universidad de león
Departamento de Ingeniería Eléctrica y de Sistemas y Automática
Departamento de Psicología, Sociología y Filosofía

Automatización del Análisis del Proceso y el Producto de la Composición Escrita Mediante Seshat Writing Research Tools

Automation of the Analysis of the Writing Process and Product Using Seshat Writing Research Tools

Tesis Doctoral

Presentada por Don Héctor Díez Caso
Dirigida por el Dr. Don Jesús Nicasio García Sánchez

León, 2014

INFORME DEL DIRECTOR DE LA TESIS DOCTORAL

El Dr. D. Jesús Nicasio García Sánchez, como Director de la Tesis Doctoral titulada *Automatización del Análisis del Proceso y el Producto de la Composición Escrita Mediante Seshat Writing Research Tools, Automation of the Analysis of the Writing Process and Product Using Seshat Writing Research Tools*, realizada por D. Héctor Díez Caso en el programa de doctorado Sistemas Inteligentes en la Ingeniería, informa favorablemente el depósito de la misma, dado que reúne las condiciones necesarias para su defensa.

Lo que firmo, en León a 31 de octubre de 2014,

El Director de la Tesis Doctoral
Fdo.: Jesús Nicasio García Sánchez

universidad de león

ADMISIÓN A TRÁMITE DE LA TESIS DOCTORAL

El órgano responsable del programa de doctorado Sistemas Inteligentes en la Ingeniería, en su reunión celebrada el día _ de _ de 2013 ha acordado dar su conformidad a la admisión a trámite de lectura de la Tesis Doctoral titulada *Automatización del Análisis del Proceso y el Producto de la Composición Escrita Mediante Seshat Writing Research Tools*, dirigida por el Dr. D. Jesús Nicasio García Sánchez, elaborada por D. Héctor Díez Caso, y cuyo título en inglés es: *Automation of the Analysis of the Writing Process and Product Using Seshat Writing Research Tools.*

Lo que firmo, en León a 31 de octubre de 2014,

Vº Bº Director Departamento	Vº Bº Secretario Departamento
Fdo.: José Luis Falagán Cavero	Fdo.: Juan José Fuertes Martínez

universidad de león

DEPÓSITO DE TESIS DOCTORAL

El Ingeniero D. Héctor Díez Caso, una vez autorizada la presentación de la Tesis Doctoral titulada *Automatización del Análisis del Proceso y el Producto de la Composición Escrita Mediante Seshat Writing Research Tools, Automation of the Analysis of the Writing Process and Product Using Seshat Writing Research Tools*, por el Director de la misma, el Dr. D. Jesús Nicasio García Sánchez, y tras la conformidad del Departamento para el inicio de los trámites,

PROCEDE al Depósito de dicha tesis en formato electrónico a través de la aplicación de depósito la Universidad de León, así como al envío de un ejemplar a cada uno de los miembros del Tribunal nombrado al efecto para su defensa.

En León, a 31 de octubre de 2014,

El Doctorando

Fdo.: Héctor Díez Caso

Solicitud **“Mención Internacional”** en el título de Doctor

D. Héctor Díez Caso con D.N.I. 71446808E, habiendo realizado la tesis doctoral titulada *Automatización del Análisis del Proceso y el Producto de la Composición Escrita Mediante Seshat Writing Research Tools, Automation of the Analysis of the Writing Process and Product Using Seshat Writing Research Tools*, dirigida por el Dr. D. Jesús Nicasio García Sánchez y cumpliendo los requisitos establecidos por el art. 15 del R.D. 99/2011 de 28 de enero,

SOLICITA:

Le sea concedida la Mención Internacional en el título de Doctor.

En León, a 31 de octubre de 2014

Fdo.: Héctor Díez Caso

SR. DECANO DE LOS ESTUDIOS DE DOCTORADO

A mis padres

Agradecimientos

A mi director, por brindarme la oportunidad de llevar a cabo esta investigación y abrir mi mente a un nuevo campo de estudio que hoy asumo como propio.

A mis compañeros y amigos, junto a quienes he navegado contra vientos y mareas y surcado las procelosas aguas de la incertidumbre sin llegar a conocer jamás los estragos de la soledad.

A Gert Rijlaarsdam, cuya extraordinaria calidad humana e impecable profesionalidad han supuesto no sólo una ayuda inestimable en esta investigación sino un ejemplo a seguir que guiará mis esfuerzos de ahora en adelante.

A Teresa Mata, sin cuyos consejos, comprensión y amistad esta tesis hubiese sido una tarea imposible.

A mis padres, cuyo apoyo incondicional e inquebrantable ánimo han inspirado en mí el aliento necesario para llevar a término la tesis que hoy les dedico desde lo más profundo de mi corazón.

Gracias a todos.

La pluma es la lengua del alma

Miguel de Cervantes

Índice general

Abstract 1

Introduction 1

Aims 5

Seshat Process: monitoring of cognitive processes 9

Seshat Product: automated essay scoring 34

Discussion and conclusion 59

1. Introducción general 67

1.1 Introducción 67

1.2 Síntesis descriptiva de los contenidos 71

1.2.1 Monitorización de procesos cognitivos 72

1.2.2 Evaluación automática de textos escritos 75

1.3 Objetivos 76

Monitorización de procesos cognitivos 83

2. Antecedentes I 85

2.1 Introducción 85

2.2 Modelos de proceso de escritura 86

2.3 Métodos de investigación 93

3. Seshat Process 99

3.1 Introducción 99

3.2 Aspectos técnicos 102

3.3 Interfaz *student* 104

3.4 Interfaz *teacher* 110
3.5 Interfaz *administrator* 117
3.6 Capacidades de análisis y gestión de resultados 117
3.6.1 Medición del uso de procesos cognitivos 118
3.6.2 Análisis evolutivo de procesos cognitivos 119
3.6.3 Estadísticos descriptivos del tiempo de respuesta ..123
3.6.4 Análisis evolutivo del tiempo de respuesta.....123
3.6.5 Descarga de resultados.....126
3.7 Validación a través de un caso de estudio real.....126
3.8 Discusión y conclusiones 131
Evaluación automática de textos escritos 133
4. Antecedentes II 135
4.1 Introducción 135
4.2 La escritura como producto.....137
4.3 Intermediación de rasgos lingüísticos de productividad y coherencia en la predicción de la calidad textual.....141
4.4 Sistemas de análisis automático de la escritura 145
4.4.1 Variables proxy 147
4.4.2 Latent Semantic Analysis.....148
4.4.3 Procesamiento del lenguaje natural 150
4.4.4 Métodos bayesianos 151
5. Seshat Product 153
5.1 Introducción 153
5.2 Protocolo de corrección 155
5.2.1 Generación de información.....159

5.2.2 Organización de la información 163
5.3 Aspectos técnicos 167
5.4 Interfaz de usuario 173
5.5 Validación 180
5.5.1 Estudio de caso: 2014 185
5.5.2 Estudio de caso: 2013 189
5.6 Limitaciones 190
5.7 Discusión y conclusiones 194
Discusión y conclusiones generales 197
6. Discusión y conclusiones generales 199
Apéndices 207
Apéndice A: Protocolo de corrección 209
A.1 Indicadores basados en el texto 203
A.1.1 Generación de información 204
A.1.2 Organización de la información 217
A.1.3 Estructuración de la información 223
A.1.4 Consideraciones adicionales 225
A.2 Indicadores basados en el lector 227
A.2.1 Textos de comparación y contraste 228
A.2.2 Textos narrativos 232
A.2.3 Textos descriptivos 236
A.3 Consideraciones adicionales 241
Apéndice B: Connexor Machinese Phrase Tagger 249
Apéndice C: Coeficiente Kappa de Cohen 253

Referencias 255
Certificados e informes internacionales 273

Lista de tablas

Tabla 3.1 Mediciones del uso de procesos y tiempo de respuesta por proceso en el curso FOCO 130

Tabla 5.1 Indicadores de generación de información.....160

Tabla 5.2 Indicadores de organización de la información..164

Tabla 5.3 Índice de acuerdo entre evaluadores en los principales indicadores del protocolo de corrección implementado por Seshat Product (N=537) 183

Tabla 5.4 Resultados estudio de caso 2014 188

Tabla A.1 Lista de perífrasis verbales 212

Tabla A.2 Clases de adverbios 214

Tabla A.3 Lista de preposiciones y locuciones 216

Tabla A.4 Lista de conjunciones y locuciones 216

Tabla A.5 Lista de pronombres personales 218

Tabla A.6 Lista de pronombres demostrativos.....219

Tabla A.7 Lista de pronombres posesivos.....219

Tabla A.8 Lista de pronombres indefinidos 220

Tabla A.9 Lista de determinantes demostrativos 221

Tabla A.10 Lista de determinantes posesivos 222

Tabla A.11 Lista de determinantes indefinidos.....222

Tabla A.12 Ejemplos de indicadores estructuradores 226

Tabla A.13 Ejemplos de indicadores conectores.....226

Tabla A.14 Lista de indicadores reformuladores 227

Tabla A.15 Lista de indicadores argumentativos 227

Tabla A.16 Leyenda de corrección manual.....247
Tabla B.1 Etiquetas morfológicas del Connexor MPT 249
Tabla B.2 Etiquetas sintácticas del Connexor MPT.....251
Tabla B.3 Otras etiquetas del Connexor MPT 251

Lista de figuras

Figura 3.1 Estructura de la base de datos.....102
Figura 3.2 Ejemplo de lista de categorías 106
Figura 3.3 Entrenamiento no supervisado 108
Figura 3.4 Ejemplo de identificación de un proceso cognitivo durante la realización el ejercicio escrito 109
Figura 3.5 Elementos del menú opciones 112
Figura 3.6 Uso de procesos cognitivos 119
Figura 3.7 Evolución de procesos cognitivos 120
Figura 3.8 Filtros de ajuste a líneas de tendencia 122
Figura 3.9 Evolución de los tiempos de respuesta.....125
Figura 5.1 Texto manuscrito analizado de forma manual ...156
Figura 5.2 Diagrama de flujo del proceso de análisis.....168
Figura 5.3 Implementación del algoritmo BMH en Java.....172
Figura 5.4 Selección de archivos para el análisis 175
Figura 5.5 Análisis en curso.....177
Figura 5.6 Resultados del análisis.....179

Abstract[1]

INTRODUCTION

The exponential growth of information available due to the widespread use of the Internet has created an increasing demand for methods of automatic analysis of natural language in modern society (Pęzik, 2010). Currently, the incorporation of computers to the field of written information processing has led to important advances in such disparate fields as economics (De Souza et al., 2007), medicine (Pakhomov, Chacon, Wicklund, & Gundel, 2011) or education (Hastings, Hughes, Magliano, Goldman, & Lawless, 2012; Magliano & Graesser, 2012). In the educational field in particular, these advances have proved to be especially significant, given that they respond to a paradigmatic change in teaching brought by the rampant technological evolution occurred during the last twenty years.

[1] The present research received funding from competitive projects awarded to Excellence Research Groups of the Regional Government of Castile and León co-funded by the European Union (GREX259; BOCyL 27 de abril de 2009), and from a MICINN competitive project (EDU2010-19250/EDUC, 2010-2013), being J. N. García the principal researcher in both projects. In addition, the present research also received funding by a competitive contract jointly funded by the Regional Government of Castile and León and the European Social Fund awarded to H. Díez Caso (EDU/1867/2009, 2009).

Abstract

It is safe to say that the rise of Information and Communication Technology (ICT) and the widespread use of internet-enabled devices over the past few years a few years has become the reality of any education system. Instant access to large amounts of information, at little or no cost, has led us to a technological revolution that has brought changes to traditional educational paradigms. In consequence, we have seen how computers have been gradually turning into an essential working tool in the classroom.

The efforts to include technology in schools are becoming more numerous and remarkable. Teaching institutions all over the world are devoting large amounts of resources to adapt to the new situation, whether by using platforms such as Moodle or WebTC or by developing their own tools tailored to specific needs (Muñoz de la Peña, Gómez-Estern & Asleep, 2012). No wonder, therefore, that increasing concern has arisen about the impact of ICT in education (Cha et al., 2011; Hernández-Ramos, Martínez-Abad, García Peñalvo, Esperanza Herrera García, & Rodríguez-Conde, 2014; Ott & Pozzi, 2011; Zhong, 2011), as well as about the development of digital skills in both teachers (Butter, Pérez, & Quintana, 2014; Koh, 2011) and students (Hatlevik & Christophersen, 2013).

Abstract

These days, the students face challenges during their academic training that differ radically from those faced in previous generations, making it necessary to instruct them in competency areas outside of the traditional curriculum such as digital learning or the search and handling of multimedia information (Trilling & Fadel, 2009). In this sense, it is undeniable that writing has assumed a leading role within this new competence framework, establishing itself as the ultimate communicative instrument in the digital world. This new reality has also fostered a complete transformation of the way students write and communicate (Herrington, Hodgson, Moran, & Eidman-Aadahl, 2009).

According to Haydel and Bucolo (2010), nowadays it is expected that students acquire skills not only in traditional disciplines such as writing research papers or argumentative essays, but also in the use of other forms of digital communication such as emails, blogs and the use of social networks, among others. Mastery of writing therefore remains a key area of the school curriculum; however, both the way the subject is taught and how it is assessed have undergone a profound transformation in recent years due to the already mentioned impact of the widespread use of computers and the Internet.

Abstract

Additionally, this new reality provides researchers with the means and the opportunity to study both the cognitive processes involved in writing and the written texts in an entirely different context. Although approaching the study of writing either as a process or a product is not new, the increase in incorporation of new technologies in the education field has prompted the development of a series of tools designed to facilitate the work of teachers and researchers in this regard. Especially relevant are the computerized monitoring of cognitive processes and the automatic analysis of writing, as the cornerstone of this thesis is the creation of separate tools to integrate these technologies as part of a common system to study both the process and the product of writing.

In the next section, a detailed description of the main objectives of this thesis is provided, as well as the causes underlying the development of all the applications of *Seshat Writing Research Tools*. This whole research, as well as all the contents presented in this report, enjoys backing from the scientific community through publications in several national and international media (Díez-Caso & García, 2011a, 2011b, 2011c, 2013, under review a, under review b; Díez-Caso, García, García Martín, & Álvarez, 2013).

AIMS

By making use of the available technological improvements in the field of distance education and automatic text processing, the aim of this project lies in the creation of a series of tools devised to study both the cognitive activities involved in writing and the final product obtained as a result of this process. The suite of applications resulting from this effort, called *Seshat Writing Research Tools*, consists of two basic components strongly interrelated, *Seshat Process* and *Seshat Product*; both of them mutually complementary and designed to satisfy each of the afforementioned writing research approaches.

In relation, therefore, to the first study presented in this thesis, the monitoring of cognitive processes, the primary objective is to create an application designed to implement the Kellogg's triple task procedure in an unsupervised environment and using visual interruption signals instead of the classical auditory ones. This application, called *Seshat Process*, is also intended to take advantage of the Internet and the current improvements in distance education in order to tackle three common limitations in this kind of studies. Namely, the need to use several instruments to be able to reproduce the complete process, the small number of individuals that may simultaneously

take part in the studies and the need to have an expert available to oversee the process at certain points. The proposed solution raises a remote work environment in which participants access the exercises via web and the system automatically records the information of their cognitive processes to perform real time analyses based on the data collected. It is therefore intended to not only provide an effective way to gather information about the cognitive processes involved in the writing, but also to create a quick and efficient way to collect the texts of the participants in any study of this nature without requiring their physical presence when performing the written composition exercises.

The second study posed in this project, the automatic evaluation of writing, focuses, in turn, on creating *Seshat Product*, one application designed to automate the analysis of written texts in order to generate comprehensive reports based on a pre-established set of well-defined rules. These rules, contained in a correction protocol created for the purpose of standardizing the manual implementation of the psycholinguistic analysis of written texts, also become essential to the research, since it is necessary to simplify and objectify them as part of the process prior to its automation.

Moreover, the creation of *Seshat Prodcut* raises a comprehensive review of the current methods, techniques and

existing solutions in the field of automatic text processing. Likewise, it also poses the use of a specific linguistic corpus in both the creation and the validation of the tool. The corpus proposed to this end is oriented to the study of writing development in schools and has been created from over twenty thousand texts written by Spanish students aged between eight and sixteen, with and without learning difficulties. This large documentary base, key to the refining of the correction protocol and search patterns used by *Seshat Product*, exists thanks to the efforts of the researchers of the Excellent Research Group of the Regional Government of Castilla y León GR259 and their collaborators, who have collected the necessary texts taking into consideration the written composition as a vehicle for studying writing with emphasis on factors of productivity, structure, coherence and general quality of the text (Arias & García, 2007; de Caso, García, Díez, Robledo, & Álvarez, 2010; Fidalgo & García, 2008; García, 2002; García & Arias-Gundín, 2004; García & de Caso, 2002, 2004, 2006a, 2006b, 2007; García & Fidalgo, 2003, 2004, 2006, 2008a, 2008b, 2008c; García & Marbán, 2003; García, Rodríguez-Pérez, Pacheco, & Díez, 2009; Pacheco, García, & Díez, 2009, 2010a, 2010b; Torrance, Fidalgo, & García, 2007).

Apart from the main purposes of the present research, this project also raises two secondary objectives inherent to the

development of the already mentioned, *Seshat Process* and *Seshat Product*, such as the contribution of these applications to the digital school and the promotion of the Spanish language.

One of the most attractive possibilities for the widespread use of tools such as those presented in this research is to provide psychologists and teachers with the means to carry out automated analysis in situ of psychological aspects associated with written texts. It is therefore intended not only to create a computer application exclusively oriented to the field of research at the university level, but an accessible and everyday use tool that can contribute to the work done in any educational institution in need of this type of systems. Thereby it is expected to provide teachers with the technology they need to carry out real-time textual analysis and generate comprehensive reports based on the experience of decades of psycholinguistic research.

The consequences derived from this research are also of the utmost importance as far as promoting the Spanish language is concerned. Considering the large number of Spanish speakers at the moment, more than four hundred ninety five millions according to the Instituto Cervantes (2012), the analysis of psychological aspects related to the written expression may be applied to sociocultural studies of various areas in the Spanish-speaking world, as well as serve as a vehicle to perform

comparative studies between native speakers and students of Spanish as a second language.

SESHAT PROCESS: MONITORING OF COGNITIVE PROCESSES

Seshat Process has been designed to implement the Kellogg's triple task procedure and analyze the writing process by adopting an approach to tackle three common limitations in this kind of studies. Namely, the need to use several instruments to be able to reproduce the complete process, the small number of individuals that may simultaneously take part in the studies and the need to have an expert available to oversee the process at certain points. The proposed solution raises a remote work environment in which participants access the exercises via web and the system automatically records the information of their cognitive processes to perform real time analyses based on the data collected. Therefore, although it follows Kellogg's ideas, also presents a new way to address the study of the writing process.

Having simplicity as a design constant throughout the development, *Seshat Process* allows the researcher to structure

the experiment as a sequence of stages or screens that integrate both the instruction of the participant and the writing exercise itself. The participant will access these screens following a logical succession of events that will guide her from the training in the processes recognition, previously categorized as convenience to the research, to the real exercise where she must identify them. As expected, this identification will be carried out using a direct retrospection method, so the participants will be exposed to certain stimuli during the writing process and she will be urged to categorize her mental process at the time they occur. As in the experiments carried out by Kellogg, the response time to such stimuli, measured in milliseconds, will be used to determine the cognitive effort at the moment. However, as a difference from Kellogg's experiences, this time, the nature of the stimulus will not be auditory, but visual. The primary objective of structuring the experiment this way is to allow the user to complete the whole task in a single session without any outside help.

It is also worth remarking that, due to its web-based design, this system also benefits from the advantages inherent to a long-distance education system, such as the fact of not requiring the physical presence of the participants in a classroom or avoiding further investment in equipment and additional software licenses. Another good example is mobility. Participants

can access the system using their own computers at any hour and are able to do scheduled tasks on their own, no matter where they are. Additionally, from a purely technical point of view, one of the biggest advantages of *Seshat Process* is that it can be accessed using conventional browsers, avoiding potential compatibility problems and ensuring its smooth operation on any up-to-date computer.

Next, some technical aspects of the tool are described as they have considered desirable to highlight in order to contribute to a better understanding of the overall system. As well as the characteristics of each of the three categories that defines the user's typology: *student*, *teacher* and *administrator*.

Technical aspects

From a purely technical point of view, *Seshat Process* is a web application hosted on one of the servers of the University of León and whose development merges the use of various programming languages, among which can be named, fundamentally: PHP, JavaScript and XML. Additionally, it also has a MySQL database and makes extensive use of AJAX technology to enhance the interactivity and optimize the general performance of the system by avoiding sporadic full page reloads. Since the database is key to understand how the whole

system works, it was considered appropriate to pay particular attention to it.

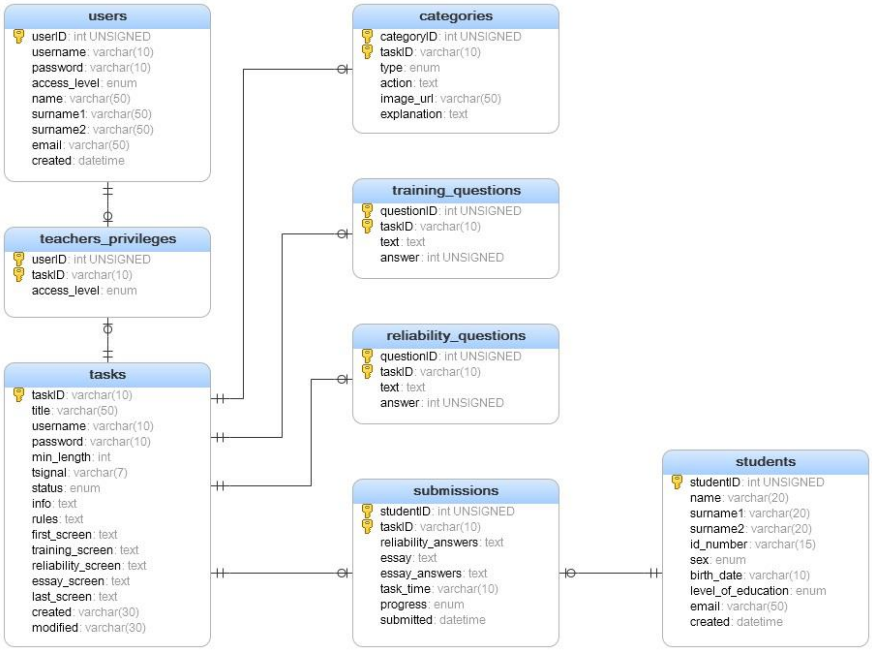


Figure 1. Database structure.

As can be seen on the diagram in Figure 1, the types of tables can be divided into two main groups, those that hold data related to tasks and those that hold the information related to the system's users. Concerning these last ones, it is especially remarkable that the *students* table is defined apart from the general users table, since this particularity defines the role of the writers in the system. Unlike the rest, the users of type *student* will not have a username and password of their own, but instead

will access the application by using previously defined generic credentials that will depend on the study they refer to. Thus, once they have accessed the system, students will be prompted to introduce their personal data; unless they have taken part in previous studies, in which case they could skip this step by simply inputting their personal ID number. This measure was adopted with the aims to ease the researcher's labor, who would have had to otherwise generate as many usernames and passwords as participants in the study.

On the other hand, as far as the data are concerned, each task is defined in terms of a series of specific attributes and three independent tables: categories, training_questions and reliability_questions. The functional relationship between these tables is explained in depth in the next section, however, from a technical perspective, it is noteworthy that there is a referential constraint between the answer fields and the categoryID field. This relationship will define, to a large extent, the operation of the main phases of the experiment, as it restricts the possible answers to the questions defined in the tables reliability_questions and training_questions to the categories established in the categories table. It also means that the participants will have to learn how to use these answers to identify their cognitive processes during the writing exercise.

Student interface

The category *student* generally refers to any individual who had participated in any study carried out with *Seshat Process*. This kind of user accesses the system using generic credentials shared with the rest of the participants involved in the same study. Once logged in, her experience in the system is defined on the basis of a sequence of ten screens that cannot be altered at time, except to carry out the planned activities. This way, it aims to ensure that all participants share not only the same credentials but the same experience, from the initial training stages to the conclusion of the final written exercise.

Initially, students will access a screen that briefly describes the process they are about to start and shows them the system's terms of use. This screen is merely informative and the user is not required to act but when decide to proceed to the next one, which will urge her to enter her personal data. Apart from the name, gender, email and personal ID number, for purely statistical purposes, she will be prompted to enter her date of birth, educational level, and the reason that led her to participate in the study. According to the researcher, this last question could be omitted or replaced by another one. However, in the event that the student was already registered in the system previous

participation in other studies, she could skip this entire screen by simply entering her personal ID number.

Category	Image	Explanation
Searching or reading information		Searching, checking and reading any information that might be helpful for writing the text (using the Internet, any related document, a book, etc.).
Reflecting on the aim of the text and its target audience		Reflecting on the purpose of the text and who is going to read it. For example, it is not the same to write a text for a teacher who is going to grade it than to write a text for a friend about personal matters.
Thinking about what to tell		Creating new stories or remembering something that has been read, seen, or told by someone to the writer, etc.
Making an outline		Writing down new ideas on a separate sheet that are going to be developed later in the text.
Writing the text		Whenever you are writing the text.
Reading the text		Reading the written text again either by reading every word, sentence or paragraph or by reading the entire text before submitting it.
Correcting the text		Correcting the existing mistakes.
Changing information		Deleting repeated words, adding information, changing words or reorganizing information.
Doing something unrelated to the task		Any action that is not related to the task at hand. For example, talking with someone, thinking about later appointments, looking out of the window at people walking down the street, etc.

Figure 2. Example of the list of categories.

After passing the first two screens, the student accesses the introduction of the study itself. In this screen, the nature of

the study is described in short and a generally hypothetical situation is presented in order to justify the necessity of writing an argumentative essay. In the same manner, it will also refer to the recognition of cognitive processes for the first time and present the list of categories that identifies them, as shown in Figure 2.

The fourth and fifth screens constitute the training phase which prepares participants to categorize their own mental activities during their future writing task. In the fourth screen the whole procedure is clearly explained. Continuing with the hypothetical situation previously proposed, the participant must interpret the actions carried out by a fictional character during a writing exercise. In the second screen, the participant must, therefore, perform a series of exercises in which she will have to identify the cognitive processes these actions involve by using the categories defined in the third screen (Figure 3). As an essential part of this training, participants will be able to instantly check whether their answers are correct or not and learn from their mistakes.

Once the training is over, the following two screens are a reliability test that evaluates the skills acquired by the participant in the training. This test also includes a set of questions, but unlike in the training phase, on this occasion it will not be

indicated if the answer is correct or not afterwards. This is intended to both evaluate the effectiveness of the proposed training and get a rough measure of the reliability of the participant's answers during the real writing exercise she will face next.

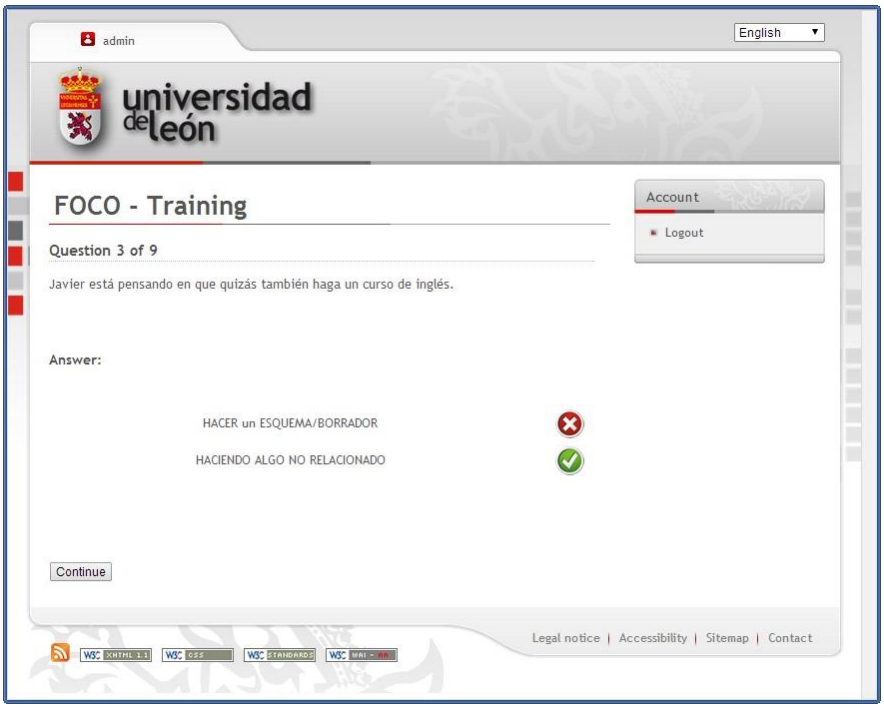


Figure 3. Unsupervised training.

As it happened in the training and reliability phases, the written exercise will be preceded by a brief explanation describing the details of the task. This brief introductory text will include the writing topic, the duration of the exercise and its extension if needed. Of course, it will also include a description

of when and how to proceed to categorize the writer's cognitive processes, specifying the nature of the interruption signal and its execution intervals. As an example, Figure 3 shows the appearance of the screen when it requests the participant to identify her cognitive process while she is writing.

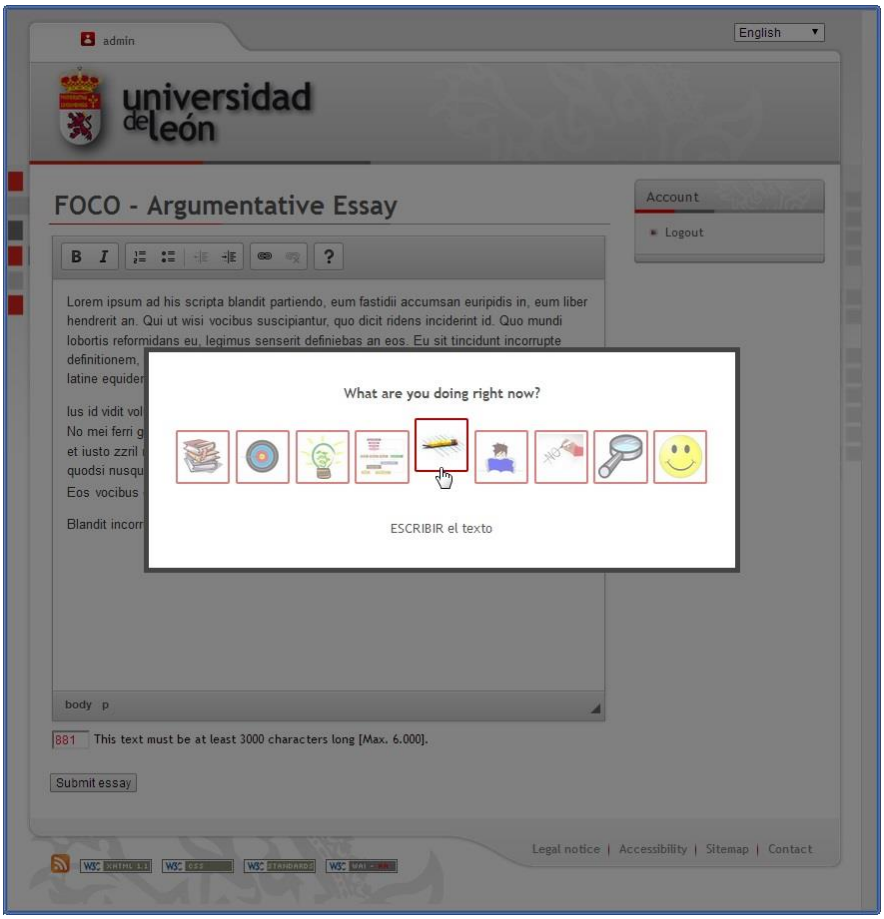


Figure 4. Example of the identification of a cognitive process during the writing exercise.

Finally, once the minimum requirements that had been established to carry out the task at hand have been satisfied, a modal dialog box will be displayed, prompting the user to proceed to save her exercise's data in the system. The tenth and final screen, that the participant will access next, is reserved for acknowledgments or any other considerations the researcher thinks appropriate.

Teacher interface

The category *teacher* defines the role of any researcher in *Seshat Process*. Unlike participants, this kind of user possesses a personal profile and can access the system using unique and nontransferable credentials. Once logged in, researchers will access the main menu of the application which presents four options, the account submenu that allows participants to modify their personal information and three general listings. First of all, the list of studies to which they have access, specifying the level of authorization in each case. Secondly, the list of participants who have taken part in all tasks available to the logged researcher. And finally, a list that contains the rest of the researchers whose respective level of authorization also allows them to access the same tasks. Although the *teacher* category defines a generic role for all researchers, according to their level

of authorization in a particular study, we can distinguish three different profiles.

Observer. The *observer* can inspect all the elements of the task, but she is not allowed to do any modifications, perform real-time analysis, download results or access the online records of the participants. However, she is able to export data from any section of the task to other tasks she can access at a higher authorization level.

Task manager. The *task manager* can inspect and modify all the elements of the task and see the online records of the participants, but, as the *observer*, she cannot modify them, perform real-time analysis or download the results of the study. The *task manager* is also able to export data from any section of the task to other tasks where she has at least the same authorization level.

Full manager. The *full manager* has full access to the task with modification options with respect to both the task and its participants. She also can perform real-time analysis and download the results of the study. Additionally, researchers with *full manager* category are able to modify the authorization level of other researchers within the study, except for those with the category.

Figure 5. Options menu from a full manager's perspective.

Regardless of the authorization level, when the researcher accesses any task on her list, the options menu will appear. Through this menu the researcher can manage each and every one of the screens of the study, access its list of participants, perform real-time analysis of the data collected and download or export the results as data matrices. Nevertheless, as expected, some of these functions will be restricted depending on the researcher's profile. The following describes every option on this

menu in detail (see Figure 5), specifying the required level of authorization when necessary.

General introduction. Once the study has been created, it allows the user to redefine basic aspects of it, such as the title, the internal description for researchers or the participants' credentials and also enables the researcher to establish the intervals of the interruption signal.

Terms of use. It allows for the definition of use directives for the system and the inclusion of a general description of the process if it was necessary. However, unlike the internal description of the previous section, it targets participants.

Task introduction. It allows for the definition of an introductory text that describes the nature of the study and presents a hypothetical scenario where the necessity of writing an argumentative essay is justified. In the participant's view, the screen to which this section refers will also present the list of categories that will be used to identify her cognitive processes during the subsequent writing exercise.

Categories. It allows for the definition of the list of categories that participants will use to explain the contents of their working memory during the written exercise. Each category consists of four elements: type, image, action and explanation. As its name suggests, type determines the general type of the

cognitive process associated with the category and may adopt four values: planning, translating, revising and others. The image is used to set a representative icon of the process that facilitates a quick identification. The action in turn defines the name of the process. And finally, the explanation consists of a brief description of the category itself and some clues about when it must be used.

Training. This section has two parts: text and questions. The first part acts as an introductory section, presenting the training phase to the participant. The second allows the participant to define the battery of questions which will constitute the training process itself. The answers to these questions will always be defined in terms of the existing categories, so it is impossible to create training questions in the absence of categories.

Reliability. Similar to the previous one, this section allows for the definition of a reliability test just as it is done with the training, so it also comprises an introductory text and a battery of questions. As for the researcher, there are no significant differences between these two sections, but participants will only be able to see the predefined answers in the case of training.

Exercise explanation. It allows for the definition of an introduction to the final exercise. It usually contains information

about the writing topic and the established restrictions for the writing task in terms of time and extension limits.

Acknowledgements. It allows for the definition of the contents that the participant will access after having finished the written exercise and recorded its results. The acknowledgements section is the last screen before leaving the system and its content can be freely modified by the researcher at her convenience.

Import data. It allows for the importing of data from any other course the researcher may access. It will only be possible to import full sections and there are also some restrictions that should be taken into consideration. Importing training of reliability questions will cause the automatic importation of the referred categories, unlike in the opposite case, it being possible to import the categories without importing any questions at all. Additionally, for security reasons, once the system has begun to register participants' results, the training, reliability and categories sections will be blocked for importing purposes unless such contributions are manually removed.

Participants. It displays the list of participants who took part in the present study, specifying their reliability score and the time it took them to finish the written exercise. Each of the names on this list will in turn grant access to the participant's records which contain her personal data, all her results and the

writing text itself. *Task manager* researchers can see the list of participants, but this section is only fully available to researchers with *full manager* category and system administrators.

Results. It allows the researcher to perform real-time analysis of the data obtained during the reliability tests and the written exercises. It also allows for the exportation of the resulting data matrices in *Microsoft Excel* format and the downloading of the overall results of the study. Given its complexity and that the capability to perform analysis is an essential feature of *Seshat Process*, this section will be discussed separately in section 3.5. Access to results is only available to researchers with *full manager* category and system administrators.

Administrator interface

As it happened with teachers, users with *administrator* status access the system using a unique name and password that do not share with any other. Administrators have the attributions of a *full manager* researcher in all existent studies and may potentially play her role if necessary. However, this category involves functions that exceed the privileges of teachers, as this kind of users is also able to manage the accounts of other users by modifying or removing them as they see fit. Additionally,

only system administrators can create new users and studies or deleting the existing ones.

Analysis capabilities and downloads

Next, the analysis capabilities of *Seshat Process* are examined throughout five sections in which the different types of possible analyses are describe in detail as well as the available options to download the results of a study.

Processes usage

The system allows the researcher to calculate the percentage of usage of each cognitive process according to the generic types in which the established categories are subdivided: planning, translating, reviewing and others. These generic types will always be constant regardless of the number and nature of the categories of a given study and cannot be changed.

As we can see in Figure 6, the final results are shown on an interactive chart that will facilitate the verification of the exact percentage of usage of any process at any time, since the data is updated in real time as the system gets new results from the participants.

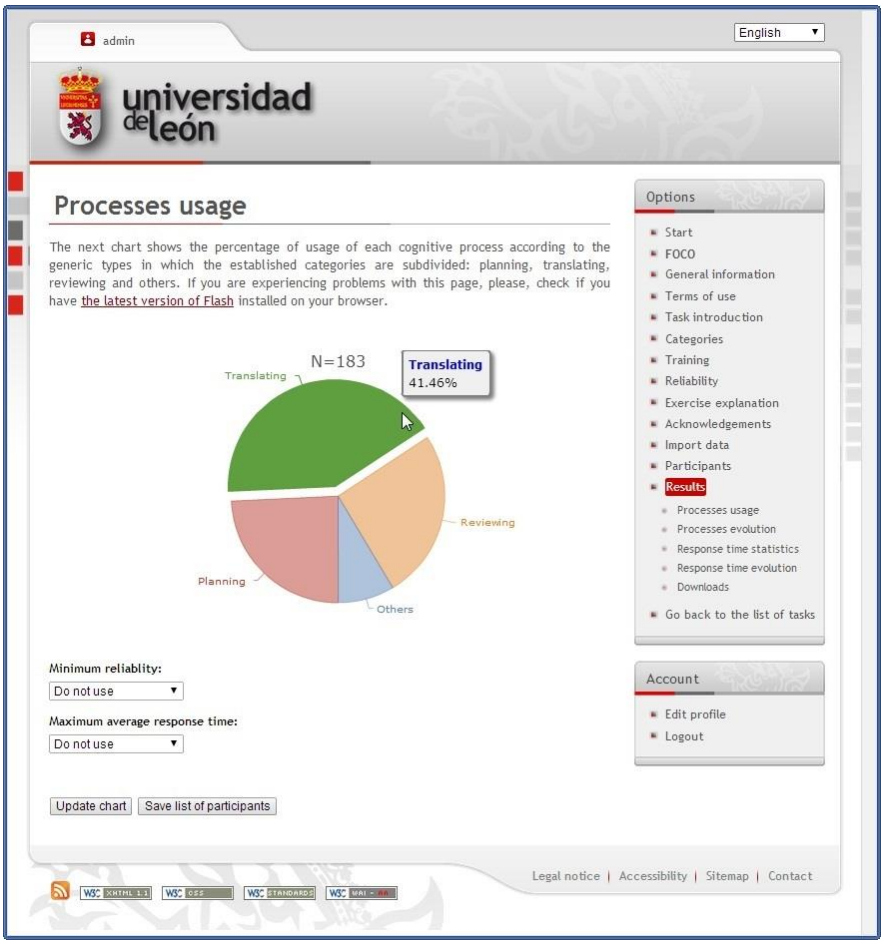


Figure 6. Processes usage chart.

This type of analysis is susceptible of two restrictions, minimum reliability and maximum average response time. The minimum reliability filter uses the results of the reliability test to establish a minimum threshold that automatically rules out all those participants who do not surpass it. In the same way, the

maximum average response time establish a maximum threshold to rule out, in turn, those participants whose average response time per beep[1] is higher than that.

Processes' evolution

The system allows the researcher to analyze the evolution of each cognitive process in relation to the general progress of the writing task. Since the time to carry out the exercise varies between participants, it is necessary to establish a reference time before the analysis. Thus, the results of each participant are extrapolated using the percentage of overall completeness of the task in relation to this reference time in order to find out what kind of process was taking place at that time in the participant's exercise. As shown in Figure 7, the chart that shows the evolution of the cognitive processes is also dynamic and highly interactive as in the previous case. This chart implements real time updates based on the participants' answers and allows the researcher to check the exact percentage of usage of any process in relation to the general progression of the task. There is also the possibility of applying the above-mentioned restrictions of minimum reliability and maximum average response time, but, in this case, the researcher can adjust the resulting curves to its respective trend lines as shown in Figure 8.

[1] Seshat Process does not use auditory signals; however, the term "beep" is still used in this context for a matter of clarity.

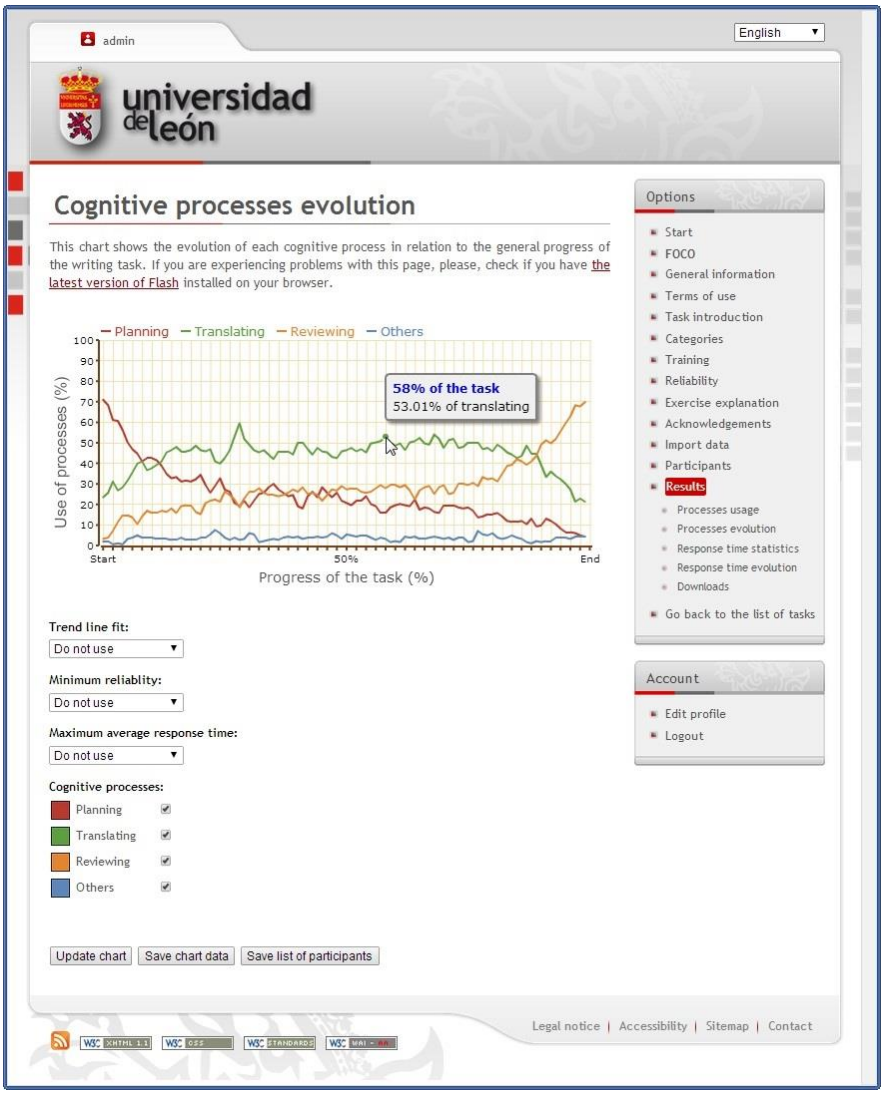


Figure 7. Cognitive processes evolution.

This filter implements a moving means algorithm and can be used at four levels of intensity: low, medium, high and very high. It is especially useful when the study has few participants.

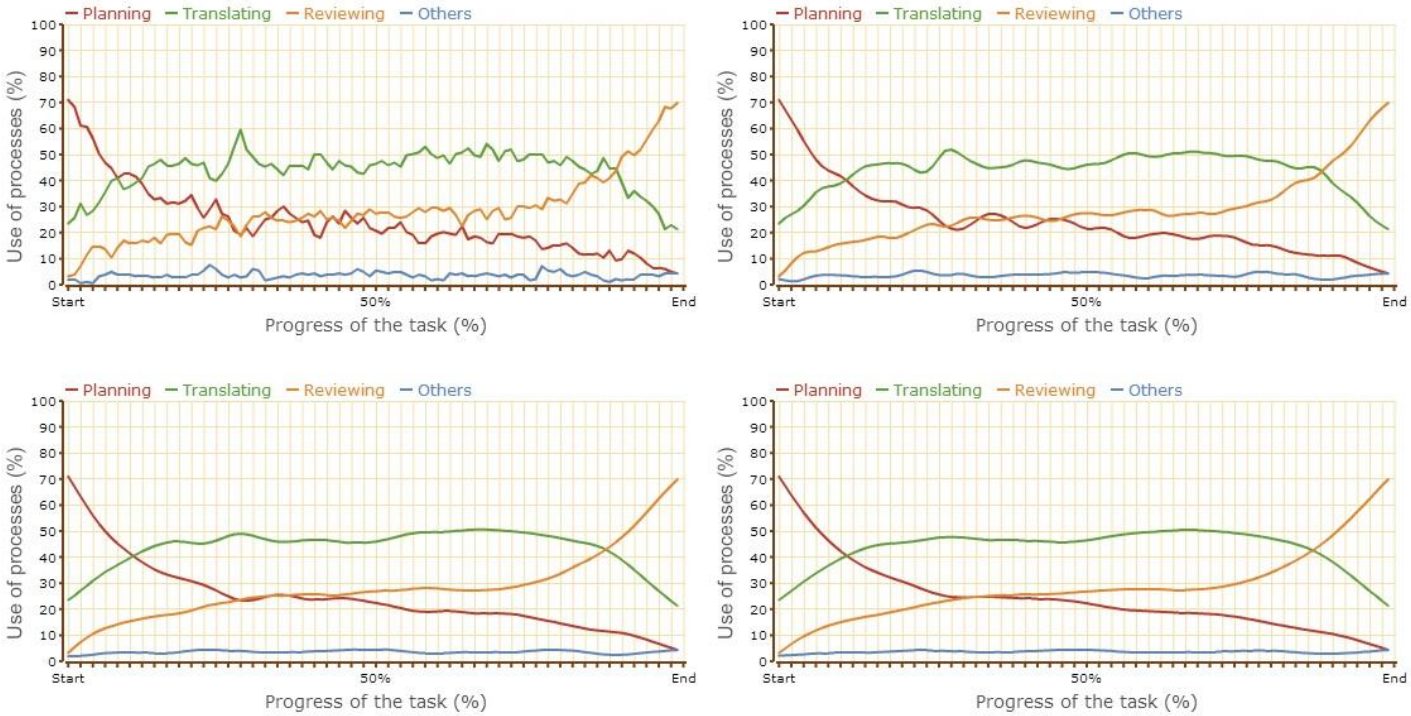


Figure 8. Trend line filters from none to high.

Additionally, this section allows for the exportation of the results in *Microsoft Excel* format. It allows the researcher to export both the chart data and the list of participants who were taken into consideration in its creation after using the desired restrictions.

Response time statistics

By means of this section it is possible to make quick statistical calculations using the response times of the participants. As in previous analysis, the researcher can also establish some restrictions as minimum reliability, maximum average response time and maximum average time per beep. Being a supervised environment, this last restriction turns out to be especially useful to avoid anomalous situations in which one participant highly exceeds the expected response times, altering the general distribution of the sample and its resulting statistics this way. The results are always presented in relation to the four generic types of cognitive processes, planning, translating, reviewing and others, and include the calculation of the mean, the standard deviation, the standard error and the confidence intervals.

Response time evolution

The system allows the researcher to analyze the evolution of the average response time of the participants in relation to the

general progress of the writing task. As it happened with the evolutionary analysis of cognitive processes, it is essential to establish a reference time in order to make the necessary calculations. Also, as we can see in Figure 9, the results are presented in an interactive graphic that will always let the researcher to know the exact average response time with respect to a certain percentage of completeness of the task. Just as in the rest of the analysis carried out by *Seshat Process*, the data used in the creation of this chart will be updated in real time based on user responses, it being possible to adjust the resulting curves to their respective trend lines by using the filters provided. Restrictions like minimum reliability, maximum average response time and maximum response time per beep can also be used and, as in previous analyses, the researcher can export the resulting data matrices in *Microsoft Excel format*.

The main peculiarity of this section lies in the possibility of making comparative analysis between various samples, up to four at the same time. Nevertheless, the current version of the platform only supports two kinds of samples, general and individual. The general type sample includes all participants who have taken part in the study, unlike the individual that, as its name suggests, only holds one. Of course, both the data and the number of participants of the selected samples may vary

according to the established restrictions, with options to make them irrelevant for analysis purposes.

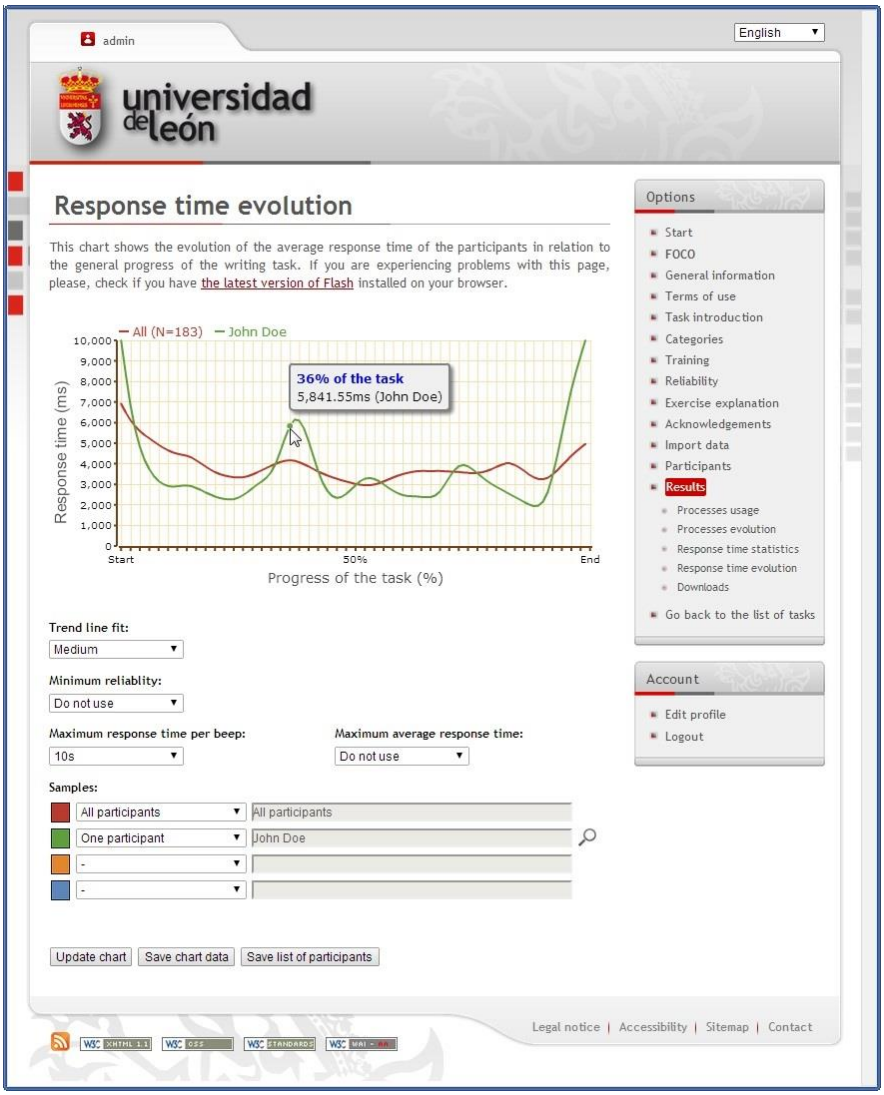


Figure 9. Response times evolution.

Downloads

This section allows the researcher to download the general results of the study in two different formats. A matrix of texts containing the ID of every participant and their respective texts and another matrix of results that contains the personal data of the participants, their reliability scores, the results of all categorization processes and the response times per beep in milliseconds.

SESHAT PRODUCT: AUTOMATED ESSAY SCORING

Seshat Product is a software application designed to automate the assessment and analysis processes of written texts in order to generate comprehensive reports based on a pre-established set of well-defined rules (Díez-Caso & García, 2011, 2013). These rules, which constitute the theoretical and practical basis of the application, are contained in a correction protocol created to establish standardized assessment criteria for the purpose of objectively assessing different aspects of texts written in Spanish (García et al., 2014). From a psycholinguistic perspective, the original leitmotiv of this research was the study

of learning difficulties in pre-university schools; however, over time, the refinement of the correction protocol has evolved *Seshat Product* towards a more general orientation. This has been possible thanks to a large and heterogeneous documentary base that contains over twenty thousand texts written by a total of seven thousand and two hundred and thirty two Spanish students, with and without learning difficulties, aged between eight and sixteen years old.

Seshat Product has been designed to ease the work of teachers and researchers enabling them to speed up the text correction process. However, unlike most of the AES systems, *Seshat Product* does not completely eliminate the human factor from the process, since it does not provide a general assessment of the writing quality, but the scores of some linguistic features that have to be interpreted by the user afterwards. The objective of the proposed solution, therefore, is not to eliminate the human action from the picture, but to provide professional educators and researchers with the necessary means to create instructional individualized strategies and a way to countersign their findings on the experience of thousands of previously analyzed texts with similar characteristics.

Abstract

Correction protocol

Since psychology of writing first begun, a growing concern about how to obtain objective indicators to assess writing in quantitative and qualitative terms has developed. In the field of educational research, it is also very common to take into consideration all indicators that are relevant to the study of aspects particularly related to this area, such as the design of instructional strategies for improving written communication skills (García et al., 2014). The psychology of writing therefore encompasses an extremely broad field that includes the analysis of written composition from both the process and the product perspective; however, the correction protocol whose presentation is addressed in this section focuses exclusively in what regards the analysis of writing as a product.

The present correction protocol is the result of over a decade of research focused on obtaining objective indicators that would allow analyzing texts written in Spanish on the basis of maximizing the inter-rater agreement between raters and the ability to distinguish between students with and without learning difficulties in basic training cycles. This approach, however, has evolved over time to adopt a more generalistic perspective. Today, with the aim of expanding its range of uses, the correction protocol implemented by *Seshat Product* has been divided into

two parts, text-based and reader-based indicators. On one hand, indicators based on the text present a purely objective analysis focusing on the generation and organization of information. On the other hand, reader-based indicators take a more subjective analysis model in which the experience and judgment of the human rater remains key to the process. Both analysis techniques have been extensively studied over the past two decades, from the early works of Spencer and Fitzgerald (1993) to consolidations of the present correction protocol (Fidalgo, Torrance, & García, 2008; Garcia & de Caso, 2004; García & de Caso, 2006; García et al., 2014; García & Fidalgo, 2006; García & Marbán, 2003; García & Pacheco, 2007; Torrance, Fidalgo, & García, 2007).

As far as *Seshat Product* is concerned, nowadays, only the text-based indicators have been implemented, since there is still plenty work to do on the refinement of the reader-based indicators, indispensable to obtain a general assessment of the writing quality. However, even under the current conditions, the partial implementation of the correction protocol has already made it possible to perform a series of studies focusing on the identification of evolutionary patterns of interest through writing analysis. The study of the predictive role of teachers in the teaching of written composition (García & Pacheco, 2012a, 2012b), the evolutionary studies in students aged between 8 and

16 years (Álvarez & García, 2014), the identification of the changes after various instructional interventions with parents and teachers (Robledo & García, 2013), the introduction of aids in rewriting tasks (Arias & García, 2012), the comparison of different levels of writing development using professional corpora (García-Martín y García, 2012) and even the realization of experimental and developmental studies (García, Rodríguez-Pérez, Pacheco, & Díez, 2009) are also noteworthy.

The following text presents the indicators currently implemented by *Seshat Product* concerning the generation and the organization of the information. As it can be seen in Tables 1 and 2, next to each indicator, the code used to identify it inside the program is shown, as well as a brief description of its nature.

Information generation

The measurements of productivity are carried out by counting specific particles, expressions and other aspects of the text. As shown in Table 1, apart from general indicators, we distinguish between content and functional words. Content words are those whose semantic information is generally related to properties associated with physical entities, and function words those whose semantic information is purely grammatical and therefore abstract and devoid of specific referents.

Table 1
Indicators of information generation

General		
ID	Name	Description
TITL	Title	Indicates the presence or absence of title.
CFIN	Final comment	Indicates the presence or absence of a final comment.
NPAR	Number of paragraphs	Counts the number of paragraphs.
NUNI	Number of textual units	Counts the number of textual units; being a textual unit a sentence delimited by periods, ellipsis, question marks or exclamation marks.
NVFP	Number of verbs in personal forms and phrasal verbs	Counts the number of verbs and phrasal verbs, except for the impersonal forms: infinitive, gerund and participle.
NDET[a]	Number of determiners	Counts the number of determiners including definite articles and demonstrative, possessive, numeral, indefinite and interrogative determiners.

Table 1
Indicators of information generation (continuation)

Content words[a] (NCON)		
ID	Name	Description
NNOM	Number of nouns	Counts the total number of nouns considering variations in gender and number as different words.
NADJ	Number of adjectives	Counts the number of qualifying adjectives.
NADV	Number of adverbs	Counts the number of adverbs and adverbial phrases.
NVER	Number of verbs	Counts the total number of verbs including impersonal forms.
NINT	Number of interjections	Counts the number of interjections.

Table 1

Indicators of information generation (continuation)

Functional words[a] (NFUN)		
ID	Nombre	Description
NPRE	Number of prepositions	Counts the number of prepositions and prepositional phrases.
NCNJ	Number of conjunctions	Counts the number of conjunctions and conjunctive phrases.
NPRO	Number of pronouns	Counts the number of pronouns including personal, relative, demonstrative, possessive, numerals, indefinite and interrogative pronouns.

[a] Seshat Product has an overall productivity indicator (PTOT) as result of the sum of determiners, content words and function words.

Information organization

The organization of the text is operationalized by means of coherence indicators. The textual coherence is a cognitive phenomenon inherent to the mental representation of the text that, on the surface, is represented through a series of marks and indicators that act as guides or processing instructions when recreating such representation (García et al., 2014).

As it can be seen in Table 2, apart from the general indicators, a coherent text consist of two types of elements, referential coherence and relational coherence. Referential coherence refers to the connection established between different elements of the text by semantic or literal repetition of both objects and subjects. Relational coherence, in turn, refers to the organization of the text based on the relationship established between the various elements of discourse by means of specific expressions and marks at a textual level.

Table 2

Indicators of information organization

General		
ID	Name	Description
NSIG	Punctuations marks	Counts the number of punctuation marks. Commas, semicolons, periods, full stops, quotes, and more.
Referential coherence[a] (CREF)		
ID	Name	Description
NREF	Referential indicators	Counts the number of exact lexical repetitions or semantic coincidences between the elements of written discourse.
NLEX	Lexical indicators	Counts the number of repetition of content words in the text.
Relational coherence[a] (CREL)		
ID	Name	Description
NMET	Meta-structural elements[b]	Counts the number of complex structures, markers of previous or subsequent content in the text that either anticipate the information that comes next or summarize what has been said before.

Table 2

Indicators of information organization (continuation)

Relational coherence[a] (CREL)		
ID	Name	Description
NEST	Structural elements[b]	Counts the number of basic organizational structures of the written discourse.
NCNT	Connective elements[b]	Counts the number of structures that establish semantic or pragmatic links between two elements of the written discourse.
NRFM	Reformulative elements[b]	Counts the number of structures that present a discourse element by the reformulation of a previous one.
NARG	Argumentative elements[b]	Counts the number of structures that condition the argumentative possibilities of the discourse element in which they are included without relating it to a previous one.

[a] Seshat Product has an overall coherence indicator (CTOTs) as result of the sum of referential and relational indicators.

[b] Discursive indicators. Invariant linguistic units sharing the same task within the discourse: guiding the inferences made in the communication process.

Technical aspects and program's workflow

Seshat Product is a desktop application implemented in Java as part of *Seshat Writing Research Tools*, a project that aims to develop a platform that integrates all means necessary for the study of writing from a process-product perspective. All tools included in this suite of applications share a common interface in order to simplify its integration with office software and more extended text processors. For this reason it was decided to make use of the *Apache POI* libraries (The Apache Software Foundation, 2014) to ensure compatibility with the formats used by *Microsoft Office*. Given the size of the project and the expected complexity of future maintenance, *Seshat Product* was designed following the modular programming paradigm, being divided into five interdependent modules: I/O management, morphological tagger, database manager, rules analyzer and graphical user interface. On Figure 10 it is shown how these parts interact with each other, shaping the general scheme of the internal operation of the program. As it can be seen, once the texts have been selected and made available to the system, they are exposed to an initial phase of morphosyntactic tagging. In this regard, it is noteworthy that the creation of *Seshat Product* was undertaken with the intention of obtaining useful results in the shortest possible time. Therefore, it was decided to avoid creating a new morphological tagger and consequently it was

necessary to begin the search for the most suitable solution to the project needs from those available in the market.

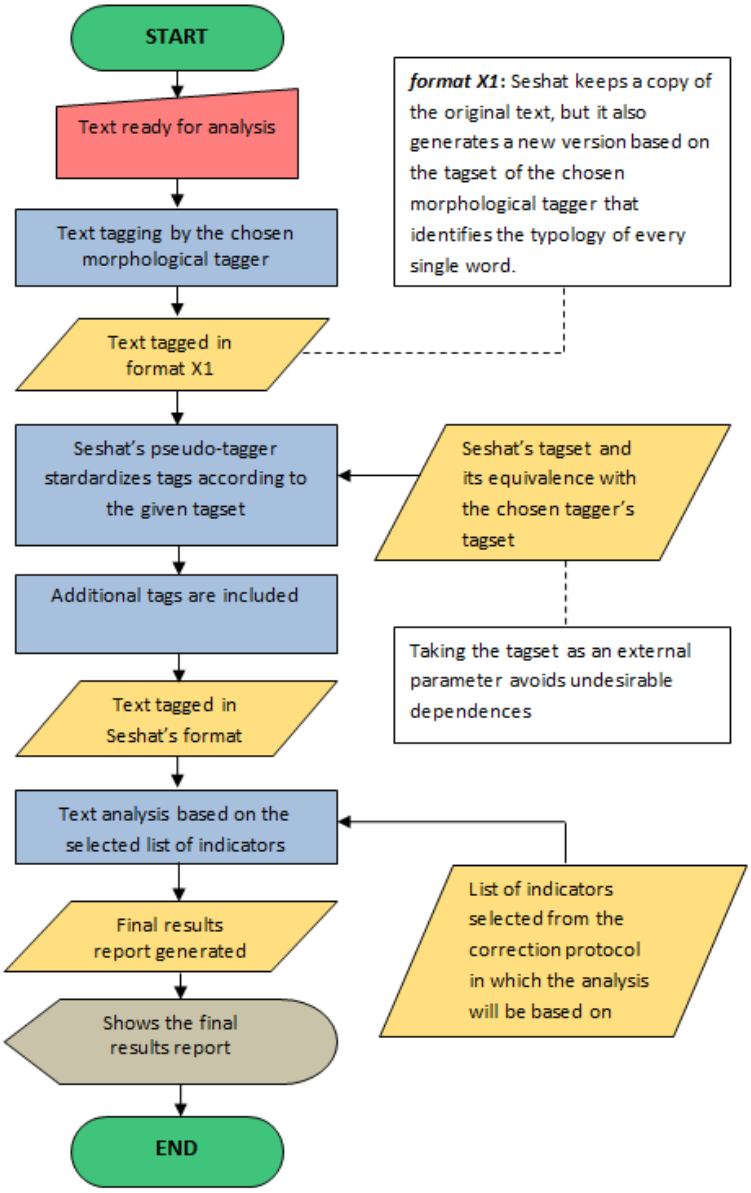


Figure 10. Flow chart diagram of the analysis process.

After a selection phase focused on reliability, price and ease of integration and maintenance, it was finally decided to acquire the *Connexor Machinese Phrase Tagger*, a professional tagger provided by the Finnish company *Connexor Oy* (2014). Above any other consideration, the crucial factor that tipped the scale in favor of this tagger was its creation as a C++ library that could be easily integrated in our system using JNI (Java Native Interface) or even replaced without undue complications if necessary. This latter aspect was particularly relevant, since the intent was to avoid excessive reliance on commercial products outside the project. It also motivated the development of a pseudo-tagger to standardize the tagset used by the system. Thereby, once tagged by the *Connexor MPT*, all texts undergo a second tagging phase in order to correct possible errors and modify the existent tags by replacing them using the standardized tagset according to the requirements of the correction protocol that governs the system's operation. Notable examples of this are the recognition of proper nouns, compound verbs and phrasal verbs as a single word, as well as the unequivocal identification of percentages and numeral adjectives regardless of how they were written. As a result, it can be said that the pseudo-tagger plays a triple role as it not only eliminates commercial dependencies, but corrects errors and simplifies future maintenance processes by providing a quick and effective way to

replace the main tagger if necessary. Moreover, its creation also provides the basis for developing a full morphosyntactic tagger in the future.

After the tagging phase, the text is available to the rules analyser, which assesses the resulting document according to a series of selected indicators of the correction protocol. The system codes each of these indicators in the form of a set of logical rules that attempt to emulate the manual correction. These rules, in turn, base their operation on the execution of one or more search patterns that explore the text based on the information contained in the program's database created from previous analyses. All rules are implemented with the intention of reaching the maximum compromise between efficiency and versatility, always trying to take advantage of existing synergies. This ensures, for example, that the user can select a particular subset of indicators at any time and make a partial analysis of the text. The system will detect if there is any relationship between the rules associated with the selected indicators. If so, it will optimize the analysis process using the results from previous analyses or by forcing the background execution of the additional rules required.

Moreover, from the point of view of efficiency, other factors have also been taken into account, such as the algorithmic

complexity of the search methods employed. As it would be expected, the execution time grows as the complexity of the algorithms involved in the search increases, which, in turn, grows depending on the established search patterns and the essay's length and complexity. This problem provokes a comparison between various search methods a priori that was eventually resolved by implementing a customized version of the Boyer Moore Horspool (BMH) algorithm adapted to the specific needs of *Seshat Product*.

As seen in Figure 1, after completing the tagging and analysis processes, a comprehensive results report is generated. This report constitutes an array of data that is presented on screen showing both the ID of the texts analyzed and the indicators according to which the analysis was performed. Finally, *Seshat Product* lets us export the final results in *Microsoft Excel* format, ensuring their compatibility with most existing statistical analysis software.

User interface

This section thoroughly describes the process of analysis from the point of view of the user. For this purpose, the graphical user interface module is also presented, i.e., the part of *Seshat Product* designed to manage how users interact either passively or actively with the system. The highest priority in creating this

module was to design and implement an intuitive and handy interface that did not require prior knowledge to use. For this reason, during the development process, it was necessary to work side by side with teachers and psychologists picking comments and suggestions so as to create a work environment that suited their specific needs and requirements. Some of these suggestions meant significant challenges that altered the preliminary designs of the program, as, for example, the multiplatform design. *Seshat Product* has been developed using Java technology in order to both enable its use in no matter what operating systems and adapt its look and feel to the specific characteristics of the running environment. Attention is drawn, however, to the fact that screenshots presented throughout this section have been taken in a *Microsoft Windows* environment only, since it is the most widespread operating system at workplaces in our days.

As shown in Figure 11, *Seshat Product* is a desktop application with a classic interface based on a windows system which has a menu bar, a toolbar with shortcuts to key features of the program, a work area and a status bar that displays short help messages. The text below describes in detail a routine analysis paying attention to each of the steps involved in the process and showing the necessary screenshots in order to illustrate the experience the experience from the point of view of the user.

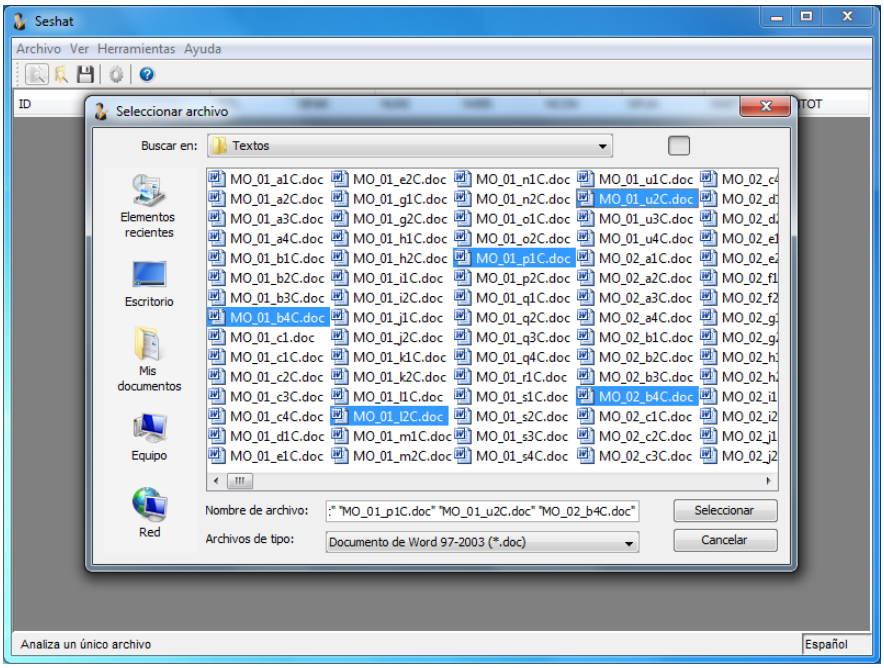


Figure 11. Selecting files for analysis.

Figure 11 shows the first step of the analysis process, the selection of files that contain the texts under study. In *Seshat Product* there are two ways of selecting files, the individual files selection and directory selection. In the first case, as shown in Figure 2, it is possible to select multiple files simultaneously. This way it is possible, for example, to conduct both evolutionary analyses of the texts of the same individual and comparative studies of several specific texts. In the second case, the selection of directories, the user can select an entire directory in order to analyze the content of all its files at once. This method lets the user modify some search criteria before starting the selection

process such as choosing the accepted file formats[1] or the possibility of making a recursive search that includes not only the files present in the selected directory, but all contained in any of its subdirectories. The latter option has been included with the intention of providing the user with a means to perform an exhaustive analysis of large amounts of information organized in a hierarchical directory structure, as it is usually the case, for example, in most linguistic corpora.

Once the selection process has finished, the system displays a modal dialog box that reports the total number of selected files and awaits the user confirmation to start the real analysis (see Figure 12). Each file in the list is independently processed, searching and extracting the texts contained inside[2] and making them available to the system in order to start their analysis as described in the previous section.

After the analysis has been initiated, the system will keep the user promptly informed about the percentage of completeness achieved and of any eventuality that may take place during the process. If errors occur, the system will alert the user indicating their nature and the files that provoked them.

---

[1] Seshat Product currently supports Apache Open Office Writer, Microsoft Excel, Microsoft Word and any plain text editor.

[2] Seshat Product can process Excel Books that may contain one or more documents organized according to a structure specifically designed for this purpose.

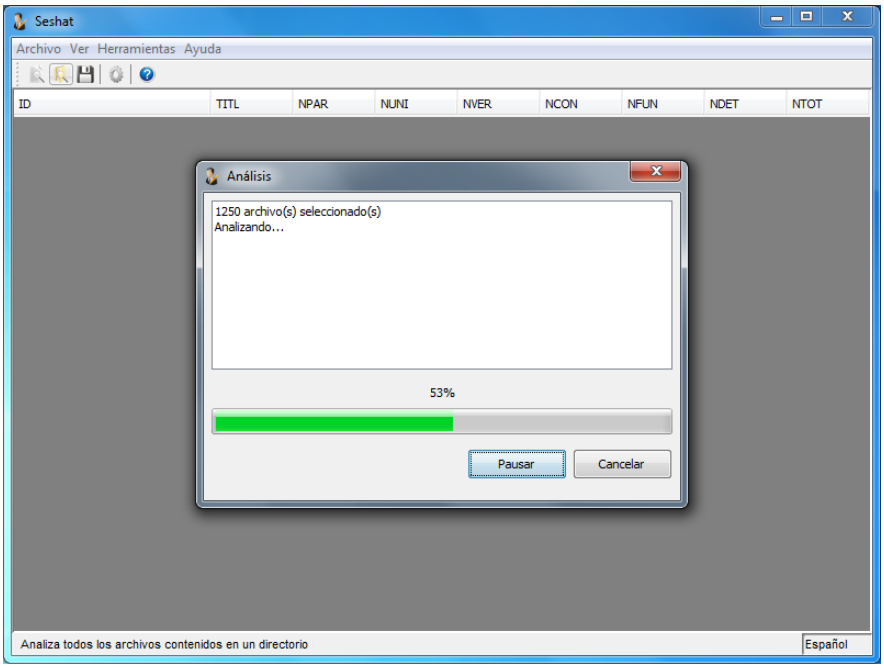
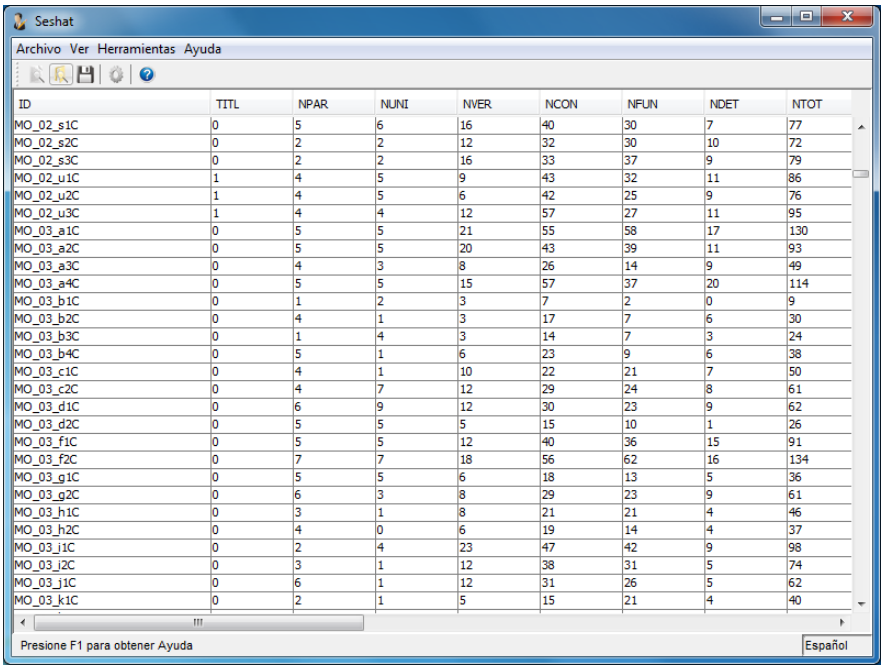


Figure 12. Analysis in progress.

Seshat Product distinguishes between four types of errors: read errors, format errors, tagging errors and fatal errors. Read errors allude to the inability of the system to access a particular file. Format errors refer to any problem that may arise from processing the content of the files analyzed. Tagging errors, in turn, refer to all eventualities that may arise during the labelling, regardless of what tagger caused it. Finally, fatal errors contemplate any unidentified problem that does not fit into any of the categories above. In any case, whether there are errors or not, the analysis process will continue until it has processed each

and every one of the selected files. At this point, the user is the only one able to stop it.

Once the analysis process is completed, a final report indicating the elapsed time, the number of analyzed files in relation to the total number of initially selected files and any incident that may have occurred during the process will be displayed on screen. At this point the user can either accept or reject the results. If accepted, the results are displayed on screen as shown in Figure 13.



ID	TITL	NPAR	NUNI	NVER	NCON	NFUN	NDET	NTOT
MO_02_s1C	0	5	6	16	40	30	7	77
MO_02_s2C	0	2	2	12	32	30	10	72
MO_02_s3C	0	2	2	16	33	37	9	79
MO_02_u1C	1	4	5	9	43	32	11	86
MO_02_u2C	1	4	5	6	42	25	9	76
MO_02_u3C	1	4	4	12	57	27	11	95
MO_03_a1C	0	5	5	21	55	58	17	130
MO_03_a2C	0	5	5	20	43	39	11	93
MO_03_a3C	0	4	3	8	26	14	9	49
MO_03_a4C	0	5	5	15	57	37	20	114
MO_03_b1C	0	1	2	3	7	2	0	9
MO_03_b2C	0	4	1	3	17	7	6	30
MO_03_b3C	0	1	4	3	14	7	3	24
MO_03_b4C	0	5	1	6	23	9	6	38
MO_03_c1C	0	4	1	10	22	21	7	50
MO_03_c2C	0	4	7	12	29	24	8	61
MO_03_d1C	0	6	9	12	30	23	9	62
MO_03_d2C	0	5	5	5	15	10	1	26
MO_03_f1C	0	5	5	12	40	36	15	91
MO_03_f2C	0	7	7	18	56	62	16	134
MO_03_g1C	0	5	5	6	18	13	5	36
MO_03_g2C	0	6	3	8	29	23	9	61
MO_03_h1C	0	3	1	8	21	21	4	46
MO_03_h2C	0	4	0	6	19	14	4	37
MO_03_j1C	0	2	4	23	47	42	9	98
MO_03_j2C	0	3	1	12	38	31	5	74
MO_03_j1C	0	6	1	12	31	26	5	62
MO_03_k1C	0	2	1	5	15	21	4	40

Figure 13. Analysis results' view.

Each of the rows in the result table represents an analyzed text identified by the name of the file that contained it or, failing that, an internal reference in the case of files that may contain more than one text. As shown on the top of Figure 4, this identifier is represented in the table header as well as the identifier of all indicators analyzed. It makes it possible to instantly check the resulting score of any indicator in relation to a particular text.

One final consideration worth mentioning with respect to the analysis process is its speed. Tests conducted proved the considerable difference between manual analysis and the analysis performed by *Seshat Product*. While the manual application of the correction protocol entails an average time of at least fifteen minutes to analyze a text of one hundred and fifty to two hundred words, *Seshat Product* can analyze one thousand texts of this size in less than fifteen seconds. Additionally, all tests perform with linguistic corpus have yielded excellent results, having come to analyze a professional corpus of pharmacology of over one million and a half words in barely four and half minutes.

Finally, once the final results report has been presented, *Seshat Product* enables the user to export this report in *Microsoft Excel* format so as to provide a means to get a data matrix that is

both easy to edit and highly compatible with most statistical analysis software used today.

Validation

Over the course of this research it was necessary to conduct two validation processes, the validation of the correction protocol, designed to provide objective indicators to analyze writing, and the validation of *Seshat Product*, the software designed in order to automate its application. The first validation was not intended to prove the effectiveness of the correction protocol, already tested in previous studies (Arias & García, 2012; García-Martín & García, 2012; García & Pacheco, 2012a, 2012b; García, Rodríguez-Pérez, Pacheco, & Díez, 2009; Robledo & García, 2013), but to check and maximize, if necessary, the inter-rater agreement between expert human raters before comparing their results with those obtained by *Seshat Product*. To this end, it was considered appropriate to start analyzing a series of samples of between forty and fifty texts and then checking the inter-rater agreement using the Cohen's Kappa coefficient with the aim of minimizing the effect of randomness on measurements. The disparity of the first results revealed the need to tackle an individualized review of the correction methods used in some of the indicators that were put to the test.

Table 3

Inter-rater agreement in the study of main productivity and coherence indicators implemented by Seshat Product (N=537).

	Human vs. Human		Human vs. Seshat Product	
	Cohen's Kappa[a]	Agreement	Cohen's Kappa[a]	Agreement
TITL[b]	1.000	Perfect	1.000	Perfect
CFIN[b]	1.000	Perfect	1.000	Perfect
NPAR	0.967	Almost perfect	0.897	Very good
NUNI	0.921	Almost perfect	0.872	Very good
NCON	0.918	Almost perfect	0.841	Very good
NFUN	0.937	Almost perfect	0.846	Very good
NTOT	0.926	Almost perfect	0.864	Very good
NREF	0.844	Very good	0.802	Very good
NLEX	0.836	Very good	0.817	Very good
CREF	0.841	Very good	0.810	Very good
NMET	0.863	Very good	0.833	Very good
NEST	0.850	Very good	0.809	Very good
NCNT	0.911	Almost perfect	0.874	Very good
NRFM	0.882	Very good	0.830	Very good
NARG	0.861	Very good	0.773	Good
CREL	0.887	Very good	0.839	Very good
NSIG	0.934	Almost perfect	0.956	Almost perfect

[a] All results are significant at $p<0.05$.

[b] All texts of this sample were pre-corrected prior to the analysis, so they incorporated special marks to identify the title and the final comment. The values of CFIN and TITL indicators do not represent, therefore, the true capacity of Seshat Product to identify these features in the text. Additional tests on unedited texts samples reported inter-rater agreements between manual and automated correction of 0.667 (Good) and 0.589 (Moderate) in TITL and CFIN respectively.

Abstract

To deal with this problem it was decided to set a series of meetings between the technical and the teaching staff in order to systematically search and fix all functional ambiguities of the correction protocol. Specific objectives were also fixed with respect to the inter-rater agreement between human raters, establishing a minimum threshold of .80 for those indicators that involved semantic considerations and of .90 for the rest. The final result of this refinement process involved the expansion of the search lists used in productivity indicators and the incorporation of more specific search patterns to improve the identification of complex grammar structures in the case of coherence indicators.

After reviewing the correction methods, another test with a larger sample was conducted in order to get the results necessary to carry out the validation of *Seshat Product*. For this purpose, paying particular attention to its heterogeneity, a sample of five hundred and thirty seven texts was selected from students aged between eight and sixteen years old. These texts underwent a rigorous correction process in order to eliminate any possible spelling mistakes before making them available to the raters. This preliminary step was necessary since it was considered desirable to check the inter-rater agreement between manual and automated correction under ideal circumstances. Once selected and reviewed, six independent raters took part in the analysis of

the sample, five well trained specialists who corrected the texts better suited to their respective areas and an expert rater especially versed in the use of the correction protocol who corrected the entire sample. Subsequently, with the aim of validating the developed tool, each indicator's mean values obtained by human raters were compared to the results obtained by *Seshat Product*. As it can be seen in Table 3 the inter-rater agreements obtained in both tests using the Cohen's Kappa coefficient were pretty good, surpassing the .80 threshold in almost every case. There is also certain uniformity between all values which is a sign of the high consistency of the final results.

DISCUSION AND CONCLUSION

In this thesis, some advances made in the field of the study of writing have been presented as part of the creation of *Seshat Writing Research Tools*, an application suite conceived to allow both the monitoring of the cognitive activity involved in the writing process as well as the automation of the psycholinguistic analysis processes of written texts. As it has been shown, this endeavour has resulted in the creation of two tools, *Seshat Process* and *Seshat Product*, which, having shown their viability in their respective fields of application, are also the

cornerstone of the present research and have become daily use programs at the Department of Psychology, Sociology and Philosophy of the Universidad de León. (Álvarez & García, 2014a, 2014b; Díez-Caso & García, 2013, under review a, under review b, Robledo & García, 2013).

Seshat Process has been designed to overcome the limitations traditionally attributed to studies performed on the field of the analysis of the writing process, such as the use of small samples caused by the need for participants to be present in the research location, and the required supervision of the process by qualified experts. In order to do this, the developed solution sets up a distributed workspace where participants access the writing composition exercises through the web and the system automatically registers the information related to their cognitive processes, allowing for the real-time analysis of the data. It has also been possible to integrate in a single separate web application all the elements required to implement Kellogg's triple task procedure, allowing the researcher to structure the experiment through a series of screens that deal with both the instruction of the participants as well as the performance of the proposed exercises.

There are, however, certain inherent limits to the nature of this system, among which it is worth pointing out the errors

arising from the lack of the required digital competence of the participants and the incorrect use of the platform itself due to lack of supervision. To avoid this, several solutions are being studied aimed both at improving the instruction process as well as simplifying the interface of *Seshat Process* in order to reduce the level of technological expertise required for its use. By way of example, we can point out the possibility of improving the implemented sequential system, allowing each of the various parts of the experiment to be carried out during separate sessions, as well as setting a maximum response time during the written exercise. Moreover, focusing on the researcher, among possible improvements to carry out in a near future we may note the increase of the system's analysis capabilities with the aim of allowing for the creation of permanent samples associated to the researcher's own profile, which could include participants from any study he has access to. This way, it is intended to grant a medium through which more complex comparative studies may be performed, using samples with differing typologies that allow the researcher to take into account new restrictions that are not currently available, such as a separation in terms of gender, level of studies or social and cultural background of the participants.

Seshat Product, in contrast, constitutes a system for the automatic processing of writing designed to automate the assessment and analysis processes of written texts according to a

series of rules created to promote the study of the written product in terms of linguistic features of productivity, coherence, structure and quality. Leaving aside the difficulty inherent to any attempt to process natural language, after several iterations of refinement, *Seshat Product* has achieved high rates of accordance with man-made corrections, surpassing an eighty percent rate on almost every single indicator implemented. This, together with the system's capability to analyze large amounts of data in mere seconds, has turned this tool into an effective method to carry out exhaustive research on specific psycholinguistic traits of productivity and coherence in Spanish (Álvarez & García, 2014a, 2014b; Díez-Caso & García, under review b, Robledo & García, 2013). However, unlike most of the current automated essay scoring systems, *Seshat Product* does not grant a direct global assessment of the quality of the analyzed text, but instead giving the user precise information on a series of particular linguistic indicators which require her interpretation. This design principle stems from the desire to grant the application a greater versatility as well as from the belief in the fundamental role that human input must play throughout the whole process of linguistic analysis, especially in such a complex field as the study of natural language.

Nowadays, *Seshat Product* only takes into account fifty percent of the indicators defined in the correction protocol that

encouraged its creation, the so-called text-based indicators. However, this limitation originates in the correction protocol itself, given the subjective nature of the remaining indicators it defines, which are intended for their manual application exclusively. Making these indicators, based on the reader's subjective criterion, objective, thus allowing for their subsequent automation, has currently become one of the main improvements to consider for the medium term, since this upgrade would not only mean a quantitative increase in the number of linguistic traits analyzed, but also a considerable increase in the tool's power to take on more complex analyses. All this being said, taking into consideration the results obtained so far, it seems undeniably necessary to simplify the correction protocol in a way that leads to a drastic cut in the extremely large number of current indicators, which, on most cases, are either repeated or do not provide the user with enough meaningful data to be an entity by themselves, at least from the point of view of an automated analysis. A possible solution might be to modify the correction protocol as far as reader-based measurements are concerned, seeing the text's final quality as a generic indicator encompassing the rest. In a similar vein, partial indicators could be interrelated; not so the basic indicators, which would remain independent so as to ease automation. In this sense, before starting any process of automatic analysis, the system should consider, for example,

the semantic field, the productivity limits, if any have been set, and the established average for the age range the studied text belongs to. In other words, the analysis of the text would be carried out by performing a regressive analysis in which the quality of the text is determined by using independent linguistic features evaluated according to the text's characteristics.

It must also be taken into account that the assessment of the texts performed by human reviewers is based on the free interpretation of certain correction guidelines which are often vague and hard to objectivize, and which end up becoming mere suggestions that the reviewer distills into a whole, using her own subjective criterion at the time of analyzing the text. It is then unlikely that we may establish general correction criteria which can be automated by taking the current reader-based indicators as a starting point, since we may assume that a human reviewer who corrects the same text twice will rarely grade it in exactly the same way whenever there is a significant time lapse between both assessments.

When considered as a whole, the *Seshat Writing Research Tools* project essentially constitutes an applications suite developed with the purpose of granting teachers and researchers an integrated solution to facilitate the study of writing from a double process-product perspective. Through the use of the

existing synergies between *Seshat Process* and *Seshat Product* this research thus intend to facilitate the performance of joint analyses that can explain the existing relationship between process and product in an environment that is constantly kept under the researcher's control; which can access the study's data in situ and analyze them in real time without having to use additional tools. This research also aims to provide writers taking part in these kinds of studies with real-time feedback containing precise information that allows them to improve their texts according to the features of those which have received higher grades, automatically identifying the main strengths and weaknesses of their writing process compared to rest of the participants involved in the same study.

Introducción general[1] 1

1.1 Introducción

El crecimiento exponencial de la información disponible debido a la generalización del uso de internet ha creado una creciente demanda de métodos de análisis automático del lenguaje natural en la sociedad moderna (Pęzik, 2010). En la actualidad, la incorporación de los ordenadores al ámbito del procesamiento de la información escrita ha supuesto importantes avances en campos tan dispares como la economía (De Souza et al., 2007), la medicina (Pakhomov, Chacon, Wicklund, & Gundel, 2011) o la educación (Hastings, Hughes, Magliano, Goldman, & Lawless, 2012; Magliano & Graesser, 2012). En el ámbito educativo en concreto, estos avances han resultado ser particularmente significativos, dado que responden a un cambio

[1] La presente investigación recibió financiación procedente de proyectos competitivos concedidos a Grupos de Investigación de Excelencia de la Junta de Castilla y León cofinanciados por la Unión Europea para el trienio 2009-2011 (GREX259; BOCyL 27 de abril de 2009) y de un proyecto competitivo del MICINN (EDU2010-19250/EDUC, 2010-2013), figurando J. N. García como investigador principal en ambos proyectos. Asimismo, la presente investigación recibió financiación procedente de un contrato competitivo cofinanciado por la Junta de Castilla y León y el Fondo Social Europeo concedido a H. Díez Caso (EDU/1867/2009, 2009).

paradigmático en la enseñanza auspiciado por la desenfrenada evolución tecnológica acaecida durante los últimos veinte años.

Se puede afirmar sin temor a equivocarse que la proliferación de las Tecnologías de la Información y la Comunicación (TICs) y el uso masivo de dispositivos con acceso a internet ha transformado la realidad de nuestro sistema educativo. El acceso instantáneo a gran cantidad de información, a bajo o ningún coste, ha dado lugar a una revolución tecnológica que ha supuesto un cambio en los paradigmas educativos tradicionales y, en consecuencia, hemos podido comprobar como el ordenador se ha ido convirtiendo, poco a poco, en una herramienta de trabajo imprescindible en el aula.

Cada vez son más numerosos y notables los esfuerzos por introducir la tecnología en las escuelas, e instituciones de enseñanza en todo mundo se apresuran a actualizar sus procedimientos para adaptarse a la nueva situación. Bien sea mediante la utilización de plataformas como Moodle o WebTC o desarrollando herramientas propias adaptadas a necesidades específicas (Muñoz de la Peña, Gómez-Estern & Dormido, 2012). No es de extrañar, por tanto, la creciente preocupación que ha surgido acerca del impacto de las TICs en la educación (Cha et al., 2011; Hernández-Ramos, Martínez-Abad, García Peñalvo, Esperanza Herrera García, & Rodríguez-Conde, 2014;

Ott & Pozzi, 2011; Zhong, 2011), así como por el desarrollo de competencias digitales tanto en profesores (Butter, Pérez, & Quintana, 2014; Koh, 2011) como en alumnos de ciclos formativos básicos (Hatlevik & Christophersen, 2013).

A día de hoy, los estudiantes se enfrentan a retos durante su formación académica que difieren radicalmente de los que se afrontaron en generaciones anteriores, haciéndose necesaria su instrucción en áreas competenciales al margen del currículum tradicional, tales como el aprendizaje digital o la búsqueda y manejo de la información y material multimedia (Trilling & Fadel, 2009). En este sentido, es innegable que la escritura ha asumido un rol predominante dentro de este nuevo marco competencial, erigiéndose como el instrumento comunicativo por excelencia en el mundo digital. Esta nueva realidad ha promovido, asimismo, una profunda transformación de la forma en que los estudiantes escriben y se comunican (Herrington, Hodgson, Moran, & Eidman-Aadahl, 2009).

De acuerdo con Haydel y Bucolo (2010), actualmente no sólo se espera que los estudiantes adquieran destreza en disciplinas tradicionales como la redacción de trabajos de investigación o ensayos argumentativos, sino también en el uso de otras formas de comunicación digitales tales como los correos electrónicos, los blogs y las redes sociales, entre otros. El

dominio de la escritura, por tanto, no sólo sigue siendo un área fundamental del currículum escolar, sino que ha cobrado renovada importancia. No obstante, tanto el modo en que esta materia se imparte, como la forma en que se evalúa, han sufrido una profunda transformación en los últimos años debido al ya mencionado impacto de la generalización del uso de los ordenadores e internet.

Asimismo, esta nueva realidad ha proporcionado a los investigadores la oportunidad de abordar el estudio tanto de los procesos cognitivos implicados en la producción escrita, como del producto final obtenido, en el contexto de una situación totalmente distinta. Si bien el planteamiento del estudio de la escritura, bien sea como proceso o como producto, no es nuevo, el auge de la incorporación de las nuevas tecnologías en el ámbito educativo ha impulsado el desarrollo de una serie de herramientas diseñadas para facilitar la labor de docentes e investigadores a este respecto. Especialmente relevantes resultan la monitorización computerizada de procesos cognitivos y el análisis automático de la escritura; siendo la integración de ambas tecnologías en un medio común a través de la creación de sendas herramientas para analizar tanto el proceso como el producto de escritura el principal objeto de interés de esta tesis.

A continuación se presenta una breve síntesis de los contenidos de este trabajo de investigación atendiendo a cada uno de los elementos que lo conforman de manera individualizada y prestando atención tanto a las revisiones teóricas realizadas como a los estudios empíricos que auspiciaron la creación de *Seshat Writing Research Tools*[1], una suite de aplicaciones informáticas diseñada para facilitar el estudio de la escritura desde una perspectiva generalista. Al final de este capítulo se proporciona, asimismo, una descripción detallada de los principales objetivos de la presente investigación así como de las causas que motivaron el desarrollo de las mencionadas aplicaciones.

1.2 Síntesis descriptiva de los contenidos

La presente memoria consta, en esencia, de dos partes complementarias nombradas cada una de ellas atendiendo a la naturaleza de los estudios realizados, a saber, la monitorización de procesos cognitivos y la evaluación automática del producto de escritura. Cada uno de estos dos grandes bloques comporta a su vez dos capítulos, un primer capítulo destinado a realizar una profunda revisión de los fundamentos teóricos que cimentan la disciplina objeto de estudio y un segundo capítulo en el que se

[1] El nombre Seshat hace referencia a diosa de la escritura y la historia en la mitología egipcia, protectora de las bibliotecas y Señora de los libros.

abordan pormenorizadamente los detalles del estudio llevado a cabo atendiendo a su concepción, metodología empleada, resultados finales obtenidos, conclusiones y prospectiva. Se cuenta asimismo con una última sección, denominada discusión y conclusiones generales, incluida a efectos de analizar la relevancia conjunta de los estudios realizados, las dificultades y limitaciones encontradas, las posibles mejoras a implementar a corto o medio plazo y el grado de consecución de los objetivos planteados.

Esta investigación en su conjunto, así como la totalidad de los contenidos dispuestos en esta memoria, cuentan con el aval de la comunidad científica, manifiesto a través de diversas publicaciones en medios tanto nacionales como internacionales (Díez-Caso & García, 2011a, 2011b, 2011c, 2013, under review a, under review b; Díez-Caso, García, García Martín, & Álvarez, 2013).

1.2.1 Monitorización de procesos cognitivos

Centrado en la captación y análisis de la actividad mental del individuo a la hora de abordar una tarea de composición escrita, este primer bloque de contenido se divide en dos capítulos atendiendo respectivamente a la revisión de los

fundamentos teóricos de este campo de estudio y a la exhaustiva presentación de la solución desarrollada a tal efecto, *Seshat Process*.

En el capítulo dos, primero de esta sección, se aborda, con intención de establecer un marco teórico apropiado, una breve revisión de la evolución y consolidación de los modelos y métodos de investigación del proceso de escritura. Se ha creído conveniente incorporar esta revisión no sólo a modo de justificación implícita de los procedimientos implementados por herramienta desarrollada, sino para poner de manifiesto los múltiples puntos de vista acerca de este tema e ilustrar las diversas aplicaciones del software desarrollado.

Seguidamente, en el capítulo tres, segundo de esta sección, se procede a realizar una exhaustiva revisión de todos los aspectos técnicos y funcionales del sistema desarrollado, *Seshat Process*, adoptando, para ello, tres perspectivas diferentes, la de cada uno de los roles implementados: *student*, *teacher* y *administrator*. Esta presentación individualizada de la tipología de usuarios se hace necesaria, dada la naturaleza del entorno de trabajo planteado, diseñado para operar de forma asíncrona, limitando, intencionadamente, la interdependencia entre cada uno de ellos. Como parte de este proceso, se profundiza igualmente en la capacidad del sistema para implementar el procedimiento

de la triple tarea de Kellogg (Kellogg, 1999; Piolat, Kellogg, & Farioli, 2001), combinando el proceso de escritura, la medición del tiempo de reacción ante determinados estímulos y la categorización de los contenidos de la memoria a corto plazo en el momento en que se producen. Asimismo, continuando con la revisión de los aspectos funcionales de *Seshat Process*, se detallarán las capacidades del sistema para procesar la información recabada, realizar análisis en tiempo real y exportar las matrices de datos resultantes en formatos compatibles con la mayoría de los softwares estadísticos actuales.

Finalmente, a modo de validación y con intención de ilustrar con mayor detalle el funcionamiento general de *Seshat Process* en un entorno de trabajo real, se presenta un caso de estudio en el que la plataforma fue puesta a prueba como parte de un curso online que contó con 366 participantes. Este estudio cumplía un doble objetivo. En primer lugar comprobar la viabilidad del sistema para operar de forma autónoma en un entorno heterogéneo, bajo nula o mínima supervisión, y en segundo lugar, poner a prueba la capacidad de la plataforma para realizar análisis en tiempo real de los datos recabados a medida que estos eran introducidos en el sistema.

1.2.2 Evaluación automática de textos escritos

El objetivo de esta sección es presentar *Seshat Product*, una herramienta de corrección automática de textos en español diseñada para facilitar el análisis de rasgos lingüísticos de productividad y coherencia en base a un protocolo de corrección diseñado para analizar determinadas particularidades de este idioma con el fin de facilitar la detección de posibles dificultades de aprendizaje. Al igual que ocurriese con el anterior, este segundo bloque de contenido consta de dos capítulos, un primer capítulo centrado en la fundamentación teórica de la investigación llevada a cabo y un segundo capítulo en el que se realiza una profunda revisión del sistema desarrollado.

En el capítulo cuatro, primero de esta sección, se aborda, con objeto de contextualizar adecuadamente la investigación realizada, una breve revisión de los fundamentos teórico-prácticos que rigen el funcionamiento de los sistemas de análisis automático de la escritura en la actualidad. Para ello se hace especial hincapié en el análisis de los indicadores de productividad y coherencia más comúnmente utilizados en este tipo de procesos. Se incluye asimismo una breve presentación de las principales soluciones disponibles en el mercado a día de hoy, así como de las técnicas y metodologías empleadas en su creación.

En el capítulo cinco, segundo de esta sección, se acomete la presentación del protocolo de corrección implementado por *Seshat Product*, así como la de la solución desarrollada en sí misma, atendiendo a una exhaustiva revisión de todos y cada uno de los aspectos técnicos y funcionales del programa. Esta presentación se realiza, al igual que ocurriese en el caso de *Seshat Process*, adoptando la perspectiva de un usuario tipo con el fin de poder analizar minuciosamente cada particularidad del sistema implementado, desde la relevancia de los rasgos lingüísticos evaluados hasta el flujo de trabajo general de la aplicación o el nivel de personalización ofrecido por su interfaz. Asimismo, a modo de validación, se aborda el análisis de una muestra de quinientos treinta y siete textos con el fin poner de manifiesto el alto índice de acuerdo alcanzado entre los resultados del análisis automático y los obtenidos por seis evaluadores humanos expertos.

1.3 Objetivos

A continuación se describen pormenorizadamente los objetivos de la presente investigación analizando conjunta e individualmente cada uno de los estudios realizados con el fin de permitir al lector evaluar el grado de consecución de las metas

planteadas a tenor de los resultados presentados a lo largo de esta tesis.

Haciendo uso de los avances tecnológicos disponibles en el campo de la educación a distancia y el procesamiento automático de la palabra escrita, la finalidad de este proyecto radica en la creación de una serie de herramientas informáticas ideadas para abordar el estudio tanto de la actividad cognitiva implicada en el proceso de escritura como del producto final obtenido fruto de dicho proceso. La suite de aplicaciones resultante de este empeño, denominada *Seshat Writing Research Tools*, consta así de dos componentes básicos fuertemente interrelacionados entre sí, *Seshat Process* y *Seshat Product*, destinados cada uno de ellos a satisfacer de forma complementaria cada una de las perspectivas planteadas en materia de estudio de la escritura.

En relación por tanto al primer estudio planteado en esta tesis, la monitorización de procesos cognitivos, el objetivo primordial consiste en la creación de una aplicación, *Seshat Process*, diseñada para acometer la implementación del procedimiento de la triple tarea de Kellogg de forma no supervisada y valiéndose señales de interrupción visuales en sustitución de las clásicas señales auditivas. Asimismo se pretende hacer uso de internet y los avances actuales en al campo

de la educación a distancia con el fin de hacer frente a tres limitaciones habituales en este tipo de estudios. A saber, la necesidad de usar varios instrumentos para poder reproducir íntegramente el procedimiento ideado por Kellogg, el reducido número de individuos que pueden participar en los estudios de forma simultánea y la necesidad de contar con la presencia de expertos que monitoricen el proceso en determinados momentos. La solución propuesta ha de pasar por tanto por el planteamiento de un entorno de trabajo distribuido en el que los participantes accediesen a los ejercicios vía web y el sistema registrase automáticamente la información de sus procesos cognitivos, permitiendo, asimismo, analizar los datos recabados en tiempo real. Se pretende por tanto no sólo plantear un medio eficaz para recabar información acerca de los procesos cognitivos implicados en el proceso de escritura, sino proporcionar una forma rápida y eficiente de recoger los textos de los participantes en cualquier estudio de estas características sin requerir su presencia física a la hora de realizar los ejercicios de composición escrita.

El segundo estudio planteado, la evaluación automática del producto escrito, se centra a su vez en la creación de *Seshat Product*, una aplicación ideada para automatizar el análisis de textos escritos a fin de generar informes exhaustivos basados en un acervo preestablecido de reglas bien definidas. Estas reglas, recogidas en un protocolo de corrección creado a efectos de

estandarizar la aplicación manual del proceso de evaluación psicolingüística, se convierten asimismo en un objetivo secundario, a la par que indispensable de la investigación, al requerirse su simplificación y objetivación como parte del proceso previo a la automatización de las mismas.

Por otra parte, la creación de *Seshat Product* plantea una revisión exhaustiva de los métodos, técnicas y soluciones implementadas hasta la fecha en el campo de la evaluación automática de la escritura. Asimismo se plantea el uso, tanto en la creación como en la validación de la herramienta, de un corpus lingüístico orientado al estudio del desarrollo de la escritura en el ámbito escolar y creado a partir de más de veinte mil textos escritos por estudiantes españoles con edades comprendidas entre los ocho y los dieciséis años, con y sin dificultades para el aprendizaje. Esta base documental, clave en el proceso de refinamiento del protocolo de corrección, así como de los patrones de búsqueda empleados por *Seshat Product*, ha sido creada fruto del esfuerzo de los diferentes investigadores del equipo de Excelencia de la Junta de Castilla y León GR259 y sus colaboradores haciendo especial hincapié en la composición escrita como vehículo para la investigación de la escritura atendiendo a factores de productividad, calidad, estructura y coherencia del texto (Arias & García, 2007; de Caso, García, Díez, Robledo, & Álvarez, 2010; Fidalgo & García, 2008;

García, 2002; García & Arias-Gundín, 2004; García & de Caso, 2002, 2004, 2006a, 2006b, 2007; García & Fidalgo, 2003, 2004, 2006, 2008a, 2008b, 2008c; García & Marbán, 2003; García, Rodríguez-Pérez, Pacheco, & Díez, 2009; Pacheco, García, & Díez, 2009, 2010a, 2010b; Torrance, Fidalgo, & García, 2007).

Al margen de los objetivos principales, este proyecto plantea asimismo dos objetivos secundarios producto inherente del desarrollo de las mencionadas aplicaciones *Seshat Process* y *Seshat Product*, como son, la contribución a la creciente digitalización del aula y la promoción del español.

Una de las posibilidades más atractivas de la generalización del uso de herramientas como las planteadas en esta investigación es proporcionar a psicólogos y docentes un medio para realizar evaluaciones automatizadas in situ de aspectos psicológicos asociados a la escritura. Se pretende por tanto, no sólo crear una aplicación informática orientada al campo de la investigación a nivel universitario, sino una herramienta accesible y de uso diario capaz de contribuir a la labor desarrollada en cualquier centro de enseñanza que precise de este tipo de sistemas. De este modo se prevé poner a disposición del profesorado la tecnología necesaria para llevar acabo análisis textuales en tiempo real y generar informes

exhaustivos basados en la experiencia de décadas de investigación psicolingüística.

Resultan asimismo de suma importancia las consecuencias derivadas de esta investigación en cuanto a la promoción del español se refiere. Atendiendo al elevado número de hispanohablantes existentes en la actualidad, más de cuatrocientos noventa y cinco millones según el Instituto Cervantes (2012), el análisis de aspectos psicológicos vinculados a la expresión escrita podrá aplicarse a diversos ámbitos socioculturales del mundo hispano, así como servir de vehículo para realizar estudios comparativos entre hispanohablantes nativos y estudiantes de español extranjeros.

Monitorización de procesos cognitivos

Antecedentes I[1] 2

2.1 Introducción

Aunque la esencia de la composición escrita se mantiene invariable, la influencia de las nuevas tecnologías en al ámbito educativo proporciona a los investigadores los medios necesarios y la oportunidad de estudiar los procesos cognitivos asociados a la escritura en un contexto totalmente diferente. En consecuencia, en los últimos años se han desarrollado multitud de métodos y herramientas para abordar la investigación de la escritura adoptando diversos enfoques. De acuerdo con Olive (2010) podemos distinguir principalmente tres: la evaluación de la fluidez mediante el análisis de pausas y periodos de ejecución, el estudio de las características funcionales de los procesos de escritura a través de think-aloud protocolos y la medición del

[1] La presente investigación recibió financiación procedente de proyectos competitivos concedidos a Grupos de Investigación de Excelencia de la Junta de Castilla y León cofinanciados por la Unión Europea para el trienio 2009-2011 (GREX259; BOCyL 27 de abril de 2009) y de un proyecto competitivo del MICINN (EDU2010-19250/EDUC, 2010-2013), figurando J. N. García como investigador principal en ambos proyectos. Asimismo, la presente investigación recibió financiación procedente de un contrato competitivo cofinanciado por la Junta de Castilla y León y el Fondo Social Europeo concedido a H. Díez Caso (EDU/1867/2009, 2009).

esfuerzo cognitivo que conllevan empleando la técnica de la doble tarea. Atendiendo a estos dos últimos enfoques, esta investigación se centra en el estudio general de los procesos cognitivos involucrados en la escritura a través de la plataforma web desarrollada a tal efecto, *Seshat Process*.

A continuación se presenta una breve revisión de los modelos de procesos y métodos de investigación de la escritura que cimentan este estudio. Este marco teórico hace especial hincapié en el cambio de paradigma que supuso la concepción de la escritura como un proceso en lugar de un producto. Así como en la importancia de entender esta actividad como una tarea global que involucra otras capacidades y recursos del individuo tales como la memoria a largo plazo y la memoria operativa.

2.2 Modelos de proceso de escritura

Sin duda, escribir es una de las tareas más complejas que puede llevar a cabo el ser humano, pues implica la coordinación de diversas actividades mentales interviniendo a distintos niveles de representación (Kellogg, 1999; Piolat, Olive & Kellog, 2005). Un primer nivel semántico que engloba todas las actividades cognitivas referentes a la planificación, es decir, la recuperación de la memoria a largo plazo de las ideas que el autor quiere

transmitir. Un segundo nivel lingüístico en el que tienen lugar los procesos cognitivos necesarios para llevar a cabo la transformación de este mensaje pre-verbal, construido en el nivel semántico, en lenguaje escrito. Un tercer nivel que comprende, a su vez, la trascripción final de dicho mensaje, operando a nivel motor. Y finalmente, un cuarto nivel, que podría llegar a involucrar a los tres anteriores, implicando todas aquellas actividades relativas a la monitorización de la calidad e idoneidad del texto escrito (Olive, 2010).

Asumiendo por tanto el punto de vista de esta perspectiva cognitivista, podemos aseverar que la escritura requiere complejos procesos de reflexión que la convierten en algo más que una mera actividad motora (Tynjälä, Mason & Lonka, 2001). Este paradigma de la psicología cognitiva se ha desarrollado durante los últimos 30 años, evolucionado a partir del modelo original de escritura presentado por Hayes y Flower en la década de los ochenta.

El modelo de Hayes y Flower (1980) supuso un cambio radical en la forma de entender la investigación de la escritura, prescindiendo por primera vez del enfoque tradicional, basado en el producto, y centrándose en describir el proceso de escritura a través de las diversas actividades cognitivas que tienen lugar en la mente del escritor. De esta forma, los estudios realizados por

Hayes y Flower para determinar la naturaleza de dichas actividades les llevaron a estructurar su modelo de acuerdo a tres componentes principales: el contexto de producción, la memoria a largo plazo del escritor y el proceso general de escritura. El contexto de producción comprendía el conjunto de circunstancias que inducen la tarea y la situación comunicativa, es decir, el tema sobre el que se escribe, el medio, la audiencia, etc. La memoria a largo plazo hacía referencia, a su vez, a todos los datos que el autor conoce sobre el tema en cuestión, en otras palabras, aquellos saberes y experiencias personales a los que recurriría a la hora de abordar la tarea. Por último, el proceso de escritura se definía en base a otros tres subprocesos cognitivos básicos: la planificación, la traducción y la revisión; todos ellos gobernados a su vez por un proceso general de monitorización que gestionaba su interacción y permitía usarlos de forma recursiva.

Este modelo es considerado la piedra angular sobre la que se ha sustentado la investigación de los procesos cognitivos en la escritura durante los últimos treinta años. No obstante, con el paso del tiempo ha sufrido numerosas modificaciones, pues a pesar de su gran aceptación, también fue objeto de revisiones críticas que le achacaban el no prestar atención a las dimensiones lingüísticas de la escritura (García Debanc y Fayol, 2002) y haberse basado exclusivamente en el los procesos cognitivos de escritores experimentados, excluyendo a aquellos cuya habilidad

para escribir estaba aún por desarrollar. En consecuencia, el propio Hayes (1996) presentó una revisión de su modelo original más de veinte años después. En esta ocasión, Hayes proponía una profunda reconceptualización de los procesos cognitivos descritos anteriormente, eliminando las distinciones externas basadas en la tarea y centrándose en tres procesos cognitivos básicos: la interpretación del texto, la reflexión y la producción del texto. Cabe destacar también que el concepto de revisión, ahora interpretación textual, cobra una especial importancia en este modelo, haciendo referencia tanto a una revisión a nivel conceptual como lingüístico por parte del autor, quien podría optar por editar el texto o no hacerlo en función de si existen o no discrepancias significativas conforme a su planteamiento inicial. Hayes considera, por tanto, que este proceso está íntimamente relacionado con el esquema general de la tarea y fuertemente influenciado por algunos recursos inherentes al individuo. Esto le induce a tratar de identificar como interaccionan determinados aspectos de la capacidad cognitiva del escritor con estas tareas, diferenciando el papel que juegan la memoria a largo plazo, la memoria operativa y la motivación.

Por otra parte, la corriente que inició el cambio de paradigma promovido por los estudios de Hayes y Flower, suscitó la aparición de otros modelos cognitivos tratando de complementar al original. Tal es el caso de los modelos

propuestos por Bereiter y Scardamalia (1987; Scardamalia y Bereiter, 1992). Bereiter y Scardamalia propusieron dos modelos basados en las diferencias existentes entre neófitos y escritores experimentados, a los que llamaron respectivamente knowledge-telling y knowledge-transforming. Knowledge-telling era un modelo que trataba de explicar los primeros estadios del aprendizaje de la escritura. Ilustraba un modo de escribir carente de planificación u organización previa en el que el escritor accedía a su memoria a largo plazo para recuperar la información que ya tuviese acerca del tema en cuestión y la plasmaba sobre el papel. La producción del texto se llevaba a cabo sin ningún tipo de esquema u objetivo, puesto que el escrito carecía de los recursos necesarios para incorporarlos a la tarea. En otras palabras, el escritor novel desarrollaba sus ideas en función de sus conocimientos previos sobre la materia sin tener en cuenta las necesidades de la audiencia a la que iba dirigido el texto. En cambio, el modelo knowledge-transforming sí tenía en consideración al lector y, en este caso, el escritor adaptaba el texto en función de la audiencia a la que iba dirigido. En este modelo, por tanto, la recuperación de información y su traducción era arbitrada por un proceso de reflexión que involucraba una interacción entre el espacio de los problemas retóricos y el de los de contenido. En otras palabras, a diferencia de los neófitos, los escritores experimentados eran capaces de

construir complejas representaciones mentales de la tarea tanto a nivel lingüístico como semántico. En consecuencia, el texto ya no era un reflejo directo de los conocimientos previos del escritor, sino un producto de la organización de ese conocimiento en función de unos objetivos concretos.

No obstante, aunque estos modelos profundizaban en el estudio de los procesos cognitivos involucrados en el desarrollo de la escritura, no servían para explicar las habilidades requeridas en los primeros estadios de su aprendizaje (Berninger, 1999). En base a esta carencia, Bereiter y Scardamalia retoman el modelo de Hayes y Flower distinguiendo dos subprocesos dentro del proceso general de textualización. La generación textual, que conlleva la transformación de las ideas del escritor en lenguaje escrito, y la transcripción, que comprende los procesos de bajo nivel como la ortografía o las habilidades grafomotoras. A raíz de esta nueva propuesta, estudios posteriores demostrarían que la falta de desarrollo de estas habilidades en las primeras etapas del desarrollo de la escritura suponía el empleo de gran cantidad de recursos atencionales en términos de memoria operativa (Torrance & Galbraith, 2006). Asimismo, se postuló que los escritores más experimentados automatizaban completamente las tareas que conllevan estas habilidades básicas, liberando recursos de la memoria operativa que podían ser destinados a otras tareas tales como la planificación o la organización del texto.

Como puede observarse, a medida que estos modelos han ido evolucionando, las continuas menciones a la memoria operativa son cada vez más frecuentes. Este concepto ha sido objeto de estudio desde que el término fuese acuñado por Miller, Galanter y Pribram en 1960 y usado por primera vez para referirse a la memoria a corto plazo por Atkinson y Shiffrin (1968). No obstante, tuvieron que pasar casi cuarenta años para que Kellogg (1996) lo incorporase por primera vez como parte de un modelo de procesamiento de la escritura que integraba, en un único modelo, tanto los procesos de escritura, como un sistema de procesamiento de la información. En la propuesta de Kellogg, dicho sistema de procesamiento de la información hacía referencia a la memoria operativa y era representado por el modelo clásico desarrollado por Baddeley (1992) que constaba fundamentalmente de tres elementos básicos. Un componente principal, llamado Ejecutivo Central, encargado de los procesos complejos como el razonamiento y la reflexión, y dos registros subordinados, la Agenda Viso-Espacial y el Buble Fonológico, que permitían mantener representaciones visuales y fonológicas (Alamargot & Chanquoy, 2001). En lo referente a la arquitectura de los procesos de escritura, Kellogg sintetizó la el modelo elaborado por Brown y McDonald, Brown y Car (1988) en el que se distinguían tres componentes principales, el componente de formulación, el de ejecución y el de monitorización. El

componente de formulación involucraba los procesos de planificación y traducción, el de ejecución la elaboración motora y la realización del mensaje y finalmente, el de monitorización implicaba la lectura del texto y su posible edición a posteriori a criterio del autor.

Aunque existen ciertas similitudes con los modelos anteriores, dado que todos ellos describen los procesos cognitivos en la escritura, el modelo de Kellogg se centra fundamentalmente en las interacciones entre dichos procesos y los diferentes componentes de la memoria operativa (Alamargot & Chanquoy, 2001). En este aspecto, resulta de especial interés que sus estudios le llevasen a postular, paralelamente a Bereiter y Scardamalia, la importancia de automatizar los procesos que conllevan ciertas habilidades básicas y la consecuente liberación de recursos de la memoria operativa que esto supone (Olive & Kellogg, 2002).

2.2 Métodos de investigación

De acuerdo con Janssen, Van Waes y Van den Bergh (1996), podemos clasificar los métodos de investigación de la escritura en métodos de observación síncronos o asíncronos atendiendo al momento en que se recoge la información. De este

modo, se consideran métodos síncronos a todos aquellos en los que la información acerca de los procesos cognitivos se recoge durante el proceso de escritura y asíncronos a aquellos en los que dicha información es recogida a posteriori. Dado que el interés de esta investigación se centra tanto en la categorización de los procesos como en la medición del esfuerzo cognitivo que conllevan, en el desarrollo de *Seshat Process* se impuso el uso de métodos de observación síncronos. Por esta razón, se decidió emplear una variante de los think-aloud protocols (TAPs), un método tradicional en la psicología cognitiva que fue modificado por Kellogg en su procedimiento de la triple tarea para permitir al escritor categorizar los procesos de motu propio.

Los TAPs han sido una de las principales maneras de abordar el estudio de los procesos cognitivos en psicología. No obstante, aunque los orígenes de este método se remontan al s. XIX, su uso no se hizo extensivo al campo de la escritura hasta la aparición del modelo de Hayes y Flower a principios de la década de los ochenta (Hayes & Flower, 1980). El cambio paradigmático que supuso este modelo en el campo de la investigación de la escritura, concebida por primera vez como un proceso, alentó al mismo tiempo el interés por los métodos de investigación que permitiesen detectar los procesos cognitivos implicados en la composición escrita.

El empleo de TAPs se basa en la asunción de que el individuo puede reconocer y expresar verbalmente algunos de sus procesos mentales así como algunas de las operaciones que estos conllevan; sin embargo, sólo los procesos controlados pueden ser objeto de verbalización. En otras palabras, el individuo sólo puede informar acerca de los procesos que operan bajo el control de su memoria operativa, pero no de aquellos que no la implican. De tal forma que, en el caso de escritores experimentados, los procesos derivados de habilidades básicas que han llegado a usarse de forma inconsciente o automatizada quedarían fuera de los límites de los TAPs. Esta limitación no es excesivamente problemática en la investigación de la escritura, dado que la mayoría de los procesos involucrados en la composición escrita interactúan con la memoria operativa (Olive, 2010). Prueba de ello es que los propios Hayes y Flower utilizaron este método para describir por primera vez el proceso de escritura, mostrándolo no como una sucesión lineal de procesos cognitivos, sino como un mecanismo recursivo en el que el escritor puede hacer uso de un proceso u otro a su conveniencia.

No obstante, a pesar de su innegable utilidad, los TAPs presentan varias limitaciones adicionales que también han de tenerse en consideración. Una de sus particularidades más criticadas ha sido su interferencia con los procesos cognitivos

normales del escritor. En otras palabras, las preguntas que se hacen al individuo mientras está escribiendo pueden llegar a interrumpir su línea de pensamiento habitual. Numerosos estudios han confirmado empíricamente esta afirmación (Krings, 2001; Jakobsen, 2003), llegando a apreciarse retrasos de entre el 25% y el 33% en la realización de la tarea (Janssen, et al., 1996). Esta demora evidencia, asimismo, que los procesos de bajo nivel, como los sintácticos, léxicos u ortográficos, son especialmente susceptibles de ser alterados por el uso de este tipo de métodos durante la redacción. Por otra parte, el hecho de que los escritores puedan llegar a activar más de un proceso al mismo tiempo (Olive & Kellogg, 2002), supone otra importante limitación de los TAPs, dado que los escritores no pueden indicar qué procesos se están ejecutando simultáneamente a tenor de que dicho paralelismo es un signo de fluidez y en cierta medida de procesos automatizados (Olive, 2010). Esta capacidad inherente al ser humano de ejecutar varios procesos cognitivos a la vez es un hecho sobradamente conocido en psicología cognitiva que alentó en su momento el desarrollo de otros métodos de estudio como el procedimiento de la doble tarea (Kahneman, 1973; Kerr, 1973; Power, 1986).

La doble tarea es una técnica que se utiliza para determinar la naturaleza de las representaciones mentales o en el caso de la escritura, de los procesos cognitivos involucrados en la

producción escrita. Asimismo, también es empleada para medir la cantidad de recursos cognitivos empleados en una determinada actividad. Esta técnica se basa en el principio de que estos recursos son limitados pues están directamente relacionados con la memoria operativa del individuo. Por lo tanto, en el supuesto de tener que ejecutar dos actividades simultáneamente, se produciría un conflicto que resultaría en una disminución del rendimiento en relación al nivel alcanzado en condiciones normales, es decir, si cada tarea se hubiese ejecutado por separado. Kellogg modificó este método creando su propia variante de la técnica de la doble tarea en lo que posteriormente se daría en llamar el método de la triple tarea (Kellogg, 1987). Este nuevo método constaba de tres elementos, un ejercicio de composición escrita, la detección de una señal auditiva que tenía lugar mientras se estaba escribiendo y una retrospección por parte del escritor, que debía describir lo que estaba haciendo coincidiendo con el momento en el que percibía el estímulo auditivo. De esta forma Kellogg combinó la doble tarea con los TAPs creando un procedimiento que permitía medir el esfuerzo cognitivo concerniente a cada proceso de la composición escrita, así como su frecuencia de uso. Actualmente existen dos variantes de la triple tarea. En la primera se hace un uso convencional de los TAPs en el que los escritores informan de las actividades que están llevando a cabo y los investigadores interpretan dicha

información. Mientras que en la segunda, los escritores son entrenados para identificar sus propios procesos cognitivos durante el ejercicio escrito. Esta segunda variante, que ha sido tomada como referencia en la creación de *Seshat Process*, fue empleada en primera instancia por Kellogg, quien probó su eficacia para lidiar con algunos de los problemas tradicionalmente atribuidos a los TAPs como evidenciarían estudios posteriores (Kellogg, 1987b; Piolat, Roussey, Olive, & Farioli, 1996; Ransdell, 1995; Piolat, Olive, Roussey, Thunin, & Ziegler, 1999; Piolat, Kellogg & Farioli, 2001). De esta forma se demostró que el procedimiento de la triple tarea ideado por Kellogg no interfería con las actividades cognitivas durante el proceso de escritura y tampoco influía en la calidad final del texto.

Seshat Process[1] 3

3.1 Introducción

Seshat Process ha sido diseñado para implementar el procedimiento de la triple tarea de Kellogg y analizar el proceso de escritura adoptando un enfoque que permita hacer frente a tres limitaciones habituales en estudios de este tipo. A saber, la necesidad de usar varios instrumentos para poder reproducir íntegramente el procedimiento, el reducido número de individuos que pueden participar en los estudios de forma simultánea y la necesidad de contar con la presencia de expertos que monitoricen el proceso en determinados momentos. La solución propuesta plantea un entorno de trabajo distribuido en el que los participantes acceden a los ejercicios vía web y el sistema registra automáticamente la información de sus procesos

[1] La presente investigación recibió financiación procedente de proyectos competitivos concedidos a Grupos de Investigación de Excelencia de la Junta de Castilla y León cofinanciados por la Unión Europea para el trienio 2009-2011 (GREX259; BOCyL 27 de abril de 2009) y de un proyecto competitivo del MICINN (EDU2010-19250/EDUC, 2010-2013), figurando J. N. García como investigador principal en ambos proyectos. Asimismo, la presente investigación recibió financiación procedente de un contrato competitivo cofinanciado por la Junta de Castilla y León y el Fondo Social Europeo concedido a H. Díez Caso (EDU/1867/2009, 2009).

cognitivos, permitiendo, asimismo, analizar los datos recabados en tiempo real. Por lo tanto, aun suscribiendo las ideas de Kellogg, se puede afirmar que la herramienta desarrollada plantea una nueva manera de abordar el estudio del proceso de escritura.

Adoptando la simplicidad como una máxima de diseño, *Seshat Process* permite al investigador estructurar el experimento en una serie de fases o pantallas que integrarán tanto la instrucción como la realización del ejercicio propuesto. El participante accederá a estas pantallas siguiendo una sucesión lógica que le guiará del entrenamiento en el reconocimiento de unos procesos mentales específicos, previamente categorizados a conveniencia de los objetivos de la investigación, a la posterior realización del ejercicio escrito en el que deberá identificarlos. Esta identificación, a su vez, se llevará a cabo mediante un método de retrospección directa, sometiendo al participante a una serie de estímulos durante la redacción e instándole a identificar lo que está haciendo en el momento en que se produzcan. Al igual que en los experimentos llevados a cabo por Kellogg, el tiempo de respuesta ante dichos estímulos, contabilizado en milisegundos, servirá para determinar el esfuerzo cognitivo asociado; no obstante, en esta ocasión, la naturaleza del estímulo no será auditiva sino visual. El objetivo primordial de plantear el

experimento de esta manera, es permitir que el participante pueda acometer la tarea en una sola sesión y sin ayuda externa.

Por otra parte, al haber sido diseñado como una plataforma web, este sistema también se beneficia por ende de las ventajas que ofrece un entorno educativo a distancia. Como por ejemplo, el hecho de no requerirse la presencia física de los usuarios en un aula, evitando, asimismo, ulteriores inversiones en equipos o licencias de software adicionales. Otro claro ejemplo es la movilidad. Los usuarios pueden acceder al sistema empleando sus propios equipos las veinticuatro horas del día sin importar dónde se encuentren, pudiendo de esta forma llevar a cabo las tareas programadas sin necesidad de supervisión por parte de los investigadores. Asimismo, desde un punto de vista puramente técnico, una de las mayores ventajas que presenta la plataforma es el acceso a través de navegadores convencionales, evitando posibles problemas de incompatibilidad y garantizando su funcionamiento en la mayoría de los sistemas actuales.

A continuación se describen algunos aspectos técnicos de la herramienta que se ha creído conveniente resaltar en aras de contribuir a una mejor comprensión del sistema en su conjunto, así como, las particularidades de cada una de las tres categorías que definen la tipología de usuarios: estudiantes, profesores y administradores.

3.2 Aspectos técnicos

Desde un punto de vista puramente técnico, *Seshat Process* es una aplicación web alojada en uno de los servidores de la Universidad de León. En su desarrollo confluye el uso de diversos lenguajes de programación entre los que podemos distinguir fundamentalmente tres: PHP, JavaScript y XML.

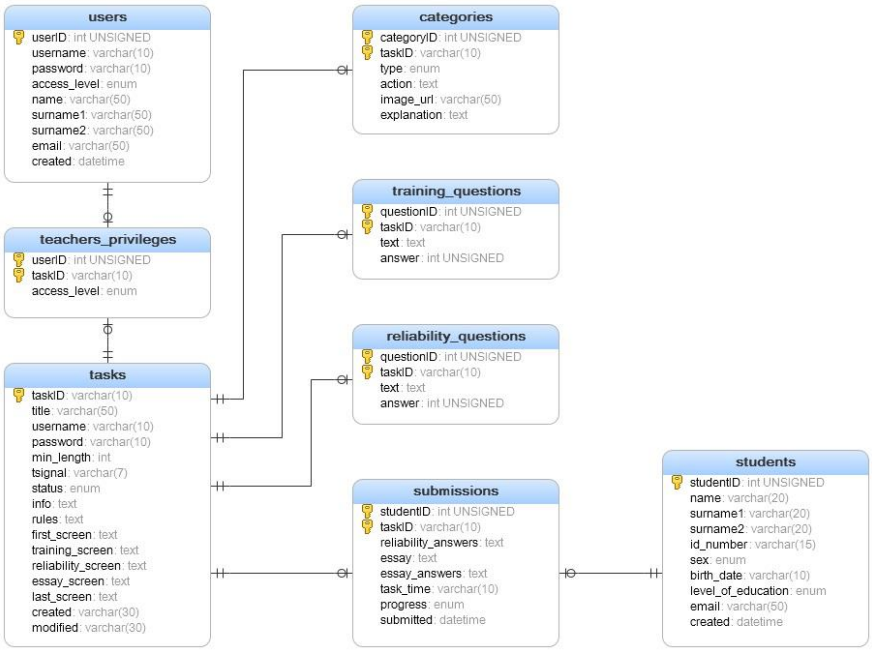


Figura 3.1. Estructura de la base de datos.

Asimismo se ha hecho uso intensivo de tecnología AJAX para mejorar la interactividad y optimizar el rendimiento general del sistema evitando eventuales actualizaciones de toda la web.

La aplicación cuenta también con una base de datos relacional desarrollada en MySQL (ver Figura 3.1) se ha creído conveniente prestar una especial atención, al ser considerada una pieza clave para comprender el funcionamiento de cada uno de los componentes del sistema.

Tal y como puede apreciarse en el diagrama de la Figura 3.1, pueden distinguirse fundamentalmente dos tipos de tablas, las que conciernen a los datos relacionados con las diversas tareas definidas y las que implican a los usuarios del sistema. En relación a éstos últimos, resulta especialmente relevante el hecho de que la tabla *students* se defina al margen de la lista general de usuarios, dado que esta particularidad define el rol de los participantes en el sistema. A diferencia del resto, los usuarios de tipo *student* no poseen un nombre de usuario y contraseña propios, sino que acceden a la plataforma empleando unas credenciales genéricas previamente definidas en función del estudio en el que participen. De este modo, una vez hayan accedido al sistema, se instará a los participantes a introducir sus datos personales; salvo que hayan intervenido en estudios previos, en cuyo caso, podrían saltarse este paso indicando únicamente su número de identificación personal. Esta medida fue adoptada con la intención de facilitar la labor del investigador, que de otra forma se hubiera visto forzado a generar

tantos nombres de usuario y contraseñas como participantes fuese a tener en su estudio.

En cuanto a la parte concerniente a datos se refiere, como puede verse en la Figura 3.1, cada tarea se define en función de una serie de atributos y tres tablas independientes: *categories*, *training_questions* y *reliability_questions*. La relación funcional entre estas tablas se explicará en profundidad en el siguiente apartado; no obstante, desde un punto de vista técnico, cabe destacar la existencia de una limitación referencial entre los campos answer y el campo *categoryID*. Esta relación definirá, en gran medida, el funcionamiento de las fases de entrenamiento de la tarea, dado que restringe las posibles respuestas de las preguntas definidas en las tablas *training_questions* y *reliability_questions* a las categorías establecidas en la tabla *categories*. Esto implica, por tanto, que los participantes tendrán que aprender a usar estas respuestas para identificar sus procesos cognitivos durante la realización del ejercicio escrito.

3.3 Interfaz *student*

La categoría *student* hace referencia de forma general a cualquier individuo que haya tomado parte en un estudio llevado a cabo con *Seshat Process*. Este tipo de usuarios accede al

sistema utilizando unas credenciales genéricas que comparte con el resto de participantes que intervienen en el mismo estudio. Una vez realizado el logged in, su experiencia en el sistema se define en base a una secuencia de diez pantallas que no podrá ser alterada en ningún momento salvo para llevar a cabo las actividades previstas. De esta forma se garantiza que todos los participantes compartan la misma experiencia, desde las fases iniciales del entrenamiento hasta la realización del ejercicio escrito final.

Inicialmente, los participantes acceden a una pantalla en la que se describe brevemente el proceso que van a iniciar y las normas de uso general del sistema. Ésta es una pantalla meramente informativa que no requiere de ningún tipo de intervención por parte del usuario a excepción del momento en el que decida avanzar a la siguiente, en la que se le instará a introducir sus datos personales. Aparte del nombre, apellidos, género, email y número de identificación personal, con fines puramente estadísticos, se le pedirá que indique su fecha de nacimiento, su nivel de estudios y el motivo que le indujo a participar en el estudio. No obstante, a juicio del investigador, esta última pregunta podría o bien omitirse o ser remplazada por otra a su elección. Por otra parte, el usuario podría evitar esta segunda pantalla introduciendo únicamente su número de

identificación personal si, previa participación en otros estudios, ya estuviese registrado en el sistema.

Category	Image	Explanation
Searching or reading information		Searching, checking and reading any information that might be helpful for writing the text (using the Internet, any related document, a book, etc.).
Reflecting on the aim of the text and its target audience		Reflecting on the purpose of the text and who is going to read it. For example, it is not the same to write a text for a teacher who is going to grade it than to write a text for a friend about personal matters.
Thinking about what to tell		Creating new stories or remembering something that has been read, seen, or told by someone to the writer, etc.
Making an outline		Writing down new ideas on a separate sheet that are going to be developed later in the text.
Writing the text		Whenever you are writing the text.
Reading the text		Reading the written text again either by reading every word, sentence or paragraph or by reading the entire text before submitting it.
Correcting the text		Correcting the existing mistakes.
Changing information		Deleting repeated words, adding information, changing words or reorganizing information.
Doing something unrelated to the task		Any action that is not related to the task at hand. For example, talking with someone, thinking about later appointments, looking out of the window at people walking down the street, etc.

Figura 3.2. Ejemplo de lista de categorías.

Una vez superadas las dos primeras pantallas, el participante accede a la introducción del estudio. En esta pantalla se describe someramente la naturaleza del mismo y se plantea

una situación, generalmente hipotética, en la que se justifica la necesidad de escribir un texto argumentativo. De igual modo, se mencionará por primera vez al reconocimiento de procesos cognitivos y se presentará la lista de categorías que los identifican, tal y como se muestra en la Figura 3.2.

Tras la introducción, se aborda la fase de entrenamiento, en la que el participante se enfrentará a otras dos pantallas. En la primera se explica el procedimiento a seguir. Continuando con la situación hipotética planteada, el participante deberá interpretar las acciones que un personaje ficticio está llevando a cabo durante la realización de un ejercicio escrito. En la segunda pantalla, el participante deberá, por tanto, realizar una serie de ejercicios en los que tendrá que identificar los procesos cognitivos a los que aluden dichas acciones haciendo uso de las categorías definidas en la introducción (Figura 3.3). Como parte esencial de este entrenamiento, el participante podrá comprobar instantáneamente si sus respuestas son correctas y aprender de sus errores en caso de no serlo.

Finalizado el entrenamiento, las dos pantallas siguientes constituyen un test de fiabilidad que pone a prueba las habilidades adquiridas por el estudiante durante la fase de entrenamiento. Esta prueba también consta de una batería de preguntas, pero a diferencia de la fase de entrenamiento, en esta

ocasión no se indicará si la respuesta ha sido correcta o no a posteriori. De esta forma se pretende tanto evaluar la eficacia del entrenamiento propuesto, como obtener una medida orientativa de la fiabilidad de las respuestas del estudiante durante la realización del ejercicio real al que se enfrentará a continuación.

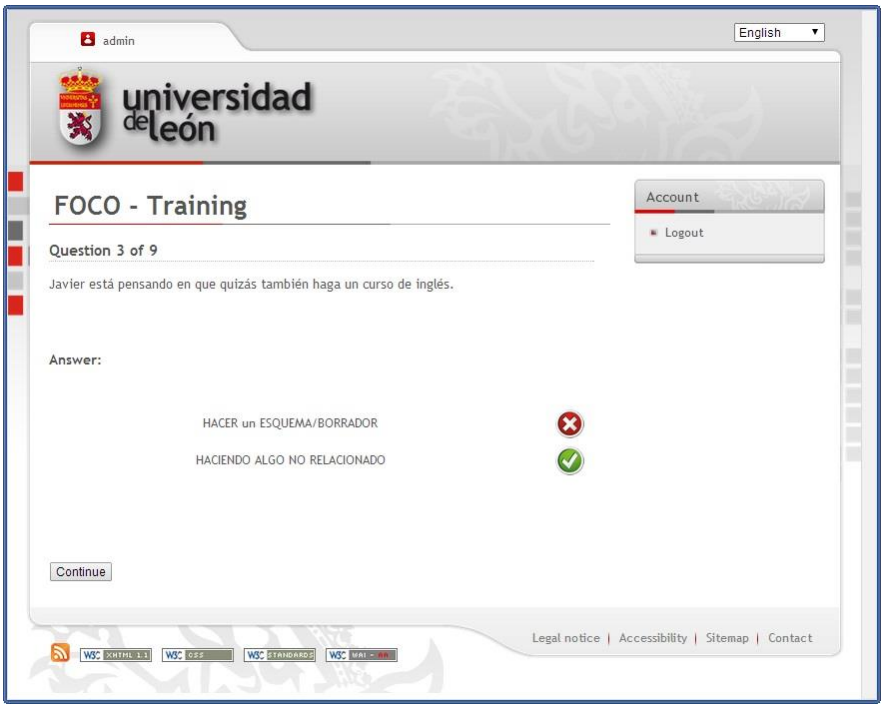


Figura 3.3. Entrenamiento no supervisado.

Al igual que en las fases de entrenamiento y fiabilidad el ejercicio escrito viene precedido de una breve explicación en la que se describirán los pormenores de la tarea a realizar. Se indicará el tema, la duración y extensiones mínima y máxima en caso de haberse especificado. Asimismo, se detallará el

procedimiento a seguir para identificar los procesos cognitivos, especificando la naturaleza de la señal de interrupción programada y los intervalos de ejecución previstos. A modo de ejemplo, en la Figura 3.4 puede observarse el momento en que un participante se enfrenta a la identificación de un proceso cognitivo.

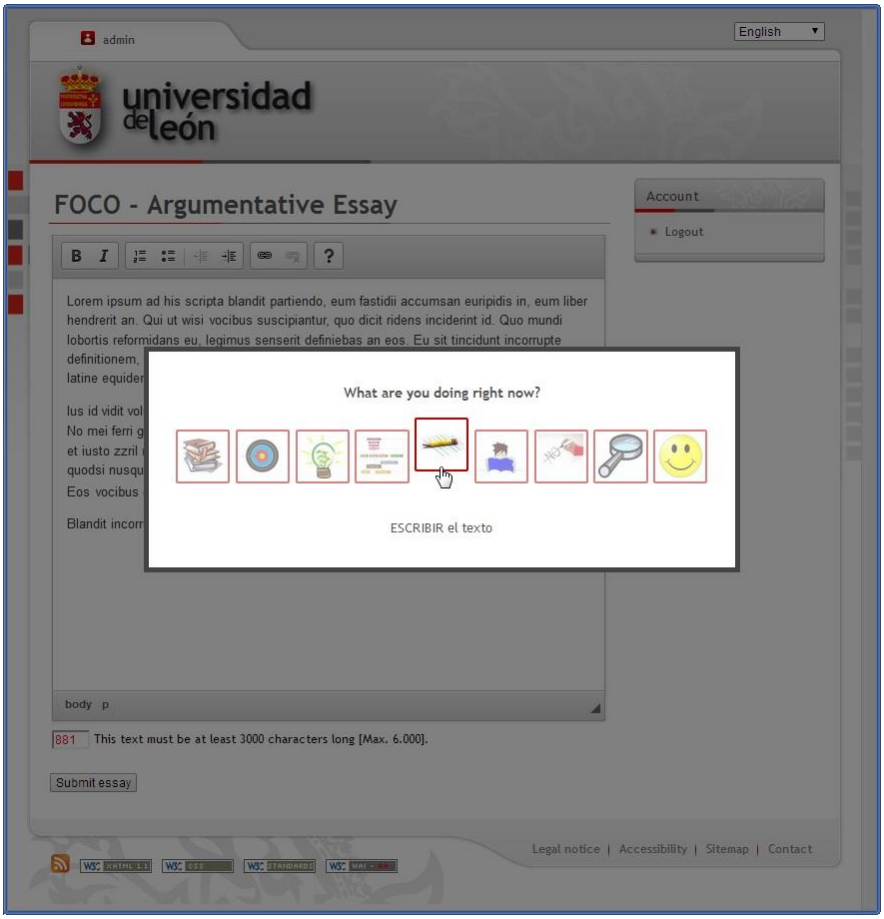


Figura 3.4. Ejemplo de identificación de un proceso cognitivo durante la realización el ejercicio escrito.

Finalmente, satisfechos los requisitos mínimos que hubiesen sido establecidos para la realización de la tarea en cuestión, se mostrará un cuadro de diálogo modal solicitando la confirmación del usuario para proceder a guardar los datos de su ejercicio en el sistema. La décima y última pantalla, a la que el participante accederá a continuación, está reservada para los agradecimientos o cualquier otra consideración que el investigador estime oportuna.

3.4 Interfaz *teacher*

La categoría *teacher* define el rol de los investigadores en *Seshat Process*. Estos usuarios, a diferencia de los participantes, gozarán de un perfil personal y podrán acceder al sistema utilizando un nombre de usuario y contraseña únicos e intransferibles. Una vez realizado el logged in, los investigadores acceden al menú principal de la aplicación que cuenta con cuatro opciones. El submenú cuenta, desde el que podrán modificar sus datos personales y tres listados de carácter general. El listado de los estudios a los que tienen acceso, indicando el nivel de autorización en cada caso, el de los participantes que han tomado parte en ellos y, finalmente, el del resto de investigadores cuyo respectivo nivel de autorización también les permite acceder a

dichos estudios. Aunque la categoría *teacher* define un rol genérico, a tenor de su grado de autorización en un determinado estudio, pueden distinguirse tres perfiles diferentes de investigador.

Observer. Puede inspeccionar todos los elementos de la tarea, pero no tiene opciones de modificación y no puede realizar análisis en tiempo real, descargar resultados, ni acceder a las fichas de los participantes. No obstante, tiene capacidad para exportar los datos de las diversas pantallas a otros estudios en los que tenga un nivel de autorización *task manager* o superior.

Task manager. Puede inspeccionar y modificar todos y cada uno de los elementos de la tarea y acceder a las fichas online de los participantes, pero, al igual que el *observer*, no puede modificarlas, realizar análisis en tiempo real ni descargar los resultados del estudio. Asimismo el *task manager* también tiene capacidad para exportar los datos de las diversas pantallas a otros estudios en los que tenga un nivel de autorización *task manager* o superior.

Full manager. Tiene acceso total al estudio con opciones de modificación tanto con respecto a la tarea como a sus participantes y puede realizar análisis en tiempo real y descargarse los resultados del estudio. Adicionalmente, los investigadores con categoría *full manager* tienen la capacidad de

modificar el nivel de autorización de otros usuarios con acceso al estudio, salvo el de aquellos con su misma categoría.

Figura 3.5. Elementos del menú opciones.

Independientemente de su nivel de autorización, en el momento en el que el investigador acceda a cualquier estudio presente de su lista, se desplegará el menú opciones. A través de este menú se podrán gestionar todas y cada una de las pantallas del estudio, acceder a la lista de sus participantes, realizar análisis en tiempo real de los datos recabados y descargar los

resultados obtenidos o exportarlos en forma de matriz de datos. No obstante, como cabría esperar, algunas de estas funciones estarán restringidas en función de la categoría del usuario. A continuación se describe detalladamente cada uno de los elementos de este menú (ver Figura 3.5), indicando, en caso de ser necesario, el nivel de autorización requerido.

Información general. Una vez creado el estudio, permite redefinir aspectos básicos del mismo, como el título, la descripción general interna, destinada únicamente a los investigadores, las credenciales genéricas del participante y el intervalo de la señal de interrupción.

Normas de uso. Permite definir unas directrices básicas de utilización del sistema e incluir una descripción general del proceso si fuese necesario. No obstante, a diferencia de la descripción interna incluida en el apartado anterior, ésta está destinada a los participantes.

Introducción. Permite añadir una breve presentación en la que se describe la naturaleza del estudio y se plantea un escenario hipotético en el que se justifica la necesidad de escribir un texto argumentativo. Asimismo deberá hacerse alusión al reconocimiento de procesos cognitivos, dado que cuando el participante acceda a esta pantalla se le mostrará por primera vez la lista de categorías planteada para identificarlos.

Categorías. Permite definir la lista de categorías que los participantes emplearán para explicitar el contenido de su memoria operativa durante la realización del ejercicio escrito. Cada categoría constará de cuatro elementos: tipo, imagen, acción y explicación. Como su propio nombre indica, el tipo determina la tipología general del proceso cognitivo asociado a la categoría, pudiendo elegirse entre cuatro: planificación, escritura, revisión y otros. La imagen permite definir un icono representativo del proceso que facilite su rápida identificación. La acción a su vez define el nombre del proceso y, finalmente, la explicación constituye una breve descripción de la naturaleza del mismo.

Entrenamiento. Esta sección consta de dos partes, texto y preguntas. La primera actúa a modo de sección introductoria, permitiendo presentar la fase de entrenamiento al participante. La segunda, por su parte, permite definir la batería de preguntas que constituirán el proceso de entrenamiento en sí. Las respuestas a estas preguntas siempre se definirán en función de las categorías existentes.

Fiabilidad. Similar a la anterior, esta sección permite definir un test de fiabilidad del mismo modo que se hizo con el entrenamiento, constando igualmente de un texto introductorio y una batería de preguntas. Desde el punto de vista del

investigador, no hay diferencias significativas entre estas dos secciones no así en el caso del participante, que sólo podrá acceder a las respuestas predefinidas en la fase de entrenamiento.

Explicación del ejercicio. Permite definir la introducción del ejercicio escrito final, planteando el tema a desarrollar y poniendo en conocimiento del usuario las restricciones establecidas en cuanto al tiempo y la extensión se refiere.

Agradecimientos. Permite definir el contenido de la pantalla a la que accederá el participante una vez finalizado el ejercicio de composición escrita y registrados sus resultados. Los agradecimientos constituyen la última pantalla que verá el participante antes de abandonar definitivamente el sistema. Esta es una sección libre que queda a disposición del investigador para usar a su conveniencia.

Importar datos. Permite importar datos de cualquier otro curso al que se tenga acceso. Sólo se podrán importar secciones completa y existen ciertas restricciones a tener en cuenta. Importar preguntas de entrenamiento y/o fiabilidad implicará la importación automática de las categorías a las que hacen referencia, no así en el caso contrario, pudiendo optarse por importar o no las preguntas asociadas a las categorías de una determinada tarea. Asimismo, por motivos de seguridad, una vez hayan comenzado a registrarse aportaciones de los participantes,

las secciones categorías, preguntas de entrenamiento y preguntas de fiabilidad, no podrán volver a ser modificadas, a no ser que dichas aportaciones sean eliminadas manualmente.

Participantes. Muestra el listado de participantes que han tomado parte en el estudio, indicando su grado de fiabilidad y el tiempo que han tardado en realizar el ejercicio escrito. Cada uno de los nombres presentes en la lista permitirá a su vez acceder a la ficha del participante al que hace referencia. En esta ficha además de sus datos personales, se mostrarán sus resultados y el texto en sí. Los usuarios con categoría *task manager* pueden acceder a la lista general de participantes, pero esta sección sólo estará totalmente disponible para usuarios con categoría *full manager* y administradores del sistema.

Resultados. Permite llevar a cabo análisis en tiempo real de los datos obtenidos durante los test de fiabilidad y la realización de los ejercicios escritos, exportar las matrices de datos resultantes en formato *Microsoft Excel* y descargar los resultados generales del estudio. Dada su complejidad y que la capacidad de realizar análisis constituye una parte esencial de *Seshat Process*, esta sección será comentada de forma independiente en el apartado 3.6. El acceso a resultados sólo estará disponible para usuarios con categoría *full manager* o superior.

3.5 Interfaz *administrator*

Al igual que ocurre con los investigadores, el usuario con categoría administrator accede al sistema empleando un nombre y contraseña únicos que no comparte con ningún otro usuario. El administrador posee todas las atribuciones de un investigador con nivel de autorización *full manager* en todos los estudios del sistema y podría llegar a desempeñar su rol llegado el caso. No obstante, la categoría de administrador comporta funciones que exceden los privilegios de los investigadores, dado que es el único capaz de gestionar las cuentas de otros usuarios, pudiendo modificarlas o borrarlas a conveniencia. Asimismo, sólo los administradores del sistema pueden crear nuevos usuarios e iniciar nuevos estudios, pudiendo también eliminar total o parcialmente el contenido de los ya existentes.

3.6 Capacidades de análisis y gestión de resultados

A continuación se examina la capacidad de análisis de *Seshat Process* a través de cinco secciones en las que se describen separadamente los distintos tipos de análisis que pueden llevarse a cabo así como las opciones de descarga de resultados.

3.6.1 Medición del uso de procesos cognitivos

El sistema permite calcular el porcentaje de uso de cada proceso cognitivo atendiendo a los tipos genéricos en los que se subdividen las categorías establecidas: planificación, escritura, revisión y otros. Estos tipos serán constantes independientemente del número y naturaleza de las categorías de un determinado estudio y no podrán cambiarse.

Tal y como puede apreciarse en la Figura 3.6, las cifras finales se muestran en un gráfico interactivo que facilitará la comprobación del porcentaje de uso exacto de cada proceso en todo momento, puesto que los datos se actualizarán en tiempo real a medida que se registren nuevos resultados de los participantes. Este tipo de análisis es susceptible de dos restricciones, fiabilidad mínima y tiempo medio de respuesta máximo. La fiabilidad mínima utiliza los resultados del test de fiabilidad de cada participante para establecer un límite mínimo que permita descartar automáticamente a todos aquellos que no lo superen. Asimismo, el tiempo medio de respuesta máximo establece un límite máximo que permite descartar, a su vez, a todos aquellos participantes cuyo tiempo de reacción medio por *beep*[1] sea superior al mismo.

[1] Seshat Process no utiliza señales de interrupción auditivas; no obstante, el término <<beep>> ha seguido usándose en este contexto por motivos de claridad.

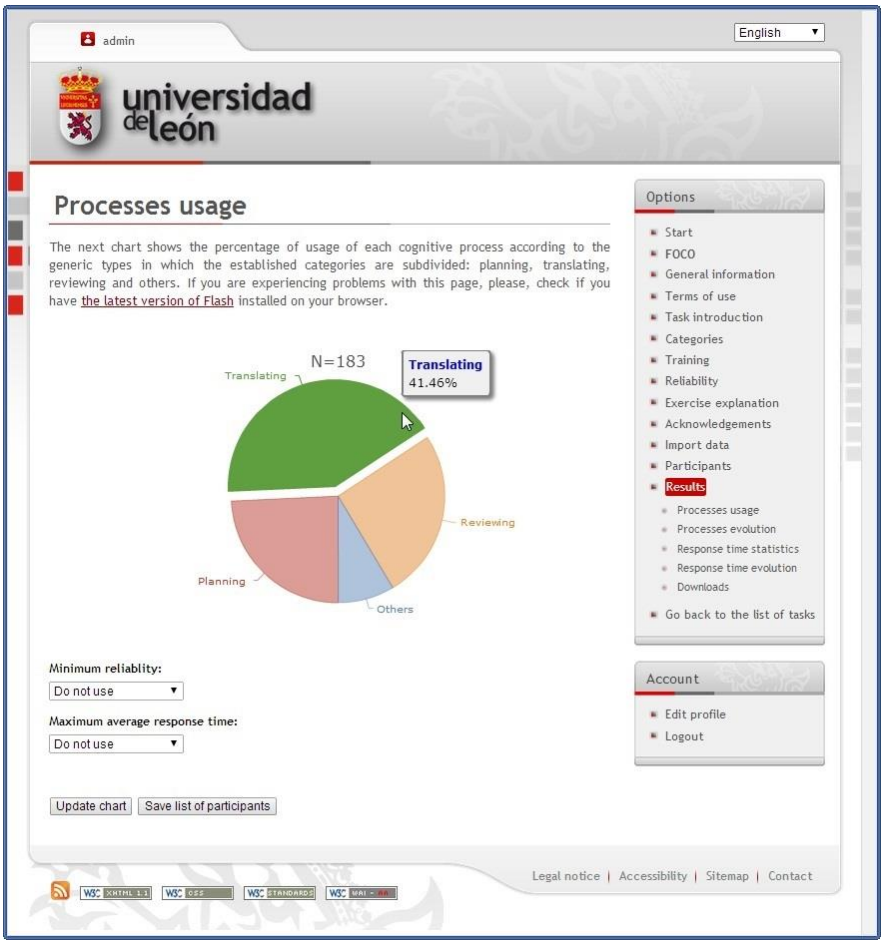


Figura 3.6. Uso de procesos cognitivos.

3.6.2 Análisis evolutivo de procesos cognitivos

El sistema permite analizar la evolución general de los procesos cognitivos en relación al desarrollo de la tarea propuesta. Dado que los tiempos de realización del ejercicio

varían entre participantes, para poder llevar a cabo los cálculos es necesario establecer previamente un tiempo de referencia.

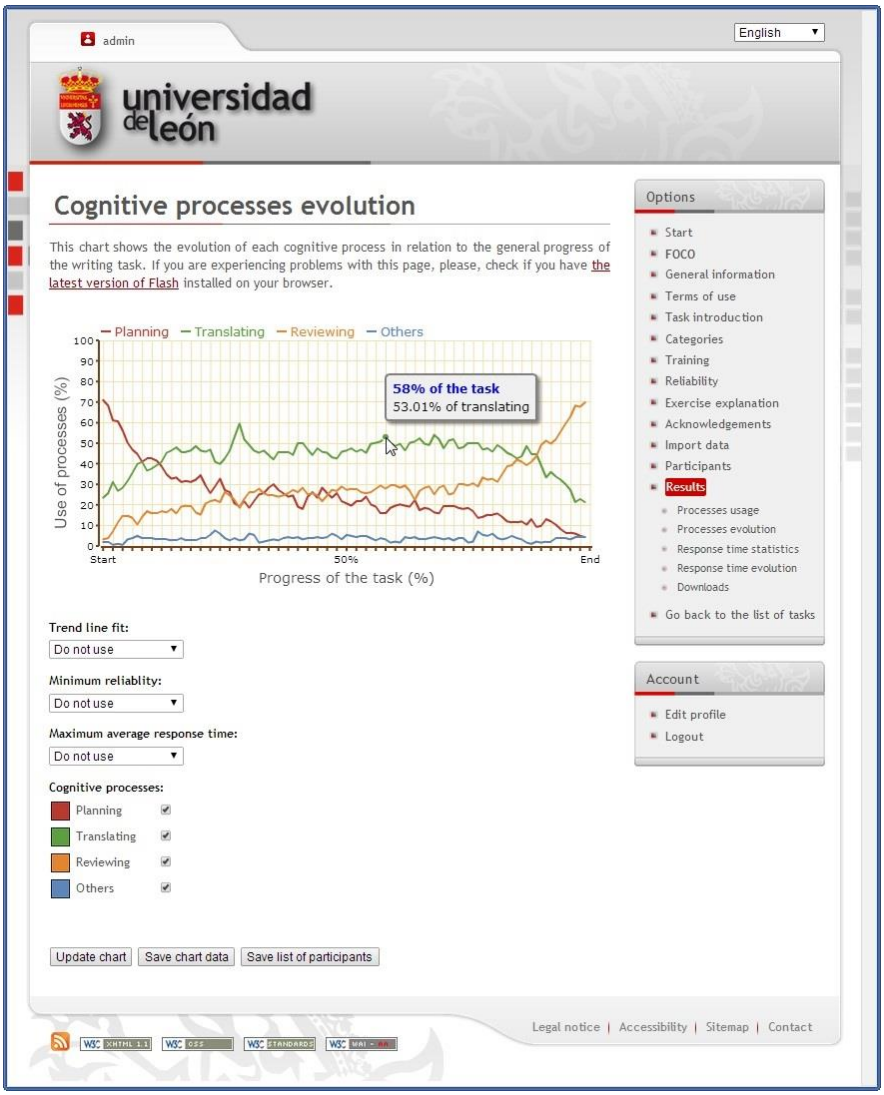


Figura 3.7. Evolución de procesos cognitivos.

De esta forma, los resultados de cada participante se extrapolan utilizando el porcentaje de completitud general de la tarea en relación a este tiempo de referencia para hallar el tipo de proceso que registraron en ese mismo momento en su ejercicio individual. Como puede observarse en la Figura 3.7, al igual que en el caso anterior, el gráfico que muestra la evolución de los procesos cognitivos también es dinámico e interactivo, actualizándose en tiempo real en función de las respuestas de los participantes y permitiendo, en todo momento, comprobar el porcentaje de uso exacto de cada proceso en relación al porcentaje general de completitud de la tarea. Existe asimismo la posibilidad de aplicar las citadas restricciones de fiabilidad mínima y tiempo medio de respuesta máximo; no obstante, en este caso, se incluye además la posibilidad de ajustar las curvas resultantes a sus respectivas líneas de tendencia tal y como se muestra en la Figura 3.8. Este filtro que consta de cuatro niveles: bajo, medio, alto y muy alto, utiliza un algoritmo de cálculo de medias móviles.

Adicionalmente, esta sección permite exportar los resultados en formato *Microsoft Excel*. De esta forma podrán exportarse tanto los datos del gráfico como la lista de respuestas de los participantes que se han tenido en cuenta en el momento de su creación una vez aplicadas las restricciones deseadas.

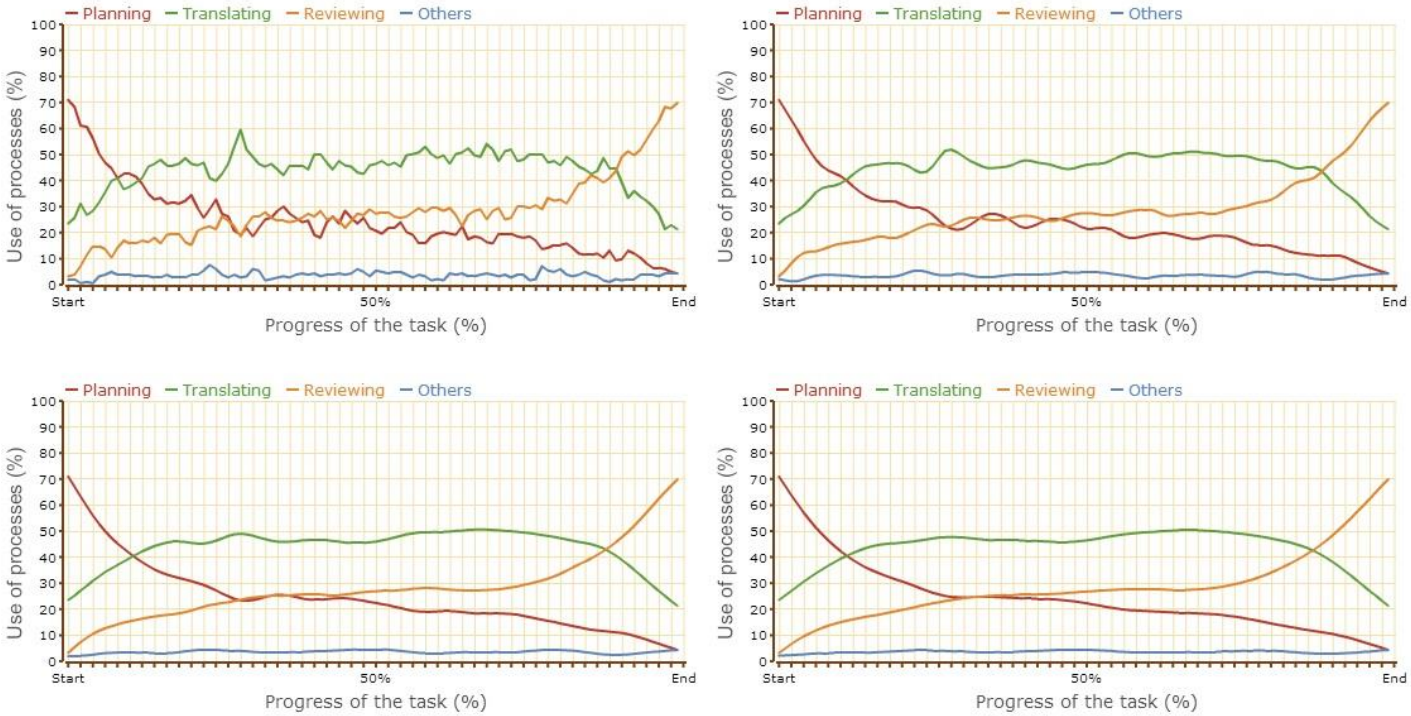


Figura 3.8. Filtros de ajuste a líneas de tendencia.

3.6.3 Estadísticos descriptivos del tiempo de respuesta

A través de esta sección se pueden realizar cálculos estadísticos rápidos en relación a los tiempos de respuesta pudiendo establecerse restricciones de fiabilidad mínima, tiempo de respuesta medio máximo y tiempo de respuesta máximo por *beep*. Al tratarse de un entorno de trabajo no supervisado, esta última restricción se hace necesaria para evitar casos anómalos en los que un participante supere por un margen inusualmente elevado los tiempos de respuesta esperados, alterando, de esta forma, la distribución general de la muestra y los estadísticos descriptivos resultantes.

Los resultados se presentarán siempre en relación a los cuatro tipos generales de procesos cognitivos establecidos: planificación, escritura, revisión y otros, e incluyen el cálculo de la media aritmética por proceso, la desviación típica, el error típico y el establecimiento de los intervalos de confianza.

3.6.4 Análisis evolutivo del tiempo de respuesta

El sistema permite analizar la evolución del tiempo de respuesta medio de los participantes en relación al porcentaje completado de la tarea. Al igual que ocurriese con el análisis de la evolución de los procesos cognitivos, para poder realizar los

cálculos necesarios, resulta indispensable establecer un tiempo de referencia fijo a priori al que poder extrapolar los resultados de todos los participantes. Asimismo, tal y como puede observarse en la Figura 3.9, los datos se presentan en un gráfico interactivo que permitirá conocer en todo momento el tiempo de respuesta medio exacto con respecto a un determinado porcentaje de completitud de la tarea. De igual forma que en el resto de los análisis llevados a cabo por *Seshat Process*, los datos empleados en la creación de este gráfico se actualizarán en tiempo real en función de las respuestas de los usuarios, pudiendo ajustarse las curvas resultantes a sus respectivas líneas de tendencia utilizando un algoritmo de cálculo de medias móviles. Existe también la posibilidad de establecer restricciones de fiabilidad mínima, tiempo de respuesta medio máximo y tiempo de respuesta máximo por *beep*, así como de exportar la matriz de datos generada en formato *Microsoft Excel*.

La principal peculiaridad de esta sección radica en la posibilidad de realizar análisis comparativos entre varias muestras, pudiendo llegar a comparar hasta un máximo de cuatro al mismo tiempo. Actualmente sólo se contemplan dos tipos de muestras, general e individual. La muestra general incluye a todos los participantes que hayan intervenido en el estudio y la individual, como su propio nombre indica, a uno solo. Por supuesto, tanto los datos como el número de participantes de las

muestras seleccionadas variarán en función de las restricciones establecidas, pudiendo llegar a no contener ningún participante y considerarse irrelevantes durante el análisis.

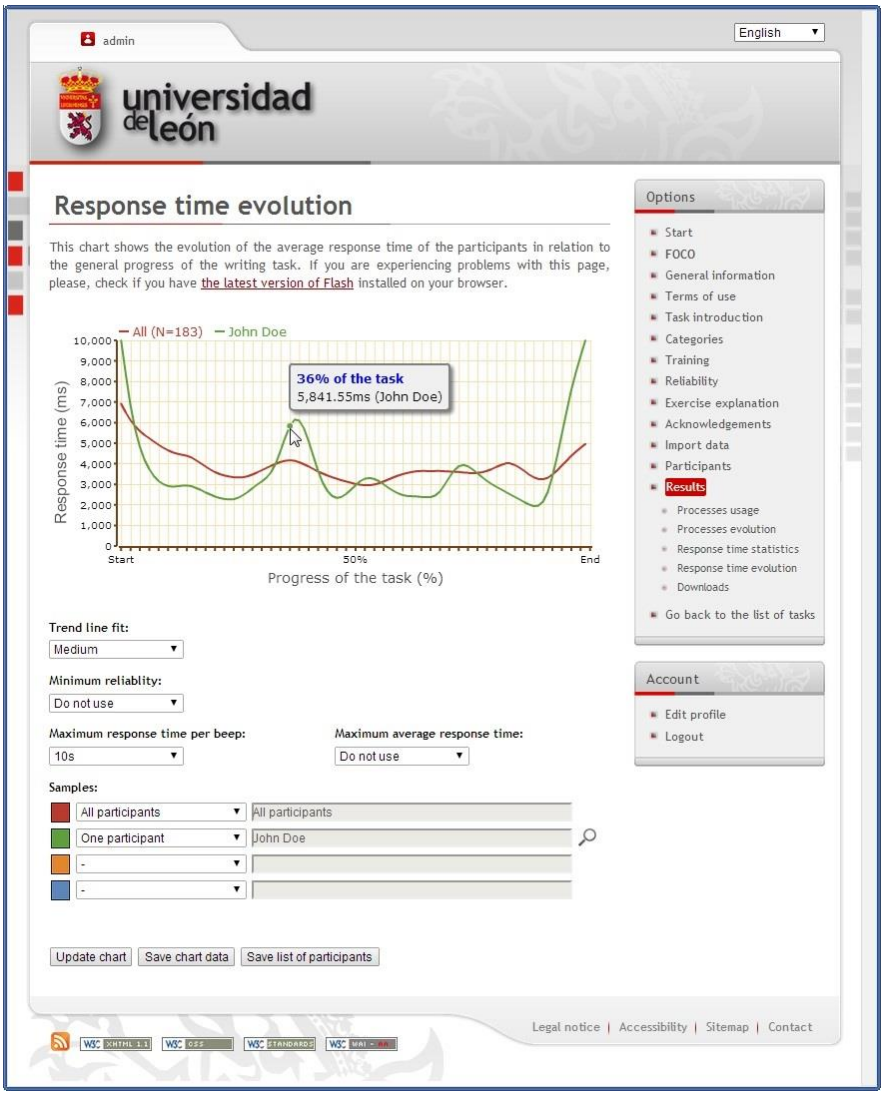


Figura 3.9. Evolución de los tiempos de respuesta.

3.6.5 Descarga de resultados

Esta sección permite descargar los resultados generales del estudio en dos formatos. Una matriz de textos que incorpora el identificador de cada participante y su respectivo ejercicio escrito y otra de resultados que contiene sus datos personales, test de fiabilidad, resultados de la categorización de procesos y tiempo de respuesta por *beep* en milisegundos.

3.7 Validación a través de un caso de estudio real

En esta sección se presenta un breve resumen de la primera prueba real llevada a cabo con *Seshat Process* en aras de explicitar sus fortalezas y debilidades. La experiencia tuvo lugar en junio de 2012, coincidiendo con el inicio de los cursos de verano de la Universidad de León. Con ánimo de poner a prueba todas las capacidades del sistema, se optó por integrar la aplicación como parte de uno de los cursos online ofertados, titulado Desarrollo de Competencias Ocupacionales. La intención inicial era comprobar si el sistema podía lidiar con las principales limitaciones que habían alentado su desarrollo. En otras palabras, se trataba de probar el éxito de la monitorización de procesos

cognitivos en un entorno integrado que pudiese gestionar la intervención simultánea de un elevado número de participantes de forma no supervisada. A tales efectos, el curso elegido contaba con 366 alumnos matriculados con edades comprendidas entre los 19 y los 58 años (M=26,67; DE=7,51) y distintas formaciones. Esta diversidad supuso a su vez una ventaja adicional, dado que permitió evaluar el grado de competencia digital requerido por la aplicación.

El estudio se planteó como un ejercicio de noventa minutos de duración dividido en tres partes. El entrenamiento y la prueba de fiabilidad, con una duración estimada de treinta minutos cada una, y un ejercicio escrito, limitado a sesenta minutos. Los participantes fueron instruidos en la identificación de nueve categorías básicas (ver Figura 3.2), cuatro para procesos de planificación, una para escritura, tres para revisión y una última categoría para hacer referencia a cualquier proceso no relacionado. Asimismo, el ejercicio propuesto consistía en la redacción de una carta motivacional de entre tres mil y seis mil caracteres en relación a una oferta de trabajo. Se eligió este argumento dado que era lo suficientemente genérico como para no requerir conocimientos previos, pretendiendo evitar así que la actividad cognitiva de los participantes pudiese verse afectada por causas ajenas al experimento.

Pasamos a analizar a continuación los resultados más significativos de este curso a la hora de evaluar la viabilidad de la solución implementada. A juzgar por los datos obtenidos (ver Tabla 3.1), el proceso de monitorización de los procesos cognitivos de los participantes no pareció verse afectado por las particularidades de la solución online, no así la medición del tiempo de respuesta. Tal y como puede observarse en la Tabla 3.1, existen significativas diferencias en los tiempos de respuesta registrados en función de las restricciones establecidas. Mientras el análisis realizado sobre la muestra completa presenta unos tiempos de respuesta inusualmente elevados y muy por encima de los valores esperados, los tiempos se normalizan una vez restringida la muestra a aquellos participantes con un mayor índice de fiabilidad. Tratando de encontrar las causas de esta anomalía, se determinó que su origen se hallaba en un reducido número de participantes que no habían tenido en consideración las instrucciones proporcionadas, en las que se les instaba a identificar sus procesos cognitivos reaccionando en el menor tiempo posible ante la señal de interrupción. No obstante, tal y como pudo comprobarse a posteriori, algunos participantes aprovecharon el bloqueo temporal de la pantalla para realizar actividades no relacionadas. Desafortunadamente, se entendió esta limitación como un problema inherente a cualquier entorno de trabajo a distancia y por ende al sistema implementado, dado

que resultaría imposible, a la par que contradictorio, llevar a cabo una supervisión directa de los sujetos que intervienen en el estudio. Cabe destacar, no obstante, que suele tratarse de casos aislados que pueden eliminarse fácilmente de la muestra aplicando filtros más específicos y menos agresivos que el de fiabilidad, como por ejemplo, el establecimiento de un tiempo medio de respuesta máximo (ver Tabla 3.1).

Durante los dos meses de duración del curso se concedió a los estudiantes acceso libre al sistema, instándoles a ponerse en contacto con el equipo ante cualquier eventualidad. En general, las impresiones fueron muy positivas y tan sólo se registraron 6 casos de estudiantes con problemas. Dos de ellos atribuidos a una falta de competencia digital y los otros cuatro relacionados con problemas de compatibilidad que fueron resueltos a posteriori. Asimismo, el método de entrenamiento no supervisado arrojo buenos resultados, alcanzándose un valor promedio de fiabilidad de 0,91 entre las dos muestras. Por su parte, los investigadores externos involucrados en estas pruebas mostraron igualmente su satisfacción con la herramienta desarrollada y propusieron algunas mejoras que supusieron la incorporación de ciertas modificaciones técnicas, entre las que destacan la inclusión de métodos de búsqueda instantánea y la de un listado general de los participantes de todos los cursos a los que tenga acceso un mismo usuario.

Tabla 3.1

Mediciones del uso de procesos y tiempo de respuesta por proceso en el curso FOCO

	Sin restricciones			Fiabilidad: 95%			TR medio[a]: 4s		
	N=183			N=109			N=167		
	UP[b]	TR[a] (en ms)		UP[b]	TR[a] (en ms)		UP[b]	TR (en ms)	
	%	M	DS	%	M	DS	%	M	DS
Planificación	24,27	8459	23858	20,17	5191	2231	19,77	3778	1020
Escritura	41,46	14770	171267	56,30	4221	2399	59,88	2194	624
Revisión	25,73	7695	30177	20,59	3747	1730	17,44	3285	962
Otros	8,54	9877	31049	2,94	3240	1150	2,91	2563	495

[a]Tiempo de respuesta.

[b]Uso de proceso.

3.8 Discusión y conclusiones

Seshat Process ha demostrado su viabilidad como herramienta para monitorizar y analizar los procesos cognitivos involucrados en la escritura, permitiendo integrar en un entorno web todos los elementos necesarios para implementar el procedimiento de la triple tarea de Kellogg. No obstante, a pesar de las ventajas que presenta este planteamiento, existen también, como ha quedado demostrado, ciertas limitaciones inherentes al sistema. Entre estas limitaciones cabe destacar, principalmente, la eventual falta de la competencia digital requerida y el uso incorrecto de la herramienta fruto de su funcionamiento autónomo. En este sentido, se barajan actualmente algunas soluciones enfocadas tanto a la mejora del proceso de instrucción de los participantes como al establecimiento de ciertas restricciones que solventen el problema de la falta de supervisión. A modo de ejemplo cabría citar, la posibilidad de incorporar un sistema que permita cumplimentar las diferentes partes del proceso en varias sesiones y la fijación de un tiempo máximo de respuesta durante la realización del ejercicio escrito.

Por otra parte, entre las posibles mejoras a abordar en un futuro cercano se plantea un aumento de la capacidad de análisis del sistema para permitir al investigador crear muestras permanentes que puedan incorporar participantes de cualquier

estudio al que tenga acceso. De esta forma, se permitiría incluir nuevas restricciones actualmente no disponibles a la hora de realizar análisis, como por ejemplo, la división por géneros, el nivel de estudios o el ámbito sociocultural.

Evaluación automática de textos escritos

Antecedentes II[1] 4

4.1 Introducción

Los sistemas de procesamiento automático de la escritura o AES, de las siglas inglesas Automated Essay Scoring, formalmente definidos como aplicaciones informáticas al servicio de la evaluación de la palabra escrita (Shermis & Barrera, 2002; Shermis & Burstein, 2003; Shermis, Raymat, & Barrera, 2003), cumplen un doble objetivo en el ámbito educativo. A saber, procurar a los profesores una forma objetiva de evaluar los ejercicios de composición escrita de los estudiantes y permitirles analizar sus fortalezas y debilidades en función de los resultados obtenidos. Este último punto resulta de especial interés en educación, dado que un análisis pormenorizado de las características del texto permite al profesor

[1] La presente investigación recibió financiación procedente de proyectos competitivos concedidos a Grupos de Investigación de Excelencia de la Junta de Castilla y León cofinanciados por la Unión Europea para el trienio 2009-2011 (GREX259; BOCyL 27 de abril de 2009) y de un proyecto competitivo del MICINN (EDU2010-19250/EDUC, 2010-2013), figurando J. N. García como investigador principal en ambos proyectos. Asimismo, la presente investigación recibió financiación procedente de un contrato competitivo cofinanciado por la Junta de Castilla y León y el Fondo Social Europeo concedido a H. Díez Caso (EDU/1867/2009, 2009).

proporcionar un *feedback* preciso e individualizado a cada uno de sus estudiantes, algo impensable en circunstancias normales, especialmente en grupos muy numerosos. Asimismo, habitualmente suelen contarse también entre las ventajas de los sistemas AES, su fiabilidad, la capacidad para tratar con los problemas habituales del análisis de la escritura y el ahorro de tiempo y consecuente reducción de costes que supone su uso (Shermis & Burstein, 2003).

La clave del funcionamiento de este tipo de herramientas radica en reducir el texto a sus dimensiones básicas para posteriormente valorarlas de forma individualizada en función de una serie de criterios previamente establecidos bien sea por el profesor o por sus desarrolladores. De este modo, tales aspectos como la gramática, la ortografía, la riqueza de vocabulario o la estructura del texto, entre otros, son valorados y ponderados individualmente por el sistema previo entrenamiento con textos corregidos de acuerdo a tales criterios. Si bien es cierto que un ordenador actualmente no puede comprender ciertos aspectos de un texto debido a la complejidad inherente al lenguaje natural, se ha demostrado que los sistemas de análisis automático ofrecen la posibilidad de realizar mediciones más precisas del constructo de escritura (Bennett, 2004).

A continuación se presenta una breve revisión de los fundamentos teóricos del procesamiento del lenguaje natural, el uso de indicadores lingüísticos de productividad y coherencia en el análisis de la calidad textual y las principales técnicas utilizadas en la actualidad por los sistemas de procesamiento automático de la escritura.

4.2 La escritura como producto

De acuerdo con Gabrielatos (2002), considerar el constructo de escritura desde el punto de vista del producto constituye un enfoque tradicionalista en el que se insta a los estudiantes a recrear un modelo textual específico. Este método de enseñanza se centra principalmente a la corrección del texto elaborado, potenciando aspectos puramente mecánicos de la composición escrita como la ortografía, la gramática o las estructuras sintácticas del texto elaborado. Se hace asimismo especial hincapié en la organización y estructura del texto, reflejándose, en el modelo proporcionado, todas aquellas cualidades deseables a este respecto que el estudiante ha de intentar extrapolar a su propia redacción. Dicho de otro modo, el texto de ejemplo constituye un marco de trabajo referencial al que los estudiantes han de saber adaptar el contenido de sus

ejercicios, tratando de organizar sus ideas de acuerdo a las estructuras planteadas. No obstante, a pesar de su gran difusión, este enfoque conlleva ciertas limitaciones y ha sido ampliamente criticado por la comunidad educativa desde la década de los ochenta.

Los principales problemas que se atribuyen a este método de enseñanza radican en el escaso papel que el estudiante juega en la toma de decisiones inherentes a todo ejercicio de composición escrita, así como en la falta de supervisión del proceso de escritura en sí mismo. Éstas y otras consideraciones propiciaron en su momento la aparición de una creciente tendencia en la práctica de la enseñanza de la escritura que instaba a centrarse en el estudio de los procesos previos a la obtención del producto final. A diferencia de los métodos tradicionales, este nuevo enfoque planteaba la enseñanza de la escritura desde una perspectiva cognitivista, prestando especial atención a la actividad mental de los estudiantes durante la realización del ejercicio escrito. Este planteamiento surge a raíz de los trabajos realizados por Hayes y Flower (1980) e inspiraría toda una nueva línea de investigación centrada en la identificación de las diferencias existentes entre los procesos cognitivos de escritores noveles y expertos.

El modelo cognitivo elaborado por Hayes y Flower, que planteaba la escritura en base a tres procesos básicos, planificación, escritura y revisión, originó asimismo una serie de estudios que pusieron de manifiesto la mayor eficacia de los métodos de enseñanza basados en el proceso en relación a los enfoques tradicionales, típicamente centrados en el producto (Hillocks, 1984, 1986; Holdzkom, Reed, Porter, & Rubin, 1982; Parson, 1985; Wesdorp, 1983). No obstante, llegados a este punto, resulta imprescindible evidenciar las diferencias existentes entre los conceptos enseñanza y evaluación, pues, si bien es cierto que la mayoría de los educadores, especialmente los profesores de lengua y literatura, tienden a hacer especial hincapié en el proceso de escritura, la evaluación sigue centrándose habitualmente en el análisis del producto final (Quinlan, Higgins, & Wolff, 2009). La esencia de este planteamiento se apoya en la premisa de que, aun no pudiendo proporcionar información precisa acerca de los procesos cognitivos del escritor, el texto escrito sí constituye una prueba válida de su destreza, aportando información significativa acerca de su habilidad para lidiar con la gramática, la ortografía y la organización estructural y semántica del texto. Desde la perspectiva del producto, por tanto, la calidad del texto suele considerarse directamente proporcional a la habilidad del escritor que lo ha redactado.

Dada la inherente naturaleza multidimensional y subjetiva de la calidad textual, se han desarrollado varias metodologías para abordar su evaluación, pudiendo hablarse, en términos generales, de métodos holísticos y analíticos (Quinlan, Higgins, & Wolff, 2009). Se entiende por métodos holísticos todas aquellas prácticas en las que los evaluadores analizan el texto de forma global, sin llegar a desglosarlo en sus dimensiones primarias, tales como el vocabulario, la gramática o la organización de las ideas, las cuales valoran y ponderan automáticamente en función de su experiencia y criterio personal. Los métodos analíticos, por el contrario, operacionalizan el análisis de la calidad del texto en función de dichas dimensiones o rasgos característicos inherentes al mismo, evaluándolos individualmente y ponderándolos a posteriori de acuerdo a una escala objetiva previamente establecida.

Tal y como puede observarse a la luz de estas diferencias, la evaluación analítica presenta ciertas ventajas respecto a los métodos holísticos que podrían aprovecharse tanto en el campo de la enseñanza, como en el de la lingüística computacional. Así pues, por un lado, mientras que los métodos holísticos dependen exclusivamente de la disposición del evaluador a favorecer el desarrollo del estudiante procurándole información precisa acerca de sus resultados, la evaluación mediante métodos analíticos le proporciona instantáneamente datos específicos

acerca de sus fortalezas y debilidades. Por otra parte, los métodos analíticos proporcionan la hoja de ruta a seguir en cuanto a procesos de automatización se refiere, dado que reducen la dependencia de criterios de evaluación subjetivos y proporcionan, por contra, una serie de indicadores específicos y a la postre objetivables, que facilitarán su implementación desde un punto de vista técnico. Adoptando esta realidad como una premisa en su diseño, la mayoría de los sistemas AES han sido creados para analizar textos escritos a partir de la valoración individualizada de determinados rasgos característicos de los mismos.

4.3 Intermediación de rasgos lingüísticos de productividad y coherencia en la predicción de la calidad textual

Tanto la psicología cognitiva como la psicolingüística contemplan una relación directamente proporcional entre los procesos cognitivos del individuo y el nivel de coherencia y productividad de la composición escrita. En términos generales, un aumento del nivel de coherencia y productividad supondría, por tanto, un incremento del nivel de conocimiento personal y metacognitivo, así como del refinamiento de las estrategias

empleadas en el proceso de escritura. El mayor grado de correlación se produce, no obstante, con respecto a la productividad total (García & Bausela, 2004).

Los componentes de producción hacen referencia a la dimensión cuantitativa del proceso de composición escrita, es decir, a la cantidad de párrafos, unidades textuales y palabras que el escritor produce en un momento determinado (Hernández & Quintero, 2001). De este modo, es posible cuantificar la productividad en base al análisis paramétrico de palabras de contenido, palabras funcionales y determinantes (Justicia, 1996). Estos indicadores se convierten asimismo en un medio para estudiar la productividad en función del nivel de detalle y elaboración de los textos analizados, permitiendo incluso llegar a discriminar entre estudiantes con y sin dificultades de aprendizaje en función de su nivel de desempeño (Graham & Harris, 1996). En cuanto a la calidad del texto se refiere, las evidencias empíricas manifiestan una clara relación entre la productividad y los resultados holísticos obtenidos por los evaluadores (Breland, Bonner, & Kubota, 1995; Collins, 1998; Crossley & McNamara, 2010; DeVillez, 2003; Mellor, 2011). Los estudios demuestran que, en condiciones normales, el efecto de aplicar un modelo de regresión libre considerando la productividad como un rasgo lingüístico independiente ponderan este indicador sobre cualquier otro a la hora de evaluar la calidad

textual (Attali & Burstein, 2006). Este hecho se debe, probablemente, a que una buena redacción requiere de cierta cantidad de información y desarrollo, detalles que, por lo general, suelen influir positivamente en la valoración de los evaluadores (Quinlan, Higgins, & Wolff, 2009). No obstante, la relación entre la productividad y la calidad textual es a menudo ambigua y presenta una progresión logarítmica. De acuerdo con (Shermis, Burstein, Higgins, & Zechner, 2010), el tamaño de la redacción es importante hasta cierto punto, más allá del cual deja de aportar información significativa en cuanto a la estimación de la calidad del texto se refiere.

Por otra parte, dada la amplitud del término, es difícil establecer una definición única y precisa de lo que se entiende por coherencia; no obstante, sí puede afirmarse que todo texto coherente posee una organización formal más o menos definida en función de las relaciones de cohesión contextual establecidas entre los elementos que lo conforman. A este respecto, se entiende por cohesión contextual la congruencia de los diversos componentes del discurso en el ámbito de toda información que pudiese llegar a determinar o condicionar, en mayor o menor medida, cualquier aspecto en relación al texto. En otras palabras, todos y cada uno de los componentes individuales del discurso contribuyen a la creación de las condiciones necesarias para ser considerados como un todo válido en su conjunto (Shen, 2010).

Se puede afirmar, por tanto, que un texto es coherente cuando los elementos que lo componen están debidamente cohesionados entre sí y su contenido se adecúa de forma efectiva tanto a la intención comunicativa del hablante como al contexto en que fuese escrito (García & Bausela, 2004).

En la actualidad son cada vez más frecuentes los estudios que emplean indicadores de coherencia para evaluar la calidad textual en términos de comprensión y legibilidad (Crossley & McNamara, 2010; McNamara, Louwerse, McCarthy, & Graesser, 2010). De este modo, la coherencia se ha convertido asimismo en uno de los factores determinantes a la hora de diferenciar entre escritores expertos y noveles (Bereiter & Scardamalia, 1992). Es lícito afirmar, por tanto, que el estudio del proceso de escritores experimentados permita, a su vez, analizar la coherencia de la composición escrita en base a cómo se generan y organizan sus ideas (Marbán, 2001). Al igual que ocurriese en el caso de la productividad, este tipo de análisis no sólo nos proporciona un medio para valorar la calidad textual, sino una forma de diagnosticar las dificultades de aprendizaje que pudiesen surgir en los distintos ámbitos de los denominados mecanismos de coherencia (García & de Caso, 2002; García & Marbán, 2002).

4.4 Sistemas de corrección automática de la escritura

La mayoría de los sistemas AES analizan los textos empleando métodos regresivos en los que cada aspecto lingüístico a considerar constituye una variable independiente ponderada mediante unos criterios fijados de ante mano en base a la tipología de análisis y a las circunstancias en las que se lleva a cabo. La eliminación del factor humano de la ecuación y por tanto de una posible valoración parcial y subjetiva del texto podría parecer a simple vista una ventaja; no obstante, ha suscitado una gran controversia a lo largo de los años. Algunos autores creen que suprimir la experiencia y la capacidad humana del proceso de evaluación hace imposible que se tengan en consideración aspectos semánticos inherentes a cualquier texto como la originalidad y la creatividad, reduciendo así su análisis a una mera valoración de aspectos puramente superficiales y dejando al margen su verdadero contenido (Yang, Buckendahl, Juszkiewicz, & Bhola, 2002). Asimismo, también se ha argumentado que los estudiantes tienden a reaccionar negativamente ante la implantación de este tipo de sistemas, acusando una menor motivación a la hora de abordar la tarea, así como cierta desconfianza ante la perspectiva de que sus textos sean analizados de forma automática sin contar con la presencia de un evaluador humano experto (Garner, 2013; Larson, 2013;

Markoff, 2013). No obstante, a pesar de sus limitaciones, este tipo de herramientas sigue contando con el respaldo de una gran parte de la comunidad educativa. Parte de su éxito radica en el haberse convertido en herramientas capaces de agilizar la labor de los profesores facilitándoles no sólo un medio adecuado para llevar a cabo el análisis simultáneo de un gran número de textos, sino también la posibilidad de proporcionar a sus estudiantes un *feedback* individualizado en función de los resultados de sus ejercicios.

Por otra parte, cabe destacar el alto índice de acuerdo que los sistemas de corrección automática han llegado a alcanzar con respecto a las puntuaciones otorgadas por evaluadores humanos (Attali & Burstein, 2006; McNamara, Ozuru, Graesser, & Louwerse, 2006; Rudner, Garcia, & Welch, 2006). Esta elevada fiabilidad es debida fundamentalmente al refinamiento de las diversas técnicas empleadas en la implementación de estas herramientas a lo largo del tiempo. A continuación se incluye un breve resumen de las cuatro técnicas más extendidas en el campo de la automatización del procesamiento de la escritura junto con los principales sistemas que las implementan.

4.4.1 Variables proxy

Las variables proxy, uno de los primeros métodos estadísticos empleados en la implementación de sistemas AES, se basan en la posibilidad de aproximar, a través de mediciones indirectas, aquellos valores que se no se tiene la capacidad de medir directamente. Este método, utilizado habitualmente en las ciencias sociales, cimenta su éxito en la correcta elección de las variables implicadas, que han de estar altamente relacionadas en el objeto de interés y ofrecer una elevada fiabilidad para poder garantizar el éxito de los resultados.

En el caso de los sistemas AES, el mejor ejemplo del uso de este tipo de variables lo constituye el Project Essay Grader, también conocido como PEG (Rudner & Gagne, 2001), un sistema pionero originalmente desarrollado por Ellis Page a mediados de la década de los sesenta. La metodología implementada por Page planteaba dos fases, una de entrenamiento y otra de evaluación. Durante la fase de entrenamiento el sistema era expuesto a una serie de textos de ejemplo que servían de referencia para poder ponderar correctamente las variables proxy seleccionas. Este proceso era largo y requería de una gran cantidad de textos para poder entrenar al sistema en un área específica. Generalmente la media

necesaria para obtener resultados válidos se situaba entre doscientos y trescientos textos.

El sistema PEG fue especialmente criticado por su bajo índice de acuerdo y su problemática vulnerabilidad ante posibles fraudes; no obstante, con el paso del tiempo fue mejorado mediante la inclusión de parsers, diccionarios y esquemas de clasificación (Page, 2003; Shermis & Barrera, 2002).

4.4.2 Latent Semantic Analysis

Creado a finales de la década de los ochenta (Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988), el Latent Semantic Analysis, o simplemente LSA, es un modelo estadístico que permite establecer comparaciones entre documentos mediante su representación a través de vectores semánticos. Adoptando como premisa que las palabras con significados afines suelen aparecer en fragmentos de texto similares, el empleo de LSA conlleva la creación de una matriz de datos construida a partir de un conjunto significativo de textos representados en forma de vectores LSA. Cada vector representará a su vez un determinado contexto semántico y estará formado por las frecuencias de aparición de cada una de las palabras contenidas en un fragmento de texto específico. Se

pretende por tanto que los vectores LSA sirvan para valorar el efecto que cada palabra tiene en el significado del pasaje que la contiene, permitiendo asimismo identificar el significado general de ese fragmento de texto en concreto y ulteriormente el de todo el texto empleando la misma técnica con los párrafos que lo componen. Como puede observarse, la idea subyacente es que el significado de cada fragmento de texto depende fundamentalmente de las palabras que lo componen y cambiar una sola palabra podría suponer una alteración significativa del mismo (Landauer, Laham, & Foltz, 2003).

Los sistemas basados en Latent Semantic Analysis se consideran altamente fiables y actualmente son empleados en multitud de campos, entre los que destacan la búsqueda de información, el reconocimiento de patrones y la educación. En cuanto a las aplicaciones más representativas se refiere, destaca el Intelligent Essay Assessor (Landauer, Laham, & Foltz, 2001), una herramienta basada en internet que permite evaluar la calidad de los textos enviados electrónicamente. Esta aplicación se centra especialmente en el análisis del contenido semántico del texto utilizando técnicas basadas en LSA; no obstante, también evalúa aspectos como la gramática, el estilo o la mecánica de la composición (Landauer, Laham, & Foltz, 2001, 2003; Streeter, Psotka, Laham, & MacCuish, 2002).

4.4.3 Procesamiento del lenguaje natural

Considerada una de las áreas más complejas de la inteligencia artificial, el procesamiento del lenguaje natural constituye a su vez la rama de la lingüística computacional que se encarga del estudio de la interacción entre ordenadores y lenguaje humano. Esta área de investigación comprende multitud de campos incluyendo el estudio de corpus, del habla, la traducción automática, la generación del lenguaje natural y la comprensión del lenguaje hablado (Salem, 2000). Desde el punto de vista de los sistemas AES, las técnicas NLP plantean medios sofisticados para analizar el texto conforme a sus múltiples dimensiones, gramática, ortografía, coherencia y estructura, entre otros. No obstante, de acuerdo con Dikli (2006), aunque eficaz, la aplicación de este tipo de técnicas entraña una gran complejidad, dado que comprenden diversas sub-tareas y niveles de procesamiento.

Actualmente existen numerosos ejemplos de aplicaciones que hacen uso de técnicas NLP; no obstante, probablemente los más significativos sean el E-rater (Burstein, Tetreault, & Madnani, 2013), desarrollado por el ETS (Educational Testing Service) para el automatizar la corrección del GMAT AWA (Graduate Management Admissions Test Analytical Writing Assessment), y el IngelliMetric (Schultz, 2013), desarrollado por

Vantage Learning e inicialmente empleado por el College Board en la corrección de pruebas de nivel (Myers, 2003).

4.4.4 Métodos bayesianos

Basados en el teorema de Bayes, los métodos bayesianos plantean un enfoque de aprendizaje inductivo que supone que las hipótesis en liza están gobernadas por distribuciones de probabilidad que permiten optimizar el proceso de toma de decisiones. En otras palabras, se asigna a cada hipótesis una probabilidad a priori que posteriormente será modificada en función de los ejemplos de entrenamiento. Por lo general estos métodos se consideran extremadamente eficientes en la ejecución de tareas de clasificación; no obstante, los métodos bayesianos requieren de un elevado número de ejemplos de entrenamiento para aumentar su fiabilidad, lo que asimismo supone un aumento del coste computacional asociado al proceso de actualización de probabilidades. El ejemplo más conocido de uso de métodos bayesianos en el campo de los sistemas AES es BETSY (Bayesian Essay Test Scoring System), un programa de evaluación y clasificación de textos desarrollado por Lawrence M. Rudner (Rudner & Liang, 2002).

Seshat Product[1] 5

5.1 Introducción

Seshat Product es una aplicación informática concebida para automatizar diversos procesos de evaluación y análisis de textos escritos con el propósito de generar informes exhaustivos basados en un acervo preestablecido de reglas bien definidas (Díez-Caso & García, 2011, 2013). Estas reglas, que constituyen en la base teórico-práctica de la aplicación, están recogidas en un protocolo de corrección creado para establecer unos criterios de evaluación estandarizados con el fin de poder valorar objetivamente diferentes aspectos de la composición escrita en español (García et al., 2014). Desde una perspectiva psicolingüística, el leitmotiv original de esta investigación ha

[1] La presente investigación recibió financiación procedente de proyectos competitivos concedidos a Grupos de Investigación de Excelencia de la Junta de Castilla y León cofinanciados por la Unión Europea para el trienio 2009-2011 (GREX259; BOCyL 27 de abril de 2009) y de un proyecto competitivo del MICINN (EDU2010-19250/EDUC, 2010-2013), figurando J. N. García como investigador principal en ambos proyectos. Asimismo, la presente investigación recibió financiación procedente de un contrato competitivo cofinanciado por la Junta de Castilla y León y el Fondo Social Europeo concedido a H. Díez Caso (EDU/1867/2009, 2009).

sido el estudio de las dificultades de aprendizaje en los ciclos formativos básicos; no obstante, con el paso del tiempo, el refinamiento del protocolo de corrección ha permitido a *Seshat Product* evolucionar hasta adoptar una orientación más generalista. Esto que ha sido posible gracias al respaldo de una amplia y heterogénea base documental compuesta por más de 20.000 textos escritos por un total de 7.232 estudiantes españoles con y sin dificultades de aprendizaje y edades comprendidas entre los 8 y los 16 años.

Todos los textos han sido recogidos por los investigadores del equipo haciendo especial énfasis en la composición escrita como vehículo para la investigación de la escritura atendiendo a factores de productividad, calidad, estructura y coherencia del texto (Arias & García, 2007; de Caso, García, Díez, Robledo, & Álvarez, 2010; Fidalgo & García, 2008; García, 2002; García & Arias-Gundín, 2004; García & de Caso, 2002, 2004, 2006a, 2006b, 2007; García & Fidalgo, 2003, 2004, 2006, 2008a, 2008b, 2008c; García & Marbán, 2003; García, Rodríguez-Pérez, Pacheco, & Díez, 2009; Pacheco, García, & Díez, 2009, 2010a, 2010b; Torrance, Fidalgo, & García, 2007). Previa digitalización, esta ingente cantidad de información ha servido para crear un corpus lingüístico abierto de más de dos millones de palabras orientado al estudio del desarrollo de la escritura en el ámbito escolar.

Seshat Product ha sido diseñado para facilitar la labor de docentes e investigadores permitiéndoles agilizar el proceso de corrección de textos. No obstante, a diferencia de una gran parte de los sistemas AES actuales, *Seshat Product* no elimina completamente el factor humano del proceso, dado que no ofrece una valoración general de la calidad del texto, sino la de una serie de aspectos lingüísticos del mismo que han ser interpretados por el usuario a posteriori. El objetivo de la solución implementada no es, por tanto, eliminar la acción humana, sino proporcionar a los profesionales de la enseñanza y la investigación un medio para crear estrategias instruccionales individualizadas y refrendar sus conclusiones en la experiencia de miles de análisis previos de textos de similares características.

5.2 Protocolo de corrección[1]

Desde los orígenes de la psicología de la escritura se ha desarrollado una creciente preocupación acerca de cómo obtener indicadores objetivos para evaluar la escritura en términos tanto cuantitativos como cualitativos. En el campo de la investigación educativa, suelen considerarse asimismo todos aquellos indicadores relevantes para el estudio de aspecto especialmente

[1] El protocolo de corrección completo, incluyendo tanto indicadores basados en el texto como en el lector, está disponible en el Apéndice A.

ligados el ámbito de la enseñanza, como es el caso del diseño de estrategias instruccionales de mejora de la competencia comunicativa escrita (García et al., 2014). La psicología de la escritura abarca, por tanto, un campo extremadamente amplio que contempla el análisis de la composición escrita tanto desde la perspectiva del proceso como del producto; no obstante, el protocolo de corrección cuya presentación se aborda en esta sección se centra exclusivamente en cuanto atañe al análisis de la escritura como producto. A continuación se muestra un texto analizado de forma manual de acuerdo a este protocolo de corrección (ver Figura 5.1); la leyenda completa de las marcas utilizadas a tal efecto se encuentra disponible en el Apéndice A.

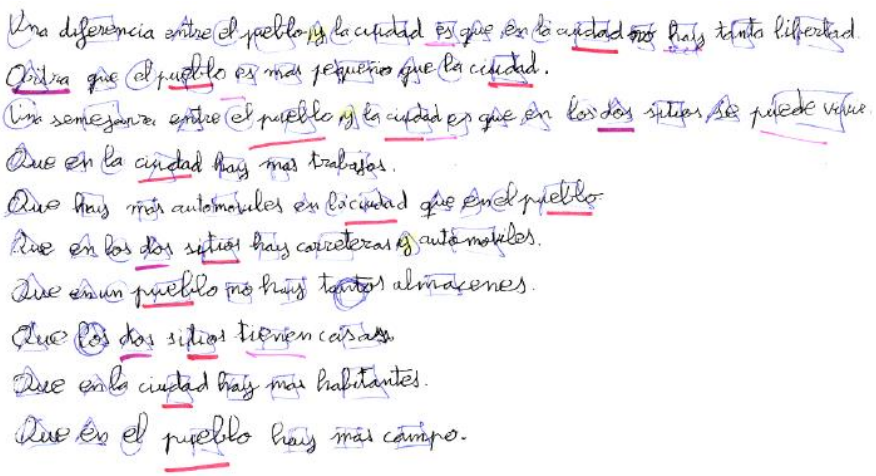

Una diferencia entre el pueblo y la ciudad es que en la ciudad no hay tanta libertad.
Otra que el pueblo es mas pequeño que la ciudad.
Una semejanza entre el pueblo y la ciudad es que en los dos sitios se puede vejar.
Que en la ciudad hay mas trabajos.
Que hay más automóviles en la ciudad que en el pueblo.
Que en los dos sitios hay carreteras y autómoviles.
Que en un pueblo no hay tantos almacenes.
Que los dos sitios tienen casas.
Que en la ciudad hay mas habitantes.
Que en el pueblo hay mas campo.

Figura 5.1. Texto manuscrito analizado de forma manual.

El actual protocolo de corrección es el fruto de más de un década de investigación centrada en la obtención de indicadores objetivos que permitiesen analizar ejercicios de composición escrita en español atendiendo a la maximización del acuerdo entre evaluadores y a la discriminación entre estudiantes con y sin dificultades de aprendizaje en ciclos formativos básicos. Este planteamiento, no obstante, fue evolucionando con el paso del tiempo hasta adoptar una perspectiva más generalista. A día de hoy, con ánimo de ampliar su espectro de aplicación, el protocolo implementado por *Seshat Product* se halla dividido en dos partes, los indicadores basados en el texto y los basados en el lector. Los indicadores basados en el texto plantean un análisis puramente objetivo centrado en la generación y organización de la información. Los indicadores basados en el lector, por su parte, adoptan un modelo subjetivo de análisis en el que la experiencia y criterio del evaluador siguen siendo la clave del proceso. Ambas técnicas de análisis han sido ampliamente estudiadas durante las últimas dos décadas, desde los trabajos iniciales de Spencer y Fitzgerald (1993) hasta la consolidación del actual protocolo de corrección (Fidalgo, Torrance, & García, 2008; Garcia & de Caso, 2004; García & de Caso, 2006; García et al., 2014; García & Fidalgo, 2006; García & Marbán, 2003; García & Pacheco, 2007; Torrance, Fidalgo, & García, 2007).

En lo que a *Seshat Product* respecta, en la actualidad únicamente se implementan los indicadores basados en el texto, dado que aún se trabaja en el refinamiento de los indicadores basados en el lector, indispensables para poder obtener una valoración general de la calidad textual. No obstante, aún en las condiciones actuales, la implementación parcial del protocolo de corrección ha facilitado ya la realización de toda una serie de estudios centrados en la identificación de patrones evolutivos de interés a través del análisis de la escritura. Destacan el estudio del papel predictivo de los profesores en la enseñanza de la composición escrita (García & Pacheco, 2012a, 2012b), los estudios evolutivos en alumnos de 8 a 16 años (Álvarez & García, 2014), la identificación de los cambios operados tras diversas intervenciones instruccionales con padres y profesores (Robledo & García, 2013), la introducción de ayudas en tareas de reescritura (Arias & García, 2012), la comparación de producciones escritas en desarrollo frente a las profesionales en corpus lingüísticos (García-Martín y García, 2012) e incluso la realización de estudios experimentales y de desarrollo (García, Rodríguez-Pérez, Pacheco, & Díez, 2009).

A continuación se presentan los indicadores actualmente implementados por *Seshat Product* en relación a la generación y organización de la información. Junto a cada indicador se adjunta

el código utilizado para su identificación en el programa, así como una breve descripción de su naturaleza.

5.2.1 Generación de información

Las mediciones de productividad se llevan a cabo mediante el recuento de determinadas partículas, expresiones y otros aspectos del texto. Como puede apreciarse en la Tabla 5.1, al margen de los indicadores de carácter general, se diferencia entre palabras de contenido y palabras funcionales; entendido por palabras de contenido aquellas cuya información semántica está generalmente relacionada con propiedades asociadas a entidades físicas, y por palabras funcionales aquellas cuyo contenido semántico es puramente gramatical y por lo tanto abstracto y carente de referentes concretos.

Tabla 5.1

Indicadores de generación de información

General		
ID	Nombre	Descripción
TITL	Título	Indica la existencia o ausencia de título.
CFIN	Comentario final	Indica la existencia o ausencia de comentario final.
NPAR	Número de párrafos	Contabiliza el número de párrafos.
NUNI	Número de unidades textuales	Contabiliza el número de unidades textuales, entendiendo por unidad textual toda oración delimitada por un punto, puntos suspensivos, interrogaciones o exclamaciones.
NVFP	Número de verbos en formas personales y perífrasis verbales	Contabiliza el número de verbos y perífrasis verbales, exceptuando las formas impersonales infinitivo, gerundio y participio.
NDET[a]	Número de determinantes	Contabiliza el número de determinantes incluyendo artículos determinados, demostrativos, numerales, indefinidos e interrogativos.

Tabla 5.1
Indicadores de generación de información (continuación)

Palabras de contenido[a] (NCON)		
ID	Nombre	Descripción
NNOM	Número de sustantivos	Contabiliza el número total de sustantivos considerando las variaciones en género y número como palabras diferentes.
NADJ	Número de adjetivos	Contabiliza el número de adjetivos calificativos.
NADV	Número de adverbios	Contabiliza el número de adverbios y locuciones adverbiales.
NVER	Número de verbos	Contabiliza el número total de verbos incluyendo las formas impersonales.
NINT	Número de interjecciones	Contabiliza el número de interjecciones.

Tabla 5.1
Indicadores de generación de información (continuación)

Palabras funcionales[a] (NFUN)		
ID	Nombre	Descripción
NPRE	Número de preposiciones	Contabiliza el número de preposiciones y locuciones prepositivas.
NCNJ	Número de conjunciones	Contabiliza el número de conjunciones y locuciones conjuntivas.
NPRO	Número de pronombres	Contabiliza el número de pronombres incluyendo personales, relativos, demostrativos, posesivos, numerales, indefinidos e interrogativos.

[a] Seshat Product cuenta con un indicador de productividad total (PTOT) producto de la suma de los determinantes, las palabras de contenido y las palabras funcionales.

5.2.2 Organización de la información

La organización del texto se operacionaliza por medio de indicadores de coherencia. La coherencia textual es un fenómeno cognitivo inherente a la representación mental del texto que a nivel superficial se representa a través de una serie de marcas e indicadores que actúan como guías o instrucciones de procesamiento a la hora de recrear dicha representación (García et al., 2014). Tal y como puede observarse en la Tabla 5.2, aparte de los indicadores generales, se diferencian dos principios constitutivos de un texto coherente, la coherencia referencial y la coherencia relacional. La coherencia referencial alude a la conexión establecida entre diferentes elementos del texto mediante la repetición literal o semántica tanto de objetos como de sujetos. La coherencia relacional, a su vez, hace referencia a la organización del texto en base a la relación establecida entre los diferentes elementos del discurso a través de determinadas expresiones y marcas a nivel textual.

Tabla 5.2

Indicadores de organización de la información

General		
ID	Nombre	Descripción
NSIG	Signos de puntuación	Contabiliza el número de signos de puntuación. Comas, puntos, puntos y comas, puntos seguidos, puntos finales, viñetas, comillas, entre otras.
Coherencia referencial[a] (CREF)		
ID	Nombre	Descripción
NREF	Indicadores referenciales	Contabiliza el uso de cualquier tipo de partícula lingüística que haga referencia a un elemento, bien anterior o bien posterior, en el discurso; exceptuando aquellos indicadores en los que la referencia se hace a partir de la repetición lexical exacta del referente al que aluden.
NLEX	Indicadores lexicales	Contabiliza el número de repeticiones lexicales exactas o coincidencias semánticas entre los elementos del discurso escrito, pero en el recuento no se cuenta la primera vez que aparece en el texto la palabra de contenido que se repite.

Tabla 5.2

Indicadores de organización de la información (continuación)

Coherencia relacional[a] (CREL)		
ID	Nombre	Descripción
NMET	Indicadores metaestructurales	Contabiliza el número de estructuras complejas, marcadoras de contenidos anteriores o posteriores en el texto, que o bien anticipan la información que se va a dar a continuación o bien resumen lo que se ha dicho anteriormente en el texto.
NEST	Indicadores estructuradores[b]	Contabiliza el número de estructuras organizacionales básicas del discurso.
NCNT	Indicadores conectores[b]	Contabiliza el número de estructuras que vinculan semántica y pragmáticamente un elemento del discurso con otro anterior.

Seshat Product

Tabla 5.2
Indicadores de organización de la información (continuación)

Coherencia relacional[a] (CREL)		
ID	Nombre	Descripción
NRFM	Indicadores reformuladores[b]	Contabiliza el número de estructuras que presentan un elemento del discurso mediante una nueva formulación de un elemento anterior.
NARG	Indicadores argumentativos[b]	Contabiliza el número de estructuras que condicionan las posibilidades argumentativas del elemento del discurso en el que se incluyen sin relacionarlo con otro anterior.

[a] Seshat Product cuenta con un indicador de coherencia total (CTOT) producto de la suma de indicadores de coherencia referencial y relacional.
[b] Indicadores discursivos. Unidades lingüísticas invariables que poseen un cometido coincidente en el discurso: guiar las inferencias que se realizan en la comunicación.

5.3 Aspectos técnicos

Seshat Product es una aplicación de escritorio multiplataforma implementada en lenguaje Java como parte del proyecto *Seshat Writing Research Tools*, cuyo objetivo es desarrollar una plataforma que integre todas las aplicaciones necesarias para abordar el estudio de la escritura desde una doble perspectiva proceso-producto. Todas las herramientas contenidas en esta suite comparten un interfaz común a efectos de facilitar su integración con las aplicaciones ofimáticas y los procesadores de textos más extendidos. Por esta razón se ha hecho uso de las librerías *Apache POI* (The Apache Software Foundation, 2014) para garantizar su compatibilidad con los formatos empleados por *Microsoft Office*.

Dada la extensión del proyecto y la previsible complejidad de su futuro mantenimiento, *Seshat Product* fue diseñado atendiendo al paradigma de la programación modular, siendo dividido en cinco módulos interdependientes: gestión de I/O, etiquetador morfológico, administrador de base de datos, analizador de reglas e interfaz gráfico. En la Figura 5.2 puede observarse como cada una de las partes del programa interaccionan entre sí conformando el esquema general del funcionamiento interno del sistema.

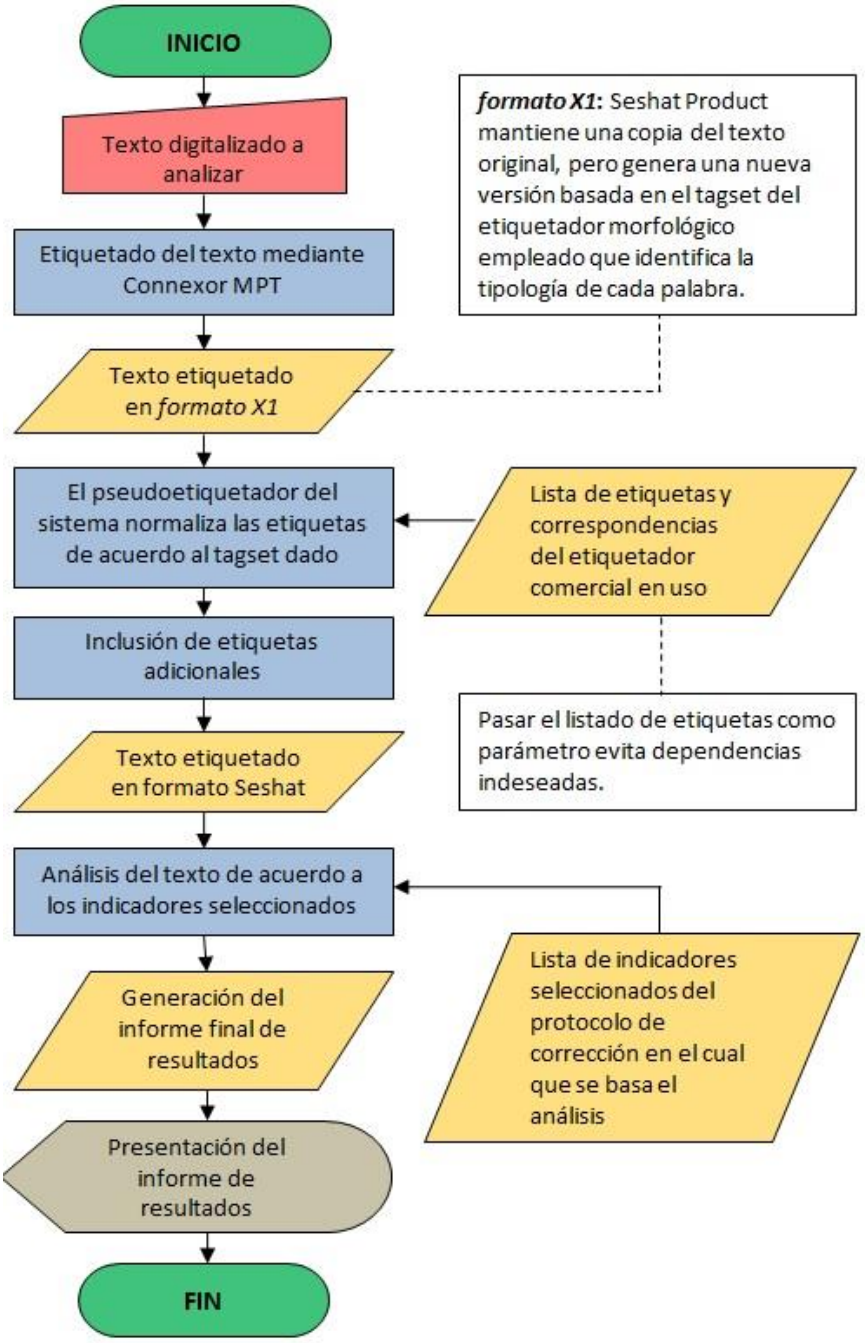


Figura 5.2. Diagrama de flujo del proceso de análisis.

Toda vez que los textos han sido seleccionados y puestos a disposición del sistema, se exponen a una fase inicial de etiquetado morfo-sintáctico. A este respecto, cabe destacar que la creación de *Seshat Product* se abordó con la intención de obtener resultados útiles en el menor tiempo posible, decidiéndose por tanto evitar la creación de un etiquetador morfológico propio e iniciándose la búsqueda de la solución más adecuada a las necesidades del proyecto de entre las disponibles en el mercado. Tras una primera fase de selección, en la que primaron aspectos como la fiabilidad, el precio y la facilidad de integración y mantenimiento, se optó por adquirir el *Connexor Machinese Phrase Tagger*[1], un etiquetador profesional suministrado por la empresa finlandesa *Connexor Oy* (2014). Por encima de cualquier otra consideración, el factor crucial que inclinó la balanza en favor de este etiquetador fue que había sido desarrollado en forma de librería C++, elemento fácilmente integrable en el sistema desarrollado mediante JNI (Java Native Interface) y que podía ser reemplazado sin excesivas complicaciones. Este último aspecto resultó especialmente relevante, dado que se pretendía evitar una excesiva dependencia de productos comerciales ajenos al proyecto, razón que asimismo motivó el desarrollo de un pseudo-etiquetador propio para

[1] La lista completa de las etiquetas o tagset utilizado por el Connexor MPT se encuentra disponible en el Apéndice B.

estandarizar el tagset empleado por el sistema. De este modo, una vez etiquetados mediante el *Connexor MPT*, los textos se someten a una segunda fase de etiquetado en la que se corrigen posibles errores, se modifican las etiquetas asignadas y añaden otras nuevas de acuerdo a los requerimientos del protocolo de corrección que rige el funcionamiento del sistema. Cabe destacar, por ejemplo, el reconocimiento de nombres propios, perífrasis verbales y verbos compuestos como una sola palabra, así como la identificación inequívoca de porcentajes y adjetivos numerales independientemente de cómo hayan sido escritos. Puede afirmarse por tanto que el pseudo-etiquetador cumple una triple función dado que no sólo permite eliminar dependencias comerciales, sino que corrige errores y simplifica futuros procesos de mantenimiento proporcionando un medio rápido y eficaz para poder reemplazar el etiquetador principal si llegado el caso fuese necesario hacerlo. Por otra parte, su desarrollo sienta asimismo las bases para la creación de un etiquetador morfo-sintáctico completo en el futuro.

Una vez finalizada la fase de etiquetado, el texto es puesto a disposición del analizador de reglas, que evalúa el documento resultante de acuerdo a los indicadores establecidos en el protocolo de corrección. Este sistema codifica cada uno de estos indicadores en forma de una serie de reglas lógicas que tratan de emular la corrección manual. Dichas reglas, a su vez, basan su

funcionamiento en la ejecución de uno o varios patrones de búsqueda que explorarán el texto en base a información contenida en la base de datos del programa procedente de análisis previos. Asimismo, cada regla es implementada con la intención de alcanzar el máximo compromiso entre eficiencia y versatilidad, tratando en todo momento de aprovechar las sinergias existentes. De este modo se garantiza, por ejemplo, que el usuario pueda seleccionar un subconjunto determinado de indicadores en cualquier momento y realizar un análisis parcial del texto. El sistema detectará si existe alguna interrelación entre las reglas asociadas a dichos indicadores y, en caso afirmativo, optimizará el proceso de análisis empleando los resultados procedentes de análisis previos si los hubiese o bien forzando la ejecución en background de las reglas adicionales requeridas.

Por otra parte, desde el punto de vista de la eficiencia, se han tenido en cuenta además otros factores, como por ejemplo es el caso de la complejidad algorítmica de los métodos de búsqueda empleados. Como cabría esperar, el tiempo de ejecución aumenta a medida que aumenta la complejidad del algoritmo involucrado en la búsqueda y ésta a su vez, aumenta en relación a los patrones de búsqueda establecidos y a la complejidad y tamaño del propio texto. Este problema, que a priori suscitase la comparación de varios métodos de búsqueda, se resolvió finalmente mediante la implementación de una

versión modificada del algoritmo Boyer Moore Horspool (BMH) adaptado a las necesidades específicas del programa (ver Figura 5.3).

```
public static ArrayList<Integer> patronEnTexto(String texto, String patron) {

        // Contendrá el resultado final una vez aplicado el algoritmo de búsqueda
        ArrayList<Integer> resultado = new ArrayList<Integer>();

        // Se transforman las cadenas de texto (Strings) en vectores de caracteres
        char[] t = texto.toLowerCase().toCharArray();
        char[] p = patron.toLowerCase().toCharArray();

        // Se obtienen la longitud del texto y del patrón
        int n = t.length;
        int m = p.length;

        // Se determina el tamaño del alfabeto
        int nAlfabeto = 256;

        // Se define la función de ocurrencia
        int[] occ = new int[nAlfabeto];

        // Se aplica el algoritmo de búsqueda Boyer-Moore en su variante Horspool
        int i, j;

            for (i=0; i<nAlfabeto; i++) occ[i]=-1;

         for (j=0; j<m-1; j++) {
             i=p[j];
             occ[i]=j;
         }

        i=0;
         while (i<=n-m) {
            j=m-1;
            while (j>=0 && p[j]==t[i+j])j--;
             if (j<0) resultado.add(i);
             i+=m-1;
             i-=occ[t[i]];
        }

        return resultado;
}
```

Figura 5.3. Implementación del algoritmo BMH en Java.

Finalizados los procesos de etiquetado y de análisis, tal y como puede observarse en la Figura 5.2, se genera un informe exhaustivo con los resultados. Este informe constituye una matriz de datos que es presentada por pantalla mostrando tanto el identificador de los textos analizados como los indicadores de acuerdo a los cuales se ha realizado el análisis. *Seshat Product* permite asimismo exportar estos resultados en formato *Microsoft Excel*, garantizando así su compatibilidad con la mayoría de los software de análisis estadístico actuales.

5.4 Interfaz de usuario

En esta sección se describe detalladamente el proceso de análisis desde la perspectiva del usuario, presentando, a tal efecto, el módulo de interfaz gráfico, es decir, la parte de *Seshat Product* destinada a gestionar el modo en que el usuario interactúa, bien pasiva o bien activamente, con el sistema. La máxima prioridad a la hora de crear este módulo fue diseñar e implementar un interfaz intuitivo y de fácil manejo que no requiriese de conocimientos previos para su uso. Por esta razón, durante el proceso de desarrollo se trabajó codo con codo con psicólogos y docentes recogiendo comentarios y sugerencias a

fin de crear un entorno de trabajo que se ajustase de forma específica a sus necesidades y requerimientos. Algunas de estas sugerencias supusieron cambios significativos en la en el diseño inicial, como por ejemplo, la compatibilidad multiplataforma. Cabe mencionar a este respecto, el hecho de que *Seshat Product* haya sido desarrollado utilizando tecnología Java con el fin de permitir su uso en distingos sistemas operativos, así como la adecuación de su apariencia a las características concretas del entorno en que se ejecute. Se advierte, no obstante, que las capturas de pantalla que se van a presentar a lo largo de este apartado han sido tomadas en un entorno *Microsoft Windows*, dado que a día de hoy es el sistema operativo más extendido en el ámbito laboral.

Tal y como puede apreciarse en la Figura 5.4, *Seshat Product* es una aplicación de escritorio con un interfaz clásico basado en un sistema de ventanas que cuenta con una barra de menú, una barra de herramientas con accesos directos a las funciones principales del programa, un área de trabajo y una barra de estado que muestra mensajes cortos de ayuda. A continuación se describe detalladamente un análisis rutinario aludiendo a cada una de las etapas que intervienen en el proceso y mostrando las capturas de pantalla necesarias para ejemplificar la experiencia desde el punto de vista del usuario.

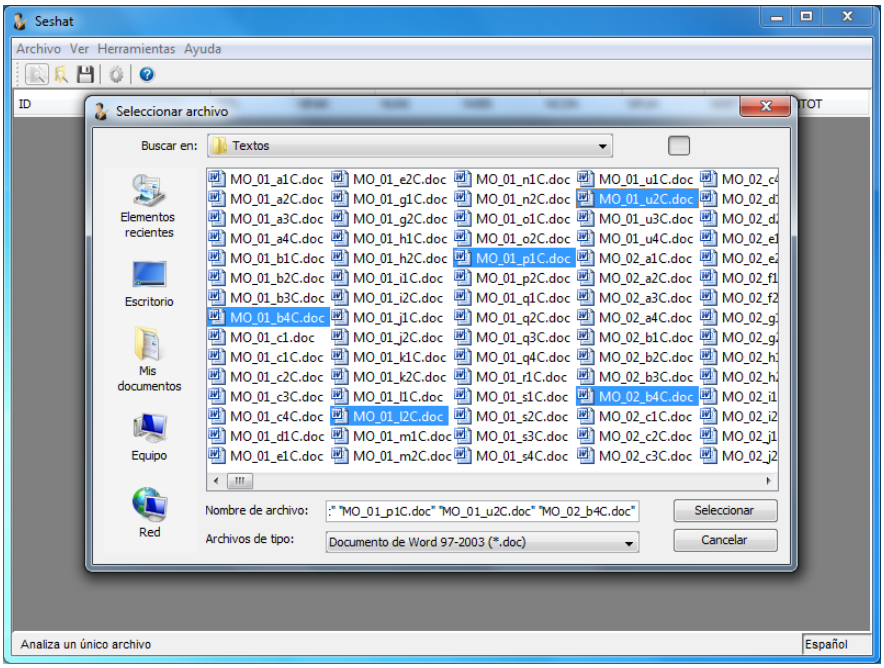


Figura 5.4. Selección de archivos para el análisis.

La Figura 5.4 muestra el primer paso del proceso de análisis, la selección de los archivos que contendrán los textos objeto de estudio. En *Seshat Product* existen dos opciones de selección, la de archivos individuales y la de directorios. En el primer caso, tal y como muestra la Figura 5.4, existe la posibilidad de seleccionar varios archivos al mismo tiempo. De este modo se pretende facilitar tanto la posibilidad de realizar un análisis evolutivo de los textos de un mismo individuo, como la de llevar a cabo estudios comparativos de varios textos específicos si fuese necesario. En el segundo caso, la selección de directorios, el usuario puede seleccionar un directorio

completo con el fin de analizar el contenido de todos sus archivos al mismo tiempo. Este método permite establecer una serie de criterios de búsqueda antes de iniciar el proceso de selección, entre los que se incluyen, la modificación de la lista de formatos de archivo aceptados[1] y la posibilidad de realizar una búsqueda recursiva ilimitada que incluya no sólo los archivos presentes en el directorio seleccionado, sino todos los que se encuentren en cualquier subdirectorio del mismo. Esta última opción ha sido incorporada con la intención de proporcionar al usuario un medio para realizar análisis exhaustivos de grandes cantidades de información organizada en una estructura jerarquizada de directorios, como suele ser el caso, por ejemplo, de la mayoría de los corpus lingüísticos.

Una vez terminado el proceso de selección, el sistema muestra un cuadro de diálogo modal que informa acerca del número total de archivos seleccionados y espera confirmación por parte del usuario para comenzar el análisis, tal y como muestra la Figura 5.5. El acceso a cada uno de los archivos de la lista se produce de forma independiente, buscando y extrayendo los textos contenidos en su interior[2] y poniéndolos a disposición

[1] Seshat Product actualmente es compatible con Apache Open Office Writer, Microsoft Excel, Microsoft Word y cualquier editor de texto plano.
[2] Seshat Product permite procesar libros Excel que contentan uno o más documentos organizados de acuerdo a una estructura diseñada de forma específica a tal efecto.

del sistema para iniciar su análisis de acuerdo a lo descrito en la sección anterior.

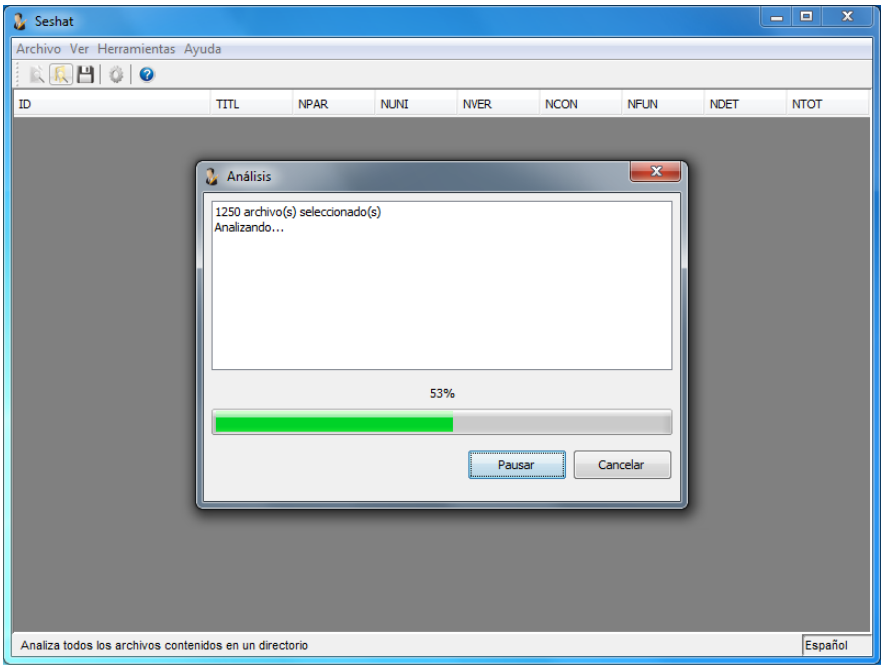
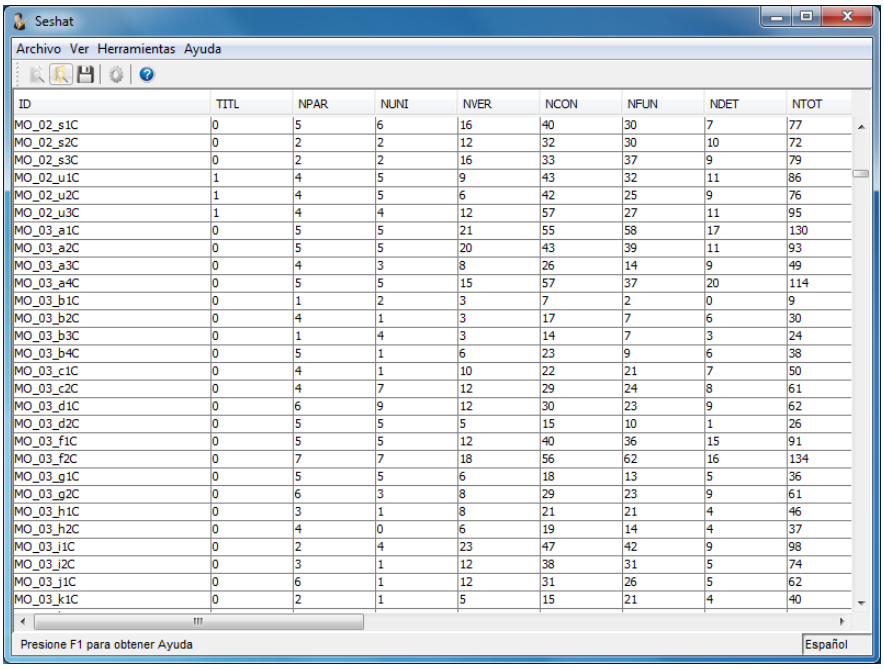


Figura 5.5. Análisis en curso.

Iniciado el análisis, el sistema mantendrá puntualmente informado al usuario acerca del porcentaje de completitud alcanzado, así como de cualquier eventualidad que pudiese tener lugar durante el proceso. En el caso de producirse errores, el sistema alertaría al usuario indicando la naturaleza de los mismos así como los archivos en los que se hubiesen originado. *Seshat Process* distingue entre cuatro tipos de errores, errores de lectura, errores de formato, errores de etiquetador y errores graves. Los

errores de lectura hacen alusión a la incapacidad del sistema para acceder a un determinado archivo. Los de errores de formato, contemplan cualquier problema que pudiera derivarse del procesamiento del contenido de los archivos analizados. Los errores de etiquetador, contemplan a su vez toda eventualidad surgida durante la fase de etiquetado, independientemente de qué etiquetador fuese el causante. Finalmente los errores graves hacen alusión a cualquier problema sin identificar que no encaje en ninguna de las categorías anteriores. No obstante, independientemente de los problemas que puedan producirse durante esta fase, el análisis continuará indefectiblemente hasta haber procesado todos y cada uno de los archivos seleccionados, siendo el usuario el único capaz de pausarlo o detenerlo completamente si fuese necesario.

Una vez completado el proceso de análisis se mostrará por pantalla un informe final indicando el tiempo transcurrido, el número de archivos analizados con respecto al número total de archivos seleccionados inicialmente, así como cualquier incidencia que haya podido tener lugar durante el proceso. Llegados a este punto el usuario podrá o bien aceptar o bien rechazar los resultados obtenidos. En caso de ser aceptados, los resultados se mostrarán por pantalla tal y como puede observarse en la Figura 5.6.



ID	TITL	NPAR	NUNI	NVER	NCON	NFUN	NDET	NTOT
MO_02_s1C	0	5	6	16	40	30	7	77
MO_02_s2C	0	2	2	12	32	30	10	72
MO_02_s3C	0	2	2	16	33	37	9	79
MO_02_u1C	1	4	5	9	43	32	11	86
MO_02_u2C	1	4	5	6	42	25	9	76
MO_02_u3C	1	4	4	12	57	27	11	95
MO_03_a1C	0	5	5	21	55	58	17	130
MO_03_a2C	0	5	5	20	43	39	11	93
MO_03_a3C	0	4	3	8	26	14	9	49
MO_03_a4C	0	5	5	15	57	37	20	114
MO_03_b1C	0	1	2	3	7	2	0	9
MO_03_b2C	0	4	1	3	17	7	6	30
MO_03_b3C	0	1	4	3	14	7	3	24
MO_03_b4C	0	5	1	6	23	9	6	38
MO_03_c1C	0	4	1	10	22	21	7	50
MO_03_c2C	0	4	7	12	29	24	8	61
MO_03_d1C	0	6	9	12	30	23	9	62
MO_03_d2C	0	5	5	5	15	10	1	26
MO_03_f1C	0	5	5	12	40	36	15	91
MO_03_f2C	0	7	7	18	56	62	16	134
MO_03_g1C	0	5	5	6	18	13	5	36
MO_03_g2C	0	6	3	8	29	23	9	61
MO_03_h1C	0	3	1	8	21	21	4	46
MO_03_h2C	0	4	0	6	19	14	4	37
MO_03_j1C	0	2	4	23	47	42	9	98
MO_03_j2C	0	3	1	12	38	31	5	74
MO_03_j1C	0	6	1	12	31	26	5	62
MO_03_k1C	0	2	1	5	15	21	4	40

Figura 5.6. Resultados del análisis.

Cada una de las filas de la tabla de resultados representa un texto analizado identificado por el nombre del archivo que lo contenía o, en su defecto, una referencia interna en caso de tratarse de un archivo que pudiese contener más de un texto. Tal y como puede apreciarse en la parte superior de la Figura 5.6, tanto este identificador como los indicadores analizados aparecen representados en el encabezado de la tabla. De este modo es posible comprobar de forma instantánea el valor resultante del análisis de cualquier indicador en relación a un determinado texto.

Una última consideración digna de mención con respecto al proceso de análisis es su rapidez. Las pruebas realizadas dan fe de la considerable diferencia que existe entre el análisis manual y *Seshat Product*. Mientras que la aplicación manual del protocolo de corrección conlleva una media quince minutos por cada texto de entre ciento cincuenta y doscientas palabras, *Seshat Product* ha puede llegar a analizar mil texto de estas características en aproximadamente diez segundos. Asimismo, las pruebas realizas con corpus lingüísticos han arrojado excelentes resultados, habiéndose llegado a analizar un corpus profesional de farmacología de más de un millón y medio de palabras en escasamente cuatro minutos y medio.

Finalmente, una vez mostrado el informe final de resultados, *Seshat Product*, ofrece al usuario la posibilidad de exportar el resultado del análisis en formato *Microsoft Excel* con el fin de proporcionarle un medio para obtener una matriz de datos fácilmente editable y altamente compatible con la mayoría de los softwares de análisis estadístico utilizados en la actualidad.

5.5 Validación

En el transcurso de esta investigación se llevaron a cabo dos procesos de validación, el del protocolo de corrección,

ideado para proporcionar métodos objetivos de evaluación de ciertos aspectos de la escritura, y el de *Seshat Product*, la herramienta diseñada a efectos de automatizar la aplicación de dicho protocolo. En el primer caso, al margen de poner a prueba la eficacia del protocolo de corrección, sobradamente contrastada en estudios previos (Arias & García, 2012; García-Martín & García, 2012; García & Pacheco, 2012a, 2012b; García, Rodríguez-Pérez, Pacheco, & Díez, 2009; Robledo & García, 2013), la validación se planteó como una forma de comprobar y maximizar, en caso necesario, el acuerdo entre evaluadores humanos antes de proceder a comparar sus resultados con los obtenidos por *Seshat Product*. Con este fin se abordó el análisis inicial de una serie de muestras de entre cuarenta y cincuenta textos, comprobándose a posteriori el índice de acuerdo entre evaluadores haciendo uso del coeficiente Kappa de Cohen (ver Apéndice C) para minimizar el efecto del azar en las mediciones. La disparidad de los resultados obtenidos evidenció la necesidad de abordar la revisión de los métodos de corrección propuestos para algunos de los indicadores del protocolo de corrección de forma individualizada. Para lidiar con este problema se decidió entonces llevar a cabo una serie reuniones entre con el equipo docente a fin de poner sistemáticamente de manifiesto todas y cada una de las ambigüedades funcionales encontradas en el protocolo. Asimismo se fijaron objetivos concretos en cuanto al

índice de acuerdo entre evaluadores humanos se refiere, estableciéndose un límite mínimo de .80 para aquellos indicadores que implicasen consideraciones semánticas y de .90 para los indicadores menos subjetivos. El resultado final de este proceso de refinamiento, que tuvo que repetirse hasta en cinco ocasiones para alcanzar los índices de acuerdo deseados, supuso la ampliación de las listas de búsqueda empleadas en los indicadores de productividad y la incorporación de patrones de búsqueda más específicos para mejorar la identificación de estructuras gramaticales complejas en el caso de los indicadores de coherencia.

Una vez finalizada la revisión de los métodos de corrección, se procedió a realizar una prueba con una muestra de mayor tamaño con el fin de obtener los resultados necesarios para llevar a cabo la validación de *Seshat Product*. A tal efecto y prestando especial atención a la heterogeneidad de la muestra, se seleccionaron quinientos treinta y siete textos escritos por alumnos de entre ocho y dieciséis años. Dichos textos fueron sometidos a un riguroso proceso de revisión previa con el fin de poder corregir cualquier error ortográfico que pudiesen contener antes de ponerse a disposición de los evaluadores. Procedimiento necesario dado que la intención de la prueba era comprobar el índice de acuerdo entre evaluadores en circunstancias ideales.

Tabla 5.3
Índice de acuerdo entre evaluadores en los principales indicadores del protocolo de corrección implementado por Seshat Product (N=537).

	Humano vs. Humano		Humano vs. Seshat Product	
	Kappa de Cohen[a]	Acuerdo	Kappa de Cohen[a]	Acuerdo
TITL[b]	1.000	Perfecto	1.000	Perfecto
CFIN[b]	1.000	Perfecto	1.000	Perfecto
NPAR	0.967	Excelente	0.897	Muy alto
NUNI	0.921	Excelente	0.872	Muy alto
NCON	0.918	Excelente	0.841	Muy alto
NFUN	0.937	Excelente	0.846	Muy alto
NTOT	0.926	Excelente	0.864	Muy alto
NREF	0.844	Muy alto	0.802	Muy alto
NLEX	0.836	Muy alto	0.817	Muy alto
CREF	0.841	Muy alto	0.810	Muy alto
NMET	0.863	Muy alto	0.833	Muy alto
NEST	0.850	Muy alto	0.809	Muy alto
NCNT	0.911	Excelente	0.874	Muy alto
NRFM	0.882	Muy alto	0.830	Muy alto
NARG	0.861	Muy alto	0.773	Alto
CREL	0.887	Muy alto	0.839	Muy alto
NSIG	0.934	Excelente	0.956	Excelente

[a] Todos los resultados son estadísticamente significativos a $p<0.05$.
[b] Fruto de la forma en que se recogieron y del proceso de corrección previo al que fueron sometidos, todos los textos de la muestra analizada incorporaban elementos diferenciadores en título y comentarios finales ajenos a su contenido original. Los valores de los indicadores TITL y CFIN no representan, por tanto, la verdadera capacidad de Seshat Product para identificar estos rasgos en el texto. Pruebas adicionales realizadas con muestras de textos sin editar reportaron unos índices de acuerdo medios entre corrección manual y automática de 0.667 (Alto) y 0.589 (Moderado) para TITL y CFIN respectivamente.

Una vez seleccionada y revisada la muestra, seis evaluadores independientes intervinieron en el proceso de análisis, cinco especialistas que corrigieron los textos más afines a sus respectivas áreas y un evaluador experto especialmente versado en el uso del protocolo de corrección que asumió la corrección de la totalidad de los textos de la muestra. Posteriormente, con ánimo de validar la herramienta desarrollada, los valores medios obtenidos para cada indicador por los evaluadores humanos se compararon con los resultados arrojados por *Seshat Product*.

La Tabla 5.3 muestra los índices de acuerdo obtenidos en ambas pruebas haciendo uso del coeficiente Kappa de Cohen. Tal y como puede observarse, existe una cierta uniformidad en los resultados que son, asimismo, excelentes, habiéndose alcanzado altos índices de acuerdo en todos y cada uno de los principales indicadores del protocolo de corrección.

Con el fin de complementar este proceso de validación, a continuación se presentan dos ejemplos que ilustran el funcionamiento de Seshat Product a través de estudios llevados a cabo por terceros investigadores en los que el sistema fue puesto a prueba en entornos de trabajo reales. Sendas investigaciones, presentadas como estudios de caso (Álvarez & García, 2014; Robledo & García, 2013), analizan tanto el modo en que los

investigadores emplearon la herramienta, como la capacidad de la misma para advertir cambios evolutivos en los textos analizados y el aprovechamiento de las sinergias existentes entre Seshat Product y la técnica de monitorización de procesos cognitivos online implementada por Seshat Process.

5.5.1 Estudio de caso: 2014 Orquestación de procesos en relación con el producto y el papel de las variables psicológicas en la composición escrita

El objetivo de esta investigación, de acuerdo con Álvarez y García (2014), se centró en el estudio de la distribución temporal de los procesos cognitivos involucrados en la escritura mediante una técnica de retrospección directa online, así como en el análisis de las diferencias existentes en el producto textual en los cursos de enseñanza obligatoria. Adoptando pues como premisa que la distribución de las actividades cognitivas durante el proceso de composición escrita no es al azar, sino deliberada (Braaksma, Rijlaarsdam, Van den Bergh, & van Hout-Wolters, 2004), Seshat Product se utilizó conjuntamente con la técnica de monitorización de procesos cognitivos implementada por Seshat Process para analizar la incidencia moduladora de la distribución de dichos procesos a lo largo del tiempo, dando fe de la capacidad de la herramienta para detectar diferentes niveles de

dominio de la escritura en función de las variables psicolingüísticas analizadas.

La muestra empleada constó de un total de 348 alumnos pertenecientes a cursos de entre 4º de Educación Primaria y 3º de Educación Secundaria Obligatoria, con edades comprendidas entre los 9 y los 16 años y un rendimiento y desarrollo normalizados. Esta muestra fue seleccionada, de acuerdo con Álvarez y García (2014), de entre un total de 1170 alumnos provenientes de 13 centros educativos de la provincia de León de acuerdo a tres criterios objetivos: capacidad intelectual, que habría de estar dentro de la normalidad, y adecuados niveles de competencia escritora y curricular.

El procedimiento empleado en la realización de este estudio constó de cuatro partes. Una primera fase centrada en la necesaria instrucción de la totalidad de los alumnos implicados, dado su absoluto desconocimiento de las tareas a desarrollar e instrumentos empleados. Una segunda fase en la que se abordó la selección de la muestra final empleada a tenor de los criterios previamente expuestos. Y, finalmente, una tercera y cuarta fases en las que se llevaron a cabo las actividades previstas y la informatización y posterior análisis de los resultados obtenidos.

En cuanto a la evaluación del producto textual se refiere, el análisis se llevó a cabo atendiendo a dos tipos de medidas

(García et al., 2014), basadas indistintamente en el lector y en el texto y centradas en última instancia en la exploración de indicadores específicos al amparo de las dimensiones de calidad, coherencia y estructura textual.

De acuerdo con Álvarez y García (2014), para garantizar la fiabilidad de las correcciones de los textos se emplearon dos evaluadores independientes, haciéndose uso de Seshat Product, en cuanto a medidas basadas en el texto ser refiere, y comparando los resultados arrojados por la herramienta con los obtenidos por un experto humano versado en la aplicación del protocolo de corrección que ésta implementa. La estimación del índice de acuerdo alcanzado se realizó mediante el empleo del coeficiente Kappa de Cohen: productividad 0.957, coherencia total 0.965, estructura 0.965 y otras medidas 0.768.

Asimismo, tal y como puede observarse en la tabla 5.4, extraída directamente del artículo publicado por Álvarez y García (2014), los resultados de las pruebas realizadas verifican la existencia de diferencias estadísticamente significativas en la totalidad de las medidas consideradas. Teniendo en cuenta por tanto, que la exactitud de dichas pruebas recae en gran medida en los resultados obtenidos por Seshat Product, puede afirmarse que la herramienta proporciona un medio fiable para advertir cambios evolutivos significativos en los textos analizados.

Tabla 5.4

Resultados en los análisis de las pruebas intersujetos, por curso, para el producto textual

	4º EP M (DT)	5º EP M (DT)	6º EP M (DT)	1º ESO M (DT)	2º ESO M (DT)	3º ESO M (DT)	*F*	*p*	η^2
Productividad	53.76 (18.24)	54.21 (23.25)	54.90 (14.40)	51.49 (18.37)	69.29 (26.82)	67.84 (30.19)	6.31	.001	.091
Coherencia total	8.28 (4.71)	8.20 (5.25)	7.67 (4.01)	6.60 (3.50)	10.14 (6.24)	9.16 (5.25)	3.33	.006	.050
Estructura	1.11 (.420)	1.11 (.312)	1.14 (.401)	1.04 (.189)	1.52 (.853)	1.44 (1.34)	4.52	.001	.067
Otras medidas	2.35 (.619)	2.54 (.808)	2.63 (.774)	2.24 (.508)	2.86 (.841)	2.78 (7.64)	5.93	.001	.086

Sólo se incluyen resultados estadísticamente significativos ($p<.05$); η^2 (eta estadística al cuadrado) = estimaciones del tamaño del efecto. La regla de Cohen (1988) señala que .01 -.06 (efecto pequeño); > .06 - .14 (efecto medio); > .14 (efecto grande).

5.5.2 Estudio de caso: 2013 Instrucción estratégica de la composición escrita en la escuela y el hogar

De acuerdo con Robledo y García (2013), esta investigación se centró en el análisis de la eficacia de un programa de intervención en composición escrita para aumentar el rendimiento de alumnos de ciclos básicos de enseñanza y fomentar el desarrollo de otras variables personales; analizando, asimismo, los posibles efectos diferenciales en la efectividad del programa en relación al entorno y las figuras responsables de su aplicación. Por lo tanto, a diferencia del estudio de caso anterior, que presentaba un ejemplo de la viabilidad de Seshat Product para detectar cambios evolutivos en la escritura en grupos numerosos a nivel de ciclo, en esta ocasión el aprovechamiento de la herramienta se concentra en su capacidad para advertir las variaciones existentes antes y después de poner en práctica un programa de intervención específico centrado en un solo individuo.

La muestra seleccionada para este estudio estuvo compuesta por 112 alumnos de Educación Primaria con edades comprendidas entre los 10 y los 13 años y sus familias. Los criterios de selección empleados se centraron en la asistencia regular de alumnos y padres a las sesiones de intervención y en la ausencia de necesidades educativas especiales, dada la necesidad

de conseguir una muestra lo más homogénea posible. Se abordó asimismo la división de los participantes en varios grupos en función del nivel de implicación de padres y profesores en la aplicación del programa de intervención propuesto. La implicación de Seshat Product en la investigación radicó por tanto en el empleo de la herramienta como medio para llevar a cabo la corrección de los productos textuales generados por los estudiantes de los diversos grupos en que fue dividida la muestra antes y después de dicha intervención.

En este caso, puede afirmarse que los resultados obtenidos por Robledo y García (2013) pusieron de manifiesto la capacidad de la herramienta para detectar diferencias estadísticamente significativas en los textos de un mismo individuo antes y después de la aplicación de un programa de intervención en escritura, verificando así mismo la utilidad de Seshat Product a la hora de comprobar la eficacia de este tipo de programas.

5.6 Limitaciones

Desde un punto de vista técnico, las limitaciones más significativas durante el proceso de desarrollo de *Seshat Product* se centran esencialmente en la implementación de las medidas basadas en el lector presentes en el protocolo de corrección

descrito en la sección 5.2. Como cabría esperar, este tipo de indicadores acarrea mayores dificultades dado que, por definición, requieren de la intervención y el juicio de un experto humano, proceso altamente complejo y extremadamente difícil de recrear. Tal y como constató la comparación realizada entre los resultados obtenidos por cada uno de los evaluadores implicados en el proceso de validación de la herramienta, resulta imposible establecer una comparación con respecto a la corrección automática debido a las divergencias existentes entre los propios evaluadores humanos, habiéndose llegando a alcanzar una disparidad superior incluso al 40% en algunos casos. Así pues, a la luz de esta nueva información se ha tomado la determinación de iniciar nuevo un proceso de revisión del protocolo de corrección, vigente en la actualidad, con el fin de aislar las causas de este problema y hallar una solución satisfactoria al mismo. A continuación se describen más detalladamente los problemas específicos encontrados así como las consiguientes dificultades en el proceso de análisis automático.

Medidas difusas. Tras la realización de una serie de pruebas a cargo de cinco evaluadores humanos independientes cuyo principal objetivo era proporcionar el *feedback* necesario para refinar el protocolo de corrección mediante la objetivación de aquellos indicadores cuya automatización resultase más

dificultosa, se comprobó que las modificaciones sugeridas por los expertos a menudo no se correspondían con las puntuaciones finales otorgadas por los mismos. Un análisis en profundidad de los resultados obtenidos reveló la causa del problema. La descripción de gran parte de las medidas basadas en el lector resultaba demasiado vaga y era susceptible de múltiples interpretaciones, algo muy común en este tipo de protocolos de corrección, donde, para alcanzar el índice de acuerdo deseado entre evaluadores, se precisa no sólo que dichos evaluadores estén familiarizados con el protocolo en sí, sino que hayan trabajado juntos en su aplicación.

Criterios de evaluación variables. Las calificaciones otorgadas por los evaluadores humanos tienden a variar incluso en circunstancias similares. Por supuesto, es de esperar que esto suceda, siempre y cuando estas fluctuaciones permanezcan dentro de unos límites aceptables. Desafortunadamente, en el caso del análisis automático este problema cobra una especial importancia, pues dificulta el establecimiento de un criterio general a la hora de valorar los indicadores del protocolo de corrección.

Rangos de calificación desmedidos. A menudo, los protocolos de corrección establecen unos rangos de posibles calificaciones para sus indicadores excesivamente precisos a

tenor de los criterios de evaluación de que se dispone. En otras palabras, resulta muy complicado establecer el valor de un indicador de carácter puramente subjetivo llegando incluso a la precisión de los decimales. Obviamente, un análisis automático acarrea la misma problemática o incluso peor, pues la implementación del sistema resulta extremadamente compleja al carecerse de criterios de evaluación objetivos.

Repetición de indicadores. En teoría todas las medidas subjetivas pueden analizarse de forma independiente, pero en la práctica resulta difícil hacer distinciones y muchos de estos indicadores se solapan entre sí. Esta situación provoca que las calificaciones arrojadas por los evaluadores humanos en ocasiones den la impresión de estar sujetas a la mera intención de establecer una diferencia forzosa entre indicadores, dado que no existe un criterio de corrección objetivo y estandarizado que la establezca de por sí. Igualmente complicado resulta automatizar de forma eficiente este tipo de indicadores si no se establecen previamente unas normas que regulen las sinergias existentes entre los mismos.

El número de indicadores en relación al tamaño del texto. La muestra base empleada en esta investigación, constituida mayoritariamente por textos escritos por alumnos de entre ocho y dieciséis años, pone de manifiesto la necesidad de adaptar el

protocolo de corrección en función de las dimensiones de la composición escrita, pues, en ocasiones, los textos no ofrecen suficiente información para ser analizados en base al total de los indicadores presentes en dicho protocolo. Esta limitación afecta tanto al análisis manual como al automático y justifica nuevamente la necesidad de hacer uso de las sinergias existentes entre indicadores.

5.7 Discusión y conclusiones

El principal objetivo de esta sección ha sido presentar *Seshat Product*, una aplicación informática diseñada para automatizar los procesos de evaluación y análisis de ejercicios de composición escrita en base a una serie de reglas concebidas con el fin de promover el estudio de la escritura desde una perspectiva psicolingüística. A este respecto, los altos índices de acuerdo obtenidos entre la corrección manual y el uso de esta aplicación, así como su capacidad para analizar ingentes cantidades de información en segundos, dan fe de la utilidad de la misma como herramienta al servicio de la actividad educativa tanto para agilizar la labor del docente, como para realizar estudios exhaustivos en español acerca de rasgos lingüísticos específicos de productividad y coherencia. Ha de tenerse en

cuenta, no obstante, que *Seshat Product*, a diferencia de la mayoría de los sistemas AES, no proporciona una valoración global directa de la calidad del texto, limitándose en su lugar a poner a disposición del usuario información acerca de una serie de indicadores lingüísticos concretos que requieren de su interpretación. Esta fórmula de diseño responde tanto al ánimo de otorgar una mayor versatilidad a la aplicación como a la creencia en el papel fundamental que el criterio humano ha de jugar en todo proceso de análisis lingüístico, especialmente tratándose de un campo tan complejo como el estudio del lenguaje natural.

Las características de *Seshat Product* lo han convertido en una herramienta especialmente útil en el ámbito de la investigación académica durante los últimos años (Álvarez & García, 2014a, 2014b; Robledo & García, 2013); no obstante, a día de hoy sólo se contemplan el cincuenta por ciento de los indicadores presentes en el protocolo de corrección empleado en su creación, los llamados indicadores basados en el texto. La incorporación del resto de los indicadores del protocolo de corrección se plantea como una mejora a medio plazo que permitiría ampliar sustancialmente la capacidad de análisis de la herramienta. Esta ampliación supondría asimismo no sólo un aumento cuantitativo con respecto al número de rasgos lingüísticos analizados, sino un considerable incremento del potencial de la herramienta para abordar análisis más complejos,

pudiendo incluso llegarse a poder valorar la calidad global del texto. Actualmente, se explora esta vía intentando automatizar las medidas basadas en la valoración subjetiva del lector haciendo uso de análisis regresivos donde cada rasgo lingüístico a considerar se pondera en función de las características del texto y de la forma en que fuese recogido.

Discusión y conclusiones generales

Discusión y conclusiones generales[1] 6

En esta tesis se han presentado los avances realizados en el campo de la investigación del proceso y producto de escritura a través de la creación de *Seshat Writing Research Tools*, una suite de aplicaciones concebida para permitir abordar tanto la monitorización de la actividad cognitiva implicada en el proceso de escritura como la automatización de los procesos de evaluación y análisis del producto escrito. Tal y como se ha expuesto, el resultado de este empeño ha dado lugar a la creación de dos herramientas singulares, *Seshat Process* y *Seshat Product*, que habiendo demostrado su viabilidad en sus respectivos campos de aplicación, constituyen asimismo la piedra angular de la presente investigación y se han convertido en programas de

[1] La presente investigación recibió financiación procedente de proyectos competitivos concedidos a Grupos de Investigación de Excelencia de la Junta de Castilla y León cofinanciados por la Unión Europea para el trienio 2009-2011 (GREX259; BOCyL 27 de abril de 2009) y de un proyecto competitivo del MICINN (EDU2010-19250/EDUC, 2010-2013), figurando J. N. García como investigador principal en ambos proyectos. Asimismo, la presente investigación recibió financiación procedente de un contrato competitivo cofinanciado por la Junta de Castilla y León y el Fondo Social Europeo concedido a H. Díez Caso (EDU/1867/2009, 2009).

uso diario en el departamento de Psicología, Sociología y Filosofía de la Universidad de León (Álvarez & García, 2014a, 2014b; Díez-Caso & García, 2013, under review a, under review b, Robledo & García, 2013).

Seshat Process ha sido diseñado para superar las limitaciones tradicionalmente impuestas por el medio en los estudios realizados en el ámbito del análisis del proceso de escritura, como son el uso de muestras reducidas que conlleva la presencia física de los participantes en instalaciones habilitadas y la necesaria supervisión del proceso por parte de expertos cualificados. A tal efecto, la solución desarrollada plantea un entorno de trabajo distribuido en el que los participantes acceden a los ejercicios de composición escrita vía web y el sistema registra automáticamente la información de sus procesos cognitivos, permitiendo analizar los datos recabados en tiempo real. Asimismo se ha logrado integrar en una única plataforma web independiente todos elementos necesarios para implementar el procedimiento de la triple tarea de Kellogg, permitiendo al investigador estructurar el experimento a través de una serie de pantallas que abordan tanto la instrucción como la realización de los ejercicios propuestos.

Existen, no obstante, ciertas limitaciones inherentes a la naturaleza de este sistema entre las que cabría destacar los

problemas derivados de la falta de competencia digital de los participantes y el uso incorrecto de la propia plataforma fruto de la falta de supervisión. En este sentido, actualmente se barajan varias soluciones enfocadas tanto a mejorar del proceso de instrucción como a la simplificación del interfaz de *Seshat Process* en aras de reducir el nivel de competencia digital requerido. A modo de ejemplo cabe destacar, la posibilidad de mejorar el sistema secuencial implementado permitiendo llevar a término cada una de las diferentes partes del experimento en varias sesiones de trabajo independientes, así como la fijación de un tiempo máximo de respuesta durante la realización del ejercicio escrito. Por otra parte, en cuanto al investigador se refiere, entre las posibles mejoras a abordar en un futuro cercano se plantea el aumento de la capacidad de análisis del sistema con el fin de permitir la creación de muestras permanentes asociadas el perfil del propio investigador que puedan incorporar participantes de cualquier estudio al que tenga acceso. De esta forma, se pretende proporcionar un medio para la realización de estudios comparativos más complejos atendiendo a muestras de diversas tipologías que pudiesen llegar a tener en cuenta nuevas restricciones actualmente no disponibles a la hora de realizar análisis, como por ejemplo, la división por sexos, el nivel de estudios o el ámbito sociocultural.

Seshat Product, por su parte, constituye un sistema de procesamiento automático de la escritura ideado para automatizar los procesos de evaluación y análisis de textos escritos en base a una serie de reglas concebidas con el fin de promover el estudio del producto de escritura atendiendo a rasgos lingüísticos de productividad, coherencia, estructura y calidad. Al margen de la consabida dificultad que entraña toda incursión en el ámbito del procesamiento del lenguaje natural, tras sucesivos procesos de refinamiento, *Seshat Product* ha logrado alcanzar altos índices de acuerdo con respecto a la corrección manual, llegando a superarse el ochenta por ciento en la práctica totalidad de los indicadores implementados. Este hecho, unido a la capacidad del sistema para analizar ingentes cantidades de información en segundos, ha convertido a esta herramienta en un medio eficaz para realizar estudios exhaustivos en español acerca de rasgos lingüísticos específicos de productividad y coherencia (Álvarez & García, 2014a, 2014b; Díez-Caso & García, under review b, Robledo & García, 2013). No obstante, a diferencia de una gran parte de los sistemas AES actuales, *Seshat Product* no proporciona una valoración global directa de la calidad del texto analizado, sino que se limita a poner a disposición del usuario información precisa acerca de una serie de indicadores lingüísticos concretos que requieren de su interpretación. Esta fórmula de diseño responde tanto al ánimo de otorgar una mayor

versatilidad a la aplicación como a la creencia en el papel fundamental que el criterio humano ha de jugar en todo proceso de análisis lingüístico, especialmente tratándose de un campo tan complejo como el estudio del lenguaje natural.

Actualmente, *Seshat Product*, sólo contempla el cincuenta por ciento de los indicadores del protocolo de corrección que originase su creación, a saber, los llamados indicadores basados en el texto. No obstante, el origen de esta limitación radica en el propio protocolo de corrección, dada la naturaleza subjetiva del resto de los indicadores presentes en el mismo, ideados para su exclusiva aplicación manual. La objetivación y posterior incorporación de estos indicadores, basados en el criterio subjetivo del lector, se ha convertido a día de hoy en una de las principales mejores a considerar a medio plazo, dado que esta ampliación supondría asimismo no sólo un aumento cuantitativo con respecto al número de rasgos lingüísticos analizados, sino un considerable incremento del potencial de la herramienta para abordar análisis más complejos. No obstante, a la vista de los resultados obtenidos hasta la fecha parece indispensable realizar una simplificación del protocolo de corrección que conlleve una drástica reducción del ingente número de indicadores actuales, los cuales, en la mayor parte de los casos o bien están repetidos o bien no aportan datos lo suficientemente significativos como para constituir una entidad en sí mismos, al menos desde el punto de

vista del análisis automático. Una posible solución sería modificar el protocolo de corrección en cuanto a las medidas basadas en el lector se refiere, tomando la calidad final del texto como un indicador general que englobe al resto. Asimismo, los indicadores parciales podrían estar interrelacionados entre sí, no así los indicadores básicos, que mantendrían su independencia en aras de facilitar la automatización. En este sentido, antes de iniciar cualquier proceso de análisis automático, el sistema debería tener en consideración tanto el campo semántico como los límites de productividad fijados, de haberlos, o el promedio establecido para el rango de edad en que se encuadre el texto objeto de análisis. En otras palabras, se plantea el análisis del texto en base a un análisis regresivo en el que la calidad textual se determinaría a partir de rasgos lingüísticos independientes ponderados en función de las características del texto.

Por otra parte, ha de tenerse igualmente en consideración que la evaluación de los textos realizada por evaluadores humanos se basa en la libre interpretación de unas medidas de corrección a menudo difusas y difícilmente objetivables que terminan por convertirse en meras directrices que el evaluador conceptualiza en su conjunto y utiliza en base a criterios subjetivos en el momento de su aplicación. Resulta por tanto poco probable que puedan establecerse criterios de corrección de carácter general susceptibles de ser automatizados en base a los

actuales indicadores basados en el lector, dado que es de suponer que un evaluador humano que corrija dos veces un mismo texto, rara vez lo puntúe de igual modo si existe una diferencia temporal significativa entre ambas correcciones.

El proyecto *Seshat Writing Research Tools*, al margen de consideraciones individuales en relación a las herramientas que lo conforman, constituye en esencia una suite de herramientas desarrolladas con la intención de proporcionar a docentes e investigadores una solución integrada para abordar el estudio de la escritura desde una doble perspectiva proceso-producto. A través de la explotación de las sinergias existentes entre *Seshat Process* y *Seshat Product* se pretende por tanto posibilitar la realización de análisis conjuntos que permitan explicar la relación existente entre proceso y producto en un entorno controlado en todo momento por el investigador, el cual puede acceder a los datos del estudio in situ y analizarlos en tiempo real sin necesidad de recurrir a herramientas adicionales. Asimismo, se pretende proporcionar a los escritores que participen en este tipo de estudios un *feedback* en tiempo real con información precisa que les permita mejorar sus textos acorde con las características de aquellos que han recibido mejores calificaciones, identificando automáticamente los puntos fuertes y débiles de su proceso de escritura en relación con el resto de los participantes del estudio.

Apéndices

Apéndice A

Protocolo de corrección

El protocolo de corrección que se presenta a continuación es fruto de más de una década de trabajo en el campo de la identificación de dificultades de aprendizaje a través de procesos de evaluación y análisis psicolingüístico de textos escritos llevado a cabo por el los investigadores del Grupo de Investigación de Excelencia de la Junta de Castilla y León GR259 y sus colaboradores (García et al., 2014).

1. Indicadores basados en el texto

Los indicadores basados en el texto plantean un análisis de rasgos lingüísticos objetivables inherentes al propio texto. Una vez localizados y recopilados, estos indicadores explicitan la evidencia textual del constructo objeto de análisis en el texto.

Pueden diferenciarse tres aspectos básicos o tipos de indicadores objetivos en el texto, la generación de información, productividad o cantidad de texto, la organización de la

información en función de indicadores de coherencia y la estructura del texto, dependiente del género textual a tratar.

1.1 Generación de información

A continuación se listan los indicadores a considerar en cuanto a la generación de información se refiere. Junto a cada indicador se adjunta el código utilizado para su identificación en *Seshat Product*, así como una breve descripción de su naturaleza y/o ejemplos si procede.

NTIT – Título

Se puntuará con un 1 si hay título y con un 0 si no lo hay o si éste forma parte del texto.

Las etiquetas «T» «/T» marcan un párrafo a efectos de ser reconocido como título por *Seshat Product*.

CFIN – Comentario final

Se puntuará con un 1 si hay comentario final y con un 0 si no lo hay o si éste forma parte del texto.

Las etiquetas «C» «/C» marcan un párrafo a efectos de ser reconocido como título por *Seshat Product.*

NPAR – Número de párrafos

Contabiliza el número de párrafos. Esté o no delimitado por punto y aparte, se considerará como párrafo todo fragmento de texto que presente una separación adecuada y/o sangría. Asimismo, los diálogos serán considerados como un único párrafo.

NUNI – Número de unidades textuales

Se contabiliza como unidad textual, todo fragmento de texto delimitado por punto seguido, punto y aparte, puntos suspensivos, signos de interrogación o signos de exclamación del texto.

En ocasiones puntuales en las que por error se haya omitido el signo de finalización al término de una oración, ésta seguirá considerándose como unidad textual siempre y cuando coincida con la finalización del párrafo.

NVFP – Número de verbos en formas personales

Se considerará cualquier tipo de verbo y cualquier tiempo verbal, ya sea simple o compuesto, exceptuando las formas impersonales infinitivo, gerundio y participio. Asimismo, tanto los tiempos verbales compuestos como las perífrasis verbales se contabilizarán como una única palabra.

Tabla A.1
Lista de perífrasis verbales

Verbo Auxiliar	*Enlace*	*Forma No Personal*
	Perífrasis Verbales Modales	
tener	que	infinitivo
deber	-	infinitivo
haber	de	infinitivo
haber	que	infinitivo
venir	a	infinitivo
deber	de	infinitivo
poder	-	infinitivo
	Perífrasis Verbales Aspectuales	
ir	a	infinitivo
estar	por	infinitivo
estar	a punto de	infinitivo
ponerse	a	infinitivo
romper	a	infinitivo
comenzar	a	infinitivo
echarse	a	infinitivo
empezar	a	infinitivo
soler	-	infinitivo
volver	-	infinitivo
dejar	de	infinitivo
acabar	de	infinitivo
terminar	de	infinitivo
llegar	a	infinitivo
cesar	de	infinitivo
estar	-	gerundio
andar	-	gerundio
venir	-	gerundio
ir	-	gerundio
seguir	-	gerundio
traer	-	participio

Tabla A.1
Lista de perífrasis verbales (continuación)

Verbo Auxiliar	*Enlace*	*Forma No Personal*
	Perífrasis Verbales Aspectuales	
dejar	-	participio
quedar	-	participio
estar	-	participio
tener	-	participio
llevar	-	participio
verse	-	participio
sentirse	-	participio

NCON – Número de palabras de contenido

Dentro de las palabras de contenido se incluyen las siguientes categorías:

a) Sustantivos (NNOM)

Palabra que designa o identifica seres animados o inanimados (ej.: hombre, casa, virtud, Caracas, etc.) (DRAE, 2001).

Las variaciones de género y número serán consideradas como palabras independientes.

b) Adjetivos calificativos (NADJ)

Palabra que acompaña al sustantivo para expresar alguna cualidad de la persona o cosa nombrada (*Ej.: bonito, difícil, grande, santo, etc.*) (DRAE, 2001).

c) Adverbios (NADV)

Palabra invariable cuya función consiste en complementar la significación del verbo, de un adjetivo, de otro adverbio y de ciertas secuencias (DRAE, 2001)

Tabla A.2
Clases de adverbios

Tipo	*Ejemplos*
Afirmación	cierto, sí, también
Cantidad	bastante, casi, más, menos, mucho, muy, poco
Duda	acaso, quizá, tal vez
Lugar	abajo, ahí, allí, alrededor, aquí, arriba, cerca, dentro, fuera, lejos
Modo	aprisa, así, bien, deprisa, despacio, mal
Negación	jamás, no, nunca, tampoco
Tiempo	ahora, ayer, después, entonces, hoy, luego, mañana, pronto, siempre, tarde

d) Verbos (NVER)

Clase de palabras que puede tener variación de persona, número, tiempo, modo y aspecto (*Ej.: cantar, comer, reír, vivir, etc.*) (DRAE, 2001).

Los verbos compuestos se consideran como una única palabra de contenido (Ej.: he cantado, has comido, ha reído, hemos vivido, etc.)

e) Interjecciones (NINT)

Clase de palabras que expresa alguna impresión súbita o un sentimiento profundo, como asombro, sorpresa, dolor, molestia, amor, etc. Sirve también para apelar al interlocutor, o como fórmula de saludo, despedida, conformidad, etc. (*Ej.: eh, hola, ay, bah, adiós, chao, etc.*) (DRAE, 2001).

NFUN – Número de palabras funcionales

Dentro de las palabras funcionales se incluyen las siguientes categorías:

a) Preposiciones y locuciones prepositivas (NPRE)

Clase de palabras invariable que introduce elementos nominales u oraciones subordinadas sustantivas haciéndolos depender de alguna palabra anterior. Varias de ellas coinciden en su forma con prefijos (DRAE, 2001).

Las locuciones prepositivas o combinaciones fijas de varios vocablos haciendo oficio de preposición se contabilizarán como una única palabra funcional; no así las llamadas preposiciones inseparables, carentes de sentido por sí solas.

Tabla A.3
Lista de preposiciones y locuciones prepositivas

Preposiciones
a, ante, bajo, cabe, con, contra, de, desde, en, entre, hacia, hasta, para, por, según, sin, so, sobre, tras, durante, mediante
Locuciones Prepositivas
acerca de, al lado de, alrededor de, antes de, a pesar de, cerca de, con arreglo a, con objeto de, debajo de, delante de, dentro de, después de, detrás de, encima de, en cuanto a, enfrente de, en orden a, en pos de, en virtud de, frente a, fuera de, gracias a, a merced de, junto a, lejos de, por culpa de, respecto a

b) Conjunciones y locuciones conjuntivas (NCNJ)

Palabra invariable que encabeza diversos tipos de oraciones subordinadas o que une vocablos o secuencias sintácticamente equivalentes (DRAE, 2001).

Las locuciones conjuntivas o combinaciones fijas de varios vocablos haciendo oficio de conjunción se contabilizarán como una única palabra funcional.

Tabla A.4
Lista de conjunciones y locuciones conjuntivas

Tipo	*Ejemplos*
Adversativas	en cambio, más, no obstante, pero, sin embargo, sino, sino que

Tabla A.4
Lista de conjunciones y locuciones conjuntivas (continuación)

Tipo	*Ejemplos*
Causales	porque, pues, puesto que, ya que
Comparativas	como, igual de, igual que, tal como, tanto como
Condicionales	si, siempre que, con tal de que
Consecutivas	así que, con que, por tanto, por lo tanto, tanto, tanto que
Copulativas	y, e, ni, que
Disyuntivas	o, u, o bien
Explicativas	es decir, es más, esto es, mejor dicho, o sea
Finales	a fin de que, para que
Temporales	ahora que, antes de que, antes que, cuando, desde que, después de que, después que, en cuanto, hasta que, luego que

c) Pronombres (NPRO)

Clase de palabras que hace las veces del sustantivo (DRAE, 2001). A continuación se detallan los distintos tipos de pronombres existentes en español con el fin de facilitar su identificación por parte del evaluador y su ulterior clasificación en subcategorías si fuese necesario.

c.1) Pronombres personales

Pronombre que designa personas, animales o cosas mediante cualquiera de las tres personas gramaticales. Generalmente, desempeña las mismas funciones del sujeto o del grupo nominal (DRAE, 2001)

Tabla A.5
Lista de pronombres personales

Persona	*Formas*	
	Singular	*Plural*
Primera	yo, mí, me, conmigo	nosotros, nosotras, nos
Segunda	tú, ti, te, contigo	vosotros, vosotras, os, ustedes
Tercera	él, ella, ello, lo, la, le, se, consigo	ellos, ellas, los, las, les, se

c.2) Pronombres relativos

Pronombre que desempeña una función en la oración a la que pertenece, inserta ésta en una unidad superior y tiene un antecedente, expreso o implícito (DRAE, 2001): *que, cual, quien, cuyo, donde.*

c.3) Pronombres demostrativos

Pronombre que señala personas, animales o cosas. Son esencialmente tres, este, ese, aquel y sus variantes (DRAE, 2001).

Tabla A.6
Lista de pronombres demostrativos

	Singular	*Plural*
Distancia corta	este, esta, esto	estos, estas
Distancia media	ese, esa, eso	esos, esas
Distancia larga	aquel, aquella, aquello	aquellos, aquellas

c.4) Pronombres posesivos

Pronombre que denota posesión o pertenencia (DRAE, 2001).

Los pronombres posesivos jamás preceden al sustantivo y concuerdan en género y número con la posesión a la que hacen referencia.

Tabla A.7
Lista de pronombres posesivos

Persona	*Formas*	
	Un poseedor	*Varios poseedores*
Primera	mío, mía, mis, míos, mías	nuestro, nuestra, nuestros, nuestras
Segunda	tú, tuyo, tuya, tus, tuyos, tuyas	vuestro, vuestra, vuestros, vuestras
Tercera	su, suyo, suya, sus, suyos, suyas	su, suyo, suya, sus, suyos, suyas

c.5) Pronombres indefinidos

Pronombre que vagamente alude a personas o cosas o expresa alguna noción que cuantifica (DRAE, 2001).

Tabla A.8
Lista de pronombres indefinidos

Singular	*Plural*
algo, alguien, algún, alguna, alguno, bastante, cualquier, cualquiera, demás, demasiada, demasiado, escasa, escaso, misma, mismo, mucha, mucho, nada, nadie, ningún, ninguna, ninguno, otra, otro, poca, poco, quienquiera, tal, tan, tanta, tanto, toda, todo, un, una, uno	algunas, algunos, bastantes, cualesquiera, demás, demasiadas, demasiados, escasas, escasos, mismas, mismos, muchas, muchos, ningunas, ningunos, otras, otros, pocas, pocos, quienesquiera, tales, tantas, tantos, todas, todos, unas, unos, varias, varios

c.6) Pronombres interrogativos

Pronombre que, sin acompañar al nombre, permite construir enunciados interrogativos u oraciones interrogativas indirectas (DRAE, 2001): *qué, cuál, cuáles, cuánto, cuántos, cuándo, dónde, por qué.*

c.7) Pronombres numerales

Pronombre que indican la cantidad exacta del sustantivo al que sustituyen. Se dividen en pronombres ordinales y cardinales dependiendo de si explicitan o no información referente a la posición del sustantivo reemplazado con respecto a una determinada organización jerárquica (*Ej. cardinales: tenemos veinte; ej. ordinales: es el séptimo*).

NDET – Número de determinantes

Palabras que necesariamente preceden al nombre, cualquiera que sea su función, y sirven para indicar a qué objeto, ya sea persona, animal o cosa, hace referencia.

A continuación se detallan los distintos tipos de pronombres existentes en español con el fin de facilitar su identificación por parte del evaluador y su ulterior clasificación en subcategorías si fuese necesario.

a) Artículos determinados

Clase de palabras de carácter átono que indica si lo designado por el sustantivo o elemento sustantivado es o no consabido (DRAE, 2001): *el, la, los, las.*

b) Determinantes demostrativos

Determinante que acompañan al nombre y lo sitúan en el tiempo o en el espacio con respecto al hablante.

Tabla A.9
Lista determinantes demostrativos

	Singular	*Plural*
Distancia corta	este, esta, esto	estos, estas
Distancia media	ese, esa, eso	esos, esas
Distancia larga	aquel, aquella, aquello	aquellos, aquellas

c) Determinantes posesivos

Determinante que acompaña al sustantivo y denota posesión o pertenencia.

Tabla A.10
Lista de determinantes posesivos

Persona	*Formas*	
	Un poseedor	*Varios poseedores*
Primera	mío, mía, mis, míos, mías	nuestro, nuestra, nuestros, nuestras
Segunda	tú, tuyo, tuya, tus, tuyos, tuyas	vuestro, vuestra, vuestros, vuestras
Tercera	su, suyo, suya, sus, suyos, suyas	su, suyo, suya, sus, suyos, suyas

d) Determinantes indefinidos

Determinante que vagamente alude a personas o cosas o expresa alguna noción que cuantifica.

Tabla A.11
Lista de determinantes indefinidos

Singular	*Plural*
algo, alguien, algún, alguna, alguno, bastante, cualquier, cualquiera, demás, demasiada, demasiado, escasa, escaso, misma, mismo, mucha, mucho, nada, nadie, ningún, ninguna, ninguno, otra, otro, poca, poco, quienquiera, tal, tan, tanta, tanto, toda, todo, un, una, uno	algunas, algunos, bastantes, cualesquiera, demás, demasiadas, demasiados, escasas, escasos, mismas, mismos, muchas, muchos, ningunas, ningunos, otras, otros, pocas, pocos, quienesquiera, tales, tantas, tantos, todas, todos, unas, unos, varias, varios

e) Determinantes interrogativos

Determinante que permite construir enunciados interrogativos u oraciones interrogativas indirectas: qué, cuál, cuáles, cuánto, cuántos, cuándo, dónde, por qué.

f) Determinantes numerales

Determinante que indican la cantidad exacta del sustantivo al que acompaña. Se dividen en determinantes ordinales y cardinales dependiendo de si explicitan o no información referente a la posición del sustantivo al que acompañan con respecto a una determinada organización jerárquica (*Ej. cardinales: tenemos veinte cromos; ej. ordinales: es el séptimo de la fila*).

PTOT – Total de productividad

Contabiliza el número total de palabras de contenido, palabras funcionales y determinantes.

1.2 Organización de la información

La organización del texto se operacionaliza por medio de indicadores de coherencia. La coherencia textual es un fenómeno cognitivo inherente a la representación mental del texto que a nivel superficial se representa a través de una serie de marcas e indicadores que actúan como guías o instrucciones de procesamiento a la hora de recrear dicha representación.

Al margen de los indicadores de carácter general, se diferencian dos principios constitutivos de un texto coherente, la coherencia referencial y la coherencia relacional. La coherencia referencial alude a la conexión establecida entre diferentes elementos del texto mediante la repetición literal o semántica tanto de objetos como de sujetos. La coherencia relacional, a su vez, hace referencia a la organización del texto en base a la relación establecida entre los diferentes elementos del discurso a través de determinadas expresiones y marcas a nivel textual.

A continuación se listan los indicadores a considerar en cuanto a la organización de la información se refiere. Junto a cada indicador se adjunta el código utilizado para su identificación en *Seshat Product*, así como una breve descripción de su naturaleza y/o ejemplos si procede.

NREF – Indicadores referenciales

Se contabiliza el uso de cualquier tipo de partícula lingüística que haga referencia a un elemento, bien anterior o bien posterior, en el discurso; exceptuando aquellos indicadores en los que la referencia se hace a partir de la repetición lexical exacta del referente al que aluden.

Ej.: «El lobo es un mamífero, su pelaje es voluminoso, posee una potente dentadura y su sentido del olfato es muy agudo.»

NLEX – Indicadores lexicales

Se contabiliza el número de repeticiones lexicales exactas o coincidencias semánticas entre los elementos del discurso escrito, pero en el recuento no se cuenta la primera vez que aparece en el texto la palabra de contenido que se repite.

Ej.: «El lobo es un mamífero carnívoro. El zorro también es un mamífero y al igual que el lobo, come carne. Tanto el lobo como el zorro tienen un pelaje voluminoso.»

CREF – Coherencia referencial

Se contabiliza el número total de indicadores de coherencia referencial sumando indicadores referenciales e indicadores lexicales.

NMET – Indicadores metaestructurales

Marcadores que incluyen toda estructura compleja, marcadora de contenidos anteriores o posteriores en el texto, que o bien anticipa la información que se va a dar a continuación o bien resume lo que se ha dicho anteriormente en el texto.

Ej.: a consecuencia de, como consecuencia de, como resultado de, el resultado de, en la imagen, en esta carta, en esta redacción, en este dibujo, en este documento, en este escrito, en este texto, en este trabajo, la consecuencia de, todo comenzó, etc.

NEST – Indicadores discursivos estructuradores

Marcadores especializados en la estructuración básica de la información.

Tabla A.12
Ejemplos de indicadores estructuradores

Tipo	*Ejemplos*
Comentadores	pues, pues bien, así las cosas, bien, etc.
Ordenadores	en primer lugar, por una parte, por otro lado, media hora después, más tarde, entonces, en ese instante, en ese mismo momento, después, luego, hasta que, el primero, el segundo, al día siguiente, al cabo del tiempo, en ese momento, etc.
Digresores	por cierto, a todo esto, a propósito, etc.

NCNT – Indicadores discursivos conectores

Marcadores que vinculan semántica y pragmáticamente un elemento del discurso con otro anterior.

Tabla A.13
Ejemplos de indicadores conectores

Tipo	*Ejemplos*
Aditivos	además, y, incluso, encima, y una vez, aparte, etc.
Consecutivos	por tanto, por consiguiente, por ende, en consecuencia, de ahí, entonces, así que, etc.
Contra-argumentativos	en cambio, por el contrario, sin embargo, no obstante, pero, con todo, etc.

NRFM – Indicadores discursivos reformuladores

Marcadores que presentan un elemento del discurso mediante una nueva formulación de un elemento anterior.

Tabla A.14
Ejemplos de indicadores reformuladores

Tipo	*Ejemplos*
Explicativos	o sea, es decir, esto es, a saber, etc.
Rectificativos	mejor dicho, mejor aún, más bien, etc.
De distanciamiento	de todos modos, en cualquier caso, en todo caso, etc.
Recapitulativos	en definitiva, en conclusión, en suma, y al final, y al fin, al fin y al cabo, etc.

NARG – Indicadores discursivos argumentativos

Marcadores que condicionan las posibilidades argumentativas del elemento del discurso en el que se incluyen sin relacionarlo con otro anterior.

Tabla A.15
Ejemplos de indicadores argumentativos

Tipo	*Ejemplos*
Refuerzo argumentativo	en realidad, en el fondo, de hecho, etc.
Concreción	por ejemplo, en particular, etc.

CREL – Coherencia relacional

Se contabiliza el número total de indicadores de coherencia relacional sumando indicadores metaestructurales y discrusivos (estructuradores, conectores, reformuladores y argumentativos).

CTOT – Coherencia total

Se contabiliza el número total de indicadores de coherencia sumando indicadores de coherencia referencial y relacional.

NSIG – Signos de puntuación

Se contabiliza el número de signos de puntuación: comas, puntos, puntos y comas, puntos seguidos, puntos finales, viñetas, comillas, paréntesis, guiones[1], etc.

Otras medidas de coherencia

Estos aspectos se puntuarán con un 1, si en el texto se refleja o aparece el indicador mencionado, y con un 0 en caso contrario.

a) Empleo de ideas pertinentes en el tema sobre el que escribe. Relación correcta entre significado y significante y contenido ajustado al contenido que se trata.

b) Enlace correcto de las ideas generadas.

[1] Los guiones se contabilizarán de forma independiente, de tal modo que las palabras compuestas por dos términos unidos mediante un guion deben, a su vez, contabilizarse por separado.

c) Párrafos coherentes y bien delimitados.

d) Hilo argumental coherente de acuerdo a una organización secuencial, temporal, lógica, etc.

1.3 Estructuración de la información

La estructura textual permite identificar la estructuración de la información, que es dependiente del género textual de que se trate.

Los géneros textuales descriptivos, expositivos, de comparación-contraste, explicativos, argumentativos, de problema-solución o ensayos de opinión se realizarán según las tres partes principales en las que se estructuran estos tipos de texto: introducción, desarrollo y conclusión:

— Si hay una introducción que presente el tema o el punto de vista del autor se puntuará con un 1, en caso contrario, se puntuará con un 0.

— Si hay un cuerpo de explicación y desarrollo en el texto se puntuará con un 1, en caso contrario, se puntuará con un 0.

— Si hay una conclusión que dé una visión del conjunto, se puntuará con un 1, en caso contrario, se puntuará con un 0.

La estructura de los textos narrativos se puntuará en función de dos grandes apartados: el marco y el episodio. A su vez, ambos se componen de subestructuras. Dentro del marco, están las subestructuras de tiempo, espacio y personajes, y dentro del episodio, las de suceso inicial, respuesta interna de los personajes, ejecución y consecuencias. Cada indicador se puntuará con un 1 cuando esté presente en el texto y con 0 en caso contrario.

a) Tiempo. Implica el uso de expresiones que ubiquen temporalmente las acciones. (*Ej.: un día de primavera, hace mucho tiempo, por la mañana, etc.*) Nótese, no obstante, que las expresiones como «había una vez, un día o érase una vez» no se consideran expresiones que ubiquen temporalmente acciones, ya que no se utilizan con ese fin al tratarse de frases hechas.

b) Espacio. Supone la presencia de expresiones que indican el lugar donde se van a desarrollar o se están desarrollando las acciones. (*Ej.: en el mar, en un bosque, en casa de su amiga, etc.*).

c) Personajes. Se considera que la narración tiene personajes, cuando se nombra a un sujeto que realiza alguna acción, independientemente de que sea o no el personaje principal.

d) Suceso inicial. Implica la existencia de alguna expresión que indique la causa por la que se generan las ejecuciones

posteriores. (*Ej.: su madre le dijo que fuera a casa de sus abuelos*).

e) Respuesta interna. Se considera que respuesta interna la interpretación de cualquier estado interior, bien ánimo, deseo o queja, de alguno de los personajes. (*Ej.: el filósofo pensó detenidamente en la trascendencia de sus palabras, la niña asustada, trató de encontrar el camino de vuelta, etc.*).

f) Ejecución. Se considera ejecución a toda aquella acción del protagonista u otros personajes que implique una nueva situación o cambios en los hechos que no se incluyan en los anteriores aspectos. Se excluyen, no obstante, las frases hechas. (*Ej.: comenzó a andar por el camino, asó el pez y se lo comió, etc.*).

g) Consecuencia. Implica un final que ponga de manifiesto una nueva situación derivada de acciones anteriores. (Ej.: y no volvieron a tener miedo a aquel animal, el ladrón no volvió a molestarles, etc.).

1.4 Consideraciones adicionales

a) Ampersand. El signo «&» se considerará equivalente a la conjunción «y».

b) Expresiones de género múltiple. En este tipo de expresiones sólo se aceptará el símbolo «/» como elemento separador,

atribuyéndosele la misma categoría gramatical que a un guion. Asimismo, las palabras que formen parte de la expresión se contabilizarán por separado.

Ej.: amigos/as.

c) Números. Serán considerados una única palabra y se les atribuirá la categoría gramatical que corresponda a dicho término.

Ej.: En el banco tenía 17323 euros. En el banco tenía diecisiete mil trescientos veintitrés euros. Ocupaba el 21er puesto. Ocupaba el vigésimo primer puesto.

d) Porcentajes. Serán considerados como una expresión numérica, atribuyéndosele a dicha expresión la categoría gramatical que le corresponda en cada momento.

Ej.: Tenía un descuento del 10%. Tenía un descuento del 10 por ciento. Tenía un descuento del diez por ciento.

e) Puntos suspensivos. Tienen la misma categoría gramatical que el punto.

f) Signos monetarios. Los signos monetarios serán considerados como palabras y se les atribuirá la categoría gramatical que les corresponda en cada momento.

Ej.: Juan tiene 10 €. (10 → determinante; € → nombre)

g) Signos genéricos. Todos aquellos signos que no correspondan a ninguna de las categorías anteriores serán contabilizados como signos genéricos.

2. Indicadores basados en el lector

A diferencia de los indicadores basados en el texto, los indicadores basados en el lector plantean el análisis del texto desde una perspectiva menos objetiva, considerando el criterio personal del evaluador como la piedra angular del proceso en aras de obtener una valoración global del texto objeto de análisis. Éste es, sin duda, el método de evaluación más extendido en la comunidad educativa, utilizado de forma habitual por profesores de todo de mundo para valorar las composiciones escritas de sus alumnos. En el caso de profesores habituados a valorar composiciones escritas de alumnos de diferentes edades, la comparación entre los alumnos de diferentes niveles les permitirá emitir un juicio rápido, válido y fiable de la calidad del mismo, de su estructura dentro del género textual de que se trate y de su adecuación o coherencia.

Este tipo de indicadores tratan, por tanto, de hacer valer la experiencia el lector para evaluar tres aspectos básicos de la composición escrita, la estructura, la coherencia y la calidad

textual. A continuación de muestran algunos de los indicadores más notables en función del género textual, adaptados a partir de los trabajos realizados por Fitzgerald (1993).

2.1 Textos de comparación y contraste

Estructura

El evaluador, para dar su juicio sobre la estructura, considerará la presencia y desarrollo de seis características:

1. Información sobre un marco o introducción que presenta el texto.
2. Marcas o señales de estructura.
3. Una introducción, que presenta el objetivo y el tema del texto, la comparación y el contraste entre los temas.
4. Un desarrollo organizado y estructurado, por ejemplo, agrupando en un bloque las semejanzas y en otro bloque las diferencias.
5. Una unidad entre los párrafos.
6. Una conclusión final al texto.

Para valorar la estructura del texto y obtener la puntuación orden se utilizará la siguiente escala:

— No estructurada = 1 punto.

— Pobremente estructurada = 2 puntos.

— Parcialmente estructurada = 3 puntos.

— Bien estructurada = 4 puntos.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(0,5) Información sobre un marco o introducción que presenta el texto.

(0,5) Marcas o señales de estructura.

(0,5) Una introducción que presenta el objetivo y el tema del texto, la comparación y el con-traste entre los temas.

(1) Un desarrollo organizado y estructurado, por ejemplo: agrupando en un bloque las semejanzas y en otro bloque las diferencias.

(1) Una unidad entre los párrafos.

(0,5) Una conclusión final al texto.

Coherencia

El evaluador considerará el grado en el que se presentan siete características:

1. Identificación del tema.
2. La exposición del tema sin digresiones.
3. Un contexto que oriente al lector.

4. Los detalles se organizan en un plan distinguible que es sustentado a lo largo de todo el texto.
5. Existen marcas o enlaces que cohesionan oraciones y párrafos.
6. El discurso fluye sin problemas.
7. La existencia de una conclusión que dé sentido de final al texto.

Para evaluar la coherencia total del texto y obtener la puntuación orden se utilizará la siguiente escala:

— Incoherente = 1.

— Algo coherente = 2.

— Coherente = 3.

— Muy coherente = 4.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(0,4) Identificación del tema.

(0,8) La exposición del tema sin digresiones (vueltas hacia atrás, embrollos...).

(0,4) Un contexto el cual oriente al lector. (0,4) Los detalles se organizan en un plan distinguible que es sustentado a lo largo de todo el texto.

(0,8) Existen marcas o enlaces que cohesionan oraciones y párrafos.

(0,8) El discurso fluye sin problemas.

(0,4) La existencia de una conclusión que dé sentido de final al texto.

Calidad

Para emitir su juicio sobre la calidad del texto, el evaluador considerará seis dimensiones:

1. Una clara secuencia de ideas.
2. Buena organización
3. Vocabulario adecuado.
4. Variedad de detalles
5. Correcta estructura de las oraciones.
6. Correcta puntuación y ortografía.

Para evaluar la calidad del texto y obtener la puntuación orden se utilizará la siguiente escala:

— Inadecuado, difícil comprensión = 1.

— Apenas adecuado = 2.

— Adecuado = 3.

— Bueno = 4.

— Muy bueno = 5.

— Excelente = 6.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(1) Una clara secuencia de ideas.

(1) Buena organización.

(1) Vocabulario adecuado.

(1) Variedad de detalles.

(1) Correcta estructura de las oraciones.

(1) Correcta puntuación y ortografía.

2.2 Textos narrativos

Estructura

Para dar su juicio sobre la estructura, el evaluador considerará la presencia y desarrollo de seis características:

1. Información del marco que desarrolle el contexto.
2. Marcas o señales de estructura.
3. Un suceso que cree un propósito, problema o complicación para el principal protagonista.
4. Esfuerzo del protagonista para resolver el suceso.
5. Resultados de los esfuerzos del protagonista.
6. Una relación causal o temporal de los eventos.

Para evaluar la estructura del texto y obtener la puntuación orden se utilizará la siguiente escala:

— No estructurada = 1.

— Pobremente estructurada = 2.

— Parcialmente estructurada = 3.

— Bien estructurada = 4.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(0,6) Información del marco que desarrolle el contexto.

(0,6) Marcas o señales de estructura.

(0,6) Un suceso que cree un propósito, problema o complicación para el principal protagonista.

(0,6) Esfuerzo del protagonista para resolver el suceso.

(0,6) Resultados de los esfuerzos del protagonista.

(0,6) Una relación causal o temporal de los eventos.

Coherencia

El evaluador considerará el grado en el que se presentan siete características:

1. Identificación del tema.

2. La exposición del tema sin digresiones.
3. Un contexto que oriente al lector.
4. Los detalles se organizan en un plan distinguible que es sustentado a lo largo de toda la historia.
5. Existen marcas o enlaces que cohesionan oraciones y párrafos.
6. El discurso fluye sin problemas.
7. La existencia de una conclusión que cree un sentido de cierre al texto.

Para evaluar la coherencia total del texto y obtener la puntuación orden se utilizará la siguiente escala:

— Incoherente = 1.

— Algo coherente = 2.

— Coherente = 3.

— Muy coherente = 4.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(0,4) Identificación del tema.

(0,8) La exposición del tema sin digresiones.

(0,4) Un contexto que oriente al lector.

(0,4) Los detalles se organizan en un plan distinguible que es sustentado a lo largo de todo el texto.

(0,8) Existen marcas o enlaces que cohesionan oraciones y párrafos.

(0,8) El discurso fluye sin problemas.

(0,4) La existencia de una conclusión que dé sentido de final al texto.

Calidad

Para emitir su juicio sobre la calidad del texto, el evaluador considerará siete dimensiones:

1. Una clara secuencia de eventos.
2. El desarrollo de la historia sin descripciones irrelevantes en la exposición.
3. Buena organización.
4. Vocabulario adecuado.
5. Variedad de detalles.
6. Correcta estructura de las oraciones.
7. Correcta puntuación y ortografía.

Para evaluar la calidad del texto y obtener la puntuación orden se utilizará la siguiente escala:

— Inadecuado, difícil comprensión = 1.

— Apenas adecuado = 2.

— Adecuado = 3.

— Bueno = 4.

— Muy bueno = 5.

— Excelente = 6.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(0,86) Una clara secuencia de eventos.

(0,86) El desarrollo de la historia sin descripciones irrelevantes en la exposición.

(0,86) Buena organización.

(0,86) Vocabulario adecuado.

(0,86) Variedad de detalles.

(0,86) Correcta estructura de las oraciones.

(0,86) Correcta puntuación y ortografía.

2.3 Textos descriptivos

Estructura

Para dar su juicio sobre la estructura, el evaluador considerará la presencia y desarrollo de seis características:

1. Información sobre un marco o introducción que presenta el texto.

2. Marcas o señales de estructura.
3. Una introducción que presenta el objetivo y el tema del texto, la comparación y el contraste entre los temas.
4. Un desarrollo organizado y estructurado, por ejemplo, agrupando en un bloque las semejanzas y en otro bloque las diferencias.
5. Una unidad entre los párrafos.
6. Una conclusión final al texto.

Para evaluar la estructura del texto y obtener la puntuación orden se utilizará la siguiente escala:

— No estructurada = 1 punto.

— Pobremente estructurada = 2 puntos.

— Parcialmente estructurada = 3 puntos.

— Bien estructurada = 4 puntos.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(0,5) Información sobre un marco o introducción que presenta el texto.

(0,5) Marcas o señales de estructura.

(0,5) Una introducción que presenta el objetivo y el tema del texto, la comparación y el con-traste entre los temas.

(1) Un desarrollo organizado y estructurado, por ejemplo, agrupando en un bloque las semejanzas y en otro bloque las diferencias.

(1) Una unidad entre los párrafos.

(0,5) Una conclusión final al texto.

Coherencia

El evaluador considerará el grado en el que se presentan siete características:

1. Identificación del tema.
2. La exposición del tema sin digresiones.
3. Un contexto que oriente al lector.
4. Los detalles se organizan en un plan distinguible que es sustentado a lo largo de todo el texto.
5. Existen marcas o enlaces que cohesionan oraciones y párrafos.
6. El discurso fluye sin problemas.
7. La existencia de una conclusión que dé sentido de final al texto.

Para evaluar la coherencia total del texto y obtener la puntuación orden se utilizará la siguiente escala:

— Incoherente = 1.

— Algo coherente = 2.

— Coherente = 3.

— Muy coherente = 4.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(0,4) Identificación del tema.

(0,8) La exposición del tema sin digresiones.

(0,4) Un contexto que oriente al lector.

(0,4) Los detalles se organizan en un plan distinguible que es sustentado a lo largo de todo el texto.

(0,8) Existen marcas o enlaces que cohesionan oraciones y párrafos.

(0,8) El discurso fluye sin problemas.

(0,4) La existencia de una conclusión que dé sentido de final al texto.

Calidad

Para emitir su juicio sobre la calidad del texto, el evaluador considerará seis dimensiones:

1. Una clara secuencia de ideas.
2. Buena organización.

3. Vocabulario adecuado.
4. Variedad de detalles.
5. Correcta estructura de las oraciones.
6. Correcta puntuación y ortografía.

Para evaluar la calidad del texto y obtener la puntuación orden se utilizará la siguiente escala:

— Inadecuado, difícil comprensión = 1.

— Apenas adecuado = 2.

— Adecuado = 3.

— Bueno = 4.

— Muy bueno = 5.

— Excelente = 6.

No obstante, se obtendrá una puntuación suma siguiendo los siguientes criterios:

(1) Una clara secuencia de ideas.

(1) Buena organización.

(1) Vocabulario adecuado.

(1) Variedad de detalles.

(1) Correcta estructura de las oraciones.

(1) Correcta puntuación y ortografía.

3. Consideraciones adicionales

A continuación se presentan los símbolos y colores utilizados en la corrección manual para identificar cada uno de los indicadores mencionados en este protocolo de corrección.

Tabla A.16
Leyenda de corrección manual

Tipo de indicador	*Símbolo*
Palabras de contenido	Rectángulo en bolígrafo azul
Palabras funcionales	Triángulo en bolígrafo azul.
Determinantes	Círculo en bolígrafo azul.
Indicadores referenciales	Subrayado en rotulador violeta.
Indicadores lexicales	Subrayado en rotulador rojo.
Indicadores meta-estructurales	Subrayado rotulador marrón.
Estructuradores	Subrayado en fluorescente naranja.
Conectores	Subrayado en fluorescente amarillo.
Reformuladores	Subrayado en fluorescente rosa.
Argumentativos	Subrayado en fluorescente verde.

Apéndice B

Connexor Machinese Phrase Tagger

A continuación se muestran los listados completos de etiquetas empleadas por el etiquetador *Connexor MPT* proporcionados por la empresa *Connexor Oy*.

Tabla B.1
Etiquetas morfológicas del Connexor MPT

Tag			*Description*
Custom			User-defined word
Part of speech	*Subfeature 1*	*Subfeature 2*	
A			adjective
	CMP		comparative
	SUP		superlative
ADV			adverb
	CMP		comparative
	SUP		superlative
CC			coordination marker
	Prop		coordination marker belongs to a proper noun
CS			clause marker
DET			determiner
	Prop		determiner belongs to a proper noun
INTERJ			interjection

Tabla B.1
Etiquetas morfológicas del Connexor MPT (continuación)

Tag			*Description*
Part of speech	*Subfeature 1*	*Subfeature 2*	
N			indicative
	Abbr		imperative
	Prop		subjunctive
		PL	infinitive
NUM			participle
	ORD		gerund
PREP	Prop		present
PRON			future
V			past
	IND		imperfect
	IMP		preterite
	SUB		conditional
	INF		progressive
	GR		perfect
		PRES	indicative
		FUT	imperative
		PAST	subjunctive
		IMPF	infinitive
		PRET	participle
		CND	gerund
		PROG	present
		PERF	future

Tabla B.2
Etiquetas sintácticas del Connexor MPT

Tag	*Description*
@ADVL	adverbial head
@AUX	auxiliary verb
@CC	coordination marker
@MAIN	main verb
@NH	nominal head
@PREMARK	preposed marker
@PREMOD	premodifier
@POSTMOD	postmodifier

Tabla B.3
Otras etiquetas del Connexor MPT

Tag	*Description*
SB	sentence break
NP-Single	single word noun phrase
NP-First	first word of a noun phrase
NP-Internal	word is within a multiword noun phrase
NP-Last	last word of a noun phrase

Apéndice C

Coeficiente Kappa de Cohen

El cálculo del índice de acuerdo entre evaluadores necesario para realizar la validación de *Seshat Product* se ha llevado a cabo haciendo uso del coeficiente Kappa de Cohen, una medida estadística especialmente indicada para reducir el efecto del azar en los resultados finales y cuyo cálculo se ajusta a la siguiente fórmula:

$$k = \sum_{1}^{k} \frac{p_0 - p_e}{1 - p_e} \times 100$$

Donde p_0 es la probabilidad observada o proporción de veces que los resultados aportados por los evaluadores coinciden:

$$p_0 = \frac{\text{Acuerdos}}{\text{Acuerdos} + \text{Desacuerdos}}$$

Y p_e es la probabilidad esperada por azar o proporción de veces que los resultaos aportados por los evaluadores coinciden fruto de la casualidad:

$$p_{e(ct1,ct2)} = \frac{O_{ct1}}{T} \times \frac{O_{ct2}}{T}$$

Siendo $\mathrm{ct1}$ el primer evaluador y $\mathrm{ct2}$ el segundo evaluador, $\mathrm{O}_{\mathrm{ct1}}$ y $\mathrm{O}_{\mathrm{ct2}}$ el número de observaciones[1] realizadas por cada uno de ellos y T el número total de observaciones realizadas.

[1] En este caso en concreto las observaciones realizadas coinciden con el número de indicadores empleados en el análisis del texto por cada evaluador.

Referencias

Adjetivo calificativo. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Adverbio. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Alamargot, D., & Chanquoy, L. (2001). *Through the Models of Writing*. Dordrecht: Kluwer Academic Press.

Álvarez, M. L., & García, J. N. (2014). Development of the Writing Process from Primary to Secondary Education. *Revista de Psicodidáctica, 19*(1), 5-26. doi: 10.1387/RevPsicodidact.6518

Álvarez, M. L., & García, J. N. (2014b). The orchestration of processes in relation to the product, and the role of psychological variables in written composition. *A, 30*(2), 1-24.

Álvarez, M. L., García, J. N., García Martín, J., & Díez-Caso, H. (2012). Contraste de dos modalidades instruccionales en competencia comunicativa escrita: producto y proceso. *International Journal of Developmental and Educational Psychology, 4*(1), 261-266.

Amiran, E., & Mann, J. (1982). Written Composition, Grades K-12: Literature Synthesis and Report. Portland, OR: Northwest Regional Educational Laboratory.

Arias-Gundín, O., & García, J. N. (2007). Eficacia de la instrucción en los aspectos mecánicos y/o semánticos de la revisión textual. *Análisis y modificación de conducta, 33*(147), 6.

Arias-Gundín, O., & García, J. N. (2012). Are help levels effective in textual revision? In D. Alamargot, M. Castelló, F. Ganier, O. Kruse, A. Mangen, L. Tolchinsky & L. van Waes (Eds.),

Learning to write effectively - Current Trends in European Research (pp. 73-75). Bingley: Emerald.

Artículo. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation, 2*, 89-195.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment, 4*(3).

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559.

Bennett, R. (2004). Moving the field forward: Some thoughts on validity and automated scoring (ETS RM 04-01). *Princeton, NJ: Educational Testing Service*.

Bereiter, C., & Scardamalia, M. (1992). Dos modelos explicativos de los procesos de composición escrita. *Infancia y aprendizaje, 58*(43-64).

Beriter, C., & Scardamalia, M. (1987). *The psychology of written composition*: Hillsdale, NJ: Erlbaum.

Berninger, V. W. (1999). Coordinating transcription and text generation in working memory during composing: Automatic and constructive processes. *Learning Disability Quarterly, 22*(2), 99-112.

Braaksma, M. A., Rijlaarsdam, G., Van den Bergh, H., & van Hout-Wolters, B. H. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and instruction, 22*(1), 1-36.

Breland, H. M., Bonner, M. W., & Kubota, M. Y. (1995). *Factors in performance on brief, impromptu essay examinations*. New York, NY: College Entrance Examination Board.

Brown, J. S., McDonald, J. L., Brown, T. L., & Carr, T. H. (1988). Adapting to processing demands in discourse production:

The case of handwriting. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 45.

Burstein, J., Tetreault, J., & Madnani, N. (2013). The E-rater® Automated Essay Scoring System. In M. D. Shermis & J. Burstein (Eds.), *Handbook of Automated Essay Evaluation: Current Applications and New Directions* (pp. 55-67). New York, NY: Routledge.

Butter, M. C., Pérez, L. J., & Quintana, M. G. B. (2014). School networks to promote ICT competences among teachers. Case study in intercultural schools. *Computers in Human Behavior, 30*(0), 442-451. doi: http://dx.doi.org/10.1016/j.chb.2013.06.024

Collins, J. L. (1998). *Strategies for struggling writers*. New York, NY: The Guilford Press.

Conjunción. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Connexor Oy. (2014). Machinese Phrase Tagger. Retrieved March 14, 2014, from http://www.connexor.com/nlplib/?q=mpt

Crossley, S. A., & McNamara, D. S. (2010). *Cohesion, coherence, and expert evaluations of writing proficiency.* Paper presented at the Proceedings of the 32nd annual conference of the Cognitive Science Society.

Cha, S. E., Jun, S. J., Kwon, D. Y., Kim, H. S., Kim, S. B., Kim, J. M., . . . Lee, W. G. (2011). Measuring achievement of ICT competency for students in Korea. *Computers & Education, 56*(4), 990-1002. doi: http://dx.doi.org/10.1016/j.compedu.2010.11.003

de Caso, A. M., García, J. N., Díez, C., Robledo, P., & Álvarez, M. L. (2010). Enhancing Writing Self-Efficacy Beliefs of Students with Learning Disabilities Improves their Writing Processes and Products. *Electronic Journal of Research in Educational Psychology, 8*(1).

De Souza, A. F., Pedroni, F., Oliveira, E., Ciarelli, P. M., Henrique, W. F., & Veronese, L. (2007, 20-24 Oct. 2007). *Automated Free Text Classification of Economic Activities Using VG-RAM Weightless Neural Networks.* Paper presented at the Intelligent Systems Design and Applications, 2007. ISDA 2007. Seventh International Conference on.

DeVillez, R. (2003). *Writing: Step by step*. Dubuque, IO: Kendall Hunt.

Díez-Caso, H., & García, J. N. (2011a). *Automatización de procesos de evaluación y análisis psicolingüísticos en textos escritos.* Paper presented at the VI Congreso Internacional de Psicología y Educación, Valladolid, Spain.

Díez-Caso, H., & García, J. N. (2011b). *Comparativa entre etiquetadores morfológicos para la automatización de análisis textuales.* Paper presented at the VI Congreso Internacional de Psicología y Educación, Valladolid, Spain.

Díez-Caso, H., & García, J. N. (2011c). Seshat: Automatic analysis of written texts. *International Journal of Developmental and Educational Psychology, 5*(1), 479-486.

Díez-Caso, H., & García, J. N. (2013). Dificultades en el análisis automático de textos escritos. *International Journal of Developmental and Educational Psychology, 2*(1), 695-702.

Díez-Caso, H., & García, J. N. (under review a). Analyzing linguistic features of productivity and coherence with Seshat Product.

Díez-Caso, H., & García, J. N. (under review b). A web-based approach to the investigation of cognitive activities in writing.

Díez-Caso, H., García, J. N., García-Martín, J., & Álvarez, M. L. (2013). Plataforma web de aprendizaje e instrucción: ilustración en el Fomento de Competencias Ocupacionales. In M. d. C. Pérez & M. d. M. Molero (Eds.), *Variables psicológicas y educativas para la intervención en el ámbito*

escolar (pp. 407-410). Almería: Asociación Universitaria de Educación y Psicología.

Dikli, S. (2006). An Overview of Automated Scoring of Essays. *The Journal of Technology, Learning, and Assessment, 5*(1).

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). *Using latent semantic analysis to improve access to textual information*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Washington, D.C., USA.

Fidalgo, R., & García, J. N. (2008). Development of written competence through metacognitive writing instruction. *Cultura y Educación, 20*(3), 325-346. doi: 10.1174/113564008785826321

Fidalgo, R., Torrance, M., & García, J.-N. (2008). The long-term effects of strategy-focussed writing instruction for grade six students. *Contemporary Educational Psychology, 33*(4), 672-693. doi: http://dx.doi.org/10.1016/j.cedpsych.2007.09.001

Gabrielatos, C. (2002). EFL Writing: Product and Process. *ERIC, ED476839*.

Garcia-Debanc, C., & Fayol, M. (2002). Apports et limites des modèles du processus rédactionnel pour la didactique de la production écrite. Dialogue entre psycholinguistes et didacticiens. *pratiques, 115*(116), 37-48.

García-Martín, E., & García, J. N. (2012). Design of an open corpus and computer tool for writing development and instruction among students 8 to 16 years old with and without learning disabilities. In D. Alamargot, M. Castelló, F. Ganier, O. Kruse, A. Mangen, L. Tolchinsky & L. van Waes (Eds.), *Learning to write effectively - Current Trends in European Research* (pp. 331-333). Bingley: Emerald.

García-Martín, J., García, J. N., Álvarez, M. L., & Diez-Caso, H. (2013). Efectos diferenciales en el desarrollo de

competencias. In J. J. Gázquez, M. d. C. Pérez, M. d. M. Molero & R. Parra Codina (Eds.), *Investigación en el ámbito escolar: un acercamiento multidimensional a las variables psicológicas y educativas* (pp. 45-48). Granada: Editorial GEU.

García-Martín, J., García, J. N., Álvarez, M. L., & Díez-Caso, H. (2012). Eficacia instruccional de diversas herramientas de la web 2.0 en la mejora de competencias ocupacionales. *International Journal of Developmental and Educational Psychology, 1*(4), 309-315.

García-Martín, J., García, J. N., Álvarez, M. L., & Díez-Caso, H. (2013a). *Efectos de variables psicológicas en un programa de fomento de competencias ocupacionales.* Paper presented at the Congreso Internacional de Convivencia Escolar: Contextos Psicológicos y Educativos, Almería, Spain.

García-Martín, J., García, J. N., Álvarez, M. L., & Díez-Caso, H. (2013b). *Mejoras en la competencia digital tras la aplicación de un programa de competencias ocupacionales.* Paper presented at the Congreso Internacional de Convivencia Escolar: Contextos Psicológicos y Educativos, Almería, Spain.

García, J. N. (2002). *Aplicaciones De Intervencion Psicopedagogica*. Madrid: Ediciones Pirámide.

García, J. N., & Arias-Gundín, O. (2004). Intervención en estrategias de revisión del mensaje escrito. *Psicothema, 16*(2), 194-202.

García, J. N., & Bausela, E. H. (2004). Relación del conocimiento de la persona, de la tarea y de la estrategia en relación a la escritura con la coherencia y productividad de la composición escrita en niños/as con y sin dificultades de aprendizaje y/o bajo rendimiento. *Revista de Educación*(334), 223-234.

Garcia, J. N., & de Caso, A. M. (2004). Effects of a motivational intervention for improving the writing of children with learning disabilities. *Learning Disability Quarterly, 27*(3), 141-159. doi: 10.2307/1593665

García, J. N., & de Caso, A. M. (2002). ¿Es posible mejorar la productividad y coherencia de los textos escritos por alumnos con dificultades de aprendizaje y/o bajo rendimiento sin que cambie su disposición reflexiva hacía la escritura? *Psicothema, 14*(2), 456-462.

García, J. N., & de Caso, A. M. (2006). Comparison of the effects on writing attitudes and writing self-efficacy of three different training programs in students with learning disabilities. *International Journal of Educational Research, 43*(4–5), 272-289. doi: http://dx.doi.org/10.1016/j.ijer.2006.06.006

García, J. N., & de Caso, A. M. (2006). Changes in Writing Self-Efficacy and Writing Products and Processes through Specific Training in the Self-Efficacy Beliefs of Students with Learning Disabilities. *Learning Disabilities: A Contemporary Journal, 4*(2), 1-27.

García, J. N., & de Caso, A. M. (2007). Effectiveness of an improvement writing program according to students' reflexivity levels. *The Spanish journal of psychology, 10*(2), 303-313.

García, J. N., Fidalgo-Redondo, R., Arias-Gundín, O., Marbán, J. M., de Caso, A. M., Pacheco, D., & Díez-Caso, H. (2014). El análisis psicolingüístico del producto textual. In J. N. García (Ed.), *Prevención en dificultades del desarrollo y del aprendizaje*. Madrid: Ediciones Pirámide.

García, J. N., & Fidalgo, R. (2003). Diferencias en la conciencia de los procesos psicológicos de la escritura: mecánicos frente a sustantivos y otros. *Psicothema, 15*(1), 41-48.

García, J. N., & Fidalgo, R. (2004). El papel del autoconocimiento de los procesos psicológicos de la escritura en la calidad de las composiciones escritas. *Revista de Psicología General y Aplicada, 57*(3), 181-298.

García, J. N., & Fidalgo, R. (2006). Effects of two types of self-regulatory instruction programs on students with learning disabilities in writing products, processes, and self-efficacy. *Learning Disability Quarterly, 29*(3), 181-211. doi: 10.2307/30035506

García, J. N., & Fidalgo, R. (2008a). Fostering the self-regulation of the recursive thinking involved in composition writing. In A. Valle, J. C. Núñez, R. González-Cabanach, J. A. González-Pienda & S. Rodríguez (Eds.), *Handbook of Instructional Resources & Applications* (pp. 171-185). New York, NY: Nova Science Publishers.

García, J. N., & Fidalgo, R. (2008b). The Orchestration of Writing Processes and Writing Products: A comparison of 6th Grade Students With and Without Learning Disabilities. *Learning Disabilities: A Contemporary Journal, 6(2)*, 77-98.

García, J. N., & Fidalgo, R. (2008c). Writing self-efficacy changes after cognitive strategy intervention in students with learning disabilities: the mediational role of gender in calibration. *The Spanish journal of psychology, 11*(2), 414-432.

García, J. N., & Marbán, J. M. (2002). *Intervención instruccional en la composición escrita*. Barcelona: Ariel.

García, J. N., & Marbán, J. M. (2003). Writing composition processes in LD/LA students: An instructional study on planning. *Infancia y aprendizaje, 26*(1), 97-113. doi: 10.1174/02103700360536455

García, J. N., & Pacheco, D. (2007). El rendimiento de los alumnos y el papel de la práctica en la composición escrita.

International Journal of Developmental and Educational Psychology, *19*(1), 265-278.

García, J. N., & Pacheco, D. (2012a). Diagnosis of teachers' practice in the teaching of written composition. In W. Sittiprapaporn (Ed.), *Learning Disabilities* (pp. 175-194). Rijeka: InTech.

García, J. N., & Pacheco, D. (2012b). The quality of teaching determines students' achievement in writing. In W. Sittiprapaporn (Ed.), *Learning Disabilities* (pp. 175-194). Rijeka: InTech.

García, J. N., Rodríguez-Pérez, C., Pacheco, D. I., & Díez, C. (2009). Influence of cognitive effort, sustained attention, working memory and ADHD symptoms in the process and product of written composition. An experimental study. *Estudios de Psicología*, *30*(1), 31-50. doi: 10.1174/021093909787536326

Garner, R. (2013). Professors angry over essays marked by computer *The Independent*.

Graham, S., & Harris, K. R. (1996). Self-regulation and strategy instruction for students who find writing and learning challenging. *The science of writing: Theories, methods, individual differences, and applications*, 347-360.

Hastings, P., Hughes, S., Magliano, J., Goldman, S., & Lawless, K. (2012). Assessing the use of multiple sources in student essays. *Behavior Research Methods*, *44*(3), 622-633. doi: 10.3758/s13428-012-0214-0

Hatlevik, O. E., & Christophersen, K.-A. (2013). Digital competence at the beginning of upper secondary school: Identifying factors explaining digital inclusion. *Computers & Education*, *63*(0), 240-247. doi: http://dx.doi.org/10.1016/j.compedu.2012.11.015

Haydel, E., & Bucolo, D. (2010). A Research-based Approach to Writing. Retrieved February 20, 2014, from http://hmhco-

v1.dev.techspa.com/~/media/sites/home/education/global/pdf/white-papers/language-arts/writing/write-source/online-technology-research-130205.pdf

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing. Theories, methods, individual differences, and applications* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum.

Hayes, J. R., & Flower, L. (1980). Identifying the organization of writing processes. *Cognitive processes in writing*, 3-30.

Hernández-Ramos, J. P., Martínez-Abad, F., García Peñalvo, F. J., Esperanza Herrera García, M., & Rodríguez-Conde, M. J. (2014). Teachers' attitude regarding the use of ICT. A factor reliability and validity study. *Computers in Human Behavior*, *31*(0), 509-516. doi: http://dx.doi.org/10.1016/j.chb.2013.04.039

Hernández, A., & Quintero, A. (2001). Comprensión y composición escrita. Estrategias de aprendizaje. Madrid: Editorial Síntesis.

Herrington, A., Hodgson, K., Moran, C., & Eidman-Aadahl, E. (2009). *Teaching the new writing: Technology, change, and assessment in the 21st-century classroom*. New York: Teachers College Press.

Hillocks, G. (1984). What works in teaching composition: A meta-analysis of experimental treatment studies. *American Journal of Education*, 133-170.

Hillocks, G. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

Holdzkom, D., Reed, L., Porter, H. J., & Rubin, D. L. (1982). *Research within reach: Oral and written communication*. St. Louis: Cemrel, Inc.

Iakobsen, A. L. (2003). Effects of think aloud on translation speed. *Triangulating translation: perspectives in process oriented research, 45*, 69-95.

Instituto Cervantes. (2012). El español: una lengua viva. Informe 2012. Retrieved May 20, 2014, from http://cvc.cervantes.es/lengua/anuario/anuario_12/i_cervantes/p01.htm

Interjección. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Janssen, D., van Waes, L., & van den Bergh, H. (1996). Effects of thinking aloud on writing processes. In C. M. L. S. Ransdell (Ed.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 233-250). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Justicia, F. (1996). Metacognición y currículo. *B. Llera y G. Genovard, Psicología de la instrucción I. Variables y procesos básicos*, 300-381.

Kahneman, D. (1973). *Attention and Effort*: Prentice Hall.

Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory & Cognition, 15*(3), 256-266. doi: 10.3758/BF03197724

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. L. S. Ransdell (Ed.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57-71). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Kellogg, R. T. (1999). *The psychology of writing*: Oxford University Press.

Kerr, B. (1973). Processing demands during mental operations. *Memory & Cognition, 1*(4), 401-412.

Koh, J. H. L. (2011). Computer skills instruction for pre-service teachers: A comparison of three instructional approaches.

Computers in Human Behavior, *27*(6), 2392-2400. doi: http://dx.doi.org/10.1016/j.chb.2011.08.002

Krings, H. P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes* (G. Shreve, Trans.). Kent: Kent State University Press.

Landauer, T. K., Laham, D., & Foltz, P. W. (2001). *The intelligent essay assessor: putting knowledge to the test.* Paper presented at the association of test publishers computer-based testing: emerging technologies and opportunities for diverse applications conference, Tucson, AZ.

Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automated Essay Scoring: A cross disciplinary perspective. In M. D. Shermis & J. Burstein (Eds.), *Automated Essay Scoring and annotation of essays with the Intelligent Essay Assessor* (pp. 87–112). Mahwah, NJ: Lawrence Erlbaum Associates.

Larson, L. (2013). Outrage over software that automatically grades college essays to spare professors from having to assess students' work, *Daily Mail*.

Levy, C. M., & Ransdell, S. (1995). Is writing as difficult as it seems? *Memory & Cognition*, *23*(6), 767-779.

Magliano, J. P., & Graesser, A. C. (2012). Computer-based assessment of student-constructed responses. *Behavior Research Methods*, *44*(3), 608-621. doi: 10.3758/s13428-012-0211-3

Marbán, J. M. (2001). *Los procesos de planificación de la escritura en alumnos con y sin dificultades de aprendizaje. Evaluación e Intervención.* Universidad de León, León.

Markoff, J. (2013). Essay-Grading Software Offers Professors a Break. *The New York Times*.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*(4), 292-330.

McNamara, D. S., Ozuru, Y., Graesser, A. C., & Louwerse, M. (2006). *Validating coh-metrix.* Paper presented at the Proceedings of the 28th annual conference of the cognitive science society.

Mellor, A. (2011). Essay Length, Lexical Diversity and Automatic Essay Scoring. *Memoirs of the Osaka Institute of Technology*, 1-14.

Muñoz de la Peña, D., Gómez-Estern, F., & Dormido, S. (2012). A new Internet tool for automatic evaluation in Control Systems and Programming. *Computers & Education, 59*(2), 535-550. doi: http://dx.doi.org/10.1016/j.compedu.2011.12.016

Myers, M. (2003). What can computers and AES contribute to a K–12 writing program? In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 3-20). Mahwah, NJ: Lawrence Erlbaum Associates.

Nombre. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Olive, T. (2010). Methods, techniques, and tools for the on-line study of the writing process. In N. L. Mertens (Ed.), *Writing: Processes, Tools and Techniques* (pp. 1-18). New York: Nova Science Publishers, Inc.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory & Cognition, 30*(4), 594-600.

Ott, M., & Pozzi, F. (2011). Towards a new era for Cultural Heritage Education: Discussing the role of ICT. *Computers in Human Behavior, 27*(4), 1365-1371. doi: http://dx.doi.org/10.1016/j.chb.2010.07.031

Pacheco, D., García, J. N., & Díez, C. (2009). Autoeficacia, enfoque y papel de la práctica de los maestros en la enseñanza de la escritura. *European Journal of Education and Psychology, 2*(1), 5-23.

Pacheco, D., García, J. N., & Díez, C. (2010a). Academic performance and the role of self-regulated practice of the teachers in writing. . In J. de la Fuente & M. A. Elissa (Eds.), *International Perspectives on Applying Self-Regulated Learning in Different Settings* (pp. 575-594). Bern: Peter Lang Publisher.

Pacheco, D., García, J. N., & Díez, C. (2010b). Self-regulation of teachers' practice in teaching writing. In J. de la Fuente & M. A. Elissa (Eds.), *International Perspectives on Applying Self-Regulated Learning in Different Settings* (pp. 553-574). Bern: Peter Lang Publisher.

Page, E. B. (2003). Project essay grade: PEG. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43-54). Mahwah, NJ: Lawrence Erlbaum Associates.

Pakhomov, S., Chacon, D., Wicklund, M., & Gundel, J. (2011). Computerized assessment of syntactic complexity in Alzheimer's disease: a case study of Iris Murdoch's writing. *Behavior Research Methods, 43*(1), 136-144. doi: 10.3758/s13428-010-0037-9

Parson, G. (1985). Hand in Hand: The Writing Process and the Microcomputer. Two Revolutions in the Teaching of Writing. *Juneau, AK: Alaska Department of Education Office of Instructional Services*.

Pęzik, P. (2010). Computational and corpus linguistics. In B. Lewandowska-Tomaszczyk (Ed.), *New Ways to Language* (pp. 433-460). Wydawnictwo Uniwersytetu Łódzkiego.

Piolat, A., Kellogg, R. T., & Farioli, F. (2001). The triple task technique for studying writing processes: on which task is attention focused? *Current Psychology Letters. Brain, Behavior and Cognition, 4*, 67-83.

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*(3), 291-312.

Piolat, A., Olive, T., Roussey, J.-Y., Thunin, O., & Ziegler, J. C. (1999). SCRIPTKELL: A tool for measuring cognitive effort and time processing in writing and other complex cognitive activities. *Behavior Research Methods, Instruments, & Computers, 31*(1), 113-121.

Piolat, A., Roussey, J.-Y., Olive, T., & Farioli, F. (1996). Charge mentale et mobilisation des processus rédactionnels: examen de la procédure de Kellogg. *Psychologie française, 41*(4), 339-354.

Power, M. J. (1986). A technique for measuring processing load during speech production. *Journal of Psycholinguistic Research, 15*(5), 371-382.

Preposición. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Pribram, K., Miller, G., & Galanter, E. (1960). Plans and the Structure of Behavior. *New York: Holt & Co*.

Pronombre demostrativo. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Pronombre indefinido. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Pronombre Interrogativo. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Pronombre personal. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Pronombre posesivo. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Pronombre relativo. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Pronombre. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Quinlan, T., Higgins, D., & Wolff, S. (2009). Evaluating the Construct-Coverage of the e-rater [R] Scoring Engine. Research Report. ETS RR-09-01. *Educational Testing Service*.

Robledo, P., & García, J. N. (2013). Strategy Instruction for Writing Composition at School and Home. Psychological Studies. *Psychological Studies, 34*(2), 1-39.

Rudner, L. M., & Gagne, P. (2001). An overview of three approaches to scoring written essays by computer. *Practical Assessment, Research & Evaluation, 7(26).*

Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric™ essay scoring system. *The Journal of Technology, Learning and Assessment, 4*(4).

Rudner, L. M., & Liang, T. (2002). Automated essay scoring using Bayes' theorem. *The Journal of Technology, Learning and Assessment, 1*(2).

Salem, A. B. M. (2000). *The potential role of artificial intelligence technology in Education.* Paper presented at the Annual Meeting of the International Conference on Technology in Mathematics Education, Beirut, Lebanon.

Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge based questioning by children. *Cognition and instruction, 9*(3), 177-199.

Schultz, M. T. (2013). The IntelliMetricTM Automated Essay Scoring Engine—A Review and an Application to Chinese Essay Scoring. In M. D. Shermis & J. Burstein (Eds.), *Handbook of Automated Essay Evaluation: Current Applications and New Directions* (pp. 89-98). New York, NY: Routledge.

Shen, Y. (2010). Qualitative Characteristics of Coherence, Substitution, and Reference by Non-English Major Chinese Students. *English Language Teaching, 3*(2).

Shermis, M. D., & Barrera, F. D. (2002). *Exit Assessments: Evaluating Writing Ability through Automated Essay Scoring*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. *International encyclopedia of education, 4*, 20-26.

Shermis, M. D., & Burstein, J. C. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Routledge.

Shermis, M. D., Raymat, M. V., & Barrera, F. (2003). *Assessing Writing through the Curriculum with Automated Essay Scoring*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Spencer, S. L., & Fitzgerald, J. (1993). Validity and structure, coherence, and quality measures in writing. *Journal of Literacy Research, 25*(2), 209-231.

Streeter, L., Psotka, J., Laham, D., & MacCuish, D. (2002). *The credible grading machine: Automated essay scoring in the dod.* Paper presented at the The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

The Apache Software Foundation. (2014). Apache POI - the Java API for Microsoft Documents. Retrieved March 14, 2014, from http://poi.apache.org/

Torrance, M., Fidalgo, R., & García, J. N. (2007). The teachability and effectiveness of cognitive self-regulation in sixth-grade writers. *Learning and Instruction, 17*(3), 265-285. doi: http://dx.doi.org/10.1016/j.learninstruc.2007.02.003

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. San Francisco: John Wiley & Sons.

Tynjälä, P., Mason, L., & Lonka, K. (2001). *Writing as a learning tool. Integrating theory and practice.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Verbo. (2001). In *Diccionario de la lengua española* (22nd ed). Madrid: Real Academia Española.

Wesdorp, H. (1983). *On the Identification of Critical Variables in Written Composition Instruction*. Paper presented at the 72nd Annual Convention of the National Council of Teachers of English, Washington, DC.

Yang, Y., Buckendahl, C. W., Juszkiewicz, P. J., & Bhola, D. S. (2002). A review of strategies for validating computer-automated scoring. *Applied Measurement in Education, 15*(4), 391-412.

Zhong, Z.-J. (2011). From access to usage: The divide of self-reported digital skills among adolescents. *Computers & Education, 56*(3), 736-746. doi: http://dx.doi.org/10.1016/j.compedu.2010.10.016

Certificados e informes internacionales[1]

Documentación requerida para la obtención de la Mención Internacional

[1] Los documentos presentados en esta sección son copias de los originales entregados en la Unidad de Doctorado de la Universidad de León.

UNIVERSITEIT VAN AMSTERDAM

Letter of Confirmation

To whom it may concern,

I hereby certify that Mr. Héctor Díez Caso, with personal ID number 71446808E, has made a stay at the University of Amsterdam, performing tasks related to his doctoral research, from April 8th to July 9th, 2013.

Signed by Martine Braaksma, PhD, Assistant Professor, at the University of Amsterdam in Amsterdam, Netherlands, 9th of July of 2013.

Universiteit Utrecht

External review in support of Héctor Díez Caso's pursuit of a PhD degree with International Mention

Huub van den Bergh
PhD, Professor
Departement Nederlands
Trans 10
3512 JK Utrecht
The Netherlands

Faculteit Geesteswetenschappen

Bezoekadres
Trans 10
Utrecht

Uw kenmerk
Ons kenmerk

Telefoon 030 253 8137
E-mail H.vandenBergh@uu.nl
Datum 11-06-2014

To whom it may concern,

I have read two articles / chapters of Héctor Díez Caso: Assessment of cognitive activities in writing: a web based approach and Analyzing linguistic features of productivity and coherence with Seshat product.

The first article / chapter presents a web-based tool for the study of writing processes. This tool is based on a so-called triple task technique, which not only allows for the measurement of delays to a secondary task, but also requires (young) writers to categorize their cognitive activities in terms of writing processes like planning, translating and reviewing. Whereas the reaction time to the secondary task is usually considered as a measure of cognitive effort, for the interpretation of the results / delays the categorization is essential.

In my opinion this article / chapter is state of the art! Such a web-based tool is a leap forward for writing research, as the collection of data becomes much easier. Furthermore, the tool allows for the analysis of writing processes in real time, which is, according to modern theories of writing processes, crucial for understanding of the writing process. The (empirical) data that can be gathered by this too undoubtedly will allow for further theory and analysis of writing processes.

The second article / chapter presents a tool for automatic essay rating. As is well known rating of essays is quite cumbersome. Therefore the writing research community is in need of tools which allows for a more objective (or intersubjective) way of rating the quality of essays. In different countries researchers work on the same type of instruments, as text quality ratings are

clearly language (and cultural) dependent. The proposed “Seshat Product” seems a good tool to analyze the productivity and coherence of the writing of young students.

I always have doubts on analytic scoring systems, in which text-characteristics are being counted, in order to get some kind of quality assessment, as I think that the quality of a text is more than some linguistic elements in isolation. I was pleasantly surprised by the stance of the author that this tool as an aid in scoring and does not provide an undebatable grade. Hence, the tool is aid for teachers to help them marking essays and giving commenting students’ texts.

In my opinion both articles / chapters are good in their own right. However, taken together they will give a boost to writing research in general. Writing research profit from booth tools, but also writing education can take advantage from both tools, as both writing processes and text quality can be assessed more objective and way more easily. Hence, I am in favor of an international note on Héctor Díez Caso’s PhD.

Prof. dr. Huub van den Bergh

External review in support of Héctor Díez Caso's pursuit of a PhD degree with International Mention

Mariëlle Leijten
PhD, Assistant professor
Department of Management
Stadscampus
Prinsstraat 13, C.458
BE 2000 Antwerp

Telephone	32 3 265 5072
E-mail	marielle.leijten@uantwerp.be
Date	25-06-2014

To whom it may concern,

I have been asked to write a review in support of Héctor Díez Caso's pursuit for a PhD degree with International Mention. Therefore I have read two chapters/articles of his PhD in English.

Chapter/Article 1: Analyzing linguistic features of productivity and coherence with Seshat Product

In this article an automated essay scoring tool called Seshat Product is presented. This tool has been designed to analyse

productivity and coherence traits in Spanish written texts. Next to the data the program analysed, I was impressed by the rigour of validating the tool. The fact that Seshat Product has been designed to ease the work of teachers and researchers enabling them to speed up the text correction process are very positive. Researchers/teachers will be in favour of a tool that will provide them with some degree of freedom to connect meaning to the automatically generated scores. On the other hand this article shows the need for on-going research in this field to get closer to a really automated scoring system (for multiple purposes and multiple languages). Although the linguistic features that are selected in Seshat are subject to further decisions by the researchers/teachers, it definitely helps them to select variables that they are interested in in further (semi-automatic and manual) analyses.

Chapter/Article 2: Assessment of cognitive activities in writing: a web based approach

In this article a web-based tool designed for studying cognitive processes involved in writing is presented: Seshat Process. It has been developed as part of a bigger project which includes a whole new suite of applications intended to perform studies from a process-product perspective (as described in article 1). The proposed system was developed to implement Kellogg's triple-

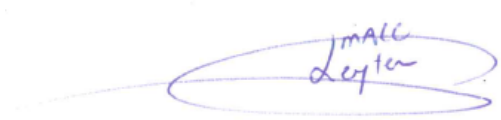
task procedure, perform online analyses of the collected data and create data matrices ready to be used by the most common statistical software. Via a referential case study the system was put to the test as part of an online course in order to both track the participants' cognitive processes and analyze the data collected. Since the tool is web based it would be worthwhile to compare the accuracy of the logging in relation to processes logged off-line (e.g. via programs such as E-prime).

In follow-up studies, I would recommend Héctor Díez Caso to strongly collaborate with other researchers that are developing both keystroke logging tools and automated essay systems. E.g. Scriptlog developed some linguistic measures for Swedish; Inputlog does the same for English and Dutch. The latest version of Scriptlog includes also a triple-task-technique and the data gathered by Scriptlog can be transferred to Inputlog for further analysis. Collaborating with the developers of the other tools will certainly strengthen the theoretical framework of the tools. Also the rigor of data analysis will benefit from this. The data matrices might be suitable to connect to Inputlog/Scriptlog. I assume that the same collaboration perspectives are in place for the developers of AEE systems.

Both articles proved that this application is a promising tool in the fields of writing and psycholinguistic research in Spanish

speaking countries. As such, this kind of research is important for the writing community. As I stated in a book review in the current issue of the Journal of Writing Research: "Concluding, I wish for future researchers to continue developing suitable language dependent and language independent measures that, importantly, can be automatized and made freely available for the research community." The tool Seshat perfectly fits in this perspective. Besides I am very impressed by the fact that Héctor has been able to integrate the theoretical frameworks at hand into a new program that is suitable for Spanish: his multidisciplinary approach is admirable. Therefore, I would like to advise the committee to approve an international mention on Héctor Díez Caso's PhD degree.

Yours Faithfully,



Mariëlle Leijten,

Assistant professor University of Antwerp/Post-doc researcher
Research Foundation – Flanders

Thomas Quinlan
32 Wiggins St.
Princeton, NJ 08540
U.S.A.

Telephone	609 977 6402
E-mail	thquinlan@gmail.com
Date	24-07-2014

To Whom It May Concern:

About one year ago, I received an email from Héctor Díez Caso, with various questions about using computers to assess writing. Since this topic figures prominently in my own research; I welcomed the opportunity to exchange ideas with a fellow scholar. I learned that Mr. Díez Caso had served an internship with Dr. Gert Rijlaarsdam, a highly-respected scholar at the University of Amsterdam. With this introduction, Mr. Díez Caso and I began a lengthy email dialog about possible dissertation topics.

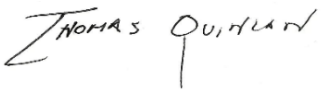
Significantly, Mr. Díez Caso and I have never met in person. All of our interactions have been through email. Somehow, this mode of interaction seems appropriate for two scholars devoted to the study of writing. Through Mr. Díez Caso's written words, I recognized someone with a keen intellect, driven by curiosity; he asked all the right questions. Further, I was impressed with his fluent command of written English.

In Mr. Díez Caso's dissertation, he demonstrates his outstanding scholarly abilities. In this field of research, i.e., using computers to assess writing, scholars often have a strength in one area or the other. That is, they may know a lot about computers and little about writing, or vice versa. Mr. Díez Caso shows strong knowledge of both. He ably summarizes the various cognitive theories of writing, while noting important distinctions between them. When he explains this computer tool, i.e., Seshat Process, his explanation strikes the right balance, neither too general nor too technical.

After designing and programming this computer tool, there arises the question, 'How well does it work?' Accordingly, Mr. Díez Caso designed and conducted a usability study, which yielded evidence to evaluate this question. With some minor revising, I could readily imagine this dissertation published in a peer-reviewed research journal.

It has been a privilege to work with Héctor Díez Caso. I wish him every success in his future endeavors.

Best regards,



Thomas Quinlan, Ph.D.