MyAnnotator: A Tool for Technology-Mediated Written Corrective Feedback

Marie-Josée Hamel,* Nikolay Slavkov,* Diana Inkpen, and Dingwen Xiao****

University of Ottawa
**Official Languages and Bilingualism Institute (OLBI) / **School of Electrical Engineering and Computer Science*
70 Laurier Ave East, Room 130 / 800 King Edward
Ottawa, ON, K1N 6N5, Canada

Marie-Josee.Hamel@uOttawa.ca; Nikolay.Slavkov@uOttawa.ca; Diana.Inkpen@uOttawa.ca; Stephen150150@gmail.com

ABSTRACT. *This article reviews some long-standing issues in the literature on written corrective feedback (WCF), discusses the potential of technology to support some of the tasks involved in the essay marking process, and then presents a new error annotation tool, MyAnnotator, developed by the authors with the purpose of facilitating technology-mediated corrective feedback. We offer an overview of different types of electronic tools that can be used in the teaching of writing, including editors, correctors and annotators, and then draw brief comparisons between MyAnnotator and other similar tools. We demonstrate some of the affordances of our tool and capitalize on its openness and flexibility in facilitating various theoretical and practical approaches to feedback. We conclude with suggestions for future work in natural language processing (NLP) and data driven learning (DDL) possibilities related to our tool.*

RÉSUMÉ. *Cet article propose un résumé des questions longuement discutées dans la littérature à propos de la rétroaction corrective écrite (RCE), illustre le potentiel de la technologie pour soutenir certaines des tâches du processus de correction de textes écrits et finalement présente un nouvel outil d'annotation d'erreurs, MyAnnotator, développé par les auteurs dans le but de faciliter la RCE à l'aide de la technologie. Y sont décrits différents types d'outils électroniques qui peuvent servir en contexte d'enseignement de l'écrit, notamment les éditeurs, les correcteurs et les annotateurs. De brèves comparaisons sont ensuite produites entre MyAnnotator et d'autres annotateurs d'erreurs à visées similaires, tout en mettant en évidence les potentialités de notre outil et misant sur sa souplesse pour faciliter diverses approches théoriques et pratiques de la RCE. Nous concluons avec des suggestions de futur développement ayant trait au traitement automatique de la langue (TAL) et aux possibilités d'apprentissage guidé par les données (data-driven learning) liées à notre outil.*

KEYWORDS: *written corrective feedback, annotations, learner corpus, error analysis.*

MOTS-CLÉS: *rétroaction corrective écrite, annotations, corpus d'apprenants, analyse d'erreurs.*

TAL. Volume 57 – n° 3/2016, pages 119 à 142

1. Introduction

In an increasingly technologically-mediated world where paper-and-pencil tasks are frequently superseded by CALL applications, language teaching practitioners and researchers face new challenges related to the teaching of writing. The use of standard word-processing technology that affords tracking of changes, comments and annotations may be useful to some extent; however, annotators designed specifically with the teaching of writing and providing feedback in mind are emerging as convenient and necessary tools that offer new opportunities to both teachers and learners. In addition, such tools may incorporate analyses of learner errors that can be helpful not only to teachers and researchers but also to writers themselves, in their journey to independent and competent life-long language learning and L2 use. This is especially true in the context of increasing class sizes and higher teacher-learner ratios (García-Yeste, 2013), which in turn necessitate higher streamlining and efficiency in classroom management practices as well as fostering higher levels of learner autonomy, without compromising quality of instruction and student achievement.

In this article, we present an in-house pilot project for the development of an error annotator, *MyAnnotator,* conceived by a team of applied linguists and computer engineers at a large North American university. The initial prototype is complete and is currently undergoing testing and validation, to be followed by a release for use in language classes and one-on-one tutoring within the university community and beyond. The overarching goal of the project is to facilitate language teaching and learning; we also highlight some potential applications based on Natural Language Processing (NLP) and data-driven learning (DDL). The tool is conceived as open access and as such has the potential of benefitting a wide range of end users.

We begin by providing some background on written corrective feedback (WCF), highlighting some controversies in the literature with regard to the usefulness and types of feedback. Then we continue with an overview of electronic tools that can be used in the teaching of writing, including editors, correctors and annotators, followed by a presentation of the features and affordances of *MyAnnotator*. In the end, we explore avenues for further development of the tool and comment on its potential applications in natural language processing and data-driven learning.

2. Background on written corrective feedback (WCF)

Various aspects of written corrective feedback in language teaching have been discussed in the literature over the past thirty years. A debate that has by now become widely-known among L2 writing researchers and teachers was spurred by Truscott (e.g., 2007, and earlier work) who made a case against grammar correction in L2 writing classes, arguing that, for both theoretical and practical reasons, such correction should be expected to be, and is in fact, ineffective. In addition, Truscott has maintained, based on critical literature synthesis and meta-analysis of various

qualitative and quantitative research studies, that grammar correction, and more generally correction in writing, has a small estimated negative effect on learners' accuracy levels in writing.

Truscott's claims have been countered by Ferris (e.g., 2011, and earlier work) and a number of other researchers (e.g., Chandler, 2004; Ellis, 2009; Van Beuginen *et al.*, 2012) who have advanced various arguments in favour of WCF. However, as many of these authors have pointed out, approaches to corrective feedback may vary along multiple dimensions and relatively little agreement exists on which particular types of feedback are most effective and most useful to learners.

In terms of a typology of WCF, some of the major categories that have been identified include focused versus unfocused feedback, direct versus indirect feedback, and metalinguistic feedback. The focused versus unfocused feedback dimension refers to the degree of comprehensiveness of WCF. Focused or selective feedback represents an approach where only certain errors, based on underlying pedagogical considerations, such as susceptibility of error to treatment, frequency and seriousness of error, as well as the learner's level and cognitive overload, are targeted and corrected by the teacher. Unfocused feedback, on the other hand, constitutes a comprehensive WCF strategy where all existing errors are targeted and corrected. Bitchener (2008), Ellis *et al.* (2008) and Sheen (2007), among others (a.o.), have provided theoretical and practical arguments in favour of focused WCF, while Ferris (2010) and Stroch (2010) have raised some concerns with regard to targeting only specific errors.

Moving on to the difference between direct versus indirect feedback, the former amounts to identifying learner errors and simply providing the correct form for them, which is perhaps the most traditional way of providing feedback; the latter approach, however, assumes that direct error correction may not be processed or internalized as deeply by learners and thus only provides indirect indications or identification of errors, encouraging learners to look for the correct forms themselves. There are different types of indirect WCF varying by degree of explicitness. For example, simply flagging an error by underlining it is less explicit than assigning different colour codes to different types of error, which is in turn less explicit than providing written codes or abbreviations for different error types, or providing full-fledged explanations of errors (see also description of metalinguistic feedback below). Once again, arguments for both direct and indirect WCF have been discussed in the literature, without necessarily offering definitive answers as to which type is superior (for a sample of perspectives see Bitchener and Knoch, 2008; Chandler, 2004; Ferris, 2010; Lalande, 1982; and Van Beuginen, 2012).

Finally, metalinguistic feedback is another type of WCF where the teacher provides an explanation (cognitive/rule-based) or examples of correct form or usage. This can be achieved in various ways. For example, a teacher can use a list of codes or annotations indicating types of errors (e.g., tns=tense; wc=word choice; wf=word form; art=article, etc.). Such lists can be extensive and specific with regard to error types (e.g., pst=past simple tense error; ft=future tense error, etc.), or shorter and more general (e.g., sem=semantic error, syn=syntactic error, etc.), and may depend

on teachers' pedagogical preferences and learners' proficiency level or learning style. Another example of metalinguistic feedback would be annotating errors in a text and providing a reference source with grammatical and/or other information by including links to outside resources for further explanations, examples of usage, or practice exercises on specific language features. As with the other categories of WCF types, researcher and teacher views with respect to metalinguistic feedback vary. Typically, proponents of more analytical/linguistic approaches to the teaching of L2 writing might view metalinguistic feedback favourably, while proponents of more strictly communicative approaches may have reservations about such feedback. Research on the effectiveness of this type of feedback is also generally inconclusive (see Lalande, 1982; Sheen, 2007; Ferris, 2011; Robb *et al.,* 1986, a.o.).

To conclude this section, having offered a brief overview of some of the literature on WCF, it is important to point out that this topic fits within more general and comprehensive issues of complexity, accuracy and fluency (CAF) in language teaching and learning (see Housen and Kuiken 2009 for a discussion). Within this broader context and based on our own pedagogical beliefs and practice, we take a relatively agnostic stance with regard to the general debate in favour or against WCF in L2 writing; we lean towards a moderate position stating that some degree of WCF is necessary and beneficial in the process of L2 writing development. We also acknowledge that while various theoretical proposals and research findings may apply on both ends of the spectrum, WCF is an important part of teacher and learner realities, and is as such a necessary component in most instruction. This need is strengthened further by both institutional and student expectations in the teaching and learning of writing (Ferris, 2011; Leki, 1991; Yeh and Lo, 2009; a.o.). Therefore, in developing *MyAnnotator*, we did not commit to a particular stance on the amount and type of feedback that needs to be provided and leave this choice to practitioners themselves, who are best positioned to make informed decisions with regard to WCF based on their personal beliefs, analysis of learner needs and preferences, and institutional requirements. Our goal in creating our free tool was to remain as flexible as possible and provide a wide spectrum of accessible and customizable WCF solutions to the end user. In the next section, we focus more on this aspect by providing an overview of technology-mediated WCF, e-tools currently available, and the related feature and affordance desiderata.

3. Technology-mediated WCF

3.1. *Technology to optimize the essay marking process*

"Marking" essays[1] can be a tedious task for language teachers which technology can optimize (Garcia-Yeste, 2013). Technology can indeed increase the efficiency of providing WCF to learners by systematizing the process, namely by making some of its components reusable. In doing so, technology can also help better control the consistency and the quality of WCF delivered to learners; it can increase its reliability and, eventually, contribute to the development of "sustainable feedback practices" (Carless *et al.*, 2011). This becomes an even more valid argument in 21st century classrooms, which tend to: increase in size (Garcia-Yeste, 2013); be hybrid in terms of modes of delivery (face-to-face and online); and have moved on to technology-supported teaching and learning environments. Several studies have put forward the advantages and effectiveness of technology-enhanced corrective feedback in language learning (e.g., Nagata, 1997; Heift, 2004), pointing out in particular learners" greater and deeper involvement in the text revision process (Tuzy, 2004; Hamel, Séror & Dion, 2015). Technology facilitates the delivery of multimodal, adaptive, individualized feedback in flexible ways: explicit-implicit, detailed-synthesized, local-global, (a)synchronous, etc. (Lai, 2005).

From a language teacher's standpoint, marking essays is a complex activity, a pedagogical intervention made of several tasks that teachers usually perform sequentially, i.e., one learner essay at a time. Here is a list illustrating the pedagogical tasks (T) performed by language teachers when marking essays:

1. Throughout this article we use *essay* as a general term to refer to any type of text produced by a language learner as part of a writing assignment or activity.

T1. *Devising essay marking rubrics* and in particular, *text (error) annotation schemes* (structured tag sets and tag definitions);
T2. *Collecting and gathering learners' texts to mark*;
T3. *Analysing learners' texts* (reading, attending to meaning and form, looking for and diagnosing errors);
T4. *Annotating learners' texts* (marking texts, inserting annotations);
T5. *Providing comments* (explanations, answers, models, examples);
T6. *Suggesting remediation actions* (revisions, post-editions, exercises, etc.)*;*
T7. *Curating and suggesting further* (linguistic, pedagogical, etc.) *resources*;
T8. *Compiling statistical information* (total number of words, sentences, errors, most frequent errors, etc.);
T9. *Grading performances* (comprehensive scores, summative comments, etc.);
T10. *Sending marked essays back to learners*.

Table 1. *Pedagogical tasks performed by language teachers when marking essays.*

Technology should enable the organisation of these pedagogical tasks into a "system" flexible enough to adapt to specific teaching and learning needs and contexts. It should support and lighten most, if not all of these pedagogical tasks, in particular those that are repetitive and time consuming. Whilst teachers want to remain the experts in charge of diagnosis (T3) and feedback (T5, T6, T7, T9), they could turn to technological help for: annotating (T4), compiling error statistics (T8), being able to keep, retrieve and reuse feedback (T1), handling text files (T2, T10). As we will show later on, we are proposing an NLP-based tool to support these tasks in particular.

Below (Figure 1) is an example of a text produced electronically by a learner of French that has been marked by a teacher. Using a word processor, she has recuperated the learner's text (T2) and annotated it (T4), using bold faced characters to highlight the errors as well as opening and closing brackets to delimit their scope. She has inserted tags (T4) to provide an indication of the error type as an outcome of her error diagnosis (T3).

Quand j'étais petite <,et **<4.2>** je ne savais pas conduire, <j'ai utilisé**e <3.6>** ma bicyclette pour aller chez mes amis et à mes leçons de piano. Parfois, il m'est arrivé de tomber de ma bicyclette et j'ai <**obtenu** une blessure**<8.5>** au genou**.** Cependant, j'ai eu <**la** courage**<2.2>** de continuer à <**monter** à bicyclette**<8.5>**.

Figure 1. *Example of a marked essay.*

Prior to inserting annotations in the learner's text, the teacher would have devised an error annotation scheme (T1) which she would have made accessible to her learners, as a key to the annotation tags inserted manually in their texts (T5). Here are some rubrics of her error annotation grid (Table 2) that detail the tags inserted in the learner's text above (Figure 1):

…
2- GROUPE DU NOM
- 2.1 GENRE/NOMBRE [intrinsèque/usuel]
- 2.2 ACCORD : NOMBRE
- 2.3 ACCORD : GENRE
- 2.4 PRONOM
- 2.5 COMPLÉMENTATION [du nom/de l'adjectif]

3- GROUPE DU VERBE
- 3.1 ACCORD : SUJET
- 3.2 CONJUGAISON
- 3.3 AUXILIAIRE
- 3.4 PLACEMENT : PRONOM
- 3.5 PARTICIPE PASSÉ
- 3.6 ACCORD : TEMPS
- 3.7 COMPLÉMENTATION [du verbe]

4- MOT-OUTILS
- 4.1 PRÉPOSITION
- 4.2 CONJONCTION
- 4.3 PRONOM RELATIF
- 4.4 CHARNIÈRE de PHRASE
- 4.5 CHARNIÈRE de PARAGRAPHE

…
8- LEXIQUE
- 8.1 CHOIX [mot inapproprié]
- 8.2 ANGLICISME/CALQUE [mot inexistant]
- 8.3 NIVEAU DE LANGUE [registre]
- 8.4 RÉPÉTITION [suremploi]
- 8.5 EXPRESSION(semi-)FIGÉE [collocation/idiomatique]
- 8.6 SUPERFLU [mot inutile]

Table 2. *Rubrics of an error annotation scheme.*

The teacher might then have compiled manually – perhaps helped with the search and find function of MS Windows – the total number of errors found in the text and provided an account of error frequencies by types (T8), and given further scaffolding feedback (T6, T7). She might also have provided a summative score for the essay and overall qualitative feedback (T9). She then would have saved the annotated file with a new name and returned it to the learner via email or the course management platform (e.g. Blackboard Learn) (T10).

Let us look at electronic tools that can support these tasks involved in the essay marking process, with the aim of increasing its overall efficiency (i.e. aiming at a better quality process: less effort and time) and effectiveness (i.e. aiming at a better quality outcome: more consistency, reusability).

3.2 *Digital tools to support WCF*

Some generic tools that are not specifically designed for (pedagogical) WCF, have nonetheless features that are useful for WCF, such as:

– the "review" features of text *editors* (e.g. MS Word) that enable teachers to insert comments, make and track changes in learners' texts, compare earlier and newer text versions, etc. With online editors (e.g., Google Docs), teachers can provide WCF synchronously, as learners are writing their texts. Such applications also facilitate and promote collaborative writing as well as dynamic and dialogic collaborative WCF (see Slavkov, 2015 for a recent overview). In addition to review features, Krajka (2002) highlights standard features of text editors that are useful for WCF, including font formatting and effects (e.g. strikethrough, bold, colour, etc.), inserting (voice) annotations, endnotes, bookmarks, auto-texts, multimedia objects (e.g. audio, image) and hyperlinks;

– the "error correction" features of *spell and grammar checkers*, mostly found integrated in editors (e.g., MS Word) or as stand-alone applications (e.g., Antidote), which can be to some degree useful for automatic detection of errors in learner texts and for obtaining error statistics. Other automatic error correction methods are discussed in Leacock *et al.* (2010). However, since their error diagnosis is limited in scope and not always reliable, and since corrections are only suggestive, teachers tend not to use them for essay marking. Having said this, correctors, and in particular grammatical analyzers that diagnose errors without correcting them (e.g., Bon Patron, http://bonpatron.com) can be valid tools to support L2 learners during the composition process, and to test their knowledge and critical thinking (Hamel, 2008);

– the "markup" features of *annotators* (e.g., Diigo (https://www.diigo.com/), Sacodeyl (http://www.um.es/sacodeyl/en/pages/software.htm); BRAT (http://brat.nlplab.org/index.html) which would enable teachers to tag learners' texts with annotations and create searchable text corpora/databases. By using text

readers (e.g., Adobe Reader), teachers can also markup learners' texts (.pdf) with annotations (e.g., highlights, post-it with comments, etc.).

Since we discuss the development of an error annotator, below we focus on such annotators in more detail, with the goal of identifying their advantages and inconveniences for WCF, as well as their useful specifications, i.e. features.

Annotators

Within hypermedia systems, annotations are generally understood as metadata (labels, content nodes) attached to text, image, or other data that serve referencing purposes (cf. Wikipedia). They are "notes a reader makes to himself/herself … a natural way to record comments and ideas in specific contexts within a document" (Yeh & Lo, 2009: 883). In the context of technology-mediated WCF, digital annotations can act as interactive cognitive tools to trigger "noticing" and eventually push learners toward producing "comprehensive language output" (cf. Chapelle, 2001; Long, 1996, a.o.). Annotations can help learners link prior and new knowledge, make internal and external connections as well as develop and consolidate metacognitive and metalinguistic awareness (Yeh & Lo, 2009). Annotations can also enhance access to external information and facilitate navigation to and from documents.

Annotators are essentially digital text *referencing tools*. They allow users to curate and build collections/libraries of electronic/online documents and to organize/classify/archive these with the help of annotations inserted in each document in order to better read, understand, remember, and go back to their content.

These include tools such as annotateit.org, diigo.com, etc. that serve generic referencing purposes. Videos can also be annotated with screen capture tools like Snagit (https://www.techsmith.com/screen-capture.html) and Jing (https://www.techsmith.com/jing.html).

Social bookmarking annotation tools such as Diigo allow users to share their personal annotated libraries and to collaborate in building and annotating collections of online documents (databases, knowledge repositories). In Diigo, annotations take the form of bookmarks, tags, highlights, sticky notes and screenshots. With an open-source framework like BRAT (Stenetorp *et al.*, 2012), annotation schemes can be defined for various purposes, including essay marking (Cf. Kutuzov & Kuzmenko, 2015), while texts can be marked up online by several users at the same time.

Some advantages of generic annotators for WCF are as follows: online availability and shareablility; easy curation of documents; possibility to create structured and searchable databases of texts. Some drawbacks of generic annotators for WCF are as follows: tricky annotation process; annotations are not structurable and not recyclable (except for BRAT, were annotations can be retrieved as separate objects from the annotated texts); no annotation statistics; no options to perform searches on annotations themselves.

3.3 *Error annotators*

Error annotators are electronic tools that have been specifically designed to support WCF in the context of teaching and learning second language writing. To our knowledge, there are few such applications. Early initiatives include that of Dagneaux, Denness and Granger (1998), who propose a Computer-aided Error Analysis (CEA) system designed to facilitate the process of error tagging of learner corpora. A comprehensive set of error types was devised for this system, as well as MS Word macros to facilitate the insertion of error tags into learner texts (for an overview of error annotation schemes and their implementation, see Lüdeling and Hirschmann, 2015, a.o.).

Recent initiatives have focused on the development of semi-automatic error annotation systems. Andersen (2011), for instance, details a project investigating the potential of pre-annotating automatically recurrent errors detected in learner corpora "to increase the productivity in the task of error annotation" (p. 17). Kutuzov & Kuzmenko (2015) also describe "a framework integrating morphological analyzer, spellchecker and Web annotation tool in order to pre-annotate learner English texts with possible errors". Although we recognise the value and the potential of such initiatives, automatic error detection falls outside the scope and the priorities of our research which concentrates on using NLP to optimize other pedagogical aspects of the essay marking process, as will be indicated in section 4.4 below.

Two error annotators specifically designed for CWF include WRITE (Yeh, Lo & Huang, 2008; Lo, Wang & Yeh, 2008; Yeh & Lo, 2009) and Markin (Holmes, 1996; Krajka, 2002; Thomas, 2004; Byrne, 2007; Garcia-Yeste, 2013). Below we offer a brief overview of these two tools to provide context for the error annotator that we are developing.

WRITE

Lo, Wang and Yeh (2008) describe what they refer to as "a Web-based online corrective feedback and error analysis tool called WRITE (Writing Revision Instrument for Teaching English)", a system enabling teachers to annotate learner essays online. WRITE's system architecture consists of several components: an error annotation editor (tools, tags, comments), a database management system (annotations and documents), a composer (error correction), an annotation query system, an error analyzer, and a viewer (annotated text, error statistics). The error annotator interestingly enables the teacher to "pencil mark" learners' texts once error tags with comments have been inserted. The error analyzer is able to provide statistical results on error distributions within a single learner's text or a collection of learners' texts (from the same learner, a group of learners, etc.). The main argument put forward by Yeh and Lo (2009) for their *Online Annotator for EFL Writing* (an evolved version of WRITE) is "that online annotation functionalities for manipulating, rearranging, search, displaying and sharing annotations can be used to support EFL error correction and corrective feedback, especially the collaboration between teachers and students outside the classroom" (p. 883).

Markin

Markin (https://www.cict.co.uk/markin/), developed by Martin Holmes in the mid-nineties (Cf. Holmes, 1996), has since evolved to Markin 4, a commercial Window program (with a free restricted version) designed to allow teachers to mark learners' texts that have been submitted electronically. It uses a three-step essay marking process, best summed up as: 1) *importation*, 2) *annotation*, and 3) *exportation*, as explained below.

1) The teacher imports the document he/she wishes to annotate, as Markin is otherwise unable to use the File|Open command to access files that are not native to it (.MRK extension), or are not html files. Text (.TXT) and rich text format (.RTF) are compatible formats; MS Word files (.DOC, .DOCX) are not.

2) After a text file has been imported, the text appears in the main window and the teacher can use the button bar to insert annotations as required. Advanced features also allow the user to create snippets, thus allowing the teacher to reuse and share (with different users of the software programme) annotations (tags and comments) later on, in other text files. Additionally, Markin also has a statistical compilation tool, thus allowing the teacher to keep track of occurrences of various errors, comments and feedback comments. Errors tagged within a text, as well as within a collection of texts (collapsed and imported into Markin as one document) can be tracked and compiled.

3) Once the text annotation is complete, the modified document can be saved, either as .MRK or .HTM file. The student can recuperate this annotated version of her text which will display all marking, comments, feedback and error statistics.

Markin was used integrated into a learning management system (LMS) in a project described by Byrne (2007) which facilitated peer collaboration. It builds on the writing environment approach taken by Wibles *et al.* (2001) which namely allowed for the automatic archiving into searching database of learner annotated texts, error tags and teacher feedback comments.

To summarize, so far we have argued that technology can support the essay marking process and targeted specific tasks of this process that can be optimized, one of them being annotating learners' texts. We have seen that generic tools such as text editors and annotators can handle certain aspects of this task but dedicated tools such as error annotators offer features that cater to a more complete set of teacher needs in terms of providing WCF. We have presented two such error annotators detailing their specific features. Before we proceed to describing the annotator developed by our team, we briefly focus on a general list of desirable features for a technology-mediated WCF tool.

3.4 *Desirable features of an "ideal" technology-mediated WCF tool*

Based on the above literature review on WCF, the tasks involved in the essay marking process and the features of some electronic generic and dedicated tools that can support WCF, we can derive a set of "desiderata" guiding the development of our proposed technology-mediated WCF tool. Three types of functions are core in this "ideal" tool: *annotation, statistics,* and *management.* Their targeted specifications are listed below.

– *Annotation functions:* to create, customize, insert, save, modify, retrieve and share annotations (with related comments and feedback, resources, and information); to enable multimodality and external links.

– *Statistics functions*: to count (error) annotations (in total and by types) and calculate error-word ratio for single or multiple texts; to provide linguistic information, including word frequencies, vocabulary frequency bands, length of text and sentences, number of sentences and paragraphs, verb-word ratio, etc.

– *Management functions*: to handle the import and export of learners' (annotated) texts; to collect and store learners' texts into a searchable database; to produce text concordances from annotation tags and word queries; to enable learners to view and revise (collaboratively) their annotated texts; to recognize multiple languages (and characters).

We believe that NLP can enhance these functions, in particular the statistics (see section 4.4 for more details). Part of speech taggers, namely, which are robust and therefore can be used reliably enough to automatically annotate learner texts with linguistic information (Cf. Meurers, 2015). The outcome, tagged learner texts, can serve pedagogical purposes (Hamel, 2008), namely that of contributing to the construction of L2 writers' profiles, informing word-frequency, lexical and morpho-syntactic category distribution.

We now move on to our own tool for technology-mediated WCF which we have called *MyAnnotator* which features the integration of some of these "ideal" functions.

4. MyAnnotator

4.1. *Overall system architecture*

Our new tool, *MyAnnotator*, is a downloadable software program that has two interface modes: teacher and student views and five main functionalities: 1) error annotation management, 2) error statistics, 3) NLP-enhanced statistics, 4) error search engine, and 5) file management. While other annotators often have such features, we believe that one of *MyAnnotator's* innovative contributions is the NLP-enhanced statistics (see sections 4.3 and 4.4 below).

4.2. *Teacher view*

In this interface mode, teachers can define and customize their error annotation scheme and their comments by creating a new error type template or by editing an existing one that they (or someone else) created previously. As shown in Figure 2, teachers can create any number of error types by defining the error names, the annotation colors, comments (brief error description), and error weights/points in a template. Once defined, a template can be saved, and opened or modified later. The templates are stored in a default file folder, for quick retrieval, unless users specify another folder.

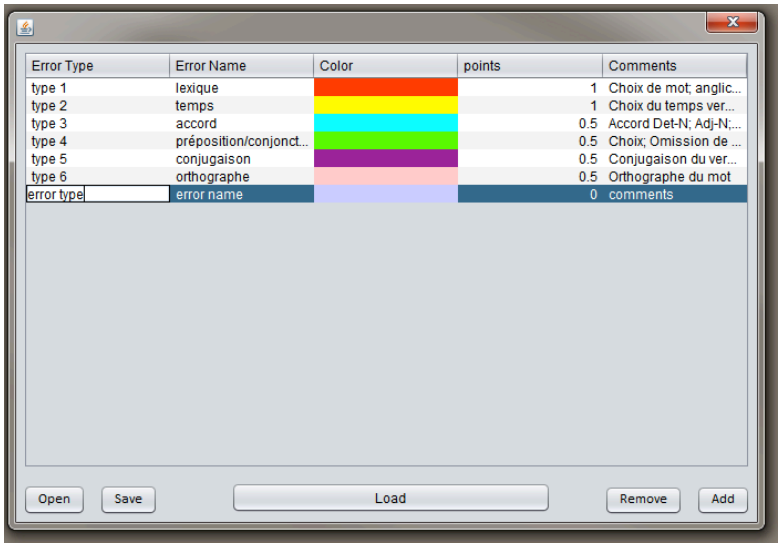

Figure 2. *Error annotation template being defined with* MyAnnotator.

Figure 3 shows the main window of our tool, which is divided into four areas: the top area is the annotation panel; the center is a text panel where the text to be annotated is displayed; the bottom area is the overall feedback panel for general comments which can be edited; the right-side area is the annotation comment panel where teacher can provide additional feedback on the error and further remediation. If users click on an annotation, the corresponding comments will be displayed in the upper part of the latter panel.

After an annotation template is loaded, its metadata is displayed on the annotation panel. An item in the panel represents an annotation type. An annotation item has four fields: the error type, the error name, its color and its weight (number of points). In the text panel, the text loaded can be annotated according to the annotation selected from the annotation panel. To maximize usability, only two steps are needed in order to annotate a sentence and add comments: 1) select a type of annotation from the annotation panel; 2) select with the mouse the part of the text to be annotated. The annotation (highlights) and the comments will be added

automatically. Errors can overlap, as long as each error is validated so it can be statistically counted. Visually, only the last colour applied in the text will show, but on browse over, the student will be able to read the comment associated with each error. Further development could include the possibility to also underline errors and link them visually, as does the annotator BRAT.

Additionally, the user interface (UI) can be translated into up to six different languages (English, French, Spanish, German, Chinese and Japanese), to offer multi-lingual support for different languages preferred by the teachers or the students.

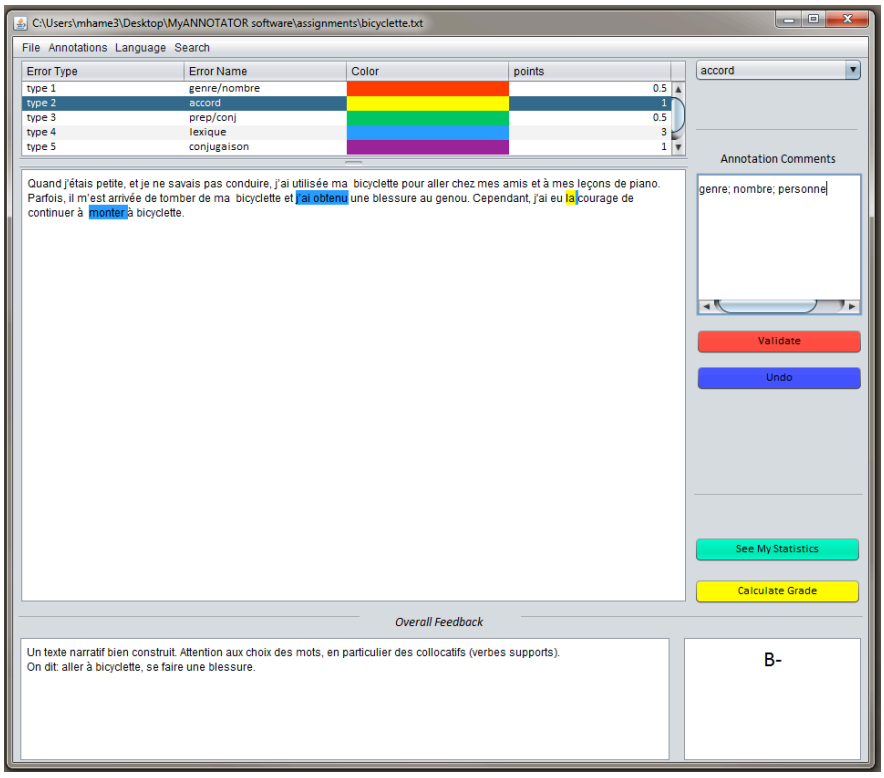

Figure 3. *Learner text being error annotated with* MyAnnotator.

4.3. *Student view*

MyAnnotator provides functionality for students to read and revise their essays. The student view mode is similar to that of the teacher view. Students need to load the error template that is created by the teacher and load their text file to see their corrected essay. The annotation panel shows all the information about the errors (error type, error name, color, and weight). The annotation panel lists all the errors the student has made and how many points they have lost for each error type (points

can also be positive). In order to see the comments, the student just needs to click on the annotated area. Then the comments are displayed in the error description area; the corresponding error name is shown in the box above the error description panel. The general comment is shown in the overall feedback panel. The main text panel is editable for students, so that they are able to revise their text dynamically, unlike with Markin, for instance, which outputs a static html file that students can view but not edit. They can also export their marked essays. Following text revision, their grade can be (optionally) recalculated by clicking on the Calculate Grade button, a useful aspect for a process approach to the teaching of writing (see also conclusion section). Clicking the "See my statistics" button, students can access their statistics, as shown in Figure 4.

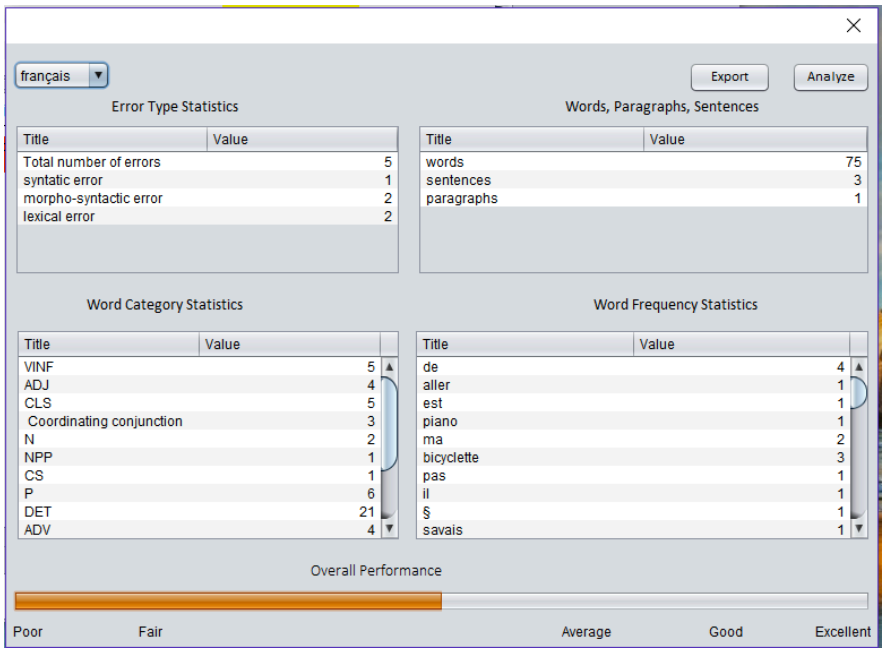

Figure 4. *Error and NLP-enhanced statistics with* MyAnnotator.

4.4. *NLP-enhanced statistical analyses*

MyAnnotator is capable of displaying four main types of statistics, as illustrated in Figure 4: 1) word category list; 2) error statistics; 3) words, sentences and paragraphs statistics; 4) word-frequency list. So far, these statistics are based on the output of a part-of-speech (POS) tagger that uses the PennTreeBank set of 45 POS tags (Marcus *et al.*, 1994). We needed to select a robust POS tagger, from the many choices available. We integrated into *MyAnnotator* the Stanford POS tagger (Toutanova and Manning, 2000). The tool achieves the best performance on the *Wall Street Journal* standard test set when compared to other existing POS taggers

and we also found it to work fast. The POS tagger is prone to errors since, as most POS taggers, it is trained on carefully edited newspaper text, and even more so due to the text with learner's errors (Van Rooy & Schäfer, 2002). Nonetheless, it proved useful in our preliminary investigation. Another advantage of the Stanford POS tagger is that it supports up to seven different languages, including the six we needed.

NLP-statistics are language dependent. The language identification (LID) routine is executed on the text before calculating POS statistics. If the language identified is wrong, users can still choose manually a language to analyze. It is important to have the correct language in order to execute the POS tagger for that language. We investigated several LID methods and tools (Gang *et al.*, 2004; Baldwin and Lui, 2010). They vary by the number of languages they recognize and by the minimum length of the text required to achieve good results. In some cases, the language can be detected directly from the character encodings. We selected Apache Tika (Mattmann and Zitting, 2010) because it has good performance even for short texts, according to a comparison of LID tools by Baldwin and Lui (2010) and it worked well on the texts that we experimented with. The users are able to export the statistics in order to do further analysis (in .csv files). The four types of statistics are saved into a single .csv file, one item per line.

We would like to provide some brief, non-exhaustive, examples of potential pedagogical applications that teachers can explore based on some of the NLP-enhanced statistics described in this section. For instance, since (academic) writing assignments often include specific requirements about length and paragraph structure, the words, sentences and paragraphs statistics can be used to verify if a learner has met the formal requirements of a particular assignment (top right window in Figure 4). Of course, this simple function is available in various general text-processing software applications, but we also integrated it into our annotator in order to provide convenient access to this information for the human annotator (teacher) during the actual WCF process (see Folse and Pugh, 2015).

The word-frequency statistics (bottom right window in Figure 4) can also be used with pedagogical purposes, especially when the former interact with specific learner errors. To provide just one example, in our experience English language learners (with various L1 backgrounds) when writing academic texts often use the noun *research* in the plural (i.e. *researches),* even though typically this noun is only used in the singular in English. The frequency of occurrence of such errors can be easily traced through a word-frequency check (bottom right window in Figure 4) where the frequency of occurrence of the words *research* and *researches* will be displayed. This can be done for a single assignment, a set of assignments for a specific learner, or for multiple learners' (e.g. a whole class) assignments taken together (see also section 4.5). Words that are repeated too often in a text and could be replaced by equivalents (e.g. synonyms, pronominal and lexical anaphora) can also be detected that way. Based on such statistics, specific targeted feedback can be offered to individual learners or to whole groups or classes. Learners can also be asked to perform their own error analyses based on specific criteria provided by the teacher and thus see for themselves how often they make a given error, and be able

to focus their efforts on eliminating it, especially if it has high rates of occurrence. Such pedagogical practices are compatible with the well-known concepts of noticing and uptake in second/foreign language teaching and learning (for an overview, see Nassaji and Fotos, 2010, a.o.) and related to the accuracy dimension of CAF-based pedagogy.

Basic POS-tagging statistics can also be helpful to teachers, learners and researchers (e.g., Granger *et al.,* 2007, a.o.) and related to CAF. For example, some researchers have found that various types of adverbs can be underused by language learners (e.g., Zinsmeister and Breckle, 2012; Chen, 2006) which affects complexity. While our tool at this point does not offer sophisticated differentiation of adverbs in terms of semantic and discourse functions, etc., the basic POS tagger output that the current version of the tool provides can be used by teachers to generate statistics on noun or verb density, use of specific verb tenses, and so on. The tagger can also be used in revision tasks where teachers ask learners to submit a second draft of a writing assignment with an increased general number of adverbs, for example. POS-tagging statistics also provide learners with occasions to focus their attention and revision efforts on specific linguistic forms (e.g. past verb tenses; relative pronouns, etc.) in their text. Some further ideas for development of the tool with regard to more sophisticated NLP and CAF measures are mentioned briefly in the conclusion of this article.

4.5. *Error search engine*

Oftentimes teachers want to find some examples from a large number of student files, and it is difficult to go through them one by one. With our search engine, teachers can search any file by specifying the essay directory and the error type or error name. After results are returned, all files in the specified folder that contain those types of errors are listed in the left panel (under error name). By clicking on a filename on the list, the original essay content and the highlighted error annotation will be displayed in the right panel. An example is shown in Figure 5.

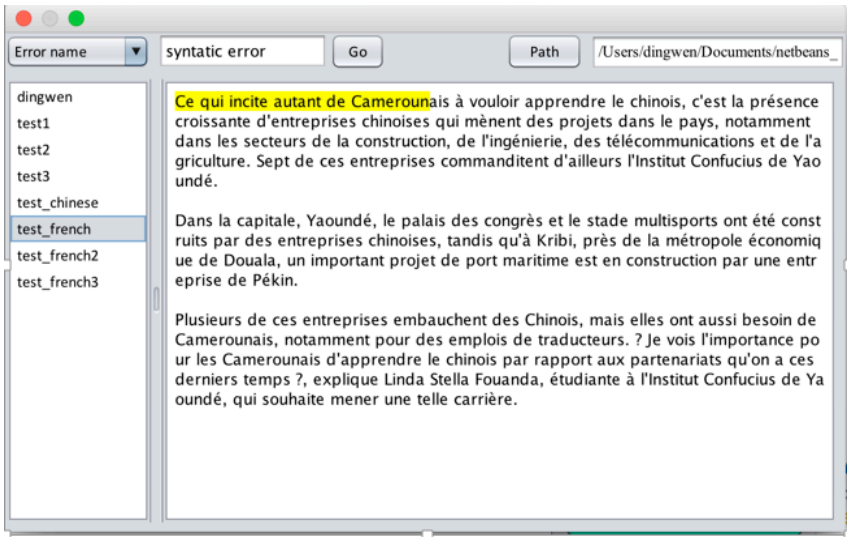

Figure 5. *The error type search engine (test data).*

4.6. *File management*

Loading, storing and finding templates, text files and statistics files, in several different file formats, can be confusing and time-consuming for users. To manage these, in our tool, the different types of files are automatically stored in different dedicated default folders. For every store or load operation, the default folder path will be shown and the files will be easy to locate. Users can find and open these default folders from the file menu. They can load texts, open the default statistics file directory, the template file folder, and so on.

5. Discussion on pedagogical affordances of *MyAnnotator*

One of our goals with *MyAnnotator* was to develop a digital tool that can facilitate and ultimately optimize some of the tasks teachers perform when marking essays; another goal that we had set was to provide language learners with opportunities to analyse their errors by number and type (based on teacher annotations and NLP) and thus engage in autonomous life-long language learning. We have seen earlier in this article that teachers can scaffold the writing process of their language learners by resorting to various forms of written corrective feedback and various techniques to deliver effective WCF. We have adopted a flexible stance in terms of most effective types and modes of feedback in light of our literature review on WCF and in order to allow for diverse language teaching styles, needs and contexts. Hence, with *MyAnnotator* teachers are able to provide direct and indirect, explicit as well as implicit types of feedback. In order to give direct

feedback, the teacher would, for instance, declare an annotation of type 1 and name it "correction", while in the text window she would correct the text directly and highlight the correction. As far as implicit feedback is concerned, she would do the same without correcting the text but providing error codes and/or metalinguistic information, as desired. As annotations have customizable descriptions, they can be formulated in terms of metalinguistic feedback or as more communicative messages, depending on pedagogical beliefs.

Coming back to the tasks involved in the process of marking essays identified in section 3.1 of this article, it is evident that *MyAnnotator* can support (with various degrees of flexibility) all those initially targeted, that is, tasks that technology can optimize and render more efficient: T1 (devising error annotation schemes); T2 (collecting and gathering learners' texts); T4 (annotating learners' texts); T5-7 (providing feedback comments and other pedagogical scaffolds); T8 (compiling statistics); T10 (making annotated texts accessible to learners).

Looking at the desired specifications of an "ideal" WCF tool, as identified in section 3.4, we note that *MyAnnotator*, in its current pilot stage, has "standard" (annotation and management) functionalities that generally match those of similar commercial error annotators such as Markin, but might not yet be as robust until fully tested with users and developed further. However, we would like to point out that some of our tool's strengths reside in its NLP-enhanced statistics search functionalities and in the planned open access release of the application, which has been a core value of the project since its inception (see also conclusion section for additional information on other planned developments, such as integrating concordances, etc.).

With regard to the second goal mentioned in the beginning of this section, that is, the opportunity for students to engage in autonomous language learning, we note that the student view of our tool is compatible with this idea. As already indicated, learners have access to a dedicated interface, which is very similar to the teacher's one. Thus, in addition to being able to revise their work in the main text panel of the tool, learners can view their error statistics and export annotated essays. Furthermore, learners can self-analyse an entire collection of previously annotated essays by (re)-importing them into the tool and running searches by error types within those. A learner may be interested in his/her average error rate on a certain type of pre-defined error over the course of a semester in which they had taken a writing class, for example. That way, learners may be able to identify areas in their writing with a high concentration of errors that may need special attention versus areas where they have relatively little to improve. In our view, placing such affordances in the hands of language learners themselves contributes to the concept of autonomy of learning (e.g., Albéro and Poteaux, 2010). To our knowledge, this opportunity for extended learner self-analysis of a collection of marked essays distinguishes our tool from some of the other annotators. Moreover, we would like to point out that the teacher and student interfaces of the tool are very similar. The only function that is not accessible in the student view of *MyAnnotator* is editing the error annotations. This assumes relatively high level of teacher-learner equality and

is consistent with views of learners as active and empowered agents in the learning process.

In addition to highlighting some of the features of our tool and how they can offer various sets of affordances to both teachers and learners, and before we move on to the conclusion of the article, we would like to mention briefly how using *MyAnnotator* fits within some broader pedagogical views on the teaching of writing. Although, as already mentioned, we do not prescribe specific pedagogic approaches or interventions, we would like to point out that our tool is consistent with the process approach to writing. The process approach assumes that writing should not be regarded as a final product submitted once by a learner and subsequently reviewed by a teacher; rather, teachers are expected to foster a collaborative process that brings learners through a multitude of prewriting activities, creation and revision of numerous drafts, including rounds of WCF, editing and proofreading, and postwriting (i.e., display or dissemination). This approach has become prevalent in writing pedagogy (Weigle, 2014; see also Slavkov, 2015 for a recent overview) and is fully compatible with technology-mediated WCF. Thus, our tool can be placed and extensively used in the middle stages of the process described (i.e. revision, editing and proofreading).

6. Conclusion and future work

As mentioned at the beginning of this article, *MyAnnotator* is the result of a small interdisciplinary collaborative project between applied linguists and language engineers. The initial development and the continuous upgrades of the tool are done primarily by supervised graduate students and supported by an internal grant. Some of the challenges involved in such research and development projects combining educational and practical purposes include limited resources and generally slower progress than in the industry. One the other hand, *MyAnnotator* is conceived as an open access tool which will soon be freely available for wider testing and use. Since it includes a range of features and can be used flexibly and creatively by both language teaching practitioners and learners, it promises a comprehensive set of affordances to a broad community of users both within and beyond the boundaries of our university.

Future work includes further use of NLP in order to provide comprehensive L2 writing performance profiles based on lexical and grammatical information found in learners' texts. Tools (e.g., Lu, 2010) such as http://aihaiyang.com/software/, http://www.lextutor.ca/ and methods (e.g., Gunnarsson, 2012) to measure CAF (Complexity, Accuracy and Fluency) in learners' texts can be plugged in/applied in the future phases of development of the tool. These instruments can namely calculate the following: word distribution into frequency bands (high to low frequents, academic list words, off-list words; n-gram analysis (co-frequency); cognates (French-English); number of T-Units (tensed sentences), verb-word ratio, etc. More information on the text profile can be gathered with the help of syntactic and discourse parsers (sentence types, discourse markers, cohesion, anaphora, etc.). The

addition of a concordancer would also enhance the error search process and be in line with data-driven pedagogy (cf. Leńko-Szymańska & Boulton, 2015; a.o.).

In its current state, *MyAnnotator* is a downloadable desktop application that, as already indicated, allows for a process approach to the teaching of writing by incorporating recursive WCF in a dialogic teacher-learner or peer-to-peer relationship over the various phases of drafting, revision and editing. The tool was originally conceived with the intention of use in a particular institutional setting where it can be installed locally on existing work stations and where server storage had to be avoided for logistical reasons. Due to this, the tool currently does not run in an online environment and cannot benefit from some of the well-known advantages of dynamic collaboration and co-authorship in real time (for a recent overview see Slavkov, 2015). We do, however, recognize the challenges that this poses in a global context where technology-mediated language teaching and learning happens in interconnected and integrated online environments. The possibility of transitioning our tool to an online platform is currently being explored.

Overall, this project represents a step forward in a research and development program aimed at being able to recycle pedagogical interventions and learner data, which is at the core of a sustainable CALL practice (Caws & Hamel, 2013). An NLP-based tool like *MyAnnotator* has a strong potential to promote and enrich such endeavours.

7. References

Albéro, B., Poteaux, N., (eds.) *Enjeux et dilemmes de l'autonomie : une expérience d'autoformation à l'université,* étude de cas, Paris, Éditions de la Maison des Sciences de l'Homme, 2010.

Andersen, O.E., "Semi-automatic ESOL error annotation", *English Profile Journal*, vol. 2, p. 1-17, 2011.

Baldwin, T., Lui, M., "Language identification: The long and the short of the matter", *In Proceedings of ACL 2010*, p. 229-237, 2010.

Byrne, T. "Marrying two existing software packages into an efficient online tutoring tool", *Computer Assisted Language Learning,* vol. 20, n° 5, p. 459-469, 2007.

Bitchener, J., "Evidence in support of written corrective feedback", *Journal of Second Language Writing*, vol. 17, n° 2, p. 102-118, 2008.

Bitchener, J., Knoch, U., "The value of a focused approach to written corrective feedback", *ELT Journal,* vol. 63, n° 3, p. 204- 211, 2009.

Caws, C., Hamel, M.-J. "From analysis to training: recycling interaction data into learning", *OLBI Working Papers,* vol. 5, n° 5, p. 25-36, 2013.

Chandler, J., "A response to Truscott", *Journal of Second Language Writing*, vol. 13, n° 4, p. 345-348, 2004.

Chapelle, C., *Computer Applications in Second Language Acquisition*, Cambridge University Press, 2010.

Chen, C.W., "The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners", *International Journal of Corpus Linguistics,* vol. 11, n° 1, p. 113-130, 2006.

Carless, D., Salter, D. Yang, M., Lam, J., "Developing sustainable feedback practices", *Studies in Higher Education*, vol. 32, n° 3, p. 285-299, 2011.

Dagneaux, E., Denness, S., & Granger, S., "Computer-aided error analysis", *System*, vol. 26, n° 2, p. 163-174, 1998.

Ellis, R., "A typology of written corrective feedback types", *ELT journal*, vol. 63, n° 2, p. 97-107, 2009.

Ellis, R., Sheen, Y., Murakami, M., Takashima, H., "The effects of focused and unfocused written corrective feedback in an English as a foreign language context", *System*, vol. 36, n° 3, p. 353-371, 2008.

Ferris, D., "Second language writing research and written corrective feedback in SLA", *Studies in Second Language Acquisition*, vol. 32, n° 2, p. 181-201, 2008.

Ferris, D., *Treatment of Error in Second Language Student Writing*, Ann Arbor, MI: University of Michigan Press, 2011.

Folse, K., Pugh, T. *Great Writing 5: From Great Essays to Research,* Boston, MA: National Geographic Learning / Cengage Learning, 2015.

García-Yeste, M., "Electronic feedback: pedagogical considerations for the implementation of software", *The EUROCALL Review*, vol. 21*,* n° 2, p. 39-48, 2013.

Garg, A., Gupta, V., Jindal M., "A survey of language identification techniques and applications", *Journal of Emerging Technologies in Web Intelligence,* vol. 6, n° 4, p. 388-400, 2004.

Granger S., O. Kraif, C. Ponton, G. Antoniadis, Zampa, V. "Integrating learner corpora and natural language processing: A crucial step towards reconciling technological sophistication and pedagogical effectiveness", *ReCALL,* vol. 19, n° 3, p. 252-268, 2007.

Gunnarsson, C., "The development of complexity, accuracy and fluency in the written production of L2 French", *in* Housen, A., Kuiken, F., & Vedder, I. (eds), *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA*, Amsterdam: John Benjamins, p. 247-276, 2012.

Krajka, J., "Correcting student work with the computer – using dedicated software and a word processor", *Teaching English with Technology*, vol. 2, n° 4, p. 46-52, 2002.

Kutuzov, A., Kuzmenko, E., "Semi-automated typical error annotation for learner English essays: Integrating frameworks", *Proceedings of the 4th Workshop on NLP for Computer Assisted Language Learning at NODALIDA 2015*, NEALT, 2015, Proceedings Series 26 /Linköping Electronic Conference Proceedings, vol. 114, p. 35-41.

Hamel, M.-J. "Natural language processing tools in CALL", *In* Ayoun D. (ed.), *Studies in French Applied Linguistics*, Amsterdam: John Benjamins, p. 63-96, 2008.

Hamel, M.-J., Séror, J., Dion, C. *Writers in Action: Modelling and Scaffolding L2 Learners' Writing Process,* Toronto: Higher Education Quality Council of Ontario, 2015.

Heift, T., "Corrective feedback and learner uptake in CALL", *ReCALL Journal*, vol. 16, n° 2, p. 416-431, 2004.

Holmes, M., "Marking student work on the computer", *The Internet TESL Journal*, vol. 2, nº 9, 1996.

Housen, A., Kuiken, F., "Complexity, accuracy, and fluency in second language acquisition", *Applied Linguistics*, vol. 30, nº 4, p. 461-473, 2009.

Lai, C. "Introduction: The Role of feedback and technology in SLA", in Zhao, Y. (ed), *Research in Technology and Second Language Education: Developments and Directions*, Greenwich, Connecticut: Information Age Publishing, p. 337-344, 2005.

Lalande, J. F., "Reducing composition errors: An experiment", *Modern Language Journal,* vol. 66, nº 2, p. 140-149, 1982.

Leacock,C., Chodorow, M., Gamon, M. & Tetreault, J., *Automated Grammatical Error Detection for Language Learners*, San Rafael, California: Morgan & Claypool, 2010.

Leńko-Szymańska, A., Boulton, A., (eds), *Multiple Affordances of Language Corpora for Data-driven Learning*, Amsterdam: John Benjamins, 2015.

Leki, I., "Twenty-five years of contrastive rhetoric: Text analysis and writing pedagogies", *TESOL Quarterly,* vol. 25, nº 1, p. 123-143, 1991.

Lo, J.-J., Wang, Yeh, Y.-C. S.-W., "Writing revision instrument for teaching English", in Z. Pan *et al.* (eds), *Technologies for E-Learning and Digital Entertainment*, Berlin: Springer-Verlag, p. 1-8, 2008.

Long, M., "The role of the linguistic environment in second language acquisition", In Ritchie, W. & Bhatia, T., *Handbook of Second Language Acquisition*. San Diego: Academic Press, p. 413-468, 1996.

Lu, X., "Automatic analysis of syntactic complexity in second language writing", *International Journal of Corpus Linguistics*, vol. 15, nº 4, 2010. p. 474-496.

Lüdeling, A., Hirschmann, H., "Error Annotation", i*n* Granger, S., Gilquin, G. & Meunier, F. (eds.), *The Cambridge Handbook of Learner Corpus Research*, Cambridge: Cambridge University Press, p. 135-158, 2015.

Marcus, M, Kim, G., Marcinkiewicz M.A., MacIntyre R., Bies, A., Ferguson M., Katz, K., Schasberger, B., "The Penn Treebank: annotating predicate argument structure", *in Proceedings of the Workshop on Human Language Technology*, ACL, 1994.

Meurers, D., "Learner corpora and natural language processing", *in* Granger, S., Gilquin, G., & Meunier, F. (eds.), *The Cambridge Handbook of Learner Corpus Research*, Cambridge: Cambridge University Press, p. 37-566, 2015.

Nagata, N., "The effectiveness of computer-assisted metalinguistic instruction: A case study in Japanese", *Foreign Language Annals*, vol. 30, nº 2, p. 187-200, 1997.

Nassaji, H., Fotos, S., *Teaching Grammar in Second Language Classrooms: Integrating Form-focused Instruction in Communicative Context*, New York, NY: Routledge, 2011.

Robb, T., Ross, S., Shortreed, I., "Salience of feedback on error and its effect on EFL writing quality", *TESOL Quarterly*, vol. 20, nº. 1, p. 83-93, 1986.

Sheen, Y., "The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles", *TESOL Quarterly*, vol. 41, nº 2, p. 255-283, 2007.

Slavkov, N., "Sociocultural theory, the L2 writing process, and Google drive: strange bedfellows?", *TESL Canada Journal,* vol. 32, nº 2 , p. 80-96, 2015.

Storch, N., "Critical feedback on written corrective feedback research", *International Journal of English Studies,* vol. 10, n° 2, p. 29-46, 2010.

Toutanova, K., Manning, C.D., "Enriching the knowledge sources used in a maximum entropy part-of-speech tagger", *In Proceedings of EMNLP/VLC-2000*, p. 63-70, 2000.

Truscott, J., "The effect of error correction on learners' ability to write accurately", *Journal of Second Language Writing*, vol. 16, n° 4, p. 255-272, 2007.

Tuzy, F., "The impact of e-feedback on the revision of L2 writers in an academic writing course", *Computers and Composition,* vol. 21, n° 2, p. 217-235, 2004.

Van Beuningen, C. G., De Jong, N. H., Kuiken, F., "Evidence on the effectiveness of comprehensive error correction in second language writing", *Language Learning*, vol. 62, n° 1, p. 1-41, 2012.

Van Rooy, B., Schäfer, L., "The effect of learner errors on POS tag errors during automatic POS tagging", *Southern African Linguistics and Applied Language Studies*, 20(4): 325-335, 2002.

Wible, D., Kuo, C.-H., Chien, F.-Y., Liu, A., Tsao, N.-L., "A Web-based EFL writing environment: Integrating information for learners, teachers, and researchers", *Computers and Education*, vol. 37, p. 297-315, 2001.

Yeh, S.-W., Lo, J.-J., "Using online annotations to support error correction and corrective feedback", *Computers and Education*, vol. 52, n° 4, p. 882-892, 2009.

Zinsmeister, H., Breckle, M., "The ALeSKo learner corpus", *in* Schmidt, T., Worner, K. (eds.), *Multilingual Corpora and Multilingual Corpus Analysis*, Amsterdam: John Benjamins, p. 71-96, 2012.