Design and Evaluation of a Social Media Writing Support Tool for People with Dyslexia

Shaomei Wu
shaomei@fb.com
Facebook

Lindsay Reynolds
lindsayreynolds@fb.com
Facebook

Xian Li
xianl@fb.com
Facebook

Francisco Guzmán
fguzman@fb.com
Facebook

ABSTRACT

People with dyslexia face challenges expressing themselves in writing on social networking sites (SNSs). Such challenges come from not only the technicality of writing, but also the self-representation aspect of sharing and communicating publicly on social networking sites such as Facebook. To empower people with dyslexia-style writing to express themselves more confidently on SNSs, we designed and implemented *Additional Writing Help* (AWH) - a writing assistance tool to proofread text produced by users with dyslexia before they post on Facebook. AWH was powered by a neural machine translation (NMT) model that translates dyslexia style to non-dyslexia style writing. We evaluated the performance and the design of AWH through a week-long field study with 19 people with dyslexia and received highly positive feedback. Our field study demonstrated the value of providing better and more extensive writing support on SNSs, and the potential of AI for building a more inclusive Internet.

CCS CONCEPTS

• **Human-centered computing → Accessibility systems and tools**;

KEYWORDS

Artificial intelligence; dyslexia; accessibility; social media

ACM Reference Format:
Shaomei Wu, Lindsay Reynolds, Xian Li, and Francisco Guzmán. 2019. Design and Evaluation of a Social Media Writing Support Tool for People with Dyslexia. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300746

ACM ISBN 978-1-4503-5970-2/19/05
https://doi.org/10.1145/3290605.3300746

1 INTRODUCTION

Although the effect of dyslexia varies widely, it often impacts the ability to process and recognize text. Previous research [46] showed that people with dyslexia face challenges when writing on social media and that writing creates a tension between the freedom of self-expression and the social stigma around "bad" writing. Thus, dyslexia can present an obstacle to one's self representation on social media.

To better support this community, we designed and evaluated a dyslexia-specific writing support tool to increase confidence and reduce anxiety associated with writing on social media. Although existing general writing tools, such as spell checkers and auto-correct, provide value to the dyslexia community [46], there are some major issues with current tools. First, designed for the general population, these tools are generally less reliable at identifying and remedying the errors that people with dyslexia are prone to making. Second, most existing tools were designed for formal writing tasks such as homework assignments and work communications, making it difficult to provide suggestions or corrections for social media's linguistic and communication style.

To address this, we designed and trained a Neural Machine Translation (NMT) based spell/grammar checking model to be sensitive to dyslexia specific errors and accustomed to the social media context. The idea is to "translate" text with common dyslexia writing issues to text without, while preserving slang, abbreviations, and hash/mention tags. This is a novel approach for both Neural Machine Translation (NMT) [2] and dyslexia spell checking [44], and we have made two major technical contributions. First, we applied a sequence-to-sequence (seq2seq) model with a character encoder. Second, using the technique of back-translations (and data augmentation in general [17]), we were able to generate a large scale synthetic training data by utilizing public text available on SNSs and injecting common dyslexia writing issues into the text to train our model.

We evaluated our model with datasets of dyslexia style text from social media and other sources [41], and performance was often comparable to mass market tools. We also deployed our model to power an *"Additional Writing Help"* (AWH) tool (Figure 1) for writing comments on Facebook and conducted a field study with 19 participants with dyslexia. AWH provides suggestions for common dyslexia style spelling mistakes (not content), but currently does not learn and accommodate individual unique patterns of spelling errors and writing style. Participants reacted positively to the tool and reported increased confidence in writing after using it.

2 LITERATURE REVIEW

Dyslexia

Dyslexia is most commonly characterized by various difficulties in word recognition, reading fluency, spelling, and writing [49]. Although the prevalence varies by language [27], it is estimated to affect 10-17% of English speakers [7, 10, 53]. Dyslexia can not only create academic or professional challenges, but also lead to social and emotional consequences [25, 47], such as low confidence in reading and writing [24].

Previous technical approaches to accessibility for dyslexia have focused on reading, including adding colored overlays, deploying special typography, increasing font size and margin space, and a combination of these changes [14, 31, 43]. Others have focused on improving reading comprehension by, for example, modifying content to be easier to read [45]. In contrast, little research has explored improving writing, and those that did have primarily explored and drawn from text samples written for school assignments [41, 42, 52].

Writing on social media

While most research on dyslexia investigates academic or professional settings, many everyday activities rely on reading and writing. One example is social media, which is thoroughly integrated into daily life, with more than 79% of Americans using Facebook and 24% using Twitter [22]. Although photos and videos are common, written text is still a key form of communication on social media. Furthermore, previous research has demonstrated that writing patterns on social media differ from business or academic writing [4, 16].

People often aim to have a positive *self-presentation* on social media (e.g., [5, 15, 57]. Self-presentation aims to "convey an impression to others which is in his interests to convey" [21]. Successfully managing a positive self-presentation may foster relationships and accumulate social capital [54].

Previous research explored how people with dyslexia experience social media, and found that writing was more challenging than reading [46]. This perception was not only because of the technical challenges of writing, but also because of the concerns about being judged for their writing quality. Due to the prevalence of text (e.g., comments, posts) on SNSs, effective self-presentation involves clearly articulating via writing, and people with dyslexia were not confident about their ability to do so. People with dyslexia reported receiving more writing-related negative feedback on Facebook than people without. Furthermore, when people with dyslexia received negative feedback, they responded more strongly, such as deleting posts or posting less in the future.

However, this study only explored the experience of people with dyslexia on social media, and did not investigate tools to improve their experience. The present study aims to extend this work by building and evaluating a writing support tool designed specifically to support writing on social media by people with dyslexia. We aim to help promote feelings of confidence and ease of writing on social media by alleviating some of the technical challenges of writing.

Dyslexia-specific writing support tools

While there are many general-use writing support tools available, such as spell check, most of them were not designed for the dyslexia community. As a result, they do not account for patterns of errors that are more common among people with dyslexia. For example, many of the popular general-use spell and grammar checkers are weakest at detecting "real-word" errors [41] (e.g., *form* instead of *from*), which comprise 17% of errors made by people with dyslexia in English [44].

There have been efforts to create dyslexia-specific spell checkers [32, 41, 44]. Pedler [40] identified sets of words likely to be confused, and enhanced an online dictionary to better detect real-word errors. *Real Check* leveraged a probabilistic language model, a statistical dependency parser and Google n-grams to detect real-word errors in Spanish [44].

However, one limitation of these dyslexia-specific spell checkers is that they were designed with academic or professional writing in mind, and may not be appropriate for the more casual writing style often found on social media. Furthermore, these systems were tested in experimental or controlled settings, and not over time and during natural use. The current study seeks to extend the research in this space by developing a dyslexia-specific spell checker to support casual writing. We also tested the tool over time and during natural use, exploring the ways in which AI technology can help improve confidence in writing on social media.

Machine learning for spell/grammar checking

There is much NLP research for automated spell/grammar checking [34, 35]. Most popular systems are rule [6, 23, 39] or statistical language model based [18, 26, 38]. Rule-based systems rely on manually-created grammar rules, and are constrained on completeness and adaptability to new grammar and word usage [36]. They are also expensive to adapt for a new language, since the rules need to be re-evaluated or re-generated. Statistical language model systems leverage large text corpora to detect and correct spelling/grammar errors based how often the original and the corrected phrase/sentence occur in the training corpus. It is unsupervised and easily portable to other languages as long as sizable training corpora exist. However, a statistical language model can be biased towards the writing style of the training corpus, and can over-fire for low-frequency words/phrases.

Both systems are limited by the types of corrections they make. Rule-based systems only make changes within the range of available rules; statistical language model based systems only make local lexical changes such as word replacement and the addition/removal of small functional words.

NMT-based seq2seq models have potential in overcoming these limitations [12, 28, 56]. Recent work demonstrated that the NMT approach could achieve above-human level performance in grammar correction [19]. However, current models are mostly trained and tested on well-formed, essay-style sentences, with few or no misspellings. It is unclear how they perform over noisy data with errors at lexical, punctuation, and sentence structural levels - as seen in dyslexia style writing on SNSs. We pioneered in this challenging problem space by proposing two new techniques for NMT-based spell/grammar checking: (1) character-level encoding to handle noisy input; (2) data augmentation to generate synthetic training data. However, our goal is *not* to establish the superiority of NMT models. We aim to understand their effectiveness on end-user experience and their potential to assist the dyslexia community, rather than evaluating performance metrics on standard benchmarking datasets alone.

3 SYSTEM

Collecting dyslexia style writing on social media

Although dyslexia covers a spectrum of conditions with different symptoms, previous research showed differences between the writing mistakes made by people with and without dyslexia [13, 41, 42]. In practice, researchers and educators have observed patterns in the writing of students with dyslexia, such as confusion between similar sounding words, contraction of several words into one, run-on sentences, and misuse of capitalization [3, 20]. To better understand common writing mistakes experienced by people with dyslexia on SNSs, and to evaluate the performance of our AI model on correcting those mistakes, we need to collect social media style writing samples with dyslexic errors.

However, existing datasets of dyslexia style writing are scarce. For English, the only publicly shared one was manually collected and annotated by Pedler [41], from students with dyslexia with writing for school assignments. This is limited for our use, because of: (1) *scale*: 673 sentences and 835 corrections in total; (2) *type of errors*: all annotated errors and corresponding corrections are at word/lexical level, with 97.8% of them being real-word errors; (3) *writing style*: the text does not reflect the linguistic style on SNSs.

To build a corpus of dyslexia style writing on SNSs, we took a similar approach as [1], using misspellings that are common to people with dyslexia to locate dyslexia style text on SNSs. Although [1] shared misspellings, most are long (avg. 9 characters/word) and infrequent on SNSs. We extended their list by mining for dyslexia-specific misspellings:

(1) Manually identifying 60 dyslexia related groups on Facebook by searching for groups with the keywords "dyslexia" or "dyslexic" in their names or descriptions.
(2) Generating a dictionary of word frequencies[1] by aggregating over all English text posts made by US-based members of the selected groups in Nov. 2016. All data was de-identified and analyzed in aggregate such that no individual's text was viewed by the researchers;
(3) Creating a *Facebook vocabulary* with tokens that appeared $\geq$10 times [2] in a random sample of 1 billion English text posts by US Facebook users in Nov. 2016. All data was de-identified and processed in aggregate.
(4) Filtering the dictionary created in 2 through the *Facebook vocabulary* to find the top 1000 most frequent words used by group members but were not in the *Facebook vocabulary*.
(5) Manually inspecting the 1000 words retrieved from previous step, filtering out acronyms and proper names, and keeping the rest as misspellings related to dyslexia.

Our approach utilized public text on Facebook to capture social media slang, acronyms, and stylistic misspellings, as well as common typos and misspellings by the general population. This gave us confidence that the misspellings we identified are relatively "unique" to people who participate in dyslexia related groups on Facebook[3]. Some examples of misspellings include: *allegrys*, *dyselxia*, *accdents* (A full list of will be shared with the publication of this paper). While our approach might not have a high recall on all misspellings made by people with dyslexia, this was appropriate for our use case since our intent was not to identify people with dyslexia but to collect a sizable sample of social media posts with dyslexia style writing mistakes.

Instead of directly finding text on Facebook containing dyslexia specific misspellings, we searched de-identified status posts to find users who used those misspellings more than 40 times[4]. We then took a de-identified random sample

[1] Restricted to tokens < 20 characters with only English letters
[2] Other thresholds such as 5 and 20 gave largely consistent results.
[3] Since groups may be used by parents/teachers, we avoided using membership as label for dyslexia, or text in groups as dyslexia-style writing
[4] Other thresholds such as 30 and 50 gave largely consistent results.

of 20K public English posts, 10 to 250 characters long, made by those users as a sample of dyslexia style text. In this way, we obtained a sample with a wider range of dyslexia style spelling and writing than the misspellings identified above.

Understanding dyslexia writing dataset

Prior work analyzing writing mistakes made by people with dyslexia [41, 42] was done with specific writing tasks (e.g., homework assignments) or types of writing errors (e.g., spelling). To better understand dyslexia writing challenges in the social media context, we designed and deployed the following human review and labeling tasks.

An annotation project manager (PM) was hired to source and train 34 US-based annotators through a third-party crowdsourcing company, requiring native English speakers with degrees in humanities. We designed two tasks for them, and no identifiable data was shared with annotators for either task[5]. During annotation, the PM reviewed 1% of each 1K labeled results and gave feedback to the annotators.

Task I: Sentence Rewrite. This task aimed to fix issues in sampled dyslexia style posts to produce a ground truth dataset for model evaluation. We asked annotators to rewrite a given post, fixing any writing issues, while preserving the meaning, style, and structure as much as possible. All annotators needed to pass an assessment by rewriting 10 posts.

To reduce individual annotator bias, we divided them into *rewriters* and *editors*, with rewriters rewriting sampled posts, and editors reviewing the original and rewritten versions and revising again if necessary. 22 annotators worked as rewriters and 14 as editors (2 dual-role). Both rewriters and editors could tag posts as incoherent, non-English, or containing symbols/emoji. We asked annotators to respect the social media writing style and leave Internet slang and memes. When a post was difficult to understand, we asked annotators to make their best efforts to correct spelling and sentence level problems without making assumptions about the meaning/intention of the post. They also completed a questionnaire after each batch (100 posts) to give us feedback.

It took ~90 seconds to process a post (~60 seconds rewriting, ~30 seconds editing). 85% of the posts in the 20K sample were changed from the original. 1% of posts were tagged as non-English, 3% as incoherent, and 5% as containing symbols.

52% of the rewritten posts were further edited. The relatively low agreement rate between rewriters and editors can be explained by the following challenges, as reflected in questionnaire responses and our examination of the results:

- Writing/editing is a subjective activity, making it difficult to agree on the "best" way of wording a sentence;
- Slang and memes are domain-specific and often requires proficiency in Internet culture. We instructed annotators to use websites such as urbandictionary.com and knowyourmeme.com when they were unsure about a term, but still could not guarantee that all slang and memes were preserved in the final version;
- Writing style is subject to individual preference. For example, some consider a sentence with all lower-case letters or no punctuation a stylistic choice, but others do not. The same goes for word choices, such as "havin" ("having"), "u" ("you") and "4" ("for"). While we want to preserve the authenticity of the voice and avoid over-correction, it is challenging since previous research showed that people with dyslexia have trouble with punctuation, capitalization and small words [3].

Such challenges occur in other text correction tasks like grammatical error detection (GED) [8, 11], and the conventional solution is to use more annotators. Two annotators per sentence is a common setup [37]. Despite these challenges, we were able to collect dyslexia style writing on SNSs with relatively high quality corrections that is almost 30 times bigger than existing dyslexia corpora [41]. The following example shows how a post was corrected.

Original: what poeple doin to day
Rewritten: what are people doing today?
Edited: What are people doing today?

Task II: Error Labeling. To understand the distribution of issues in dyslexia style writing, annotators tagged correction type, for a sub-sample of 2300 posts modified in task I.

We provided the annotators with a fixed set of tags, covering spelling, punctuation, capitalization, and grammar. In each category, there were sub-tags for specific issues, such as "spelling - omitting letters" or "spelling - inserting letters", informed by previous research on dyslexia writing errors [41].

The distribution of issues was: 62% punctuation, 36% spelling, 35% capitalization, and 21% sentence structure. A more detailed break down can be found in table 1.

Model architecture

Our goal is to detect dyslexia style writing errors and propose appropriate corrections. Given a sequence of words s which may contain errors (e.g., "She went to watch speiderman"), we want to produce a new sequence s' in which errors are corrected (e.g., "She went to watch Spiderman"). This is akin to a *translation* task where the source language is dyslexia

[5]Identifiable text like @name, full name, or phone number was removed for annotation through regular expression pattern matching and review by the PM. We didn't collect post author information in training data. We also programmatically replaced @name with a general @MENTION token and ensured it wasn't modified in data augmentation, so the model is trained to not alter names in @name form.

	Type of corrections made	%
Spelling	Spacing and apostrophes	37
	Omitting letters	33
	Confusing homophones or similar soundings words	19
	Inserting letters	18
	Swapping order of letters/syllables	9
	Confusing letters visually	5
	Foreshortening words	4
	Similar looking words spelled with similar letters	3
	Omitting or misusing prefixes or suffixes	3
	Misusing small and common words	2
	Other	34
Punctuation	Missing ending punctuation	51
	Missing comma	38
	Extra or missing spaces without creating spelling error	15
	Other	17
Capitalization	Improperly over-capitalized in the middle of a sentence	10
Grammar	Sentence structure	48
	Misuse of verb tense or missing verb	18
	Misuse or missing articles	10
	Misuse of singular or plural	8
	Misuse or missing preposition	4

Table 1: Corrections made to sampled posts. Posts can contain multiple corrections.

style English and the target language is non-dyslexia style English, while preserving the sentence's meaning.

Neural Machine Translation using seq2seq models has achieved state-of-the-art translation quality[2, 50]. It is trained on parallel sentences, where each example contains a source sentence $x_1, ..., x_n$ and target sentence $y_1, ..., y_m$.

Our model consists of two-layer bi-directional LSTM for the encoder with hidden layer sizes of 256 and 512 units. For the decoder, we used a two-layer bi-directional LSTM with hidden layer sizes of 512 units each. Dropout rate is 0.2 throughout all layers. We used a special type of LSTM cell that does multiplicative integration [55]. The embedding size for the encoder and decoder is 128 and 256 dimensions, respectively. The encoder computes hidden states $h_1, ...h_n$ to represent the semantic meaning of the source sentence, while the decoder produces the translated version by scoring the most likely tokens y_t at each time step until an "End of Sentence" token is generated. The model was trained using SGD with learning rate starting at 0.5 and a decay factor of 0.95 after 40K batches (batch size=256). During training, we used a 2K leave-out sample as a validation set (for early-stopping and hyper-parameter selection) and a 2K leave-out dev-test set as a model selection set. The model is trained with 20M posts with programmatically injected errors (see *Data Augmentation*), and tested using 2 datasets described in *Model Evaluation.*

Character-level encoding. A classical NMT model encodes source and target sentences with a fixed-size vocabulary, capped by computational and memory constraints. Thus, NMT models struggle with rare or out of vocabulary (OOVs) words [33, 56]. OOVs are often treated as special *UNK* tokens, and must be handled in an ad-hoc post-processing step. However, this is not a robust solution to our problem.

Most spelling errors made by people with dyslexia are intra-word errors (see Table 1), where characters are flipped, omitted, or added at incorrect places. This makes it difficult for a neural encoder to represent a sentence given that OOVs are transformed into *UNK* tokens. For example, the sentence "**She went to watch speiderman**" would be represented internally as "**She went to watch UNK**" making it impossible for the model to correct the *UNK* token.

Our solution is to use character-level encoders[6]. For example, we can represent the previous sentence as: "**S h e EOW w e n t EOW t o EOW w a t c h EOW s p e i d e r m a n EOW**", where *EOW* is an end of word marker. This allows us to learn uncommon misspellings and fix them, but increases model latency, due to a dramatic increase in sequence length. To reduce sequence length, we only use the character model in the encoder and use the word model in the decoder. This works for our application since we only need to handle *UNK* tokens in the source and decode them using in-vocabulary words.

Data Augmentation. Data augmentation is a technique to generate large scale parallel data by pairing monolingual training with an automatic back-translation [48], and has been widely used for translating low-resource languages, where labeled parallel data is scarce [17]. We combined data augmentation with data noising to generate synthetic data for training. For each sentence x from the reference corpus, we applied different noising augmentations to generate its perturbed version x' containing dyslexia style writing errors. This way, we were able to generate a large parallel corpus for training the seq2seq model.

Based on the distribution of errors in table 1, we injected one of following errors in 20% of words[7] in a random sample of 20M US English 10-250 character posts on Facebook:

(1) Letter confusion: substitute similar-looking letters (e.g., *b* v.s *p*). e.g., *"My best friend!"* became *"My pest friend"*;
(2) Homophone: replace a word with its homophones. e.g., *"Here we go"* became *"Hear we go"*;
(3) Confusion set: using frequently confused words from our labeled dataset (e.g., "your", "you're", "you"), randomly replace a word with another word in its confusion set based on the frequency of confusion. e.g., *"You're welcome!"* became *"Your welcome!"*.

All generated training data were de-identified and processed in aggregate by the model.

[6]Our character-level encoders use a CNN with convolution filters with widths 1, 2, and 3 and max pooling.

[7]The error rate was informed by [41]. We also noticed the model performed worse when error rate is above 40%.

	Raw sentences	*Aligned sentences*				
Original	im going to.	im	going	[X]	to	.
Human	I'm going, too.	**I'm**	going	**,**	**too**	.
Model	Im going too.	**Im**	going	[X]	**too**	.

Table 2: How we aligned original, human corrected, model corrected sentences to locate changes (bold). The model made one correct change (*to*->*too*), one incorrect one (*im*->*Im*), and missed one (adding comma). The precision is 50% and the recall is 30%.

Model evaluation

Datasets. We used the following to measure the performance of the seq2seq model.

(1) *FB sample*: 1K sample of human annotated public Facebook posts with dyslexia style writing errors, containing 9163 corrections made by annotators; the distribution of corrections is largely consistent with Table 1.
(2) *Public sample*: dyslexia dataset shared in [41].

Metrics. We measured token-level precision and recall. *Precision* is the percentage of corrections made by the model also made by the annotators/researchers; *recall* is the percentage of corrections made by annotators also made by the model.

To compare changes made by annotators and the model at token-level, we need to programmatically align the original sentence with both human and model corrected versions to know which token was changed in what way (see Table 2). To get token-level alignment, we tokenized the sentence into a sequence of word/punctuation tokens (instead of characters), and implemented a version of the Needleman-Wunsch algorithm using Levenshtein edit distance matrix [29]. We only allowed three types of edits when computing the edit distance between sequences: *add*, *delete*, and *substitute*.

Results. We calculated the precision and recall of our model and benchmarked against Microsoft Word (widely used by people with dyslexia for spell checking [46]). To get corrections made by Microsoft Word (Office 2018 on Windows), annotators copy-and-pasted the text from *FB sample* into Word, right-clicked every place that was underlined, and picked the top suggestion provided. Similar benchmarking data for the *public sample* was collected and shared by a third party tech blogger [51], which we used for our evaluation[8].

As shown in Table 3 *c.s.* columns, in many cases our model's performance is close to or better than those we benchmarked against (although Word 2018 outperformed the model on the FB dataset at times). After analysis of the false positives (i.e., incorrect changes the model made) and false negatives (i.e., human corrections that the model missed), a

[8]Our numbers are different from what was shown in [51] because they did not count false positives (i.e. changes to correct text).

Sample	*Model*	***P*** (%)		***R*** (%)	
		c.s.	c.i.	c.s.	c.i.
FB	seq2seq	21.1	47.5	12.1	7.7
	Word 2018 - (Win)	31.2	28.2	15.3	14.7
Public	seq2seq	37.2	65.9	21.3	21.5
	Word 2007 - (Win)	39.1	42.5	30.5	30.8
	Word 2008 - (Mac)	14.3	15.5	10.6	10.6
	MacOS 10.6.1 checker	27.1	34.0	7.8	8.1

Table 3: Model precision (P) and recall (R) on different datasets, with or without ignoring casing. Columns *c.s.* show the performance when case sensitive, and columns *c.i.* when case insensitive.

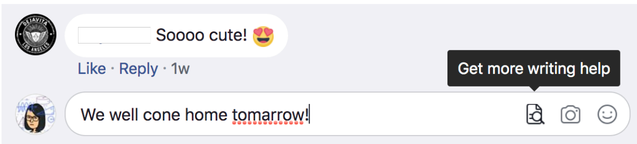
large proportion of errors involved capitalization and punctuation. In particular, the model often normalized the capitalization and lower-cased sentences even when they are in all caps. It also had difficulty figuring out whether and where to add/remove punctuation, and how to correct the irregular use of punctuation (e.g., changing ". . ." to "..."). In fact, when we ignored capitalization issues, our model out-performed popular systems in its precision (see Table 3 *c.i.* columns).

4 DESIGNING A DYSLEXIA SUPPORT TOOL FOR COMPOSING COMMENTS ON FACEBOOK

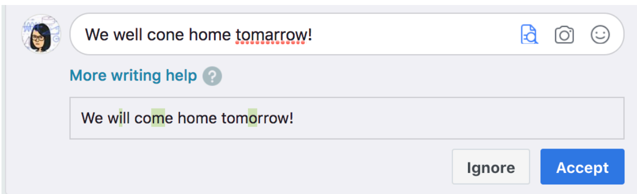
To deploy and test our model for social media for our study participants, we designed and prototyped the "Additional Writing Help" (AWH) tool for composing comments on the web version of Facebook (see Figure 1). The goal of AWH is to boost confidence for people with dyslexia when writing on Facebook. We started with comments because we believe that commenting is an important form of social exchange that often relies on text. Also, by commenting, a person would expose her writing to anyone who can see the original post, which implicates notions of self-representation documented in previous dyslexia research [46].

We took several design considerations into account:

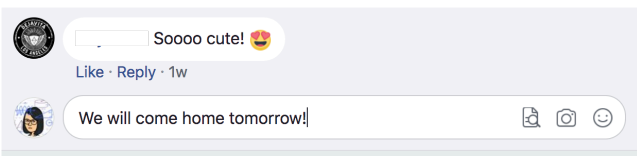
(1) Control: To ensure control over whether and when to use this feature, we do not flag misspellings while typing and the feature would only be invoked by clicking its icon. We will not analyze the text unless the user requests to do so. We will not modify the original text unless the user explicitly accepts the changes we suggested. Overall, we want to respect individual writing style and not interrupt the composing flow.
(2) Self-representation needs: People with dyslexia have concerns about writing mistakes being archived and visible in the "edit history", as it undermines efforts to correct errors for a more positive self-image [46]. We designed AWH to provide suggestions **before** the comment is posted, reducing further edits.



(a) A user can trigger AWH by clicking the "Get more writing help" icon in the comment composing box under a Facebook post



(b) Suggestion shown in the "More writing help" area below the comment composer, with changes highlighted. If the model cannot generate changes, "No suggestion" is shown instead.



(c) Once "Accept" is clicked, the suggested version replaces what was there originally. The comment is *not* posted automatically, so that the comment can be checked and edited before posting.

Figure 1: Additional Writing Help (AWH) design and flow

(3) Features of seq2seq model: Our model translates the entire sequence rather than making local changes. Therefore, inputting an entire sentence helps performance.
(4) Complement existing tools: As most people with dyslexia already use spell/grammar checking tools when writing on social media [46], we want to complement their current writing strategies instead of replacing them. With most existing tools doing real-time corrections, we position AWH as a *proofreading* system that specializes in areas such as homophone confusions.

5 USER EVALUATION

Method

In June 2018, we evaluated AWH with a seven-day diary study and follow-up interviews with a subset of diary study participants. Our goal was to evaluate the tool's effectiveness and its effect on confidence in writing content on Facebook.

Diary Study. We recruited 24 English-speaking American adults (5 male and 19 female), who all self-reported having dyslexia, writing content on Facebook, and using the web version of Facebook at least three times per week. We informed participants that responses would be anonymized and may be published, and they were compensated for their time. Of 24 initial participants, one had difficulty locating AWH during the study and was excluded from analysis. For the remaining 23, 19 of them stayed through the week-long study, and completed the final survey at day seven[9].

The study consisted of three parts: an initial survey and brief introduction to the tool, a seven day diary study with daily surveys, and a longer survey on the final day. The initial survey took about 10 minutes to complete, and asked about learning disability diagnoses, perceptions of writing difficulty and confidence, spell-check tool use, and Facebook posting behavior. Response options were randomized and scales were randomly flipped when appropriate. Participants were then shown a short tutorial explaining how to use the tool. This tutorial was also emailed to all participants so that it could be reviewed at any point during the study.

Participants had access to AWH for one week and were asked to use AWH on at least four of those days. However, they were not obligated to post any comments.

To collect qualitative feedback, we sent a short survey each afternoon, asking participants whether they used AWH and their experience with it that day. We asked them to fill in the survey every day, even if they did not use AWH. On the final day, the survey included not only the questions from the daily surveys, but also questions about their overall opinions of AWH, and asked again about their perceptions of writing difficulty and confidence. For quantitative feedback, we logged certain AWH usage statistics, including the number of times a user triggered the feature, the length of the text input[10], whether our model provided any suggestions for given text, and whether the user accepts or ignores the suggestions, if provided.

Interviews. After the completion of the diary study, we conducted semi-structured interviews with 11 participants. Interview participants were chosen from the diary study participants based on the following three criteria: (1) being a relatively active user of AWH during the diary study; (2) having a change in perceived writing difficulty and confidence between the initial and final diary study surveys; or (3) having a write-in response during the diary study that warranted further discussion.

Each phone interview lasted approximately one hour, and was audio recorded and transcribed for analysis. Participants were compensated for their time.

We began by asking about their overall impressions of AWH, what worked well and what challenges they encountered. We then asked about specific aspects, such as their perceptions of accuracy and some of its mechanics. We also

[9] "After diary study" results only use data from these 19 participants.
[10] For privacy, we did not log text before it was posted on Facebook.

asked about their perceptions of writing difficulty and confidence, and the ways in which AWH and other spell check tools affect those.

We analyzed the transcribed interviews using inductive qualitative methods drawn from grounded theory [9]. Interviews were reviewed and notes were taken about these key ideas. The authors reviewed the notes and discussed key ideas, one author then coded all transcripts, then reconvened to review and discuss the coded material. We used affinity diagrams to organize ideas into the themes we discuss below.

Results

Use of AWH

Log data showed that the 24 participants used AWH 165 times over the course of the study (defined as clicking the "Get more writing help" button in the comment box). Four participants used it exactly once, and eventually dropped out of the study. Half of the remaining participants (10) used it more than 9 times, and the top person used it 20 times.

However, one could click the button without requesting a suggestion (e.g., nothing was typed in the composer). According to log data, suggestions were requested 152 times, and 72% of the times (110/152) we returned one; otherwise we returned "No suggestion." This is largely consistent with self-report data (75%) on how often a suggestion was returned.

When a participant reported seeing a suggestion in the daily survey, we asked how helpful it was. 18% of the time, suggestions were rated as *"Extremely helpful"*, 44% *"Very helpful"*, 18% *"Somewhat helpful"*, 10% *"A little helpful"*, and 10% *"Not helpful at all."* And as a result, 91% of suggestions were accepted. Participants echoed this in the follow-up interviews as well; for example, that they *"worked well to check my spelling and the capitalization"* (P9). One interesting observation is that participants modified only 25% of suggestions – we expected that to be higher.

Overall experience

In the final survey, we asked additional questions about the overall experience/perception of AWH. As shown in Figure 2, of the 19 people who completed the study, 13 reported being somewhat or very satisfied, 2 reported being somewhat or very dissatisfied, and 4 reported neutral satisfaction.

Interview study participants supported this result. Nine interview participants had a positive experience, reporting: *"[AWH] just made it a positive experience of commenting, to make sure that it was right, because everybody wants to post correct grammar on a post or comment"* (P7) and *"I think it would be a great added benefit to Facebook users"* (P1).

In fact, a few were surprised by how positive their experience was. For example, P3 stated: *"About midway through I went 'Okay, I can see the benefits. Yeah, I got it' because as*

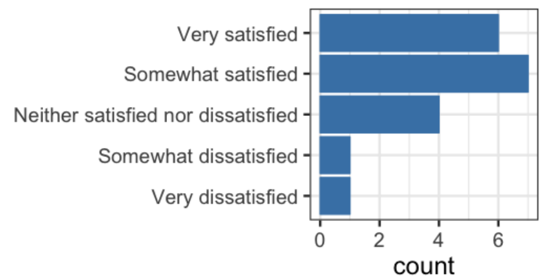


Figure 2: Responses to the question "Overall, how satisfied or dissatisfied were you with AWH?"

I tried to do a post, there was something that I missed." P1 explained *"I actually really enjoyed using the tool, and I didn't think I was going to."* Her low expectations were a result of previous bad experiences with assistive technologies: *"a lot of times being in the disabilities community, they're always coming up with new, shiny things to make everybody's life better, but lots of times those shiny things are just kind of crappy at the end of the day."* However, two interview participants reported that AWH was not novel enough: *"[I wish] it would have offered something more...it really didn't offer me anything that made me go, 'Oh boy!'"* (P5).

Interest in future use. In the final survey, we asked about interest in using AWH in the future. Of the 18 participants who answered the question, nine were *"Extremely interested"* to use it in the future, seven were "Somewhat interested", one was "A little interested", and one was "Not at all interested". Several interview participants lamented that they no longer had access to AWH: *"I'm very sad to see it go and I'm definitely hoping Facebook brings this to everyone"* (P1).

Perceived accuracy. Performance and precision are critical to spelling support, and we sought to understand how they were perceived in AWH. We asked diary study participants about AWH's precision, and it was rated favorably: 12 of 17 participants who answered this question reported it as 80% or higher, two between 60-70%, and three between 30-50%. Interview participants confirmed this and also provided context about when AWH was inaccurate. Many of the issues mentioned were noted in our model evaluation as well, but hearing their experience helped us understand the severity of different errors and their impact on user experience.

Several noted that AWH often failed to accurately provide suggestions when comments had slang or abbreviations, such as regional words: *"I'm from Florida, so I use words like ain't and y'all...And it wants to change it to 'you all.' What good is that?"* (P5), and *"It didn't like LOL a lot; it kept trying to find lots of other words, like 'look' or 'loop' or something like that"* (P1). However, P2 noted that if she was using slang, she would just not click "Get more writing help."

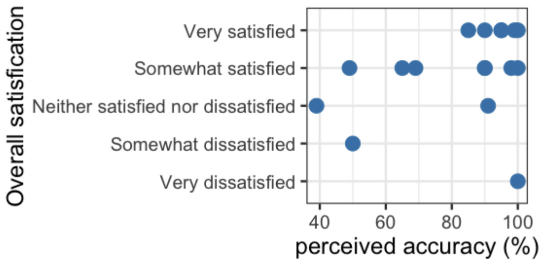


Figure 3: Relationship between perceived accuracy (x-axis) and overall satisfaction level with AWH (y-axis).

AWH also struggled with numbers: *"I had posted recipes online, and I noticed several times it had changed significantly, like a quarter of a cup to 114 cups...as I reread it, it really looked like something that somebody drunk would have written"* (P8). This is a known issue with the current model, which does not guarantee preserving the order of numbers and entities (e.g., names, places) in the "translation", although could avoid changing the numbers and entities themselves.

Perceived accuracy and satisfaction. Surprisingly, participants' perceptions of accuracy were only very weakly correlated to their satisfaction with the tool. As shown in Figure 3, while there is a cluster of people who perceived high accuracy and reported high satisfaction with AWH, those who felt "Somewhat satisfied" or neutral with the tool varied in their perceptions of accuracy.

The interviews provided insight on this. The design features helped manage against accuracy concerns because accepting suggestions before words were changed provided an opportunity to review changes, rather than it happening automatically. For example, P9 explained *"autocorrect sometimes doesn't give the word that you're looking for...It'll give you those weird words, and then you just assume it's the right word when it's way not."* P8 added: *"I liked that it gave me the option of seeing the ways I could change it, but it didn't force me to change it."* In this case, the extra step of clicking a button to provide a writing suggestion was not a hindrance, but rather a way to give more control to the writers.

Perceived accuracy and AI. Five interview participants believed (erroneously) that AWH would learn their spelling patterns over time and provide more accurate suggestions. For example, P2 stated, *"[AWH] got into learning a little bit, like wow, how did it know that thought was what I was trying to say?"* This affected their perceptions of accuracy, because they expected that it would become more accurate: *"I only had five days with this. If I had more time, it might become more helpful and more automatic"* (P10) and *"But knowing that it's machine learning, I think over time, it would correct itself"* (P6). In fact, a few thought AWH had adapted over the course of the study. P1 said *"I felt like after a while it kind of caught on to what I was doing; this is going to sound silly, but it kind of got smarter as it was going along."*

This provided an opportunity to explore thoughts regarding adaptive writing support tools. They highlighted the potential to be both transformative and invasive. Some were excited by the prospect of more useful support: *"I think it should [learn habits]. Because for people like myself and my son and my daughter, who have multiple learning disabilities, it's going to be a really good thing"* (P1) and *" If AI works the way I think it does, I mean it'll be significantly better [than current writing support tools]"* (P3). Despite the excitement, there were concerns about privacy implications. P1 wondered, *"Are you guys kind of really watching what I'm saying or doing? How closely are you guys paying attention to me?"*

Incorporating AI for provide more personalized and context-dependent writing suggestions is an intriguing and promising direction for the design of writing support tools, but it's important to be thoughtful about the data sensitivity and social implications around training an individualized model.

Confidence. We asked participants about their confidence and perceived difficulty in writing before and after the study, and saw a shift toward "more confident" (Figure 4) and "easier" (Figure 5). Of the 18 who answered these questions before and after the study, seven reported an increase in their writing confidence while eight reported the same; and 13 people reported writing to be easier while 4 reported the same. This was a key topic in the follow-up interviews.

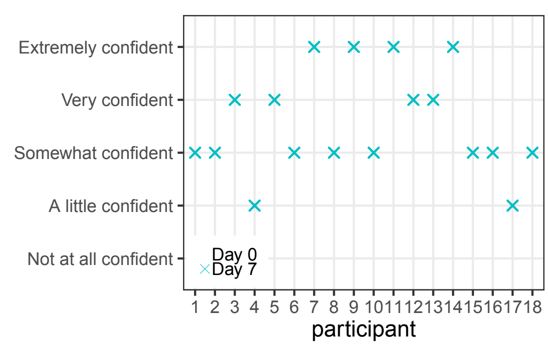


Figure 4: Participants' responses to "How confident are you in your ability to write clearly?", before and after field study

Overall, participants described AWH as a *"safety net"* which helped bolster confidence. For example, P2 explained that *"I just felt good to know this safety net was behind me"* and P6 reinforced this: *"Just that reassurance is good, even though I can pretty much do everything I want to say correctly. But having that gives me a little bit more confidence."*

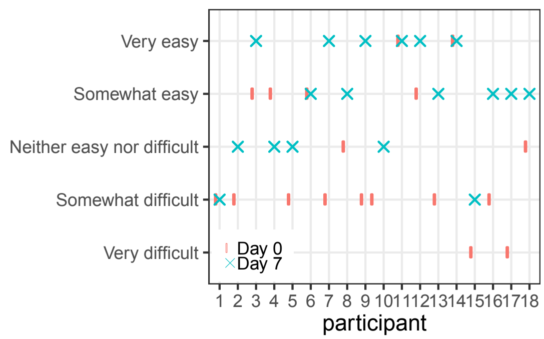


Figure 5: Participants' responses to "How easy or difficult is writing for you, in general?", before and after field study

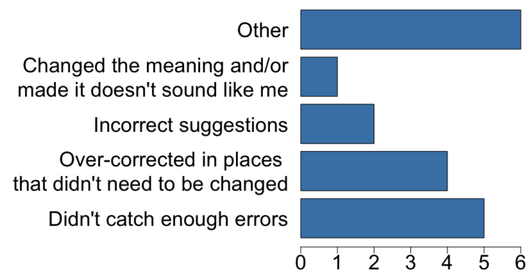


Figure 6: Distribution of responses to "Overall, what was the TOP challenge with using AWH?"

There are several reasons why AWH helped promote confidence. *"It tells you if you spell it wrong. And it gives you just a little more confidence to post something so you can express the way you feel without having the fear of not looking as intelligent"* (P7). P1 said that her confidence increased because she *"felt like [her writing] was not too garbled; it was coming across as a more clear and concise sentence."* Furthermore, it was discreet: *"I think the technology - no one knows that I'm asking somebody for help. I'm just doing it."* (P9). Essentially, AWH allowed participants to overcome anxieties about posting on Facebook, which was documented previously [46].

Confidence and writing tools. Unfortunately, confidence did not last after the participants lost access to AWH. *"Once it, sometimes discussions have come up and I'm like, 'Gee, I don't know if I'm going to able to convey this,' so instead of saying something I don't say anything, and that's just been kind of a bummer"* (P1) and *"I guess [my confidence] kind of went away when the tool went away. But I guess when I answered [the survey], I was thinking, oh, that's going to be there forever. Then it went away"* (P9).

AWH wasn't the only writing support tool that our participants relied on, and we learned that their writing confidence was tied to those tools as well. P4 depended on having access to Grammar.ly: "My confidence would absolutely go with it."P2 and P6 relied on built-in spell check tools on their phones and computers *"So if spell check suddenly disappeared I would have bad anxiety. That is a nightmare."* (P2) and *"Man, my life would be so much harder [without spell check]."* (P6)

Building tools for building confidence. As we aim to bolster writing confidence in people with dyslexia, we were curious to learn how a writing tool could provide a lasting effect.

Some participants highlighted aspects of AWH that could improve their confidence and writing ability: P1 felt that more time with AWH might lead to more benefits: *"I have contributed more [since the diary study], but not as much as I did when I was using the tool. I mean for me it takes more than a week – it just takes me longer to get there."* Others thought that having to accept AWH's suggestions promotes learning: *"I also really liked that it was like an accept or reject and it didn't change it for you once you clicked that button. So not only did it help you to correct but it also will make people more aware of the mistakes they're making."* (P2).

A couple of participants also had suggestions for longer-term writing support. One involved using AI to provide personalized writing suggestions: *"You would have to know my habits, what words I don't do well on. Then you're gonna try to teach me because that's the thing that spelling does that spellchecks don't do right now. They don't teach. They just fix it for you. They're enablers."* (P3) Another involved promoting different writing styles based on context: *"If you're using a more relaxed voice with your friends, and you don't really care so much about some of these adjustments, you can make those evaluations and understand in what context you should be more cautious about what you say and how you say it"* (P8).

However, four participants questioned whether designing for longer-term writing support was a useful or worthwhile goal. They felt it was unrealistic, and that a well executed in-the-moment writing support tool was more beneficial. P1 explained that *"Spelling independently really isn't a long-term option for me. I'm 43 now, and this is probably about as good as it's going to get."* P10 explained that a longer-term writing support tool wouldn't be practical for her daily use: *"You just need something to tell you what the right answer is."* However, while P4 didn't want a tool like this for herself, she saw benefits for her son, who also has dyslexia: *"For myself, I'm fine with having a crutch. If it's for my son, I'd rather he have building blocks and training tools to help him go along. I want more for him, and I think if it's a building block it gives you more of a capability to independent."*

Top challenges with AWH. In the final survey, we asked participants about the top challenge they experienced with AWH. The top (5/18) concern is low recall, and the second concern (4/18) is over-correction (high false positive) (Figure 6).

Six people indicated that their top challenge was "Other", and their issues mostly concentrated on *More product surfaces and platforms* (e.g.,*"not on my phone to use", "I noticed the tool was only available in certain aspects of Facebook"*) and *Interaction design* (e.g., *"Interfere with Grammarly", "I thought that it posted the comments when I accepted the corrections, but it took an extra step to actually post the comment...it should be just a one-click accept and post."*)

We asked about how best to improve AWH in the interviews. The results were consistent with the survey findings - participants suggested improvements related to usability (e.g., more visible entry points, multiple edit suggestions returned) the model (e.g., better corrections for slang and abbreviations, corrections for other languages), and the availability of the feature (e.g., on mobile and for more than comments).

6 DISCUSSION

We evaluated the effectiveness of AWH and found that many had a positive experience. It also provides us an opportunity to reflect on larger issues regarding the development of writing support tools, including: the impact of design choices, the role of confidence, and the potential of AI.

Reflections on design

We made several design choices in building AWH that affected participants' experiences. We chose to (1) not automatically correct language and spelling, but to instead only offer suggestions after "Get Writing Help" was pressed and (2) to require an additional step in order for the suggested edits to take effect.

Many interview participants compared AWH to tools like autocorrect, and highlighted how these differences impacted their perceptions. Specifically, the AWH "pull" model AWH gave participants greater control than autocorrect's "push" model . This seems counterintuitive, as the "pull" model takes more time and effort, but participants appreciated that it gave them the choice both when and whether to seek writing help. The "pull" model also helped increase confidence in AWH's ability to suggest appropriate grammar changes, because AWH would check the entirety of the text at once. Although tools like autocorrect can also make some grammar suggestions, its design of incremental changes *appeared* to be less well suited to detect grammar and sentence structure.

This can inform writing support tool design considerations because they run counter to conventional wisdom. Instead of valuing speed and automation, we suggest instead valuing giving people the control to decide when and whether to edit their writing. This is especially important for social media, as it provides an opportunity to promote a positive self-presentation [5], and therefore writing support tools should not only correct language, but also ensure that people feel that they have control over how they present themselves.

Tool dependence vs self-confidence

One goal for AWH was to improve confidence about and ease of writing. The diary study indicates early promise that AWH may improve writing confidence and perceived ease of writing, and interview participants discussed how writing support tools (including AWH) affect their writing confidence. However, any confidence gains results from access to the tool, and when the tool is removed, confidence goes away with it.

We wondered about design opportunities to develop a writing support tool to help people foster more confidence in themselves. When we raised this with interview participants, we learned that providing opportunities to learn through raising suggestions and highlighting errors rather than just correcting mistakes may be a promising avenue. Currently, AWH provides suggestions and requires them to be accepted before they take effect, allowing the opportunity to review the difference between the original text and the suggested edits. Yet future designs should consider other ways in which teaching can occur within an everyday writing support tool.

However, a few participants questioned this goal, and didn't expect that a tool could or should provide lasting benefits. An analogy would be that some who are near-sighted want the lasting benefits of LASIK surgery, while others prefer the temporary benefits of glasses. This raises the question of whether promoting confidence is always the best goal. While we believe that longer-term support is possible, we also want to avoid imposing our value beliefs about the ideal writing support tool. This is an opportunity for future research to better understand the needs and desires of this community, and to challenge the design community to consider all relevant value beliefs. Furthermore, we should consider people with dyslexia not as a monolithic group, but also consider the importance of other factors which may impact needs and wants from assistive technology (e.g., age, severity of dyslexia, level of education).

AI-powered adaptive writing support tool

Some participants believed that AWH would leverage machine learning/AI to adapt and evolve, to learn and accommodate their unique patterns of spelling errors and writing style (e.g., slang, deliberate capitalization, etc.). While our current model does not have this capability, it points to a promising direction.

One benefit of such a tool is the potential to be more universal – instead of being marketed to people with dyslexia, it can support writing challenges faced by others, such as people using a second language or who rely on voice input. By covering a broader population, we can destigmatize the use of assistive writing tools and improve writing for all.

However, for the model to learn writing style, it needs to be trained with personal data, which surfaced the heavily discussed tension between personalization and data privacy [30]. We want to be especially careful about the design and use of a personalization algorithm, offering people control over what type of data is used for training the model.

7 LIMITATIONS AND FUTURE WORK

There are a few limitations of our work. In terms of the model, our data collection is limited by assumptions we made about writing by people with dyslexia. In the future, it would be beneficial to have people with dyslexia donate their writing samples with fixes made by someone they know and trust.

In terms of the diary study, all of our participants were US-based English speakers, and the AWH tool was only available for comments on posts (as opposed to the many other places on Facebook in which a person can write text). In the future, this work can expand to other populations within the dyslexia community. Furthermore, given the small sample size, we were unable to perform statistical analysis on survey responses.

Dyslexia affects people differently in terms of the way it manifests and its severity. AWH is not designed to address all possible manifestations of dyslexia, but rather a first attempt to address only the challenges in writing on social media that were surfaced in previous research [46].

The diary study helped us understand the impact of different types of model errors and informed how to better prioritize the model improvement work and what to aim for (e.g., precision vs recall). Furthermore, the potential of utilizing machine learning/AI in order to create a personalized and adaptive writing support tool seems like a promising area for future research and innovation.

There is also room for our NMT model to improve. For example, we will annotate numbers with its order in the source sentence to guarantee the order information was passed and preserved in the translation result. We will also increase the vocabulary size to contain words in different capitalization forms, so that the model will learn the semantic of all caps form. NMT models can make broader changes (e.g., adding/removing words, switching word orders), obscuring individual changes between input and output sentences. We could leverage the attention matrix output by the decoder to align target token(s) with corresponding source token(s) and allow the users to accept individual changes instead of the whole sentence.

8 CONCLUSION

We designed and evaluated a new writing support tool (AWH) for people with dyslexia on Facebook. Our tool was powered by an neural machine translation model that is customized for the dyslexia and social media use case. We tested this tool with 19 people with dyslexia on Facebook and observed increased level of confidence and ease with writing on social media. Our field study demonstrated the value of a better writing support tool on social media sites, and the potential of empowering people with dyslexia using AI technologies.

9 ACKNOWLEDGEMENTS

We thank Eye to Eye National for their assistance with recruiting, and Roman Rädle for his assistance with formatting.

REFERENCES

[1] Ricardo Baeza-Yates and Luz Rello. 2011. Estimating Dyslexia in the Web. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '11)*. ACM, New York, NY, USA, Article 8, 4 pages. https://doi.org/10.1145/1969289.1969300

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR* abs/1409.0473 (2014). arXiv:1409.0473 http://arxiv.org/abs/1409.0473

[3] Eileen Bailey. 2018. How Dyslexia Impacts Writing Skills. Retrieved September 17, 2018 from https://www.thoughtco.com/how-dyslexia-impacts-writing-skills-3111195.

[4] Timothy Baldwin, P Cook, M Lui, Andrew MacKinlay, and Li Wang. 2013. How noisy social media text, how diffrnt social media sources? *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)* (01 2013), 356–364.

[5] Natalya Bazarova, Jessie G. Taft, Yoon Hyung Choi, and Dan Cosley. 2013. Managing Impressions and Relationships on Facebook: Self-Presentational and Relational Concerns Revealed Through the Analysis of Language Style. 32 (06 2013), 121–141.

[6] Andrew Bredenkamp, Berthold Crysmann, and Mirela Petrea. 2000. Looking for Errors: A Declarative Formalism for Resource-adaptive Language Checking. In *LREC*.

[7] Nicola Brunswick. 2010. Unimpaired reading development and dyslexia across different languages. *Reading and dyslexia in different orthographies* (2010), 131–154.

[8] Christopher Bryant and Hwee Tou Ng. 2015. How Far are We from Fully Automatic High Quality Grammatical Error Correction?. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 697–707.

[9] Philip Burnard. 1991. A method of analysing interview transcripts in qualitative research. *Nurse education today* 11, 6 (1991), 461–466.

[10] D. Carnine. 2003. IDEA: focusing on improving results for children with disabilities. (2003).

[11] Martin Chodorow, Markus Dickinson, Ross Israel, and Joel Tetreault. 2012. Problems in evaluating grammatical error detection systems. *Proceedings of COLING 2012* (2012), 611–628.

[12] Shamil Chollampatt, Kaveh Taghipour, and Hwee Tou Ng. 2016. Neural network translation models for grammatical error correction. *arXiv preprint arXiv:1606.00189* (2016).

[13] Chris Coleman, Noël Gregg, Lisa McLain, and Leslie W. Bellair. 2009. A Comparison of Spelling Performance Across Young Adults With and Without Dyslexia. *Assessment for Effective Intervention* 34, 2 (2009), 94–105. https://doi.org/10.1177/1534508408318808

[14] Vagner Figueredo de Santana, Rosimeire de Oliveira, Leonelo Dell Anhol Almeida, and Marcia Ito. 2013. Firefixia: An accessibility web browser customization toolbar for people with dyslexia. In *Proc. of W4A*.

[15] Joan Morris Dimicco and David R. Millen. 2007. Identity ManagementâĂŕ: Multiple Presentations of Self in Facebook. (2007).

[16] Jacob Eisenstein. 2013. What to do about bad language on the internet. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Atlanta, Georgia, 359–369. http://www.aclweb.org/anthology/N13-1037

[17] Marzieh Fadaee, Arianna Bisazza, and Christof Monz. 2017. Data Augmentation for Low-Resource Neural Machine Translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL '17)*.

[18] Michael Gamon, Jianfeng Gao, Chris Brockett, Alexander Klementiev, Bill Dolan, Dmitriy Belenko, and Lucy Vanderwende. 2008. Using Contextual Speller Techniques and Language Modeling for ESL Error Correction. https://www.microsoft.com/en-us/research/publication/using-contextual-speller-techniques-and-language-modeling-for-esl-error-correction/

[19] T. Ge, F. Wei, and M. Zhou. 2018. Reaching Human-level Performance in Automatic Grammatical Error Correction: An Empirical Study. *ArXiv e-prints* (July 2018). arXiv:cs.CL/1807.01270

[20] Vincent Goetry and Dyslexia International. 2018. Common mistakes in reading and spelling. Retrieved September 17, 2018 from https://www.dyslexia-international.org/ONL/EN/Course/S2-3-2.htm.

[21] Ervin Goffman. 1959. *The presentation of self in everyday life.* Anchor Books, New York.

[22] Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Social Media Update 2016. (11 November 2016).

[23] George E. Heidorn, Karen Jensen, Lance A. Miller, Roy J. Byrd, and Martin S Chodorow. 1982. The EPISTLE text-critiquing system. *IBM Systems Journal* 21, 3 (1982), 305–326.

[24] Beve Hornsby and M Farrar. 1990. Some effects of a dyslexia centred teaching programme. *Children's Difficulties in Reading, Spelling and Writing* (1990), 173–96.

[25] S. Gunnel Ingesson. 2007. Growing Up with Dyslexia. *School Psychology International* 28, 5 (2007), 574–591. https://doi.org/10.1177/0143034307085659

[26] Aminul Islam and Diana Inkpen. 2009. Real-word Spelling Correction Using Google Web IT 3-grams. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3 (EMNLP '09)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 1241–1249. http://dl.acm.org/citation.cfm?id=1699648.1699670

[27] Juan E Jiménez, Remedios Guzmán, Cristina Rodríguez, and Ceferino Artiles. 2009. Prevalencia de las dificultades específicas de aprendizaje: la dislexia en español. *Anales de psicología* 25, 1 (2009).

[28] Marcin Junczys-Dowmunt, Roman Grundkiewicz, Shubha Guha, and Kenneth Heafield. 2018. Approaching Neural Grammatical Error Correction as a Low-Resource Machine Translation Task. *CoRR* abs/1804.05940 (2018). arXiv:1804.05940 http://arxiv.org/abs/1804.05940

[29] Dan Jurafsky. 2018. CS124 Lecture Notes: Minimum Edit Distance. Retrived September 21, 2018 from https://web.stanford.edu/class/cs124/lec/med.pdf.

[30] Alfred Kobsa. 2007. Privacy-enhanced Personalization. *Commun. ACM* 50, 8 (Aug. 2007), 24–33. https://doi.org/10.1145/1278201.1278202

[31] Peter Korn. 2016. New Accessibility Features for Kindle E-Readers.

[32] Alberto Quattrini Li, Licia Sbattella, and Roberto Tedesco. 2013. Polispell: an adaptive spellchecker and predictor for people with dyslexia. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 302–309.

[33] Minh-Thang Luong, Ilya Sutskever, Quoc V. Le, Oriol Vinyals, and Wojciech Zaremba. 2015. Addressing the Rare Word Problem in Neural Machine Translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Beijing, China, 11–19. http://www.aclweb.org/anthology/P15-1002

[34] James H Martin and Daniel Jurafsky. 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* Pearson/Prentice Hall.

[35] Roger Mitton. 1994. Spellchecking by computer. Retrieved September 17, 2018 from http://www.dcs.bbk.ac.uk/~roger/spellchecking.html.

[36] Daniel Naber. 2003. A rule-based style and grammar checker. (2003).

[37] Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. 2014. The CoNLL-2014 shared task on grammatical error correction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*. 1–14.

[38] Peter Norvig. 2007. How to Write a Spelling Corrector. Retrieved September 17, 2018 from https://norvig.com/spell-correct.html.

[39] Jong C Park, Martha Stone Palmer, and Clay Washburn. 1997. An English Grammar Checker as a Writing Aid for Students of English as a Second Language. In *ANLP*. 24.

[40] Jennifer Pedler. 2001. Computer spellcheckers and dyslexics—A performance survey. *British Journal of Educational Technology* 32, 1 (2001), 23–37.

[41] Jennifer Pedler. 2007. *Computer Correction of Real-word Spelling Errors in Dyslexic Text.* Ph.D. Dissertation. London University, Birkbeck College.

[42] Luz Rello. 2014. *DysWebxia. A Text Accessibility Model for People with Dyslexia.* Ph.D. Dissertation. Universitat Pompeu Fabra, Barcelona, Spain.

[43] Luz Rello, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013. Frequent words improve readability and short words improve understandability for people with dyslexia. In *IFIP Conference on Human-Computer Interaction*. Springer, 203–219.

[44] Luz Rello, Miguel Ballesteros, and Jeffrey P. Bigham. 2015. A Spellchecker for Dyslexia. In *Proc. of ASSETS*. https://doi.org/10.1145/2700648.2809850

[45] Luz Rello, Horacio Saggion, and Ricardo Baeza-Yates. 2014. Keyword highlighting improves comprehension for people with dyslexia. In *Proceedings of PITR*.

[46] Lindsay Reynolds and Shaomei Wu. 2018. "I'm Never Happy with What I Write": Challenges and Strategies of People with Dyslexia on Social Media. In *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018*. 280–289. https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17812

[47] B. Riddick. 2009. *Living With Dyslexia: The social and emotional consequences of specific learning difficulties/disabilities.* Taylor & Francis. https://books.google.com/books?id=NveMAgAAQBAJ

[48] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural Machine Translation of Rare Words with Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL '16)*. 1715–1725.

[49] Sally E. Shaywitz, Michael Escobar, Bennett Shaywitz, Jack Fletcher, and Robert Makuch. 1992. Evidence That Dyslexia May Represent the Lower Tail of a Normal Distribution of Reading Ability. 326 (02 1992), 145–50.

[50] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.

[51] After the Deadline. 2010. Measuring the Real Word Error Corrector. Retrieved September 21, 2018 from https://blog.afterthedeadline.com/2010/04/09/measuring-the-real-word-error-corrector/.

[52] W Tops, C Callens, E Van Cauwenberghe, J Adriaens, and M Brysbaert. 2013. Beyond spelling: the writing skills of students with dyslexia in higher education. *Reading and Writing* (2013).

[53] US Interagency Committee on Learning Disabilities. 1987. *Learning Disabilities: A Report to the U.S. Congress.* https://books.google.com/books?id=OG-7PQAACAAJ

[54] Joseph B Walther. 2007. Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior* 23, 5 (2007), 2538–2557.

[55] Yuhuai Wu, Saizheng Zhang, Ying Zhang, Yoshua Bengio, and Ruslan R Salakhutdinov. 2016. On Multiplicative Integration with Recurrent Neural Networks. (2016), 2856–2864. http://papers.nips.cc/paper/6215-on-multiplicative-integration-with-recurrent-neural-networks.pdf

[56] Zheng Yuan and Ted Briscoe. 2016. Grammatical error correction using neural machine translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 380–386.

[57] Shanyang Zhao, Sherri Grasmuck, and Jason Martin. 2008. Identity Construction on Facebook: Digital Empowerment in Anchored Relationships. *Computers in Human Behavior* (09 2008).