workers to evaluate readability is not new [6, 9]. Feblowitz and Kauchak [9] used Amazon's Mechanical Turk[7] to evaluate the output of an automatic text simplification system working at sentence level. 10 judges per sentence were asked to assign a 1 to 5 Likert [20] score to each sentence. The judges were requested to be living in the US. De Clercq et al. [6] used a their own developed crowdsourcing tool 1) to rank texts per degree of simplicity and 2) for pairwise comparison of the simplicity of two texts.

In our case, each tweet has been annotated by 5 workers, and similarly to [9], we allowed only participants living in countries with a majority of native English-speakers: Australia, Canada, New Zealand, United Kingdom, and United States. The participants were asked to assign each tweet as belonging to one of the three categories: *Is very CLEAR - easy to understand*, *Needs slight IMPROVEMENT to be clear*, and *Is very UNCLEAR - hard to understand*, and optionally, suggest how the tweet could be improved. The experiment was preceded by instructions, providing 2 examples of each tweet category, then by a training phase, which showed 5 examples per category, followed by re-writing explanations for the tweets of the category *Needs slight IMPROVEMENT to be clear*. The instructions did not mention the potential readability problems applicable for tweets, listed in Table 2. Figure 1 shows an example of a crowdsourcing task.

Figure 1: Example crowdsourcing task, displaying one tweet about the 2013 Singapore haze crisis, and asking annotators to indicate how clear is this message. Optionally, annotators can also indicate how to improve or re-write the tweet.



After the crowdsourcing annotation was complete, for the purposes of the analysis described in Sections 4 and 5, we selected tweets annotated with a relatively high confidence score. In the CrowdFlower platform this is a weighted measure of agreement among workers, where each worker is weighted according to how much they agreed with the label given to the test questions. We apply a threshold of $\theta = 0.66$. We asked for a confidence threshold larger than 0.5 to avoid borderline cases in which the label of the message was disputed. Each message has a minimum of 5 labels by different workers, but some messages have more, as they can be presented to more workers by the platform. Heuristically, we used a threshold of 2/3 for the confidence; this value is related to agreement but considers others aspects including worker trust, which is related to their performance on previous tasks. We note that varying this threshold yields minor variations in the proportion of tweets in different categories, and hence in their statistical properties. Table 4 summarizes the data annotation.

Table 4: Characteristics of our annotated dataset applying a minimum threshold of $\theta = 0.66$.

All tweets with confidence $>= \theta$	301	100.0%
Is very CLEAR - easy to understand	247	82.1%
Needs slight IMPROVEMENT to be clear	36	12.0%
Is very UNCLEAR - hard to understand	18	6.0%

[5]http://crisislex.org/tweet-collections.html accessed on January 22, 2015.
[6]http://crowdflower.com/ accessed on January 23, 2015.
[7]https://www.mturk.com/ accessed on January 23, 2015.

4. READABILITY ANALYSIS OF TWITTER COMMUNICATIONS

Based on the annotated dataset obtained in Section 3, in this section we present a series of problems identified in the dataset we analyze, including example tweets, why they are problematic, and how they can be fixed. Sections 4.1. and 4.2. present very unclear and needing improvement tweets, as annotated by the crowdsourcing workers, along with their suggestions, while Section 4.3. provides examples of tweets, re-written by the authors of this paper, following annotators' re-writing of similar tweets.

4.1 Very Unclear Examples (Annotators Suggestions)

Example 1: *More news on our #NCCARF report RT: @ozmining News: Queensland mines were not ready for floods [URL] ...*

– Annotator suggestion: *Queensland mines were not ready for floods.*

– Rewrite factors: The annotator removed the superfluous phrase at the start of the tweet; there is no need to say "more news on our NCCARF report," particularly since the acronym NCCARF is domain-specific and possibly not well known to many readers. The rewritten tweet expresses the important message at the beginning, and avoids any needless extra verbiage.

Example 2: *RT @QPSmedia Major flooding occurring in Lockyer Valley. Evacuations underway #bigwet*

– Annotator suggestion: *Major flooding in Lockyer Valley. Evacuations ongoing. @QPSmedia*

– Rewrite factors: The annotator removed the 'RT' at the start of the tweet, and changed some vocabulary. The annotator deleted the verb "occurring" and simplified the first phrase, while keeping the original meaning. In the second phrase the annotator used what is likely perceived as a more common, simpler word - swapping "underway" for "ongoing."

4.2 Examples Needing a Slight Improvement (Annotators Suggestions)

Example 3: *#SGHaze: PSI now at 250 as of 11pm in Singapore*

– Annotator suggestion: *#SGHaze: As of 11pm Singapore Haze now at 250 pounds per inch*

– Rewrite factors: The annotator changed the word order to make the time of the reported haze level more clear by moving "as of" to the front of the phrase. They also spelled out the acronym "PSI," which is potentially helpful to those readers who are unaware of what PSI is. Unfortunately, the suggested spelling of the acronym is incorrect, as it should be "pollutant standards index." However, we include this example here due to the intent of the annotator regarding replacing an acronym with its full label.

Example 4: *#coflood CDOT opened 7 highways between yesterday & today following flood repairs thanks to maintenance crews & contractors*

– Annotator suggestion: *7 highways opened by CDOT after repairs made by maintenance crews and contractors.*

– Rewrite factors: The annotator removed the hashtag from the tweet completely; it is no longer at the front of the tweet, nor found elsewhere. The annotator also changed the tense to passive voice, and in doing so, highlighted the "7 highways opened" aspect of the message, which is the most important factor to communicate. In addition, the annotator changed the "&" sign to "and" to make it easier to read.

Below are further examples of very unclear tweets, and tweets that require slight improvement. Here, instead of annotator suggestions, we provide our own improvements based on the changes we observed the annotators made.

4.3 Very Unclear Examples (Authors' Suggestions)

In these examples we make our own suggestions on how to improve tweets, as crowdsource workers did not always offer a suggestions regarding those tweets they saw as very unclear.

Example 5: *3-hour PSI is 48. Issued 6am. [URL] #sghaze*

– Authors' suggestion: *At 6am, the Singapore Haze for the past 3 hours is 48 on the Pollutant Standards Index (PSI). [URL] #sghaze*

– Rewrite factors: We start with the specific time the measurement of the haze was reported, and explain what "3-hour PSI" is by detailing that "3-hour" means "over the past three hours," and that PSI is a measurement of pollutants. We kept the URL and hashtag in the same position at the end of the tweet.

Example 6: *For all in Sydney and most of Illawarra "@Nyx2701: Sydney Water statement: #BlueMountains #bushfires [URL] #NSWFires*

– Authors' suggestion: *Sydney Water statement for those in Sydney and Illawarra: [URL] #BlueMountain #NSWFires*

– Rewrite factors: We begin with the main idea of the tweet, and then indicate that it is information for those in Sydney and Illawarra; it is not necessary to mention that it is "most of" Illawarra. The important point is that readers are aware that Illawarra is an affected area, and that people in that area should read the water statement. We kept the URL at the end of the tweet, but we removed one of the hashtags, and kept the two that were most salient to this particular emergency. Those appear at the end of the tweet so as not to disrupt the reading flow.

5. QUANTITATIVE ANALYSIS

In this section we present some preliminary quantitative observations about the labeling done by crowdsourcing workers. This analysis is based on the data annotated in Section 3.

We measure a series of superficial characteristics of text. The intent is not to be comprehensive, but to evaluate if there are some obvious observations from the data labeling. Table 5 summarizes these statistics, with the last column providing the statistical significance p-values (* corresponds to the p-value of 0.10, ** – to the p-value of 0.05, and *** – to the p-value of 0.01).

From Table 5 we note, first, that many of the tweets requiring improvements, or labeled as "Unclear" were not written in English. A manual analysis of these tweets showed that some of these tweets were written in a mixture of languages. Beyond this observation, we note that tweets that are problematic to read tend to include more acronyms (almost double the amount of them), and more user mentions.

Table 5: Characteristics of selected tweets in our dataset. In this table, "Clear" means labeled as "Very Clear", and "Unclear" means labeled as "Needs Slight Improvement" or "Very Unclear"

	Clear	Unclear	p-values
Average length	108.6	93.1	***
Average num. of words	15.5	14.0	**
Average num. of English words	12.0	7.7	***
Average word length	6.3	6.1	
Average number of acronyms	0.3	0.7	***
Average number of mentions	0.3	0.5	*
Average number of hashtags	1.1	1.2	
Fraction with acronyms	25.5%	64.8%	***
Fraction with mentions	23.5%	38.9%	**
Fraction with URLs	56.3%	22.2%	***
Fraction with URLs in the middle	29.2%	11.1%	***
Fraction with ellipsis	17.8%	14.8%	
Fraction with hashtags (#)	68.8%	87.0%	***
Fraction with # at the beginning	6.1%	37.0%	***
Fraction with # in the middle	31.6%	35.2%	
Fraction with # at the end	37.3%	25.9%	*

Acronyms are probably used to shorten tweets and be able to say more in the character limitation of Twitter, however they tend to be associated to messages that are harder to read.

User mentions (e.g. @FEMA) often refer to institutions such as emergency response, media, or government organizations, but they seem to be perceived as making readability of tweets harder.

The usage of hashtags is where we observe the most evident differences. Tweets that are considered problematic to read include more hashtags, especially more hashtags placed at the beginning of the tweet. From these observations, it is clear that hashtags placed at the beginning of the tweet are impairing the readability of crisis tweets.

6. CONCLUSIONS AND RECOMMENDATIONS

We have presented a preliminary investigation regarding the readability of tweets posted by formal organizations and agencies (government, NGOs, and mainstream media) in crisis situations, focusing on crises that occurred in countries with large English-speaking populations. We employed crowdsource workers who ascertained how easy or difficult tweets were to read and understand, and then performed qualitative and quantitative analyses.

Recommendations. On the basis of the crowdsourced annotations, and on the readability issues hypothesized to be applicable to tweets (listed in Table 2) we have the following recommendations for writing easy-to-understand tweets during crisis events:

- Message length:
  - Include a maximum of 1 or 2 main points per tweet.
  - Write brief, concise sentences.
  - Remove superfluous words.
  - Write fully-formed sentences; avoid writing incomplete thoughts, or incomplete messages.
- Vocabulary:
  - Use only simple and familiar words.
  - Use abbreviations and acronyms with care, i.e. only if they are more understandable to the public than their expanded form.
- Twitter-specific elements:
  - Place all hashtags at the end of the tweet; do not write more than 2 hashtags.
  - Avoid mentions (e.g. "@user").

In future work, we plan to deepen the tweet readability analysis to include features that are more difficult to detect, as well as develop a readability metrics for crisis tweets. In addition, we plan to analyse crisis tweets for features which affect the difficulty of tweets' automatic processing by Natural Language Processing (NLP) applications.

Reproducibility. Our dataset, along with the instructions for CrowdFlower workers, is available for research purposes at: http://chato.cl/2015/readability/.

7. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Yelena Mejova for the help with calculating statistical significance.

References

[1] P. Bellot, T. Chappell, A. Doucet, S. Geva, S. Gurajada, J. Kamps, G. Kazai, M. Koolen, M. Landoni, M. Marx, et al. Report on INEX 2012. In *ACM SIGIR Forum*, volume 46, pages 50–59. ACM, 2012.

[2] C. Bravo-Lillo, L. F. Cranor, J. Downs, S. Komanduri, and M. Sleeper. Improving computer security dialogs. In *Human-Computer Interaction–INTERACT 2011*, pages 18–35. Springer, 2011.

[3] D. P. Coppola. *Introduction to international disaster management*. Butterworth-Heinemann, 2006.

[4] E. Dale and J. S. Chall. A formula for predicting readability. *Educational research bulletin*, pages 11–28, 1948.

[5] J. R. Davenport and R. DeLine. The readability of tweets and their geographic correlation with education. *arXiv preprint arXiv:1401.6058*, 2014.

[6] O. De Clercq, V. Hoste, B. Desmet, P. Van Oosten, M. De Cock, and L. Macken. Using the crowd for readability prediction. *Natural Language Engineering*, pages 1–33, 2013.

[7] Z. C. De Linde. *Linguistic and visual complexity of televison subtitles.* PhD thesis, University of Bristol, 1997.

[8] W. H. DuBay. The principles of readability. 2004.

[9] D. Feblowitz and D. Kauchak. Sentence simplification as tree transduction. In *Proc. of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pages 1–10, 2013.

[10] R. Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948.

[11] T. François and E. Miltsakaki. Do NLP and machine learning improve traditional readability formulas? In *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations*, pages 49–57. Association for Computational Linguistics, 2012.

[12] A. C. Graesser, Z. Cai, M. M. Louwerse, and F. Daniel. Question Understanding Aid (QUAID) a web facility that

tests question comprehensibility. *Public Opinion Quarterly*, 70(1):3–22, 2006.

[13] S. Guo, G. Zhang, and R. Zhai. Integrating readability index into Twitter search engine. *British Journal of Educational Technology*, 42(5):E103–E105, 2011.

[14] M. Harbach, S. Fahl, T. Muders, and M. Smith. Towards measuring warning readability. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 989–991. ACM, 2012.

[15] T. A. Harley. *The psychology of language: From data to theory*. Psychology Press, 2013.

[16] M. Heilman, K. Collins-Thompson, and M. Eskenazi. An analysis of statistical models and features for reading difficulty prediction. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 71–79. Association for Computational Linguistics, 2008.

[17] T. Kanungo and D. Orr. Predicting the readability of short web summaries. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 202–211. ACM, 2009.

[18] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, DTIC Document, 1975.

[19] D. Kiwan, A. Ahmed, and A. Pollitt. The effects of stress on text comprehension and performance in examinations. In *Paper presented at the BPS London Conference*, 1999.

[20] R. Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[21] B. A. Lively and S. L. Pressey. A method for measuring the vocabulary burden of textbooks. *Educational administration and supervision*, 9(389-398):73, 1923.

[22] I. Lorge. Predicting readability. *The Teachers College Record*, 45(6):404–419, 1944.

[23] J. Lundberg and M. Asplund. Communication problems in crisis response. In *Proceedings of the 8th International ISCRAM Conference: Lisbon, Portugal, May 2011*, 2011.

[24] D. Manning. Writing readable health messages. *Public Health Reports*, 96(5):464, 1981.

[25] F. May. Communicating the message: Characteristics of train station notice board texts and implications thereof for translation. Master's thesis, University College London, UK, 2011.

[26] G. H. McLaughlin. Smog grading: A new readability formula. *Journal of reading*, 12(8):639–646, 1969.

[27] D. S. McNamara, M. M. Louwerse, P. M. McCarthy, and A. C. Graesser. Coh-metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47(4):292–330, 2010.

[28] M. Ogrizek and J.-M. Guillery. *Communicating in Crisis:[a Theoretical and Practical Guide to Crisis Mangement]*. Transaction Publishers, 1999.

[29] A. Olteanu, S. Vieweg, and C. Castillo. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of CSCW 2015 (forthcoming)*, 2015.

[30] E. Pitler and A. Nenkova. Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 186–195. Association for Computational Linguistics, 2008.

[31] B. S. Plake. Application of readability indices to multiple-choice items on certification/licensure examinations. *Educational and psychological measurement*, 48(2):543–551, 1988.

[32] P. J. Pym. Pre-editing and the use of simplified writing for MT: an engineer's experience of operating an MT system. *Translating and the Computer*, 10:80–96, 1990.

[33] J. Renahy, D. Devitre, A. Dziadkiewicz, I. Thomas, et al. Controlled language norms for the redaction of security protocols: Finding the median between system needs and user acceptability. In *Proceedings of the 11th International Symposium on Social Communication*, 2009.

[34] D. E. Rose, D. Orr, and R. G. P. Kantamneni. Summary attributes and perceived search quality. In *Proceedings of the 16th international conference on World Wide Web*, pages 1201–1202. ACM, 2007.

[35] S. E. Schwarm and M. Ostendorf. Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 523–530. Association for Computational Linguistics, 2005.

[36] M. W. Seeger, T. L. Sellnow, and R. R. Ulmer. Communication, organization, and crisis. *Communication yearbook 21*, page 231, 2012.

[37] L. Spaggiari, F. Beaujard, and E. Cannesson. A controlled language at Airbus. *EAMT-CLAW 2003*, 2003.

[38] I. Temnikova. *Text Complexity and Text Simplification in the Crisis Management domain*. PhD thesis, University of Wolverhampton, UK, 2012.

[39] M. Vogel and C. Washburne. An objective method of determining grade placement of children's reading material. *The Elementary School Journal*, pages 373–381, 1928.

[40] M. V. Wegner and D. C. Girasek. How readable are child safety seat installation instructions? *Pediatrics*, 111(3): 588–591, 2003.

[41] L. Winerman. Crisis communication. *Nature*, 457:376, 2009.

[42] D. Yajuan, C. Z. Weif Uru, Z. M. Heung, and Y. Shum. Twitter topic summarization by ranking tweets using social influence and content quality. In *Proceedings of the 24th International Conference on Computational Linguistics*, pages 763–780, 2012.