

Foreign Language Literacy through Fanfiction Writing and Text Mining

Patrícia da Silva Campelo Costa, Eliseo Berni Reategui
Federal University of Rio Grande do Sul PPGIE, Brazil

Abstract

This paper presents a novel approach on the use of text mining as a support for foreign language literacy. The mining tool used in our research is able to extract recurring terms from a given text and represent them in the form of a graph. Such graphical representation can help students to focus on specific features of the text and create narratives in a foreign language. We carried out an experiment to investigate how foreign language learners deal with the process of creating fanfiction narratives by using the mining tool. Most of the students that participated in the experiment stated that the graphs provided by the mining tool helped them to reflect and create a clearer picture of the stories they were about to write.

1. Introduction

Digital tools which may be effective in the process of developing literacy in a foreign language can work as mediator resources for pedagogical tasks. Several scholars have established useful links among issues which concern technology, language, and literacy [1, 2, 3]. This way, it is important that teachers who act in these areas use electronic resources in order to create tasks which can consider the process of literacy in a foreign language. In this research proposal, literacy is considered the process through which foreign language learning can occur based on the development of socially meaningful reading practice and written production [4]. Having as a starting point the reading and analysis of the textual genre fanfiction (stories published on the Internet, produced by fans of media icons such as movies and TV shows), the literacy process in a foreign language is observed, which can take place through the relationship between technology and language learning. Hence, a discussion regarding cognitive and social processes related to literacy on the web arises as it is examined how reading and writing can be developed through the use of a text mining tool.

Numerous studies have emphasized the use of digital tools in the process of foreign language learning [5]. An issue that remains to be researched, however, is the extent to which certain tools, such as

the one for mining in the current study, can aid the literacy process so that a real internet genre (e.g. fanfiction) can be produced. This work intends to foster the process of foreign language literacy focused on reading and writing via the following means: a) student involvement with a pedagogical task through the creation of a literary internet genre b) support for written production through the use of a mining tool which may aid the student in the creation of a new text.

For the purposes of this research, reading can be promoted through the contact with websites about fanfiction, in a phase previous to the one when readers/writers analyze these texts as models for the creation of new narratives. When analyzing the fanfictions, students are able to use a device which can help the reader identify linguistic patterns that are recurrent in fanfictions. This way, learners will have carried out an exam of this textual structure before starting the creation of their text *per se*. The text emphasized in this work is an internet based genre which is focused on narratives written by fans of pop culture icons (represented by *animes*, movies, series, etc). When producing their fanfics, the authors/fans base the plots on original stories while modify and add things to the script, presenting, for instance, new relationships between the characters [6]. Thus the goal is to explore the process of literacy in a foreign language by means of inspecting the reading and writing of fanfictions by learners who may face this practice as a moment of meaningful textual creation.

This paper presents an approach to foster textual production in English as a foreign language through the use of particular digital resources. This work is based on previous research on the use of text mining to support the evaluation of texts produced collaboratively [7] and on the studies by Black [8,9] on the incorporation of an internet genre (fanfiction) in language learning.

2. Literacy and Technology

Literacy is not only reading and writing, but also a meaningful process of social practice individuals go through, so that they interact in real contexts of use [3]. Similarly, according to Street [10], literacy practices may be defined as specific ways of reading, writing and reflecting about these same processes in specific contexts of use. Consistent with Gee, Hull

and Lankshear [11], to read and write specific texts people should immerse in social practices in which participants not only read but also carry certain attitudes and values in relation to these documents, in order to interact socially with and through these texts. According to Scribner and Cole [12], social practice refers to socially developed and standardized ways of using technology and knowledge to perform tasks. Thus, literacy involves the application of knowledge such as reading and writing in specific contexts of use for specific purposes.

Since digital natives can often be proficient users of multiple semiotic systems [13], the processes of literacy supported by digital technology moves reading beyond the linear and sequential procedure, focusing on the interaction with multimodal hypertexts, which may be read without an exact order.

Hull and Schultz [13] propose a study of new literacies based on sociocultural theory, so that there is a focus on the relationship between social practice, production, exchange and construction of meaning. Typical examples of the new literacies are video games, production blogs, social practices related to the use of mobile technology and fanfiction writing [3]. Hence, new literacies involve practices that tend to be more developed in extra-curricular contexts.

3. Text Mining

This work investigates how foreign language literacy can be fostered through the use of a text mining tool which can help students in the analysis of textual genres. Through the use of a particular text mining tool called SOBEK [7], recurring terms can be extracted from text and represented in the form of a graph. Such graphical representation can help students to focus on specific features of the text and, from that analysis, to create narratives in a foreign language. SOBEK identifies relevant concepts in a text by identifying frequent terms and relationships among them. A specific algorithm was implemented based on statistical information extracted from texts and represented in a graph model called n-simple distance [14].

The tool operates as follows:

- (1) A text is copied onto the area where data is inserted. Alternatively, it is possible to upload a file in one of the formats: txt, doc or pdf.
- (2) From this text, the tool automatically creates a base of concepts through the process of text mining. These concepts correspond to a set of words and terms that occur more frequently in the text (aside from articles, prepositions and other words which are not deemed relevant).
- (3) When the concepts are extracted, a graph is generated with the most common concepts and their

relationships. Figure 1 shows an example of a graph drawn by SOBEK from a fanfiction based on the movie Avatar.

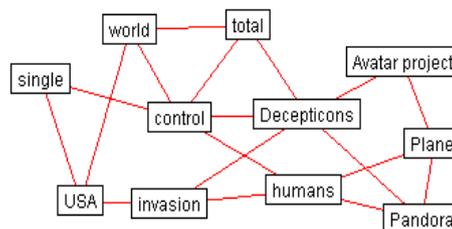


Figure 1. Graph based on a fanfiction about the movie Avatar

From the point of view of text writing, such graphs may be used to help students reflect on their readings as they analyze the terms included in the graph. They may create and undo associations among concepts and employ the tool and its graphs for the pre-writing stage.

The next section presents the genre fanfiction, emphasizing its use in the process of language learning.

4. The Genre Fanfiction

Black [8] exposes the concept of a fanfiction as an internet-based genre focused on narratives written by fans of some of the icons from popular culture, such as TV shows, *animes* (Japanese animation), movies, etc. In the production of fanfiction, authors/fans, having as a starting point the original story, can enhance or change the stories script, adding new characters and changing relationships among the existing ones. Though fanfiction has existed for years in printed form [15], new ICTs (Information and Communication Technologies) offer the fans the opportunity of having online meetings during which authors and readers can write, exchange texts, criticize and discuss changes in their fanfictions collaboratively [8]. At such moments, the authors have the chance of receiving feedback from their interlocutors, write and interact collaboratively with social groups alike. Black's studies [6,8,9] demonstrated the effectiveness of writing fanfiction for the process of language learning, by showing that the informants who participated in her research developed confidence and motivation to continue writing due to the interest and scaffolding provided by their readers.

5. A writing approach supported by text mining

Our fanfiction writing approach so as to promote language learning starts with the reading of

fanfictions. Students are advised to enter a particular website and look for a fanfiction that interests them. After reading the selected fanfiction, students may use SOBEK to extract a graph from the text which was read.

By analyzing the results of the text mining process, the students may confront the knowledge represented in the graph with a preliminary mental model they have built for the fanfiction they have read. In this process, the students are no longer accommodated, since they need to formulate interpretations about the graph to bring it closer to their viewpoint. While the students examine the graphs and confront them with their ideas, they may also edit the graphs, inserting or eliminating nodes, adding or deleting relationships between them. Therefore, in this process of imbalance – accommodation, students may develop a deeper understanding of the situation and facts they want to write about.

When talking about memory and mental models, Sternberg [16] states that concepts can be organized into schemas, which are mental structures that represent knowledge involving a number of interrelated concepts in a meaningful organization. Such structures resemble those of the graphs used to represent concepts and their relationships.

Furthermore, from the extraction of concepts, students may have as a resource for the draft of their text, graph models which can help them in the structuring of a new text. As Torrance and Galbraith [17] point out, the strategy of production in which there is the composition of a draft may assist a more fluid process of writing, since the individual has a first draft on which to rely. Thus, the author would have more opportunity to organize his/her ideas prior to writing, so as to leave his/her attention exclusively devoted to putting those ideas into words - the stage of creating texts per se. Moreover, according to Graham [18], it is relevant that many teachers in their classrooms provide opportunities for discussion aimed at writing strategies. For the author, at least three cognitive processes should be incorporated – *planning*, *development* and *editing* of a text – so as to provide an effective and independent use of such procedures by the students. Thus, in this research, the stage of *planning* is contemplated by the extraction, analysis and editing of the graphs. The stage of *development* is achieved by putting into words the ideas represented in the graph, creating an initial draft. The stage of *editing* is finally accomplished by the actual writing of the final version.

6. An experiment based on a text writing task

The experiment was conducted so as to investigate how the process of writing texts online

could be aided by the use of a text mining tool (SOBEK). The study was carried out with a group of undergraduate students from the Federal University of Rio Grande do Sul (UFRGS), who had an upper-intermediate level of English. The group was composed of five students with ages ranging from 19 to 24 years, and one computer was available for each participant.

The class prior to the experiment consisted of discussions about the textual genre fanfiction and the practical usage of the text mining tool. Along these lines, the main features of the tool were presented so that students could grasp the procedures which were based on the extraction of concepts and generation of graphs.

Then, learners were introduced to the website *fanfiction.net*, which is considered the largest database of fanfiction stories on the web. At this stage, students could explore further the topic *fanfiction* and choose their favorite stories to read.

The students were drawn to take notice of the central genre features, such as the fierce reader interference in the text, since fanfictions tend to be published in chapters and the author often relies on the feedback from readers to continue their story. Subsequently, the proposal was that the students use the tool SOBEK to extract recurrent terms and to build a graph from the text chosen.

At this stage, the learners had the opportunity to add or delete concepts and relationships in the graph. Through the analysis of the graphs, students could see, for instance, the recurrence of verbal tenses which are typical in narratives, as well as the regularity of adjectives used to describe characters and settings. According to the task proposed in this research, a student A should save the resulting graph and forward it to a colleague (student B). Thereafter, this classmate should create a fanfiction based on the graph received from student A, using a regular text editor.

From this task analysis, we observed that the graphs generated by the text mining tool have played an important role in the analysis of the fanfictions and subsequent production of texts. Most of the students made comments as to how the analysis of the graphs made them go back and forth to the original texts in order to better understand them and to create a clearer picture of the stories for their classmates. The actual text writing departing from the graphs has also demonstrated to be an interesting way to instigate students to create new stories based on a few central ideas.

7. Data Analysis

The task proposed to the students may be classified as well-structured, having specific steps for reading and writing development. This way, the task of writing may gain support, based on a collaborative

process of text production. This collaborative work should not displace individual work but “both approaches can complement each other and expand the writers’ linguistic experiences in a holistic manner” [19]. The creation is collective because two students work on different stages while developing the text.

Our main question, therefore, is the following: How can this tool support textual production? We have focused on the following cues:

- a) The number of times the students returned to the original text as to edit concepts before generating a graph;
- b) The number of times the students checked the graph while writing;
- c) Recurrence of terms in the text, which were also present in the graph;
- d) Topics in the text related to concepts from the graph;

By analyzing the results of the text mining process, the students may confront the knowledge represented in the graph with a preliminary mental model they have built for the fanfiction they have read. In this process, the students are no longer accommodated, since they need to formulate interpretations about the graph to bring it closer to their viewpoint. While the students examine the graphs and confront them with their ideas, they may also edit the graphs, inserting or eliminating nodes, adding or deleting relationships between them. Therefore, in this process of imbalance, students may develop a deeper understanding of the situation and facts they want to write about.

Below, there is the description of some actions carried out by student A as she edited the concepts bar.

Apparently, the students analyze their own text comprehension, as they go back and forth, rereading the original text and comparing the concepts presented in the graph to the semantic relations established among the terms in the story. Therefore, they return to the text and practice a more attentive rereading. The students need to return to the text, so that they can establish relations among the terms highlighted by the mining software. After using the mining process, the reading seems to be more thoughtful because they scroll down the page and take a longer while to read the story in a way that the concepts they are supposed to edit become more interlinked. As we can observe below, there is a process of more attentive reading after the tool is used and the graph is built, since students return to the text as they decide to reread the story.

Table 1. Text 1 - Students A and B

Time	Student’s actions
Footage 1¹	
15:27 – 15:35	Student A chooses the fanfiction <i>Jumping to Conclusions</i> ² .
15:36 – 18:04	Student A copies this new story and pastes it on the initial page of the SOBEK device.
18:05 – 18:15	Student A clicks on <i>Extract Concepts</i> and analyses the relations which were established by the mining software.
18:16 – 20:00	Student A edits the column <i>Concepts</i> , providing relations among terms and removing some of the concepts.
20:01 – 20:38	Student A clicks on <i>Create Graph</i> and observes the image.
20:39 – 21:12	Student A returns to the <i>Concept Base</i> and edit more concepts.
21:13 – 21:39	Student A clicks on <i>Create Graph</i> and observes the image.
21:40 – 23:57	Student A returns to the original text on the webpage and seems to reread more attentively.

Table 2. Text 1 - Students A and B

23:58 – 24:04	Student A checks the section about <i>Friends</i> once more and looks for a new fanfiction on this theme.
24:05 – 24:10	Student A chooses the fanfiction <i>TOW The Central Park receipt</i> ¹ .
24:11 – 24:31	Student A copies this new story and pastes it on the initial page of the tool SOBEK.
24:32 – 24:39	Student A clicks on <i>Extract Concepts</i> and analyses the new relations which were established by the mining software.
24:40 – 32:56	Student A returns to the original text on the webpage and starts rereading it more attentively.

It was observed that these concepts were edited more than once and then later compared to a new graph. It seems that the main goal was to provide a clear and more organized graph for their peer, so that they would have a comprehensible illustration to support writing. Besides, the terms which were present in the graph, were recurrently used by the authors of new texts. Thus, the students would not face a blank page – which is a restraint many writers have to deal with. Such evidence demonstrates that the tool also served to provide an initial prompt for a more fluid writing process.

When faced with a graph built by SOBEK, the students are no longer accommodated, since they need to formulate interpretations about that graph to bring it closer to the internal representation they have of a given situation/theme. This process of imbalance – accommodation – leads students to reflection, making them build richer and more solid foundations in their writings processes.

Since Sobek can provide graphs with recurrent concepts from a specific text, this resource may be useful to foster reading and writing. The users can ponder about their reading after analyzing the terms included in the graph, they can create associations among these concepts and/or employ the tool and its graphs for the pre-writing stage. From the extraction of concepts, the student may have, as a resource for the draft of their text, graph models which can assist the structuring of a new text. As Torrance and Galbraith [17] point out, the strategy of production in which there is the composition of a draft may assist a more fluid process of writing, since the individual has a first draft on which to rely. Thus, the author would have more opportunity to organize their ideas prior to writing, so as to leave their attention exclusively devoted to putting those ideas into words – the stage of creating texts *per se*. Moreover, according to Graham [18], it is relevant that many teachers in their classrooms provide opportunities for discussion aimed at writing strategies. For the author, at least three cognitive processes should be incorporated – planning, development and editing of a text – so as to provide an effective and independent use of such procedures by the students. Thus, in this research, the stage of *planning* is contemplated by the extraction, analysis and editing of the graphs. The stage of *development* is achieved by putting into words the ideas represented in the graph, creating an initial draft. The stage of *editing* is finally accomplished by the actual writing of the final version.

8. Conclusions

This paper presented an approach to foster textual production in English as a foreign language through the use of particular digital resources. From this task analysis, we observed that the graphs generated by

the text mining tool have played an important role in the analysis of the fanfictions and subsequent production of texts. The data provided by the students that participated in the experiment indicated that the graphs produced by the mining tool helped them reflect and create a clearer picture of the stories they were about to write. The analysis of the graphs made them go back and forth to the original texts in order to better understand them and to create a more thorough representation of the stories for their peers. The actual text writing departing from the graphs has also demonstrated to be an engaging way to instigate students to create new stories based on a few central ideas.

According to Sternberg [16], the fundamental unit of symbolic knowledge is the concept, which can be organized into categories and schemas in our mind. The author states that the reader creates a mental model of the text being read. This mental model can be considered a kind of internal working model of the situation described in the text, as understood by the reader. In other words, the reader creates a mental representation that contains in itself the main elements of the text. Still, as said by Sternberg [16], the construction of mental models illustrates that, in addition to understanding themselves, we also need to comprehend how they combine into integrated representations of meaning when reading narratives, for instance. This way, our hypothesis is that a text mining tool such as SOBEK may act as a resourceful device to materialize this mental model into a visual graph which actually represents a text through connections among concepts. This way, the tool may promote foreign language development. Since the learner can work with the miner as to write authentic texts, such as the genre fanfiction, the process of text production can be linked to a literacy event, which, in this case, consists of meaningful writing shared in online communities.

9. Acknowledgements

This project has been partially funded by CNPq (grant 476398/2010-0), CAPES (grant 00.889.834/0001-08) and FAPERGS (grant 1018248).

10. References

- [1] Kern, R. & Warschauer, M. (2000). 'Theory and practice of network-based language teaching', in Warschauer, M. and Kern, R. (Eds.), *Network-based language teaching: concepts and practice* (pp.1-19). Cambridge University Press, New York.
- [2] Warschauer, M. (2001). 'Millennialism and media: language, literacy, and technology', *AILA Review 14: Applied Linguistics for the 21st century*, 44, pp. 49-59.

- [3] Lankshear, C. and Knobel, M. (2007). 'Sampling "the new" in new literacies', in Knobel, M. and Lankshear, C. (Eds.), *A new literacies sampler* (pp. 1-24). Peter Lang Publishing: New York.
- [4] Barton, D.; Hamilton, M. (2000). 'Literacy practices', in Barton, D.; Hamilton, M. and Ivanic, R. (Eds.), *Situated literacies: reading and writing in context* (pp.7-15). Routledge: New York.
- [5] Wible, D.; Kuo, C. and Tsao, N. (2004) 'Contextualizing language learning in the digital wild: tools and a framework', in ICALT, pp.231-235, Fourth IEEE International Conference on Advanced Learning Technologies (ICALT'04).
- [6] Black, R. (2007). 'Digital design: English language learners and reader reviews in online fanfiction', in Knobel, M. and Lankshear, C. (Eds.), *A new literacies sampler* (pp.115-136). Peter Lang Publishing: New York.
- [7] Macedo, A. L., Reategui, E. R., Behar, P. A. (2009) 'Using Text-Mining to Support the Evaluation of Texts Produced Collaboratively', in *Education and Technology for a Better World: Selected papers of the 9th World Conference on Computers in Education*, Springer: Berlin.
- [8] Black, R. W. (2006). 'Language, culture, and identity in online fanfiction'. *E-learning*, 3 (2), pp. 170-184.
- [9] Black, R. (2009). 'Online fanfiction, global identities, and imagination'. *Research in the teaching of English*, 43(4), pp. 397-425.
- [10] Street, B. (2001). *Literacy and development: ethnographic perspectives*. Routledge: London.
- [11] Gee, J.; Hull, G.; Lankshear, C. (1996). *The new work order: behind the language of the new capitalism*. Westview Press: Boulder.
- [12] Scribner, S.; Cole, M. (1981). *The psychology of literacy*. Harvard University Press: Cambridge.
- [13] Hull, G. & Schultz, K. (2001). 'Literacy and learning out of school: A review of theory and research', *Review of Educational Research*, 71(4), pp. 575-611.
- [14] Schenker, A. (2003). *Graph-Theoretic Techniques for Web Content Mining*. PhD thesis, University of South Florida.
- [15] Jenkins, H. (1992). *Textual poachers: television, fans and participatory culture*. Routledge: New York.
- [16] Sternberg, R. (2008). *Cognitive psychology*. Cengage: Belmont.
- [17] Torrance, M.; Galbraith, D. (2006). 'The processing demands of writing', in Macarthur, C.; Graham, S.; Fitzgerald, J. (Eds.), *Handbook of writing research* (pp. 67-80). The Guilford Press: New York.
- [18] Graham, S. (2006). 'Strategy instruction and the teaching of writing: a meta-analysis', in Macarthur, C.; Graham, S.; Fitzgerald, J. (Eds.), *Handbook of writing research* (pp. 67-80). The Guilford Press: New York.
- [19] Elola, I. (2010). 'Collaborative writing: fostering foreign language and writing conventions development', *Language Learning & Technology*, 14 (3), 51-71.