J Canc Educ (2018) 33:134–140
DOI 10.1007/s13187-016-1059-5

Making Cancer Health Text on the Internet Easier to Read for Deaf People Who Use American Sign Language

Poorna Kushalnagar[1] · **Scott Smith**[2] · **Melinda Hopper**[2] · **Claire Ryan**[3] · **Micah Rinkevich**[2] · **Raja Kushalnagar**[1]

Published online: 8 June 2016

Abstract People with relatively limited English language proficiency find the Internet's cancer and health information difficult to access and understand. The presence of unfamiliar words and complex grammar make this particularly difficult for Deaf people. Unfortunately, current technology does not support low-cost, accurate translations of online materials into American Sign Language. However, current technology is relatively more advanced in allowing text simplification, while retaining content. This research team developed a two-step approach for simplifying cancer and other health text. They then tested the approach, using a crossover design with a sample of 36 deaf and 38 hearing college students. Results indicated that hearing college students did well on both the original and simplified text versions. Deaf college students' comprehension, in contrast, significantly benefitted from the simplified text. This two-step translation process offers a strategy that may improve the accessibility of Internet information for Deaf, as well as other low-literacy individuals.

Keywords Deaf · American Sign Language · Text simplification · Health text

✉ Poorna Kushalnagar
poorna.kushalnagar@gmail.com

[1] Gallaudet University, Washington, DC, USA

[2] Rochester Institute of Technology, Rochester, NY, USA

[3] University of Texas at Austin, Austin, TX, USA

Background

Health communication has been rapidly evolving to address barriers to health information usage across various populations. Using "plain language" is recommended when creating text-based health information for the public [1]. In addition, the National Institute of Health (NIH) recommends that online health materials be written at the fourth to sixth grade level. To date, this recommendation has not been fully implemented. Recent readability studies of cancer-related materials online estimate their reading level to be between 9.2 and 14.2 [2, 3]. Another readability study of ten widely used websites for breast reconstruction reported the grade level for online health materials to be between 10.7 and 15.8 [4]. Of the 32 websites on dental care for patients with cancer, none had health materials that could be read at sixth grade level or lower [5].

Online health information written in English can especially be difficult for some deaf adults whose first language is not English. A nationwide standardized testing study has shown that 50 % of deaf 18-year olds read at a fourth grade level or lower [6]. A more recent cohort study found that this average reading level remained unchanged [7]. Additionally, deaf people often miss out on information that is typically acquired through incidental learning in a primarily spoken language society [8]. Deaf college students were found to have smaller English vocabularies than hearing peers, thought to be caused by the inability to overhear peripheral conversation in spoken language [9]. For these populations of deaf adults, presenting health information in ASL or using high-frequency vocabulary and simplified grammar in English may be beneficial.

A group of researchers [10] recommended "planning, implementing, and evaluating a free, web-based repository of reliable health resources specifically designed for delivery in both ASL and English print" as the next steps toward promoting accessible health information for a population of deaf people who use ASL. In the past decade, there have been increasing efforts to develop and post ASL cancer health education videos on the Internet (see [11] for a review). Yet, it remains unclear who the users of these ASL health videos are. A recent health website usability research with a group of primarily college-aged Deaf students revealed that some ASL health videos were not easy to find [12]. In another website, usability study of ten mid-aged Deaf adults in Spain, researchers found that simplified text resulted in higher rates of task completion and less frustration for deaf and hard-of-hearing participants [13]. In reality, abundant health website resources are rapidly expanding across specialized topics, yet accessible health information is difficult to find.

Automated Text Simplification

Many consumers who are not fluent in English are unable to easily understand online information, which motivates us to investigate the feasibility of text simplification of web pages. Text simplification is aimed at reducing the reading complexity of text while retaining its original meaning, by rephrasing the text at multiple levels: word, sentence, paragraph, discourse, or semantic levels. For example, word-level strategies can include replacing difficult words with simpler synonyms or replacing or medical jargon with definitions. Similarly, sentence level strategies can include splitting longer sentences with embedded clauses into simpler, shorter sentences or eliminating peripheral information. Finally, semantic strategies include simplification of meaning or rephrasing to include terms familiar to the reader's background. Simplicity is intuitively obvious, yet hard to define. For instance, simplified versions of text often are shorter in terms of words and syllable count, but this is not always true because concise, complicated descriptions may need further clarifications. In light of this, text simplification is its own niche in natural language processing and draws from related areas such as text summarization and machine translation [14].

Early research in text simplification focused on either manual simplification or automatically applying rules to text. For instance, Chandrasekar and Srinivas [15] describe a process for automatically generating simplification rules from annotated training data. Unfortunately, text simplification remains largely an unattained goal and has not been widely adopted in industry and applications because the chief limiting factor in these systems is their inaccuracy [16–18].

Automated text simplification is an important area of research because it can be enormously beneficial to people who may need assistance in reading nontrivial documents, such as those with autism spectrum disorder [19] and dyslexia [20] and those who are second-language learners [21]. One area of text simplification that has not been explored much in the research community is its application to some deaf people who do not either have the fund of information or English fluency to understand typical health information in print.

Purpose of Study

The purpose of the study is to explore the benefit of text simplification to make existing cancer health text on the Internet easier to read and understand by deaf college students whose primary language is American Sign Language. If information presented in text online could be translated into simpler English that is understood by deaf people who are relatively less proficient in English, then health information would be more accessible for everyone with a wide range of reading levels.

Because current text simplification algorithms are not yet mature enough to accurately simplify medical information, we used the text simplification algorithm output to support rapid rewriting of the cancer health text to a representation that used less jargon and shorter sentences in a simplified text condition. We report the results of a study in which we asked participants to read both the original text and simplified text, and answer questions related to the medical information in the text.

Methods

Iterative Simplification

We reviewed several public health websites and selected materials from www.medicinenet.org. We decided to use breast cancer slideshow materials at this website because they included short slide shows with pictures and text that was intended for the general public. We chose this topic specifically due to its high importance as a public health concern. At least 1 in 8 females in the USA would have breast cancer in their lives [22]; therefore, it is likely that a majority of the population will be affected by this disease directly or indirectly. Original slide and cancer health text materials at this website were evaluated using a Flesch–Kincaid readability test [23] and found to be written at a range of 7.7 to 14.2 grade levels.

The first step in the text simplification process was to extract the paragraphs on each slide to a text file. The paragraphs were then fed into an automated text simplification program called MOSES, which read and rewrote the paragraphs into simplified English text. The resulting simplified English paragraphs were imperfect and required a second manual step to be consistent and accurate. In the second step, an English professor with expertise in educating deaf students reviewed the automatically generated simple English sentences. The goal in the second step was to convert from imperfect simplified English to easy-to-read, grammatically accurate sentences. Following this iteration, a physician with expertise in Deaf health and health literacy reviewed the simplified text version for concept equivalency with the original text. Below, we provide an example of the two-step text simplification process.

Slide: Finding a Lump
Original text on slide: Remember that the majority (about 80%) of breast lumps are not due to cancer. Cysts, benign tumors, or changes in consistency due to the menstrual cycle can all cause benign breast lumps. Still, it's important to let your doctor know about any lumps or changes in your breast that you find. Early detection of breast cancer is associated with high cure rates. Grade level, 7.6.
Pass 1, Text after automated text simplification: Majority (about 80%) breast lumps are not due to cancer. Cysts, benign tumors, or consistency changes due to the menstrual cycle can cause benign breast lumps. Important to let your doctor know about any lumps or changes in your breast. Early detection of breast cancer gives high cure rates.
PASS 2, Text after manual text simplification: Remember most breast bumps are not from cancer (about 4 of 5). Cysts, non-cancer tumors, or changes because of period can make breast bumps. It is important to let your doctor know if you find any bumps or changes in your breast. Early catch breast cancer can help you become healthy again. Grade level, 5.4 (see Fig. 1).

The simplified version of the breast cancer slideshow materials had a Flesch–Kincaid readability of between 3.5 and 9.7 grade levels, with an average grade level of 6.4 for each slide. For the simplified slides that had higher grade levels (grades 7.0 through 9.7), these slides included low-frequency medical terms that could not be further simplified. Examples of such medical terms would be "tumor" and "chemotherapy." In these cases, pictures were essential in providing added contextual information to these words. For example, the slide that contained the word "tumor" had an x-ray photo of the breast that showed what the tumor looked like. The chemotherapy slide showed a woman who was connected to an IV pump.

Usability sessions were conducted with four deaf college students. Study participants were recruited through word of mouth. At time of participation, these students were enrolled in an entry-level analytical reading and writing English course. All answered a baseline question about the prevalence of breast cancer compared to other diseases. Following this, two student participants were assigned to original, unsimplified text condition (condition 1), in which they were asked to read a short paragraph below a slide of pictures related to breast cancer. When the slides ended, the participants completed a multiple-choice quiz. The other two participants were assigned to a simplified text condition (condition 2) and followed the same procedure. Students were given an option to review the slides as they answered the questions, a behavior that would be similar to as if they were browsing for information on the Internet. All participants took advantage of this option to refer back to the slides to assist with completing the quiz.

Following quiz completion, the researcher displayed corresponding slides from both of the original and simplified conditions for the participant to compare. Participants were asked which condition appeared to be easier to read and why. For each slide, the researcher recorded the participants preferred condition and their descriptive reasoning as to why the preferred condition was easier to read. The qualitative responses from each participant were carefully reviewed and taken into consideration for improving the simplified text condition's readability. Regardless of the order of presentation, deaf college student participants preferred the simplified version (condition 2). Common responses in favor of the simplified version included (1) shorter sentences, (2) high-frequency words, and (3) added contextual information from pictures.

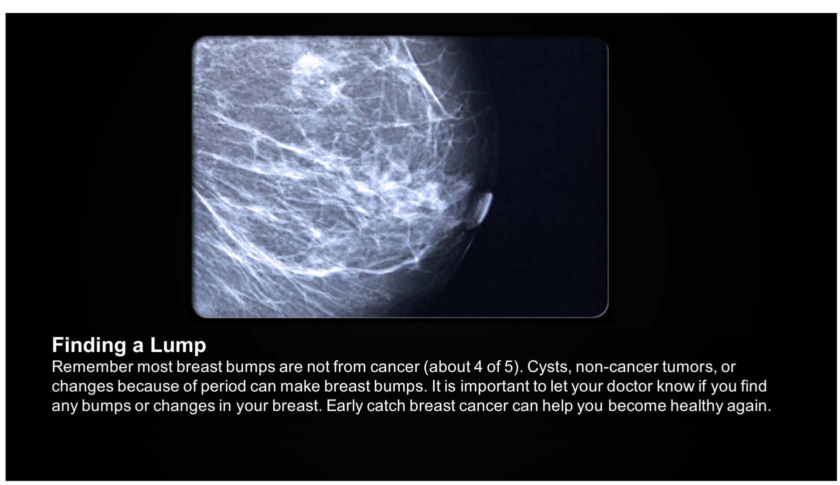

Fig. 1 Example of slide after two-step automated and manual text simplification, grade level: 5.4

Slide Materials

Slides contained a paragraph of text, each covering a different topic about breast cancer, such as likelihood of contracting breast cancer, symptoms, or treatment. The breast cancer slide show with ten slides was divided in half to form two separate slide shows (5 slides per condition). In condition 1, the first half of the slide show contained original text whereas the second half contained simplified text. This order was reversed in condition 2. This allowed us to use a 2 × 2 crossover design where deaf and hearing (control) participants were exposed to both original and simplified conditions. We used counterbalancing to provide some control for carryover effects: half of the participants began with the original condition and ended with the simplified condition, and this order was reversed for the other half.

Procedure

After review and approval from the institution's human subject board, recruitment was done through word of mouth and flyers posted on campus. Each session began with informed consent and explanation of the study procedure in ASL. Participants were then handed an English version to read and sign. All participants were assigned a unique alphanumeric ID to preserve confidentiality. After the participant gave signed consent, we began the session with background, a question about breast cancer risk and prevalence, and functional health literacy questionnaires. The question about breast cancer risk and prevalence was used to establish a pre-knowledge baseline control for both deaf and hearing samples. We added this pre-knowledge baseline to factor out existing knowledge about breast cancer that might override the benefit of the simplified text in the breast cancer slides. For the functional health literacy questionnaire, we used the paper-based Short Test of Functional Health Literacy Assessment (STOFHLA) [24] that required the participant to read two short health-related passages and fill in missing words in 7 min. We included this as a measure of the individual's health-related reading skills to control for variation in English proficiency.

Once both STOFHLA passages were completed or the time passed 7 min, participants proceeded to the breast cancer slides. Participants read the first half of the breast cancer slides in either original or simplified condition then took a quiz based on the information included in the slides. Following completion of the first quiz, participants read the second half of the breast cancer slides in the alternative condition and took a second quiz on the material they just read. The participants were allowed to refer back to the slides while answering questions. We allowed this so to reflect real-world experience where Internet surfers naturally switch between web pages to review information. At the end of the session, the researcher provided answers to the quizzes to ensure the participant left with accurate knowledge of breast cancer information. The total time to complete the entire session was 30 min. Participants were compensated for their time and participation.

Study Sample

Seventy-four deaf and hearing college students were recruited to participate in the study. As shown in Table 1, 49 % were from racial/ethnic minority groups and 48 % were female. Within the deaf subgroup, 48 % had mothers with high school degree or lower. In contrast, only 25 % of the hearing subgroup had mothers with high school degree or lower. This is consistent with profile of our private institution that includes a majority of hearing students from educated families and a majority of deaf students who are supported on social security

Table 1 Descriptive statistics for the sample of deaf and hearing participants ($N = 74$)

Variable	Total $N = 74$ M (SD)	Deaf ($n = 38$) M (SD)	Hearing ($n = 36$) M (SD)
Age	21 (3)	22 (3)	20 (2)
Parent years of education	14.1 (2.8)	13.6 (3.3)	14.9 (2.0)
Functional health literacy	33.7 (3.8)	32.4 (4.3)	34.9 (2.7)
		%	%
Race			
Non-White	49 %	53 %	42 %
White	51 %	47 %	58 %
Gender			
Female	49 %	53 %	44 %
Male	51 %	47 %	56 %
Parent hearing status			
Deaf	22 %	42 %	0 %
Hearing	78 %	58 %	100 %

Table 2 Test of within-subject effects for condition

Source	Type III sum of squares	*df*	Mean square	*F*	*p* value	Partial eta-squared
Text condition	.703	1	.703	.01	NS	.001
Text condition × functional hearing literacy	6.619	1	6.619	.05	NS	.000
Text condition × hearing status	967.386	1	967.386	6.51[a]	<.01	.086
Error (text condition)	10,260.408	69	148.702			

[a] Post hoc analyses showed that deaf sample significantly benefitted from the simplified text condition, whereas hearing sample did well on both simplified and original conditions

income or vocational rehabilitation services. For functional health literacy, the hearing college student sample had significantly higher score than the deaf college student sample ($F = 8.88$; $p < .01$). The mean score and standard deviations for each group are shown in Table 1.

Results

Descriptive analyses were conducted to examine the distribution of quiz score across hearing status and text conditions. We also examined correlations between background variables (e.g., pre-knowledge of cancer risk and prevalence; functional health literacy) and quiz score-dependent variable. Functional health literacy was the only variable that was significantly associated with the quiz score for both deaf and hearing samples and therefore entered as a covariate in the model. Race, gender, parent's hearing status (e.g., deaf or hearing), and baseline knowledge of breast cancer prevalence and risk were not associated with the quiz score outcome measure for both samples and therefore not entered in the model.

One-way repeated measures ANOVA was used to compare original and simplified text condition quiz scores across deaf and hearing groups. Significant within-subject effects were followed up by post hoc analyses in each hearing group. In the post hoc analyses, paired-samples *t* tests were used to compare quiz scores between original and simplified text conditions in each hearing group.

After controlling for functional health literacy, there were no statistically significant differences between deaf and hearing group scores for the original and simplified conditions. As shown in Table 2, significant within-subject effects emerged for the text condition by hearing status. Generally, after controlling for functional health literacy, all participants did better on the simplified text condition ($M = 91.4$; SE = 1.4) compared to the original text condition ($M = 88.6$; SE = 1.8). When we conducted post hoc analyses to examine within-subject effects for the text condition, we found that deaf participants did significantly better on the simplified text condition quiz compared to original text condition quiz. Hearing participants, on the other hand, performed similarly on both conditions. Paired samples *t* test results for both groups are shown in Table 3.

Discussion

Because the language used at the health website was developed for and by hearing people, the hearing readers in our study did well on the original and simplified condition quizzes. For deaf readers in our study, the simplified condition was especially helpful. Simplifying the text on future websites would be beneficial not only for some deaf people but also other readers who have relatively low English proficiency. For example, in the past, closed-captions on television was thought to be designed for deaf viewers but in time, it came to our realization the closed-captioning was an asset to all readers with a range of English proficiency. Simplifying text information would include limiting to fewer concepts at one time, reducing clauses or phrases, using basic coordinating conjunctions, and providing synonyms for uncommon or health-laden terms.

Regardless of whether the deaf participants began with the original text and ended with the simplified text or vice versa, they performed significantly better on the simplified condition quiz than the original condition quiz even after factoring out

Table 3 Text condition quiz score comparisons by hearing status

	Mean		95 % CI of the mean difference between original and simplified text quizzes		Paired samples *t* test
	Original	Simplified	Lower	Upper	*t*
Deaf	82.43	91.89	−16.08	−2.84	−2.90*
Hearing	93.89	91.11	−2.345	7.90	1.10

*$p > .01$

functional hearing literacy skills. This was not the case for hearing readers as they performed comparably well on both original and simplified text condition quizzes regardless of the order. We believe this is because hearing participants are able to draw upon their access to incidental learning and prior knowledge to decipher or interpret newer or more difficult concepts. As for the deaf participants in our study, because of being situated in spoken language-privileged environments, it is possible that they have been relegated as bystanders most of their lives where they do not access to a shared and common language, their choices of participating and discussing important topics such as health issues are limited [8]. In addition, hearing participants who have access to health-related discussions with their physician typically have better knowledge and skill to navigate novel health information. Furthermore, because of their privileged daily and accumulative access to English language usage, hearing college students may well have higher health literacy levels than deaf college students. If simplified text can serve to increase or ensure comprehension among these deaf students, it is likely to have an even more profound impact for other language minority groups such as hearing English language learner college students who are learning English as their second, third, or fourth language. Youth and seniors, regardless of hearing status, may also benefit from the simplified text, particularly among those who have unmet health needs. Many deaf youth do not attend college and text simplification of internet information might be particularly valuable for those deaf youth and adults who do not.

Since the majority of healthcare providers are not fluent in ASL, some deaf individuals who use ASL have limited access to engage in health-related discussions with their physician due to language barriers. To accommodate, these deaf individuals must rely on alternative sources of health information [25]. This makes it all the more important for health text on the Internet to be written in a language that everyone, including deaf readers, are able to understand.

In the above study, we used two steps to simplify the text materials taken from a health website designed for the general public. This process has been proven to be beneficial for both deaf and hearing college students. Replicating this method with endless text on the Internet can be quite time consuming and expensive. One method to automate text simplification is by treating it as an English-to-English translation task and using statistical machine translation software trained with an aligned corpus of "normal" English and "simple" English text. Currently, however, the largest aligned data set for simple English text consists of 167,000 paired sentences from Simple and normal English Wikipedia, and it is too small to form a robust translation model. Therefore, to effectively perform text simplification; it makes sense to use a two-stage process. The first stage would apply a statistical translation software such as Moses [26] to generate a simplified version of the text. The second stage would involve human editors who clean up the text to make it grammatically correct or to possibly suit it for a target audience. Content experts can then check for concept equivalency. The hope is that the use of the statistical translation software as a first step will ease the burden on the human editors to make text simplification happen almost in real-time.

References

1. Stableford S, Mettger W (2007) Plain language: a strategic response to the health literacy challenge. J Public Health Policy 28(1):71–93
2. Ellimoottil C, Polcari A, Kadlec A, Gupta G (2012) Readability of websites containing information about prostate cancer treatment options. J Urol. doi:10.1016/j.juro.2012.07.105
3. Pruthi A, Nielsen ME, Raynor MC, Woods ME, Wallen EM, Smith AB (2015) Readability of American online patient education materials in urologic oncology: a need for simple communication. Urology 85:351–356
4. Vargas CR, Ricci JA, Chuang DJ, Lee BT (2015) Online patient resources for liposuction. Ann Plast Surg. doi:10.1097/SAP.0000000000000438
5. Patton LL, George SF, Hollowell RP (2014) Content, quality, and readability of website information on dental care for patients with cancer. Oral Surg Oral Med Oral Pathol Oral Radiol. doi:10.1016/j.oooo.2014.03.019
6. Traxler CB (2000) Measuring up to performance standards in reading and mathematics: achievement of selected deaf and hard-of-hearing students in the national norming of the 9th Edition Stanford Achievement Test. J Deaf Stud Deaf Educ 5:337–348
7. Qi S, Mitchell RE (2011) Large-scale academic achievement testing of deaf and hard-of-hearing students: past, present, and future. J Deaf Stud Deaf Educ. doi:10.1093/deafed/enr028
8. Hopper MJ (2011) Positioned as bystanders: deaf students' experiences and perceptions of informal learning phenomena (doctoral dissertation). University of Rochester, Rochester, NY
9. Convertino C, Borgna G, Marschark M, Durkin A (2014) Word and world knowledge among deaf learners with and without cochlear implants. J Deaf Stud Deaf Educ. doi:10.1093/deafed/enu024
10. Smith CE, Massey-Stokes M, Lieberth A (2012) Health information needs of d/Deaf adolescent females: a call to action. Am Ann Deaf 157:41–47
11. Kuenburg A, Fellinger P, Fellinger J (2016) Health care access among deaf people. J Deaf Stud Deaf Educ 21(1):1–10
12. Kushalnagar P, Naturale J, Paludneviciene R, Smith SR, Werfel E, Doolittle R, Jacobs S, DeCaro J (2015) Health websites: accessibility and usability for American sign language users. Health Commun. doi:10.1080/10410236.2013.853226
13. Pascual A, Ribera M, Granollers T (2014) Impact of web accessibility barriers on users with hearing impairment. Proceedings of the XV international conference on human computer interaction. ACM, Puerto de la Cruz, Tenerife, Spain, pp. 1–2
14. Siddharthan A (2014) A survey of research on text simplification. Int J Appl Linguistics 165(2):259–298
15. Chandrasekar R, Srinivas B (1997) Automatic induction of rules for text simplification. Knowl-Based Syst 10:183–190
16. Xu W, Callison-Burch C, Napoles C (2015) Problems in current text simplification research: new data can help. Trans Assoc Comput Linguist 3:283–297 Retrieved from https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/549

17. Štajner S, Mitkov R, Saggion H (2014) One step closer to automatic evaluation of text simplification systems. In: Proceedings of the 3rd workshop on predicting and improving text readability for target reader populations (PITR)@ EACL pp 1–10
18. Shardlow M (2014) A survey of automated text simplification. Int J Adv Comput Sci Appl 4(1). doi:10.14569/SpecialIssue.2014.040109
19. Štajner S, Evans R, Orasan C, Mitkov R (2012) What can readability measures really tell us about text complexity. In: Proceedings of workshop on natural language processing for improving textual accessibility pp 14–22
20. Rello L (2012) DysWebxia: a model to improve accessibility of the textual web for dyslexic users. In: SIGACCESS accessible computing 102: 41–44. ACM, New York
21. Petersen SE, Ostendorf M (2007) Text simplification for language learners: a corpus analysis. In: Proceedings of workshop on speech and language technology for education
22. Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Altekruse SF, Kosary CL, et al. (2012) SEER cancer statistics review, 1975–2009 (Vintage 2009 Populations). National Cancer Institute, Bethesda Retrieved September 7, 2012
23. Kincaid JP, Fishburne RP Jr, Rogers RL, Chissom BS (1975) Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy-enlisted personnel (No. RBR-8-75). Naval technical training command Millington TN research branch
24. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J (1999) Development of a brief test to measure functional health literacy. Patient Educ Couns 38(1):33–42
25. Margellos-Anast H, Hedding T, Miller L (2004) Improving access to health and mental health for Chicago's deaf community: a survey of deaf adults. Sinai Health System and Advocate Health Care, Chicago
26. Koehn P, Hoang H, Birch A, Callison-Burch C, Federico M, Bertoldi N, Cowan B et al (2007) Moses: open source toolkit for statistical machine translation. In: Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions (pp. 177–180). Association for computational linguistics