# Supporting Spanish Writers using Automated Feedback

Aoife Cahill<sup>1</sup>, James V. Bruno<sup>1</sup>, James T. Ramey<sup>2</sup>, Gilmar Ayala Meneses<sup>2</sup>,

Ian Blood<sup>1</sup>, Florencia Tolentino<sup>1</sup>, Tamar Lavee<sup>1</sup> and Slava Andreyev<sup>1</sup>

<sup>1</sup> Educational Testing Service, Princeton, NJ 08541, USA

<sup>2</sup> Universidad Autónoma Metropolitana, Cuajimalpa, Mexico City, Mexico

### Abstract

We present a tool that provides automated feedback to students studying Spanish writing. The feedback is given for four categories: topic development, coherence, writing conventions, and essay organization. The tool is made freely available via a Google Docs add-on. A small user study with post-secondary level students in Mexico shows that students found the tool generally helpful and that most of them plan to continue using it as they work to improve their writing skills. In an analysis of 6 months of user data, we see that a small number of users continue to engage with the app, even outside of planned user studies.

### 1 Motivation and Background

There are a multitude of writing support tools available for students who wish to improve their English writing (e.g. Grammarly,<sup>1</sup> Writing Mentor,<sup>2</sup> Ginger,<sup>3</sup> Microsoft Word, or Revision Assistant<sup>4</sup>). These tools for English vary in complexity from basic feedback on spelling errors to advanced feedback about structure, register, topic development, and use of evidence to support claims. In the context of writing feedback for Spanish, there are automatic grammar checkers available (e.g. Language- $Tool^5$  or SpanishChecker<sup>6</sup>). However, there are no tools for Spanish that offer the kind of comprehensive writing feedback that tools such as Writing Mentor and Grammarly offer. There is a huge native Spanish-speaking population (almost 500 million globally according

to Wikipedia<sup>7</sup>) that we could potentially support by providing advanced NLP tools to help improve writing skills.

# 2 Related Work in Automated Feedback

Studies have shown that automated feedback on student writing can have a positive impact on their learning (Attali, 2004; Shermis et al., 2004; Nagata and Nakatani, 2010; Cotos, 2011; Roscoe et al., 2014). The NLP technologies used to provide feedback on writing have often gone hand-in-hand with the development of automated scoring systems. The intuition is that if the system is "measuring" some aspect of writing in order to be able to grade it, it could also use that same measurement in order to give feedback. However, there are also studies with mixed, or less favourable outcomes when students use tools that provide automated feedback on their writing (Choi, 2010; Bai and Hu, 2017; Ranalli et al., 2017). This is an active area of research, and one that requires significant resources to conduct valid user studies and evaluations.

# 3 Spanish Writing Mentor

Building on previous work that developed comprehensive automated writing feedback for English, we have developed a tool to similarly support Spanish writers by providing automated feedback. mentormywriting.org/es. html contains information about the tool, links to the download page, a video describing the main features of the tool, as well as an FAQ section. The tool is implemented as a Google Doc add-on (front-end), freely available to download from the app store, with a server-based

<sup>&</sup>lt;sup>1</sup>www.grammarly.com

<sup>&</sup>lt;sup>2</sup>mentormywriting.org

<sup>&</sup>lt;sup>3</sup>www.gingersoftware.com

<sup>&</sup>lt;sup>4</sup>www.revisionassistant.com/

<sup>&</sup>lt;sup>5</sup>https://languagetool.org

<sup>&</sup>lt;sup>6</sup>https://spanishchecker.com/

<sup>&</sup>lt;sup>7</sup>https://en.wikipedia.org/wiki/Spanish\_ language

back-end processing student texts and computing feedback. It is an extension of the original work done for English (Burstein et al., 2018; Madnani et al., 2018a) and the add-on allows users to select either English or Spanish on a per-document basis. Figure 1 shows the language selection screen when the add-on is first started.



Figure 1: Writing Mentor offers feedback in both English and Spanish. The user makes a selection for each document they write.

The server-based back-end of the tool is implemented using a micro-service framework based on Apache Storm<sup>8</sup> – see Madnani et al. (2018b) for more details. The framework allows for robust, scalable, fault-tolerant processing (automatically restarting components if they fail). The back-end engine takes a text as input and returns a JSON representation of feedback. The feedback is computed using a network of micro-services. Each micro-service is defined in terms of inputs (prerequisites) and outputs, and data flows in parallel (automatically managed by Storm) to the final component (Storm *bolt*) that sends JSON-encoded feedback to the front-end of the tool for display to the user.

The design of the back-end feedback engine was based on the corresponding English one. However, in terms of the implementation, much of the functionality naturally differs in order to account for the language differences. Furthermore, we introduce some new functionality – most notably the section on well-organized writing – that could potentially also be made available for the English version of the tool.

We take advantage of a number of publicly available tools to build our feedback components. We use the Spanish Stanford Core NLP<sup>9</sup> for tokenization, tagging and constituency parsing (Manning et al., 2014). We use the Spacy<sup>10</sup> Spanish dependency parser (Honnibal et al., 2020), aligning the dependency relations to the tokenization provided by the Stanford tools. We use the standalone version of Spanish LanguageTool<sup>11</sup> to compute a subset of the feedback relating to writing conventions (spelling and grammatical errors).

# 3.1 Feedback Components

Spanish Writing Mentor gives feedback on four broad areas of writing: topic development, coherence, writing conventions, and essay organization. Figure 2 shows the tool when the user loads the app on an open document.

### 3.1.1 Topic Development

As in the English version of Writing Mentor, we include feedback on topic development (Beigman Klebanov and Flor, 2013), which relies on a database of pointwise mutual information (PMI) values. In this instance, we are able to re-use the code from the English implementation and simply substitute a Spanish PMI database. We build the database by retokenizing the raw version of the Spanish Billion Word Corpus (Cardellino, 2019) with the Stanford tools. This corpus is a union of Spanish resources in a wide range of domains and formats, including legal, financial and medical

<sup>&</sup>lt;sup>9</sup>https://stanfordnlp.github.io/CoreNLP/

download.html We use version 3.9.2 which has the linguistically desirable feature of separating clitics from the words they depend on during tokenization. Sadly this feature is not available in the most recent versions of the Stanford Spanish tools, as they have switched to UD tokenization.

<sup>&</sup>lt;sup>10</sup>https://spacy.io/models/es

<sup>&</sup>lt;sup>11</sup>https://github.com/languagetool-org/ languagetool/

#### Writing Mentor



Figure 2: The Spanish Writing Mentor tool has four categories of feedback: topic development, coherence, writing conventions, and essay organization

documents, books, and movie subtitles. This variety of domains makes it a suitable background database for topic detection in essays on many different subjects. Topic words are identified if they have PMI values higher than a set threshold when paired with all other words in the text, i.e. only the PMI of words in the student's response are considered for this feature. The threshold for identified topic words was tuned by experimenting with a range of values and manually inspecting the output to judge the appropriateness of the topics detected. The tuning was done on a small set of 96 essays written by native-speaker university students as well as a sample of 60 essays representing various levels of proficiency from a publicly available corpus of non-native Spanish essays.<sup>12</sup> A lower threshold yielded many word pairs that were unrelated, while a higher threshold yielded far fewer word pairs. Further threshold tuning and assessment of the background database on a broader dataset remains for future work. The main topic of the essay is identified as the topic word that participates in the most pairs of words over the PMI threshold. As in the English version, users are also able to provide their own topic terms. If these are provided

they are highlighted and considered in the automatic identification of the main topic regardless of their PMI values. Users may also manually identify topics of their own choosing, and related topic words are highlighted according to the same rules.

#### 3.1.2 Coherence

Spanish Writing Mentor gives feedback on the following aspects of coherence:

- Flow of ideas The same topic words that are highlighted in the *Topic Development* component are also highlighted in this component, color-coded according to topic. This enables the user to visually understand the extent to which their topics are elaborated in various parts of the document, and they are advised that the most important topics should be represented throughout the entirety of the text.
- **Transition terms** We have a fixed list of 100 words and phrases<sup>13</sup> that we highlight. This is intended to prompt users to consider over/under use of transition terms that link ideas and arguments. Examples

<sup>&</sup>lt;sup>12</sup>https://github.com/ucdaviscl/cowsl2h/blob/ master/README.md

<sup>&</sup>lt;sup>13</sup>Adapted from https://modlang.unl.edu/docs/ STC/Palabras\_para\_ordenar\_el\_discurso\_escrito. pdf

include porque (because), primeramente (firstly), en conclusión (in conclusion), etc.

- Title and section headers We identify titles and section headers using a set of regular-expression-based rules. These are used both to visually prompt the user about the identified structure of their essay, as well as identify sections of the essay that we do not want to give certain kinds of feedback on. For example, we do not want to highlight spelling or grammatical errors in a list of references.
- Sentence/paragraph length We highlight complex sentences, which we consider to be sentences containing 2 or more dependent clauses as identified by the constituency parse. This is intended to highlight sentences that could perhaps be broken up to make the text more readable. Using the number of sentences identified by the tokenizer, we also highlight paragraphs that are either too short (choppy, <4 sentences), or too long (>9 sentences), to prompt users to think about elaborating their claims without losing coherence. This extends what is available in the English version which only gives feedback at the sentence level.
- **Pronoun use** We highlight a subset of pronouns to help prompt the user to make sure that the references that the pronouns refer to are clear. The POS tags are used to identify the pronouns.

# 3.1.3 Writing Conventions

We give the following types of feedback on Spanish Writing Conventions:

Grammar, Usage and Mechanics We follow a similar categorization of error types to the English Writing Mentor. Some of these errors come directly from the LanguageTool library, though only a subset of errors detected are displayed to the user. We include accent errors, agreement errors, contraction errors, comma errors, spelling errors, and incorrect word usage errors. We also implemented new grammatical error detectors. For example, rules were written to identify fragments and run-ons based on subordinating and coordinating conjunctions and their dependents identified by the dependency parses. Our initial work focused on trying to include only feedback for errors for which we were confident we could achieve reasonable precision, though of course no system is perfect. Future work would extend the coverage of these detectors.

- **Unnecessary Words** Related to the concept of pobreza léxica (*lit.* lexical poverty), we highlight occurrences of unnecessary words. These words, when over-used, lead to imprecise and poor writing. This is done by simple regular expression matching from a list that includes words like *absolutamente* (absolutely) and *muy* (very). Future work would build out this functionality to account for more specific guidelines related to this topic.
- **Contractions** We highlight sequences of words that should be contracted in Spanish, e.g. *de el* should be written as *del*.
- Accents We highlight errors related to accent use. This category of errors is new for the Spanish version of the app. These errors are identified using dictionary resources and rules encoded in Language Tool.

# 3.1.4 Essay organization

A novel aspect of the Spanish tool is that we give feedback on essay organization in the form of a questionnaire. There are 9 main questions, each with a corresponding follow-up question (18 questions total), that prompt the user to think about how they have structured the arguments in their essay. This questionnaire draws on concepts from various rhetoric and composition studies textbooks (e.g. Ramage et al. (2015), Lunsford (2008), Hacker (2006), and Crews (1992)). The questions were chosen to implement insights and recommendations from the writing literature. Figure 3 shows the tool prompting the user to highlight the sentence in the essay containing the main claim. When the user has completed the survey, they are presented with a summary of the aspects that they highlighted, schematized in Figure 4. An obvious extension of this component will be



7/18: Por favor, resalta la oración completa en la que aparece la afirmación principal del ensayo.

Figure 3: Users are prompted to think about the structure of their essay in the *Well-Organized* questionnaire. Here the users are asked to select the sentence containing the main claim of the essay. This is question 7 of 18 total.

- I. Introduction. Thesis Statement: [Thesis statement as highlighted by the user]
- II. Argumentative Paragraph 1: [Topic sentence as highlighted by the user]
- III. Argumentative Paragraph 2: [Topic sentence as highlighted by the user]
- IV. Argumentative Paragraph 3, etc...
- V. Conclusion Paragraph: [Conclusion paragraph as highlighted by the user]

Figure 4: A schematic of what the user sees after completing the organization questionnaire.

to automate the detection of organizational elements and present an automated *sentence outline* (a formal representation of an essay draft) to the user in the future.

# 3.2 Paragraph-Writing Support

Analogous to the English app, the Spanish app also provides support for paragraph writing. The idea behind the paragraph-writing part of the tool is to support less proficient writers; for example adult learners. The paragraph-writingsupport tool includes motivational badges, and provides a subset of the feedback available in the main tool. The focus in the paragraphwriting tool is to help the user understand what aspects of writing lead to a well-written paragraph. Figure 5 shows a screenshot of the paragraph-writing help, which provides scaffolding and guidelines for writing a well-structured paragraph in response to an argumentative question. There are a number of questions available to students to help them practice. The questions come from the New York Times Teaching Resources. (They are translations of



Figure 5: Users are prompted to think about the structure of a paragraph. This interface provides scaffolding to help them write a wellstructured paragraph in response to an argumentative prompt.

the original questions supplied by the NYT who approved the translations for use).

# 4 User Study

We conducted a user study to collect initial usage and perception data from the tool. Our participants were students in the Universidad Autónoma Metropolitana, Cuajimalpa, in Mexico City. Participants were recruited from two groups: (1) a group of students taking optional courses in the university Writing Center, which provides support to students who want to improve their writing skills and (2) a group of 3rd and 4th year undergraduate students taking an elective course in Latin American Literature. Participation in the study was optional, and each student who took part received a certificate of participation upon completion. Participants were asked to use Spanish Writing Mentor to support their regular coursework writing assignments. No changes were made to the assignments. Users were given instructions

on how to use the tool three weeks before the end of the trimester, and could choose how to use the tool (if at all) during those three weeks. The user study focused only on extended writing. An investigation into the usefulness of the paragraph writing component remains for future work.

Our user study consists of two components: (1) a measure of writing ability before and after using the tool and (2) a user survey completed after the three weeks. In order to measure writing ability before and after using the tool, each participant completed a standardized assessment of writing ability. The assessment is usually administered as a placement test in the Writing Center to assign students to one of four levels: low (0-49), moderate (50-69), acceptable (70-89) or optimum (90-100). Our user survey consisted of 13 questions (see Appendix A) and participants were asked to complete it after they had handed in their final assignments.

# 5 User Data Analysis

Table 1 gives an overview of our participants. We have 13 students in total who completed the entire study; 6 from the Spanish Writing course, and 7 from the Latin American Literature course. All students take the standardized test before their course and after, and receive a score in the range 0-100. We see that the writing ability of all participants, as measured by the standardized test, increases between the pre- and post-tests. Of course, this improvement can be attributed to the content of the courses, and at this point we have no way to measure the direct impact (if any) of using the Writing Mentor tool. A fully randomized controlled experiment would be needed to study this in more depth. For comparison, the average scores of all students (n=94) in the pre-test was 38.7 and this increased to 51.9 in the posttest (n=80 students). The writing proficiency of our participants was, on average, higher than the general population in these classes.

The main findings from the 13 questions in the user survey were as follows:

• The average score for how useful the participants found the tool was 3.7 (on a scale from 1-5, 5 being the most useful; min=2, max=5).

Student	Course	Initial	Final
1	Writing	39	64
2	Writing	44	58
3	Writing	58	65
4	Writing	37	45
5	Writing	72	76
6	Writing	38	50
7	Literature	36	53
8	Literature	28	42
9	Literature	58	63
10	Literature	72	76
11	Literature	50	52
12	Literature	58	62
13	Literature	51	57
Average		49.3	58.6

Table 1: User Study Participants. Initial and Final are the written evaluation results on a standardized test (0-100 scale).

- 12 of 13 participants indicated that by using the tool they had learned something to help them improve their writing.
- The features that were selected as being the most useful were: Coherence (flow of ideas) - 10/13; Well organized - 10/13; Topic development - 9/13.
- The features that were selected as being the least useful were: Coherence (Title and Section headers) – 4/13; Coherence (Use of pronouns) – 3/13; Coherence (Sentence/paragraph length) – 3/13.
- The most useful help article was the one on Coherence (Flow of ideas) 10/13.
- 12/13 participants plan to use Spanish Writing Mentor again, and 11/13 planned to recommend it to others.
- 11/13 participants indicated that one of the main aspects they liked LEAST about the tool was the interface, but only 2/13 participants commented that the functionality provided by the tool (i.e. what it was presenting as feedback) were what they liked least.

### 5.1 User Behavior

In addition to the user study, we also analyze 6 months of application log data. Figure 6a



(b) The average length of a Spanish Writing Mentor usage session (in minutes) by month.



(c) The distribution of the time (in seconds) for each section of the app for the time period October to March.

Figure 6: Analysis of 6 months of user data

shows the number of unique users each month between October 2020 and March 2021. We see that the number of users peaked during our user study, but did not drop off entirely once the study was over. Figure 6b shows the average time (in minutes) for an active session (i.e. we exclude sessions where no text was entered). We see differences in average usage across months, but for the months with the most users, the average time spent using the app was between 15 and 30 minutes. Finally, Figure 6c shows the distribution of the time (in seconds) for each section of the app from the time period October 2020 to March 2021. We restrict the plot to the interquartile range, since there were many extreme outliers (probably due to users switching away from the app and coming back later). Even still, we see quite a range of values for the medial time spent per section. Sections such as contractions, grammar errors, word choice, transition terms, well organized and long sentences engaged the users for longer times, while sections such as flow of ideas, pronoun use, title and section headers engaged the users less.

#### 6 Conclusions

We presented a tool to support writers of Spanish by providing them automated feedback within a free Google Docs add-on. The tool was built by adapting an existing tool for English, and implementing Spanish-languagespecific components. We conducted a small study with 13 post-secondary level students in Mexico City, and in general found that they considered the tool helpful and were planning to continue using it and also recommend it to others. We see that their writing ability, as measured by a standardized test, improved between the pre- and post-tests, though we cannot yet say whether the Spanish Writing Mentor app contributed to this improvement. We find that a small number of users engage with the app, even outside of planned user studies, which is encouraging.

#### References

- Yigal Attali. 2004. Exploring the Feedback and Revision Features of *Criterion*. Paper presented at the National Council on Measurement in Education (NCME), Educational Testing Service, Princeton, NJ.
- Lifang Bai and Guangwei Hu. 2017. In the face of fallible awe feedback: how do students respond? *Educational Psychology*, 37(1):67–81.
- Beata Beigman Klebanov and Michael Flor. 2013. Word association profiles and their use for automated scoring of essays. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1148–1158.
- Jill Burstein, Norbert Elliot, Beata Beigman Klebanov, Nitin Madnani, Diane Napolitano,

Maxwell Schwartz, Patrick Houghton, and Hillary Molloy. 2018. Writing mentor: Writing progress using self-regulated writing support. *Journal of Writing Analytics*, 2:280–284.

- Cristian Cardellino. 2019. Spanish Billion Words Corpus and Embeddings.
- Jaeho Choi. 2010. The Impact of Automated Essay Scoring (AES) for Improving English Language Learner's Essay Writing. Ph.D. thesis, University of Virginia Charlottesville, VA.
- Elena Cotos. 2011. Potential of automated writing evaluation feedback. *Calico Journal*, 28(2):420– 459.
- Frederick Crews. 1992. The Random House Handbook, Sixth edition. McGraw Hill.
- Diana Hacker. 2006. *The Bedford Handbook*, Seventh edition. Bedford/St. Martin's.
- Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.
- Andrea A. Lunsford. 2008. St. Martin's Handbook, Sixth edition. Bedford/St. Martin's.
- Nitin Madnani, Jill Burstein, Norbert Elliot, Beata Beigman Klebanov, Diane Napolitano, Slava Andreyev, and Maxwell Schwartz. 2018a. Writing mentor: Self-regulated writing feedback for struggling writers. In Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations, pages 113–117, Santa Fe, New Mexico. Association for Computational Linguistics.
- Nitin Madnani, Aoife Cahill, Daniel Blanchard, Slava Andreyev, Diane Napolitano, Binod Gyawali, Michael Heilman, Chong Min Lee, Chee Wee Leong, Matthew Mulholland, and Brian Riordan. 2018b. A robust microservice architecture for scaling automated scoring applications. *ETS Research Report Series*, 2018(1).
- Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In Association for Computational Linguistics (ACL) System Demonstrations, pages 55–60.
- Ryo Nagata and Kazuhide Nakatani. 2010. Evaluating Performance of Grammatical Error Detection to Maximize Learning Effect. In *Proceedings of COLING (Posters)*, pages 894–900, Beijing, China.
- John D Ramage, John C Bean, and June Johnson. 2015. The Allyn & Bacon guide to writing, Seventh edition. Pearson.

- Jim Ranalli, Stephanie Link, and Evgeny Chukharev-Hudilainen. 2017. Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. Educational Psychology, 37(1):8–25.
- Rod D Roscoe, Laura K Allen, Jennifer L Weston, Scott A Crossley, and Danielle S McNamara. 2014. The writing pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, 34:39–59.
- Mark D. Shermis, Jill C. Burstein, and Leonard Bliss. 2004. The Impact of Automated Essay Scoring on High Stakes Writing Assessments. In Annual Meeting of the National Council on Measurement in Education.

# Appendix

#### A User Survey

The following questions were included (in Spanish) in our final user survey:

1. Overall, how useful was Writing Mentor in helping you with your writing	5 point scale (5 be-
assignment?	ing most useful)
2. What (if anything) did you like most about Writing Mentor?	free text
3. What (if anything) did you like the least about Writing Mentor?	free text
4. Do you think you learned anything by using Writing Mentor that will help	Yes/No
you improve your writing?	
a. If so, what?	free text
5. Please mark the features that you found MOST useful (if any):	Multiple-select
	from list of all
	possible sections in
	the tool
6. Please mark the <i>Writing Help</i> articles that you found MOST useful (if any)	Multiple-select
	from list of all help
	articles
7. Please mark the features that you found LEAST useful (if any):	Multiple-select
	from list of all
	possible sections in
	the tool
8. Please mark the Writing Help articles that you found LEAST useful (if	Multiple-select
any):	from list of all help
	articles
9. What feature(s) should be added to writing mentor (if any) and why?	free text
10. What other things should be improved in writing mentor (if any), and	free text
why?	
11. Do you plan to use Writing Mentor again in the future to help you with	Yes/No
your writing assignments?	
12. Do you plan to recommend Writing Mentor to others?	Yes/No
13. Is there anything else you would like to share about your experience using	free text
writing mentor?	

 Table 2: Final User Survey Questions