

El diseño y arquitectura de herramientas para la redacción asistida de textos en inglés

ELENA BÁRCENA

Dpto. Filologías Extranjeras y sus Lingüísticas, UNED

TIMOTHY READ

Dpto. Lenguajes y Sistemas Informáticos, UNED

Recibido: 15-01-07 / Aceptado: 18-04-07

ISSN: 1697-7467

RESUMEN: Este artículo describe las principales líneas y trabajos investigadores llevados a cabo hasta la fecha en el campo de la corrección textual (semi)automática, con especial atención a la labor realizada desde una perspectiva didáctica, esto es, de aprendizaje de lenguas extranjeras. Tras un análisis de las conclusiones obtenidas por los expertos a lo largo de los años sobre las causas del grado de desarrollo de los correctores y su posición relativa con respecto a otras aplicaciones del procesamiento del lenguaje natural, se presenta una nueva herramienta de asistencia en la composición, denominada VAT (Virtual Authoring Tool). Como se desprende de dicha presentación, el VAT no es un corrector en el sentido estricto ya que, debido a su finalidad fundamentalmente pedagógica, no pretende subsanar automáticamente los errores cometidos por el estudiante, sino ayudarle a que él mismo los corrija. Se trata de una herramienta didáctica multifuncional, flexible y sensible al contexto, de ayuda en la composición para estudiantes de inglés de nivel intermedio.

Palabras claves: composición asistida por ordenador, correctores lingüísticos, enseñanza de lenguas asistida por ordenador.

ABSTRACT: This article describes the main research activities undertaken up until now in the field of (semi)automatic textual correction, paying special attention to this work from a didactic perspective, that is to say, learning foreign languages. An analysis is carried out of the research produced in the field, relating the state of development of correctors to other natural language processing applications. Subsequently, a new tool, called the VAT (Virtual Authoring Tool), is presented which is not strictly speaking a corrector, due to its inherent pedagogic nature, because it does not attempt to automatically correct the errors committed by a student. It is a multifunctional didactic tool which provides flexible context sensitive assistance to the student at an intermediate level of the English language, while writing a composition.

Key words: computer textual correction, linguistic correctors, computer assistance for foreign language learning.

1. INTRODUCCIÓN

Existen decenas de clasificaciones sobre los aspectos a destacar en la tarea de revisión de la redacción de un estudiante de una lengua extranjera¹. Ferris y Hedgcock (1998), por ejemplo, identifican niveles que incluyen contenido, organización, uso de lenguaje, mecánica y vocabulario, y Hutchins y Somers (1992), tres: ortografía, gramática y estilo. Aunque la corrección de la totalidad de los errores presentes en la redacción de un estudiante plantea al profesor problemas únicamente de naturaleza práctica, los intentos de automatizar este proceso han demostrado la inviabilidad de producir una herramienta capaz de funcionar adecuadamente en todas estas tareas, si bien sí es posible alcanzar algunas y hacer que la herramienta ayude de alguna manera, modesta pero consistente y fiable, a los estudiantes de la lengua extranjera de que se trate; es decir, aunque no produzca un análisis completo del *input* del usuario, su elaboración y uso quedarán justificados si puede servir para señalar algunos errores superficiales comunes entre los estudiantes.

Debido a la naturaleza integrada de los distintos niveles lingüísticos, cabe prever que una herramienta útil de revisión de la composición abarque indistintamente procesos de corrección ortográfica (como los errores de capitalización), morfológica (como la falta de concordancia personal entre el sujeto y el verbo principal y los errores en la conjugación verbal), sintáctica (como el uso de un pronombre relativo incorrecto y la ausencia de sujeto) y estilística (como la excesiva longitud de una oración y el comienzo de una oración con una conjunción coordinada). Esto, naturalmente, no impide que la arquitectura del corrector esté dividida en módulos siguiendo criterios lingüísticos.

La condiciones esenciales para que la herramienta sea pedagógicamente aceptable son: primero, que no proporcione información errónea, es decir, que no detecte como error algo que no lo es y, segundo, que advierta precisamente al usuario de que sólo cubrirá algunos errores, para que éste no asuma que lo que no se señala como error está necesariamente bien. No sería imprescindible que una herramienta de este tipo tuviera ni conocimiento semántico ni del mundo real para ser útil, porque hay un amplio conjunto de errores que se pueden detectar con un análisis gramatical rudimentario. Tampoco sería un requisito fundamental que tuviera capacidades productivas en lenguaje natural, ya que no sería pedagógicamente aceptable que realizara sustituciones de texto de forma completamente automática (autónoma), sino solamente capacidades de análisis y, como es sabido, las técnicas analíticas se encuentran mucho más avanzadas que las de síntesis o generación, cuestiones. Ambos aspectos reducen ostensiblemente a priori las demandas lingüístico-computacionales para la elaboración de una herramienta automática de corrección o revisión.

2. LA CORRECCIÓN AUTOMÁTICA VS. LA TUTORIZACIÓN PERSONAL

Keobke y otros (1995) encontraron un problema de tipo pedagógico con uno de los tipos de herramientas más básicos de asistencia a la escritura como son los correctores de ortografía que funcionan automáticamente a iniciativa del sistema, porque no permiten que el usuario tenga un recuerdo consciente de qué palabra ha cambiado de forma. Además, como mencionaban los anteriores autores, tanto los que funcionan por propia iniciativa como aqué-

llos que destacan visualmente el error encontrado para que el usuario decida si desea o no modificarlo no aportan razonamiento alguno o éste es muy rudimentario, particularmente para que sea empleado con aprovechamiento por lingüistas. Basándose en estos y otros argumentos concluyen radicalmente que es mejor no usar los correctores y que es un tutor quien debe encargarse de la tarea de la corrección de los errores de sus alumnos. Yao y Warden (1996) añaden que una herramienta debería no solamente ayudar al alumno a corregir sus errores sino a aprender a no producirlos más en el futuro.

Por su parte, Pennington y Brock (1989) habían llevado a cabo años atrás un análisis entre dos grupos de alumnos, uno de los cuales usaba un corrector y el otro, un tutor profesional. Sus resultados fueron que los textos producidos por los alumnos que empleaban el corrector acababan con oraciones más cortas después de menos etapas de revisión que los alumnos que trabajaban con el tutor. Tras este análisis, sin embargo, no parece adecuada su conclusión de que los correctores sean inútiles porque no se puede comparar su actuación con el conocimiento y magisterio de un tutor (Liou, 1993b). Adair-Hauck y otros (2000), sin embargo, establecieron una comparación entre la labor de los correctores y los tutores, destacando que consideraba muy útiles algunas de las funciones de aquéllos por su línea de no intervención, consistente en buscar palabras y partes de frases y dirigir al alumno en su tratamiento (artículos, preposiciones, verbos con partícula, pronombres, vocabulario especial, etc.). Clayton (1995) demostró más adelante que un corrector modificable reducía el número de errores de sus alumnos de ingeniería entre 30% y 50%. Así, los profesores podían concentrar sus esfuerzos docentes en ayudar a sus alumnos con los aspectos más sofisticados y creativos de su redacción.

3. LOS CORRECTORES Y LA LINGÜÍSTICA COMPUTACIONAL

Entre los programas correctores comerciales y no comerciales disponibles se puede ver que hay varias maneras de corregir un texto (Brock, 1990): el acoplamiento de patrones (en el cual se buscan palabras y frases dentro del texto), el análisis estadístico (en el que se hacen cálculos de la longitud de ciertas estructuras o el uso repetido de una palabra dada, para prevenir la composición de textos inferiores estilísticamente) y, el más sofisticado lingüística y computacionalmente, el análisis realizado por un *parser* (Holland y otros, 1993). Los resultados disponibles de estos tipos de programas muestran que podrían servir de gran ayuda a varios tipos de estudiantes en ciertos contextos de uso, aunque ninguno de ellos puede llegar a prescindir por completo del profesor debido a sus limitaciones causadas por las dificultades que conlleva el tratamiento y el análisis automáticos del lenguaje (Bolt, 1992).

Con todo, según autores como Oliva (1995), los correctores automáticos constituyen una de las aplicaciones más plausibles dentro del ámbito de la lingüística computacional. Además de la relativa simplicidad técnica y didáctica con la que se pueden desarrollar herramientas útiles de este tipo en comparación a otro tipo de aplicaciones como los generadores textuales o los traductores automáticos, hay también datos procedentes de estudios que muestran el interés por ellas de los estudiantes de lenguas extranjeras. Brock (1990) ha argüido que los resultados obtenidos en esta línea de investigación y desarrollo son muy variados y que existe una cantidad de estudios que muestran que el alcance y la cantidad de revisión que los estudiantes de lenguas extranjeras llevan a cabo con un texto aumenta cuando trabajan con

un procesador de texto y un corrector, en lugar de un lápiz y papel (van Zaanen y van Huyssteen, 2003; Phinney, 1989). Jacobs y Rodgers (1999) también han argüido a favor de un uso selectivo de los correctores, ya que al escribir con un procesador de textos, el análisis o corrección textual es la segunda actividad más común después de la composición en sí. Kenning (1991) por su parte ha confirmado en un estudio que los estudiantes prefieren los programas de manipulación de texto a otros tipos de programas y de ayudas en línea.

Liou (1991, 1993a, 1993b) ha realizado multitud de estudios sobre correctores y, aunque sus grupos de estudiantes siempre son muy pequeños, los resultados han mostrado que los grupos que trabajan con las herramientas de enseñanza de lenguas asistida por ordenador mejoran más que los grupos de control (Chen, 2000). Healey (1992) ha examinado también programas de este tipo y ha descubierto que la extensión de las reglas existentes supone un gran trabajo para el profesor, aunque concluye que a pesar de que estos programas no encuentren todos los errores, sirven al menos para aumentar el interés del alumno en el problema, lo cual no es poco. Liou (1993a) también ha descubierto que un grupo de alumnos usando herramientas como los correctores era capaz de corregir más errores durante el proceso de modificación de sus textos y cometía menos errores que un grupo sin acceso a ellas.

Al plantearse la pregunta: ¿qué factores en una lengua dificultan el funcionamiento de los correctores?, se pueden distinguir varios. En primer lugar, un problema para los ordenadores es que el significado de un texto está presente en muchos niveles, como explica Peng (1993) y, aunque se puede controlar el nivel más superficial con los correctores de ortografía, existen también errores al nivel estructural de difícil detección y solución. Por ejemplo:

“The doctor could tell that the patient was going to dye/die.”

Una solución sin pasar por un análisis lingüístico profundo sería comprobar con la ayuda de un corpus que combinación de palabras es más frecuente: “patient – dye” or “patient – die”. Lamentablemente, encontrar un corpus representativo es un problema añadido. Además, este tipo de proceso suele ser muy lento y, por lo tanto, caro computacionalmente.

Podría decirse que los principales retos de los correctores incluyen los siguientes (<http://englishplus.com/news/news0299.htm>):

- El significado está presente en todos los niveles y unidades lingüísticas. Incluso la posición de los signos de puntuación y la capitalización o no de una palabra pueden causar distinciones semánticas.
- Incluso la corrección de los errores tipográficos, una tarea fundamentalmente mecánica, puede no llevarse a cabo correctamente por problemas de identificación o incluso se pueden producir errores donde no los había (por ejemplo, corregir “peace” como ortografía errónea de “piece”, ambas palabras existentes en inglés). Pennington (1993) ha sido especialmente crítico con los correctores por indicar errores donde no hay y no indicar errores donde hay.
- Las relaciones gramaticales a veces están entrelazadas con otros fenómenos que las afectan y condicionan. Se trata de mecanismos heterogéneos, lingüísticos y paralingüísticos como, por ejemplo, el sarcasmo o la ironía.
- Hay variantes en la estructura del lenguaje dentro una amplia comunidad lingüística debido a razones de lugar, tema, tipo textual, etc.

- El uso aproximativo, poco explícito y ambiguo que hacemos los hablantes (consciente o inconscientemente).

Otros problemas con los correctores destacados en la literatura son los siguientes (Brock, 1990; Liou, 1993b):

- Suelen producir una dependencia en el usuario, quien quizás pierda seguridad en sí mismo, al estar acostumbrado a contar con la ayuda de la herramienta, e incluso puede que pierda el hábito mecánico de prestar atención a las cuestiones de forma.
- La mayoría de los programas comerciales producen un análisis incorrecto debido a limitaciones fundamentales en el procesamiento del lenguaje natural. Esto es peligroso cuando el usuario no es plenamente competente en la lengua pertinente, ya que seguramente aceptará las sugerencias incorrectas del programa.
- Los correctores suelen seguir un enfoque orientado hacia un producto o tarea específica y operar a nivel de la cadena lingüística superficial.
- Los correctores suelen estar preparados para un sub-grupo de usuarios, como los empresarios o escritores técnicos y en su afán de simplificar las explicaciones, suelen contener serias incorrecciones teóricas.
- Los profesores no suelen poder modificar los parámetros de operación de los programas.

Al analizar la simplicidad técnica que caracteriza a la mayoría de los correctores existentes, se constata que apenas incorporan las técnicas de ingeniería lingüística disponibles. Se plantea entonces la siguiente pregunta: ¿qué grado de sofisticación lingüística debe tener un módulo de este tipo? Lo evidente es que, por ahora, lo que funciona de un modo estable es un sistema de ayuda conectado a un editor de texto. Aunque se suele decidir no aspirar a una gran sofisticación computacional para la elaboración de la primera versión de un sistema de este tipo, sí se considera deseable que aumenten sus capacidades y el tipo de procesamiento subyacente para futuras versiones, como se puede ver claramente en un par de ejemplos de conocidos argumentos:

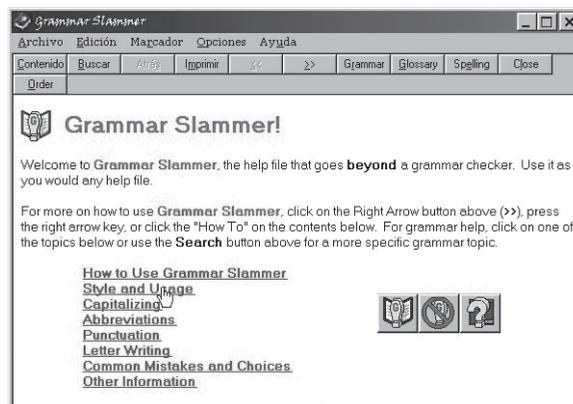
- Según Nagata (1996), la motivación para el desarrollo de sistemas inteligentes (ICALI) (*Intelligent Computer-Assisted Language Instruction*) es analizar el input de los alumnos para poder facilitar un análisis de errores más sofisticado y la consecuente generación de retroalimentación para el usuario. La indicación de un fallo en alguna parte de la oración es fácil para un sistema CALI (*Computer-Assisted Language Instruction*) (basándose en el acoplamiento de patrones) pero la ventaja de los sistemas ICALI es la capacidad de deducir qué fragmentos de lenguaje son realmente errores y de generar mensajes útiles sobre su naturaleza (Juozulynas, 1994).
- Golding y Roth (2001) han mostrado que los errores de ortografía detectables por el contexto forman entre el 25% y el 50% de todos los errores de este tipo (Kukrich, 1992), por ejemplo: "I'd like a peace/piece of cake, please.". Los programas sofisticados de comprobación de ortografía como el de estos autores pueden identificar muchos de estos errores como, por ejemplo, los siguientes:

accept,	cite, sight,	its, it's	peace, piece	than, then
except	site	lead, led	principal,	their, there,
affect,	country,	maybe	principle	they're
effect	county	may be	quiet, quite	weather,
among,	fewer, less	passed,	raise, rise	whether
between	I, me	past		your, you're
amount,				
number				
begin, being				

4. UNA MUESTRA DE CORRECTORES DISPONIBLES

Según el autor de <http://englishplus.com/news/news0299.htm>, de todos los programas disponibles para tareas de corrección, los más útiles (en cuanto a su cobertura y a no cometer errores) son los que combinan un sencillo editor de textos y un archivo de ayuda relacionado con los problemas más comunes de los usuarios. Véanse un par de ejemplos. En primer lugar, el programa Grammar Slammer que aparece en la figura 1. Este “programa” consiste en un archivo en formato de Ayuda de Windows (y, por lo tanto, sólo funcionará en la plataforma Windows) que permite al usuario encontrar ayuda sobre los problemas comunes de la gramática inglesa.

Figura 1. La interfaz del programa Grammar Slammer



En este archivo de ayuda se cubre un gran número de puntos gramaticales que suelen causar dificultades a los estudiantes. Se puede ver una lista de los temas tratados bajo cada sección a continuación:

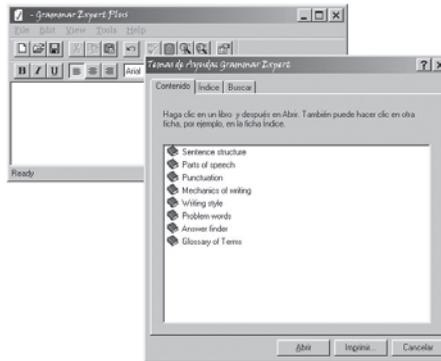
Section	Contents		
Style and Usage Contents	Sentence Fragments Run-on Sentences Dangling Modifiers Misplaced Modifiers Pronoun Case Possessive Pronouns Pronouns with Than and As	Subject-Verb Agreement Indefinite Pronouns Pronouns Ending in -self Pronoun-Antecedent Problems The Subjunctive ...	Comparisons Irregular Comparisons Comparison Problems Negatives Tricky Plurals The Verb To Be
Rules for Capitalization Contents	Capitalizing Sentences Capitalizing Quotations Special Cases for Capitalizing	Capitalizing Proper Nouns Names Not Capitalized Capitalizing Proper Adjectives	Capitalizing Personal Titles Capitalizing Titles of Things Capitalizing in Letters Scientific Nomenclature
Rules for Abbreviations Contents	Abbreviated Names and Social Titles Abbreviated Rank and Academic Titles Abbreviations After a Name Abbreviations of Geographical Features State and Province	Abbreviations Abbreviations of Units of ... Measure Abbreviations of Units of Time Common Latin Abbreviations Abbreviations in Names of	Businesses Acronyms and Pronounced Abbreviations Scientific Nomenclature
Punctuation Rules	Periods Question Marks Exclamation Points Commas Semicolons and Colons	Quotation Marks Italicizing and Underlining Dashes and Parentheses Hyphens	Apostrophes The Ellipsis Brackets The Virgule
Commas	The Three Most Common Comma Rules Commas in Compound Sentences Commas in a Series Commas with Paired Adjectives Commas and Introductory Words Commas After Introductory Phrases Commas After Introductory Clauses Commas with Interrupting Expressions Commas with Nonrestrictive Modifiers	Commas with Geographical Names Commas with Dates Commas with Titles that Follow Names Commas in Addresses Commas in Letter Writing Commas in Numbers Commas with Certain Words Omitted	Commas with Quotations Adding Commas for Clarity Commas with Adjectives Following Nouns When Not to Use Commas
Semicolons and Colons	Semicolons with Clauses Semicolons in a Series Colons with Lists Colons Before Quotations	Colons Separating Independent Clauses	Colons with Appositives Special Cases Using Colons
Quotation Marks	Quotation Marks in Direct Quotations Question Marks and Exclamation Points in Quotations	Other Punctuation Marks with Quotation Marks Quotation Marks in Dialogue Quotation Marks in Titles	Quotation Marks with Slang Definitions in Quotation ... Marks Single Quotation Marks
Italicizing and Underlining	Underlining Titles Underlining Names Underlining Foreign Words or Abbreviations	Underlining Words for Emphasis Underlining Items Which Name Themselves	Titles Which Take No Punctuation Scientific Nomenclature
Dashes	Using Dashes	Using Parentheses	Numbers and Letters

Section	Contents		
and Parentheses	Dashes with Nonrestrictive... Modifiers ...	Parentheses with Certain	Punctuation Inside Parentheses
® Hyphens	Numbers Written Out with Hyphens	Hyphenated Prefixes / Suffixes Hyphenated Compound Words	Hyphens for Clarity Dividing at the End of a Line
® Apostrophes	Apostrophes Showing Possession Plural Possessives	Possessives with More than One Owner Apostrophes with Underlined or Italicized Items	Apostrophes with Verb Contractions Apostrophes with Other Contractions
Letter Writing Rules	Business Letters Business Letter Styles Friendly Letters	Friendly Letter Format Envelopes	Envelope Format Folding a Standard Letter
Common Mistakes and Tricky Choices	A A/An Absolute Modifiers Accept/Except Accuse/Allege Across/Acrossed/Cross Adapt/Adopt Administer/Administrate Aggravate/Irritate Ain't A hold/Ahold A lot/Allot/Alot Already/All Ready Alright/All Right Altogether/All Together Always/All Ways American Grammar vs. British Grammar Among/Between Amount/Number Ante-/Anti- Anxious/Eager Anymore Anyone/Any One Anyway/Any Way Anyways Anywheres Apart/A Part As/Like As To At after Where A Ways Awhile/A While B Bad/Badly Because after Reason Being As or Being That Beside/Besides Between/Among Bring/Take British Grammar vs.	H Hanged/Hung Have or Had plus Ought Have after Could, Would, Should, or Will Healthful/Healthy Height or Heighth Hers/Her's Historic/Historical Hopefully Hypo-/Hyper- Hypocritical/Hypercritical I I Hope/Hopefully Immigrate/Emigrate Imply/Infer In after Want Incredible/Incredulous Indeterminate/Indeterminable Irregardless/Regardless Irritate/Aggravate Its/It's J Jiggle/Joggle/Juggle Judicious/Judicial/Juridical Just, Use of K Kind Of, Use of L Lay/Lie Leave/Let Lend/Loan Less/Fewer Less/Littler (More Little) Like/As Lie/Lay Littlest/Least Loath/Loathe	Q Quash/Squash Quote/Quotation/Quotation Mark R Raise/Rise Real/Really Reason with Because Regardless/Irregardless Renown/Reknow S Says/Said Seen/Saw Sensual/Sensuous Set/Sit Should Have/Should Of Solid/Stolid Somewheres Sort Of Special(ly)/Especial(ly) Squash/Quash Supposed to/Suppose to Sure/Surely T Take/Bring Tenet/Tenant Than/Then That/Where That There and This Here That/Which/Who Them/Those There/Their/They're There's/Theirs/Their's To/Too/Two Tortuous/Torturous/Tortured Try And/Try To Turbid/Turgid U

Section	Contents		
	American Grammar Blond/Blonde Burst/Bust/Busted	Luxuriant/Luxurious M	Unequivocal/Unequivocable Uninterested/Disinterested Unique (and Other Absolute Modifiers) Use To/Used To
	C Can/May Can't Help But Clipped Words Compose/Comprise Continual/Continuous Could Have/Could Of Cross/Across/Acrossed Cult/Occult	Macro-/Micro- Magnificent/Munificent Manic/Maniac May/Can Maybe/May Be Morale/Moral N Nowheres Number/Amount	W Want followed by In, Out, Off, Down, or Up Warranty/Warrantee/Warrant Ways after A Well/Good Went/Gone Where Followed by At Where/That Who/Which/That Who/Whom Who's/Whose Will Have/Will Of -Wise (Suffix) Would Have/Would Of
	D Different From/Different Than Diffuse/Defuse Disinterested/Uninterested Done as Verb Doesn't/Don't Due To Due To the Fact That	O Occult/Cult Of, Use of Of after Would, Could, Should, or Will Official/Officious Orient/Oriente Only, Use of Ought with Have or Had	
	E Economic/Economical Emigrate/Immigrate Enthused/Enthusiastic Equivocal/Unequivocable Especial(ly)/Special(ly) Everyone/Every One Everywheres Exalt/Exult Except/Accept	P Parameter/Perimeter Perimeter/Parameter Persecute/Prosecute Perspective/Prospective Prescribe/Proscribe Proscribe/Prescribe Prosecute/Persecute Prospective/Perspective	
	F Farther/Further Fewer/Less For Free Fortunate/Fortuitous		
	G Gone/Went Good/Well		

Este programa es realmente un sistema de ayuda hipertextual con el que un usuario puede consultar sus dudas gramaticales mientras escribe. Otro ejemplo de un programa de este tipo, aunque algo más sofisticado, es el Grammar Expert Pro que, además de ofrecer un sistema de ayuda hipertextual, tiene un sencillo editor de textos donde un usuario puede preparar sus composiciones. Se puede ver la interfaz del programa en la figura 2 a continuación:

Figura 2. La interfaz del programa Grammar Expert Pro.



Al igual que Grammar Slammer, Grammar Expert Pro ofrece una lista de las secciones de ayuda que se incluye a continuación:

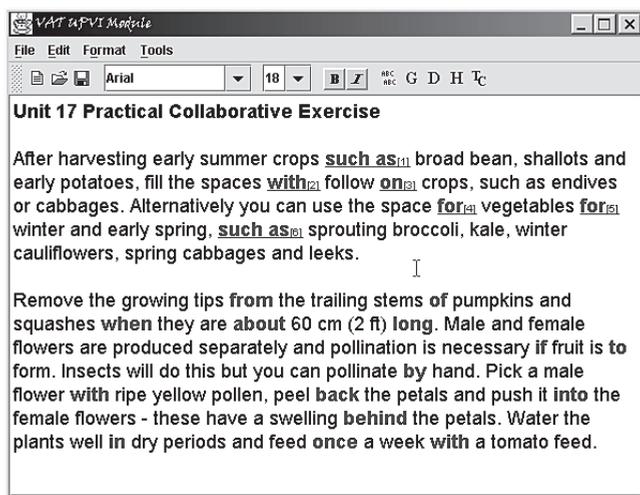
Section	Contents		
Sentence structure	Clauses Dependent clauses Independent clauses	Objects Paragraphs Phrases	Predicates Subjects
Parts of speech	Adjectives Adverbs Articles	Conjunctions Interjections Nouns	Prepositions Pronouns Verbs
Punctuation	Apostrophe Colon Commas: Basic comma use Using commas to: ... separate adjectives ... prevent ambiguity ... isolate conjunctive adverbs	... separate dependent and independent clauses ... isolate descriptive clauses ... separate independent clauses ... join interdependent clauses ... isolate introductory phrases ... separate items in a series ... substitute for omitted words ... isolate parenthetical phrases	Comma splice error Dashes Ellipses Exclamations Parentheses Periods Questions Quotation marks Semicolons Slashes
Mechanics of writing	Abbreviations ... Acronyms ... Contractions ... time and date references ... Initialisms ... of Latin phrases ... for measures ... organization names	... Plural abbreviations ... Possessive abbreviations ... Abbreviating degrees and titles Capitalization Dates and times Hyphenating words Hyphenating ... adverbs and adjectives	... compound adjectives ... compound nouns ... numbers ... prefixes Using hyphens to break long words Lists Numbers Titles

Section	Contents		
Writing style	Clarity and directness ...Clarity and directness ...Avoiding ambiguity ...Stating the subject explicitly Choosing words ...Archaic words ...Awkward words and phrases ...Classified singulars ...Confusing terms ...Double negatives ...Jargon	...Qualified absolutes ...Redundant terms ...Stilted terms ...Vague terms ...Wordy phrases ...Gender-specific terms Effective sentences ...Sentence ...beginnings ...endings ...length	Writing letters ...Letter format ...Complimentary closing ...Letter salutations Organized writing Writing ...to inform ...to persuade ...effective paragraphs ...Reviewing
Problem words	Ability, Capacity Accept, Except Adapt, Adopt Advise, Advise Affect, Effect All ready, Already All together, Altogether Allude, Elude Allusion, Illusion Alternatively, Alternately Alumna, Alumnae, Alumni, Alumnus Among, Between Amoral, Immoral Amount, Number Anxious, Eager Appraise, Apprise As, Like Ascent, Assent Assume, Presume Beside, Besides Biannual, Biennial Blatant, Flagrant Born, Borne Breach, Breech Can, May Canvas, Canvass Cite, Sight, Site Coarse, Course	Complement, Compliment Comprise, Compose Console, Consul, Counsel, Council Continuous, Continual Convince, Persuade Credible, Credulous, Creditable Criteria, Criterion Decisive, Incisive Deduce, Deduct Definite, Definitive Delusion, Illusion Discreet, Discrete Disinterested, Uninterested Elicit, Illicit Elusive, Illusive Emigrate, Immigrate Eminent, Imminent Ensure, Insure Envelop, Envelope Envious, Jealous Explicit, Implicit Farther, Further Few, Less Flair, Flare Flaunt, Flout Hoard, Horde	Imply, Infer Incredible, Incredulous Ingenious, Ingenuous Its, It's Lay, lie Literally Loath, Loathe Memoranda, Memorandum Ordinance, Ordinance Passed, Past Peak, Peek Peal, Peel Pedal, Peddle Principal, Principle Prostate, Prostrate Ravage, Ravish Sensual, Sensuous Stationary, Stationery Strategic, Tactical Straight, Strait Their, There, They're To, Too, Two Weather, Whether Whose, Who's Your, You're
Answer finder	Help with ... punctuation ...writing personal titles	...choosing the right word ...choosing between confusing words	... I and me, him and he, etc.

5. VAT

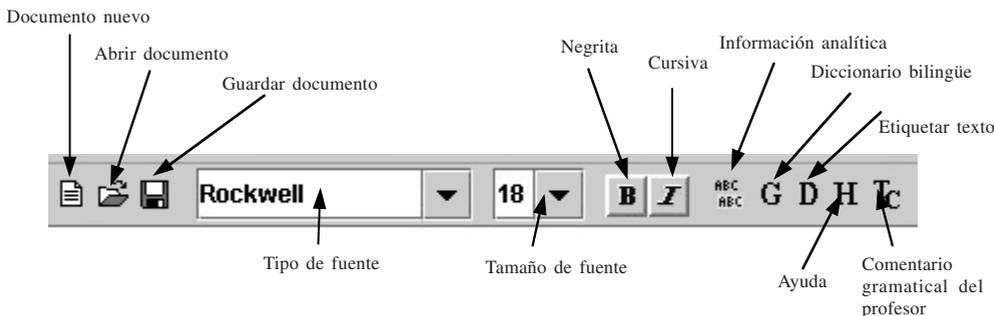
Es evidente que la aplicación didáctica de los correctores automáticos ha recibido bastante crítica, pero la comunidad de investigadores y profesores sigue trabajando en el campo porque existe verdaderamente un problema que solucionar: la persistencia en la producción de errores lingüísticos al escribir, sobre todo (aunque no sólo) por parte de los estudiantes de un idioma secundario. Después de un extenso proceso de análisis de la literatura sobre los correctores se ha diseñado el VAT cuyo interfaz se anticipa en la figura 3 a continuación:

Figura 3. La interfaz del VAT.



Antes de explicar la utilidad y el funcionamiento del VAT, véanse sus funciones principales a través de la barra de herramientas etiquetada, en la figura 4 a continuación:

Figura 4. Barra de herramientas del VAT



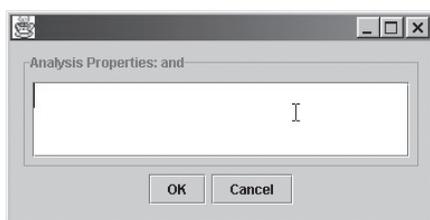
Además de las funciones normales de un editor de texto (tales como abrir y guardar documentos, en los formatos RTF y HTML, seleccionar un tipo, tamaño y estilo de fuente, etc.), son destacables los cuatro botones a la derecha de la barra para efectuar lo siguiente:

- añadir comentarios de análisis gramatical a las palabras o frases,
- acceder a la información gramatical y léxica contenida,
- añadir comentarios especiales (en una fuente no disponible normalmente, en color rojo) o simplemente marcar porciones del texto (en rojo),
- destacar todas las ocurrencias en un texto de una categoría o sub-categoría gramatical a petición del usuario.

A continuación se describe cada función en más detalle:

Primero, la función de añadir comentarios de análisis gramatical a las palabras o frases. Al pinchar en este botón se abrirá una ventana donde el usuario puede apuntar información analítica sobre la palabra (o frase) pinchada en el texto, como se puede ver en el ejemplo en la figura a continuación:

Figura 5. Ventana de información analítica para la palabra “and”.

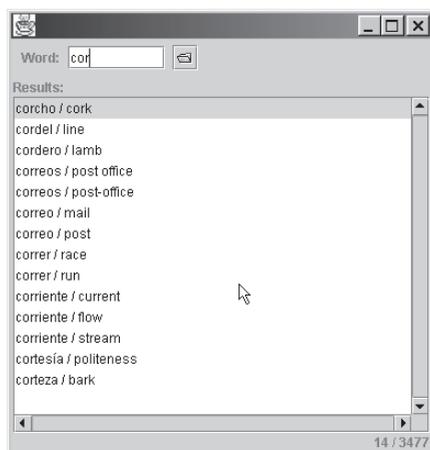


La información que el usuario puede anotar en esta ventana no se introduce en el texto, sino que es guardada por el programa (la palabra en cuestión está subrayada por el programa junto con un número de identificación, como en el ejemplo de la palabra “and” dentro de la oración: “The boys and[1] girls”). El usuario puede acceder a su análisis simplemente pinchando sobre la palabra y este botón otra vez. Si el archivo se convierte al formato HTML, la información de los análisis será incluida como notas a pie de página. Las palabras correspondientes tienen el mismo aspecto visual, pero el subrayado es en realidad un enlace, en el que si pincha el usuario, se le llevará a la nota a pie de página correspondiente.

Esta función es muy útil para la realización de los ejercicios de tipo ensayo y también para la preparación de textos propios por parte de un usuario dubitativo, al que las anotaciones sobre un punto del que no está seguro le pueden servir para volver luego sobre ellas al finalizar su composición.

Segundo, la función de acceder a la ayuda léxica y gramatical. El primer tipo de ayuda consiste en un diccionario bilingüe rudimentario en el sentido de que es unidireccional, sólo traduce palabras aisladas y no permite la adición de nuevas palabras, y pequeño (sólo contiene 3.477 palabras). Sin embargo, puede constituir por ahora cierta ayuda inmediata al estudiante hispanohablante que no se acuerda del equivalente inglés de lo que está intentando expresar en un momento dado:

Figura 6. Diccionario bilingüe español-inglés.



Al pinchar en este botón se accederá al sistema de ayuda gramatical. Este sistema no está en el formato de archivos de ayuda de Windows (como los de los programas Grammar Slammer y Grammar Expert Pro) sino del formato HTML (para ser transportable a otros sistemas operativos como los mismos módulos del UPVI en Java).

La gramática se divide en tres partes que corresponden a tres unidades lingüísticas fundamentales: el texto, la oración y la palabra. En cada parte se encuentran todos los principios y elementos que se detectan y estudian mejor dentro de la unidad lingüística correspondiente. Veamos algunos ejemplos. El nivel textual incluye las cuestiones pragmáticas de estructura retórica y los conectores supraoracionales. El nivel oracional trata cuestiones como las estructuras sintácticas que destacan la información nueva y los signos de puntuación. El nivel de la palabra cubre los niveles léxico y semántico, tratando todas las categorías sintácticas a excepción de la conjunción. Aunque las conjunciones constan a menudo de una sola palabra, su función no es referencial (no tienen apenas contenido semántico) sino sintáctica (son transpositoras de oraciones) y es, por lo tanto, preferible estudiarlas a nivel oracional.

Aunque hasta la fecha el sistema piloto del VAT, que ha sido ya finalizado, consta de un número reducido de reglas, la siguiente versión en la que estamos trabajando está basada en un menú que contiene una larguísima jerarquía de puntos gramaticales. Por esto, dada la gran extensión de la sección de gramática que cubre el sistema, recomendamos al estudiante que tenga a su lado una copia impresa del índice de *The Virtual Authoring Tool* durante las primeras sesiones, hasta que se familiarice con los contenidos de la gramática y la localización de los mismos. Si no lo hace, es posible que no sepa cómo ni dónde solicitar ayuda para un tema determinado, y ni siquiera si el sistema lo contiene, corriendo el riesgo de acabar haciendo un uso repetitivo, pobre y breve del VAT, sin explotar todo el potencial de este sistema y, por tanto, reduciendo lamentablemente las posibilidades de aprendizaje de la composición de ensayos en inglés que le ofrece. Se pueden ver los tres niveles y las entradas y sub-entradas para cada nivel a continuación:

Analysis Level	Menu	Sub-Menu	Analysis Level	Menu	Sub-Menu
Text Level	Introduction and background Arguments and counterarguments Giving examples Conclusion Narration Description Dialogue		Word Level	Determinatives	Articles
					Possessives
					Demonstratives
					Interrogatives
					Indefinites
Sentence Level	Structures Coordinative Relative Nominal				Relatives
				Exclamatives	
	Adverbial	Time Place Manner Reason Purpose		Predeterminatives	
	Condition Comparison			Postdeterminatives	
	Punctuation	Punctuation marks Capitals Italics		Pronouns	Personal Reflexives Reciprocals Possessives Demonstratives Indefinites Relatives Interrogatives
	Constituent order Constituent presence			Adverbs	Time Location Manner Quantity Assertion Negation Doubt Intensification
				Auxiliaries	Be Have Do Modals
				Prepositions	

Cuando el estudiante haya utilizado el sistema varias veces, sin duda habrá adquirido la suficiente intuición o familiaridad para saber dónde dirigirse. Naturalmente, tenemos previsto en futuras versiones del VAT aumentar la sofisticación de la interfaz, de tal modo que no esté basada en un menú, sino en el lenguaje natural, y el estudiante pueda teclear su duda concreta directamente al sistema y sea éste quien se ocupe de las tareas de interpretarla y buscarla en su base de conocimiento.

El orden de presentación de los tres niveles de la gramática del VAT revela en cierta medida el orden de importancia que otorgamos a cada uno de ellos. Teniendo en cuenta factores como el hecho de que se trata de una herramienta de asistencia en la composición y el nivel de conocimiento del inglés que se le supone al estudiante medio, consideramos que es el nivel textual el que más le ayudará en la redacción de su ensayo y en el que más recursos expresivos nuevos aprenderá. Los niveles de la oración y la palabra son fundamentalmente de apoyo, aunque hay que distinguir en ellos las cuestiones meramente formales (por ejemplo, la posición de un modificador concreto con respecto al nombre que acompaña) de aquéllas en las que forma y función van estrechamente relacionadas, siendo la cadena lingüística superficial determinada por la función pragmática o comunicativa del mensaje. Aunque ambos aspectos lingüísticos son necesarios para producir un texto correcto en inglés, no cabe duda del especial interés que conlleva para nuestros estudiantes el aprender las distintas maneras posibles que hay de manipular el lenguaje en el acto comunicativo escrito con el fin de expresar un mensaje con todas sus implicaciones, matices intencionales, etc.

Tercero, la función de añadir comentarios especiales por parte del profesor. Este botón está pensado solamente para el profesor que desea escribir algún comentario o quiere marcar una parte del texto escrito por un alumno. Así, un alumno puede enviar fácilmente un ejercicio o texto que ha hecho al profesor (por correo electrónico o a través de la Web y el profesor puede corregirlo y devolvérselo).

Cuarto y último, la función de destacar todas las ocurrencias en un texto de una categoría o sub-categoría gramatical a petición del usuario. Veamos cómo funciona. El estudiante ha de componer su texto normalmente en el editor de la herramienta y recomendamos que no solicite la ayuda gramatical hasta que lo haya finalizado. Existe evidencia científica de la existencia diferenciada de los niveles lingüísticos en la mente del hablante, además de las limitaciones plásticas de nuestra capacidad de atención, por lo que recomendamos al estudiante que, tras terminar el ensayo, repase una vez su contenido semántico y a continuación, todas las cuestiones de forma (puntuación, capitalización, orden de los constituyentes, morfología, etc.). Una vez finalizado el proceso de autocorrección es cuando el estudiante debe solicitar ayuda gramatical al VAT, si lo considera necesario.

Al pinchar en este botón aparecerá un menú despegable que ofrece los tres niveles de análisis donde el usuario debe elegir la categoría lingüística que desea comprobar. Al seleccionar una categoría, el programa cambia el color del fondo de todas las palabras en el texto de esta categoría al color amarillo para facilitar su tratamiento por parte del usuario y abre una ventana de información gramatical sobre esta categoría para ayudar al usuario a comprobar si la está usando correctamente.

Una vez que el estudiante ha seleccionado con el ratón un aspecto gramatical sobre el que desea saber más, el sistema le muestra un pequeño texto explicativo sobre dicho punto, acompañado de ejemplos. Además, el VAT señala todas las palabras y signos del ensayo relacionados con la norma consultada para que, tras leerla detenidamente, el estudiante pueda

dirigirse fácilmente y sin pérdida de tiempo a las unidades señaladas para comprobar si siguen la norma, o en su defecto, hacer las oportunas modificaciones.

La versión actual del VAT identifica los fragmentos del ensayo que corresponden al principio lingüístico consultado por el estudiante, a través de un simple emparejamiento de palabras. El sistema cuenta con una base de datos léxica para cada principio lingüístico. Por ejemplo, si el estudiante ha consultado los artículos, el sistema sabe que se refiere a las palabras “a(n)” y “the”, las cuales destacará ópticamente en el ensayo para que el estudiante compruebe con facilidad su seguimiento de la norma. Esto quiere decir que, en su actual versión, el sistema puede detectar en el texto todos los principios lingüísticos que corresponden a una sola palabra o signo. Esto incluye la mayoría de los miembros integrantes de las categorías sintácticas cerradas (determinantes, pronombres, preposiciones, conjunciones, algunos tipos de adverbio y signos ortográficos).

Este procedimiento es rudimentario desde el punto de vista de la Lingüística Computacional y tiene un considerable número de limitaciones en su funcionamiento. Siguiendo con el mismo sencillo ejemplo, el VAT puede ayudar a identificar visualmente los artículos en su uso específico, determinado e indeterminado, pero no en su uso genérico, que en ocasiones no tiene un marcador (un punto en el que los estudiantes nativos de español suelen errar por la falta de equivalencia biunívoca con su lengua). El VAT tampoco puede, en su versión actual, detectar las conjunciones con correlato en otra cláusula, como por ejemplo “although ... nevertheless”, ni identificar las palabras multifuncionales, como es el caso de varios determinantes y pronombres, también de algunos adverbios, conjunciones y preposiciones.

En nuestro afán de hallar vías explicativas lo más pedagógicas y prácticas posible, la gramática del VAT no se adscribe a ninguna corriente ni escuela lingüística, sino que es deliberadamente ecléctica. Así pues, hemos elaborado las sub-clasificaciones categoriales que nos han parecido más útiles en el contexto de nuestra herramienta. Además, aunque por ejemplo aparezcan términos afines a la gramática chomskiana como “orden canónico”, etc., nuestro propósito no ha sido otro que enfocar cada fenómeno lingüístico del modo que nos parece más efectivo para su asimilación; en el ejemplo dado, estudiar la relativa flexibilidad del inglés en cuanto al orden sintáctico obedeciendo criterios de tipo pragmático (un concepto, por otra parte, de corte claramente funcionalista). Por otra parte, dada la diversidad de enfoques y la falta de estandarización terminológica existentes en el campo de la Lingüística, cualquier interpretación o uso idiosincrásico por razones metodológicas, aparece debidamente justificado en el apartado correspondiente del menú.

El VAT está diseñado para funcionar interactivamente con el usuario, con la doble intención de aumentar su eficacia didáctica y operativa. Para ello, proporciona al estudiante-usuario el conocimiento que estima necesario para que él mismo lleve a cabo las correcciones de su propio texto si lo desea. Por ahora, el VAT presenta los puntos que sabe que entrañan mayores dificultades al estudiante medio de inglés, para que él mismo compruebe su corrección en el texto personalmente si así lo desea. La gramática del VAT está expresada en inglés, decisión generalizada entre los lingüistas que en el tema del aprendizaje de una segunda lengua defienden el concepto de inmersión lingüística y la función metalingüística de las propias explicaciones. Cada punto lingüístico va, además, complementado por varios ejemplos, cuya importancia se considera fundamental para la comprensión, asimilación y retención del concepto en la memoria.

El nivel y los contenidos de que iba a constar la gramática del sistema se ha establecido siguiendo una metodología empírica a partir de un corpus de ensayos elaborados por estudiantes de Filología Inglesa de la UNED en las pruebas de evaluación a distancia y presenciales de dos asignaturas de primer ciclo. Hemos seleccionado los fallos más recurrentes con el fin de incluir las normas correspondientes en la gramática del sistema, desechando aquellos principios y fenómenos que la mayoría de los estudiantes parecía tener asimilados. Evidentemente, no todos los estudiantes de un mismo curso tienen un nivel de conocimiento similar, por lo que hemos optado por diseñar el sistema en términos de un perfil de usuario, que podríamos calificar como prototípico.

Como se ha comentado anteriormente, la gramática del VAT se centra en el inglés de los ensayos académicos e incluye, por ejemplo, información sobre la expresión en inglés de la autoría literaria y artística, aunque también compara expresiones y normas propias de los ensayos académicos y de otros tipos textuales, como los de tipo técnico. En efecto, a veces el sistema apoya sus explicaciones haciendo uso de recursos comparativos o contrastivos (en especial, el contraste ocasional del inglés de los textos literarios, académicos y técnicos pretende ser útil al licenciado de Filología Inglesa en su heterogéneo futuro profesional). Debe mencionarse que, debido a la propia historia y al uso actual de la lengua inglesa en todo el mundo, el VAT no ofrece en ningún caso reglas absolutas que cubran todas las posibles ocurrencias – y así se le indica al estudiante en el botón “Read me” –, sino que proporciona algunos de los usos correctos y corrientes de cada aspecto lingüístico. Por ello, si el estudiante tiene absoluta certeza de la existencia de un uso determinado en el subgénero de los ensayos académicos, no ha de desecharlo a pesar de que no esté incluido en el sistema.

Por último, el VAT ha sido diseñado para trabajar con archivos en formato HTML, es decir, el mismo formato que las páginas Web. Por lo tanto, se pueden distribuir los archivos producidos con el VAT por la Web y verlos en cualquier navegador Web, como Netscape Navigator o Internet Explorer. El hecho de que el VAT emplee este formato es muy útil porque así se pueden compartir archivos producidos con el VAT con usuarios que no disponen de tal herramienta. Además, se puede trabajar con documentos en formato RTF (*Rich Text Format*), lo cual es muy útil a la hora de transferir el documento a un procesador de textos normal o a un programa de correo electrónico.

6. PRÓXIMOS PASOS

El VAT no es un corrector en el sentido estricto ya que, debido a su finalidad fundamentalmente pedagógica, no pretende subsanar automáticamente los errores cometidos por el estudiante, sino ayudarle a que él mismo los corrija. Al objeto de lograr una adaptación y flexibilidad máximas del sistema a las necesidades y requisitos de los distintos estudiantes, se ha optado de momento por que sean ellos, y no el sistema, quienes lleven la iniciativa de la interacción, seleccionando sólo los puntos que ellos consideren oportunos. Se ha previsto, sin embargo, replantearnos esta cuestión para futuras versiones del VAT y quizás optar por un sistema de los llamados de “iniciativa mixta”, es decir, en los que el usuario y el sistema puedan intervenir espontáneamente, si el primero así lo desea.

Además, como se ha explicado en la sección anterior, la versión actual del VAT es piloto y, por lo tanto, su cobertura limitada. El VAT no puede, en su versión actual, detectar las

conjunciones con correlato en otra cláusula, como por ejemplo “although ... nevertheless”, ni identificar las palabras multifuncionales, como es el caso de varios determinantes y pronombres, también de algunos adverbios, conjunciones y preposiciones. Para ampliar la cobertura del sistema y que incluya, por ejemplo, las categorías abiertas, o que realice desambiguaciones léxicas, incorporaremos un etiquetador que asigne una categoría sintáctica a cada palabra en su contexto oracional (ya existen varios etiquetadores disponibles para inglés con aceptables niveles de éxito). Por último, hay cuestiones más complejas, como las de orden estructural, en las que interviene la función sintáctica, para cuya detección será indispensable integrar un analizador o *parser* en el sistema. Contar con más información estructural sobre el texto le permitirá no sólo ampliar la gama de errores que puede identificar, sino también ser más exacto en la clasificación del tipo de error de que se trata.

7. REFERENCIAS

- Adair-Hauck, B., Willingham-McLain, L., Youngs, B.E. (2000). “Evaluating the Integration of Technology and Second Language Learning”. *CALICO Journal*, 17, 2: 269-306.
- Bolt, P. (1992). “An evaluation of grammar-checking programs as self-help learning aids for learners of English as a foreign language”. *Computer Assisted Language Learning*, 4, 3: 183-189.
- Brock, M.N. (1990). “Customizing a computerized text analyser for ESL writers: Cost versus gain”. *CALICO Journal*, 8: 51-60.
- Chen, J.F. (2000). “Computer Generated Error Feedback and Writing Process: A Link”. *TESL-EJ* 2, 3 (enero). <http://www-writing.berkeley.edu:16080/TESI-EJ/ej07/a1.html>.
- Clayton, T. (1995). “Using a Customized Grammar Checker to Improve Writing and Enhance Self-Directed Learning”. *Proceedings of the Fourth International Literacy Education Research Network Conference*. Townsville, Australia: 241-254.
- Ferris, D. y Hedgcock, J.S. (1998). *Teaching ESL Composition: Purpose, Process, and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates
- Golding, A.R. y Roth, D. (2001). “A Window-Based Approach to Context-Sensitive Spelling Correction”. *Machine Learning (Special Issue on Natural Language Learning)*, 34: 107-130.
- Healey, D. (1992). “Where’s the beef? Grammatical practice with computers”. *CAELL Journal*, 3, 1: 10-16.
- Holland, M., Maisano, R. Alderks, C. y Martin, J. (1993). “Parsers in Tutors: What Are They Good For?” *CALICO Journal*, 11, 1: 28-46.
- Hutchins, J. y Somers, H.L. (1992). *An Introduction to Machine Translation*. CUP.
- Jacobs, G. y Rodgers, C. (1999). “Treacherous Allies: Foreign Language Grammar Checkers”. *CALICO Journal*, 16, 4: 509-529.
- Juozulynas, V. (1994). “Errors in the Composition of Second Year German Students: An Empirical Study for Parse-Based ICAL”. *CALICO Journal*, 12, 1: 5-17.
- Keobke, K.N., Chiu, K. y Talbot, D. (1995). *Dictionaries and computer spelling programs: implications of substitution*. Hong Kong: División of Language Studies.
- Kenning, M. (1991). “CALL evaluation the learner’s view”. *Computer Assisted Language Learning*, 4, 1: 21-27.

- Kukrich, K. (1992). "Techniques for automatically correcting words in texts". *ACM Computing Surveys*, 24, 4: 377-439.
- Liou, H. (1991). "Development of an English grammar checker a progress report". *CALICO Journal*, 9, 2: 57-70.
- Liou, H. (1993a). "Integrating text-analysis programs into classroom writing revision". *CAELL Journal*, 4, 1: 21-27.
- Liou, H. (1993b). "Investigation of Using Text-Critiquing Programs in a Process-Oriented Writing Class". *CALICO Journal*, 9, 2: 57-70.
- McEnery, T., Baker, J.P. y Wilson, A. (1995). "A statistical analysis of corpus based computer versus traditional human teaching methods of part of speech analysis". *Computer Assisted Language Learning*, 8: 259-274.
- Nagata, N. (1996). "An Effective Application of Natural Language Processing in Second Language Instruction". *CALICO Journal*, 13, 1: 47-67.
- Oliva, K. (1995). "Grammar-Based Grammar Checker: A feasible practical implementation of high-level language technologies". *Proceedings of the 5th German Student Conference on Computational Linguistics*, Saarbrücken: 54-60.
- Peng, Y. (1993). "Answer markup on computer assisted language learning". *CAELL Journal*, 10, 3: 31-40.
- Ferris, D. y Hedgcock, J.S. (1998)
- Pennington, M.C. (1993). "Computer-assisted writing on a principled basis: the case against computer-assisted text analysis for non-proficient writers". *Language and Education*, 7, 1: 43-59.
- van Zaanen, M.M. y van Huyssteen, G.B. (2003). "Various uses of a spelling checker project: practical experiences, teaching, and learning". *Southern African Linguistics and Applied Language Studies*, 21, 4: 325-338.
- Yao, Y. y Warden, C. (1996). "Process Writing and Computer Correction: Happy Weddings or Shotgun Marriage?" *CALL Electronic Journal*, 1, 1. <http://www.blackwell-synergy.com/links/doi/10.1111/0023-8333.00141/abs>.