

ECGI

Un algoritmo de Inferencia Gramatical mediante Corrección de Errores

Héctor Miguel Rulot Segovia

Trabajo presentado en la Universitat de València para optar al grado de
Doctor en Ciencias Físicas. Realizado bajo la dirección de

D. Enrique Vidal Ruiz y D. Francisco Casacuberta Nolla.

Valencia, 1992.

**Este trabajo ha sido soportado en parte por el proyecto TIC-0448/89 de la
CICYT.*

-Magnífico –le dije–. No imagino nada más repulsivo que escribir una tesis, pero tu eres capaz de disfrutar con ello.

-Gracias. Representará un año de trabajo de todos modos.

William Sloane, en *"El tiempo de la noche"*.

Al Dr. Héctor Rulot[†],
y también, al Dr. Héctor Rulot[†].
Sin ellos, este trabajo nunca
habría sido posible.

Agradecimientos

Tengo que agradecer aquí la inestimable ayuda, consejo y aliento que me ha prestado a lo largo de todo este trabajo Enrique Vidal Ruiz, que ha excedido con creces sus obligaciones como codirector de esta tesis, demostrándome con ello una gran e inmerecida amistad.

Igualmente y cómo no, agradezco a Francisco Casacuberta Nolla, también codirector de la misma tesis, su paciencia incansable y su colaboración siempre presente y nunca denegada.

Tengo aquí un extraordinario recuerdo para todos los componentes del Grupo de Investigación en Reconocimiento del Habla de Valencia, que primero en el Centro de Informática de la Universitat Literaria, y luego en el Departamento de Sistemas Informáticos y Computación de la Universidad Politécnica de Valencia, me han animado y ayudado a lo largo de muchos años para que diera fin a este trabajo.

Asimismo, debo mencionar a mis amigos y compañeros de trabajo del Centro de Informática por su paciencia conmigo durante todo este tiempo.

Lamento no disponer aquí de espacio suficiente para mencionar la aportación de todos y cada uno de aquellos que han dedicados horas y más horas al desarrollo y mejora del ECGI (Natividad Prieto, Emilio Sanchis, Asunción Castaño, Francesc Torró, Isabel Alfonso) así como aquellos que me han permitido compararlo con sus propios métodos (Pedro García, Isabel Galiano, Encarna Segarra), todos los cuales han significado un continuo incentivo para que prosiguiera mi trabajo. Todos ellos y muchos más han colaborado, incluso poniendo su voz, en el

duro trabajo de recolección y parametrización de las muestras (especialmente José Miguel Benedí).

Mención aparte merecen José Miguel Valiente y Gabriela Andreu del Departamento de Ingeniería de Sistemas, Computadores y Automática, pues sin ellos hubiera sido imposible la realización de los experimentos de reconocimiento de imágenes.

Universitat de València, 1992.

Héctor Rulot

Anecdótico

Este trabajo empezó como consecuencia de una de esas tardes de apasionado debate que se dan en todo grupo de investigación.

Por aquel entonces, algunos de los compañeros del autor, entre los que se encontraban Enrique Vidal y Emilio Sanchis, estaban desde hacía tiempo trabajando en el proyecto "Tabarca" [Benedi,89], en el que era necesario definir un autómata que reconociera dígitos a partir de una cadena de símbolos más o menos "difusos" provenientes de la etapa de parametrización. La tarea de construir una red de estados que optimizara el reconocimiento les había llevado de cabeza, ya que había sido necesario realizarla a mano aplicando el arduo método de «ensayo y error».

Aquel día, poco después de que Tabarca proporcionara resultados esperanzadores, la polémica derivó hacia lo ideal que sería el que la maravillosa red, de la que dependía en un 50% la eficiencia del reconocimiento, se pudiera inferir automáticamente a partir de las muestras de aprendizaje. Fue entonces cuando el autor, sin imaginarse las consecuencias que ello le acarrearía, recordó una idea¹ que alguna vez había rondado por aquellas reuniones: utilizar programación dinámica para comparar las secuencias, y construir incrementalmente una red añadiendo sólo los segmentos que fueran diferentes.

La cuestión obvia que inmediatamente planteó tal sugerencia volvió a ser la que alguna vez ya había sido: ¿cómo exactamente?. Pero esta vez el interrogante dió pie a una manifestación irreflexiva, típica del presente autor: «Humm..., no sé, pero no parece difícil». Manifestación que cerró la discusión ipso facto con un predecible «¡Pues manos a la obra!», puntuado de irónicas sonrisas. Claramente, aquella intervención no era seria y no impidió que la reunión terminara, como la mayoría de ellas, con un mínimo de nuevas conclusiones.

Sin embargo, algo debió quedar rondando por alguna de las circonvoluciones de detrás de las gafas del autor, porque un sábado de primavera, de esos en los que poco hay que hacer que se tengan ganas de hacer, se sentó a su escritorio y escribió los diez folios de rigor de los que consta el programa mínimo. Como era de esperar, cuando al lunes siguiente, durante un rato perdido lo introdujo en el ordenador, aquello no funcionó.

Pero en ese momento ya no era cuestión de dejarlo. El autor ya estaba convencido de que la idea era factible, y no en vano había reinventado para el caso el "algoritmo de alineamiento de una secuencia con una red mediante programación dinámica" (la P.D. como vulgarmente se conoce, era, y probablemente lo es aún, la única técnica de programación que realmente maneja el autor) [Bellman,57]. Unas cuantas vueltas de reloj y unas decenas de tachones condujeron a una simplificación drástica del algoritmo que construía la red (lugar

¹ De origen incierto, derivada quizás de las "discriminative networks" de [Moore,83]

en donde está concentrada toda la carga heurística del método), reduciendo la longitud del programa de un par de hojas. Unas horas más de lucha con el compilador, y... ¡Funcionó!. Funcionó tan bien que ahora, estimado lector, tiene usted en sus manos esta tesis, que como ve, es el resultado inesperado de un momento de cabezonería.

Cuando el autor mostró sus primeros resultados, asombrosamente aceptables a pesar de la baja calidad del corpus experimental utilizado, Enrique Vidal (que todavía no dirigía ninguna tesis) comprendió que allí había algo interesante y forzó (es la palabra) al autor a tomarse aquello en serio y a completar los experimentos. En consecuencia, unos meses más tarde se escribió a partir de ese trabajo el primer artículo realmente serio firmado por este autor; y en los sucesivos años transcurridos desde entonces, la idea original y sus mejoras han dado lugar a 16 artículos en congresos y revistas, a 8 trabajos de fin de carrera, a esta tesis, y... a alguna más que está en curso y que usted probablemente tenga ocasión de leer [Alfonso,91] [Arévalo,90] [Carpí,90] [Casacuberta,90] [Casacuberta,90a] [Castaño,90] [Galiano,90] [Galiano,91] [Miralles,91] [Prieto,88] [Prieto,91] [Prieto,91a] [Rulot,87] [Rulot,88] [Rulot,89] [Sanchis,90] [Sanchis,91] [Tarazona,91] [Torró,89] [Torró,90] [Vidal,88] [Vidal,91] [Vidal,92] [Vidal,92a].

A lo largo de estos años, muchos son los recuerdos que la realización de esta obra ha dejado en la memoria del autor. Pero, descontando las muchas horas de trabajo interesante, lleno de las típicas emociones del «¿funcionará o no funcionará?», y las inacabables noches (durante meses) que fueron imprescindibles para producir este volumen, sólo dos situaciones se imponen al recuerdo del firmante mientras escribe estas líneas. La primera de ellas, cuando, bien mediada una turbia mañana en un lujoso hotel de las Ardenas, el sobradamente conocido Dr. Thomason se puso a describir su nuevo método de inferencia gramatical... a primera vista idéntico al que el autor iba a presentar a la tarde siguiente, en su primera aparición en un congreso, con su primer discurso en lengua inglesa,... y con la aguda impresión (afortunadamente, resultó ser sólo una impresión) de que le habían pisado su maravilloso invento.

De la misma manera, no menos claramente, aunque en segundo lugar, acude a la mente del firmante la imagen de aquel sargento que, a eso de las diez de una invernal noche cartagenera, entró imponente con su escolta, abarrotando el diminuto reducto en el que el autor acaparaba el único ordenador (PC-XT) del cuartel de instrucción de la infantería de marina española. «¿Qué estás haciendo aquí?» -tronaron los galones- «Jugando a marcianitos, ¿verdad?...¿No?. A ver si me explicas entoces lo que estabas haciendo. ¡Y cuidado con lo que cuentas!, que entiendo de esto»... Curiosamente, se marchó antes de que el autor pudiera terminar de explicarle cómo se resolvía el problema del "análisis sintáctico corrector de errores en una gramática regular con circuitos".

En fin, estimado lector, para terminar con este anecdotario sólo queda decir que el firmante espera, aunque sea pecar de ambicioso, que en el (¿lejano?) futuro en el que los sistemas reconocedores del habla (o de otra cosa) sean moneda corriente en nuestra vida diaria, y todos nosotros dispongamos en nuestro hogar de uno (o varios) reconocedores, se utilice en ellos alguna de las ideas referidas a continuación. Si no... bueno, tal vez el tomo sea lo bastante grueso para calzar una mesa [Wollman,70].

H.R.S., el autor.

INDICE

Prólogo	I
En pocas palabras.....	I
¿De qué se trata?.....	I
¿Cómo se hace?.....	II
Los Capítulos.....	II
Capítulo 1: El Reconocimiento de Formas	1
1.1 Fundamentos.....	1
1.1.1 Las etapas de un reconocedor.....	2
1.1.2 Representación de objetos y formas.....	4
1.1.3 Composición de reconocedores.....	5
1.1.4 Reconocimiento Geométrico y Reconocimiento Estructural.....	5
1.1.4.1 Los modelos geométricos.....	6
1.1.4.2 Los modelos estructurales.....	7
1.1.4.3 Métodos geométricos versus métodos estructurales.....	7
1.1.5 Segmentación.....	8
1.1.6 El tamaño de las subestructuras.....	11
1.1.7 El Aprendizaje Automático.....	12
1.1.7.1 Taxonomía del aprendizaje.....	12
1.1.7.2 Factibilidad del aprendizaje Inductivo.....	12
1.2 El Reconocimiento Automático del Habla.....	13
1.2.1 Adquisición, parametrización y etiquetado en RAH.....	15
1.2.2 Estado del Arte.....	17
1.3 El Reconocimiento Automático de Caracteres.....	18
1.3.1 Adquisición, reducción de ruido y parametrización en RAC.....	19
1.3.2 Estado del Arte.....	20
Capítulo 2: Los Métodos Sintácticos	23
2.1 Introducción.....	23
2.2 Los lenguajes y sus gramáticas.....	23
2.3 Reconocedores de lenguajes.....	25
2.4 El problema de la imprecisión.....	27
2.4.1 Análisis sintáctico corrector de errores.....	28
2.4.2 Gramáticas estocásticas.....	31
2.4.2.1 Definiciones.....	31
2.4.2.2 Autómatas estocásticos.....	33
2.4.2.3 Aprendizaje de gramáticas estocásticas.....	34
2.4.3 Análisis sintáctico corrector de errores estocástico.....	35
2.5 Gramáticas SANSAT.....	37
2.5.1 Gramática SANSAT equivalente a una gramática.....	37
2.5.2 Autómata de estados etiquetados (LAS).....	39
2.5.3 LAS y Modelos Ocultos de Markov (HMM).....	40
Capítulo 3: Inferencia Gramatical	43
3.1 Introducción.....	43
3.2 Especificación del problema de IG.....	44
3.2.1 Dominio de formas a inferir.....	44
3.2.2 Espacio de hipótesis o representaciones.....	44
3.2.3 Método de presentación de ejemplos.....	44
3.2.4 Método de inferencia.....	45
3.2.5 Criterio de éxito.....	46
3.3 Métodos enumerativos.....	46
3.4 Presentación positiva.....	47
3.5 Métodos Heurísticos y Métodos Caracterizables.....	48
3.6 Métodos constructivos de IG.....	49
Capítulo 4: Programación Dinámica	51
4.1 Introducción.....	51

4.2	Principio de optimalidad	52
4.3	Relación de recurrencia, versión iterativa.....	53
4.4	Programación Dinámica y autómatas	56
4.4.1	Grafos multi-etapa	56
4.4.2	Camino mínimo en un grafo multietapa.....	57
4.4.3	Trellis	57
4.5	El algoritmo de Viterbi	59
4.6	Versión iterativa del algoritmo de Viterbi.....	60
Capítulo 5:	Corrección de errores.....	63
5.1	Reglas de error y trellis	63
5.2	Autómatas libres de circuitos.....	65
5.2.1	El algoritmo ViterbiCorrector	67
5.3	Autómatas con circuitos.....	68
5.3.1	El trellis tiene circuitos.....	68
5.3.2	Camino óptimo en un grafo con circuitos.....	69
5.3.3	El algoritmo CICLICO.....	72
5.4	Camino y traza de edición.....	74
Capítulo 6:	El método ECGI	77
6.1	Motivación.....	77
6.2	Método	79
6.2.1	Un algoritmo trivial.....	79
6.2.3	ECGI	80
6.3	Algoritmo	83
6.4	Propiedades de la gramática inferida.....	85
6.5	Heurísticos	87
6.5.1	Las asunciones de ECGI.....	87
6.5.2	No se añaden las reglas de error.....	90
6.5.3	Heurísticos no esenciales.....	91
6.5.3.1	Prioridad a la sustitución.....	92
6.5.3.2	Inserción en paralelo	94
6.5.3.3	Otras consideraciones	95
6.6	(Di)similitudes	96
6.6.1	Menos errores.	97
6.6.2	Relativamente menos errores.....	97
6.6.3	Más aciertos	98
6.6.4	Relación entre criterios	98
6.6.5	Nadie es perfecto.....	100
6.7	Reconocimiento.....	101
6.8	Aprendizaje continuo	102
6.9	Sólo sustituciones	104
6.9.1	Consideraciones prácticas	105
6.10	Implementación	106
6.10.2	Representación Gráfica de un LAS de ECGI	107
6.10.3	Version LAS de la construcción	108
Capítulo 7:	ECGI Estocástico.....	113
7.1	Ambigüedad y probabilidad	114
7.2	Construcción estocástica.....	114
7.3	Reconocimiento estocástico	115
7.4	La estocasticidad de la gramática expandida.....	118
7.4.1	Las probabilidades de deformación.....	118
7.4.2	Estocasticidad según ECGI	120
7.4.3	En la práctica	121
7.5	Sólo sustitución	124
Capítulo 8:	Resultados experimentales.....	125
8.1	Reconocimiento del Habla	126
8.1.1	Representación simbólica de la señal vocal	126
8.1.1.1	Corpus piloto: Dígitos monolocutor.....	127
8.1.1.2	Corpus principal: Dígitos.....	129
8.1.1.3	Corpus difícil: letras.....	131
8.1.2	Experimentos de reconocimiento.....	131
8.1.2.1	Corpus piloto.....	131
8.1.2.2	Corpus principal.....	135
8.1.2.2.1	Experimentos sencillos.....	135

8.1.2.2.2	Leaving-k-out	136
8.1.2.3	Corpus de Letras	139
8.1.2.3.1	Experimentos sencillos	139
8.1.2.3.1	Leaving-k-out	140
8.2	Reconocimiento de imágenes planas	142
8.2.1	Representación simbólica de imágenes planas	143
8.2.3	Dígitos Manuscritos	144
8.2.3.1	Los experimentos	145
8.2.3.2	Los resultados	146
8.2.4	Dígitos Impresos	148
8.2.4.1	Los experimentos	149
8.2.4.2	Los resultados	150
8.2.4.3	El uno sin base	152
8.2.6	Invarianza a la rotación y otras posibles extensiones	154
8.3	Experimentos de síntesis	155
8.4	Resumen de resultados	157
Capítulo 9:	ECGI: Estudios Empíricos	159
9.1	Introducción	159
9.2	¿Minimizar errores o maximizar aciertos?	159
9.3	Convergencia del aprendizaje	162
9.4	Ambigüedad de las gramáticas	165
9.5	Sólo Substituciones	167
9.6	ECGI estocástico y no estocástico	169
9.6.1	Ignorar la frecuencia de las reglas	170
9.6.2	Ignorar las frecuencias de los errores	171
Capítulo 10:	Caminos y Simplificación de Automatas	173
10.1	Introducción	173
10.2	Caminos en un grafo sin circuitos	173
10.2.1	Numero de caminos por vértice	174
10.2.2	Número de caminos por arista	175
10.2.3	Camino más largo y más corto	176
10.2.4	Algoritmos no recursivos	176
10.2.5	Relación Tráfico/Probabilidad	177
10.3	Simplificación de autómatas	179
10.3.1	Método	180
10.3.1.1	Estado menos significativo	180
10.3.1.2	Criterios de detención	181
10.3.1.3	El algoritmo	181
10.3.2	Aplicación práctica y resultados	183
Capítulo 11:	ECGI Determinista y Cadena Mediana	187
11.1	Introducción	187
11.2	Gramáticas Deterministas	187
11.2.1	Método	188
11.2.2	Implementación	190
11.2.3	Resultados Experimentales	192
11.3	Cadena mediana	194
11.3.1	Mediana generalizada y mediana de un conjunto	194
11.3.2	La cadena más probable	195
11.3.3	Consideraciones prácticas	195
11.3.4	La cadena más frecuente	196
11.3.5	Comprobación empírica	197
Capítulo 12:	Comparaciones	201
12.1	Métodos similares de IG	201
12.1.1	Método de Thomason	201
12.1.2	Método de Falaschi	203
12.1.3	Método de Chirathamjaree	206
12.1.4	Método de Marco	206
12.2	Métodos en Reconocimiento del Habla	207
12.2.1	Modelos Ocultos de Markov	207
12.2.2	Alineamiento Temporal No lineal, K-Vecinos	209
12.3	Métodos en Reconocimiento de Imágenes	210
Epílogo		213
	Conclusiones	213

Perspectivas	216
Apéndice A: Otros trabajos sobre ECGI.....	A-I
Apéndice B: Tablas complementarias.....	B-I
B.1 Número de estados de los autómatas inferidos	B-I
B.1.1 Dígitos Hablados:.....	B-I
B.1.2 Letras Habladas:.....	B-II
B.1.3 Dígitos Manuscritos:	B-III
B.1.4 Dígitos Impresos:.....	B-IV
B.2 Matrices de confusión	B-V
B.2.1 Dígitos Manuscritos (sólo substitución)	B-V
B.2.2 Dígitos impresos (sólo substitución).....	B-VI

Prólogo

En pocas palabras

¿De qué se trata?

En el presente trabajo se introduce un nuevo método de inferencia gramatical, el algoritmo INGRACE (INferencia GRAMatical mediante Corrección de Errores).

Desafortunadamente para los hispano-parlantes, al INGRACE se le ha dado a conocer sobre todo según la versión inglesa de su nombre: «**Error-Correcting Grammatical Inference algorithm**», con lo cual a partir de ahora utilizaremos para referirnos a él el impronunciable acrónimo «**ECGI**».

Como todos los algoritmos de inferencia gramatical, la aplicación principal de ECGI se halla en inferir la estructura (gramática) que mejor represente una determinada forma (lenguaje) a partir de muestras de objetos pertenecientes a dicha forma (cadenas). Es pues un algoritmo cuya finalidad original se halla en el reconocimiento de formas.

En concreto, el (buen) funcionamiento de ECGI se ha comprobado en dos casos muy diferentes del reconocimiento de formas: el reconocimiento de la **palabra hablada**, campo al que se dedica nuestro grupo investigador, y el reconocimiento de **formas planas** (en imágenes bidimensionales), campo que nos interesó, pues es también poco usual en él el empleo de métodos de inferencia gramatical.

Ello no es limitativo, y en principio ECGI es aplicable a cualquier problema de reconocimiento de formas en el que se puedan representar las mismas mediante cadenas de símbolos.

Cabe notar sin embargo, el que la naturaleza inherentemente heurística del método restringe el campo de aplicación a aquellos casos en los que las características diferenciadoras de las formas se centran en la concatenación unidimensional de unos u otros (grupos de) símbolos y en la distinta longitud (duración) de los grupos de símbolos (las subformas o subcadenas).

¿Cómo se hace?

ECGI construye una gramática regular (un autómata de estados finitos) mediante un procedimiento incremental, que en principio permite efectuar simultáneamente inferencia y reconocimiento.

El procedimiento de construcción se apoya en un método de Programación Dinámica, el cual determina, para cada nueva muestra, la secuencia de reglas de la gramática que la generan con un mínimo número de reglas de error (de reglas **no** pertenecientes a la gramática). A partir de estas reglas de error, se determinan las reglas a añadir a la gramática actual, de forma que en lo sucesivo la muestra forma parte del lenguaje de la misma.

La minimización explícita permite controlar el crecimiento de la gramática, a la vez que induce una generalización que ha demostrado ser especialmente idónea para capturar la variabilidad presente en las (sub)estructuras de las cadenas consideradas, así como en la concatenación y la longitud de estas subestructuras.

Los experimentos realizados en reconocimiento del habla, destino inicial para el que se ideó ECGI, han demostrado una elevada capacidad de aprendizaje, que conduce a prestaciones en reconocimiento superiores a la conseguidas cuando los autómatas se construyen a mano, y muy similares (o superiores) a las obtenidas mediante otros métodos de aprendizaje, ya sean o no de inferencia gramatical.

En cuanto al reconocimiento de formas planas, los experimentos se centraron en formas muy típicas y sobre las cuales existe una abundante literatura: los dígitos impresos y/o manuscritos. Aquí también ECGI demostró ser comparable y en muchos casos superior a métodos clásicos en ese campo (momentos centrales normalizados con distancia de Mahalanobis).

Los Capítulos

El texto de la presente exposición se distribuye en 12 capítulos:

- El primer capítulo resume en unas páginas qué es y de qué trata la disciplina del "reconocimiento de formas", desarrollando brevemente algunos conceptos de interés para situar este trabajo, tales como el reconocimiento geométrico (estadístico) y sintáctico de formas y el problema de la longitud de las subestructuras

A continuación el capítulo se centra en el "reconocimiento automático del habla", del que se explican muy por encima los métodos básicos y el estado del arte, ambos necesarios para poder

entender y comparar los resultados experimentales que se presentarán al final de este volumen.

De la misma manera y por la misma razón, al final de capítulo se ofrece una introducción al reconocimiento de formas (imágenes) planas y más concretamente al reconocimiento de dígitos impresos/manuscritos.

- El segundo capítulo se presenta como una introducción a la teoría de lenguajes formales, pero se centra casi de inmediato en los lenguajes y gramáticas regulares, así como en sus respectivas versiones estocásticas, todos ellos imprescindibles para explicar mínimamente el funcionamiento y aplicación de ECGI.
- El tercer capítulo intenta plantear formalmente el problema de la inferencia gramatical. En él, tras introducir los conceptos básicos sobre inferencia y gramáticas, se exponen toda una serie de resultados teóricos que van constriñendo el campo de lo que es posible inferir y con qué procedimientos puede hacerse. Ello permite delimitar claramente el dominio sobre el que trabajará nuestro algoritmo, así como tipificarlo y situarlo con relación a los otros métodos existentes en inferencia gramatical.
- El cuarto capítulo procede a una exposición rápida de los métodos y algoritmos de la Programación Dinámica, técnica de optimización en la que se basa ECGI a la hora de minimizar el crecimiento de la gramática que infiere. El capítulo se fija principalmente en el algoritmo de Viterbi, imprescindible cuando se manejan gramáticas estocásticas y correctoras de errores.
- En el quinto capítulo aparecen las primeras aportaciones originales de este trabajo. En este capítulo se lleva a cabo un análisis detallado de los algoritmos correctores de errores para el análisis sintáctico regular (estocástico o no), prestando particular atención a los problemas que se presentan con las gramáticas con circuitos. Se proponen, comprobándolos en la práctica, varios algoritmos para solventar estos problemas. Finalmente se introduce un concepto básico para comprender ECGI: el camino de derivación.
- El sexto capítulo presenta el método ECGI, el algoritmo y sus propiedades, así como la propiedades de las gramáticas inferidas. Se estudian detalladamente los heurísticos en que se basa el algoritmo, su utilización en reconocimiento de formas y su implementación real.
- En el séptimo capítulo se introduce la extensión estocástica de ECGI y se describen los detalles del procedimiento seguido para aprovechar la información estadística contenida en la frecuencia de utilización de las reglas. Se propone y estudia una nueva manera de definir la estocasticidad de la gramática expandida cuando se efectúa un análisis sintáctico corrector de errores.

- El octavo capítulo expone extensivamente los múltiples experimentos de reconocimiento de formas que han sido necesarios tanto para demostrar la viabilidad de ECGI y su eficacia como reconocedor, como para estudiar su comportamiento y propiedades en distintos casos. Se describen los corpus de datos utilizados (dígitos hablados e impresos), los métodos de parametrización aplicados y los resultados obtenidos.
- El noveno capítulo se dedica a profundizar en ECGI mediante una serie de estudios empíricos. Se muestran los resultados de distintos experimentos destinados a analizar la convergencia de ECGI, la ambigüedad de las gramáticas que genera y la cantidad de información que realmente aportan los distintos tipos de frecuencias que se utilizan en la extensión estocástica.
- En el décimo capítulo se introduce una serie de cantidades, todas ellas relacionadas con los caminos entre dos vértices de un grafo sin circuitos. Estas cantidades se emplean luego para reducir la complejidad de los modelos generados por ECGI (simplificación de autómatas).
- En el capítulo once se expone un heurístico (subóptimo) con el que se puede obligar a ECGI a generar gramáticas deterministas, lo que permite eliminar la ambigüedad de las mismas. Asimismo se presenta un método para obtener una aproximación a la cadena mediana de un conjunto de cadenas a partir del correspondiente modelo inferido por ECGI.
- El duodécimo capítulo y último compara ECGI con otros métodos similares de inferencia gramatical (basados también en la corrección de errores) y con diversas otras metodologías utilizadas generalmente en reconocimiento del habla y de formas planas.

Finalmente, unas páginas de epílogo, seguidas de algunos apéndices y de la bibliografía concluyen el presente trabajo.