



MODELS ALGEBRAICS ANALÍTICS DEL LLENGUATGE. UN EXEMPLE: LES GRAMÀTIQUES CONTEXTUALS; FORMALITZACIÓ I ESTUDI DE LES SEVES APLICACIONS

Joan Miquel Vergés

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

UNIVERSITAT ROVIRA I VIRGILI
FACULTAT DE LLETRES
DEPARTAMENT DE FILOGIES ROMÀNIQUES

MODELS
ALGEBRAICS ANALÍTICS
DEL
LLENGUATGE

UN EXEMPLE: LES GRAMÀTIQUES CONTEXTUALS;
FORMALITZACIÓ I ESTUDI DE LES SEVES APLICACIONS

TESI DOCTORAL
PRESENTADA PER
JOAN MIQUEL I VERGÉS
DIRIGIDA PER
DR. CARLOS MARTÍN VIDE

Tarragona, 1997

UNIVERSITAT ROVIRA I VIRGILI

FACULTAT DE LLETRES

DEPARTAMENT DE FILOLOGIES ROMÀNIQUES

**MODELS
ALGEBRAICS ANALITICS
DEL
LLENGUATGE**

**UN EXEMPLE: LES GRAMÀTIQUES CONTEXTUALS;
FORMALITZACIÓ I ESTUDI DE LES SEVES APLICACIONS**

TESI DOCTORAL

PRESENTADA PER
JOAN MIQUEL i VERGÉS

DIRIGIDA PER
DR. CARLOS MARTIN VIDE

Tarragona, 1997

*Súnion! T'evocaré de lluny amb un crit d'alegria,
tu i el teu sol lleial, rei de la mar i del vent:
pel teu record, que em dreça, feliç de sal exaltada,
amb el teu marbre absolut, noble i antic jo com ell.
Temple mutilat, desdenyós de les altres columnes
que en el fons del teu salt, sota l'onada rient,
dormen l'eternitat! Tu vetlles, blanc a l'altura,
pel mariner, que per tu veu ben girat el seu rumb;
per l'embriac del teu nom, que a través de la nua garriga
ve a cercar-te, extrem com la certesa dels déus;
per l'exiliat que entre arbredes fosques t'albira
súbitament, oh precís, oh fantasmal! i coneix
per ta força la força que el salva als cops de fortuna,
ric del que ha donat, i en sa ruïna tan pur.*

A una nació, *Romania.*

A un déu, *Solomon Marcus.*

A un profeta, *Gheorghe Păun.*

A un mestre, *Carlos Martín Vide.*

A un amor, *Mati.*

A un poeta, *Carles Riba.*

PÀG.	LÍN.	ON DIU ...	HA DE DIR ...
7	7	de l'Aacadèmia	de l'Acadèmia
8	4	ITSM	ITESM
8	8	Min-isterio	Mi-nisterio
15	5	(n, k) -C-autòmat	(n, k) -C-autòmats
31	1	$X_e(F)$	$X_d(F)$
43	9	de L_3	d' L_3
47	13	Sreider	Šreider
50	4	Tomas-Marcu	Toma-Marcu
50	8	(Coardo, s, 1976)	(Coardos, 1976)
50	16	Jurča	Jurcă
54	5	Uspenskiĭ	Uspenskij
54	9	Uspenskiĭ	Uspenskij
54	15	Gladkiĭ	Gladkij
55	1	grammatical	gramatical
55	10	Horecký	Horecký
55	12	Gladkiĭ	Gladkij
59	13	relacionat amb amb eines	relacionat amb eines
66	14	considerada per	considerat per
67	13	de galois	de Galois
73	21-22	gramat-icals	grama-ticals
80	6	de L_0	d' L_0
80	10	de L_0	d' L_0
82	14	la de l' <i>esquerra</i>	el de l' <i>esquerre</i>
82	14-15	la de la <i>dreta</i>	el de la <i>dreta</i>
83	14	de L_0	d' L_0
83	18-19	gen-erats	gene-rats
85	16-17	corre-spondre's	corres-pondre's
85	21	i g de H	i g d' H
88	2	de S_i	d' S_i
88	3	de S_i	d' S_i
88	5	de S_i	d' S_i
88	5	de S_1	d' S_1
88	8	de S_i	d' S_i
88	8	de S_i	d' S_i
88	10	de S_i	d' S_i
88	10	de S_1	d' S_1
88	12-13	simplic-itat	simpli-citat
89	3	discutda	discutida
91	11-12	prob-lemes	pro-blemes
91	17	de S_1	d' S_1

(continua)

PÀG.	LÍN.	ON DIU ...	HA DE DIR ...
339	13	Teorema 10.1.2 de la pàgina 391	Teorema 9.2.2 de la pàgina 334
339	17	pàgina 335	pàgina 339
340	11	pàgina 306	pàgina 310
351	9	$L_i = \partial^1$	$L_i = \partial^e$
352	2	$X^1 \rightarrow uY^1, X^1 \rightarrow uY^r, X^1 \rightarrow uY^f$	$X^e \rightarrow uY^e, X^e \rightarrow uY^d, X^e \rightarrow uY^f$
352	3	$X^r \rightarrow Y^1v, X^r \rightarrow Y^rv, X^r \rightarrow Y^fv$	$X^d \rightarrow Y^ev, X^d \rightarrow Y^dv, X^d \rightarrow Y^fv$
352	8	$(\{X^l, X^r, X^f \mid x \in N\}, T, S^l, P')$	$(\{X^e, X^d, X^f \mid x \in N\}, T, S^e, P')$
363	20	Lemes 10.1.1 de la pàgina 375	Lemes 9.1.4 de la pàgina 303
366	10	Lema 10.1.1 de la pàgina 375	Lema 9.1.4 de la pàgina 303
370	3-4	supercadena estricte	supercadena estricta
384	13-14	Lema 10.1.2 de la pàgina)).	Lema 6.1.2 de la pàgina 211).
391	10	pàgina 376	pàgina 380
391	11	Figura 9.2	Figura 10.1.2
394	11	Teorema 10.1.2 de la pàgina 391	Teorema 9.2.2 de la pàgina 334
394	17	pàgina 158	pàgina 162
397	8	X_e	X_d
404	20-21	Lema 11.3.1 de la pàgina 449	Lema 9.1.6 de la pàgina 308
406	5	pàgina 399	pàgina 403
406	17	pàg. 197	pàg. 201
407	15	pàgina 154	pàgina 159
409	19-20	Teorema 10.1.2 de la pàgina 391	Teorema 9.2.2 de la pàgina 334
425	6	pàgina 373	pàgina 377
425	6	pàgina 383	pàgina 387
425	7	pàgina 390	pàgina 394
425	13	pàgina 399	pàgina 403
425	14	pàgina 403	pàgina 407
425	15	pàgina 409	pàgina 413
425	18	pàgina 416	pàgina 420
435	17-18	con-struïm	cons-truïm
442	13	esugerra	esquerra
442	14	Chomski	Chomsky
446	2-3	gen-eració	gene-ració
446	4	amb derivació paral·lela poden ser trobat.	amb derivació paral·lela.
451	16-17	ree-scrivint	re-escrivint
457	9	Dirèctament	Directament
464	14	“with marked derivatios”	“with marked derivations”
472	9-10	ree-scriptura	re-escriptura
476	12	$\partial^1(\partial^r$	$\partial^e(\partial^d$

(continua)

PÀG.	LÍN.	ON DIU ...	HA DE DIR ...
476	13	$\partial^1(\partial^r$	$\partial^e(\partial^d$
476	15	$\partial^1(\partial^r$	$\partial^e(\partial^d$
482	19	lema 10.1.2 de la pàgina 380	lema 9.2.2 de la pàgina 334
483	8	pàgina 473	pàgina 477
483	17	pàgina 117	pàgina 121
486	7	pàgina 226	pàgina 230
486	12	pàgina 399	pàgina 403
486	17	pàgina 479	pàgina 483
486	18	pàgina 477	pàgina 481
486	21	pàgina 470	pàgina 474
488	12	pàgina 149	pàgina 153
489	1	pàg. 483	pàg. 487
505	13	conjunts de cades,	conjunts de cadenes,
508	1	pàgina 133	pàgina 137
508	13	“contraction automats”	“contraction automata”
509	14-15	ver-ificadors	veri-ficadors
511	21-22	de-sprés	des-prés
511	22	in-strucció	ins-trucció
519	1	pàgina 138	pàgina 142
519	9	pàgina 150	pàgina 154
521	19	pàgina 514	pàgina 518
527	17	pàgina 519	pàgina 523
531	7	capac-itats	capaci-tats
531	8	pàgina 133	pàgina 137
537	20	Figura ??	Figura 14.2.1
539	5	estat de K_B	estat de K_C
544	13	pàgina 524	pàgina 528
544	17	pàgina 520	pàgina 524
545	18	pàgina 524	pàgina 528
546	19	pàgina 520	pàgina 524
549	4	$s \in K_B$	$s \in K_C$
549	9	de K_B	de K_C
556	12	pàgina 555	pàgina 553
558	1	pàgina 14.4.2	pàgina 553
558	1	pàgina 14.4.4	pàgina 556
563	9	coentífic	científic
587	7	$L_2 = \{ 23, 24, 25, 26, 78, 87 \}$	$L_2 = \{ 23, 24, 25, 26, 78 \}$
587	8	$L'_2 = L_2 \cup \{ 7887 \dots \}$; $\text{card}(L'_1) = \infty$. La infinitud ... Són <i>Astèrix</i> i <i>Obèlix</i> , respectivament.	$L'_2 = L_2$; $\text{card}(L'_2) = 5$.

(continua)

PÀG.	LÍN.	ON DIU ...	HA DE DIR ...
598	19	l'Exemple ? ?	l'Exemple 16.0.3
602	5	l'Exemple ? ?	l'Exemple 16.0.3
603	5	l'Exemple anterior	l'Exemple 16.0.3 anterior
606	13	una darrera l'altre	una darrera l'altra
608	4	munic-ipals	municipals
620	19-20	ob-stant	obs-tant
621	11-12	inter-essant	inte-ressant
623	4	pàgina 375	pàgina 371
625	19-20	ree-scrit	re-escrit
626	6	Galiukschov sota	Galiukschov (Galiukschov, 1981) sota
631	7	mitjançs	mitjans
631	9	d'us	d'ús
632	1	produint extrems extrems	produint extrems
632	12	rep e nom	rep el nom
636	11	defined per	definida per
636	13	per ò l'operació	però l'operació
643	4-5	re-stringida	res-tringida
648	17	“la exposició donada	“l'exposició donada
674	9	Coardo, s	Coardos
677	3	anf	and
684	3	To appear	Pendent de publicació
692	1	l'tude	l'étude
693	1	de kunze	de Kunze
695	15	Traucció	Traducció
695	17	Math'ematique	Mathématique
696	1	Math'ematique	Mathématique
696	4	? ' por qu las	¿por qué las

Bibliografia addicional

Serrano, S (1979). Teatre i matemàtica. *Quaderns de Ponent*, 1, pàgines 40-58, Lleida.

AGRAIMENTS

Voldria expressar el meu profund agraïment a totes aquelles persones i institucions que han tingut l'avinentesa d'ajudar-me durant aquests anys de formació doctoral i de redacció d'aquesta tesi que avui presento.

En especial, vull significar la importància que han tingut en la meva formació acadèmica i humana el Prof. Solomon Marcus de la Facultat de Matemàtiques de Bucarest (Romania), el Prof. Gheorghe Păun de l'Institut de Matemàtiques de l'Acadèmia Romanesa de Ciències de Bucarest (Romania) i, finalment, el Prof. Carlos Martín Vide de la Universitat Rovira i Virgili de Tarragona (catedràtic de Lingüística del Departament de Filologies Romàniques i Cap del Grup de Recerca en Lingüística Matemàtica i Enginyeria del Lenguatge - GRLMC).

També agrair als professors Mihai Dinu i Monica Tătărâm de la Universitat de Bucarest (Romania); als professors Elisabet Engdahl, Matt Crocker, Alexander Holt, Mercè Prat, Paul Schweizer, Richard Shillcock i Enric Vallduví del *Centre for Cognitive Science - CCS & the Human Communication Research Centre HCRC* de la Universitat d'Edinburgh (Escòcia); als professors Arto Salomaa, Lucian Ilie, Alexandru Mateescu i Valeria Mihalache del *Department of Mathematics of the University of Turku and the Turku Centre*

for Computer Studies - TUCS (Finlàndia); a la Prof. Erzébet Csuhaj-Varjú de la *Hungarian Academy of Sciences & Computer and Automation Institute* (Hongria); als professors Fidel Chávez, Alva Canizal, Víctor Garza i Yolanda Pérez del *Instituto Tecnológico y de Estudios Superiores de Monterrey - ITESM* (Mèxico); i, finalment, als doctors, llicenciats, estudiants i amics, en Joan Busquets, Lluís Fondevila, Rosa Ma. Hidalgo, M. Dolores Jiménez, Rudolf Ortega, Rosa Ma. Pérez i Jesús Vidal del *Grup de Recerca en Lingüística Matemàtica i Enginyeria del Llenguatge - GRLM*. I a la Pepita, i en el seu nom al *Ministerio de Educacion y Ciencia*, per fer-me partícip d'una beca d'investigació pre-doctoral. A tots vosaltres: moltes gràcies per tot !

Només em queda agrair a en Jordi Reiss i a la Mati Recolons la seva constant disposició anímica i personal, que tant ha contribuït a dur a bon terme aquesta tesi, amb qui estaré sempre en deute.

Per a vosaltres, i tots aquells a qui no he anomenat, el meu sincer agraïment per tot el que heu fet per mi.

Arenys de Munt, El Maresme, febrer de 1997.

* * * * *

CONTINGUT

1 Els models algebraics analítics del llenguatge	33
1.1 La Fonologia i la segmentació	36
1.2 Les relacions de dominació	41
1.3 Clausura i connexions de Galois	46
1.4 Les parts de l'oració i altres aspectes	50
1.5 L'espai contextual. La Topologia	56
2 Les gramàtiques contextuais	59
2.1 L'origen de les gramàtiques contextuais	61
2.2 Motivacions de les gramàtiques contextuais	63
2.3 Cadenes, contextos i clausura de Sestier	66
2.4 Modelització de les categories	70
2.5 L'aproximació contextual en la perspectiva generativa	74
2.6 Generació de cadenes i contextos	78

2.7	Interrelació de cadenes i contextos	83
2.8	Aprofundint en cadenes i contextos	89
2.9	Un nivell més alt d'abstracció	93
2.10	El poder generatiu	96
2.11	Altres línies de recerca	97
3	Teoria de Llenguatges Formals	103
3.1	Nocions bàsiques	103
3.2	Operacions amb llenguatges	105
3.3	Gramàtiques de Chomsky	111
3.4	Gramàtiques regulades	119
3.4.1	Gramàtiques controlades	121
3.4.2	Gramàtiques matricials	123
3.4.3	Gramàtiques programades	126
3.4.4	Gramàtiques condicionals	128
3.5	Sistemes de Lindenmayer	130
3.6	Autòmats i transductors	133
3.7	Problemes de decidibilitat	150
3.8	Condicions necessàries	152
3.9	Representacions	158
3.10	El teorema de l'espai de treball	160

3.11 Formes normals	161
3.12 Ambigüitat	162
3.13 Complexitat descriptiva	163
4 Gramàtiques contextuais bàsiques	167
4.1 Definicions	167
4.2 Exemples	175
5 Capacitat generativa de les famílies de gramàtiques contextuais bàsiques	183
5.1 Condicions necessàries	184
5.2 Jerarquies de llenguatges contextuais	197
5.3 Relacions amb la jerarquia de Chomsky	200
6 Propietats de clausura de les famílies de gramàtiques contextuais bàsiques	209
6.1 Operació d'unió	210
6.2 Operació d'intersecció amb conjunts regulars	211
6.3 Operació d'intersecció	213
6.4 Operació de complementarietat	215
6.5 Operació de concatenació	216
6.6 Operació de Kleene + ...	218

6.7 Operació de morfismes (λ -lliure)	219
6.8 Operació de substitució	221
6.9 Operació de substitució finita	223
6.10 Operació de morfismes inversos	224
6.11 Operació de barreja	227
6.12 Operació d'imatge mirall	228
6.13 Altres operacions de clausura: operació de prolongació a la dreta	230
7 Propietats de decidibilitat de les famílies de gramàtiques contextuals bàsiques	233
7.1 Propietat d'equivalència	235
7.2 Propietat d'inclusió	236
7.3 Propietat de pertanyença	236
7.4 Propietat de buidor	237
7.5 Propietat de finitud	237
7.6 Altres propietats de decidibilitat: l'ambigüitat	240
8 Altres propietats rellevants de les famílies de gramàtiques contextuals bàsiques	243
8.1 La semilinealitat i propietats relacionades	243

8.2 La complexitat descriptiva	248
8.3 L'ambigüïtat	273
9 Gramàtiques contextuais amb selecció restringida	301
9.1 Definicions i resultats bàsics	302
9.2 Gramàtiques contextuais internes amb selecció finita	327
9.3 Gramàtiques contextuais externes amb selecció regular	340
10 Gramàtiques contextuais amb restriccions en els components	365
10.1 Gramàtiques contextuais amb ús minimal i maximal dels selectors	366
10.1.1 Definicions bàsiques i el poder de l'ús minimal dels selectors	368
10.1.2 El poder de les gramàtiques contextuais amb ús maximal dels selectors	381
10.2 Gramàtiques contextuais amb contextos esborrats	396
10.3 Gramàtiques contextuais amb contextos unilaterals	401
10.4 Gramàtiques contextuais deterministes	414
11 Gramàtiques contextuais amb modificacions en les seves relacions de derivació	427
11.1 Gramàtiques contextuais amb derivació màxima a l'esquerra	428

11.2 Gramàtiques contextuais amb derivació paral·lela	444
11.3 Gramàtiques contextuais amb derivació bloquejada	446
11.4 Gramàtiques contextuais amb derivació de primera profunditat	451
11.5 Gramàtiques contextuais amb derivació marcada	464
12 Gramàtiques contextuais regulades	471
12.1 Gramàtiques contextuais matricials	473
12.2 Gramàtiques contextuais programades	479
12.3 Gramàtiques contextuais controlades (regularment)	483
12.4 Altres gramàtiques contextuais regulades	489
13 Altres tipus de gramàtiques contextuais	491
13.1 Gramàtiques contextuais difuses o borroses	491
13.2 Gramàtiques contextuais amb infinits contextos	495
13.3 Gramàtiques contextuais basades en patrons	496
13.4 Gramàtiques contextuais de dos nivells	500
14 Autòmats Contextuais	507
14.1 Autòmats contextuais vs. autòmats d'esborrament	509
14.1.1 Definicions bàsiques	509

14.1.2 Relació entre els autòmats d'esborrament amb operació de reiniciació i els autòmats contextuais	517
14.2 Autòmats contextuais vs. autòmats de contracció	530
14.2.1 Definicions bàsiques	531
14.2.2 Relació entre els (n, k) - C -autòmats i els autòmats contextuais	535
14.3 Caracterització de llenguatges contextuais usant famílies de DR -autòmats	542
14.4 Caracterització de famílies de llenguatges contextuais usant famílies d'autòmats basades en autòmats de contracció	547
14.4.1 Resultats preliminars	548
14.4.2 Resultats de caracterització	550
14.4.3 Derivacions marcades	556
14.4.4 Llenguatges contextuais programats	558
14.4.5 Altres caracteritzacions	559
15 Caracterització de llenguatges associats a l'acció dramàtica (teatre i còmic) mitjançant gramàtiques contextuais	561
15.1 Llenguatges associats al teatre	561
15.2 Llenguatges associats al còmic	580
16 Caracterització de llenguatges associats als "sistemes d'accions" mitjançant gramàtiques contextuais	591

17	Caracterització de llenguatges amb “computació natural”	
	mitjançant gramàtiques contextuais	619
17.1	L’adjunció (i l’esborrament)	620
17.2	La inserció (i l’esborrament)	625
17.3	L’empalmament	631
17.4	Consideracions finals	638
18	Caracterització dels llenguatges/llengües naturals	
	mitjançant gramàtiques contextuais	639
18.1	Un model formal global de les llengües naturals:	
	llenguatges suaument dependents del context	639
18.1.1	Exemples de “construccions” no-independents del context	
	en els llenguatges/llengües naturals	649
18.2	Un model formal de les llengües naturals: els llenguatges	
	contextuals amb ús maximal de selectors	655
19	Conclusions finals	665
*	Bibliografia	669

QUESTIONS METODOLÒGIQUES

1 Distribució capitular

La tesi en sí està dividida en 19 capítols. El **Capítol 1** i el **Capítol 2** corresponen a una introducció als models algebraics analítics del llenguatge i a les gramàtiques contextuais, respectivament. El **Capítol 19** mostra les conclusions finals de la tesi.

La resta d'informació està agrupada en dos grans blocs. Per una part el model formal, teòric, matemàtic que inclouria els capítols 3 – 14 i, per altra part el model aplicat, pràctic, que inclouria els capítols 15 – 18.

1.1 Marc teòric

Dins del marc teòric comencem per una introducció a la teoria de llenguatges formals (**Capítol 3**), per tal de fer el text (tan) independent (com es pugui) dels coneixements en teoria de llenguatges formals, especificant una sèrie de nocions bàsiques i de resultats en aquesta àrea; d'aquesta manera també fixarem la notació que usarem en tota la tesi.

A continuació introduïrem les classes de gramàtiques-llenguatges contextuais bàsics (**Capítol 4**), donant les seves definicions i alguns exemples al respecte; estudiarem les seves capacitats generatives (**Capítol 5**), comparant després, amb les famílies de la jerarquia de Chomsky i buscarem condicions

necessàries per tal que un llenguatge pertanyi a una família donada. Seguidament (**Capítol 6**) estudiarem les seves propietats o operacions de clausura (unió, intersecció amb conjunts regulars, intersecció, complementarietat, concatenació, Kleene +, morfismes λ -lliures, substitució, substitució finita, morfismes inversos, barreja, imatge mirall i d'altres operacions de clausura pròpiament contextuais com, p.e., l'operació dòmino o de prolongació a la dreta). També estudiarem (**Capítol 7**) les seves propietats o problemes de decidibilitat bàsics en teoria de llenguatges: la buidor, la finitud, l'equivalència, la inclusió i la pertanyença. Finalment, estudiarem altres propietats (lingüístiques) rellevants com és el cas de la semilinealitat i l'ambigüïtat (**Capítol 8**). En tots els casos (capacitat generativa, propietats de clausura i propietats de decidibilitat) agruparem els resultats obtinguts en un teorema-taula final per tal de facilitar el seu ús posterior en la tesi.

En la definició de gramàtica contextual introduïda en el capítol 4, les gramàtiques contextuais tenen una concepció infinitària: la funció de selecció en la presentació funcional és definida per a totes les cadenes de V^* (anàlogament, els selectors en la presentació modular poden ser conjunts infinits). No obstant, la noció d'una "gramàtica" suposa algun tipus de mecanisme finit que descrigui la sintaxi d'un llenguatge. Per tant, es fan necessàries algunes restriccions en la definició general d'una gramàtica contextual (en la funció de selecció i en els selectors). La idea més natural és la d'imposar certa regularitat en la definició de la selecció. En el **Capítol 9** veurem un estudi sistemàtic d'aquestes variants.

A continuació veurem algunes idees de com modificar la definició de les gramàtiques contextuais o dels llenguatges generats per elles. Els capítols 10–13 estan dedicats a l'estudi d'aquestes variants. Algunes d'elles es corresponen amb idees clàssiques ja explorades en la teoria de llenguatges formals, especialment pel cas de les gramàtiques lliures del context, però moltes d'altres són específiques de les gramàtiques contextuais. En primer lloc (**Capítol 10**) considerarem variants de les gramàtiques contextuais, on imposarem certes restriccions en els seus components. Veurem les gramàtiques contextuais: amb ús maximal/minimal dels selectors, (amb contextos) unilaterals i deterministes. Seguidament considerarem un seguit de modificacions en les relacions de derivació definides respecte a una gramàtica contextual (**Capítol 11**). Estudiarem les gramàtiques contextuais: amb derivació màxima a l'esquerra, amb derivació paral·lela, amb derivació bloquejada, amb derivació de primera profunditat i, finalment, amb derivació marcada.

En les variants de les gramàtiques contextuais considerades fins ara, l'ordre d'ús de les produccions d'una gramàtica contextual és lliure. No obstant, podem restringir aquesta llibertat, controlant la derivació a l'estil de la reescriptura regulada per a les gramàtiques de Chomsky. Veurem (**Capítol 12**) els casos de les gramàtiques contextuais: matricials, programades i controlades (regularment).

Per acabar amb aquest bloc de variants de les gramàtiques contextuais, veurem (**Capítol 13**) algunes altres variants introduïdes dins el marc de les

gramàtiques/llenguatges contextuais: les gramàtiques contextuais difuses o borroses, les gramàtiques contextuais amb infinits contextos, les gramàtiques contextuais basades en patrons, les gramàtiques contextuais de dos nivells i, finalment, les n -gramàtiques contextuais.

Acabarem aquest estudi metodològic amb una caracterització dels llenguatges contextuais mitjançant autòmats (**Capítol 14**). Constitueix, bàsicament, la meua aportació original a l'aparell teòric de les gramàtiques i llenguatges contextuais (principalment pel que fa a la caracterització del autòmats contextuais mitjançant autòmats de contracció – Secció 14.2). Al meu entendre aquesta caracterització obre uns nous horitzonts d'investigació molt amplis i esperançadors dins del camp dels llenguatges contextuais ja que podem pensar en la seva aplicació computacional.

1.2 Marc aplicat

El marc aplicat, pràctic, inclou els capítols 15 – 18. Els capítols 15 – 17 corresponen a caracterització de llenguatges mitjançant gramàtiques contextuais en àmbits no estrictament lingüístics: en llenguatges associats a l'acció dramàtica (**Capítol 15**) com el teatre i el còmic (aquest últim cas és també una aportació original), en models de sistemes d'accions (**Capítol 16**) i, finalment, en llenguatges amb “computació natural” (**Capítol 17**). Ja per acabar aquest bloc aplicat, el **Capítol 18** mostra una caracterització del llenguatge/llengua natural mitjançant un tipus determinat de gramàtiques/llenguatges contextuais: les gramàtiques/llenguatges contextuais amb ús maximal dels seus selectors.

És potser l'aplicació que ha estat menys estudiada i, en canvi, la que segurament té més futur.

L'últim capítol de la tesi (**Capítol 19**) mostra les conclusions finals de la mateixa.

2 Corpus lexicogràfic

He utilitzat el català com a llengua de redacció de la tesi adaptant alguns termes existents només en anglès, francès o romanès al català. Això vol dir que he hagut d'adoptar un seguit de criteris pel que fa a la traducció: he intentat que la traducció s'ajustés al màxim al terme original (anglès, francès o romanès) i a la gramaticalitat catalana; tot i així, normalment incloc també la terminologia original (bàsicament anglesa) al costat de l'adoptada en català o, quan és necessari, quan no existeix una traducció sinònima, he inclòs una nota a peu de pàgina explicativa al respecte. Malgrat traduir els conceptes fonamentals, he mantingut les abreviacions originals (bàsicament angleses) dels mateixos per tal de facilitar el seu reconeixement dins els documents originals i uniformitzar el marc de treball. Per exemple :

Definició 5.1.1 de la pàgina 184. Direm que un llenguatge $L \subseteq V^*$ té la propietat de *pas de derivació externa acotat*, que denotarem per *EBS* (d'“external bounded step”), si existeix una constant p tal que per a cada $x \in L$, $|x| > p$, existeix $y \in L$ tal que $x = u y v$ i $0 < |uv| \leq p$.

Tot i que he consultat diccionaris generals i especialitzats i he demanat consell

pel que fa a certes traduccions, sóc conscient que, tal vegada, algunes de les traduccions adoptades poden no ajustar-se a la idea expresada per la definició original. En qualsevol cas, però, aquesta ha estat una elecció arbitrària per la meua part i, com a tal, susceptible de ser canviada en qualsevol moment.

Per si pot servir d'orientació a continuació enumeraré el material bibliogràfic consultat al respecte:

- Dictionaris generals monolingües:

- Català:

- * *Diccionari de l'Enciclopèdia. Diccionari de la llengua catalana*, Enciclopèdia Catalana S.A., Barcelona, 1994.
 - * *Gran Enciclopèdia catalana*, 18 vol., Enciclopèdia Catalana S.A., Barcelona, 1974 (1er. supl. 1983, 2on. supl. 1989, 3er. supl. 1993).

- Anglès:

- * *Webster's Encyclopedic Unabridged Dictionary of the English Language*, Portland House, New York, 1989.

- Dictionaris especialitzats monolingües:

- Català:

- * *Diccionari de sinònims i antònims*, Ed. Teide, Barcelona, 1970.
 - * *Diccionari d'abreviacions*, Ed. Teide, Barcelona, 1970.

- * *Els verbs catalans conjugats*, Ed. Claret, Barcelona, 1991.
- * *170 vocables d'Informàtica*, PPU, Barcelona, 1989.
- * *Criteris lingüístics*, Servei de Llengua Catalana, Barcelona, 1989.
- * *Proposta de traducció de noms propis*, Servei de Llengua Catalana, Barcelona, 1991.
- * *Proposta d'abreviatures, sigles i símbols*, Servei de Llengua Catalana, Barcelona, 1989.
- Dictionaris generals multilingües:
 - Anglès-català:
 - * *Diccionari anglès-català*, Enciclopèdia Catalana S.A., Barcelona, 1983.
 - Català-anglès, anglès-català:
 - * *Diccionari de gramàtica generativo-transformacional*, PPUAB, Bellaterra, 1991.
 - Català-castellà-francès, castellà-català-francès, francès-català-castellà:
 - * *Diccionari trilingüe*, Ed. Casal i Vall, Andorra, 1988.
 - Francès-romanès:
 - * *Dicționar Francez-Român*, Ed. Gramar, București, 1993.
 - Romanès-francès:
 - * *Dicționar Român-Francez*, Ed. Gramar, București, 1994.

- Dictionaris especialitzats multilingües:
 - Català-castellà-anglès, anglès-català-castellà, castellà-català-anglès:
 - * *Vocabulari matemàtic*, Ed. Eumo, Estudis Universitaris de Vic, 1990.

3 Maquetació

Aquesta tesi s’ha compost mitjançant el sistema $\text{\LaTeX}$ 2_ε amb la classe de document (“documentclass”) **book** de **12pt**. S’han usat els paquets de definicions (“usepackage”) estàndars:

- **a4**: Pel fet que els tipus de fitxers originals del $\text{\LaTeX}$ estan basats en les mesures de pàgina americanes, els usuaris europeus han desenvolupat diferents paquets de definicions que adapten els paràmetres de composició de la pàgina per a mesures mètriques. En aquest cas, el paquet **a4** genera pàgines del tipus DIN-A4 (210mm×290mm).
- **amstex**: L’ $\mathcal{A}\mathcal{M}\mathcal{S}\text{-}\text{\TeX}$ és una extensió del sistema $\text{\TeX}$ per a compondre construccions matemàtiques que dóna una sèrie de marques i entorns que simplifiquen l’escriptura de certes construccions matemàtiques, com ara una matriu de matrius o tota una paraula com a subíndex d’una fórmula, d’acord amb les normes d’estil de l’*American Mathematical Society*. L’activació del paquet de definicions **amstex** dóna com a resul-

tat el sistema $\mathcal{A}\mathcal{M}\mathcal{S}\text{-}\text{\LaTeX}$, un sistema que incorpora la funcionalitat de l' $\mathcal{A}\mathcal{M}\mathcal{S}\text{-}\text{\TeX}$ en el sistema $\text{\LaTeX}$.

- **apalike**: Permet usat el sistema *autor-data* de citació bibliogràfica i l'estil bibliogràfic *chicago*. No obstant, però, l'he modificat per tal d'adaptar-lo al català.
- **qbezier**: Permet al $\text{\LaTeX}$ 2_ε la construcció de corbes de Bézier.
- **theorem**: Ofereix una extensió del mecanisme d'elaboració de teoremes del $\text{\LaTeX}$ i permet la manipulació de la composició de teoremes mitjançant un estil determinat.

La caixa de text del document té unes mesures de 140mm × 180mm, amb un interlineat d'1.5 línies i una separació entre paràgrafs de 4mm (amb un queixal d'entrada de 7mm).

Els gràfics i dibuixos més senzill estan fets usant $\text{\TeX}cad$ v. 2.8 i els més elaborats usant $CorelDraw$ v. 5.0 per a Windows.

La bibliografia està confeccionada usant el gestor de bases de dades bibliogràfiques $Bib\text{\TeX}$ usant el paquet de definicions **apalike**, que permet emprar l'estil bibliogràfic *chicago*, amb el sistema *autor-data* de citació bibliogràfica – convenientment modificat per al català.

NOTACIÓ GENERAL

Usarem la següent notació:

1. V = vocabulari,
2. $x \in A$ = x és un element d' A o x pertany a A ,
3. $x \notin A$ = x no és un element d' A o x no pertany a A ,
4. $A \subseteq B$ = A està inclòs a B , A és un subconjunt de B ,
5. $A \subset B$ = A inclòs estrictament a B , A és un subconjunt propi de B ,
6. 2^X , $\mathcal{P}(X)$ = el conjunt de parts d' X , el conjunt potència d' X , el conjunt de subconjunts d' X ,
7. V^* = el monoide lliure generat per V , conjunt de totes les cadenes sobre V , llenguatge universal sobre V ,
8. $V^+ = V^* - \{\lambda\}$,
9. $\text{card}(A)$ = nombre d'elements d' A ,
10. λ = la cadena buida,
11. $|x|$ = la longitud de la cadena x ,
12. $|x|_a$ = el nombre d'ocurrències del símbol $a \in V$ en la cadena $x \in V^*$,

13. $|x|_U$ = el nombre d'ocurrències de símbols de $U \subseteq V$ en la cadena $x \in V^*$,
14. $length(L)$ = el conjunt longitud d' L ,
15. $L \subseteq V^*$ = qualsevol llenguatge sobre V ,
16. $L_1 \cup L_2$ = la unió d' L_1 i d' L_2 ,
17. $\bigcup_{i=1}^n L_i$ = la unió finita d' L_i , $1 \leq i \leq n$,
18. $\bigcup_{i=1}^{\infty} L_i = \bigcup_{i \geq 1} L_i$ = la unió infinita d' L_i , $i \geq 1$,
19. $L_1 \cap L_2$ = la intersecció d' L_1 i d' L_2 ,
20. $\bigcap_{i=1}^n L_i$ = la intersecció finita d' L_i , $1 \leq i \leq n$,
21. $\bigcap_{i=1}^{\infty} L_i = \bigcap_{i \geq 1} L_i$ = la intersecció infinita d' A_i ,
22. $L_1 - L_2$ = la diferència entre L_1 i L_2 ,
23. L' = el complementari del conjunt L ,
24. $L_1 \cdot L_2$ = la concatenació, multiplicació, multiplicació directa o producte d' L_1 per L_2 ,
25. $L^k = \underbrace{L \cdot L \cdot \dots \cdot L}_k$,
26. $L^0 = \{\lambda\}$
27. $mi(x)$ = la imatge mirall d' x ,

- 28. $Sub(L)$ = el conjunt de subcadenaes de cadenaes d' L ,
- 29. $Pre(L)$ = el conjunt de prefixos de cadenaes d' L ,
- 30. $Suf(L)$ = el conjunt de sufixos de cadenaes d' L ,
- 31. $PPref(L)$ = el conjunt de prefixos propis de cadenaes d' L ,
- 32. $PSub(L)$ = el conjunt de subcadenaes pròpies de cadenaes d' L ,
- 33. $PSuf(L)$ = el conjunt de sufixos propis de cadenaes d' L ,
- 34. $L^* = \bigcup_{i \geq 0} L^i$,
- 35. $L^+ = \bigcup_{i \geq 1} L^i$,
- 36. $L_1 \uplus L_2$ = la barreja dels llenguatges L_1 i L_2 ,
- 37. $L_1 \setminus L_2$ = el quocient per l'esquerra d' L_1 respecte a L_2 ,
- 38. L_1 / L_2 = el quocient per la dreta d' L_1 respecte a L_2 ,
- 39. $\partial_y^e(L_1)$ = la derivada per l'esquerra d' $L_1 \subseteq V^*$ respecte a $y \in V^*$,
- 40. $\partial_y^d(L_1)$ = les derivades per la dreta d' $L_1 \subseteq V^*$ respecte a $y \in V^*$,
- 41. $pdret(A)$ = la prolongació a la dreta d' A ,
- 42. $cl(X)$: = la clausura d' X
- 43. $L_{left}(G)$ o bé $L_{lf}(G)$ = derivació més-a-l'esquerra,

- 44. REG, LIN, CF, CS, RE = les famílies de gramàtiques/llenguatges regulars, lineals, lliures o independents del context, sensibles o dependents del context i recursivament enumerables, respectivament,
- 45. FIN = la família de gramàtiques/llenguatges finits,
- 46. LIN_1 = la família de gramàtiques/llenguatges lineals minimal (gramàtiques lineals amb només un símbol no terminal),
- 47. Ψ_V = la funció de Parikh associada al vocabulari V ,
- 48. $alph(L)$ = el conjunt de símbols que apareixen en les cadenes d' L ,
- 49. $D_{a,b}$ = el llenguatge de Dyck sobre el vocabulari $\{a, b\}$,
- 50. EC, IC, ECC, ICC, TC = les famílies de gramàtiques/llenguatges contextuais externes sense selecció, internes sense selecció, externes amb selecció, internes amb selecció i totals, respectivament,
- 51. $X(F)$, sent X qualsevol gramàtica/llenguatge contextuais usual i $F \in \{REG, LIN, CF, CS, RE\}$ = les famílies de gramàtiques/llenguatges contextuais del tipus X amb selecció F ,
- 52. $X_c(F)$ = les famílies de gramàtiques/llenguatges contextuais computables,
- 53. $X_\alpha(F)$, $\alpha \in \{ml, mg, Ml, Mg\}$ = les famílies de gramàtiques/llenguatges contextuais amb ús minimal local (ml), minimal global (mg), maximal local (Ml) i maximal global (Mg) dels selectors,

- 54. $X_d(F)$ = les famílies de gramàtiques/llenguatges contextuais amb contextos esborrats,
- 55. $1X$ i $11X$ = les famílies de gramàtiques/llenguatges contextuais (unilaterals) amb només contextos de la forma (λ, v) i les famílies de gramàtiques/llenguatges on només son permesos els contextos de la dreta (λ, v) i els contextos de l'esquerra (u, λ) (però, no així, els contextos (u, v) amb u i v no-buits); respectivament,
- 56. $DX(F)$ = les famílies de gramàtiques/llenguatges contextuais deterministes,
- 57. $X_\alpha(F)$, $\alpha \in \{lf, pr, lfM, prM\}$ = les famílies de gramàtiques/llenguatges contextuais, amb derivació més a l'esquerra (lf), usant un selector com a prefix en la derivació (pr), amb derivació màxima a l'esquerra i ús maximal dels selectors (lfM) i, finalment, usant un selector com a prefix en la derivació en mode maximal (prM),
- 58. $X_p(F)$ = les famílies de gramàtiques/llenguatges contextuais amb derivació paral·lela,
- 59. $BX_\alpha(F)$ = les famílies de gramàtiques/llenguatges contextuais bloquejats,
- 60. $X_{df}(F)$ = les famílies de gramàtiques/llenguatges contextuais amb primera derivació,

- 61. $X_{mk}(F)$ = les famílies de gramàtiques/llenguatges contextuais marcades,
- 62. $MX(F)$ = les famílies de gramàtiques/llenguatges contextuais matri-cials,
- 63. $PX(F)$ = les famílies de gramàtiques/llenguatges contextuais progra-mades,
- 64. $KX(F)$ = les famílies de gramàtiques/llenguatges contextuais contro-lades regularment,
- 65. $QX(F)$ = les famílies de gramàtiques/llenguatges contextuais Q -difuses (o Q -borroses),
- 66. $X_\infty(F)$: = les famílies de gramàtiques/llenguatges contextuais amb in-finits contextos,
- 67. $P_{atrons}IC_\alpha$, $\alpha \in \{f, ml, mg, Ml, Mg\}$ = les famílies de gramàtiques/llen-guatges contextuais amb patrons,
- 68. $TLC(F_1, F_2)$ = les famílies de gramàtiques/llenguatges contextuais de dos nivells,
- 69. X_n : = les famílies de n -gramàtiques/llenguatges contextuais, $n \geq 1$.

Capítol 1

Els models algebraics analítics del llenguatge

Els models algebraics analítics del llenguatge van néixer als anys 60. Les seves fonts eren lingüístiques – principalment relacionades amb el Cercle Lingüístic de Praga i amb la Lingüística Descriptiva Americana –, matemàtiques – relacionades amb la teoria de semigrups, especialment amb la teoria de semigrups lliures –, lògiques – Escola Polonesa – i computacionals – traducció automàtica de llenguatges. *Analític* s’ha d’entendre aquí en contrast amb *generatiu* – en el sentit chomskià de la paraula. Els models analítics estan relacionats amb les propietats internes, intrínseques dels llenguatges – com ara, les relacions entre les subcadenaes d’una cadena donada, la possibilitat de substituir una cadena per una altra en un context donat, etc. –, mentre que els models generatius estan relacionats amb diversos tipus de mecanismes, concebuts com a sistemes formals, que generen o reconeixen un llenguatge. En molts casos,

aquests mecanismes usen alguns elements auxiliars a mode de metasímbols de significat categorial – veure, p.e., l’alfabet o vocabulari no-terminal en les gramàtiques de Chomsky de la Definició 3.3.3 de la pàg. 112. Per aquesta raó, els models generatius es consideren externs, en contraposició a l’aproximació analítica, que es considera interna. Normalment, un mecanisme generatiu es concep com un algorisme i, per aquesta raó, els models generatius són, en molts casos, algorítmics, a diferència dels models algebraics analítics que emfatitzen l’aspecte axiomàtic – deixarem de banda els models analítics estadístics. Però aquest no és un tret general; algunes gramàtiques generatives generalitzades poden perdre la qualitat d’un algorisme, mentre que alguns models analítics poden aproximar-se a models generatius. La frontera entre analític i generatiu esdevé algunes vegades molt difusa, però, en qualsevol cas, aquests dos aspectes diferents del llenguatge estan sempre interactuant. Més endavant veurem alguns exemples al respecte.

A diferència dels models generatius, que tenen una pobra tradició en la investigació lingüística dels últims dos segles, els models analítics tenen uns arrels molt riques en l’estudi dels llenguatges naturals. Contràriament, els models generatius chomskians han demostrat ser molt útils més enllà dels llenguatges naturals – p.e., en l’estudi de la sintaxi i la semàntica dels llenguatges de programació –, a diferència de l’aproximació analítica, que ha tingut poca rellevància en el cas dels llenguatges artificials – amb algunes excepcions significatives.

Una interessant discontinuïtat apareix entre la història de les aproximacions generativa i analítica. Chomsky va desenvolupar la seva teoria generativista com una polèmica reacció a l'Estructuralisme Americà – representat principalment per l'escola lingüística descriptiva, distribucional, de Z. Harris – malgrat el fet que Chomsky fos un deixeble del mateix Harris. Per altra banda, Chomsky invoca tradicions no americanes com és el cas de la gramàtica Panini India i la lingüística Cartesiana. Pel que fa als mitjans tècnics usats en la seva aproximació, Chomsky usa explícitament els sistemes Combinatoris de Post, però hom considera que es basà, essencialment – malgrat el fet que ell mai no ho va reconèixer en els seus escrits –, en la teoria de Hilbert relativa als sistemes formals. De fet, una gramàtica de Chomsky és un cas particular de sistema formal de Hilbert. Si Chomsky desenvolupa tradicions típicament no americanes, els models analítics nascuts a l'est d'Europa els anys 60 es basen en les idees de la Lingüística Descriptiva Americana, on la informació contextual és la principal, si és que no és la única, en ser considerada rellevant.

La primera visió sistemàtica relativa als models analítics va ser presentada als monogràfics (Marcus, 1967a; Marcus, 1967d) per Solomon Marcus, però ambdós han quedat desfasats per la gran quantitat de noves aportacions sorgides en aquests últims trenta anys. Desafortunadament, aquesta línia d'investigació va ser, i encara ho és, ignorada per molts investigadors en lingüística formal i computacional. No obstant, alguns models algebraics analítics tenen un enorme potencial per tal de capturar diversos fenòmens de la sintaxi i la semàntica dels llenguatges naturals, en la representació del coneix-

xement i en altres aspectes diversos de la intel·ligència artificial, en el camp de les accions humanes, en diversos tipus de competència humana, en llenguatges de programació, etc.

Per tal de donar una visió la més àmplia possible dels models algebraics analítics ens fixarem en les següents àrees de recerca:

- la fonologia i la segmentació,
- les relacions de dominació,
- els operadors de clausura i les connexions de Galois,
- les parts de l'oració, el gènere, el cas, i la tipologia lingüística,
- l'espai contextual i els models topològics.

Altres aspectes com la relació amb altres teories matemàtiques, les configuracions sintàctiques, la dependència sintàctica i la projectivitat, les interaccions entre aspectes analítics i generatius, la praxi del llenguatge o la universalitat dels models lingüístics no seran considerats.

1.1 La Fonologia i la segmentació

Destacarem tres línies d'investigació. La primera d'elles relacionada amb l'aproximació lògica, que tracta de reconstruir el camí des de la naturalesa

contínua de la realitat fonètica fins a la naturalesa discreta de les representacions fonològiques. És el cas de les investigacions de T. Batog a (Batog, 1967), basades en la mereologia de Lesniewski.

La segona línia d'investigació desenvolupa les idees de Harris, considerant el fonema com una classe de cadenes sonores. Una investigació típica segons aquesta aproximació va ser donada per S. Kanger a (Kanger, 1964). Una investigació posterior d'aquest model – veure el segon capítol de (Marcus, 1967d) – mostra el sorprenent i delicat fet que una classe de cadenes sonores que defineix un fonema no és una classe d'equivalència – la propietat transitiva no se satisfà.

Finalment, la tercera línia d'investigació està relacionada amb la manera com Jakobson considera un fonema com una classe de trets binaris distintius. A partir del conjunt de trets acústics, o articulatoris, d'un so, hem de seleccionar aquells trets que tinguin una rellevància lingüística. A partir d'un aspecte purament paradigmàtic hem de passar a una estratègia que inclogui els eixos combinatoris, p.e., la distinció entre cadenes ben formades i cadenes mal formades. El model que S. Marcus va presentar al respecte – sintetitzat a (Marcus, 1967d) – va ser després desenvolupat i millorat per Nebeský (Nebeský, 1966b; Nebeský, 1966a), Mazzaroli (Mazzaroli, 1969), Brainerd (Brainerd, 1969; Brainerd, 1971; Brainerd, 1974) i Kortlandt (Kortlandt, 1972; Kortlandt, 1973). Per Nebeský, un conjunt de trets rellevants d'una unitat de llenguatge, un so, ha d'acomplir aquests tres requeriments:

- 1) Ha de ser capaç, en tots els casos, de diferenciar dues unitats o sons, dins el conjunt de les cadenes d'unitats o sons ben formades; i on és necessari diferenciar també, entre cada dues cadenes ben formades, en el cas que es diferenciïn mitjançant els conjunts de tots els trets de les unitats.
- 2) Els conjunts han de posseir alguna propietat "raonable";
- 3) Pel què fa als requeriments 1) i 2), el sistema de conjunts de trets rellevants dels sons, així com els conjunts mateixos, han de tenir el menor nombre possible d'elements.

Els requeriments 1) i 3) no es compleixen en el model de Marcus. Aquesta mancança és transgredida per Nebeský. Brainerd, per la seva part, generalitza el model fonològic de Marcus en un sistema amb aplicacions en altres nivells lingüístics, més enllà de la fonologia: el sistema fonètic potencial s'exté a un concepte general de sistema ètic – incloent, p.e., els estudis fonètics i els estudis d'afinitat. Brainerd mostra que, donat un sistema ètic arbitrari, sempre és possible reduir-lo mitjançant el colapse dels trets, d'una manera natural, en un sistema on cada tret és, o bé redundant, o bé marcat, o bé forma part d'un parell contrastiu. Aquesta definició de fonema requereix que un tret sigui contrastiu respecte a un altre per tal que sigui rellevant, o distintiu, per a algun fonema. Si s'usen massa trets, aleshores no hi ha suficients parells contrastius de trets i, per tant, no hi ha suficients candidats pel paper de trets rellevants. A partir d'un sistema ètic concebut sota la forma d'una quàdrupla $\langle F, V, f, P \rangle$, on F és el tret de conjunt finit, V és el conjunt finit designat

(disjunt amb F), f és la funció d'assignació definida a V que pren valors al conjunt de subconjunts d' F – que, d'ara endavant, denotarem per $\mathcal{P}(F)$ o bé 2^F – i, finalment, P és una partició $\{F_1, F_2, \dots, F_k\}$ d' F tal que per a cada $1 \leq i \leq k$ i cada x de V la intersecció entre $f(x)$ i F_i conté com a màxim un element; Brainerd obté un sistema ètic parcial afegint un sistema sinònim $\langle W, L, s \rangle$, on W és un vocabulari finit, L és un llenguatge sobre W i s és una relació d'equivalència en L . Com a sistema sinònim particular, Brainerd obté el sistema de fonemes consonàntics de l'anglès nord-americà, els pronoms de l'anglès i el de les notacions decimals pels números naturals del francès-anglès bilingüe.

Kortlandt comença la seva anàlisi fonèmica amb un sistema $P = [K, R, S]$, on K és el conjunt de possibles dimensions (anterior-posterior, oral-nasal, etc.) en el contínuum dels sons que poden ser produïts pels òrgans humans de la parla, R és una relació entre seqüències de sons que reflecteix la possibilitat dels informants de reconèixer i identificar formes lingüístiques, p.e. paraules, que pertanyin al seu propi llenguatge – els elements de K són conjunts de trets homogenis en el sentit de (Marcus, 1989b). Kortlandt és també l'autor de la més comprensiva presentació crítica de diversos intents de l'aproximació matemàtica al fonema (Kortlandt, 1972), mitjançant una discussió molt acurada de la hipòtesi fonamental de Marcus: si es reemplaça un so x en una seqüència certificada per un so absolutament equivalent a ell, aleshores la nova seqüència obtinguda és també certificada. Es consideren quatre interpretacions diferents del conjunt S de cadenes certificades:

- 1) S és el conjunt de seqüències certificades de fet (S és finit);
- 2) S conté, no només les seqüències certificades de fet, sinó també les noves entitats observades les quals no poden ser distingides de les seqüències certificades de fet, ja que ambdues tenen les mateixes característiques físiques;
- 3) S és el conjunt de seqüències de sons abstractes;
- 4) les seqüències que pertanyen a S són establertes per l'ús d'un informant.

Kortland va parlar molta atenció a alguns dels fets més controvertits en la fonètica teòrica, com són el fet de la significació, els trets rellevants i els trets redundants, la neutralització i l'arxifonema i la indispensabilitat de la relació de sinonímia.

L'estudi grafo-teòric dels grups consonàntics inicials i finals va ser iniciat per Marcus i Vasiliu (Marcus i Vasiliu, 1960a) i desenvolupat i/o usat posteriorment per Horecký (Horecký, 1964), Lekomceva (Lekomceva, 1966), Lekomcev (Lecomcev, 1966), Piotrovskii (Piotrovskii, 1966), Revzin (Revzin, 1966), Saumjan (Šaumjan, 1968) i Afendras (Afendras, 1969). Un model algebraic de la segmentació fonètica va ser proposat per Naftanail (Naftanail, 1976). Un sistema automàtic de segmentació morfològica va ser proposat per Cazimir (Cazimir, 1976). El concepte de morfologia regular usat per Marcus (Marcus, 1967d) va ser generalitzat per Schweiger (Schweiger, 1972) mitjançant el concepte de profunditat dels prefixos d'un llenguatge L definit de la següent

manera: considerem l'operador prefix $pL = \{x \wedge y \mid x <> y, x, y \in L\}$ on $x \wedge y$ és el prefix comú d' x i y . Si definim

$$p^0 L = L, \text{ i } p^1 L = pL, \dots, p^{n+1} L = p(p^n L),$$

aleshores, la profunditat d'un prefix, o $d(L)$, d'un llenguatge L és el menor n tal que $p^n L = 0$. La morfologia regular es correspon amb $d(L) = 2$. Similarmet, podem definir l'operador sufix i la profunditat d'un sufix. Posteriors investigacions van ser fetes per Ciobotaru (Ciobotaru, 1974), Orman (Orman, 1970) i Klein (Klein, 1971).

1.2 Les relacions de dominació

L'objectiu d'aquesta àrea de recerca és el d'expressar diversos tipus d'ambigüitat contextual, a diferents nivells lingüístics, i modelitzar el concepte general de categoria gramatical. La relació de dominació $\rightarrow$, considerada per Dobrušin (Dobrušin, 1957) al nivell del vocabulari i per les categories gramaticals elementals, va ser extesa a llenguatges i categories gramaticals generals – Capítol 5 de (Marcus, 1967d). Donats x, y sobre un vocabulari V , i el llenguatge L sobre V , tenim que $x \rightarrow y (L)$ si, qualsevol context (u, v) que accepta x a L , accepta també y a L – i.e., si $uxv \in L$ aleshores també $uyv \in L$. Això vol dir que l'ambigüitat contextual d' x respecte a L no és més gran que l'ambigüitat contextual d' y . La classe distribucional $D(x)$ d' x és el conjunt $\{y \mid y \rightarrow x (L), x \rightarrow y (L)\}$. La categoria gramatical $G(X)$ generada pel

conjunt X de cadenes sobre V es defineix pel conjunt $\{y \mid x \rightarrow y (L)\}$ per a qualsevol $x \in X$. Si X és una classe distribucional, aleshores $G(X)$ és una categoria gramatical elemental. Qualsevol $G(X)$ és la unió d'alguna categoria gramatical elemental i sempre tenim que $G(G(X)) = G(X)$. Existeix una gran bibliografia relacionada amb les categories gramaticals. Degut principalment a raons lingüístiques, aquest model presentat anteriorment va ser generalitzat en diverses direccions, pel tal d'incloure una gran varietat de categories gramaticals concretes i tipus de sincretismes. Per això, Nebeský (Nebeský, 1965) introdueix el que va anomenar un “reemplaçament condicional de cadenes, o paraules”. Donades dues cadenes x i y sobre V i dos llenguatges M i N sobre V , escriurem $x \rightarrow y(M/N)$ quan, per a qualsevol context (u, v) que accepta x i y a N , si $uxv \in M$, aleshores també $uyv \in M$. Anomenarem aquest tipus de dominació la “dominació de Nebeský”. En el cas particular que $N = V^*$ (el llenguatge universal sobre V), obtenim la relació de dominació ordinària.

Un altre tipus de relació de dominació respecte a dos llenguatges va ser proposada per Franclova (Franclova, 1969). La dominació de Franclova $x \rightarrow y(M//N)$ es defineix mitjançant la següent condició: qualsevol context (u, v) que accepta x a M accepta y a N . En el cas particular que $M = N$ obtenim la relació de dominació usual. Si V és un conjunt finit de símptomes i si M és un conjunt de cadenes sobre V , interpretat com a síndromes de la mateixa malaltia, mentre que N és un conjunt de cadenes sobre V que defineix una altra malaltia, aleshores la dominació de Franclova significa que qualsevol context crític que emplaça x en la malaltia M també emplaçarà y en la malaltia N . Una

altra possible interpretació estaria relacionada amb diversos nivells possibles de gramaticalitat que succeeix, p.e., en l'estudi del llenguatge poètic. Començant per un conjunt L_1 de cadenes ben formades sobre V , i afegint a L_1 totes les cadenes que són subcadenes de les cadenes d' L_1 , obtenim un conjunt més gran L_2 , al qual podem afegir-li les cadenes paràsites, i.e., les cadenes de V^* que no pertanyen a L_2 . La poesia sovint es basa en algunes substitucions que ens traslladen des d'un univers estàndard com pot ser L_1 cap a un univers més gran com pot ser L_2 , on la semi-gramaticalitat és acceptada i algunes vegades som transferits cap a un univers absurd localitzat en algun lloc d' L_3 .

Un altre interessant tipus de relació de dominació va ser introduïda, independentment, per Sestier (Sestier, 1961) i Kunze (Kunze, 1967). Anomenarem a aquesta dominació la dominació de Sestier-Kunze. Per a qualsevol llenguatge L sobre V , definim l'operador contextual α_L , que associa a cada conjunt X de cadenes sobre V el conjunt $\alpha_L(X)$, format per tots els contextos (u, v) que accepten a L qualsevol cadena d' X – i.e.. per a qualsevol $x \in X$ la cadena $uxv \in L$. Donats dos conjunts de cadenes X i Y sobre V , la dominació de Sestier-Kunze, representada per $X \rightarrow Y(SK)$ es defineix mitjançant el requeriment sobre $\alpha(X)$ que estigui contingut a $\alpha(Y)$. Per això, a diferència dels altres tipus de dominació definits fins ara que estaven definits en primer lloc per a cadenes individuals i només en posteriors passos per a conjunts de cadenes, la SK-dominació està definida directament per a conjunts i perd l'interès tan aviat com aquests conjunts degeneren en cadenes individuals, pel fet que en aquest cas és idèntic a la dominació usual.

De fet, la terminologia usada aquí és diferent a la utilitzada per Sestier, guiat per motivacions purament algebraiques, i Kunze, les raons principals del qual estaven relacionades amb algunes situacions específiques de l'alemany, on conjunts com $\{ich, du\}$ i l'estatus dels pronoms personals no podien ser descrits mitjançant la dominació de Dobrušin, però en canvi sí podien ser capturats mitjançant l'aproximació de Sestier-Kunze.

Trybulec (Trybulec, 1967b; Trybulec, 1967a) va proposar un altre tipus de dominació entre cadenes individuals, on la partició P del vocabulari és rellevant. Escrivem $x \rightarrow y(P)$ i l'anomenarem la dominació de Trybulec respecte a P o P -dominació si, per a qualsevol context (u, v) que accepta x a L , amb $u = a_1 a_2 \dots a_{i-1}$, $v = a_{i+1} \dots a_n$, existeix un context (u', v') , amb $u' = a'_1 a'_2 \dots a'_{i-1}$, $v' = a'_{i+1} \dots a'_n$, tal que $a'_1 \in P(a_1), \dots, a'_{i-1} \in P(a_{i-1}), a'_{i+1} \in P(a_{i+1}), \dots, a'_n \in P(a_n)$ i $u' y v' \in L$.

L'aproximació de Marcus a la dominació de Kunze va ser donada a (Marcus, 1969c). Altres diverses investigacions estaven relacionades amb combinacions de dos tipus diferents de relacions de dominació: la dominació de Kunze-Franclova – per Cazimir (Cazimir, 1975c; Cazimir, 1975a; Cazimir, 1975b) –, la dominació Trybulec-Kunze – per Dascălu (Dascălu, 1975) – i la dominació Franclova-Trybulec – per Fertig (Fertig, 1974). Altres estudis relacionats amb la dominació de Trybulec van ser donats per Popescu (Popescu, 1972) i Schwartz (Schwartz, 1972) – amb una anàlisi comparativa de les dominacions de Dobrušin i Trybulec. S'observa que la dominació de Dobrušin entre

conjunts de cadenes implica la seva dominació de Kunze, pero l'invers no és cert – la dominació de Kunze és un altre nom que es dona a la dominació de Sestier-Kunze.

Després de combinar en parells algunes de les relacions de dominació considerades, ens podem preguntar si seria possible obtenir un marc de treball més comprensiu on tinguéssim la possibilitat d'obtenir com a situació particular cadascun dels tipus de relacions de dominació obtingudes. Una resposta interessant a aquesta qüestió va ser proposada per Schwarz (Schwartz, 1976). Considerem les funcions $A : M \longrightarrow 2^N$ i $A^{-1} : N \longrightarrow 2^M$. Tenim l'equivalència entre $A^{-1}(y) = \{x \in M \mid y \in A(x)\}$ i $x \in A^{-1}(y) \iff y \in A(x)$. Tenim també l'equivalència entre $(A^{-1})^{-1}(x) = \{y \in N \mid x \in A^{-1}(y)\}$ i $y \in (A^{-1})^{-1}(x) \iff x \in A^{-1}(y)$. Tenim que $y \in (A^{-1})^{-1}(x) \iff y \in A(x)$ i $(A^{-1})^{-1} = A$. Considerem ara dues funcions A i B entre M i el conjunt de subconjunts d' N (2^N). Definim les funcions $A \cup B$ i $A \cap B$ entre M i 2^N mitjançant $(A \cup B)(x) = A(x) \cup B(x)$ i $(A \cap B)(x) = A(x) \cap B(x)$. Definim també la funció $A' : M \longrightarrow 2^N$ mitjançant $A'(x) = N - A(x)$ i podem afirmar que la família $\mathcal{A} = \{A \mid A : M \longrightarrow 2^N\}$ és una àlgebra booleana. Aleshores tenim que $(A \cup B)^{-1} = A^{-1} \cup B^{-1}$, $(A \cap B)^{-1}(x) = A^{-1} \cap B^{-1}$, i $(A')^{-1} = (A^{-1})'$. Mitjançant $A < B$ significarem que $A(x) \subset A(x)$ per a qualsevol $x \in M$. La relació $<$ és un ordre parcial a $\mathcal{A}$. Tenim que $A < B \iff A^{-1} < B^{-1}$.

Podem estendre de dues maneres l'operador A . Considerem els operadors

A_s i A_p entre 2^M i 2^N , definits de la següent manera:

$$A_s(X) = \bigcup_{x \in X} A(x) \text{ i } A_p(X) = \bigcap_{x \in X} A(x).$$

Considerem ara dos operadors contextuais α i β entre 2^{V^*} (el llenguatge universal sobre el vocabulari finit V) i $2^{v^* \times V^*}$. La relació de dominació introduïda pels operadors α i β , denotada per $\rightarrow(\alpha, \beta)$, es defineix de la següent manera: si X i Y són conjunts de cadenes sobre V i si $\alpha(X) \subset \beta(Y)$, aleshores $X \rightarrow Y(\alpha, \beta)$.

Considerem ara les restriccions $A = \alpha/V^*$ i $B = \beta/V^*$. Si impossem que $\alpha = A_s$ o $\alpha = A_p$ o bé $\beta = A_s$ o $\beta = A_p$ obtenim, per a un parell (A, B) d'operadors, quatre tipus de relacions de dominació: $\rightarrow(A_s, B_p)$, $\rightarrow(A_p, B_p)$, $\rightarrow(A_s, B_s)$, $\rightarrow(A_p, B_s)$.

La primera de les relacions de dominació implica les altres tres relacions de dominació, mentre que l'última dominació és implicada per la segona i la tercera.

1.3 Clausura i connexions de Galois

L'interès pels operadors de clausura en lingüística algebraica va començar amb els treballs pioners de Kulagina (Kulagina, 1958), on s'introdueix el concepte de derivació d'una partició. Donada una partició P d'un vocabulari V , denotem per $P(a)$ el terme de P que conté l'element a de V . Per a $a, b \in V$, els conjunts

$P(a)$ i $P(b)$ són o bé idèntics o bé disjunts i, per tant, el conjunt $\{P(a) \mid a \in V\}$ és finit i pot jugar el paper d'un nou vocabulari $P(A)$. Sobre aquest vocabulari considerarem que una cadena $P(a_1)P(a_2) \dots P(a_n)$ és ben formada, respecte al llenguatge L sobre V , si existeix una cadena $x = b_1b_2 \dots b_n$ a L tal que $b_i \in P(a_i)$ per a qualsevol i entre 1 i n . En aquesta situació, podem definir una relació de P -dominació (respecte a L) entre dos elements P_1 i P_2 de $P(A)$: P_1 P -domina P_2 si, per a dues cadenes qualsevol $\mathcal{P}_3$ i $\mathcal{P}_4$ tal que $\mathcal{P}_3P_1\mathcal{P}_4$ és ben formada (respecte a L), la cadena $\mathcal{P}_3P_2\mathcal{P}_4$ és també ben formada. Començant per la partició P de V , definim la partició derivada P' de la següent manera: per a qualsevol $a \in A$, $P'(a)$ és la unió de tots els termes $P(b)$ que P -dominen $P(a)$ i són P -dominats per $P(a)$ (respecte a L). Per això, P és sempre més fina que P' . Està provat que la segona derivada $P'' = (P')'$ és sempre idèntica a P' . Šreider (Šreider, 1970; Šreider, 1973) prova que les propietats:

- 1) $P \subset P'$;
- 2) si $P \subset Q \subset P'$, aleshores $Q' = P'$;
- 3) si $P \subset Q$, $P'Q = PQ'$, aleshores $P' \subset Q'$;

que són satisfetes per l'operador derivada, no el caracteritzen i proposa un altre sistema d'axiomes. En qualsevol cas, l'operador derivada és un operador de clausura en el sentit algebraic de la paraula – veure també (Vasilevskiĭ, 1973).

El següent operador de clausura important en l'estudi del llenguatge va ser introduït per Sestier (Sestier, 1961). Donat un conjunt X de cadenes sobre

V , definim $\alpha_L(X)$ com el conjunt de tots els contextos que accepten qualsevol cadena d' X del llenguatge L sobre V . Donat un conjunt $\mathcal{C}$ de contextos sobre V , definim $\alpha_L^{-1}(\mathcal{C})$ com el conjunt de totes les cadenes d' L per a cada context de $\mathcal{C}$. Considerem ara el que anomenem la clausura de Sestier d' X :

$$\varphi_L(X) = \alpha_L^{-1}(\alpha_L(X)).$$

Es pot provar que:

- i) per a qualsevol conjunt X de cadenes sobre V i per a qualsevol llenguatge L sobre V , X està contingut (algunes vegades estrictament) a $\varphi_L(X)$;
- ii) si X està contingut a Y , aleshores $\varphi_L(X)$ està contingut a $\varphi_L(Y)$;
- iii) es compleix $\varphi_L(\varphi_L(X)) = \varphi_L(X)$.

Per tant, $\varphi_L(X)$ és un operador de clausura. És l'operador contextual més important i natural considerat en la lingüística algebraica. Sestier accentua la connexió de Galois existent entre cadenes i contextos, entre α i α^{-1} (el subíndex L pot ser omès si no hi ha cap confusió possible).

Aleshores se'ns presenta una qüestió natural: quina és la relació entre $\varphi_L(X)$ i $G(X)$ – la categoria gramatical generada per X ? De fet, la definició d'una categoria gramatical, donada en l'apartat anterior, és normalment reforçada amb la següent condició: el generador X de la categoria $G(X)$ hauria de ser inicial, i.e., no pot existir cap x que no pertanyi a X , que domini qualsevol $y \in X$. No obstant, per tal de facilitar la següent discussió, ometrem

aquesta condició. Hi ha dos teoremes que estableixen les relacions bàsiques entre φ i G : el primer afirma que, per a qualsevol X , $G(X)$ està continguda (algunes vegades estrictament) a $\varphi(X)$; el segon ens diu que, per a qualsevol X , existeix un Y tal que $\varphi(X) = G(Y)$; en altres paraules, qualsevol clausura de Sestier és una categoria gramatical, però l'invers no és cert. Podem afegir-hi un tercer teorema: tenim $\varphi(X) = \varphi(Y) \iff \alpha(X) = \alpha(Y)$. Un conjunt X de cadenes és una clausura de Sestier si, i només si, $X = \varphi(X)$. A la vista de tots aquests teoremes, les clausures de Sestier són un dels millors models de categories gramaticals en els llenguatges naturals, com s'observa principalment per a l'alemany – veure (Kunze i Priess, 1970).

Podem definir altres operadors contextuais relacionats. Denotem per $\beta_L(X)$ el conjunt de contextos que accepten, com a mínim, una cadena d' X i per $\beta_L^{-1}(X)$ el conjunt de cadenes acceptades per, com a mínim, un context de $\mathcal{C}$. Obtenim tres nous operadors contextuais: $\Psi(X) = \beta^{-1}(\beta(X))$, $\lambda(X) = \beta^{-1}(\alpha(X))$, $\mu(X) = \alpha^{-1}(\beta(X))$. D'entre els cinc operadors $G, \varphi, \Psi, \lambda, \mu$, els únics parells permutables són (G, φ) , (G, λ) , (G, μ) , (φ, λ) , i.e., $G(\varphi(X)) = \varphi(G(X))$, $G(\lambda(X)) = \lambda(G(X))$, etc. La diferència entre $G(X)$ i $\varphi(X)$ és que $G(X)$ s'obté a partir d' X mitjançant la dominació de Dobrušin, mentre que $\varphi(X)$ s'obté d' X mitjançant la dominació de Sestier-Kunze. El “privilegi” de les categories gramaticals és el d'abraçar tots els altres quatre operadors: per a qualsevol X existeixen els conjunts Y, Z, K, W tal que $\varphi(X) = G(Y)$, $\Psi(X) = G(Z)$, $\lambda(X) = G(K)$, $\mu(X) = G(W)$. Per a una consulta més exhaustiva de totes aquestes propietats i d'altres no explicitades, relacionades

amb les relacions entre diversos operadors contextuals i les categories gramaticals, per una banda, i els operadors de clausura i les connexions de Galois, per altra banda, hom es pot remetre a Marcus (Marcus, 1967b; Marcus, 1967c; Marcus, 1968; Marcus, 1969c), Mayoh (Mayoh, 1967), Toma-Marcu (Toma-Marcu, 1969), Raischi (Raischi, 1970), Popescu (Popescu, 1976), Holban (Holban, 1972), Crăciun (Crăciun, 1972; Crăciun, 1973), Revzin (Revzin, 1977), Schwarz (Schwartz, 1975), Raischi (Raischi, 1973), Pană (Pană, 1975), Coardos (Coardos, 1976) i Petrescu (Petrescu, 1977).

1.4 Les parts de l'oració i altres aspectes

A continuació veurem alguns aspectes relacionats amb les parts de l'oració, el gènere, el cas i la tipologia lingüística. El primer model matemàtic de parts de l'oració (nom, pronom, verb, etc.) va ser proposat per Kulagina (Kulagina, 1958), sota la forma de la derivada d'una partició del vocabulari en formes flexionals. Una explicació lingüística d'aquest model va ser donada en el capítol III de (Marcus, 1967a), mentre que un desenvolupament matemàtic d'aquest model es deu a Popescu (Popescu, 1971), Revzin (Revzin, 1977) i Jurcă (Jurcă, 1970). Un important inconvenient d'aquest model és la impossibilitat per una forma flexional de pertànyer a diverses parts de l'oració, a diferència del que succeeix, p.e., en llenguatges com l'anglès. Aquesta és una de les raons de la no-concordança entre l'estatus de parts de l'oració de la gramàtica tradicional i el seu estatus en el model matemàtic. En molts casos, dins la mateixa part

de l'oració tradicional, distingim diverses parts de l'oració lògiques. Algunes vegades, hem de combinar el model de parts de l'oració amb el sintàctic – veure el càlcul sintàctic de Lambek (Lambek, 1961). Però per tal de millorar l'adequació i la rellevància del model de parts de l'oració, esdevingué molt important introduir una tipologia lingüística molt més refinada i subtil que la tradicional. El model de Kulagina aconsegueix de descobrir una profunda i sorprenent relació lògica entre les formes flexionals i les parts de l'oració, però el preu que hem de pagar per aquest aventatge és massa gran, mitjançant l'increment algunes vegades amb escreix de l'error d'aproximació. Es feia necessari un compromís. Aquest compromís consisteix en definir una tipologia lingüística on en alguns tipus de llenguatges, principalment en llenguatges adequats i homogenis, la importància dels models matemàtics de parts de l'oració i gènere o del cas sigui més gran que per altres tipus de llenguatges. Veurem ara els principals conceptes d'aquesta tipologia i les classificacions corresponents.

La tipologia considerada s'aplica a llenguatges definits sobre un vocabulari V , una partició P de V i un conjunt L de cadenes finites sobre V . Denotarem per S la partició de V en classes distribucionals i per T la partició de V en tipus, i.e., $T = P'$. Un llenguatge $\langle V, P, L \rangle$ és:

- *amorf*, si $P(x) = \{x\}$ per a qualsevol $x \in V$;
- *simple*, si a partir d' $x \in s(y) \cap P(y)$ se segueix que $x = y$;
- *perfecte*, si $p(x) \subset S(x)$ per a qualsevol $x \in V$;

- *homogeni*, si a partir de la no-buidor de $S(x) \cap P(y)$ se segueix la no-buidor de $S(y) \cap P(x)$;
- *aglutinatiu*, si de $x \rightarrow y$ se segueix que $x \in S(y)$;
- *adequat*, si per a qualsevol $x \in V$, $S(x) \subset T(x)$;
- *completament adequat*, si a partir de $x \rightarrow y$ se segueix que $x \in T(y)$;
- *ben adequat*, si a partir de $x \rightarrow y$ ($x, y \in V$) se segueix que $P(x)$ P -domina $P(y)$.

És important també recordar alguns tipus d'isomorfismes. Una funció un-a-un (o bijectiva) f d' A_1 a A_2 és:

- un P -isomorfisme entre $\langle A_1, P_1, L_1 \rangle$ i $\langle A_2, P_2, L_2 \rangle$, si $P_2(f(x)) = f(P_1(x))$, per a qualsevol $x \in A_1$;
- un S -isomorfisme entre $\langle A_1, P_1, L_1 \rangle$ i $\langle A_2, P_2, L_2 \rangle$, si $S_2(f(x)) = f(S_1(x))$;
- un PS -isomorfisme entre $\langle A_1, P_1, L_1 \rangle$ i $\langle A_2, P_2, L_2 \rangle$, si és a la vegada un P -isomorfisme i un S -isomorfisme.

Per a cada tipus de llenguatge considerem la corresponent versió absoluta que consisteix en la invariabilitat del tipus respecte respecte a un PS -isomorfisme.

Per exemple,

- $\langle A, P, L \rangle$ és *absolutament adequat* si qualsevol llenguatge PS -isomorf a ell mateix és també adequat;
- $\langle A, P, L \rangle$ és *absolutament complet* si qualsevol llenguatge PS -isomorf a ell mateix és també completament adequat; etc.

Es coneixen diversos teoremes al respecte,

Teorema 1.4.1 Un llenguatge L és absolutament ben adequat si, com a mínim, una de les següents condicions es compleix:

- $P(x) \cap S(y) = \emptyset$ per a qualsevol $a, y \in V$;
- $P(x) \subset S(x)$ per a qualsevol $x \in V$.

Teorema 1.4.2 Si qualsevol llenguatge PS -isomòrfic a $\langle A, P, L \rangle$ és adequat, aleshores $\langle A, P, L \rangle$ és homogeni; i a l'inrevés.

Considerem $x, y \in V$; x està sintàcticament relacionat amb y si per a cada cadena $u = u_1 \dots u_{i-1} x u_{i+1} \dots u_n \in L$ i per a cada i ($1 \leq i \leq n$), existeix una cadena $v = v_1 \dots v_{i-1} y v_{i+1} \dots v_n \in L$. Si x està sintàcticament relacionada amb y i y està sintàcticament relacionada amb x , escriurem xQy (x és *cognat* a y). Per a cada $x \in V$ aleshores $S(x) \subset Q(x)$. El llenguatge $\langle A, P, L \rangle$ és *quasibé perfecte* si $P(x) \subset Q(x)$ per a qualsevol $x \in V$. Qualsevol llenguatge perfecte és quasibé perfecte. Un llenguatge és quasibé perfecte si, i només si, $T(x) \subset Q(x)$ per a qualsevol $x \in V$. Revzin conjectura sobre el fet que en

l'anglès i en el francès totes les preposicions formen una única Q -classe, una altra Q -classe estaria formada per tots els noms, i encara una altra formada per adjectius. Per això aquests llenguatges semblen ser quasibé perfectes.

El primer intent de construir l'anterior model de parts de l'oració es deu a Uspenskij (Uspenskij, 1957b). Entre d'altres contribucions posterior trobem les de Zelinka (Zelinka, 1966; Zelinka, 1967a), Revzin (Revzin, 1967b; Revzin, 1967a), Trybulec (Trybulec, 1967a), Kasai (Kasai, 1970), Popescu (Popescu, 1970) i Petrescu (Petrescu, 1976).

El model del cas gramatical de Kolmogorov-Uspenskij (Uspenskij, 1957a) és el punt de partida del model del cas de Marcus – veure el Capítol 6 de (Marcus, 1967d). El cas és concebut com una classe d'equivalència de *contextos gramaticals congruents* sobre el vocabulari V , respecte a L . Els contextos γ_1 i γ_2 són *congruents*, respecte a L , si existeix una seqüència finita de contextos $\delta_1, \delta_2, \dots, \delta_n$ amb $\delta_1 = \gamma_1, \delta_n = \gamma_2$ i, δ_i i δ_{i+1} , estan en *concordança* respecte a L , i.e., existeix $a_i \in V$ acceptat a L per δ_i i δ_{i+1} ($1 \leq i \leq n$). Gladkij (Gladkij, 1973) no accepta treballar amb el conjunt de totes les cadenes ben formades degut a la seva nul·la efectivitat i el seu potencial infinit. En comptes d'això usa alguns predicats finits.

Més controvertit és el problema de construir un model matemàtic del gènere gramatical. Kiefer (Kiefer, 1968) expressa el seu escepticisme al respecte, exclamant que “el nostre coneixement sobre el gènere gramatical és insuficient per a justificar o permetre la seva formalització”. Però, a l'igual

que en el cas de les parts de l'oració o del cas gramatical, el problema no és trobar un argument matemàtic per a cada comportament singular; l'objectiu de la formalització és el d'identificar una estructura lògica genèrica que expressi l'essència del pas entre allò que és natural al gènere gramatical. El model proposat per Marcus – veure Capítol IV de (Marcus, 1967a) – usa un llenguatge amb estructures paradigmàtiques $\langle A, P, L \rangle$, on P és una partició del vocabulari V , mentre que L és un conjunt de cadenes ben formades que depèn del nivell de gramaticalitat seleccionat. Aquest model va ser desenvolupat posteriorment per Revzin (Revzin, 1967a), Karpinskaja (Karpinskaja, 1966; Revzina, 1974) i Horecký (Horecký, 1966). Una aproximació diferent, que està relacionada amb el fet de construir el model del gènere independentment del model del cas, va ser donada per Gladkij (Gladkij, 1973). Si en l'aproximació de Marcus la longitud d'algunes cadenes és essencial per a determinar la distància entre gèneres diferents, Karpinskaja considera que, per als llenguatges eslaus, no és la longitud sinó el nombre de cadenes que podem construir entre noms de gèneres diferents el que és significatiu. Karpinskaja determina les distàncies entre gèneres en el polonès, rus, serbi i txec. Teòricament, la distància més gran possible entre dos gèneres és igual a 1. Això difereix de l'aproximació de Marcus (Marcus, 1970a), on la distància entre dos gèneres diferents és sempre un nombre enter positiu més gran que 1.

1.5 L'espai contextual. La Topologia

Donat un llenguatge L sobre V , una cadena entre la cadena x i y és una seqüència finita de cadenes $x_0, x_1, \dots, x_n$ tal que $x_0 = x, x_n = y$ i, per a qualsevol i , $0 \leq i \leq n-1$, existeix un context no-nul que accepta x_i i x_{i+1} a L . La longitud d'aquesta cadena és $n-1$. La distància contextual $d_L(x, y)$ és la longitud de la cadena més curta entre x i y . Si no hi ha cap ambigüitat entre el llenguatge L , podem escriure solament $d(x, y)$. És fàcil de veure que el llenguatge universal V^* està organitzat, mitjançant d_L , com un espai mètric, al qual anomenarem *espai contextual*. Va ser introduït per Marcus en el Capítol 1 de (Marcus, 1967d) i investigat posteriorment per Zelinka (Zelinka, 1967b), que va mostrar l'existència d'un llenguatge que genera una finita, però no acotada, distància contextual. Dincă (Marcus, 1973a; Marcus, 1973b; Marcus, 1976) millora la distància de Marcus donant-li la possibilitat de prendre valors racionals que no siguin necessàriament enters. Ito (Ito, 1971; Ito, 1972b; Ito, 1972a) defineix l'extensió i la restricció associada a l'espai contextual. I Calude (Calude, 1973; Calude, 1976b; Calude, 1976c; Calude, 1976a) obté una posterior millora usant la distància de Pompeiu-Hausdorff entre conjunts tancats i la numeració de Gödel. Una transferència a la distància contextual semàntica va ser realitzada per Iosifescu (Iosifescu, 1975).

La primera aproximació de Marcus a la topologia de llenguatges va ser donada a (Marcus, 1961). Valk (Valk, 1972) va indicar una manera de connectar la topologia associada a una relació binària amb la teoria d'autòmats.

Aquest fet va ser usat per Ardelean (Ardelean, 1977b; Ardelean, 1977a) i aplicat a la relació de dominació i als llenguatges lineals. Una important topologia mètrica va ser proposada per Bodnarčuk (Bodnarčuk, 1965), mentre que la seva aplicació a l'aprenentatge va ser donada per Dumitrescu-Vianu (Dumitrescu i Vianu, 1979). La coherència d'un text va ser investigada mitjançant l'estructura topològica de connectivitat – per Salini-Trybulec (Salini i Trybulec, 1974), per Lipsky (Lipsky, 1974) i per Marcus (Marcus, 1980). Espais de tolerància – un cas particular de topologies de Čech – van ser aplicades diverses vegades a les estructures lingüístiques – veure Pogonovski (Pogonovski, 1981). La rellevància de la topologia en els processos d'aprenentatge de llenguatges va ser discutida per Marcus (Marcus, 1989a).

Notes bibliogràfiques. A més a més de les referències bibliogràfiques donades anteriorment, hom pot trobar una acurada introducció als models algebràico-analítics, amb una detallada relació bibliogràfica annexa, a (Marcus, 1994). Altres introduccions es poden trobar a (Marcus et al., 1966), a (Serrano, 1975; Serrano, 1977) i a (Miquel-Vergés, 1995b).

Capítol 2

Les gramàtiques contextuals

La investigació sistemàtica dels llenguatges/llengües naturals mitjançant models algebraics, combinatoris i de teoria de conjunts comença als anys 60 conjuntament a Europa i als EEUU. Un any molt important al respecte és l'any 1957, quan Chomsky, (Chomsky, 1957), publica el seu llibre pioner relatiu a la nova aproximació generativa a les estructures sintàctiques i alguns matemàtics russos proposen un marc de treball conceptual per a l'estudi de les categories morfològiques generals (Dobrušin, 1957), per a l'estudi de la categoria del cas gramatical – A. N. Kolmogorov; veure (Uspenskij, 1957a) – i l'estudi de la categoria de les parts de l'oració (Uspenskij, 1957b), sent el començament del desenvolupament dels models analítics dels llenguatges.

Cadascun d'aquests dos intents està relacionat amb eines matemàtiques específiques. Mentre les gramàtiques generatives estan inspirades en la combinatòria, els sistemes formals i la teoria de monoides lliures, els models analítics

estan basats en algunes idees provinents de la teoria de conjunts, relacions binàries, monoides lliures i operadors de clausura.

Malgrat les seves diferències respecte a l'aspecte matemàtic, aquests dos intents es diferencien també en dos aspectes més:

- a) Ambdós tenen tradicions lingüístiques diferents. Chomsky invoca la gramàtica India de Panini i el que ell anomena “Lingüística Cartesiana” (relacionada amb René Descartes), mentre que els models analítics, inicialment motivats pel treball desenvolupat en el camp de la traducció automàtica de llenguatges, estan formalitzant directament idees provinents de la lingüística estructural (Cercle de Praga, lingüística descriptiva Americana, etc.).
- b) Les gramàtiques generatives i els models analítics tenen estratègies complementàries. Els primers busquen mecanismes generatius que aproximïn el procés d'obtenir cadenes ben formades. Els segons assumeixen que és donat un cert nivell de bona formació i analitzen l'estructura corresponent. Les gramàtiques de Chomsky fan ús d'alguns mitjans externs al llenguatge (concebut aquest com a un conjunt de cadenes sobre un vocabulari terminal), com pot ser el vocabulari auxiliar o no-terminal, mentre que els models analítics adopten un punt de vista intern, intrínsec.

Les gramàtiques generatives són una ruptura de la tradició lingüística de la primera meitat del segle XX, mentre que els models analítics constitueixen pre-

cisament el seu desenvolupament, la continuació d'aquesta tradició. Semblava natural esperar un esforç per tal d'unir ambdues concepcions. Aquest esforç va provenir d'ambdues parts i, com veurem més endavant, les gramàtiques contextuais són un component d'aquest procés.

2.1 L'origen de les gramàtiques contextuais

Les Gramàtiques Contextuais (*CGs*) tenen el seu origen en l'intent de transformar en mecanismes generatius alguns processos desenvolupats dins el marc de treball dels models analítics. La idea de connectar d'aquesta manera l'estudi analític amb l'aproximació generativa als llenguatges naturals va ser un dels problemes principals investigat en la lingüística matemàtica en el període des de 1957 a 1970. Al respecte, podríem posar l'exemple del càlcul sintàctic de Lambek (1961), de les gramàtiques categorials que segueixen la tradició de l'Escola Polonesa de Lògica, de les gramàtiques de dependència que tenen el seu origen en el llibre de Tesnière, (Tesnière, 1959) i de les gramàtiques configuracionals estudiades per Gladkij, (Gladkij, 1963); veure al respecte els capítols corresponents de (Marcus, 1967a).

Les Gramàtiques Contextuais tracten d'explotar dues idees que es trobaven en els fonaments de la tradició de la lingüística descriptiva distribucional en els EEUU als anys 50 i 60: per una banda, la idea d'una cadena d'un vocabulari donat V , finit i no-buit; i, per altra banda, la idea d'un context sobre V ,

concebut com a un parell ordenat (u, v) de cadenes sobre V . El concepte *context* és usat sovint d'una manera poc efectiva, per la qual cosa en alguns casos aquest concepte és substituït per algun terme equivalent com *entorn* o bé *veïnatge*. L'objectiu principal de la lingüística descriptiva era el de descriure els llenguatges naturals utilitzant mitjans sintàctics, deixant de banda en la mesura del possible la implicació dels aspectes semàntics. Un exemple típic al respecte és el llibre de Harris (Harris, 1951), on l'autor comença considerant, com únic factor semàntic acceptat en la seva anàlisi, la distinció entre cadenes ben formades (*wfs*) i cadenes mal formades (*nwfs*). Aquesta distinció és assumida com un axioma, com a una convenció primitiva, que té una base intuïtiva empírica, que resulta de l'experiència de qualsevol parlant natiu d'una llengua. El problema principal per a Harris era el d'observar i descriure les interrelacions entre les cadenes i les seves subcadenes, per una banda, i amb els corresponents contextos, per altra banda; i establir què és el que les operacions contextuais i de concatenació, aplicades a les cadenes ben formades, aporten a les cadenes ben formades. En altres paraules, quines transformacions d'una cadena deixen invariant la seva bona formació.

Cal remarcar que aquesta creença en les enormes capacitats de les operacions sintàctiques i contextuais d'abraçar la quasi totalitat de les necessitats de l'anàlisi lingüística va ser una presuposició implícita en l'aproximació adoptada pels primers investigadors en traducció automàtica de llenguatges; més tard es va veure clar que els factors semàntics i pragmàtics eren molt més importants del que inicialment s'havia considerat.

2.2 Motivacions de les gramàtiques contextuais

Existeix un acord quasibé general sobre l'estatus de vaguetat del conjunt de cadenes ben formades en un llenguatge/llengua natural (per altra banda, sembla ser que aquest acord no és tan general en el cas dels llenguatges de programació). Es coneixen nombroses condicions necessàries per tal que una cadena sigui ben formada, però no es coneix cap condició suficient (i, a posteriori, cap condició necessària i suficient) al respecte. També hi ha un consens general pel que fa a la infinitud del conjunt de cadenes ben formades en un llenguatge/llengua natural – pel que fa a la seva numerabilitat, les investigacions actuals dins la lingüística matemàtica mantenen que els llenguatges/llengües naturals són conjunts numerables de frases; no obstant, Martín-Vide prova a (Martín-Vide, 1990) que, des de cert punt de vista, les llengües naturals són conjunts infinits no-numerables d'ordre 2. Per això, qualsevol anàlisi lingüística efectiva consisteix en l'aproximació del conjunt total de cadenes ben formades mitjançant un subconjunt concret del mateix; aquest subconjunt s'acostuma a anomenar un *nivell de gramaticalitat* (lg).

Donat un cert nivell de gramaticalitat, i.e., un conjunt concret L de cadenes ben formades sobre V , direm que una cadena s sobre V és acceptada a L per un context (u, v) sobre V si la cadena usv pertany a L . Assumim que és més convenient precisar primer un nivell de gramaticalitat (el conjunt L), i després distingir diversos tipus de cadenes i de contextos, d'acord al test d'acceptació a L . Però, si volem aconseguir un nivell de gramaticalitat més

interessant, en el sentit de la infinitud d' L , apareix una qüestió natural: podem obtenir un conjunt infinit L a partir d'un nivell de gramaticalitat inferior, expressat mitjançant un conjunt finit de cadenes ben formades ? Aquesta qüestió immediatament en provoca una altra: podem obtenir la transformació d'un nivell de gramaticalitat més baix en un nivell de gramaticalitat superior només usant operacions contextuais (i.e., emplaçant algunes subcadenes en alguns contextos) aplicades un nombre finit de vegades ? En la recerca de la resposta a aquestes qüestions, obtenim, successivament, el concepte de gramàtica contextual simple – sense selecció – (*SCG*) i el concepte de gramàtica contextual amb selecció (*CGC*). A continuació resseguirem aquest itinerari.

Existeixen algunes cadenes molt estàndards, com és el cas de “Aquest cotxe és blanc” o “Tots els homes són mortals”, que són acceptades com a cadenes ben formades sense cap dubte, per qualsevol parlant natiu del català. Per això, podem assumir que un conjunt finit d'aquestes cadenes poden constituir el nostre punt de partida; anomenarem a aquestes cadenes *cadenes ben formades primitives* (*wfs*). Cal observar que el conjunt finit L_0 de cadenes ben formades primitives pot ser escollit de diverses maneres. Aquest conjunt es troba sotmès a dues tendències conflictives: per una banda, el desig de tenir L_0 tan petit com sigui possible, per tal de simplificar el procés d'emplaçament de cadenes d' L_0 en diversos contextos possibles; i, per altra banda, el desig de tenir L_0 el més gran possible, per tal d'incloure-hi una gran varietat de tipus de cadenes ben formades.

Una vegada fixat el conjunt L_0 , la següent qüestió està relacionada amb la selecció d'un conjunt finit adequat C_0 de contextos que puguin ser aplicats a cadenes d' L_0 i a cadenes obtingudes a partir d' L_0 d'aquesta manera. En altres paraules, l'operació d'aplicar contextos de C_0 comença amb cadenes d' L_0 , però poden ser aplicats iterativament un nombre limitat de vegades. Aquesta situació ens condueix a un conjunt L de cadenes que és el més petit possible que satisfà els següents requeriments: qualsevol cadena d' L_0 pertany a L i, per a qualsevol cadena x d' L i per a qualsevol context (u, v) de C_0 , la cadena uxv pertany a L .

De fet, els conjunts finits L_0 i C_0 no haurien de ser seleccionats consecutivament sinó concomitantment, ja que les cadenes i contextos apareixen conjuntament i uns depenen dels altres. La doble restricció que actua sobre el conjunt L_0 actua també sobre C_0 i, d'alguna manera, cadascun dels conjunts C_0 i L_0 pot ser millorat, l'un a expenses de l'altre, però al mateix temps; i un increment de complexitat d' L_0 requereix el corresponent increment de complexitat de C_0 .

L'experiència amb llenguatges/llengües naturals mostra que qualsevol cadena selecciona un tipus de context i qualsevol context selecciona un tipus de cadena. Per exemple, un adjectiu numeral *tres* selecciona el conjunt de tots els grups de noms que poden contenir *tres*, mentre que un context de la forma (λ, nom) , on λ és la cadena buida, selecciona tots els adjectius numerals acceptats en català per aquest context. No fa falta dir que aquesta selecció es fa en

base al nivell de gramaticalitat; en l'exemple anterior, el nivell de gramaticalitat considerat estava constituït per tots els grups nominals del català. Du Feu i Marcus (Du Feu i Marcus, 1973) van donar una anàlisi contextual detallada de l'anglès, respecte a diferents nivells de gramaticalitat. Una anàlisi similar, en el cas del francès, va ser donada per Vasilescu (Vasilescu, 1973; Vasilescu, 1974; Vasilescu, 1975); mentre que, l'hongarès i el romanès, així com alguns llenguatges de programació, són considerats sota aquest aspecte a (Marcus, 1981; Marcus, 1983). La idea principal que guia aquest procés és el fet que, malgrat la vaga natura de la propietat de bona formació, l'aproximació del conjunt difús de totes les cadenes ben formades d'un subconjunt concret de cadenes ben formades pot ser sempre millorat o reformat.

De fet, no hi cap evidència que l'estatus de les cadenes ben formades es reflecteixi de la millor manera mitjançant el concepte de conjunt difús (“fuzzy set”) considerat per Zadeh, (Zadeh, 1965). Sembla ser que el concepte de conjunt bast, introduït per Pawlak, (Pawlak, 1982), és un millor candidat al respecte; aquesta conjectura, però, encara no ha estat verificada – al respecte, es pot consultar (Miquel-Vergés, 1995a; Miquel-Vergés, 1996a).

2.3 Cadenes, contextos i clausura de Sestier

El concepte de gramàtica contextual amb selecció (*CGC*) té en compte la capacitat de qualsevol cadena de seleccionar una classe de contextos prefer-

encials. No obstant, aquesta capacitat és només una part d'un fenomen més comprensiu, la dualitat entre cadenes i contextos.

Donat un llenguatge L sobre V , podem associar a cada conjunt S_0 de cadenes sobre V la classe $C(S_0)$ de contextos que accepten a L qualsevol cadena d' S_0 . De manera similar, podem associar a qualsevol conjunt C_0 de contextos sobre V , el conjunt $S(C_0)$ de totes les cadenes acceptades a L per qualsevol context de C_0 . És fàcil de veure que, si S està contingut en el conjunt T de cadenes, aleshores $C(S)$ conté el conjunt $C(T)$, mentre que si C està contingut al conjunt D de contextos, aleshores $S(C)$ conté el conjunt $S(D)$ (per això, els dos operadors que acabem de definir són antimonòtons). A més a més, S_0 està contingut a $S(C(S_0))$, mentre que C_0 està contingut a $C(S(C_0))$. Aquesta situació mostra que la interrelació entre cadenes i contextos es realitza mitjançant una correspondència de Galois o connexió de Galois. Aleshores – veure, p.e., (Riguet, 1948) –:

Definició 2.3.1 Donats dos conjunts ordenats X i Y , una funció f d' X a Y i una funció g d' Y a X , direm que el parell (f, g) defineix una *connexió de Galois* entre X i Y quan se satisfan els següents requisits:

- i) si u i v pertanyen a X i u precedeix v , aleshores $f(v)$ precedeix $f(u)$;
- ii) si y i z pertanyen a Y i y precedeix z , aleshores $g(z)$ precedeix $g(y)$;
- iii) $x \in X$ sempre precedeix $g(f(x))$, mentre que $y \in Y$ sempre precedeix $f(g(y))$.

Prendrem ara, com a X , la família de diversos conjunts de cadenes sobre V i, com a Y , la família de diversos conjunts de contextos sobre V . Associem a cada conjunt S de cadenes sobre V el conjunt $f(S)$ de contextos que accepten a L qualsevol cadena d' S . Associem a cada conjunt C de contextos sobre V el conjunt $g(C)$ de cadenes acceptades a L per qualsevol context de C . Prendrem com a relació d'ordre a X i a Y la relació d'inclusió de la teoria de conjunts. Segons les consideracions anteriors, se segueix que el parell (f, g) de funcions que acabem de definir és una connexió de Galois entre X i Y , i.e., entre la família de conjunts de cadenes i la família de conjunts de contextos sobre V . Aquesta connexió de Galois és relativa a un llenguatge donat L sobre V . Ja que el conjunt de llenguatges sobre V és no només infinit sinó també no-comptable, se segueix que el conjunt de connexions de Galois que podem definir entre conjunts de cadenes i conjunts de contextos sobre V és no-comptable (i, per tant, també podem fer esment al fet que la família de conjunts de cadenes i la família de conjunts de contextos sobre V són ambdós no-comptables).

L'anterior dualitat entre cadenes i contextos va ser explorada per primera vegada per Sestier, (Sestier, 1961). Per a tot llenguatge L sobre V , va definir l'operador contextual c que associa a cada conjunt X de cadenes sobre V el conjunt $c(L, X)$ de tots els contextos que accepten a L totes les cadens de X . Similarment, Sestier defineix un operador "invers" s que associa a cada conjunt Y de contextos sobre V el conjunt $s(L, Y)$ de cadenes acceptades a L per qualsevol context d' Y . Aleshores, Sestier considera la composició $\varphi(X, L) =$

$s(L, c(L, X))$, anomenada avui en dia la *clausura de Sestier* d' X respecte a L . És, en efecte, un operador de clausura i va ser definit, p.e., per Ward, (Ward, 1942) de la següent manera:

Definició 2.3.2 Donat un conjunt ordenat Z , un *operador de clausura* γ a Z és una funció γ de Z a Z que satisfà les condicions:

- i) qualsevol x de Z precedeix, en el sentit d'ordre, l'element $\gamma(x)$;
- ii) si x i y pertanyen a Z i x precedeix y , aleshores $\gamma(x)$ precedeix $\gamma(y)$ (propietat de monotonia);
- iii) per a qualsevol x pertanyent a Z , $\gamma(\gamma(x)) = \gamma(x)$ (la iteracció de γ esdevé estacionària des del primer pas de la seva aplicació). A més a més, si Z té un ínfim 0 , es requereix que $\gamma(0) = 0$.

En el nostre cas, Z és el conjunt de tots els llenguatges possibles sobre V i la relació d'ordre és la relació d'inclusió. La qualitat de l'operador de Sestier φ de ser un operador de clausura pot ser verificada directament, però també se segueix a partir d'un teorema donat per Ore, (Ore, 1944):

Teorema 2.3.1 Donats dos conjunts ordenats X i Y i una connexió de Galois (f, g) entre X i Y , la composició, en qualsevol ordre, d' f i de g , defineix un operador de clausura a X (si comencem per f) o a Y (si comencem per g).

El fet més interessant de l'operador de Sestier és que, malgrat la propietat de la funció s de ser, d'alguna manera, la inversa de la funció c , aquestes dues

funcions no es neutralitzen l'una a l'altra (com passa amb l'addició i la resta, o amb la multiplicació i la divisió); i, parlant d'una manera general, $\varphi(X)$ és diferent d' X . Quan $\varphi(X) = X$, es diu que X és un conjunt tancat (podem reemplaçar $\varphi(X, L)$ per $\varphi(X)$ quan no existeix cap ambigüitat pel que fa al llenguatge considerat L), i.e., un llenguatge tancat.

2.4 Modelització de les categories

Sestier es va guiar per motivacions algebraiques associades amb una implícita sensació del que podria ser lingüísticament rellevant. No obstant, com es va veure més tard, (Marcus, 1967a; Marcus, 1967c; Marcus, 1968; Marcus, 1969c), les clausures de Sestier són el model més comprensiu de les categories morfològiques, tal com van ser investigades dins el marc de treball dels models analítics del llenguatge. Per tal d'arribar a aquest resultat, hem de considerar el treball fet independentment al de Sestier, relacionat directament amb els problemes que apareixen en els llenguatges/llengües naturals.

El primer model de categories morfològiques va ser proposat per Dobrušin, (Dobrušin, 1957). El seu punt de partida és, malgrat no ser explicitat, el concepte de classe distribucional, considerat en la lingüística descriptiva americana. Per a cada a de V , Dobrušin defineix la classe distribucional (dc) d' a respecte a L com el conjunt $D_L(a)$ que conté tots els elements b de V de manera que a i b siguin acceptats a L exactament pels mateixos contextos.

Aleshores considera el que ell anomena la relació de dominació: donats a i c de V , a domina c respecte al llenguatge L si, qualsevol context que accepta a a L , accepta també c a L . Per exemple, si V és el vocabulari del català i L és el conjunt de grups nominals en català, aleshores l'adjectiu qualificatiu *gros* domina els adjectius qualificatius *petit* i *gran*. No obstant, mentre que *petit* domina *gros* respecte a L (per això, *petit* pertany a la classe distribucional de *gros*), *gran* no domina *gros*, ja que, p.e., en el grup nominal ben format *una dona gran*, la substitució de *gran* per *gros* ens dóna una cadena que és mal formada **una dona gros*.

En el següent pas, Dobrušin defineix el concepte de classe distribucional inicial (*idc*), com una classe distribucional tal que qualsevol element de V que la domina pertany a la classe. Per exemple, si prenen la interpretació anterior d' L , $D(\textit{gros})$ és una classe distribucional inicial, però $D(\textit{gran})$ no ho és, ja que *gros* està dominant *gran* però no pertany a $D(\textit{gran})$. Aquest exemple, així com molts altres considerats – en el cas del francès, veure (Marcus, 1981; Marcus, 1983) –, mostren que la relació de dominació entre paraules permet de comparar-los des del punt de vista de la seva homonímia morfològica (*mh*): l'homonímia entre les formes flexionals de *gros* és més petita que l'homonímia entre les formes flexionals de *gran* (aquesta última és la mateixa forma pel masculí i pel femení, mentre que amb *gros* no succeeix igual). Més generalment: dues cadenes o paraules de la mateixa classe distribucional tenen la mateixa homonímia morfològica. Si a domina b aleshores l'homonímia morfològica de b és més gran que la d' a , però l'invers no és cert; si ni a domina b ni b domina

a , aleshores a i b no són comparables.

En aquesta situació, sembla molt natural l'últim pas de l'aproximació de Dobrušin: defineix la categoria morfològica elemental (*emc*) com la unió entre una classe distribucional inicial $D(a)$ i el conjunt de tots els elements de V que són dominats per $D(a)$. Per exemple, en el cas d' $a = \text{gros}$, aquest seria el conjunt de totes les formes flexionals dels adjectius qualificatius que tenen valors de masculí singular.

No obstant, molts altres tipus de categories morfològiques que no són “elementals” necessiten la seva corresponent generalització. La investigació sistemàtica al respecte va ser feta per Marcus, (Marcus, 1967d), mitjançant la substitució d'elements de V per cadenes de V i les classes distribucionals inicials per conjunts arbitraris de cadenes. D'aquesta manera, el model obtingut anava més enllà de la frontera de la morfologia i esdevenia un procediment general per a capturar el fenomen de l'homonímia contextual (Marcus, 1981; Marcus, 1983).

Però fins i tot si ens quedem en el nivell morfològic de les formes flexionals, Kunze va observar, (Kunze, 1967), que la relació de dominació de Dobrušin, tan eficient pels llenguatges eslaus, pel llatí, francès, i romanès, és menys eficient per l'alemany, on, per exemple, un conjunt com $\{\text{ich}, \text{du}\}$ (pronoms personals de primera i segona persona) requereix una relació de dominació més general. Si per a Marcus, (Marcus, 1962), la categoria gramatical $G(X)$ generada per un conjunt X de cadenes és la unió d' X amb el conjunt de totes les cadenes

dominades per qualsevol cadena a X , per a Kunze, (Kunze, 1967), la categoria $K(X)$ generada per X (anomenarem a aquesta categoria la categoria de Kunze) és la unió d' X amb tots els conjunts Y de cadenes. De manera que, tot context que accepta a L totes les cadenes a X , també accepta a L totes les cadenes d' Y .

La categoria gramatical, la categoria de Kunze i la clausura de Sestier estan en relació molt estreta. Es verifiquen els següents teoremes, (Marcus, 1969a):

Teorema 2.4.1 Donat el vocabulari V i el llenguatge L sobre V , la categoria de Kunze $K(X)$ respecte a L , generada per un conjunt X de cadenes sobre V , és idèntica a la clausura de Sestier $\varphi(X)$ respecte a L .

Teorema 2.4.2 La categoria gramatical $G(X)$ respecte a L està sempre continguda a $K(X)$ i, per això, la categoria de Kunze és més extensa que la categoria gramatical (però, quan X és una classe distribucional, tenim que $G(X) = K(X)$).

Teorema 2.4.3 Per a qualsevol conjunt X de cadenes existeix un conjunt Y de cadenes tal que $\varphi(X) = G(Y)$; en altres paraules, qualsevol clausura de Sestier és una categoria gramatical i, a partir del primer teorema (2.4.1), qualsevol categoria de Kunze és una categoria gramatical. L'invers no és cert.

Aquests teoremes mostren que la categoria gramatical i la clausura de Sestier inclouen una gran varietat de fenòmens morfològics i altres fenòmens grama-

tics, de caràcter més general, relacionats amb el comportament contextual de les cadenes i la seva ambigüitat contextual. Cal observar que, ni en la lingüística tradicional ni en la lingüística estructural o generativa, existeix un clar estatus conceptual del que podria ser una categoria morfològica general o una categoria gramatical general.

2.5 L'aproximació contextual en la perspectiva generativa

En vistes a la important significació lingüística de les clausures de Sestier i de la categoria gramatical, com hem pogut observar en les consideracions prèvies, val la pena investigar les seves rèpliques generatives.

Prenem primer els llenguatges sobre V que siguin categories gramaticals respecte a un llenguatge donat L ; com ja hem vist, entre aquests llenguatges trobarem també totes les clausures de Sestier (Teorema 2.4.3 de la pàgina 73). Per a qualsevol llenguatge X sobre V , tenim $G(G(X)) = G(X)$ (Marcus, 1967d). Se segueix que una condició necessària i suficient per tal que un llenguatge X sobre V sigui una categoria gramatical respecte a L (i.e., que existeixi un llenguatge U tal que $X = G(U)$), és que $X = G(X)$. Això vol dir que l'homonímia contextual de cadenes d' X és d'alguna manera maximal; no pot existir cap cadena que no pertanyi a X l'homonímia contextual de la qual sigui igual o més gran que l'homonímia contextual de qualsevol cadena d' X . Si, seguint Dobrušin, volguéssim afegir, per a la categoria gramatical

generada per un llenguatge X , la condició de ser una categoria gramatical inicial (i.e., requerir d' X ser inicial: que qualsevol cadena que domini totes les cadenes d' X pertanyi a X), aleshores la condició $X = G(X)$ significaria que X no és comparable, des del punt de vista de l'homonímia contextual, amb cap altre llenguatge. Si ens imaginem un graf, els nodes del qual són llenguatges sobre V i una fletxa que va des d' X fins a Y , on cada cadena a X domina cada cadena a Y (respecte a un llenguatge prefixat donat L), aleshores tota categoria gramatical X seria un node terminal, mentre que tota categoria gramatical inicial seria un node isolat.

Podem preguntar-nos: quin lloc ocuparien en la jerarquia de Chomsky (i en els seus posteriors refinaments) aquests llenguatges que són categories gramaticals o categories gramaticals inicials? La formulació precisa d'aquesta qüestió seria: donat un llenguatge L sobre V , existeixen categories gramaticals i categories gramaticals inicials respecte a L en qualsevol classe de la jerarquia de Chomsky extesa (incloent-hi, per exemple, no només els llenguatges regulars, lliures del context i sensibles al context, sinó també els seus coneguts refinaments)? La resposta dependria, a la vegada, del lloc d' L en la jerarquia de Chomsky extensa.

Direm que dos llenguatges sobre V estan *gramaticalment conjugats* si cadascun d'ells és una categoria gramatical respecte a l'altre; els anomenarem *gramaticalment conjugats des de l'inici* si ambdós són gramaticalment conjugats i inicialment conjugats (i.e., cadascun d'ells és inicial respecte a l'altre). Po-

dem estendre aquests conceptes, de manera òbvia, a conjunts arbitraris de llenguatges: un conjunt de llenguatges sobre V és gramaticalment conjugat (resp., gramaticalment conjugat des de l'inici) si qualsevol parell de llenguatges del conjunt és gramaticalment conjugat (resp., gramaticalment conjugat des de l'inici).

Pel fet que qualsevol clausura de Sestier és un operador de clausura, tenim, respecte a un llenguatge prefixat L sobre V , $\varphi(\varphi(X)) = \varphi(X)$, per a qualsevol llenguatge X sobre V . Se segueix fàcilment que X és una clausura de Sestier (i.e., existeix U tal que $X = \varphi(U)$) si, i només si, $X = \varphi(X)$. Anomenarem a aquests llenguatges *llenguatges tancats*. Cal recordar que les clausures de Sestier són categories lingüístiques més comprensives que les categories gramaticals, ja que, per a qualsevol X , $G(X)$ està contingut a $\varphi(X)$, mentre que algunes vegades (veure la situació en alemany), $\varphi(X)$ no està contingut a $G(X)$. Per això, seria lingüísticament rellevant esbrinar com estan distribuïts els llenguatges tancats entre les diferents classes de la jerarquia de Chomsky extensa. En aquest cas, també la resposta dependria del lloc que ocupa el llenguatge donat, L , en aquesta jerarquia. A més a més, pel fet que qualsevol llenguatge tancat és una categoria gramatical, la distribució dels llenguatges tancats (la seva tipologia generativa) està relacionada amb la tipologia generativa de les categories gramaticals.

Considerarem que dos llenguatges són *conjugats tancats* si cadascun d'ells és un llenguatge tancat respecte a l'altre. Un *conjunt de llenguatges* s'anomena

conjugat tancat si cadascun dels llenguatges del conjunt és conjugat tancat respecte a qualsevol altre llenguatge del conjunt. Introduïrem també el concepte de *llenguatge inicial tancat* (que significaria: que és a la vegada tancat i inicial; però en aquest cas, *inicial* no en el sentit de Dobrušin, sinó en el sentit de Kunze: X és inicial segons Kunze si qualsevol llenguatge Y que Kunze-domina X – respecte al llenguatge donat L – està contingut a X). Podríem considerar els *llenguatges conjugats tancats des de l'inicial* com a conjunts de llenguatges on cada parell de llenguatges és tal que, cada llenguatge del parell, és inicialment tancat respecte als altres llenguatges del parell.

Es va deixar de banda un cas especial : la possibilitat, per a un llenguatge, de conjuguar-se (en els diversos aspectes considerats anteriorment) ell mateix. En aquest cas s'observa l'important paper que juga el context buit. Podríem, en algunes situacions, excloure'l de les nostres consideracions.

Apareixen interessants qüestions relatives a conjunts conjugats de llenguatges. A continuació veurem alguns exemples (la paraula *conjugat* ha de ser interpretada, consecutivament, en cadascun dels quatre significats introduïts anteriorment: gramaticalitat, gramaticalitat inicial, tancament, conjunts conjugats tancats inicials de llenguatges). Per a quins valors d' n podem trobar conjunts conjugats d' n llenguatges en cada classe de la jerarquia de Chomsky extesa ? Per a quins valors d' n podem trobar conjunts conjugats d' n llenguatges situats en, com a mínim, dues classes diferents en la jerarquia de Chomsky extesa ? Podem fer-nos la mateixa qüestió quan *dues* és substituït per *tres*,

quatre o *cinc*. La qüestió més simple d'aquest tipus seria si existissin parells conjugats de llenguatges pertanyents a diferents tipus generatius i si existissin tripletes conjugades d'aquests llenguatges.

2.6 Generació de cadenes i contextos

A la vista dels procediments contextuals essencials usats en la definició d'una categoria gramatical, la dominació de Kunze i les clausures de Sestier, podríem esperar una relativa simple connexió entre elles, d'una banda, i amb les gramàtiques contextuals – en les seves diverses variants proposades per Marcus (Marcus, 1969b) i altres autors (Păun, 1982a; Păun, 1994) – per altra banda. Això vol dir:

- i) la identificació d'aquests tipus de gramàtiques contextuals, les quals generen categories morfològiques, categories gramaticals i llenguatges tancats, categories gramaticals inicials i llenguates tancats inicials (respecte a un llenguatge donat, respecte als seus tipus contextuals o respecte als seus tipus chomskians) i, finalment,
- ii) la caracterització d'aquests parells de conjunts de gramàtiques contextuals que generen diversos tipus de llenguatges conjugats i conjunts conjugats de llenguatges

Un aspecte més específic de les gramàtiques contextuais és la seva manera natural de portar-nos a dos tipus diferents de processos generatius: un d'ells generant cadenes i l'altre generant contextos. Donat un vocabulari V , considerem un parell de funcions (f, g) , on f associa a cada conjunt S de cadenes sobre V un conjunt $f(S)$ de contextos sobre V ; i g associa a cada conjunt C de contextos sobre V un conjunt $g(C)$ de cadenes sobre V . Sota quines condicions d' f i g existeix un llenguatge L sobre V , respecte al qual el parell (f, g) defineix una connexió de Galois entre la família de conjunts de cadenes i la família de conjunts de contextos sobre V ? L'existència d' L és sempre única?

Direm que (f, g) és una connexió cadena-context de Galois (respecte a L). Una gramàtica contextual associada a (f, g) podria generar, per una banda, un conjunt S de cadenes i, per altra banda, un conjunt C de contextos, de manera que $f(S) = C$. Aquesta gramàtica és *f-compatible amb la connexió de Galois* (f, g) . Si $g(C) = S$, aleshores la gramàtica és *g-compatible amb* (f, g) . Si tenim els dos casos $f(S) = C$ i $g(C) = S$, aleshores la gramàtica és *fg-compatible amb* (f, g) . Per tant, donada una connexió cadena-context de Galois (f, g) , quines restriccions operen en la família $\mathcal{F}$ de gramàtiques contextuais que són *f-compatibles*, *g-compatibles*, *fg-compatibles amb* (f, g) ? Quina és la possible tipologia de $\mathcal{F}$? Podria ser buida? Podria incloure tots els parells associats $(S, f(S))$, $(g(C), C)$, $(g(C), f(S))$?

Hem donat suficients motivations per a respondre ara a la qüestió bàsica: com pot una gramàtica contextual generar cadenes i contextos? Considerem

en primer lloc una gramàtica contextual simple $G = (V, L_0, C_0)$, on L_0 és un llenguatge finit sobre V , mentre que C_0 és un conjunt finit de contextos sobre V . Recordem que el llenguatge generat per G és el conjunt mínim L de cadenes sobre V tal que L_0 és contingut a L i per a qualsevol cadena x d' L i qualsevol context (u, v) de C_0 la cadena uxv pertany a L_0 . Si ens aprofitem de la similitud de la situació a G , d' L_0 i de C_0 , podem canviar els seus papers en el procés de generació i, reemplaçant les cadenes pels contextos i els contextos per cadenes, obtenim que: el conjunt de contextos generats per G és el conjunt mínim C de contextos tal que C_0 està contingut a C i, per a qualsevol context (u, v) de C i per a qualsevol cadena x d' L_0 , com a mínim un dels contextos (ux, v) , (u, xv) , (xu, v) , (u, vx) pertany a C .

Podem ser més específics i restrictius, mitjançant la definició de:

- *gramàtiques contextuais internes per l'esquerra simples (SLICG)* on la variant (ux, v) és considerada;
- *gramàtiques contextuais internes per la dreta simples (SRICG)* on escolim la variant (u, xv) ;
- *gramàtiques contextuais externes per l'esquerra simples (SLECG)* corresponents a la variant (xu, v) ; i
- *gramàtiques contextuais externes per la dreta simples (SRECG)* associades amb la variant (u, vx) .

Aleshores, podem definir:

- les *gramàtiques contextuais internes simples* (*SICG*) com les gramàtiques que tenen com a mínim un dels contextos $(ux, v), (u, xv)$ a C ;
- *gramàtiques contextuais internes simples fortes* (*SSICG*) si són a la vegada (*SLICG*) i (*SRICG*);
- *gramàtiques contextuais externes simples* (*SECG*) quan com a mínim un dels contextos $(xu, v), (u, vx)$ pertany a C ;
- *gramàtiques contextuais externes simples fortes* (*SSECG*) si són a la vegada (*SLECG*) i (*SRECG*);
- *gramàtiques contextuais per l'esquerra simples* (*SLCG*) si com a mínim un dels contextos $(ux, v), (xu, v)$ pertany a C ;
- *gramàtiques contextuais per la dreta simples* (*SRCG*) si, com a mínim, un dels contextos $(u, xv), (u, vx)$ pertany a C ;
- *gramàtiques contextuais per l'esquerra simples fortes* (*SSLCG*) si són a la vegada (*SLICG*) i (*SLECG*);
- *gramàtiques contextuais per la dreta simples fortes* (*SSRCG*) si són a la vegada (*SRICG*) i (*SRECG*);
- *gramàtiques contextuais simples fortes* (*SSCG*) si són a la vegada (*SSICG*) i (*SSECG*) o, equivalentment, si són a la vegada (*SSLCG*) i (*SSECG*).

En totes aquestes variants, l'estructura de la gramàtica és la mateixa; només canvia el procés de generació, amb la possibilitat de tenir diferents conjunts de contextos generats.

Ara el problema és: donada una gramàtica contextual simple G d'un dels tipus anteriors, denotem per S el conjunt de cadenes generades per G i per C el conjunt de contextos generats per G . Aleshores, existeix un llenguatge L tal que G és f -compatible, g -compatible o fg -compatible amb la connexió de Galois (f, g) associada a L (i.e., tenir $f(S) = C, g(C) = S$ o ambdues igualtats) ? Si aquest llenguatge L existeix, és únic ? Si no és únic, què podem dir sobre la família de tots els llenguatges L amb aquesta propietat ? Òbviament, la resposta dependrà del tipus de gramàtica contextual simple considerada, i.e., del tipus de procés de generació que ens porta als contextos generats per la gramàtica.

Cada context (u, v) té dos components: el de l'*esquerre*, u , i el de la *dreta*, v . Cada component de l'esquerra o de la dreta s'anomena un *component unilateral*. Podem definir el tipus generatiu chomskià de l'esquerre (dreta, unilateral) d'un conjunt de contextos com el tipus chomskià del conjunt de tots els components de l'esquerre (dreta, unilateral) d'aquests contextos. D'aquesta manera, podem associar a cada gramàtica contextual G dos tipus chomskians: un relacionat amb el conjunt de cadenes generades per G i l'altre relacionat amb el conjunt de contextos generats per G . Mentre que el conjunt inicial és únic, el final depèn del tipus de gramàtica contextual simple que usem. Se-

ria interessant determinar d'alguna manera en quina mesura aquests dos tipus chomskians poden ser diferents. Podem considerar també el tipus chomskià del conjunt de contextos de l'esquerra i el tipus chomskià del conjunt de contextos de la dreta i investigar en quina mesura poden ser diferents.

Si, en comptes de referir-nos al tipus chomskià, ens referim a la tipologia de les gramàtiques contextuais, establerta, per exemple, per (Păun, 1994), aleshores ens trobem de nou enfrontats amb problemes similars als anteriors.

2.7 Interelació de cadenes i contextos

Considerem ara les gramàtiques contextuais amb selecció (*CGC*). A diferència de les gramàtiques contextuais simples (*SCG*), la definició de les quals és simètrica respecte a cadenes i a contextos, la definició de gramàtiques contextuais amb selecció no posseeix aquesta propietat de simetria. Per això, juntament amb el tipus $G = (B, L_0, C_0, f)$ – on (B, L_0, C_0) és una gramàtica contextual simple, i f és una funció que associa a cada cadena s d' L_0 un subconjunt $f(s)$ de C_0 –, introduïrem un tipus dual de gramàtica contextual amb selecció, com un sistema $H = (B, L_0, C_0, g)$, on els primers tres objectes són els mateixos que abans, mentre que g és una funció que associa a cada context c de C_0 un subconjunt $g(c)$ de L_0 . El conjunt $C(H)$ de contextos generats per H és, per definició, el conjunt més petit C' de contextos tal que C_0 està contingut a C' i, si $c = (u, v)$ pertany a C' i la cadena x pertany a $g(c)$, aleshores

com a mínim un dels contextos $(ux, v), (u, xv), (xu, v), (u, vx)$ pertany a C' . Com en el cas de les gramàtiques contextuais simples, podem considerar altres variants diferents, mitjançant la imposició d'un o d'alguns d'aquests quatre contextos. Aquesta vegada, no obstant, canviarem parcialment la possible tipologia i aquest canvi (que ens condueix a una gran varietat de possibilitats) també pot ser considerat per a gramàtiques contextuais simples. Aleshores, tenim:

- gramàtiques contextuais selectives internes per l'esquerra (*LICGC*);
- gramàtiques contextuais selectives externes per l'esquerra (*LECGC*);
- gramàtiques contextuais selectives internes per la dreta (*RICGC*);
- gramàtiques contextuais selectives externes per la dreta (*RECGC*);
- gramàtiques contextuais selectives internes bilaterals (*BICGC*) que són a la vegada (*LICGC*) i (*RICGC*);
- gramàtiques contextuais selectives externes bilaterals (*BECGC*) que són a la vegada (*LECGC*) i (*RECGC*);
- gramàtiques contextuais selectives per l'esquerra (*LCGC*) que són a la vegada (*LICGC*) i (*LECGC*);
- gramàtiques contextuais selectives per la dreta (*RCGC*) que són a la vegada (*RICGC*) i (*RECGC*);

- gramàtiques contextuais selectives internes dèbils ($WICGC$) quan tots els contextos generats són interns, alguns d'ells per l'esquerra, i d'altres per la dreta;
- gramàtiques contextuais selectives externes dèbils ($WECGC$) obtingudes a partir de ($WICGC$) canviant “intern” per “extern”;
- gramàtiques contextuais selectives per l'esquerra dèbils ($WLCGC$) quan tots els contextos són generats per l'esquerra, alguns d'ells interns, d'altres externs;
- gramàtiques contextuais selectives per la dreta dèbils ($WRCGC$) obtingudes a partir de ($WLCGC$) canviant “esquerra” per “dreta”.

Podem plantejar-nos, en posteriors passos, similars qüestions a les formulades per a gramàtiques contextuais simples respecte a $L(G)$ i a $C(H)$. Sota quines condicions existeix un llenguatge M tal que $L(G)$ i $C(H)$ es corresponen l'un a l'altre dins el marc de treball de la connexió cadena-contextos de Galois respecte a M ? Què succeeix amb la unicitat d' M ? Pel fet que la connexió de Galois està donada per un parell (h, p) de funcions, $L(G)$ i $C(H)$ poden correspondre's l'una a l'altra de tres maneres diferents i obtenim tres tipus de compatibilitats entre el parell (G, H) de gramàtiques contextuais selectives i la connexió de Galois (h, p) : h -compatibilitat, p -compatibilitat i hp -compatibilitat. Podríem preguntar-nos qüestions més específiques: com podríem expressar en termes de les funcions de selecció f de G i g d' H la condició d'existència

d'un llenguatge M tal que el parell (G, H) de gramàtiques contextuais selectives sigui h -compatible, p -compatible o hp -compatible respecte a la connexió cadena-contextos de Galois associada a M ? Si M no és únic, hem de cercar el M més simple possible; per exemple, M regular serà preferit a M no-regular sempre i quan, però, sigui lliure del context. Podem obtenir una jerarquia de parells (G, H) respecte a la complexitat del llenguatge M que determina una connexió cadena-contextos de Galois g -compatible, p -compatible o hp -compatible a (G, H) . Alguns suggeriments que provenen de l'anglès o del francès poden ser obtinguts quan examinem classes distribucionals i classes contextuais en aquests llenguatges, a diferents nivells de gramaticalitat (veure (Marcus, 1981)). Seria una correspondència entre $f(s)$ i $g(c)$, quan s pertany a $g(c)$ i c pertany a $f(s)$.

Ens referirem ara a les gramàtiques contextuais selectives en la seva forma més general (Păun, 1982a; Păun, 1994): $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$, on V és el vocabulari, B és un llenguatge finit sobre V , $S_1, \dots, S_n$ són llenguatges sobre V i $C_1, \dots, C_n$ són conjunts finits de contextos sobre V . Una interpretació natural d' $S_1, \dots, S_n$ és la d'una seqüència de diferents nivells de gramaticalitat i, al respecte, el cas més interessant, en l'estudi dels llenguatges naturals, és aquell en el qual $S_1, \dots, S_n$ formen una cadena ascendent: cada S_i està contingut en el següent nivell de gramaticalitat S_{n+1} (i entre 1 i $n - 1$). Això és el que succeeix a (Du Feu i Marcus, 1973) i a (Marcus, 1981; Marcus, 1983). Com a pas preliminar, podríem algunes vegades treballar amb conjunts aparellats disjunts $S_1, \dots, S_n$, però en un pas posterior hem de prendre suc-

cessivament la seva unió sota la forma de $S_1, S_1 \cup S_2, \dots, S_1 \cup S_2 \cup S_3, \dots$ i finalment aconseguim una cadena ascendent de nivells de gramaticalitat.

La rèplica contextual de la forma generalitzada anterior de les gramàtiques contextuais selectives (*GCGC*) pot ser definida com un sistema $H = (V, C_0, (C_1, S_1), \dots, (C_n, S_n))$ on: V és el vocabulari; C_0 és un conjunt finit de contextos sobre V ; $C_1, \dots, C_n$ són conjunts de contextos sobre V ; mentre que $S_1, \dots, S_n$ són conjunts finits de cadenes sobre V . El conjunt C de contextos generats per H s'obté com el conjunt mínim C de contextos que satisfan els següents requirements:

- i) C_0 està contingut a C .
- ii) Donat un context $c = (u, v)$ de C i una cadena x d'algun selector S_i , podem estendre internament (u, v) , mitjançant x , si $u = u'u_i, v = v_iv'$; on (u_i, v_i) és un context de C_i . El context extès que pertany a C serà, o bé (ux, v) o bé (u, xv) .

Similarment podem definir el context extès d'altres maneres, usant, com en els casos anteriors, la combinatòria de les distincions esquerra-dret, intern-extern, dèbil-fort.

Les condicions relacionades amb l'ús maximal local de selectors (*Mlus*), ús maximal global de selectors (*Mgus*) en una *GCGC* així com les corresponents condicions minimal (*mlus, mgus*) van ser introduïdes a (Martín-Vide et al., 1995a) i poden ser expressades de la següent manera:

- *Mlus*: ús maximal local de selectors. Per a cada i a partir d'1 fins a n , cap cadena-selector d' S_i pot ser una subcadena d'una altra cadena-selector d' S_i ;
- *Mgus*: ús maximal global de selectors. Per a cada i , cap cadena-selector d' S_i pot ser una subcadena d'una altra cadena-selector d' $S_1 \cup S_2 \cup \dots \cup S_n$;
- *mlus*: ús minimal local de selectors. Per a cada i , cap cadena-selector d' S_i pot ser una super-cadena d'una altra cadena-selector d' S_i ;
- *mgus*: ús minimal global de selectors. Per a cada i , cap cadena-selector d' S_i pot ser una super-cadena d'una altra cadena-selector d' $S_1 \cup S_2 \cup \dots \cup S_n$.

Aquestes condicions expressen requeriments naturals d'economia i de simplicitat. De fet, voldríem tenir conjuntament *Mgus* i *mgus* i això és el que realment passa en l'anàlisi contextual de l'anglès de Du Feu, Marcus, (Du Feu i Marcus, 1973), on es consideren tres nivells diferents de gramaticalitat (per tant, $i = 1, 2, 3$). Ja que la manera usual en una anàlisi lingüística és prendre $S_1, \dots, S_n$ tal que formin una cadena ascendent, *Mgus* és equivalent, en aquest cas, a *Mlus* i umgs és equivalent a umls i, per tant, la distinció local-global no és rellevant més enllà (és millor dir que és neutralitzada).

2.8 Aprofundint en cadenes i contextos

Tots els fenòmens relacionats amb *GCGC* tenen la seva correspondència respecte a la seva rèplica contextual, discutida en l'apartat anterior. Els tòpics que vam apuntar sobre gramàtiques contextuais amb selecció poden ser extesos a *GCGC* i, *mutatis mutandis*, a la seva rèplica contextual. Aquesta transferència és part d'un principi més ampli, la dualitat entre cadenes i contextos. Donat un llenguatge L sobre V i dos contextos c_1 i c_2 sobre V , c_1 L -domina c_2 si qualsevol cadena acceptada per c_1 a L també és acceptada per c_2 a L . La categoria gramatical $G(C)$ generada pel conjunt C de contextos sobre V és, per definició, la unió de C amb el conjunt de tots els contextos L -dominats per qualsevol context de C . Per altra banda, si denotem per $S(C)$ el conjunt de totes les cadenes acceptades a L per qualsevol context de C , definim la clausura de Sestier $\varphi(C)$ com el conjunt de tots els contextos que accepten a L qualsevol cadena d' $S(C)$. Per tant, tots els problemes relacionats amb les gramàtiques contextuais i la clausura de Sestier tenen la seva equivalència quan les cadenes són reemplaçades pels contextos.

No obstant, podem procedir d'una manera més radical i reemplaçar el llenguatge L , el qual és un conjunt de cadenes, per un conjunt K de contextos sobre V . La relació de dominació entre contextos tindrà, en aquest cas, una varietat més gran. Per exemple, si $c_1 = (u_1, v_1)$ and $c_2 = (u_2, v_2)$, podem definir:

- la dominació interna per l'esquerra de c_2 per c_1 , quan per a qualsevol cadena x sobre V , tal que (u_1x, v_1) pertany a K , (u_2x, v_2) pertany a K ;
- la dominació externa per l'esquerra de c_2 per c_1 , quan per a qualsevol cadena x , tal que (xu_1, v_1) pertany a K , (xu_2, v_2) pertany a K ;
- la dominació interna per la dreta i la dominació externa per la dreta que s'obtenen similarment.

També podem definir, d'una manera òbvia,

- la dominació interna bilateral,
- la dominació externa bilateral ,
- la dominació per l'esquerra (quan es donen a la vegada les dominacions interna per l'esquerra i externa per l'esquerra),
- la dominació per la dreta, i alguns altres tipus de dominació.

Per a cadascuna d'elles obtenim el corresponent concepte de categoria gramatical. Per altra banda, donat un conjunt C de contextos sobre V , podem definir una varietat més gran de clausures de Sestier de C respecte a K : interna per l'esquerra, externa per l'esquerra, interna per la dreta, externa per la dreta, etc. En aquest cas, tenint en compte la identitat entre la clausura de Sestier i les categories de Kunze, la clausura de Sestier de C serà la unió de C amb tots els conjunts de contextos que són Kunze-dominats per C respecte a K .

No obstant, la dominació de Kunze entre C i un altre conjunt D de contextos pot ser vista de diferents maneres, seguint la combinatòria de distincions esquerra-dreta, intern-extern, fort-dèbil. Per exemple, la dominació interna per l'esquerra de Kunze entre C i D ve donada per la següent formulació: donada una cadena x tal que (ux, v) pertany a K per a qualsevol (u, v) de C , tindrem $(u'v, v')$ pertanyent K per a qualsevol (u', v') de D .

Òbviament, aquesta multiplicitat de possibilitats ens porta a una tipologia molt rica de situacions contextuais i de connexions de Galois. Però aquesta riquesa és secundada per la realitat dels llenguatges/llengües naturals. Pel fet que una similar riquesa tipològica també es troba en les gramàtiques contextuais, se segueix que, per a cada tipus per separat, podem incrementar els problemes ja discutits per a gramàtiques contextuais simples i per a gramàtiques contextuais amb selecció.

Existeix també la possibilitat de considerar, seguint (Martín-Vide et al., 1995a), l'ús dels selectors dispersament i estendre la noció de context a un conjunt finit ordenat de cadenes.

Malgrat la interpretació d' $S_1, \dots, S_n$, respectivament de $C_1, \dots, C_n$ com a nivells de gramaticalitat definits, en el primer cas, mitjançant els conjunts de cadenes i, en el segon cas, mitjançant els conjunts de contextos, existeixen també altres interpretacions possibles, que van més enllà dels llenguatges/llengües naturals. És suficient, al respecte, amb adoptar una nova interpretació dels elements del vocabulari V . Per exemple, si aquests elements són

síntomes¹ de diverses malalties possibles, aleshores les cadenes sobre V poden definir diverses síndromes² possibles (veure Celan-Marcus a (Celan i Marcus, 1973), i Popescu a (Popescu, 1977)). Una anàlisi clínica implica un grau d'ambigüïtat en l'establiment d'un diagnòstic; els símptomes patognomònics³ (aquells que estan desproveïts d'ambigüïtat) són assequibles només algunes vegades. Normalment, una síndrome ens condueix a un estat de dubte entre n malalties possibles diferents i els selectors $S_1, \dots, S_n$ estan modelant precisament aquestes diferents malalties, cadascuna d'elles representada pel conjunt de síndromes corresponents.

Finalment, podem observar que el requeriment expressat a (Marcus et al., 1995a) de tenir els selectors el més simple possibles (finites o, com a màxim, regulars) se satisfà completament en ambdues interpretacions dels selectors considerades abans: els nivells de gramaticalitat i els conjunts de síndromes d'algunes malalties concretes. En els llenguatges/llengües naturals, pràcticament treballem amb nivells finits de gramaticalitat i després, mitjançant inferència gramatical, procedim a una aproximació amb conjunts infinits; al respecte, cal veure (Savitch, 1993) per a una anàlisi més acurada de la infinitud dels llenguatges naturals.

¹Un *síntoma* és un fenomen perceptible provocat en l'organisme per una malaltia.

²Una *síndrome* és un conjunt de símptomes i signes que defineixen clínicament una malaltia.

³Un *síntoma patognomònic* és un símptoma característic d'una determinada malaltia, amb el qual podem establir d'una manera segura la seva diagnosi.

2.9 Un nivell més alt d'abstracció

Nivells més alts d'abstracció i nous aspectes lingüístics de les gramàtiques contextuals poden ser obtinguts mitjançant l'explotació de la idea presentada en l'última part de la secció anterior: reinterpretar d'una manera significativa els elements del vocabulari V . En molts apartats de la discussió anterior, els elements de V eren interpretats com a morfemes o formes flexionals. Considerem ara les interpretacions dels elements de V com a paradigmes flexionals. Això vol dir que cada element de V serà el conjunt de totes les formes flexionals d'una unitat lèxica. Per exemple, una forma nominal estaria en el mateix conjunt que totes les seves formes flexionals que reflecteixen la variabilitat respecte a diverses categories morfològiques (cas, nombre, etc.). Òbviament, aquesta operació és menys rellevant en anglès, en vistes a la seva pobra morfologia. També el francès, degut a la seva preferible declinació analítica, té classes flexionals pobres, però els paradigmes verbals són més rics. L'alemany és més interessant en aquest aspecte. Molt més rellevants són els llenguatges amb una morfologia sintètica, com és el cas del llatí i de diversos llenguatges eslaus. Aquesta situació explica el perquè el primer model matemàtic que té com a punt de partida la partició de V en classes flexionals va ser proposat per un erudit rus, Kulagina, (Kulagina, 1958), i desenvolupat posteriorment per diversos autors de països eslaus (Uspenskij a (Uspenskij, 1957b), Revzin a (Revzin, 1962; Revzin, 1967b), Shreider a (Shreider, 1970; Shreider, 1973), Trybulec a (Trybulec, 1967b), Semeniuk-Polkowska (Semeniuk-Polkowska, 1982). Un de-

senvolupament matemàtic posterior d'aquest model va ser proposat per Marcus, (Marcus, 1962), que va aplicar-lo a la tipologia dels llenguatges (Marcus, 1963a; Marcus, 1963c) i a l'anàlisi comparativa del gènere gramatical en llengües romàniques i eslaves (Marcus, 1970a).

Òbviament, si considerem que les formes flexionals ens duen a una partició del vocabulari adoptem una simplificació de la realitat lingüística, ja que algunes formes flexionals són comuns amb diferents paradigmes; aquest fet és molt freqüent en l'anglès, però menys freqüent en llenguatges amb morfologia sintètica, pels quals el model considerat va ser concebut principalment.

L'importat descobriment fet per Kulagina, (Kulagina, 1958), és el fet que, quan els elements de V són interpretats com a paradigmes flexionals, aleshores les corresponents classes distribucionals estan modelant les parts de l'oració (nom, verb, adjectiu qualificatiu, etc.). Començant per una partició P de V , Kulagina defineix la derivada P' de P respecte a L , com una nova partició de V , els termes de la qual són unions de termes L -equivalents de P ; en altres paraules, els termes de P' són les classes distribucionals respecte a L , quan els elements del vocabulari deixen de ser els elements de V per a passar a ser els termes de P . Una cadena $P_1 P_2 \dots P_n$ d'aquest nou vocabulari és L -ben-format (L és un llenguatge donat a V) si existeix una cadena $a_1 a_2 \dots a_n$ sobre V que pertany a L i tal que a_1 pertany a P_1 , a_2 pertany a P_2 , $\dots$, a_n pertany a P_n . Si tractem d'iterar aquesta operació, no ho aconseguirem, ja que un important teorema donat per Kulagina afirma que $P'' = P'$.

Per a un refinament matemàtic de la derivada d'una partició es pot consultar Solomon, (Solomon, 1973). Mitjançant el model de Kulagina, les gramàtiques contextuais de diversos tipus i les seves relacions amb operacions de clausura i connexions de Galois poden rebre noves interpretacions lingüístiques, que abracen no només el comportament contextual de les cadenes lingüístiques, sinó també aspectes paradigmàtics importants.

Altres aspectes de les connexions de Galois i els operadors de clausura són considerats per Păun a (Păun, 1975a), Coardos a (Coardos, 1976) i Schwartz a (Schwartz, 1982); mentre que un concepte general de la dominació contextual, incloent-hi totes les particulars nocions conegudes al respecte va ser proposada per Schwarz a (Schwartz, 1976). Per altra part, Nebeský a (Nebeský, 1968) va proposar un concepte general de sistema de llenguatges. Altres aspectes contextuais van ser considerats per Dobrušin a (Dobrušin, 1961), Ito a (Ito, 1971; Ito, 1972b), Dincă (Dincă, 1976) i Calude a (Calude, 1976a). Altres aspectes de les relacions de dominació van ser analitzats per Mayoh a (Mayoh, 1967), Nebeský a (Nebeský, 1965) i Cazimir a (Cazimir, 1975a; Cazimir, 1975b). Finalment, una caracterització completa de les classes de Chomsky, en termes de models analítics, va ser donada per Novotný a (Novotný, 1971), el qual també va donar un marc de treball general per a l'estudi dels llenguatges a (Novotný, 1972).

2.10 El poder generatiu

Les úniques contribucions relacionades amb la rellevància generativa de les gramàtiques contextuals respecte als llenguatges naturals són les de (Marcus et al., 1995a; Marcus et al., 1996b; Marcus et al., 1996a); on s’observa que les gramàtiques contextuals internes amb ús maximal de selectors són les més apropiades (respecte a altres classes de gramàtiques contextuals) per tal de modelitzar la capacitat generativa dels llenguatges naturals (i també d’alguns llenguatges artificials), ja que són capaces de descriure totes les restriccions usuals que apareixen en aquests llenguatges que provoquen la seva no independència del context: reduplicació, dependències a llarga distància i concordanances.

Els principals candidats per a modelitzar els aspectes generatius dels llenguatges naturals ja van ser considerats en les gramàtiques indexades “indexed grammars”^a (Aho, 1968), les gramàtiques amb nucli “head grammars”^a (Pollard, 1984), i les gramàtiques d’adjunció d’arbres “tree adjunct grammars”^a (Joshi et al., 1975); veure també (Joshi, 1985)), així com algunes de les seves generalitzacions. Comparant els arguments donats a (Marcus et al., 1995a; Marcus et al., 1996b; Marcus et al., 1996a) amb els arguments donats a favor dels anteriors tipus de gramàtiques per Gazdar i Pullum a (Gazdar i Pullum, 1985) i Partee, Meulen i Wall a (Partee et al., 1990), se segueix que les gramàtiques contextuals no es queden curtes pel que fa a la seva rellevància en la capacitat de generació dels llenguatges naturals. També tindrem en compte

el fet que, més enllà de les seves capacitats generatives, les gramàtiques contextuais estan, per les seves motivacions inicials i per les seves estretes relacions amb els aspectes morfològics, distribucionals i contextuais, en una interacció genuïna amb els llenguatges/llengües naturals (veure també Manaster Ramera (Manaster-Ramer, 1994)).

2.11 Altres línies de recerca

A continuació veurem algunes altres línies de recerca com poden ser les gramàtiques contextuais restringides, els sistemes de gramàtiques i els esquemes d'empalmament contextuais.

Definim el concepte de gramàtica contextual restringida simple com un sistema $G = (V, \mathcal{L}, \mathcal{C})$, on $\mathcal{L}$ és un conjunt finit de llenguatges finits $L_1, L_2, \dots, L_n$ sobre V , mentre que $\mathcal{C}$ és un conjunt finit de conjunts finits $C_1, C_2, \dots, C_p$ de contextos sobre V . Aquesta gramàtica genera cadenes i contextos. Per tal de generar cadenes, definim el llenguatge $L(G)$ generat per G com el llenguatge més petit L sobre V tal que: cada L_i ($1 \leq i \leq n$) està contingut a L ; si $x \in L$, aleshores existeix i entre 1 i p i j entre 1 i n , tal que la cadena $uxyv$ pertany a L per a qualsevol (u, v) de C_i i qualsevol y de L_j . Per tal de generar contextos, definim el conjunt $C(G)$ de contextos generat per G com el conjunt més petit C de contextos sobre V tal que: cada C_i ($1 \leq i \leq p$) està contingut a C ; si (u, v) pertany a C , aleshores existeix i entre 1 i n pel qual el context (ux, v)

pertany a C per a qualsevol x de L_i .

Hem definit abans una gramàtica contextual restringida simple que treballa en mode intern per l'esquerra, per la qual cosa podem anomenar a aquesta gramàtica una gramàtica contextual restringida simple interna per l'esquerra. D'igual manera podem definir les altres variants relacionades amb les distincions esquerra-dreta, interna-externa, forta-dèbil, etc. També es pot estendre el tipus restringit a gramàtiques contextuais amb selecció i les seves diverses generalitzacions.

La motivació lingüística del tipus restringit de gramàtiques contextuais està relacionada amb el fet que les unitats elementals de l'anàlisi contextual són les classes distribucionals de cadenes (i.e., cadenes que tenen el mateix comportament contextual) i les classes d'equivalència de contextos (i.e., contextos que accepten exactament les mateixes cadenes); aquests dos tipus de situacions es reflecteixen mitjançant els conjunts $\mathcal{L}$ i $\mathcal{C}$.

Aleshores, com podríem enllaçar aquestes idees amb la d'un sistema de gramàtiques ("grammar system") (veure (Csuha-j-Varju et al., 1994)) ? Si la idea bàsica d'un sistema de gramàtiques és la d'una cadena de gramàtiques on, sortides d'una d'elles, esdevenen entrades per a una altra; aleshores aquesta idea està potencialment continguda en la de la clausura de Sestier. En aquest cas, l'operador c de $\varphi(X) = s(c(X))$ té com a entrades diversos llenguatges sobre V i com a sortides conjunts de contextos sobre V , mentre que el següent operador s té com a entrades precisament les sortides del primer operador c .

D'aquesta manera, c i s fan de mediadors entre les entrades i les sortides de l'operador de clausura de Sestier φ . El corresponen sistema de gramàtiques contextuals simulària aquesta situació. També hem de tenir en consideració les següents alternatives a la clausura de Sestier: definim els operadors $e(C) =$ el conjunt de cadenes acceptades per com a mínim un context de C i $f(S) =$ el conjunt de contextos que accepten com a mínim una cadena d' S i després considerar els operadors compostos $s(f(S))$, $e(f(S))$ i $e(c(S))$.

Una altra possible perspectiva per a gramàtiques contextuals està relacionada amb el concepte d'*esquema d'empalmament* ("splicing scheme"), introduït per Head, (Head, 1987) (veure també (Head et al., 1996)). Un esquema d'empalmament ("splicing scheme") és una tripla $S = (V, T, P)$, sobre V és un conjunt finit, T és un subconjunt finit d' $V^* \times V^* \times V^*$ i P és una relació binària a T que satisfà la següent condició: Si p, x, q, u, y, v pertanyen a V^* , (p, x, q) i (u, y, v) pertanyen a T , i $(p, x, q)P(u, y, v)$, aleshores $x = y$. El conjunt V s'anomena el *vocabulari*, T les *triples* i P la relació d'*acoblament* de l'esquema. D'una manera, l'esquema d'empalmament (V, T, P) actua sobre cada parell de cadenes de la forma $hpxqk$ i $wuxvz$, on cadascuna de les lletres és una cadena sobre V^* i $(p, x, q)P(u, x, v)$ per tal de produir el parell $hpxvz$ i $wuxqk$.

A partir de les definicions anteriors de (Head, 1987) es podria donar una versió contextual, que permetés agafar avantatge de l'aparell relacionat amb l'anàlisi contextual algebraic (connexions de Galois, clausures de Sestier, diversos tipus de dominacions contextuals, etc). Sigui V un vocabulari no-buit, C

un conjunt finit de contextos sobre V , L un llenguatge finit obtingut emplaçant algunes cadenes de V^* en alguns contextos de C i P una relació binària a L , tal que si $pxqPuyv$ (amb (p, q) i (u, v) pertanyent C), aleshores $x = y$. Anomenarem al sistema $G = (V, C, L, P)$ un *esquema d'empalmament contextual*. El llenguatge $L(G)$ generat per G és el conjunt de tots els parells de cadenes de la forma $kpxvz$ i $wuxqk$, on (p, q) i (u, v) pertanyen a C , i pxq i uxv són cadenes d' L que estan en la relació P , mentre que h, k, w , i z són cadenes sobre V . És fàcil de veure que els esquemes d'empalmament contextuais són equivalents als esquemes d'empalmament inicialment proposats per Head a (Head, 1987; Head, 1992). Falta verificar com podem explotar aquesta versió contextual d'un esquema d'empalmament. Per exemple, la relació binària P podria ser introduïda a C en comptes de ser definida a L : dos contextos (p, q) i (u, v) de C estan en relació P si per a qualsevol cadena x acceptada per (p, q) a L i per a qualsevol cadena y acceptada a L per (u, v) tenim $x = y$.

Com a observació final, podríem afegir que el monogràfic compost per dos volums de van Helden, (van Helden, 1993), desenvolupa, en el transcurs de més de tres-centes pàgines, una anàlisi molt acurada de diversos llenguatges/llengües naturals, usant precisament les eines contextuais discutides abans.

Notes bibliogràfiques. A més a més de les referències bibliogràfiques donades anteriorment, hom pot trobar una extensa introducció a l'aplicació dels models matemàtics als llenguatges naturals i artificials (tipus, bases, problemes inherents, etc.). a (Marcus, 1996). Per a una introducció detallada a la relació

entre models algebraics analítics, gramàtiques contextuais i llenguatges naturals es pot consultar el prefaci donat per Păun a (Păun, 1996a) i, sobretot, el donat per Marcus a (Păun i Rozenberg, 1996). Altres introduccions podrien ser (Miquel-Vergés, 1994a; Miquel-Vergés, 1994b; Miquel-Vergés, 1995b).

Capítol 3

Teoria de Llenguatges Formals

Per tal de fer el text (tan) independent (com es pugui) dels coneixements en teoria de llenguatges formals, especificarem aquí una sèrie de nocions bàsiques i resultats en aquesta àrea; d'aquesta manera, també fixarem la notació i terminologia que usarem en tota la tesi.

3.1 Nocions bàsiques

Qualsevol conjunt no-buit i finit de símbols abstractes es denomina un *alfabet* o *vocabulari*. Si $V = \{a_1, a_2, \dots, a_n\}$ és un vocabulari, aleshores qualsevol seqüència de la forma $x = a_{i_1}a_{i_2} \dots a_{i_m}$, d'elements de V , es denomina una *cadena* o *paraula* (sobre o de V). El nombre de símbols d'una cadena x (m per a l'anterior x) és la *longitud* d'aquesta cadena i es denota per $|x|$. El

nombre d'ocurrències de símbols d'un vocabulari U en una cadena x sobre V , per a $U \subseteq V$ (és a dir, el nombre de vegades que cada un dels símbols del vocabulari apareix en una determinada cadena), es denota per $|x|_U$; quan $U = \{a\}$ escriurem $|x|_a$ en comptes de $|x|_{\{a\}}$. La cadena buida es denota per λ . El conjunt de totes les cadenes sobre V es denota per V^* . Per a $x, y \in V^*$, $x = a_{i_1} \dots a_{i_m}$, $y = a_{j_1} \dots a_{j_r}$, $m, r \geq 0$, la cadena $a_{i_1} \dots a_{i_m} a_{j_1} \dots a_{j_r}$ es denota per xy i es denomina la *concatenació* d' x i y . D'aquesta manera, V^* és el monide lliure generat per V respecte a l'operació de concatenació i a la identitat λ (l'operació de concatenació és associativa i $x\lambda = \lambda x = x$, per a tot $x \in V^*$). Denotem per $V^+ = V^* - \{\lambda\}$ (és el conjunt de cadenes no-buides sobre V).

En general, usarem com a sinònims els termes “alfabeti” “vocabulari”, així com “paraula” “cadena”.

Si $x, y \in V^*$ i $y = uxv$, per algun $u, v \in V^*$, aleshores direm que x és una *subcadena* (o subparaula) d' y ; si $y = xv$, aleshores x és un *prefix* d' y , i si $y = ux$, aleshores x és un *sufix* d' y . Per a $y \in V^*$, denotem per $Sub(y)$, $Pref(y)$, $Suf(y)$ el conjunt de subcadenes, prefixos, i sufixos d' y , respectivament. Una subcadena (prefix o sufix) d'una cadena y es diu que és *pròpia* si és diferent d' y ; els conjunts de subcadenes pròpies, prefixos propis, i sufixos propis d'una cadena y es denoten per $PSub(y)$, $PPref(y)$, i $PSuf(y)$, respectivament.

Cada subconjunt L de V^* es denomina un *llenguatge* sobre V ; V^* és el *llenguatge universal* sobre V . Un llenguatge $L \subseteq V^+$ es denomina *λ -lliure* (és a dir, que no conté la cadena buida). Si $L \subseteq V^*$, aleshores el $U \subseteq V$

més petit (en el sentit de la relació d'inclusió) tal que $L \subseteq U^*$ el denotarem per $\text{alph}(L)$. També escriurem $\text{alph}(w)$ en el cas del vocabulari més petit que conté la cadena w .

Un parell (u, v) de cadenes sobre algun vocabulari V es denomina un *context*. Direm que el context (u, v) *accepta* la cadena x respecte a un llenguatge $L \subseteq V^*$ (o, equivalentment, que x és acceptat per (u, v)) si $uxv \in L$.

3.2 Operacions amb llenguatges

Sigui $V = \{a_1, \dots, a_n\}$ un vocabulari donat. Com que els llenguatges sobre V són conjunts de cadenes, podem definir, entre d'altres, les operacions d'*unió*, *intersecció*, *diferència*, i *complementarietat* (respecte a V^*) de manera equivalent a la teoria de conjunts. Aleshores tenim:

Definició 3.2.1 Si denotem per “ $\cup$ ” l'operació d'*unió*, aleshores, la unió de dos llenguatges $L_1, L_2 \subseteq V^*$ és el llenguatge

$$\cup(L_1, L_2) = L_1 \cup L_2 = \{x \mid x \in L_1 \vee x \in L_2\}.$$

Definició 3.2.2 Si denotem per “ $\cap$ ” l'operació d'*intersecció*, aleshores, la intersecció de dos llenguatges $L_1, L_2 \subseteq V^*$ és el llenguatge

$$\cap(L_1, L_2) = L_1 \cap L_2 = \{x \mid x \in L_1 \wedge x \in L_2\}.$$

Definició 3.2.3 Si denotem per “ $-$ ” l’operació de *diferència*, aleshores, la diferència entre dos llenguatges $L_1, L_2 \subseteq V^*$ és el llenguatge

$$-(L_1, L_2) = L_1 - L_2 = \{x \mid x \in L_1 \wedge x \notin L_2\}.$$

Definició 3.2.4 Si denotem per “ $'$ ” l’operació de *complementarietat*, aleshores, la complementarietat (o el complementari) d’un llenguatge $L \subseteq V^*$ és el llenguatge

$$'(L) = L' = \{x \mid x \in U \wedge x \notin L\}, \text{ on } U \text{ denota l'univers del discurs.}$$

Hi ha també una sèrie d’operacions específiques dels llenguatges.

En general, qualsevol operació monària α entre cadenes, definida per $\alpha : V^* \longrightarrow 2^{U^*}$ (on, per a un conjunt X , denotem per 2^X el conjunt de tots els subconjunts d’ X ; és a dir, $\mathcal{P}(X)$ o conjunt de parts d’ X), pot ser extesa a llenguatges sobre V mitjançant

$$\alpha(L) = \bigcup_{x \in L} \alpha(x),$$

mentre que una operació binària β entre cadenes, definida per $\beta : V^* \times V^* \longrightarrow 2^{U^*}$, pot ser extesa a una operació entre llenguatges mitjançant

$$\beta(L_1, L_2) = \bigcup_{x \in L_1, y \in L_2} \beta(x, y).$$

Per exemple:

Definició 3.2.5 Si denotem per “ $\cdot$ ” l’operació de *concatenació* (o multiplicació, o multiplicació directa, o producte), aleshores la concatenació de dos llenguatges $L_1, L_2 \subseteq V^*$ és el llenguatge

$$\cdot(L_1, L_2) = L_1 \cdot L_2 = \{xy \mid x \in L_1, y \in L_2\}.$$

Per definició, $L^k = \underbrace{L \cdot L \cdot \dots \cdot L}_k$, i $L^0 = \{\lambda\}$

Definició 3.2.6 Si denotem per “ *mi* ” l’operació d’*imatge mirall* (o, simplement, mirall), i si $x = a_{i_1}a_{i_2} \dots a_{i_r}$ és una cadena sobre V , $a_{i_j} \in V, 1 \leq j \leq r$, aleshores la imatge mirall d’ x és la cadena

$$mi(x) = a_{i_r}a_{i_{r-1}} \dots a_{i_2}a_{i_1}.$$

Aleshores, per a un llenguatge $L \subseteq V^*$, la seva imatge mirall és el llenguatge

$$mi(L) = \{mi(x) \mid x \in L\}.$$

Les operacions de *Sub*, *Pref*, i *Suf* també poden ser exteses a llenguatges de la mateixa manera. Aleshores,

Definició 3.2.7 Si denotem per “ *Sub* ” l’operació de *subcadenaes*, aleshores, el conjunt de subcadenaes d’un llenguatge $L \subseteq V^*$ és el llenguatge

$$Sub(L) = \{z \mid \exists x, y \in V^* \text{ tal que } xzy \in L\}.$$

Definició 3.2.8 Si denotem per “ *Pref* ” l’operació de *prefixos*, aleshores, el conjunt de prefixos d’un llenguatge $L \subseteq V^*$ és el llenguatge

$$Pref(L) = \{x \mid \exists y \in V^* \text{ tal que } xy \in L\} = L/V^*.$$

Definició 3.2.9 Si denotem per “ *Suf* ” l’operació de *suffixos*, aleshores, el conjunt de suffixos d’un llenguatge $L \subseteq V^*$ és el llenguatge

$$Suf(L) = \{y \mid \exists x \in V^* \text{ tal que } xy \in L\} = V^* \setminus L.$$

Definició 3.2.10 Si denotem per “ $*$ ” l’operació de *clausura de Kleene* (o iteració, o clausura concatenativa, o estrella de Kleene), aleshores, la clausura de Kleen d’un llenguatge $L \subseteq V^*$ és el llenguatge

$$*(L) = L^* = \bigcup_{i \geq 0} L^i.$$

Per definició, $L^0 = \{\lambda\}$ i $L^{i+1} = L^i L, i \geq 1$.

Podem definir també la clausura de Kleen λ -lliure “ $+$ ” (o iteració truncada, o clausura concatenativa λ -lliure) com el llenguatge

$$+(L) = L^+ = \bigcup_{i \geq 1} L^i.$$

Es pot observar que aquestes notacions són consistents amb les definicions de V^* i de V^+ donades prèviament a 3.1.

Definició 3.2.11 Si denotem per “ $\sqcup$ ” l’operació de *barreja* (o mescla), aleshores, la barreja de dues cadenes $x, y \in V^*$ es defineix com

$$x \sqcup y = \{x_1 y_1 \dots x_n y_n \mid x = x_1 \dots x_n, y = y_1 \dots y_n, x_i, y_i \in V^*, 1 \leq i \leq n, n \geq 1\}.$$

Aleshores, la barreja de dos llenguatges $L_1, L_2 \subseteq V^*$ és el llenguatge

$$\sqcup (L_1, L_2) = L_1 \sqcup L_2 = \bigcup_{x \in L_1, y \in L_2} x \sqcup y.$$

Definició 3.2.12 Considerem ara dos vocabularis, V and U , i una funció s definida per $s : V \longrightarrow 2^{U^*}$. Extenem aquesta funció a V^* mitjançant

$$s(\lambda) = \{\lambda\}, s(xy) = s(x)s(y), x, y \in V^*.$$

Aleshores, direm que s és una *substitució*, i ho extendrem de manera natural a llenguatges:

Donat un llenguatge $L \subseteq V^*$, aleshores

$$s(L) = \bigcup_{x \in L} s(x).$$

Si $\text{card}(s(a)) < \infty$ per a tot $a \in V$, aleshores s és una *substitució finita*.

Si $\text{card}(s(a)) = 1$ per a cada $a \in V$, aleshores s és un *morfisme* i escriurem $s(a) = x$ en comptes de $s(a) = \{x\}$.

Si $\lambda \notin s(a)$ per a tot $a \in V$, aleshores direm que s és una substitució λ -lliure (morfisme, respectivament). Un morfisme $h : V^* \longrightarrow U^*$ tal que $|h(a)| \leq 1$ per a tot $a \in V$ es denomina una *codificació dèbil*; una codificació dèbil λ -lliure es denomina una *codificació* (en aquest cas és una funció lletra-a-letra). Una funció $h : V^* \longrightarrow U^*$ per a la qual existeix una constant k tal que $k|h(x)| \geq |x|$ per a tot $x \in V^+$ es diu que és *amb esborrament lineal*. Un morfisme $h : (V \cup \{c\})^* \longrightarrow V^*$ tal que $h(a) \in V^+$ per a $a \in V$ i $h(c) = \lambda$ es denomina *k-restringit* sobre cada llenguatge $L \subseteq (V\{\lambda, c, c^2, \dots, c^k\})^*$. Un morfisme *k-restringit* sobre un llenguatge, per algun k , es denomina *restringit*. Clarament, un morfisme restringit és una funció amb esborrament lineal.

Si $h : V^* \longrightarrow U^*$ és un morfisme, aleshores la funció $h^{-1} : U^* \longrightarrow 2^{V^*}$ definida per $h^{-1}(x) = \{y \in V^* \mid h(y) = x\}$, $x \in U^*$, es denomina el *morfisme invers* associat a h , que denotarem per “ h^{-1} ”.

Definició 3.2.13 Si denotem per “ $\setminus$ ” l’operació de *quocient per l’esquerra* (o divisió per l’esquerra), aleshores el quocient per l’esquerra d’un llenguatge $L_1 \subseteq V^*$ respecte d’un altre llenguatge $L_2 \subseteq V^*$ és el llenguatge

$$\setminus(L_1, L_2) = L_2 \setminus L_1 = \{x \mid yx \in L_1, y \in L_2\}.$$

Si $L_2 = \{y\}$, $y \in V^*$, aleshores escriurem $\partial_y^e(L_1)$ en comptes de $\{y\} \setminus L_1$ i direm que aquesta és la *derivada per l’esquerra*, denotada per “ ∂^e ”, d’ L_1 respecte a y

$$\partial_y^e(L_1) = \{x \mid yx \in L_1\}.$$

De la mateixa manera es defineixen el quocient (o divisió) per la dreta “ $/$ ”, com el llenguatge

$$/(L_1, L_2) = L_1 / L_2 = \{x \mid xy \in L_1, y \in L_2\},$$

i la derivada per la dreta, denotada per “ ∂^d ”, d’ L_1 respecte a y , com el llenguatge

$$\partial_y^d(L_1) = \{x \mid xy \in L_1\}.$$

Definició 3.2.14 Sigui F una família de llenguatges, α una operació monària entre llenguatges i β una operació binària entre llenguatges. Direm que la

família F és *tancada sota* l'operació α (β , respectivament) si per a cada $L \in F$ ($L_1, L_2 \in F$, respectivament), tenim que $\alpha(L) \in F$ ($\beta(L_1, L_2) \in F$).

Una família F és *tancada sota substitució* amb llenguatges d'una família F' si per a cada $L \in F$, $L \subseteq V^*$, i $s : V^* \longrightarrow 2^{U^*}$ una substitució tal que $s(a) \in F'$ per a cada $a \in V$, tenim que $s(L) \in F$. La família F s'anomena simplement *tancada sota substitució* si és tancada sota substitució amb llenguatges de la mateixa família.

3.3 Gramàtiques de Chomsky

En general, una gramàtica és un mecanisme finit mitjançant el qual podem *generar* els elements d'un llenguatge. Molts tipus de gramàtiques són casos particulars de sistemes de reescriptura.

Definició 3.3.1 Un *sistema de reescriptura* és un parell $\gamma = (V, P)$, on V és un vocabulari i P és un subconjunt finit de $V^* \times V^*$; els elements (u, v) de P s'escriuen de la forma $u \rightarrow v$ i s'anomenen regles de reescriptura o produccions (o, simplement, regles o produccions). Per a $x, y \in V^*$ escriurem $x \Longrightarrow_\gamma y$ si $x = x_1 u x_2, y = x_1 v x_2$, per algun $u \rightarrow v \in P$ i $x_1, x_2 \in V^*$. Si el sistema de reescriptura γ se sobreentén, aleshores escriurem $\Longrightarrow$ en comptes de $\Longrightarrow_\gamma$. La clausura reflexiva i transitiva de $\Longrightarrow$ es denota per $\Longrightarrow^*$.

Definició 3.3.2 Si afegim un *axioma* a un sistema de reescriptura i totes les regles $u \rightarrow v$ tenen $u \neq \lambda$, aleshores obtenim la noció de *gramàtica pura*. Per a una gramàtica pura $G = (V, w, P)$, on $w \in V^*$ és l'axioma, definim el *llenguatge generat* per G , denotat per $L(G)$, com

$$L(G) = \{x \in V^* \mid w \Longrightarrow^* x\}.$$

Definició 3.3.3 Una *gramàtica de Chomsky* és una quàdrupla $G = (V_N, V_T, S, P)$, on V_N i V_T són vocabularis disjunts, $S \in V_N$, i P és un subconjunt finit de $(V_N \cup V_T)^* V_N (V_N \cup V_T)^* \times (V_N \cup V_T)^*$. Els elements de V_N s'anomenen *no-terminals* o *auxiliars*, mentre que els de V_T s'anomenen *terminals*, S és l'*axioma* de la gramàtica, i P és el conjunt de *regles de producció*. Igual que en el cas dels sistemes de reescriptura escriurem les regles de la forma $u \rightarrow v$. Cal notar que u conté com a mínim un no-terminal.

Definició 3.3.4 Respecte a una gramàtica G definim la relació $\Longrightarrow_G$ sobre $(V_N \cup V_T)^*$ d'igual manera que pel cas del sistema de reescriptura $(V_N \cup V_T, P)$: per a $x, y \in (V_N \cup V_T)^*$ escriurem $x \Longrightarrow_G y$ si $x = x_1 u x_2$, $y = x_1 v x_2$, per algun $x_1, x_2 \in (V_N \cup V_T)^*$ i $u \rightarrow v$ una regla de G . Si G se sobreentén, aleshores escriurem $\Longrightarrow$ en comptes de $\Longrightarrow_G$. Aleshores, direm que x deriva directament (o produeix) y en la gramàtica G . Si a $x_1 u x_2 \Longrightarrow x_1 v x_2$ tenim que $x_1 \in V_T^*$, aleshores la derivació és la *més-a-l'esquerra* possible i escriurem $\Longrightarrow_{left}$ o bé $\Longrightarrow_l$. La clausura reflexiva i transitiva de la relació $\Longrightarrow$ es denota per $\Longrightarrow^*$. El llenguatge generat per G , denotat per $L(G)$, es defineix com $L(G) = \{x \in V_T^* \mid S \Longrightarrow^* x\}$. Quan usem la derivació més-a-l'esquerra, aleshores ho denotem mitjancament $L_{left}(G)$ o bé $L_l(G)$.

Definició 3.3.5 D'acord amb la forma de les regles, les gramàtiques de Chomsky es clasifiquen de la següent manera:

- a. gramàtiques *de tipus 0* (o gramàtiques d'estructura de frase): són gramàtiques com les anteriors, sense cap tipus de restricció;
- b. gramàtiques *de longitud creixent* (o monòtones): són gramàtiques amb $|u| \leq |v|$ per a cada regla $u \rightarrow v$;
- c. gramàtiques *de tipus 1* (o sensibles al context, o “context-sensitive”): són gramàtiques on totes les regles són de la forma $w_1Aw_2 \rightarrow w_1xw_2$, per a $A \in V_N, w_1, w_2 \in (V_N \cup V_T)^*, x \in (V_N \cup V_T)^+$;
- d. gramàtiques *de tipus 2* (o lliures del context, o “context-free”): són gramàtiques on totes les regles són de la forma $A \rightarrow x$, per a $A \in V_N, x \in (V_N \cup V_T)^*$;
- e. gramàtiques *lineals*: són gramàtiques on totes les regles són de la forma $A \rightarrow x, A \rightarrow x_1Bx_2$, per a $A, B \in V_N, x, x_1, x_2 \in V_T^*$;
- f. gramàtiques *lineals a la dreta*: són gramàtiques on totes les regles són d'una de les dues següents formes $A \rightarrow x, A \rightarrow xB$, per a $A, B \in V_N, x \in V_T^*$;
- g. gramàtiques *lineals a l'esquerra*: són gramàtiques on totes les regles són d'una de les dues següents formes $A \rightarrow x, A \rightarrow Bx$, per a $A, B \in V_N, x \in V_T^*$;

- h. gramàtiques *de tipus 3* (o regulars): són gramàtiques on totes les regles són d'una de les dues següents formes $A \rightarrow a$, $A \rightarrow aB$, per a $A, B \in V_N$, $a \in V_T$.

Pel fet que en les gramàtiques monòtones, sensibles al context, i regulars no podem tenir una regla de la forma $A \rightarrow \lambda$ (i, per tant, no podem generar la cadena buida), s'adopta la següent *convenció de compleció*: en qualsevol gramàtica permetrem una regla de la forma $S \rightarrow \lambda$, que faciliti que S , l'axioma, no aparegui en la banda dreta de cap regla (la regla $S \rightarrow \lambda$ només introdueix la cadena buida, però no pot esborrar el símbol S d'una cadena generada en, com a mínim, un pas en la derivació). Una gramàtica que conté només regles λ -lliure, exceptuant possiblement la regla de compleció $S \rightarrow \lambda$ definida anteriorment, es denomina λ -lliure.

Dues gramàtiques que generen el mateix llenguatge es diu que són *equivalents*.

És conegut que les gramàtiques de tipus b i c generen la mateixa família de llenguatges; això també és cert per a gramàtiques de tipus f, g, i h. Denotem per RE (o $\mathcal{L}_0$), CS (o $\mathcal{L}_1$), CF (o $\mathcal{L}_2$), LIN , i REG (o $\mathcal{L}_3$) les famílies de llenguatges generades per gramàtiques de tipus-0, sensibles al context (o de tipus-1), lliures del context (o de tipus-2), lineals, i regulars (o de tipus-3), respectivament. També denotarem per FIN la família de llenguatges finits, i, per ARB (d'arbitrària), la família de tots els llenguatges. Algunes vegades, usarem les abreviacions RE, CS, CF, LIN, REG per a identificar les correspo-

nents classes de gramàtiques.

Es compleix la següent inclusió estricta:

$$FIN \subset REG \subset LIN \subset CF \subset CS \subset RE \subset ARB.$$

Anomenarem a aquesta seqüència d'inclusions *la jerarquia de Chomsky*.

Pel que fa a la derivació més-a-l'esquerra, se sap que, per a cada gramàtica del tipus-0, G , que conté només regles $u \rightarrow v$, amb $u \in V_N^+$, tenim que $L_{esq.}(G) \in CF$. Mentre que, si G és lliure del context, aleshores $L_{esq.}(G) = L(G)$.

També les gramàtiques pures poden ser classificades, d'acord amb la forma de les seves regles, en arbitràries, de longitud creixent ($|u| \leq |v|$ per a cada regla $u \rightarrow v$), lliures del context (totes les regles de la forma $a \rightarrow v, a \in V$), i λ -lliure lliures del context (si $a \rightarrow v$ és una regla, aleshores $v \neq \lambda$).

Definició 3.3.6 Una important noció relacionada amb les gramàtiques lliures del context és la d'*arbres de derivació*. Donada una gramàtica lliure del context $G = (V_N, V_T, S, P)$, un arbre dirigit τ amb els nodes etiquetats amb elements de $V_N \cup V_T$ o de λ és un arbre de derivació associat a una derivació δ a G si:

1. l'arrel de τ està etiquetada mitjançant S , tots els nodes intermedis estan etiquetats mitjançant símbols de V_N , i els extrems estan etiquetats mitjançant símbols de V_T o de λ ;

2. si un node de τ està etiquetat per un no-terminal A i els seus descendents estan etiquetats, per ordre, mitjançant $\alpha_1, \dots, \alpha_k$, per a $\alpha_i \in V_N \cup V_T$, $1 \leq i \leq k$, $k \geq 1$, aleshores $A \rightarrow \alpha_1 \dots \alpha_k$ és una producció de P ;
3. si un node de τ està etiquetat per un no-terminal A i els seus descendents només estan etiquetats mitjançant λ , aleshores $A \rightarrow \lambda$ és una producció de P .

Per exemple, si $G = (\{S\}, \{a, b\}, S, \{S \rightarrow SaS, S \rightarrow b\})$, aleshores els arbres de la Figura 3.3.1 són arbres de derivació per a les derivacions

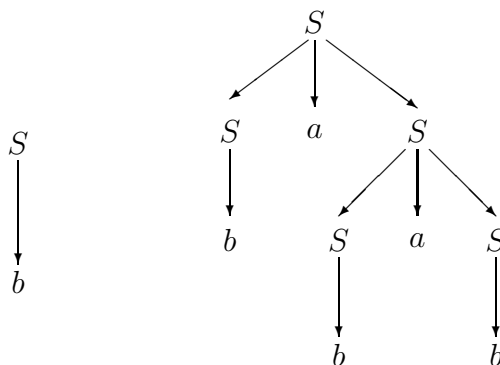
$$\delta_1 : S \Longrightarrow b,$$

$$\delta_2 : S \Longrightarrow SaS \Longrightarrow baS \Longrightarrow baSaS \Longrightarrow babaS \Longrightarrow babab.$$

Normalment, es dibuixen els arbres de derivació amb l'arrel a sobre i els descendents del node etiquetat d'esquerra a dreta amb els símbols a la banda dreta de la producció corresponent.

La *frontera* d'un arbre de derivació és la cadena obtinguda concatenant les etiquetes dels seus extrems, d'esquerra a dreta. En l'exemple anterior, les fronteres dels dos arbres són b , $babab$, respectivament.

Cal observar que el mateix arbre pot descriure diverses derivacions, a saber, totes aquelles obtingudes canviant l'ordre d'ús de les regles.

Figura 3.3.1 Arbres de derivació per a les derivacions δ_1 i δ_2 definides.

Una altra definició de les derivacions d'una gramàtica com a arbres etiquetats, el que hem anomenat arbres de derivació, és la donada per Keenan i Mönnich a (Keenan i Mönnich, 1993) i reformulada per Martín-Vide a (Martín-Vide, 1993).

A continuació donarem algunes definicions bàsiques al respecte, però sense entrar en detalls.

Definició 3.3.7 Un *arbre* (no etiquetat i no ordenat) T és un parell (N, D) on N és un conjunt, els elements del qual s'anomenen nodes o vertexs, i D és una relació binària a N , denominada dominació, que satisfà:

- (i) D és un ordre dèbil; és a dir, es reflexiu, antisimètric i transitiu;
- (ii) condició de l'arrel: existeix algun $r \in N$ tal que, per a tot $b \in N$, aleshores rDb ; és a dir, r és únic i s'anomena l'arrel de T ;

- (iii) condició de ramificació (d'una cadena): per a tot $b \in N$, aleshores el conjunt $\{a \in N \mid aDb\}$ és una cadena. Per tant, si dos nodes $a, a' \in N$ dominen ambdós un altre node $b \in N$, aleshores o bé aDa' o bé $a'Da$.

Definició 3.3.8 Un *arbre etiquetat (no ordenat)* T és una tripla (N, D, L) , que satisfà:

- (i) (N, D) és un arbre (no etiquetat i no ordenat) i
- (ii) L és una funció amb domini N .

Definició 3.3.9 Un *arbre (no etiquetat) terminalment ordenat* T és una tripla $(N, D, <)$, que satisfà:

- (i) (N, D) és un arbre (no etiquetat i no ordenat) i
- (ii) $<$ és un ordre lineal dels nodes terminals de (N, D) ; és a dir, és asimètric, antisimètric, transitiu i connectat.

Definició 3.3.10 Un *arbre etiquetat terminalment ordenat* T és una quàdrupla $(N, D, <, L)$, que satisfà:

- (i) (N, D) és un arbre (no etiquetat i no ordenat),
- (ii) $<$ és un ordre lineal dels nodes terminals de (N, D) i
- (ii) L és una funció amb domini N .

3.4 Gramàtiques regulades

Algunes restriccions en la derivació de gramàtiques lliures del context han estat introduïdes. Són les denominades *gramàtiques amb mecanismes de control*, *gramàtiques amb derivacions controlades*, o *gramàtiques regulades*. La idea que s'amaga darrera de tots els mecanismes de control és la següent: donada una gramàtica lliure del context, la restringim mitjançant el control de l'aplicació de les seves regles, de manera que algunes derivacions que són possibles en el procés usual de derivació de les gramàtiques lliures del context, són evitades. En aquest cas, el conjunt de cadenes generades per les restriccions és un subconjunt del llenguatge lliure del context generat de la manera usual. Aquests subconjunts generats poden ser llenguatges no-lliures del context. Aquests mecanismes són més poderosos que les gramàtiques lliures del context.

Alguns exemples d'aquestes gramàtiques són: les *gramàtiques matricials* i similars a matrius (les produccions lliures del context estan agrupades en “matrius” o “vectors”, i les produccions de cada grup són usades conjuntament); les *gramàtiques programades* i les *gramàtiques controlades* (la restricció és imposada en la seqüència de regles usades); les *gramàtiques de context aleatori* (una regla és usada solament en presència o absència de certs símbols no-terminals); les *gramàtiques condicionals* (una condició global per tal que una cadena pugui ser reescrita ha de ser satisfeta per a poder usar la mencionada producció); les *gramàtiques paral·leles* (totes les ocurrències de símbols no-terminals i terminals poden ser reescrites en un mateix pas de derivació); les

gramàtiques de valències (cada regla té associada un valor i una derivació és acceptada només si la suma o el producte de les valències de les regles usades en aquesta derivació és igual al valor prescrit), etc. Definicions, resultats, i referències bibliogràfiques poden ser trobades en el monogràfic (Dassow i Păun, 1989); resultats recents poden ser trobats a (Dassow et al., 1996a). També a (Martín-Vide, 1994) es pot trobar una introducció al respecte, molt clara i entenedora.

A continuació veurem uns quants exemples de gramàtiques regulades o amb mecanismes de control: amb derivacions regulades mitjançant “seqüències prescrites” amb derivacions regulades mitjançant “condicions del context”.

En el cas de les gramàtiques regulades per seqüències prescrites, considerarem alguns mecanismes que defineixen seqüències de produccions, i prendrem només aquelles cadenes del llenguatge generat que puguin ser obtingudes mitjançant una derivació on les produccions són aplicades d’acord a les seqüències definides. Veurem els exemples de:

- les gramàtiques controlades,
- les gramàtiques matricials i
- les gramàtiques programades.

3.4.1 Gramàtiques controlades

En les gramàtiques controlades la restricció és imposada en la seqüència de regles usades.

Definició 3.4.1 i) Una *gramàtica (lliure del context) controlada regularment amb comprovació de les aparicions* és una 6-tupla

$$G = (V_N, V_T, S, P, R, F),$$

on

- V_N, V_T, P i S són especificades com en una gramàtica de Chomsky lliure del context (el vocabulari no-terminal, el vocabulari terminal, l'axioma inicial, i el conjunt de regles o produccions, respectivament; veure la Definició 3.3.3 de la pàgina 112),
- R és un conjunt regular sobre P i
- F és un subconjunt de P .

ii) Per a una producció o regla $p = A \longrightarrow w \in P$ i $x, y \in V_G^* = (V_N \cup V_T)^*$, definim l'aplicació

$$x \Longrightarrow_p^{ca} y$$

d'una producció p , en el mode de comprovació de les aparicions (ca), mitjançant

$$x = x_1 A x_2, \text{ i } y = x_1 w x_2$$

o

$$x = y, A \text{ no apareix a } x, \text{ i } p \in F.$$

iii) El llenguatge $L(G)$ generat per G (amb comprovació de les aparicions en l'aplicació de les produccions) consisteix en totes les cadenes $w \in V_T^*$ tal que existeix una derivació

$$S \Rightarrow_{p_1}^{ca} w_1 \Rightarrow_{p_2}^{ca} w_2 \Rightarrow_{p_3}^{ca} w_3 \dots \Rightarrow_{p_n}^{ca} w_n = w$$

amb

$$p_1 p_2 p_3 \dots p_n \in R.$$

iv) Direm que G és una *gramàtica controlada regularment sense comprovació de les aparicions* si, i només si, $F = \emptyset$.

Cal notar que els modes, d'amb i sense comprovació de les aparicions, coincideixen si la producció no pertany a F i si el símbol no-terminal de la banda esquerra apareix en la forma sentencial. En aquest cas podem escriure $x \Rightarrow_p y$ en comptes de $x \Rightarrow_p^{ca} y$.

Òbviament, si $R = P^*$, aleshores no hi ha cap restricció respecte a les seqüències de produccions; és a dir, la gramàtica treballa com una gramàtica lliure del context.

Per rC_{ca}^λ denotem la família de gramàtiques regularment controlades amb comprovació de les aparicions i amb produccions o regles lliures del context

arbitràries. Si ens restringim a gramàtiques regularment controlades sense comprovació de les aparicions aleshores ometem el subíndex ca , i si ens restringim a gramàtiques regularment controlades sense sense regles d'esborrament ometem el superíndex λ .

Teorema 3.4.1 Per a una família X de gramàtiques, si denotem per $\mathcal{L}(X)$ la família associada de llenguatges, aleshores es compleix

- i) $\mathcal{L}(CF) \subset \mathcal{L}(rC) \subset \mathcal{L}(rC_{ca}) \subset \mathcal{L}(CS)$,
- ii) $\mathcal{L}(CF) \subset \mathcal{L}(rC) \subset \mathcal{L}(rC^\lambda) \subset \mathcal{L}(rC_{ca}^\lambda) = \mathcal{L}(RE)$,
- iii) La família $\mathcal{L}(rC^\lambda)$ és incomparable amb $\mathcal{L}(CS)$ i amb $\mathcal{L}(rC_{ca})$.

Teorema 3.4.2 Tots els llenguatges de $\mathcal{L}(rC^\lambda)$ sobre un vocabulary unari són regulars.

3.4.2 Gramàtiques matricials

Les produccions o regles lliures del context estan agrupades en “matrius” o “vectors”, i les produccions de cada grup són usades conjuntament.

Definició 3.4.2 i) Una *gramàtica matricial amb comprovació de les aparicions* és una quintupla ordenada

$$G = (V_N, V_T, S, M, F),$$

on

- V_N, V_T i S tenen el mateix significat que en el cas d'una gramàtica de Chomsky (el vocabulari no-terminal, el vocabulari terminal, i l'axioma inicial, respectivament; veure la Definició 3.3.3 de la pàgina 112),
- $M = \{m_1, m_2, \dots, m_n\}$, $n \geq 1$ és un conjunt finit de seqüències finites no-buides, anomenades matrius, $m_i = (p_{i_1}, \dots, p_{i_{k(i)}})$, $k(i) \geq 1$, $1 \leq i \leq n$, on cada p_{i_j} , $1 \leq i \leq n$, $1 \leq j \leq k(i)$, és una producció o regla lliure del context (és a dir, els elements de p_i són parells ordenats (w, u) , amb $w \in (V_N \cup V_T)^* V_N (V_N \cup V_T)^*$ i $u \in (V_N \cup V_T)^*$, i s'escriuen $w \Rightarrow u$).
- F és un subconjunt de totes les produccions que tenen lloc en els elements d' M , i.e. $F \subset \{p_{i_j} \mid 1 \leq i \leq n, 1 \leq j \leq k(i)\}$.

ii) Per a m_i , $1 \leq i \leq n$, i $x, y \in V_G^* = (V_N \cup V_T)^*$ definim $x \Rightarrow_{m_i} y$ mitjançant

$$x = x_0 \Rightarrow_{p_{i_1}}^{ca} x_1 \Rightarrow_{p_{i_2}}^{ca} x_2 \Rightarrow_{p_{i_3}}^{ca} \dots \Rightarrow_{p_{i_{k(i)}}}^{ca} x_{k(i)} = y,$$

on l'aplicació de les produccions en el mode de comprovació de les aparicions està definida a iii).

iii) Per a una producció o regla $p = A \rightarrow w \in P$ i $x, y \in V_G^*$, definim l'aplicació

$$x \Rightarrow_p^{ca} y$$

d'una producció p , en el mode de comprovació de les aparicions (ca), mitjançant

$$x = x_1 A x_2, \text{ i } y = x_1 w x_2$$

o

$$x = y, \quad A \text{ no apareix a } x, \text{ i } p \in F.$$

iv) El llenguatge $L(G)$ generat per G (amb comprovació de les aparicions en l'aplicació de les produccions) es defineix com el conjunt de totes les cadenes $w \in V_T^*$ tal que existeix una derivació

$$S \Longrightarrow_{m_{j_1}} y_1 \Longrightarrow_{m_{j_2}} y_2 \Longrightarrow_{m_{j_3}} \dots \Longrightarrow_{m_{j_s}} w,$$

per a algun $s \geq 1$, $1 \leq j_i \leq n$, $1 \leq i \leq s$.

v) Direm que M és una *gramàtica matricial sense comprovació de les aparicions* si, i només si, $F = \emptyset$.

Òbviament, si cada matriu conté exactament una sola producció o regla, la gramàtica matricial treballa com una gramàtica lliure del context.

Per M_{ca}^λ denotem la família de gramàtiques matricials amb comprovació de les aparicions i amb regles d'esborrament. Ometrem els índexs ca i/o λ , si ens restringim a gramàtiques matricials sense comprovació de les aparicions i/o sense regles d'esborrament, respectivament.

Teorema 3.4.3 Per a una família X de gramàtiques, si denotem per $\mathcal{L}(X)$ la família associada de llenguatges, aleshores es compleix

$$\mathcal{L}(M) = \mathcal{L}(rC),$$

$$\mathcal{L}(M^\lambda) = \mathcal{L}(rC^\lambda),$$

$$\mathcal{L}(M_{ca}) = \mathcal{L}(rC_{ca}),$$

$$\mathcal{L}(M_{ca}^\lambda) = \mathcal{L}(rC_{ca}^\lambda).$$

3.4.3 Gramàtiques programades

En les gramàtiques programades la restricció és imposada en la seqüència de regles usades.

Definició 3.4.3 i) Una *gramàtica programada* és una quàdrupla ordenada

$$G = (V_N, V_T, S, P),$$

on

- V_N, V_T i S tenen el mateix significat que en el cas d'una gramàtica de Chomsky (el vocabulari no-terminal, el vocabulari terminal, i l'axioma inicial, respectivament; veure la Definició 3.3.3 de la pàgina 112) i
- P és un conjunt finit de triples $r = (p, \sigma, \phi)$, on p és una producció o regla lliure del context (és a dir, els elements de p són parells ordenats (w, u) , amb $w \in (V_N \cup V_T)^* V_N (V_N \cup V_T)^*$ i $u \in (V_N \cup V_T)^*$, i s'escriuen $w \Rightarrow u$), i σ i ϕ són subconjunts de P .

ii) El llenguatge $L(G)$ generat per G es defineix com el conjunt de cadenes $w \in V_T^*$ tal que existeix una derivació

$$S = w_0 \Rightarrow_{r_1} w_1 \Rightarrow_{r_2} w_2 \Rightarrow_{r_3} \dots \Rightarrow_k w_k = w,$$

$k \geq 1$, i existeix $r_{k+1} \in P$ tal que, per a $r_i = (A_i \longrightarrow v_i, \sigma, \phi)$, $1 \leq i \leq k$, es compleix una de les següents condicions:

$$w_{i-1} = w'_{i-1} A_i w''_{i-1}, \quad w_i = w'_{i-1} v_i w''_{i-1}, \quad \text{per a algun } w'_{i-1}, w''_{i-1} \in V_G^* \text{ i } r_{i+1} \in \sigma_i.$$

o

$$A_i \text{ no apareix a } w_{i-1}, \quad w_{i-1} = w_i \text{ i } r_{i+1} \in \phi_i.$$

iii) Si $r = (p, \sigma, \phi)$ es compleix per a cada $r \in P$, aleshores direm que G és una *gramàtica programada sense comprovació de les aparicions*. En qualsevol altre cas, G és una *gramàtica programada amb comprovació de les aparicions*.

Si $r = (p, \sigma, \phi)$, aleshores σ i ϕ s'anomenen *camp d'encert* i *camp de fallida* d' r , respectivament.

Clarament, si $\sigma_r = P$ i $\phi_r = \emptyset$, per a tot $r \in P$, aleshores la gramàtica programada treballa com una gramàtica lliure del context.

Per P_{ca}^λ denotem la família de gramàtiques programades amb comprovació de les aparicions i amb regles d'esborrament. Ometrem els índexs ca i/o λ , si ens restringim a gramàtiques programades sense comprovació de les aparicions i/o sense regles d'esborrament, respectivament.

Teorema 3.4.4 Per a una família X de gramàtiques, si denotem per $\mathcal{L}(X)$ la família associada de llenguatges, aleshores es compleix

$$\mathcal{L}(P) = \mathcal{L}(rC),$$

$$\mathcal{L}(P^\lambda) = \mathcal{L}(rC^\lambda),$$

$$\mathcal{L}(P_{ca}) = \mathcal{L}(rC_{ca}),$$

$$\mathcal{L}(P_{ca}^\lambda) = \mathcal{L}(rC_{ca}^\lambda).$$

Pel que fa a les gramàtiques regulades per condicions del context considerarem alguns mecanismes on el control funciona de la següent manera: una producció només pot ser aplicada a una forma sentencial si la forma satisfà certes condicions. Per exemple, mitjançant aquesta condició podem requerir que certs símbols hagin d'aparèixer en la forma sentencial.

Mitjançant aquest control la seqüència de produccions aplicables no és descrita “a priori”; és controlada per la seqüència generada de formes sentencials. Veurem només l'exemple de les gramàtiques condicionals.

3.4.4 Gramàtiques condicionals

En les gramàtiques condicionals, una condició global per tal que una cadena pugui ser reescrita ha de ser satisfeta per a poder usar la mencionada producció.

Definició 3.4.4 i) Una *gramàtica condicional* és una quàdrupla ordenada

$$G = (V_N, V_T, S, P),$$

on

- V_N, V_T i S tenen el mateix significat que en el cas d'una gramàtica de Chomsky (el vocabulari no-terminal, el vocabulari terminal, i l'axioma inicial, respectivament; veure la Definició 3.3.3 de la pàgina 112) i
 - P és un conjunt finit de parells $R = (p, R)$, on p és una producció o regla lliure del context (és a dir, els elements de p són parells ordenats (w, u) , amb $w \in (V_N \cup V_T)^* V_N (V_N \cup V_T)^*$ i $u \in (V_N \cup V_T)^*$, i s'escriuen $w \Rightarrow u$) i R és un conjunt regular sobre $V_G = (V_N \cup V_T)$.
- ii) Per a $x, y \in V_G^*$, direm que x deriva directament a y , i ho denotarem per $x \Rightarrow y$, si existeix un parell $r = (A \rightarrow w, R) \in P$ tal que $x = x'Ax''$ i $y = x'wx''$ per a algun $x', x'' \in V_G^*$ i $x \in R$ (també usarem la notació $x \Rightarrow_r y$ o $x \Rightarrow_p y$ si volem especificar el parell o les produccions que es aplicat).

iii) El llenguatge $L(G)$ generat per G es defineix com

$$L(G) = \{w \in V_T^* \mid S \Rightarrow^* w\}.$$

Per definició, podem aplicar el parell $r = (p, R)$ (i, per tant, la producció p) a x , si x està contingut en el conjunt regular associat amb p .

Si el conjunt regular associat amb cada producció és el conjunt de totes les cadenes sobre V_G , aleshores la condició per a la seva aplicació és sempre satisfeta. Per tant, en aquest cas, la gramàtica condicional treballa com a una gramàtica lliure del context.

Per C^λ i C denotem la família de totes les gramàtiques condicionals amb i sense regles d'esborrament, respectivament.

Teorema 3.4.5 Per a una família X de gramàtiques, si denotem per $\mathcal{L}(X)$ la família associada de llenguatges, aleshores es compleix

$$\mathcal{L}(C) = \mathcal{L}(CS),$$

$$\mathcal{L}(C^\lambda) = \mathcal{L}(RE).$$

3.5 Sistemes de Lindenmayer

Els *sistemes de Lindenmayer* o *sistemes-L* són construccions matemàtiques molt pròximes a les gramàtiques. Els seus orígens cal buscar-los en els treballs d'A. Lindenmayer (Lindenmayer, 1968; Lindenmayer, 1971; Lindenmayer, 1975), del qual reben el seu nom.

L'objectiu original d'aquesta teoria era el d'oferir models matemàtics del desenvolupament dels organismes filamentosos simples. En un principi, els sistemes-L foren definits com a sèries lineals d'autòmats finits; més tard, no obstant, foren reformulats en termes d'una estructura més apropiada de construccions del tipus de les gramàtiques. De llavors ençà, la teoria dels sistemes-L s'ha anat desenvolupant com una branca de la teoria de llenguatges formals.

El tret fonamental dels sistemes-L és que la reescriptura d'una cadena

es fa de manera paral·lela, a diferència de la reescriptura seqüencial de les gramàtiques. Això significa que en cada pas del procés de reescriptura d'un sistema-L s'han de reescriure totes les lletres, mentre que en el cas d'una gramàtica només se substitueix una part de la cadena.

Aquest paral·lisme en la reescriptura posa de manifest la motivació biològica bàsica que s'amaga darrera dels sistemes-L. S'està intentant modelitzar el desenvolupament d'un organisme, que s'esdevé d'una manera paral·lela, simultàniament en tot el seu conjunt. La reescriptura seqüencial no és adequada per a aquests menesters.

La versió més simple dels sistemes-L assumeix que el desenvolupament d'una cèl·lula és independent de la influència de les altres cèl·lules. Habitualment, aquest tipus de sistema-L s'anomena 0-sistema-L (0 fa referència al fet que la comunicació entre les cèl·lules té lloc per 0-bandes, o que hi ha 0-iteracions entre les cèl·lules).

Aleshores, bàsicament, un 0-sistema-L és una gramàtica pura lliure del context amb derivacions paral·leles: $G = (V, w, P)$, on V és un vocabulari, $w \in V^*$ (axioma), i P és un conjunt finit de regles de reescriptura de la forma $a \rightarrow v$ amb $a \in V, v \in V^*$. S'assumeix també que, per a cada $a \in V$, existeix com a mínim una regla $a \rightarrow v$ a P (direm que P és *complet*). Per a $w_1, w_2 \in V^*$ escriurem $w_1 \Longrightarrow w_2$ si $w = a_1 \dots a_n, w_2 = v_1 \dots v_n$, per a $a_i \rightarrow v_i \in P, 1 \leq i \leq n$. Aleshores, el llenguatge generat per G consta de totes les cadenes que es poden derivar a partir de w usant les regles de P d'una

manera paral·lela; és a dir,

$$L(G) = \{x \in V^* \mid w \Longrightarrow^* x\}.$$

Si per a cada regla $a \rightarrow v \in P$ tenim que $x \neq \lambda$, aleshores direm que G és *propagador* (és no-esborrador); si per a cada $a \in V$ existeix una sola regla $a \rightarrow v$ a P , aleshores es diu que G és *determinista*. Si distingim un subconjunt T de V i definim $L(G)$ com a $L(G) = \{x \in T^* \mid w \Longrightarrow^* x\}$, aleshores direm que G és *extesa*. La família de llenguatges generada per 0-sistemes-L es denota per $0L$; afegim les lletres P, D, E davant de $0L$ si s'usen sistemes propagadors, deterministes, o extesos, respectivament.

Un 0-sistema-L *taulat* (o amb taules), abreujadament T0L, és un sistema $G = (V, w, P_1, \dots, P_n)$, tal que cada tripla $(V, w, P_i), 1 \leq i \leq n$, és un 0-sistema-L (també anomenat una *taula*). El llenguatge generat es defineix mitjançant

$$L(G) = \{x \in V^* \mid w \Longrightarrow_{P_{j_1}} w_1 \Longrightarrow_{P_{j_2}} w_2 \Longrightarrow_{P_{j_3}} \dots \Longrightarrow_{P_{j_m}} w_m = x, \\ m \geq 0, 1 \leq j_i \leq n, 1 \leq i \leq m\}.$$

(Cada pas de la derivació es realitza mitjançant regles de la mateixa taula.)

Un sistema T0L és determinista quan cadascuna de les seves taules és determinista. Els trets de propagació i d'extensió es defineixen de manera natural.

La família de llenguatges generada per T0L es denota per $T0L$; les famílies

$ET0L$, $EDT0L$, etc. s'obtenen de la mateixa manera que $E0L$, $ED0L$, etc.

La família $D0L$ és incomparable amb FIN , REG , CF , CF , mentre que $E0L$ inclou estrictament la família CF ; $ET0L$ és la família més gran de llenguatges de Lindenmayer i està estrictament inclosa a CS . Resultats posteriors referents a aquestes famílies poden ser trobats a (Rozenberg i Salomaa, 1980; Rozenberg i Salomaa, 1996a).

3.6 Autòmats i transductors

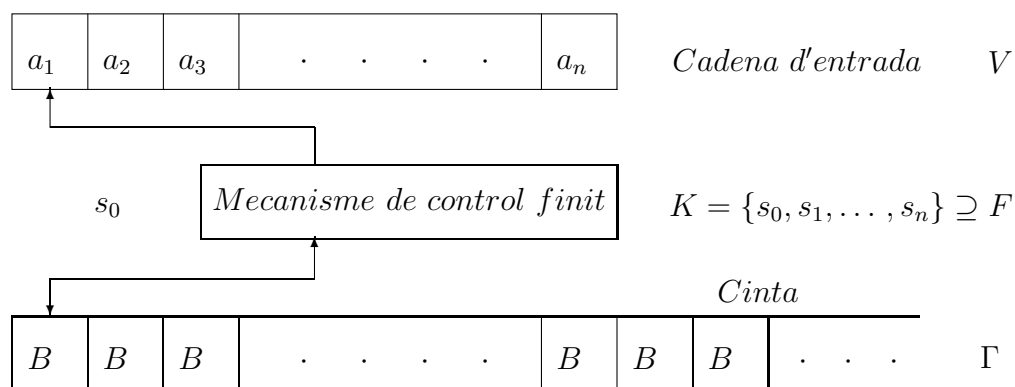
Les cinc famílies bàsiques de llenguatges en la jerarquia de Chomsky, REG , LIN , CF , CS , RE , poden ser també caracteritzades mitjançant autòmats reconeixadors. Aquests autòmats són: l'autòmat finit (“finite automaton”), que el denotarem per FA; l'autòmat de pila d'un sol torn o direcció (“one-turn push-down automaton”), que el denotarem per OPDA; l'autòmat de pila (“push-down automaton”), que el denotarem per PDA; l'autòmat de memòria linealment limitada (“linear-bounded automaton”), que el denotarem per LBA; i, la màquina de Turing (“Turing machine”) que la denotarem per TM, respectivament.

Definició 3.6.1 Intuïtivament, una *màquina de Turing* té una cinta dividida en cel·les, les quals poden emmagatzemar símbols del vocabulari de la cinta Γ (cada cel·la només pot emmagatzemar un únic símbol de Γ). La cinta està limitada, per l'esquerra (existeix la cel·la que és la de més a l'esquerra) i, en

canvi, no està limitada per la dreta. La màquina té un mecanisme de control finit i un capçal de lectura/escriptura, que pot llegir o escriure en una cel·la de la cinta en un determinat moment. La cadena d'entrada és una cadena sobre V i se situa, o emmagatzema, en la cinta, començant per la cel·la més a l'esquerra possible, i amb totes les altres cel·les contenint el símbol blanc B .

Inicialment, el capçal es troba en la cel·la més a l'esquerra possible i el mecanisme de control finit en l'estat s_0 . La màquina pot realitzar moviments. Un moviment depèn de la cel·la explorada, "escanejada", en aquell precís moment i de l'estat actiu en el mecanisme de control. Un moviment consisteix en: canviar d'estat, escriure un símbol de V en la cel·la activa, i moure el capçal una cel·la a l'esquerra (E) o a la dreta (D). Direm que una cadena d'entrada és *acceptada* si, i només si, després d'un nombre finit de moviments, la màquina de Turing entra en un estat final (d' F).

Figura 3.6.1 Descripció d'una màquina de Turing.



Formalment, una *màquina de Turing* (TM) és un sistema ordenat

$$M = (K, V, \Gamma, \delta, s_0, B, F)$$

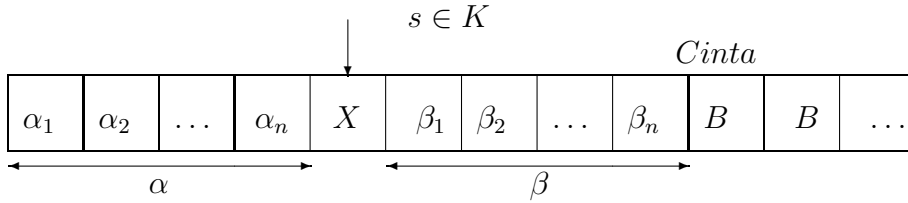
on:

- K és un conjunt finit d'*estats*,
- V és un vocabulari, denominat *vocabulari d'entrada*,
- Γ és un vocabulari, denominat *vocabulari de la cinta*, $V \subset \Gamma$,
- $s_0 \in K$ és l'*estat inicial*,
- $B \in \Gamma - V$ és el símbol *blanc*,
- $F \subseteq K$ és el conjunt d'*estats finals* i
- $\delta : K \times \Gamma \Longrightarrow \mathcal{P}(K \times \Gamma \times \{D, E\})$ és la *funció de transició*.

El conjunt de *descripcions instantànies* o *configuracions*, denotat per C_M , d'una màquina de Turing M , és una seqüència o cadena $\alpha s X \beta \in \Gamma^* K \Gamma^*$, on $\alpha, \beta \in \Gamma^*$, $s \in K$, $X \in \Gamma$. La cadena té codificada la descripció de la màquina de Turing en un moment determinat de la següent manera: s és l'estat actiu del mecanisme de control finit, X és el contingut de la cel·la activa de la cinta, α és el contingut de la cinta a la banda esquerra de la cel·la activa, mentre que β és el contingut de la cinta a la banda dreta de la cel·la activa fins l'última cel·la de la banda dreta que és no-blanca.

Per exemple,

Figura 3.6.2 La descripció d'una màquina de Turing en un moment determinat



correspon a la configuració $\underbrace{\alpha_1 \alpha_2 \dots \alpha_n}_\alpha s X \underbrace{\beta_1 \beta_2 \dots \beta_n}_\beta \in C_M$.

Una màquina de Turing M defineix una relació de *transició directa* entre configuracions, denotada per $\vdash_M$, i.e. $\vdash_M \subset C_M \times C_M$, definida per

$$\alpha s' \beta \vdash_M \alpha Y s \beta \text{ si, i només si, } (s, Y, D) \in \delta(s', X)$$

$$\alpha s' \vdash_M \alpha Y s \text{ si, i només si, } (s, Y, D) \in \delta(s', B)$$

$$\alpha Z s' \beta \vdash_M \alpha s Z Y \beta \text{ si, i només si, } (s, Y, E) \in \delta(s', X)$$

$$\alpha Z s' \vdash_M \alpha s Z Y \text{ si, i només si, } (s, Y, E) \in \delta(s', B)$$

on, $\alpha, \beta \in \Gamma^*$, $X, Y, Z \in \Gamma$, $s, s' \in K$, i E significa *esquerra* i D significa *dreta*.

La *relació de transició*, denotada per $\vdash_M^*$, és la clausura reflexiva i transitiva de $\vdash_M$. Per exemple, si $c, c' \in C_M$ aleshores $c \vdash_M^* c'$ si, i només si, existeixen

$c_1, \dots, c_n \in C_M$ tal que

$$c = c_1 \vdash_M c_2 \vdash_M \dots \vdash_M c_n = c'$$

El llenguatge acceptat per M , denotat per $L(M)$, és:

$$L(M) = \{w \in V^* \mid s_0 w \vdash_M^* \alpha s \beta, \text{ per algun } s \in F, \alpha, \beta \in \Gamma^*\}.$$

Cal observar que $L(M) \subseteq V^*$.

Si per a a cada $(s, X) \in K \times \Gamma$, $\text{card}(\delta(s, X)) \leq 1$, aleshores M és una màquina de Turing *determinista*. En cas contrari, M és una màquina de Turing *no-determinista*.

Observació 3.6.1 Un llenguatge L és acceptat per una màquina de Turing no-determinista si, i només si, L és acceptat per una màquina de Turing determinista.

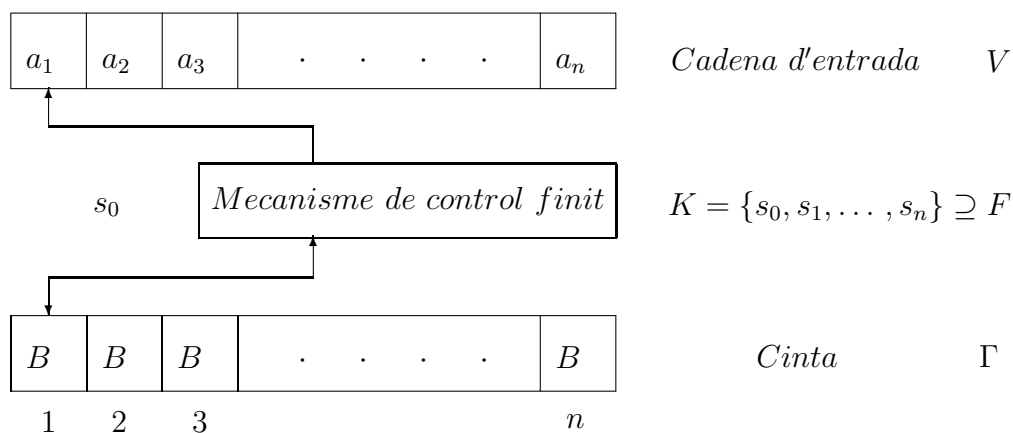
Teorema 3.6.1 Sigui L_0 un llenguatge. Aleshores, existeix una gramàtica de Chomsky G de tipus-0 tal que $L(G) = L_0$ si, i només si, existeix una màquina de Turing M tal que $L(M) = L_0$.

Veurem ara algunes restriccions de la màquina de Turing.

Definició 3.6.2 Un *autòmat de memòria linealment limitada*, denotat per LBA , és una màquina de Turing M amb la següent restricció: per a cada cadena d'entrada $w \in V^*$ amb $|w| = n$, M usa com a màxim n cel·les de la cinta.

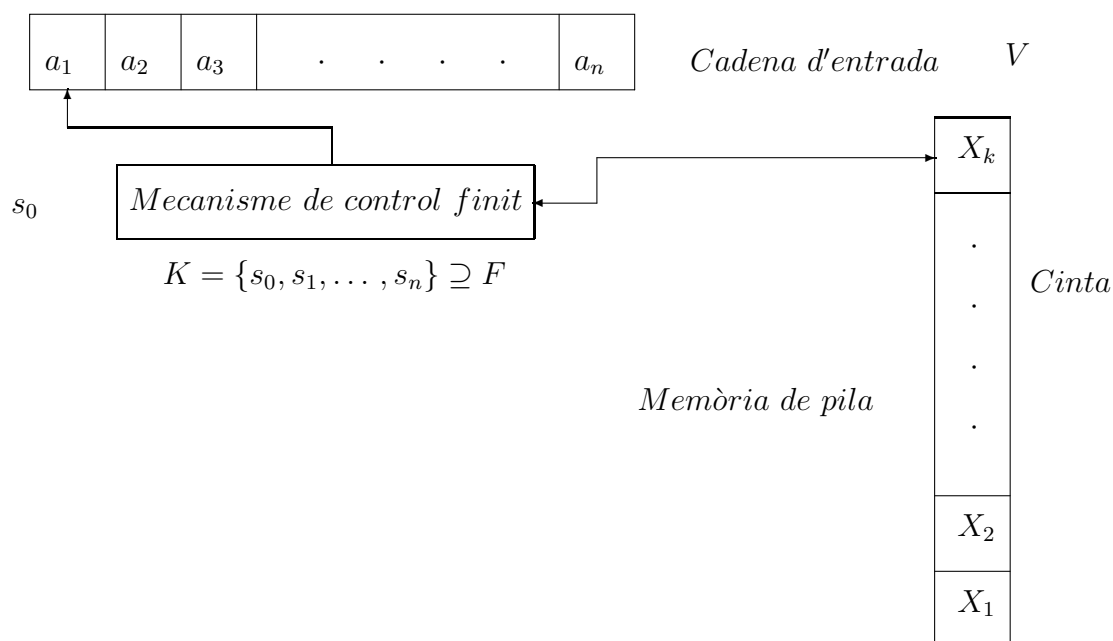
Això és

Figura 3.6.3 Autòmat de memòria linealment limitada.



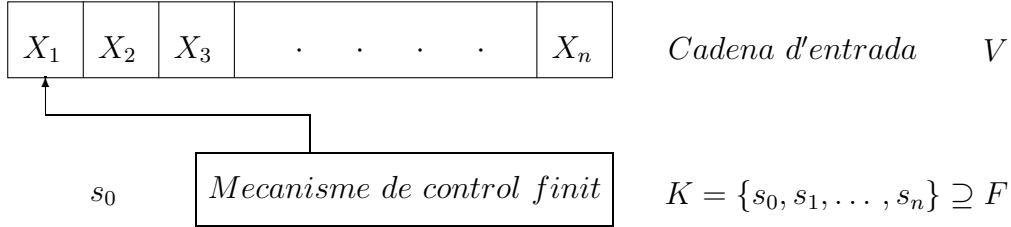
Teorema 3.6.2 Sigui L_0 un llenguatge. Aleshores, existeix una gramàtica de Chomsky G de tipus-1 (sensibile al context o CS) tal que $L(G) = L_0$ si, i només si, existeix un autòmat de memòria linealment limitada (LBA), M , tal que $L(M) = L_0$.

Definició 3.6.3 Un *autòmat de pila*, denotat per PDA , és una màquina de Turing M que té com a cinta una memòria denominada “pila”. Aquesta pila és un tipus d’emmagatzement en el que només es poden llegir les informacions en l’ordre invers en que han estat escrites, seguint el principi de “l’últim d’entrar serà el primer de sortir”. D’aquí ve que també se la denomini “memòria LIFO” (*last-in, firts-out*). Les dues operacions possibles amb una pila són la d’introduir (escriure) o la d’extreure (llegir).

Figura 3.6.4 Autòmat de pila.

Teorema 3.6.3 Sigui L_0 un llenguatge. Aleshores, existeix una gramàtica de Chomsky G de tipus-2 (lliure del context o CF) tal que $L(G) = L_0$ si, i només si, existeix un autòmat de pila (PDA), M , tal que $L(M) = L_0$.

Definició 3.6.4 Intuïtivament, un *autòmat finit* o *autòmat d'estats finits*, denotat per FA o bé FSA , és una màquina de Turing sense cinta.

Figura 3.6.5 Autòmat finit.

Per tant, formalment, un autòmat finit no-determinista, denotat per *NFA*, serà una construcció

$$M = (K, V, \underbrace{\Gamma}_{no}, \delta, s_0, \underbrace{B}_{no}, F),$$

on K, V, s_0, F estan definits de la mateixa manera que en el cas de la màquina de Turing (veure la Definició 3.6.1 de la pàgina 133); i δ , la funció de transició, està definida per $\delta : K \times V \longrightarrow 2^K$.

Si $\text{card}(\delta(s, a)) \leq 1$ per a tot $s \in K, a \in V$, aleshores direm que l'autòmat és *determinista* (és a dir, si δ està definida per $\delta : K \times V \longrightarrow K$), i el denotarem per *DFA*.

Podem estendre δ a una nova funció δ' definida per $\delta' : K \times V^* \longrightarrow K$ on:

- i) $\delta'(s, \lambda) = s, \forall s \in K,$
- ii) $\delta'(s, aw) = \delta'(\delta(s, a), w), \forall a \in V, w \in V^*, s \in K.$

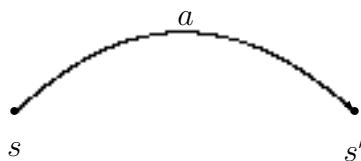
Observem que $\delta'(s, a) = \delta(s, a), \forall a \in V, s \in K$. Per això algunes vegades δ' es denota també com a δ .

Aleshores, el llenguatge de les cadenes acceptades per l'autòmat M és

$$L(M) = \{w \in V^* \mid \delta(s_0, w) \in F\}.$$

Observació 3.6.2 Un *graf de transició* és un graf orientat amb els vèrtexs o nodes etiquetats mitjançant estats, i, la línia entre ells, anomenada arc del graf, etiquetada mitjançant símbols d'entrada.

Figura 3.6.6 Aleshores, si $\delta(s, a) = s'$, podem representar-ho mitjançant el graf



on $s, s' \in K$ i $a \in V$.

Al conjunt $K \times V^*$ podem definir la relació $\vdash$ de la següent manera : per a $s, s' \in K, a \in V, w \in V^*$, escriurem $(s, aw) \vdash (s', w)$ si $s' \in \delta(s, a)$. Per definició, considerarem que $(s, \lambda) \vdash (s, \lambda)$. Si $\vdash^*$ és la clausura reflexiva i transitiva de la relació $\vdash$, aleshores, el llenguatge de les cadenes acceptades per l'autòmat M és

$$L(M) = \{w \in V^* \mid (s_0, w) \vdash_M^* (s, \lambda), s \in F\}.$$

Observació 3.6.3 Un llenguatge L és acceptat per un autòmat finit no determinista (NFA) si, i només si, L és acceptat per un autòmat finit determinista (DFA).

Teorema 3.6.4 Sigui L_0 un llenguatge. Aleshores, existeix una gramàtica de Chomsky G de tipus-3 (regular o REG) tal que $L(G) = L_0$ si, i només si, existeix un autòmat finit (FA), M , tal que $L(M) = L_0$.

Podem resumir els resultats obtinguts fins ara en la següent taula:

Taula 3.6.1 Relacions entre llenguatges, gramàtiques i autòmats en la jerarquia de Chomsky.

Famílies de lleng.	Gram. de Chomsky	Autòmats	Determ. vs. no-determ.
$\mathcal{L}_0(RE)$	tipus-0	TM	determ. $\equiv$ no-determ.
$\mathcal{L}_1(CS)$	tipus-1	LBA	determ. $? \subset$ no-determ.
$\mathcal{L}_2(CF)$	tipus-2	PDA	determ. $\subset$ no-determ.
$\mathcal{L}_3(REG)$	tipus-3	FA	determ. $\equiv$ no-determ.

Definició 3.6.5 Els autòmats que hem considerat fins ara no ofereixen altres resultats que el fet d'estar o no en un estat final; és a dir, només són capaços d'acceptar o rebutjar dades. Podem considerar ara determinats mecanismes, anomenats *transductors*, que poden oferir cadenes com a resultats; és a dir, són capaços de traduir cadenes a cadenes. Ens fixarem només en el transductor

que es correspon a un autòmat finit. El que es correspon a un autòmat de pila es defineix de manera anàloga.

Un sistema de reescriptura (V, P) és un *transductor seqüencial*, que denotarem per ST (de “sequential transductor”) si, i només si, compleix les tres condicions següents:

- i) V està dividit en dos vocabularis disjunts K i $(V_I \cup V_O)$. Els conjunts K , V_I i V_O reben, respectivament, els noms de vocabulari d'estats, vocabulari de dades (o vocabulari d'entrada o “input”) i vocabulari de resultats (o vocabulari de sortida o “output”). Aquests dos últims són no-buits però no necessàriament disjunts.
- ii) S'especifiquen $s_o \in K$ i $F \subseteq K$, l'estat inicial i el conjunt d'estats finals, respectivament.
- iii) Les produccions de P tenen la forma: $s_i w \vdash u s_j$, amb $s_i, s_j \in K$, $w \in V_I^*$ i $u \in V_O$.

Si, a més a més, $w \neq \lambda$ en totes les produccions, aleshores el sistema de reescriptura rep el nom de *màquina seqüencial generalitzada*, que denotarem per “*gsm*” (de “generalized sequential machine”).

Per tant, un transductor seqüencial és un sistema

$$g = (K, V_I, V_O, s_o, F, \delta),$$

on K, s_0, F es defineixen d'igual manera que en els autòmats finits, V_I, V_O són vocabularis (el vocabulari d'entrada o “input” i el de sortida o “output”, respectivament), i $\delta : K \times V_I \longrightarrow 2^{V_O^* \times K}$. Si $\delta(s, a) \subseteq V_O^+ \times K$ per a tot $s \in K, a \in V_I$, aleshores g s'anomena λ -lliure. Per a $s, s' \in K, a \in V_I, y \in V_I^*, x, z \in V_O^*$, escriurem $(x, s, ay) \vdash (xz, s'y)$ si $(z, s') \in \delta(s, a)$. Aleshores, per a $w \in V_I^*$, definim

$$g(w) = \{z \in V_O^* \mid (\lambda, s_0, w) \vdash^* (z, s, \lambda), s \in F\}.$$

La funció g pot ser extesa, de manera natural, a llenguatges sobre V_I .

S'observa que una gsm pot ser vista com un autòmat finit amb sortides. També és fàcil de veure que si una família de llenguatges és tancada sota funcions gsm , aleshores també es tancada sota substitucions finites (i, per tant, també sota morfismes), així com sota les operacions de *Sub*, *Pref*, *Suf*.

Algunes vegades és molt útil presentar la funció de transició dels autòmats finits i de les gsm com un conjunt de regles de reescriptura. Aleshores, escriurem $sa \rightarrow as'$ en comptes de $s' \in \delta(s, a)$ en el cas d'autòmats finits, i, $sa \rightarrow zs'$ en comptes de $(z, s') \in \delta(s, a)$ en el cas de gsm . En aquest cas, les relacions $\vdash, \vdash^*$ són exactament $\implies, \implies^*$ en el sistema de reescriptura obtingut. Per a una gsm g i un llenguatge $L \in V_I^*$, tenim que

$$g(L) = \{z \in V_O^* \mid s_0 w \implies^* zs, w \in L, s \in F\}.$$

Per a un autòmat finit tenim el cas particular:

$$L(M) = \{x \in V^* \mid s_0 x \implies^* xs, s \in F\}.$$

Definició 3.6.6 Un marc de treball per tal d'investigar operacions amb llenguatges és la teoria de *famílies abstractes de llenguatges*, que denotarem per AFL (d'“abstract family of languages”). Una família de llenguatges s'anomena *no-trivial* si conté com a mínim un llenguatge no-buit diferent de $\{\lambda\}$. Una família no-trivial s'anomena un *trio* si és tancada sota λ -lliure morfismes, morfismes inversos, i intersecció amb llenguatges regulars. Un trio tancat sota unió s'anomena una *semi-AFL*. Una semi-AFL que sigui tancada sota concatenació i Kleene + s'anomena una *AFL*. Un trio (semi-AFL, o AFL) que sigui tancat sota morfismes arbitraris s'anomena *complet*. Una família de llenguatges que no sigui tancada sota cap de les sis operacions d'AFL (és a dir, unió, concatenació, Kleene +, morfismes λ -lliure, intersecció amb llenguatges regulars, i morfismes inversos) s'anomena una *anti-AFL*.

Les sis operacions AFL no són independents les unes de les altres. Per exemple, tenim que:

Teorema 3.6.5 Tota semi-AFL (completa) tancada sota Kleene + és una AFL (completa).

Teorema 3.6.6 Si una família F de llenguatges conté només llenguatges λ -lliure i és tancada sota concatenació, morfismes λ -lliure, i morfismes inversos, aleshores també és tancada sota unió i intersecció amb llenguatges regulars; per tant, F és una semi-AFL. (Si F és també tancada sota Kleene +, aleshores és una AFL.)

Corol·lari 3.6.1 Si una família de llenguatges és tancada sota concatenació, morfismes λ -lliure, morfismes inversos, Kleene $+$, diferència i unió amb $\{\lambda\}$, aleshores és una AFL.

Teorema 3.6.7 Si una família de llenguatges és tancada sota intersecció amb conjunts regulars, unió amb conjunts regulars, substitució amb llenguatges regulars λ -lliure, i morfismes restringits, aleshores també és tancada sota morfismes inversos.

Usant aquests resultats es pot provar que una família de llenguatges donada és una AFL sense comprovar la clausura sota totes les sis operacions. Per altra banda, coneixent que una família és una AFL podem inferir altres propietats de clausura de la mateixa:

Teorema 3.6.8 Tota semi-AFL és tancada sota substitució amb llenguatges regulars λ -lliure i funcions *gsm* λ -lliure. Tota semi-AFL completa és tancada sota substitucions amb llenguatges regulars, funcions *gsm* arbitràries, i sota quocients a l'esquerra i a la dreta amb llenguatges regulars.

Les propietats de clausura de les famílies de la jerarquia de Chomsky es mostren en la Taula 3.6.2 de la pàgina 148 (S indica la clausura de la família situada per sobre de la columna corresponent sota l'operació situada a l'esquerra de la corresponent fila). S'observa que *RE*, *CF*, *REG* són AFL completes, *CS* és una AFL, i *LIN* és una semi-AFL completa (no és tancada sota concatenació ni sota clausura de Kleene). De fet, *REG* és el trio complet

més petit.

La família dels llenguatges regulars sobre un vocabulari V és la mateixa que la família dels llenguatges obtinguts a partir dels llenguatges atòmics $\{\lambda\}$ i $\{a\}$, amb $a \in V$, mitjançant un nombre finit d'aplicacions de les operacions regulars: unió, concatenació i clausura concatenativa. La fórmula que expressa com obtenir un determinat llenguatge regular a partir de llenguatges atòmics mitjançant operacions regulars s'anomena *expressió regular*. En aquesta tesi sovint usarem aquesta presentació de llenguatges regulars mitjançant expressions regulars.

Donat un vocabulari V , el conjunt d'expressions regulars sobre V es defineix mitjançant les dues regles següents:

1. $\emptyset, \lambda$, i tots els elements de V són expressions regulars;
2. si u i v són expressions regulars, aleshores $(u) \cup (v)$, $(u)(v)$, i $(u)^+$ són expressions regulars.

Si tenim una expressió regular e , podem associar-li un llenguatge $L(e)$, d'acord amb les regles següents:

1. $L(\emptyset) = \emptyset$,
 $L(\lambda) = \{\lambda\}$,
 $L(a) = \{a\}$, $a \in V$;

$$2. L((u) \cup (v)) = L(u) \cup L(v),$$

$$L((u)(v)) = L(u)L(v),$$

$$L((u)^+) = (L(u))^+,$$

per a totes les expressions regulars u, v .

Pel fet d'assumir que l'operació de $+$ té prioritat sobre l'operació de concatenació, i la concatenació té prioritat sobre la unió, aleshores, quan escrivim una expressió regular, només especificarem els parèntesis que siguin necessaris. Per exemple, a $(a)(b)$ els parèntesis són innecessaris i, per tant, escriurem ab ; a $(a \cup b)(c)$ el primer parell és necessari, mentre que el segon parell no ho és, i, per tant, escriurem $(a \cup b)c$.

Tenim que:

Teorema 3.6.9 (Teorema de Kleene) La família de llenguatges definida mitjançant expressions regulars és igual a REG.

En general, identificarem una expressió regular amb el llenguatge definit per ella. Per exemple, direm que $a^+b \cup a^+ba^+$ és un llenguatge, i no pas una expressió que identifica el llenguatge $L(a^+b \cup a^+ba^+)$.

Taula 3.6.2 La taula mostra les propietats de clausura de (les gramàtiques associades a) les cinc famílies de llenguatges de la jerarquia de Chomsky: N indica que la família de llenguatges és no-tancada sota l'operació esmentada i S indica que la família de llenguatges és tancada sota l'operació esmentada.

	<i>RE</i>	<i>CS</i>	<i>CF</i>	<i>LIN</i>	<i>REG</i>
Unió	S	S	S	S	S
Intersecció	S	S	N	N	S
Complementarietat	N	S	N	N	S
Concatenació	S	S	S	N	S
Kleene +	S	S	S	N	S
Intersecció amb llenguatges regulars	S	S	S	S	S
Substitució	S	N	S	N	S
Substitució λ -lliure	S	S	S	N	S
Morfismes	S	N	S	S	S
Morfismes λ -lliure	S	S	S	S	S
Morfismes inversos	S	S	S	S	S
Funcions <i>gsm</i>	S	N	S	S	S
Funcions <i>gsm</i> λ -lliure	S	S	S	S	S
Quocient per l'esquerra/dreta	S	N	N	N	S
Quocient per l'esquerra/dreta amb llenguatges regulars	S	N	S	S	S
Derivada a l'esquerra/dreta	S	S	S	S	S
Barreja	S	S	N	N	S
Imatge mirall	S	S	S	S	S

3.7 Problemes de decidibilitat

Donada una classe $\mathcal{G}$ de gramàtiques, podem formular-nos les següents preguntes bàsiques relatives a elements arbitraris G_1, G_2 de $\mathcal{G}$:

- *Equivalència*: són G_1 i G_2 equivalents ? (És a dir, $L(G_1) = L(G_2)$?)
- *Inclusió*: està $L(G_1)$ inclosa a $L(G_2)$? (És a dir, $L(G_1) \subseteq L(G_2)$?)
- *Pertanyença*: és una cadena arbitrària x un element de $L(G_1)$? (És a dir, $x \in L(G_1)$?)
- *Buidor*: és $L(G_1)$ buit ? (És a dir, $L(G_1) = \emptyset$?)
- *Finitud*: és $L(G_1)$ finit ? (És a dir, $L(G_1)$ és finit ?)
- *Regularitat*: és $L(G_1)$ un llenguatge regular ? (És a dir, $L(G_1) \in REG$?)

Els llenguatges amb la pregunta de pertanyença decidible s'anomenen *recursius* i la seva família està emplaçada estrictament entre *CS* i *RE*.

Ens fixarem en el fet que un problema de decidibilitat com els anteriors es refereix a gramàtiques arbitràries d'una classe donada $\mathcal{G}$; per a determinades gramàtiques de $\mathcal{G}$ podria ser possible resoldre un problema, mitjançant un algorisme *ad-hoc*, mentre que aquest problema és indecidible en general. En poques paraules, per a cada cas del problema necessitem un algorisme específic. El problema de pertanyença formulat anteriorment s'anomena algunes

vegades *uniforme*: x i G_1 són, ambdues, entrades arbitràries del problema. Si G_1 està fixada i només x és arbitrària, aleshores obtenim el problema de pertanyença *no-uniforme*. Nosaltres treballarem amb la variant uniforme d'aquest problema.

Taula 3.7.1 Aquesta taula mostra les propietats de decidibilitat de (les gràmiques associades a) les cinc famílies de llenguatges de la jerarquia de Chomsky: I indica un problema indecidible i D indica un problema decidable.

	<i>RE</i>	<i>CS</i>	<i>CF</i>	<i>LIN</i>	<i>REG</i>
Equivalència	I	I	I	I	D
Inclusió	I	I	I	I	D
Pertanyença	I	D	D	D	D
Buidor	I	I	D	D	D
Finitud	I	I	D	D	D
Regularitat	I	I	I	I	D

Els problemes de decisió aconsegueixen un paper molt important dins la teoria de llenguatges formals. El mètode usual per tal de provar que un problema és indecidible consisteix en reduir-lo a una altra indecidibilitat ja coneguda. Al respecte, l'eina més útil pels problemes de la teoria de llenguatges formals és el *problema de correspondència de Post*.

Sigui V un vocabulari que contingui com a mínim dos símbols i siguin x, y dues n -tuples de cadenes no-buides sobre V , $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n)$.

Direm que el problema de correspondència de Post té una solució per a x, y si existeixen els índexs $i_1, \dots, i_m, m \geq 1$, amb $1 \leq i_j \leq n$ per a tot $1 \leq j \leq m$, tal que $x_{i_1} \dots x_{i_m} = y_{i_1} \dots y_{i_m}$. Clarament, si existeix aquesta solució, aleshores existeix un nombre infinit de solucions, obtingudes iterant les solucions “primitives”.

Sabem per (Post, 1946) que el problema de correspondència de Post és indecidible.

3.8 Condicions necessàries

Per tal de provar que un llenguatge no pertany a una família donada, generalment s’usen condicions necessàries.

Sigui $L \subseteq V^*$ un llenguatge. Com a la Secció 3.1, denotem per $c(L, x)$ el conjunt de tots els contextos que accepten x respecte a L , que és el conjunt $\{(u, v) \mid u, v \in V^*, uxv \in L\}$. Per a $x, y \in V^*$ escriurem $x \sim_L y$ si $c(L, x) = c(L, y)$. Aquesta relació és d’equivalència (és a dir, és reflexiva, simètrica, i transitiva). Denotem per $[x]_L$ les classes d’equivalència de $x \in V^*$ respecte a $\sim_L$; i, per $V^*/\sim_L$ el conjunt de totes les classes d’equivalència, $V^*/\sim_L = \{[x]_L \mid x \in V^*\}$ (aquest conjunt s’anomena també el *monoide sintàctic* del llenguatge L).

Teorema 3.8.1 (Teorema de Myhill-Nerode) Un llenguatge $L \subseteq V^*$ és regular

si, i només si, $V^*/\sim_L$ és un conjunt finit.

Definició 3.8.1 Per a un vocabulari $V = \{a_1, \dots, a_n\}$ (amb els símbols ordenats d'aquesta manera) definim la funció (denominada la *funció de Parikh*) $\Psi_V : V^* \longrightarrow \mathbb{N}^n$ per $\Psi(x) = (|x|_{a_1}, \dots, |x|_{a_n})$, $x \in V^*$. Podem estendre aquesta funció, de forma natural, a llenguatges. Dos llenguatges $L_1, L_2 \subseteq V^*$ tal que $\Psi_V(L_1) = \Psi_V(L_2)$ s'anomenen *lletra-equivalents*.

Teorema 3.8.2 Tot llenguatge lliure del context és lletra-equivalent a un llenguatge regular.

Considerem ara les operacions relacionades amb l'addició i la multiplicació per una constant, sobre el conjunt de vectors naturals de dimensió donada. Un subconjunt M de $\mathbb{N}^n$ s'anomena *lineal* si existeixen els vectors $v_0, v_1, \dots, v_m$, $m \geq 0$, tal que $M = \{v_0 + \sum_{i=1}^m p_i v_i \mid p_i \in \mathbb{N}\}$. La unió finita de conjunts lineals s'anomena un *conjunt semilineal*.

Teorema 3.8.3 (Teorema de Parikh) Si $L \subseteq V^*$ és un llenguatge lliure del context, aleshores $\Psi_V(L)$ és un conjunt semilineal.

Corol·lari 3.8.1 (i) Tot llenguatge lliure del context sobre un vocabulari d'una-letra és regular. (ii) El conjunt longitud, $length(L) = \{|x| \mid x \in L\}$, de qualsevol llenguatge lliure del context conté una progressió aritmètica infinita.

Considerem ara els tan coneguts i usats *lemes de bombeig*:

Teorema 3.8.4 (Lema de Bar-Hillel, o lema de $uvwxy$, o lema de bombeig per a llenguatges lliures del context) Si $L \in CF$, $L \subseteq V^*$, aleshores existeixen dues constants $p, q \in \mathbb{N}$ tal que cada cadena $w \in L$ amb $|w| > p$ pot ser escrita de la forma $w = u_1w_1u_2w_2u_3$, amb $u_1, w_1, u_2, w_2, u_3 \in V^*$, $|w_1u_2w_2| \leq q$, $w_1w_2 \neq \lambda$, i $u_1w_1^i u_2w_2^i u_3 \in L$ per a tot $i \geq 0$.

O en altres paraules, tota paraula suficientment llarga w d'un llenguatge lliure del context L es pot escriure de la forma següent: $w = u_1w_1u_2w_2u_3$, amb $w_1w_2 \neq \lambda$, on per a cada $i \geq 0$ la paraula $u_1w_1^i u_2w_2^i u_3 \in L$.

Teorema 3.8.5 (Lema de bombeig per a llenguatges lineals) Si $L \in LIN$, $L \subseteq V^*$, aleshores existeixen dues constants $p, q \in \mathbb{N}$ tal que cada cadena $w \in L$ amb $|w| > p$ pot ser escrita de la forma $w = u_1w_1u_2w_2u_3$, amb $u_1, w_1, u_2, w_2, u_3 \in V^*$, $|u_1w_1w_2u_3| \leq q$, $w_1w_2 \neq \lambda$, i $u_1w_1^i u_2w_2^i u_3 \in L$ per a tot $i \geq 0$.

Teorema 3.8.6 (Lema de bombeig per a llenguatges regulars) Si $L \in REG$, $L \subseteq V^*$, aleshores existeixen dues constants $p, q \in \mathbb{N}$ tal que cada cadena $w \in L$ amb $|w| > p$ pot ser escrita de la forma $w = u_1w_1u_2$, amb $u_1, w_1, u_2 \in V^*$, $|u_1w_1| \leq q$, $w_1 \neq \lambda$, i $u_1w_1^i u_2 \in L$ per a tot $i \geq 0$.

O en altres paraules, tota paraula suficientment llarga w d'un llenguatge regular L es pot escriure de la forma següent: $w = u_1w_1u_2$, amb $w_1 \neq \lambda$, on per a cada $i \geq 0$ la paraula $u_1w_1^i u_2 \in L$.

Teorema 3.8.7 (Lema d'Ogden, o lema de bombeig amb posicions marcades) Si $L \in CF$, $L \subseteq V^*$, aleshores existeix $p \in \mathbb{N}$ tal que per a cada $w \in L$, i per

a cada com a mínim p ocurrencies marcades de símbols a w , podem escriure $w = u_1 w_1 u_2 w_2 u_3$, on

- (i) o bé cada u_1, w_1, u_2 o bé cada u_2, w_2, u_3 conté com a mínim un símbol marcat,
- (ii) $w_1 u_2 w_2$ conté com a màxim p símbols marcats,
- (iii) $u_1 w_1^i u_2 w_2^i u_3 \in L$ per a tot $i \geq 0$.

Una altra condició necessària per tal que un llenguatge sigui lineal és:

Teorema 3.8.8 Si $L \subseteq V^*$ és un llenguatge lineal, aleshores existeixen dos llenguatges regulars $L_1, L_2 \subseteq V^*$ tal que

- (i) $L \subseteq L_1 L_2$,
- (ii) per a cada $x \in L_1$ ($y \in L_2$) existeix $y \in L_2$ ($x \in L_1$, resp.), tal que $xy \in L$.

Usant aquestes condicions necessàries podem demostrar les següents relacions:

$$L_1 = \{a^n b^n \mid n \geq 1\} \in LIN - REG,$$

$$L_2 = L_1 L_1 \in CF - LIN,$$

$$L_3 = \{a^n b^n c^n \mid n \geq 1\} \in CS - CF,$$

$$L_4 = \{xcx \mid x \in \{a, b\}^*\} \in CS - CF,$$

$$L_5 = \{a^{2^n} \mid n \geq 1\} \in CS - CF,$$

$$L_6 = \{a^n b^m c^n d^m \mid n, m \geq 1\} \in CS - CF,$$

$$L_7 = \{a^n b^m \mid n \geq 1, 1 \leq m \leq 2^n\} \in CS - CF,$$

$$L_8 = \{a^n b^m c^p \mid 1 \leq n \leq m \leq p\} \in CS - CF,$$

$$L_9 = \{x \in \{a, b\}^* \mid |x|_a = |x|_b\} \in CF - LIN,$$

$$L_{10} = \{x \in \{a, b, c\}^* \mid |x|_a = |x|_b = |x|_c\} \in CS - CF.$$

Cap de les condicions dels Teoremes 3.8.2 – 3.8.8 és també suficient per tal de determinar si un llenguatge és del tipus mencionat. Per tant, aquestes condicions no caracteritzen ni els llenguatges lliures del context, ni els llenguatges lineals, ni tampoc els llenguatges regulars.

Coneixent les propietats de clausura d'una família F de llenguatges i condicions necessàries per tal que un llenguatge pertanyi a F , una estratègia usual a seguir per tal de provar que un llenguatge donat L no pertany a F és la següent: usant operacions sota les quals F sigui tancada, modifiquem L , transformant-lo en un llenguatge L' . Si podem provar que $L' \notin F$, possiblement usant condicions necessàries, aleshores podrem inferir que $L \notin F$.

Aquest esquema és el que s'ha usat la majoria de vegades per tal d'intentar provar que certs llenguatges naturals no són lliures del context. A continuació

explicitarem un dels primers intents (si bé refutat actualment, veure, p.e., (Gazdar i Pullum, 1985)), aparegut a (Bar-Hillel i Shamir, 1964):

Considerem el conjunt L de totes les frases de l'anglès de la forma

“John, Mary, David, ... are a widower, a widow, a widower, ... , respectively”.

Acceptem el fet que aquesta frase és gramatical (no només “lògicament correcta”) si, i només si, la llista de noms té la mateixa longitud que la llista de paraules “a widower”, “a widow”; i, a més a més el i -èsim nom de la primera llista té el mateix gènere que el i -èsim adjectiu de la segona llista. Des del punt de vista de la gramàtica anglesa, el conjunt d'aquestes frases és infinit.

Considerem ara la gramàtica lineal a la dreta (i, per tant, regular) $G = (\{S, X, Y\}, V_T, S, P)$, on $V_T = \{a\ widower, a\ widow, are, respectively, coma, punt\} \cup N$, essent N el conjunt donat de noms, i contenint el conjunt P les següents regles:

$$S \rightarrow nom\ coma\ X, \text{ per a } nom \in N,$$

$$X \rightarrow nom\ coma\ X, \text{ per a } nom \in N,$$

$$X \rightarrow nom\ are\ Y, \text{ per a } nom \in N,$$

$$Y \rightarrow a\ widow\ coma\ Y,$$

$$Y \rightarrow a\ widower\ coma\ Y,$$

$$Y \rightarrow respectively\ punt.$$

Usem *coma* i *punt*, en comptes de “,” i “.”, per tal d'evitar confusions

entre aquests símbols i el metallenguatge de les comes i els punts. Clarament, interseccionant l'anglès amb $L(G)$ obtenim el llenguatge L . Si l'anglès fos lliure del context, aleshores, ja que $L(G)$ és regular i CF és tancat sota intersecció amb llenguatges regulars (c.f. Taula 3.6.2 de la pàgina 148), se segueix que L és lliure del context. Ara bé, això no és veritat: prenguem una *gsm* g que substitueixi cada nom masculí pel símbol a , cada nom femení per b , substitueixi la paraula “are” pel símbol c , substitueixi “a widower” per a i “a widow” per b , i esborri totes les comes, “respectively”, i el període final. aleshores, obtenim

$$g(\text{Anglès} \cap L(G)) = g(L) = \{xcx \mid x \in \{a, b\}^*\},$$

que és el llenguatge L_4 considerat abans. Aquest és un llenguatge que no és lliure del context, la qual cosa és una contradicció amb la hipòtesi que l'anglès és lliure del context.

Un argument similar (intersecció amb un llenguatge regular, després una *gsm* o només un morfisme per tal d'esborrar/modificar les cadenes) s'usa en molts altres llocs.

3.9 Representacions

Sovint, els llenguatges d'una família poden ser obtinguts a partir de llenguatges més simples (és a dir, de llenguatges d'una subfamília pròpia) usant certes operacions. A continuació farem esment d'alguns resultats d'aquest tipus:

Teorema 3.9.1 Per a cada llenguatge $L \subseteq V^*$, $L \in RE$, existeixen dos símbols $c, d \notin V$ i un llenguatge $L' \subseteq Lcd^*$, $L' \in CS$, tal que per a cada $x \in L$ existeix $i \geq 0$ amb la propietat $xcd^i \in L'$.

Corol·lari 3.9.1 Tot llenguatge $L \in RE$ pot ser escrit de la forma $L = h(L')$, on $L' \in CS$, h és un morfisme esborrador, i L' és un llenguatge regular.

Teorema 3.9.2 Tot llenguatge $L \in RE$ pot ser escrit de la forma $L = h(L_1 \cap L_2) = L_3/L_4$, on tots els llenguatges L_1, L_2, L_3, L_4 són lineals i h és un morfisme esborrador.

El llenguatge generat per la gramàtica lliure del context $G = (\{S\}, \{a_1, b_1, \dots, a_n, b_n\}, S, \{S \rightarrow SS, S \rightarrow \lambda\} \cup \{S \rightarrow a_i S b_i \mid 1 \leq i \leq n\})$ s'anomena *llenguatge de Dyck* (sobre n parells de símbols) i es denota per D_n . Es pot observar que, interpretant cada parell (a_i, b_i) com un parell de parèntesis, D_n és el conjunt de les expressions ben formades d' n parèntesis. Podem definir també aquest llenguatge com el conjunt de totes les cadenes sobre $\{a_1, b_1, \dots, a_n, b_n\}$ les quals poden ser reduïdes a λ mitjançant l'ús iteratiu de les regles de reducció $a_i b_i \rightarrow \lambda, 1 \leq i \leq n$. Quan $n = 1$ considerarem D_1 com un llenguatge sobre $\{a, b\}$ i el denotarem per $D_{\{a,b\}}$.

Teorema 3.9.3 (Teorema de Chomsky-Schützenberger) Tot llenguatge lliure del context L pot ser escrit de la forma $L = h(D_n \cap R)$, on h és un morfisme, $D_n, n \geq 1$, és un llenguatge de Dyck, i R és un llenguatge regular.

Òbviamment, els llenguatges $D_n, n \geq 1$, són no lineals.

3.10 El teorema de l'espai de treball

Del Teorema 3.9.1 de la pàgina 159 inferim que tot llenguatge recursiu enumerable pot ser aproximat mitjançant un llenguatge sensible al context amb una cua. El següent resultat ens diu, en principi, que si aquesta cua no és “massa llarga”, aleshores esborrant-la encara roman en la família de llenguatges sensibles al context.

Sigui $G = (V_N, V_T, S, P)$ una gramàtica de tipus-0 i considerem una derivació a G

$$\delta : S = w_0 \Longrightarrow w_1 \Longrightarrow \dots \Longrightarrow w_m = y, y \in V_T^*, m \geq 1.$$

Definim el *marc de treball*, que denotarem per WS (de “workspace”), com

$$WS(\delta, G) = \max_{0 \leq i \leq m} |w_i|,$$

i, per a $y \in L(G)$, expressem

$$WS(y, G) = \min\{WS(\delta, G) \mid \delta : S \Longrightarrow^* y\}.$$

Direm que G té un marc de treball acotat si existeix una constant k tal que, per a cada $x \in L(G) - \{\lambda\}$ tenim que $WS(x, G) \leq k|x|$.

Teorema 3.10.1 Si G té un espai de treball acotat, aleshores $L(G) \in CS$.

3.11 Formes normals

Treballar amb gramàtiques d'una determinada forma, sense perdre el seu poder de generació, és en general molt profitós. Existeixen alguns resultats que garanteixen l'existència d'aquestes formes normals.

Teorema 3.11.1 (Forma normal de Chomsky) Per a tota gramàtica G lliure del context, podem construir una nova gramàtica $G' = (V_N, V_T, S, P)$, amb les regles de P de les formes $A \rightarrow a$ i $A \rightarrow BC$, per a $A, B, C \in V_N, a \in T$, tal que $L(G') = L(G) - \{\lambda\}$.

Teorema 3.11.2 (Forma normal de Greibach) Per a cada gramàtica G lliure del context, podem construir una nova gramàtica $G' = (V_N, V_T, S, P)$, amb les regles de P de la forma $A \rightarrow a\alpha$, per a $A \in V_N, a \in T, \alpha \in (V_N \cup V_T)^*$, tal que $L(G') = L(G) - \{\lambda\}$.

Si volem obtenir que $L(G') = L(G)$ en els dos teoremes anteriors, aleshores també haurem d'admetre una regla de compleció $S \rightarrow \lambda$.

Teorema 3.11.3 (La forma super-normal) Per a cada tripla (k, l, m) d'enters no negatius i per a cada gramàtica lliure del context G , podem construir una gramàtica equivalent $G' = (V_N, V_T, S, P)$ que contingui regles de les formes següents:

- (i) $A \rightarrow xByCz$, amb $A, B, C \in V_N, x, y, z \in V_T^*$, i $|x| = k, |y| = l, |z| = m$,

(ii) $A \rightarrow x$, amb $A \in V_N, x \in V_T^*, |x| \in \text{length}(L(G))$.

Noves variants de les formes normals de Chomsky i de Greibach poden ser obtingudes particularitzant els paràmetres anteriors k, l, m .

Teorema 3.11.4 (Forma normal de Kuroda) Per a cada gramàtica de longitud creixent G , podem construir una gramàtica $G' = (V_N, V_T, S, P)$, amb les regles de P de les formes $A \rightarrow BC, A \rightarrow a, AB \rightarrow CD$, per a $A, B, C, D \in V_N, a \in V_T$ (més una regla de compleció, $S \rightarrow \lambda$, si és necessari), tal que $L(G) = L(G')$.

S'obté un resultat similar per a gramàtiques de Chomsky arbitràries, on es permeten també regles del tipus $A \rightarrow \lambda, A \in V_N$, en la gramàtica G'

3.12 Ambigüitat

Definició 3.12.1 Una gramàtica G és *ambigua* si, i només si, existeix alguna cadena $x \in L(G)$ amb, com a mínim, dues derivacions diferents a G (és el mateix que dir que té dos arbres de derivació diferents). En qualsevol altre cas, G és *no-ambigua*.

Definició 3.12.2 Una gramàtica G és *ambigua de grau k* (k és un nombre natural o ∞) si, i només si, tota cadena $x \in L(G)$ té, com a màxim, k derivacions diferents a G (és el mateix que dir que té k arbres de derivació diferents),

i alguna cadena $x \in L(G)$ posseeix exactament k derivacions (que és el mateix que dir que té k arbres de derivació).

Definició 3.12.3 Un llenguatge L és *ambigu de grau k* si, i només si, $L = L(G)$, per a alguna gramàtica G ambigua de grau k , i no existeix cap altra gramàtica G' ambigua de grau menor que k tal que $L = L(G')$.

Definició 3.12.4 Un llenguatge L és *no-ambigu* si, i només si, $L = L(G)$, per a alguna gramàtica G no-ambigua. En qualsevol altre cas, L rep el nom d'*inherentment ambigu*.

Teorema 3.12.1 i) Existeixen llenguatges lineals inherentment ambigus.

ii) Tot llenguatge regular és no-ambigu.

iii) El problema de l'ambigüitat per a gramàtiques lliures del context és indecidible.

3.13 Complexitat descriptiva

Un llenguatge donat pot ser generat per moltes, sovint infinites, gramàtiques diferents. Per això, sembla natural buscar aquelles gramàtiques que siguin les més simples possibles; i, per això, necessitem mesures relatives a complexitat de gramàtiques.

Donada una classe $\mathcal{G}$ de gramàtiques, una mesura de complexitat descriptiva (també l'anomenarem *mesura de complexitat sintàctica*) és una funció $K : \mathcal{G} \rightarrow \mathbb{N}$ la qual pot ser extesa a llenguatges generats per elements de $\mathcal{G}$ mitjançant

$$K(L) = \min\{K(G) \mid L = L(G), G \in \mathcal{G}\}.$$

Si és necessari, també escriurem $K_{\mathcal{G}}(L)$, per tal d'especificar la classe de gramàtiques usada.

Les tres mesures bàsiques per a llenguatges lliures del context són les següents: donada una gramàtica lliure del context $G = (V_N, V_T, S, P)$ definim

$$Var(G) = card(V_N),$$

$$Prod(G) = card(P),$$

$$Symb(G) = \sum_{r \in P} Symb(r), \text{ on } Symb(r : A \rightarrow x) = |x| + 2.$$

Una mesura de complexitat K s'anomena *no-trivial* si per a cada n existeix una gramàtica G_n tal que $K(L(G_n)) > n$; K s'anomena *connectada* si existeix n_0 tal que per a cada $n \geq n_0$ existeix G_n amb $K(L(G_n)) = n$.

En una mesura de complexitat descriptiva K (de gramàtiques d'una classe $\mathcal{G}$) donada, sembla natural formular una sèrie de qüestions o problemes relatius a la computabilitat i a la decidibilitat (veure (Gruska, 1973)):

Q1. És $K(G)$ computable per a una gramàtica arbitrària G ?

Q2. És $K(L(G))$ computable per a una gramàtica arbitrària G ?

- Q3. És o no és decidible quan $K(L(G)) = n$ per una gramàtica G i un valor d' n arbitrari ?
- Q4. És o no és decidible quan $K(G) = K(L(G))$ per una gramàtica arbitrària G ? (És o no és decidible quan una gramàtica arbitrària és la més simple, respecte al paràmetre K , pel llenguatge generat ?)
- Q5. Donada una gramàtica arbitrària G , podem construir algorímicament una gramàtica G' tal que $K(G') = K(L(G))$? (Podem construir una gramàtica que sigui la de menor complexitat, respecte al paràmetre K , per un llenguatge arbitrari ?)

Per a les mesures prèvies *Var*, *Prod*, *Symb*, la qüestió Q1 té una resposta afirmativa, mentre que per a la resta de qüestions la resposta és negativa. Aquestes tres mesures estan connectades (fins i tot en la família dels llenguatges regulars). En general, dues mesures de complexitat sintàctica no poden ser acomplides simultàniament: si trobem una gramàtica que sigui la més simple des del punt de vista d'una mesura, aleshores aquesta gramàtica serà més complexa des del punt de vista de les altres dues mesures. Una situació similar es dona entre complexitat i ambigüitat.

Notes bibliogràfiques. No presentarem aquí més elements de la teoria de llenguatges formals. Algunes nocions i resultats més singulars seran introduïts quan sigui necessari. Pel cas de referències generals en teoria de llenguatges

formals hom es pot remetre a alguns monogràfics en aquesta àrea, com poden ser (Ginsburg, 1966), (Harrison, 1978), (Hopcroft i Ullman, 1969; Hopcroft i Ullman, 1979), (Salomaa, 1973); una font comprensiva d'informació és (Rozenberg i Salomaa, 1996b) o bé (Păun, 1996a). Per a l'àrea de reescriptura regulada ens podem remetre a (Dassow i Păun, 1989) i resultats més recents poden ser trobats a (Dassow et al., 1996b); també a (Martín-Vide, 1994) es pot trobar una introducció al respecte. Pel que fa a la teoria d'AFL es pot consultar (Ginsburg, 1975), per a l'àrea de sistemes de gramàtiques (Csuha-Jarju et al., 1994) i, finalment, per a Sistemes de Lindenmayer (Rozenberg i Salomaa, 1980).

Capítol 4

Gramàtiques contextuais bàsiques

4.1 Definicions

Quan introduïm diversos tipus de gramàtiques contextuais, podem usar el mateix estil que en el cas de les gramàtiques de Chomsky, començant amb les gramàtiques més complexes i particularitzant la definició per tal d'obtenir les classes més senzilles. O també podem usar l'estratègia inversa, començant amb les gramàtiques més senzilles. Nosaltres seguirem en aquest cas una estratègia intermèdia, introduint primer la classe de gramàtiques que és més usada en aquesta tesi.

Definició 4.1.1 Una *gramàtica contextual amb selecció* és una construcció

$$G = (V, B, C, \varphi),$$

on V és un vocabulari finit no-buit, B (el conjunt base o conjunt d'axiomes) és un llenguatge finit sobre V , C és un subconjunt finit de $V^* \times V^*$, i $\varphi : V^* \longrightarrow 2^C$ és una funció.

Les cadenes de B s'anomenen *axiomes*, els elements (u, v) de C són anomenats *contextos* i φ és la funció de *selecció/elecció*.

Definició 4.1.2 Respecte a una gramàtica contextual com l'anterior, definim dues relacions a V^* : per a $x, y \in V^*$, escrivim

$$\begin{aligned} x \Longrightarrow_{ex} y \quad &\text{sii} \quad y = uxv, \text{ per algun } (u, v) \in \varphi(x), \\ x \Longrightarrow_{in} y \quad &\text{sii} \quad x = x_1x_2x_3, y = x_1ux_2vx_3, \text{ per algun} \\ &x_1, x_2, x_3 \in V^*, (u, v) \in \varphi(x_2). \end{aligned}$$

Direm que $\Longrightarrow_{ex}$ és una derivació *externa* a G , mentre que $\Longrightarrow_{in}$ és una derivació *interna*. Denotant, de manera usual, per $\Longrightarrow_{ex}^*, \Longrightarrow_{in}^*$ la clausura reflexiva i transitiva d'aquestes relacions, podem associar a G dos llenguatges,

$$L_\alpha(G) = \{x \in V^* \mid w \Longrightarrow_\alpha^* x, \text{ per algun } w \in B\},$$

per a $\alpha \in \{ex, in\}$.

Direm que $L_{ex}(G)$ és el llenguatge extern generat per G i $L_{in}(G)$ és el llenguatge intern. També direm que una gramàtica usada en el mode extern

és una *gramàtica contextual externa* i una gramàtica usada en el mode intern és una *gramàtica contextual interna*.

Podem observar que les cadenes d'aquests llenguatges són, o bé axiomes, o bé s'han obtingut afegint als axiomes un nombre finit de contextos (al final de la corresponent cadena en el cas del llenguatge extern i afegint a una subcadena en el cas del llenguatge intern) depenentment de la funció de selecció. A més a més, també podem definir els llenguatges $L_\alpha(G)$ de la següent manera (*algebraica*): Donada una gramàtica contextual (amb selecció) $G = (V, B, C, \varphi)$, el llenguatge $L_{ex}(G)$ és el llenguatge $L \subseteq V^*$ més petit (en el sentit de la relació d'inclusió) tal que

1. $B \subseteq L$,
2. si $x \in L$ i $(u, v) \in \varphi(x)$, aleshores $uxv \in L$,

mentre que el llenguatge $L_{in}(G)$ és el llenguatge més petit $L \subseteq V^*$ amb les propietats

1. $B \subseteq L$,
2. si $x \in L$, $x = x_1x_2x_3$, per algun $x_1, x_2, x_3 \in V^*$, i $(u, v) \in \varphi(x_2)$, aleshores $x_1ux_2vx_3 \in L$.

Pel llenguatge extern, també podem donar una definició *explícita*:

$$\begin{aligned} L_{ex}(G) = B \cup \{ & u_n \dots u_1 w v_1 \dots v_n \mid n \geq 1, w \in A, \\ & (u_1, v_1) \in \varphi(w), (u_i, v_i) \in \varphi(u_{i-1} \dots u_1 w v_1 \dots v_{i-1}), \\ & i = 2, 3, \dots, n \}. \end{aligned}$$

En els exemples i demostracions usarem aquesta definició dels llenguatges generats per una gramàtica contextual ja que és més adequada en aquests casos en concret. De fet, també la definició intrínseca de la gramàtica pot ésser presentada en dos models.

Per exemple, per a una gramàtica contextual $G = (V, B, C, \varphi)$, $\varphi : V^* \longrightarrow 2^C$, la relació $(u, v) \in \varphi(x)$ pot ésser interpretada com a una regla de reescriptura $x \rightarrow u x v$. A més a més, una gramàtica contextual amb derivació interna pot ésser considerada com a una gramàtica pura amb un nombre arbitrari de produccions (de longitud creixent) de la forma $x \rightarrow u x v$, amb uv de longitud acotada.

Això suggereix la *presentació modular* d'una gramàtica contextual, això és, com un sistema

$$G = (V, B, P),$$

on V és un vocabulari, B és un llenguatge finit sobre V , i P és un conjunt finit de parells (S, C) , amb $S \subseteq V^*$, $S \neq \emptyset$ i C un subconjunt finit de $V^* \times V^*$. Els parells (S, C) s'anomenen *produccions*, S és el *selector* d'aquestes produccions, i C és el conjunt de *contextos*. La interpretació és equivalent a $C \subseteq \varphi(x)$ per a

tot $x \in S$: els contextos de C poden ésser afegits a cadenes d' S . Malgrat P és considerat un *conjunt*, sovint l'escriurem com a una seqüència d'aquestes produccions, donada la gramàtica de la forma $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$.

Quan presentem una gramàtica en la forma modular $G = (V, B, P)$ sovint usarem la funció φ definida per $\varphi(x) = \{(u, v) \mid (u, v) \in C \text{ per a algunes produccions } (S, C) \in P \text{ tal que } x \in S\}$. Per a $x \in V^* - \{y \in S \mid (S, C) \in P\}$ direm que $\varphi(x) = \emptyset$.

Inversament, si tenim una gramàtica contextual $G = (V, B, C, \varphi)$ en la presentació funcional, és fàcil produir gramàtiques equivalents en la presentació modular: per a cada $C' \subseteq C$ podem considerar la producció

$$(\{x \in V^* \mid C' \subseteq \varphi(x)\}, C').$$

Part dels selectors $\{x \in V^* \mid C' \subseteq \varphi(x)\}, C' \subseteq C$, són buits; i és clar que, en aquests casos, les corresponents produccions són ignorades.

En general, donada una gramàtica en la *presentació funcional*, $G = (V, B, C, \varphi)$, especificarem la funció φ només per a cadenes x tals que $\varphi(x) \neq \emptyset$; queda implícitament sobreentès que $\varphi(x) = \emptyset$ per a totes les x no especificades.

En els exemples i demostracions usarem la presentació de les gramàtiques contextuais més convenient per a cada marc de treball en particular; fins i tot en el cas de la presentació modular usarem la funció de selecció φ definida com abans, sense definir-la explícitament una altra vegada.

La presentació modular suggereix, d'una manera natural, una classificació de les gramàtiques contextuais que depèn del tipus de selectors en una jerarquia específica de llenguatges. Direm que una gramàtica $G = (V, B, P)$, amb $P = \{(S_1, C_1), \dots, (S_n, C_n)\}$, és de tipus F (o és amb F selecció/elecció) si $S_i \in F$, $1 \leq i \leq n$, per a una família F de llenguatges donada. Aquesta classificació és cabdal per a la teoria de gramàtiques contextuais i la estudiarem en detall en les subsegüents seccions.

En una gramàtica contextual interna, el context adjuntat depèn d'una subcadena de la cadena derivada. Si imposem la condició que el context adjuntat depèn de la cadena íntegra, obtenim la següent classe més robusta de gramàtiques contextuais:

Definició 4.1.3 Una *gramàtica contextual total* és una construcció

$$G = (V, B, C, \varphi),$$

on V és un vocabulari, B és un llenguatge finit sobre V , C és un subconjunt finit de $V^* \times V^*$ i $\varphi : V^* \times V^* \times V^* \longrightarrow 2^C$.

El llenguatge generat és

$$L(G) = \{x \in V^* \mid w \Longrightarrow^* x, \text{ per algun } w \in B\},$$

on $\Longrightarrow^*$ és la clausura reflexiva i transitiva de la relació $\Longrightarrow$ definida per

$$x \Longrightarrow y \quad \text{sii} \quad x = x_1 x_2 x_3, \quad y = x_1 u x_2 v x_3, \text{ per algun}$$

$$x_1, x_2, x_3 \in V^*, (u, v) \in C, \text{ tal que } (u, v) \in \varphi(x_1, x_2, x_3).$$

És fàcil de veure que les gramàtiques contextuais totals són una generalització de les gramàtiques internes i externes. Per això, particularitzant la forma de la funció φ en una gramàtica contextual total, podem obtenir els altres tipus particulars:

- (i) Una gramàtica contextual externa és una gramàtica contextual total $G = (V, B, C, \varphi)$ amb $\varphi(x_1, x_2, x_3) = \emptyset$, per a tot $x_1, x_2, x_3 \in V^*$ tal que $x_1 x_3 \neq \lambda$.
- (ii) Una gramàtica contextual interna és una gramàtica contextual total $G = (V, B, C, \varphi)$ amb $\varphi(x_1, x_2, x_3) = \varphi(x'_1, x_2, x'_3)$, per a tot $x_1, x'_1, x_2, x_3, x'_3 \in V^*$.

Una posterior (més basta) simplificació de les gramàtiques contextuais (totals) és ignorar la selecció; aleshores les gramàtiques contextuais totals coincideixen amb les gramàtiques contextuais internes:

Definició 4.1.4 Una gramàtica $G = (V, B, C, \varphi)$ s'anomena *sense selecció* si $\varphi(x) = C$, per a tot $x \in V^*$. (En la presentació modular tenim que $G = (V, B, C, (V^*, C))$.)

En aquest cas en particular, la funció φ pot ésser ignorada, i escriurem la gramàtica de la forma $G = (V, B, C)$. Novament, definim dues relacions de

derivació:

$$x \Rightarrow_{ex} y \quad \text{sii} \quad y = uv \text{ per algun } (u, v) \in C,$$

$$x \Rightarrow_{in} y \quad \text{sii} \quad x = x_1x_2x_3, y = x_1ux_2vx_3 \text{ per algun } (u, v) \in C.$$

D'aquesta manera, obtenim cinc famílies bàsiques de llenguatges:

TC = la família de llenguatges generada per gramàtiques contextuais totals,

ECC = la família de llenguatges generada externament per gramàtiques contextuais amb selecció,

ICC = la família de llenguatges generada internament per gramàtiques contextuais amb selecció,

EC = la família de llenguatges generada externament per gramàtiques contextuais sense selecció,

IC = la família de llenguatges generada internament per gramàtiques contextuais sense selecció.

En tots els casos de gramàtiques amb selecció, podem imposar també la restricció natural de tenir una funció computable φ . Les famílies corresponents són denotades per TC_c, ECC_c, ICC_c , respectivament.

4.2 Exemples

Els exemples següents tenen la intenció d'il·lustrar les definicions anteriors i la manera de treballar i el poder dels diversos tipus de gramàtiques contextuais. També volen ser un punt de referència per a altres seccions d'aquesta tesi.

Exemple 4.2.1 Tot llenguatge finit, B , pertany a cadascuna de les famílies de llenguatges contextuais: prenent $G = (V, B, \emptyset)$ tenim que $L_{ex}(G) = L_{in}(G) = B$.

Exemple 4.2.2 Per a cada vocabulari V , el llenguatge V^* pertany a cadascuna de les famílies de llenguatges contextuais: per a $G = (V, \{\lambda\}, \{\lambda\} \times V)$ tenim que $L_{ex}(G) = L_{in}(G) = V^*$.

Si prenem $G' = (V, V, \{\lambda\} \times V)$, aleshores $L_{ex}(G) = L_{in}(G) = V^+$.

En general, per a $G_k = (V, V^k, \{\lambda\} \times V)$, $k \geq 0$, obtenim que $L_{ex}(G_k) = L_{in}(G_k) = V^k V^*$.

Exemple 4.2.3 Considerem la gramàtica $G = (\{a, b\}, \{\lambda\}, \{(a, b)\})$. Òbviament,

$$L_{ex}(G) = \{a^n b^n \mid n \geq 1\},$$

mentre que $L_{in}(G) = D_{\{a, b\}}$, el llenguatge de Dyck sobre el vocabulari $\{a, b\}$. Aquesta segona igualtat pot ésser demostrada per inducció sobre la longitud de les cadenes, usant la següent caracterització dels llenguatges de Dyck (que

serveix també per a la Secció 2.2): $D_{\{a,b\}}$ és el llenguatge més petit $D \subseteq \{a,b\}^*$ que posseeix les tres propietats següents

- (i) $\lambda \in D$,
- (ii) si $x, y \in D$, aleshores $xy \in D$,
- (iii) si $x \in D$, aleshores $x_1ax_2bx_3 \in D$ per a tot $x_1, x_2, x_3 \in \{a,b\}^*$ tal que $x = x_1x_2x_3$.

Insertant un parell de símbols ab en la cadena de $D_{\{a,b\}}$ obtenim una cadena pertanyent al mateix llenguatge. Inversament, prenent una cadena no-buida $w \in D_{\{a,b\}}$, podem escriure-la de la forma $w = w_1abw_2$ amb $w_1, w_2 \in \{a,b\}^*$, de manera que $w_1w_2 \in D_{\{a,b\}}$. La cadena més curta que pertany a $D_{\{a,b\}}$, que és λ , és un axioma de G .

Cal notar que $L_{ex}(G)$ és no regular i $L_{in}(G)$ és no lineal.

Exemple 4.2.4 Per a

$$G = (\{a, b, c\}, \{\lambda\}, \{(\alpha\beta, \gamma), (\alpha, \beta\gamma) \mid \{\alpha, \beta, \gamma\} = \{a, b, c\}\}),$$

obtenim

$$L_{in}(G) = \{x \in \{a, b, c\}^* \mid |x|_a = |x|_b = |x|_c\}.$$

La inclusió $\subseteq$ és òbvia (cada context incrementa el nombre d'ocurrències a, b, c amb una més); la inclusió inversa pot ésser demostrada mitjançant inducció

sobre la longitud de les cadenes (cada $x \in \{a, b, c\}^*$ amb $|x|_a = |x|_b = |x|_c$ conté una subcadena $\alpha\beta$, per a $\alpha, \beta \in \{a, b, c\}$, $\alpha \neq \beta$, i també una ocurrència de γ , abans o després de $\alpha\beta$, tal que $\{\alpha, \beta, \gamma\} = \{a, b, c\}$; eliminant aquests símbols α, β, γ – es corresponen a un context de G – obtenim una cadena estrictament més curta, x' , amb $|x'|_a = |x'|_b = |x'|_c$).

Cal notar que el llenguatge anterior és no-lliure del context (algunes vegades és anomenat llenguatge de Bach).

Exemple 4.2.5 La gramàtica $G = (\{a\}, \{\lambda, a^3\}, \{(\lambda, a), (\lambda, a^2)\}, \varphi)$, amb la funció φ definida per

$$\varphi(a^i) = \begin{cases} (\lambda, a), & \text{si } i+1 \neq 2^k, k \geq 0, \\ (\lambda, a^2), & \text{si } i+1 = 2^k, k \geq 0, \end{cases}$$

per a tot $i \geq 1$, genera

$$L_{ex}(G) = \{a^n \mid n \neq 2^k, k \geq 0\},$$

$$L_{in}(G) = a^* - \{a, a^2\}$$

(en el mode intern, cada cadena a^i , $i \geq 3$, pot ésser prolongada amb un símbol a , usant $(\lambda, a) \in \varphi(a^2)$, per alguna subcadena a^2 de a^i).

Exemple 4.2.6 Prenem un llenguatge arbitrari $L \subseteq V^*$ i un símbol $c \notin V$, i construïm la gramàtica

$$G_L = (V \cup \{c\}, \{\lambda\}, \{(\lambda, \alpha) \mid \alpha \in V \cup \{c\}\}, \varphi),$$

$$\varphi(x) = \begin{cases} \{(\lambda, a) \mid a \in V\}, & \text{per a } x \in V^* - L, \\ \{(\lambda, a) \mid a \in V\} \cup \{(\lambda, c)\}, & \text{per a } x \in L. \end{cases}$$

Obtenim $L_{ex}(G_L) = V^* \cup L\{c\}$.

Cal notar que si L és no recursiu, aleshores la funció associada φ és no computable. A més a més, si $L \notin F$ per a una família F de llenguatges tancada sota derivació a la dreta, aleshores també $L_{ex}(G) \notin F$. Ja que $L = \partial_c^d(L_{ex}(G_L))$.

Exemple 4.2.7 Prenem $G = (\{a, b\}, \{a\}, \{(\lambda, a), (\lambda, b)\}, \varphi)$, amb

$$\varphi(a^i) = \begin{cases} \{(\lambda, a), (\lambda, b)\}, & \text{si } i = 2^p 3^q 5^r 7^s, p^s + q^s = r^s, s \geq 3, \\ \{(\lambda, a)\}, & \text{en qualsevol altre cas,} \end{cases}$$

per a tot $i \geq 1$. Obtenim

$$L_{ex}(G) = a^+ \cup \{a^n b \mid n = 2^p 3^q 5^r 7^s, (p, q, r, s) \text{ que correspon a un contraexemple de l'últim teorema de Fermat}\}.$$

Exemple 4.2.8 Considerem la gramàtica

$$G = (\{a, b, c\}, \{babccab\}, \{(\lambda, a), (a, a), (b, b), (c, c)\}, \varphi),$$

$$\varphi(ab^n cca) = \{(a, a)\}, \quad n \geq 1,$$

$$\varphi(bcca^n b) = \{(b, b)\}, \quad n \geq 1,$$

$$\varphi(ba^n b^m c) = \{(\lambda, a)\}, \quad n, m \geq 1,$$

$$\varphi(a^n b^n caca^m b^m) = \{(c, c)\}, \quad n, m \geq 1.$$

Podem comprovar que obtenim

$$L_{in}(G) = \{ba^n b^m ca^p ca^n b^m \mid n, m \geq 1, p \geq 0\} \cup \\ \cup \{bc^i a^n b^n caca^n b^n c^i \mid n, i \geq 1\}.$$

Exemple 4.2.9 Considerem ara un exemple més complicat, d'una gramàtica presentada en la forma modular. Per a

$$G = (\{a, b, c, d\}, \{a\}, (\{a\}, \{(\lambda, b), (\lambda, c)\}), \\ (\{bc\}, \{(\lambda, d)\}), (\{dc\}, \{(c, d)\})),$$

tenim

$$length(L_{in}(G) \cap ab(cd)^+) = \{2^n \mid n \geq 2\}. \quad (*)$$

Ja que $\varphi(x) \neq \emptyset$ només per a un nombre finit de cadenes x , la gramàtica G és equivalent a la gramàtica pura $G' = (\{a, b, c, d\}, a, P)$ amb el conjunt P consistint en les següents produccions:

$$\pi_1 : a \rightarrow ab, \pi_2 : a \rightarrow ac, \pi_3 : bc \rightarrow bcd, \pi_4 : dc \rightarrow cdcd.$$

Examinem les cadenes del conjunt $L = L(G') \cap ab(cd)^+$. Per tal de produir una cadena $z = ab(cd)^m, m \geq 1$, hem d'usar una vegada la producció π_1 (en z només existeix una ocurrència de b) i el mateix nombre de vegades les produccions π_2, π_3 ; π_2 no pot ésser usada després d'usar π_1 (b ha de romandre adjacent a a) i π_3 no pot ésser usada abans d'usar π_1 . Per tant, comencem per usar n vegades π_2 , després π_1 , i després π_3 (n vegades). Després d'usar π_1 tenim

la cadena abc^n , $n \geq 1$. Les produccions π_3 poden ésser usades només per afegir d a bc , i bc no canvia mai; l'ocurrència més a l'esquerra de c és sempre la mateixa durant tota la derivació. Ja que podem canviar l'ordre d'ús de π_3 i π_4 sense modificar la cadena generada, podem assumir que després d'haver produït abc^n , hem produït $abcd^n c^{n-1}$ usant π_3 . A partir d'aquest moment, només podem usar produccions del tipus π_4 (d'altra manera obtenim una cadena que no pertany a L).

Per inducció sobre n podem provar que si la cadena obtinguda a partir de $abcd^n c^{n-1}$ pertany a $ab(cd)^+$, aleshores aquesta és igual a $ab(cd)^{2^n-1}$, la qual cosa conclou la demostració de la igualtat (*).

Notes bibliogràfiques. Les gramàtiques contextuais externes, amb i sense selecció, foren introduïdes a (Marcus, 1969a), amb el llenguatge generat definit de la següent manera: si $G = (V, B, C, \varphi)$ com abans, aleshores $L(G)$ és el llenguatge més petit $L \subseteq V^*$ que conté B i posseeix la propietat que, per a cada $x \in L$, $(u, v) \in \varphi(x)$, i $z \in B$, totes les cadenes uxv , zx , i xz pertanyen a L . Això sembla molt restrictiu, ja que implica, per exemple, que $B^+ \subseteq L(G)$. La definició ha estat modificada en l'estil presentat en la Secció 4.1 de (Păun, 1974b), usada també en tots els articles posteriors. Les gramàtiques contextuais internes foren introduïdes a (Nguyen i Păun, 1980). Les gramàtiques contextuais totals foren considerades a (Păun, 1982a), seguint els suggeriments de (Nguyen, 1980c; Nguyen, 1980d; Nguyen, 1981b; Nguyen, 1981c). Aquests articles tenen en compte n -gramàtiques contextuais on els “context-

tos” són n -tuples, adjuntats de dependentment de $n + 1$ subcadena de la cadena derivada (la Secció X d’aquest llibre està dedicada a aquestes gramàtiques). La presentació modular va ser considerada per primera vegada a (Păun et al., 1994b). Gramàtiques amb selecció restringida van ser considerades en primer lloc a (Istrail, 1978a).

Una variant en la definició de les gramàtiques contextuais és considerar C com un subconjunt finit de $V^*\$V^*$, on $\$$ és un símbol especial no existent en V . Aleshores els elements de C , els contextos, tenen la forma $u\$v$. Aquesta definició, usada a (Ehrenfeucht et al., 1996a), està fonamentada en la recent introducció d’una nova branca en la teoria de llenguatges formals: la computació de l’ADN, basada en l’operació d’empalmar (cf. (Head, 1987) i (Head et al., 1996)).

Capítol 5

Capacitat generativa de les famílies de gramàtiques contextuals bàsiques

El test bàsic usual per tal d'avaluar la capacitat generativa de qualsevol nova classe de mecanismes generatius introduïts és la jerarquia de Chomsky. En aquest capítol compararem les famílies de llenguatges contextuals, les unes amb les altres i després, amb les famílies de la jerarquia de Chomsky. Els exemples del capítol previ ens seran molt útils per aquest objectiu, però, per tal de trobar contraexemples necessitarem també condicions per tal que un llenguatge pertanyi a una família donada (condicions necessàries). Començarem per buscar aquestes eines.

5.1 Condicions necessàries

De fet, part de les condicions considerades aquí són a la vegada necessàries i suficients perquè un llenguatge sigui d'un cert tipus segons l'anterior classificació dels llenguatges contextuals. Començarem per considerar algunes propietats específiques.

Definició 5.1.1 Direm que un llenguatge $L \subseteq V^*$ té la propietat de *pas de derivació externa acotat*, que denotarem per EBS (d'“external bounded step”), si existeix una constant p tal que per a cada $x \in L$, $|x| > p$, existeix $y \in L$ tal que $x = uyv$ i $0 < |uv| \leq p$.

Definició 5.1.2 Un llenguatge $L \subseteq V^*$ té la propietat de *pas de derivació interna acotat*, que denotarem per IBS (d'“internal bounded step”), si existeix una constant p tal que per a cada $x \in L$, $|x| > p$, existeix $y \in L$ tal que $x = x_1ux_2vx_3$, $y = x_1x_2x_3$ i $0 < |uv| \leq p$.

Definició 5.1.3 Un llenguatge $L \subseteq V^*$ té la propietat d' *increment de longitud acotat*, que denotarem per BLI (de “bounded length increase”), si existeix una constant p tal que per a cada $x \in L$ existeix $y \in L$ amb $-p \leq |x| - |y| \leq p$.

Clarament, si un llenguatge té la propietat d'EBS o bé d'IBS, aleshores té també la propietat de BLI, però l'invers no és cert: el llenguatge $\{a^n b^n c^n d^n e^n \mid n \geq 1\}$ té la propietat de BLI (amb la constant p igual a 5), però no podem esborrar cap context d'una cadena pertanyent a aquest llenguatge de manera

que la cadena resultant pertanyi encara a aquest llenguatge.

Observació 5.1.1 El context buit, $(u, v), uv = \lambda$, no és productiu per a gramàtiques contextuais del tipus considerat fins ara. D'aquí el fet que sigui ignorat. D'ara endavant, per convenció, suposarem sempre que tots els contextos usats són diferents de (λ, λ) .

Lema 5.1.1 Un llenguatge pertany a la família ECC si, i només si, té la propietat d'EBS.

Demostració. Per a una gramàtica contextual $G = (V, B, C, \varphi)$, denotem

$$p_1 = \max\{|x| \mid x \in B\}, \quad p_2 = \max\{|uv| \mid (u, v) \in C\}.$$

Aleshores $L_{ex}(G)$ té la propietat d'EBS per a $p = \max\{p_1, p_2\}$. (Si $z \in L_{ex}(G)$, $|z| > p$, aleshores $z \notin B$, ja que $z = uz'v$ per algun $(u, v) \in \varphi(z'), z' \in L_{ex}(G)$.)

I a la inversa, considerem un llenguatge $L \subseteq V^*$ que té la propietat d'EBS per a una constat p . Construïm la gramàtica $G = (V, B, C, \varphi)$ amb

$$\begin{aligned} B &= \{x \in L \mid |x| \leq p\}, \\ C &= \{(u, v) \mid u, v \in V^*, 0 < |uv| \leq p\}, \\ \varphi(x) &= \{(u, v) \in C \mid uxv \in L\}, x \in V^*. \end{aligned}$$

La igualtat $L = L_{ex}(G)$ se segueix fàcilment de la definició de B, C, φ . $\square$

Lema 5.1.2 Un llenguatge pertany a la família TC si, i només si, té la propietat d'IBS.

Demostració. Exactament com abans. Per exemple, si $L \subseteq V^*$ té la propietat d'IBS per a una constant p , aleshores podem construir la gramàtica $G = (V, B, C, \varphi)$ amb

$$B = \{x \in L \mid |x| \leq p\},$$

$$C = \{(u, v) \mid u, v \in V^*, 0 < |uv| \leq p\},$$

$$\varphi(x_1, x_2, x_3) = \{(u, v) \in C \mid x_1 u x_2 v x_3 \in L\}, \quad x_1, x_2, x_3 \in V^*.$$

La igualtat $L(G) = L$ és òbvia, i, per tant, $L \in TC$. □

Corol·lari 5.1.1 Es compleix que

- (i) Tots els llenguatges contextuais tenen la propietat de BLI.
- (ii) Cadascun dels llenguatges de la família ICC té la propietat d'IBS.

Conseqüències:

$$L_1 = \{a^{2^n} \mid n \geq 1\} \notin TC \quad (\text{per la propietat de BLI}),$$

$$L_2 = ab^+a \notin ECC \quad (\text{per la propietat d'EBS}),$$

$$L_3 = \{a^n b a^n b a^n \mid n \geq 1\} \notin TC \quad (\text{per la propietat d'IBS}),$$

$$L_4 = \{a^n b^n c^n d^n e^n \mid n \geq 1\} \notin TC \quad (\text{per la propietat d'IBS}).$$

Cal observar que tant L_3 com L_4 ténen la propietat de BLI, d'aquí que aquesta propietat no sigui suficient per tal que un llenguatge pertanyi a TC . El fet que $L_4 \notin TC$ pot ser expressat de la forma “les gramàtiques contextuais totals no poden comptar fins a cinc o més enllà”. Però, pel fet que $\{a^n b^n c^n d^n \mid$

$n \geq 1\} \in TC$ (aquest llenguatge té la propietat d'IBS per a $p = 4$), “les gramàtiques contextuais totals poden comptar fins a quatre” (com hem vist, totes les famílies bàsiques de gramàtiques contextuais estan incloses a TC i, per tant, cap gramàtica contextual pot “comptar fins a cinc”).

Definició 5.1.4 Direm que un llenguatge $L \subseteq V^*$, $\text{card}(V) \geq 2$, té la propietat de *mixtura*, que denotarem per *mix*, si existeix una constant p tal que si existeixen $a, b \in V$, $a \neq b$, i $x, y \in L$ amb $|x|_a > p, |y|_b > p$, aleshores per a cada $n \geq 1$ existeix $z_n \in L$ amb $|z_n|_a \geq n, |z_n|_b \geq n$.

Definició 5.1.5 Un llenguatge $L \subseteq V^*$ amb $\text{card}(V) \geq 2$ té la propietat de *Mixtura*, que denotarem per *Mix*, si existeix una constant p tal que si existeixen $a, b \in V$, $a \neq b$, i $x, y \in L$ amb $|x|_a > p, |y|_b > p$, aleshores per a cada $n \geq 1$ existeix $z_n \in L$ que pot ser escrita com a $z_n = x_1 x_2 \dots x_n$, per a $x_i = x_{i,1} a x_{i,2} b x_{i,3}$, $1 \leq i \leq n$ (podem trobar-hi cadenes amb un nombre arbitrari d'ocurrències del símbol a , seguides per ocurrències de b).

Clarament, la propietat de *Mix* implica la propietat de *mix*.

Lema 5.1.3 Tot llenguatge de la família EC té la propietat de *mix*; mentre que tot llenguatge de la família IC té la propietat de *Mix*.

Demostració. Per a una gramàtica contextual sense selecció $G = (V, B, C)$, denotem

$$p = \max\{|w| \mid w \in B\}.$$

Clarament, si existeixen $x, y \in L_\alpha(G)$, $\alpha \in \{ex, in\}$, tal que $|x|_a > p$, $|y|_b > p$, aleshores existeixen els contextos $(u, v), (u', v')$ pertanyents a C tal que $|uv|_a > 0$, $|u'v'|_b > 0$. Possiblement, $(u, v) = (u', v')$. Aquests contextos poden ser usats arbitràriament conjuntament i, per tant, podem produir cadenes que continguin un nombre arbitrari d'ocurrències d' a i b . En el mode intern, cada context pot ser usat en llocs arbitraris i, per tant, un nombre arbitrari d'ocurrències d' a i b mesclades poden ser produïdes. $\square$

Conseqüències:

$$L_1 = a^+ \cup b^+ \notin EC \cup IC,$$

$$L_2 = a^+b^+ \notin IC,$$

$$L_3 = \{a^n b^n \mid n \geq 1\} \notin IC.$$

Notem que aquests llenguatges tenen les propietats d'EBS i d'IBS (de $L_3 \notin IC$ inferim que les gramàtiques contextuais internes sense selecció “no poden comptar fins a dos”; de l'Exemple 4.2.3 de la pàgina 175 sabem que les gramàtiques contextuais externes sense selecció – i per tant, tampoc les que tenen selecció – “poden comptar fins a dos”).

Definició 5.1.6 Direm que un llenguatge $L \subseteq V^*$, $\text{card}(V) \geq 2$ té la propietat d'*ordenació*, que denotarem per *ord*, si existeix una constant p tal que si existeixen $a, b \in V$, $a \neq b$ i $x, y \in L$ amb $|x|_a > p$, $|y|_b > p$, aleshores existeix un nombre arbitrari de cadenes $z \in L$ tal que $z = z_1 a z_2 b z_3 = z'_1 b z'_2 a z'_3$, per algun $z_1, z_2, z_3, z'_1, z'_2, z'_3 \in V^*$ (existeixen cadenes on els símbols a, b apareixen en els dos ordres).

Òbviamment, tenim

Lema 5.1.4 Tots els llenguatges d' IC tenen la propietat d' ord .

Això prova novament que els llenguatges a^+b^+ i $\{a^n b^n \mid n \geq 1\}$ no pertanyen a IC .

Tal com hem vist a l'Exemple 4.2.6, per a cada llenguatge $L \subseteq V^*$, i $c \notin V$, el llenguatge $L' = V^* \cup L\{c\}$ no pertany a ECC . Això vol dir que els llenguatges pertanyents a ECC (i per tant també a TC) no tenen propietats de bombeig (si prenem $L = \{a^{2^n} \mid n \geq 1\}$, no podem bombejar cap subcadena d' $a^{2^n}c \in L$, sigui quin sigui el tipus de bombeig usual considerat). No obstant, tenim que

Lema 5.1.5 Si $L \subseteq V^*$, $L \in ICC$, aleshores existeixen dues constants p, q tal que tot $z \in L$, $|z| > p$, pot ser escrit de la forma $z = uvwxy$ amb $u, v, w, x, y \in V^*$, $0 < |vx| \leq q$ i $uv^iwx^iy \in L$ per a $i \geq 0$.

Demostració. Donada una gramàtica $G = (V, A, C, \varphi)$, definim

$$p = \max\{|w| \mid w \in B\},$$

$$q = \max\{|uv| \mid (u, v) \in C\}.$$

Si $x \in L_{in}(G)$ tal que $|x| > p$, aleshores $x \notin A$, ja que existeix $y \in L_{in}(G)$ tal que $y \Rightarrow_{in} x$, amb $y = x_1x_2x_3$, $x = x_1ux_2vx_3$. Això vol dir que $(u, v) \in \varphi(x_2)$, ja que tota derivació de la forma $x_1u^ix_2v^ix_3 \Rightarrow_{in} x_1u^{i+1}x_2v^{i+1}x_3$, $i \geq 0$ és correcta respecte a G . Totes aquestes cadenes pertanyen a $L_{in}(G)$. $\square$

Conseqüència: $L = \{a^n b^n c^n \mid n \geq 1\} \notin ICC$.

Aquesta conseqüència podria ser reescrita dient que “les gramàtiques contextuais internes – amb selecció arbitrària – no poden comptar fins a tres”, la qual cosa mostra que aquestes gramàtiques no tenen suficient poder per a les nostres necessitats d’abraçar específiques construccions no-lliures del context en llenguatge natural.

Òbviament, la propietat de bombeig del Lema 5.1.5 implica la propietat d’IBS (ja que implica la propietat de BLI), però l’invers no és cert: el llenguatge $L = a^+ \cup \{a^n b a^m b a^p \mid n + m + p = 2^k, k \geq 1\}$ té la propietat d’IBS, però cap cadena o parell de subcadenaes d’ $a^n b a^m b a^p$ pot ser bombejada.

Lema 5.1.6 Si $L \subseteq V^*$, $L \in EC$, aleshores existeixen dues constants p, q tal que tot $z \in L$, $|z| > p$, pot ser escrit de la forma $z = uyv$ amb $u, y, v \in V^*$, $0 < |uv| \leq q$, i $u^i y v^i \in L$ per a tot $i \geq 0$.

Demostració. És exactament la mateixa que en el cas del lema anterior, però en el cas específic d’una gramàtica contextual externa sense selecció. $\square$

L’existència d’una propietat de bombeig implica la següent propietat (que ja havíem apuntat al Capítol 3, com a conseqüència de la semilinealitat):

Definició 5.1.7 Un llenguatge $L \subseteq V^*$ té la propietat de *progressió aritmètica infinita*, que denotarem per *IPA* (d’“infinite arithmetical progression”), si el conjunt longitud d’ L conté una progressió aritmètica infinita.

Corol·lari 5.1.2 Tot llenguatge de $IC \cup EC$ té la propietat d'IAP.

Clarament, la propietat d'IAP implica també la propietat de BLI, però l'invers no és cert. Per exemple, considerem el llenguatge

$$L = \{a^n \mid n \in \bigcup_{r=0}^{\infty} T(r)\},$$

on

$$T(2k) = \{n \in \mathbb{N} \mid 2^{2k} \leq n \leq 2^{2k+1}, n \text{ parell}\},$$

$$T(2k+1) = \{n \in \mathbb{N} \mid 2^{2k+1} \leq n \leq 2^{2k+2}, n \text{ imparell}\}, \quad k \geq 0.$$

Pel fet que per a cada $x \in L$ existeix $y \in L$ tal que $|y| - |x| \in \{1, 2\}$, se segueix que L té la propietat d'increment de longitud acotat. Per altra banda, el conjunt $\text{length}(L)$ no conté una progressió aritmètica infinita. Per tal de demostrar-ho, suposem el contrari i prenem la progressió $\{p + s \cdot i \mid i \geq 0\} \subseteq \bigcup_{r=0}^{\infty} T(r)$, per algun $p \geq 0, s \geq 1$. Existeix k tal que

$$2^{2k+1} - 2^{2k} = 2^{2k} > 2s$$

i, per tant, existeixen dos termes de la forma $p + si, p + s(i+1)$ pertanyents a $T(2k)$. Això vol dir que s és parell, i, per tant, també p és parell. Similarment, existeix k tal que $T(2k+1)$ conté dos termes de la forma $p + si, p + s(i+1)$ de la progressió. Sent s parell, aleshores tenim que p és imparell; la qual cosa és una contradicció.

La propietat que segueix relativa als llenguatges contextuais externs amb selecció es relaciona amb la propietat d'EBS.

Definició 5.1.8 Per a un llenguatge $L \subseteq V^*$ denotem

$$\text{Min}_0(L) = \{x \in L \mid \text{PSub}(x) \cap L = \emptyset\},$$

$$\text{Min}_i(L) = \text{Min}_{i-1}(L) \cup \text{Min}_0(L - \text{Min}_{i-1}(L)), \quad i \geq 1.$$

Lema 5.1.7 Si $L \subseteq V^*, L \in \text{ECC}$, aleshores tots els conjunts $\text{Min}_i(L), i \geq 0$ són finits.

Demostració. Sigui $G = (V, B, C, \varphi)$ una gramàtica contextual tal que $L = L_{\text{ex}}(G)$. Considerem els conjunts

$$\begin{aligned} K_i(G) = \{ & u_j \dots u_1 w v_1 \dots v_j \mid 0 \leq j \leq i, w \in B, (u_1, v_1) \in \varphi(w), \\ & (u_k, v_k) \in \varphi(u_{k-1} \dots u_1 w v_1 \dots v_{k-1}), 2 \leq k \leq j \}, \end{aligned}$$

per a $i \geq 0$ (quan $i = 0$ tenim que $K_0(G) = B$).

Tots aquests conjunts són finits (B i C són finits). Podem provar, mitjançant inducció sobre i , que $\text{Min}_i(L) \subseteq K_i(G), i \geq 0$, la qual cosa implica que també $\text{Min}_i(L)$ és finit.

Per a $i = 0$ la inclusió és evident (òbviament tenim $\text{Min}_0(L) \subseteq B$). Suposem que és cert per a tot $i \leq j$ i considerem el cas $\text{Min}_{j+1}(L_{\text{ex}}(G)) \subseteq K_{j+1}(G)$.

Prenem $z \in \text{Min}_{j+1}(L_{\text{ex}}(G))$. Si $z \in \text{Min}_j(L_{\text{ex}}(G))$, aleshores tenim que $z \in K_j(G)$ per hipòtesi d'inducció i, pel fet que $K_j(G) \subseteq K_{j+1}(G)$, tenim que $z \in K_{j+1}(G)$. Suposem que $z \notin \text{Min}_j(L_{\text{ex}}(G))$, amb la qual cosa $z \in$

$Min_0(L_{ex}(G) - Min_j(L_{ex}(G)))$. Se segueix que $z \in L_{ex}(G) - Min_j(L_{ex}(G))$ i no hi ha cap subcadena d' x pertanyent a $L_{ex}(G) - Min_j(L_{ex}(G))$. Suposem que $z \notin K_{j+1}(G)$. Aleshores, existeix $k > j + 1$ tal que $z = u_k \dots u_1 x v_1 \dots v_k$, per a $x \in B$, $(u_1, v_1) \in \varphi(x)$, $(u_r, v_r) \in \varphi(u_{r-1} \dots u_1 x v_1 \dots v_{r-1})$, $r = 2, 3, \dots, k$ i no hi ha cap escriptura de z d'aquesta forma usant menys de $j + 2$ contextos. Totes les cadenes de la forma $u_s \dots u_1 x v_1 \dots v_s$, $0 \leq s \leq k$, pertanyen a $L_{ex}(G)$. Existeixen $k + 1 > j + 1$ subcadenaes d'aquest tipus les quals són comparables dues a dues en el sentit de la relació de subcadena. Conseqüentment, la cadena z té diferents subcadenaes pertanyents a més de $j + 2$ conjunts $Min_i(L_{ex}(G))$, la qual cosa vol dir que té com a mínim una subcadena d'aquest tipus que no pertany a $Min_j(L_{ex}(G))$. A més a més, existeixen subcadenaes de z pertanyents a $L_{ex}(G) - Min_j(L_{ex}(G))$, la qual cosa implica que z no és una cadena de $Min_0(L_{ex}(G) - Min_j(L_{ex}(G)))$. Això és una contradicció. Conseqüentment, la cadena z pertany a $K_{j+1}(G)$, la qual cosa completa l'argument d'inducció.

Conseqüències:

$$\begin{aligned} L_1 &= a^*ba^*ba^* \notin ECC && (\text{tenim que } Min_0(L) = ba^*b), \\ L_2 &= \{a^n b^n c^n \mid n \geq 1\} \notin ECC && (\text{tenim que } Min_0(L) = L). \end{aligned}$$

A més a més, les gramàtiques contextuais externes amb selecció arbitrària “no poden comptar fins a tres”.

Tot llenguatge construït a partir d'un vocabulari d'una-lletra té la propietat del Lema 5.1.7. Ja que no existeixen tals llenguatges en la família ECC, se segueix que aquesta condició no és suficient per tal que un llenguatge pertanyi

a *ECC*.

Presentarem ara una condició necessària d'un tipus diferent a les anteriors.

Definició 5.1.9 Un llenguatge $L \subseteq V^*$ es diu que és *k-magre*, per algun enter k , (Păun i Salomaa, 1995) si $\text{card}(L \cap V^n) \leq k$ per a tot $n \geq 0$. Un llenguatge és *magre* si és *k-magre* per algun k . Un llenguatge 1-magre s'anomena *prim*¹.

Lema 5.1.8 Si $L \in IC$ és un llenguatge infinit tal que $\text{alph}(L) \geq 2$, aleshores L és no-magre.

Demostració. Sigui $G = (V, B, C)$ una gramàtica amb $L_{in}(G) = L$ i $\text{card}(V) \geq 2$.

(1) Si $C \subseteq (a^*, a^*)$ per algun $a \in V$, aleshores prenem $(u, v) \in C$, amb $|uv| = a^i, i \geq 1$, i $w \in B, \text{alph}(w) \neq \{a\}$. Totes les cadenes

$$z_{n,m} = (uv)^m w (uv)^{n-m}, \quad 0 \leq m \leq n, n \geq 1$$

estan en L , $|z_{n,m}| = |w| + n|uv|$ i $z_{n,m} \neq z_{n,p}$, per $0 \leq m, p \leq n, m \neq p$. Per tant, L és no-magre.

(2) Si existeix $(u, v) \in (a^*, a^*)$ i $(u', v') \in (b^*, b^*)$, per $a \neq b$, aleshores considerem les cadenes de L

$$z_{n,m} = (uv)^{n-m|u'v'|} (u'v')^m |uv| w, \quad 0 \leq m \leq \left\lfloor \frac{n}{|u'v'|} \right\rfloor, n \geq 1$$

¹Els termes emprats en l'article original en anglès, “slender” i “thin”, s’han traduït al català, en aquest cas concret, com a “magre” i “prim”, respectivament.

per algun $w \in B$. Novament $|z_{n,m}| = |w| + n|uv|$, $z_{n,m} \neq z_{n,p}$ per a tot $m \neq p$, ja que L és no-magre.

(3) Si existeix $(u, v) \in C$ amb $u = a^i, v = b^j, i, j \geq 1$, i $a \neq b$, aleshores considerem les cadenes de L

$$z_{n,m} = (uv)^m u^{n-m} v^{n-m} w, \quad 0 \leq m \leq n, n \geq 1$$

per algun $w \in B$. Tenim que $|z_{n,m}| = |w| + n|uv|$ i $z_{n,m} \neq z_{n,p}, m \neq p$ ($z_{n,m} = (a^i b^j)^m a^{i(n-m)} b^{j(n-m)} w$), ja que L és no-magre.

(4) Si no es dona cap dels casos anteriors, aleshores existeix $(u, v) \in C$ amb un u, v contenint ocurrencies de dos símbols $a, b \in V, a \neq b$. Suposem que $v = v_1 a b^i, i \geq 1$. Quan u té aquesta propietat l'argument és el mateix. Considerem les cadenes de L

$$z_{n,m} = w(uv)^m u^{n-m} (v_1 a)^{n-m} (b^i)^{n-m}, \quad 0 \leq m \leq n, n \geq 1$$

per $w \in B$. Ja que $|z_{n,m}| = |w| + n|uv|$ i $z_{n,m} \neq z_{n,p}$ per $m \neq p$, se segueix que L és no-magre.

Aquests són tots els casos possibles. □

Conseqüències:

$$L_1 = a \cup b^+ \notin IC,$$

$$L_2 = \{a^n b^n \mid n \geq 1\} \notin IC.$$

Notem que totes les altres famílies de llenguatges contenen llenguatges magres (fins i tot primers): per exemple, L_2 pertany a la vegada a EC i a ICC (per a $G = (\{a, b\}, \{ab\}, \{\{\{ab\}, \{(a, b)\}\})\})$, tenim $L_2 = L_{in}(G)$).

Tancarem aquesta secció amb un contraexemple que no pot ser refutat per les condicions necessàries prèvies per tal que un llenguatge pertanyi a la família ICC .

Lema 5.1.9 El llenguatge $L = \{x \, mi(x) \mid x \in \{a, b\}^*\} \notin ICC$.

Demostració. Suposem que $L = L_{in}(G)$ per a una gramàtica contextual $G = (\{a, b\}, B, P)$. Pel fet que L és infinit, existeix $(u, v) \in \varphi(x)$ per algun $x \in Sub(L)$, amb $|uv| > 0$. Prenem aquest context, (u, v) .

Suposem que $|v|_a > 0$; el cas $|v|_b > 0$ és anàleg. Considerem la cadena

$$w = xb^k b^k \, mi(x)$$

per a $k > 2|uv|$. Ja que $w \in L = L_{in}(G)$ i $w \Rightarrow_{in}^* w' = uuxvvb^k b^k \, mi(x)$, tenim que $w' \in L_{in}(G)$. No obstant, $w' \notin L$: si $w' = y \, mi(y)$, aleshores $y = uuxvvb^{k-|uv|}$, $mi(y) = b^{|uv|} b^k \, mi(x)$, en aquest cas $uuxvvb^{k-|uv|} = xb^{k+|uv|}$, això és $v \in b^+$ (tenim que $k + |uv| \geq k - |uv| + 2|v|$). La qual cosa és una contradicció.

Si $v = \lambda$, suposem que $|u|_a > 0$; el cas $|u|_b > 0$ és anàleg. Considerem la cadena

$$w = mi(x) b^k b^k x,$$

per a $k > 2|u|$ i obtenim $w \Rightarrow_{in}^* w' = mi(x)b^k b^k uux \in L_{in}(G)$. No obstant, $w' \notin L$: com abans, prenem $u \in b^+$, la qual cosa és una contradicció. El llenguatge L no pot ser generat per una gramàtica contextual interna. $\square$

Es pot comprovar que aquest llenguatge té les propietats d'IBS i de *mix*, la propietat de bombeig del Lema 5.1.4 de la pàgina 189 i, a més a més, és no-magre. Per la qual cosa les condicions necessàries definides prèviament no són suficients per tal de provar que $L \notin ICC$.

5.2 Jerarquies de llenguatges contextuais

En aquesta secció intentarem clarificar les relacions entre les vuit famílies de llenguatges contextuais, mentre que en la secció següent examinarem la seva situació respecte a la jerarquia de Chomsky. Formularem els resultats en lemes i els agruparem en un teorema-diagrama final.

Les inclusions següents són òbvies, conseqüència de les definicions:

Lema 5.2.1 Es compleix que

$$(i) \quad IC \subseteq ICC_c \subseteq ICC,$$

$$(ii) \quad EC \subseteq ECC_c \subseteq ECC,$$

$$(iii) \quad TC_c \subseteq TC,$$

$$(iv) \quad ICC_c \subseteq TC_c,$$

$$(v) \quad ECC_c \subseteq TC_c,$$

$$(vi) \quad ICC \subseteq TC,$$

$$(viii) \quad ECC \subseteq TC.$$

Lema 5.2.2 Totes les inclusions del lema previ són pròpies.

Demostració. Moltes de les inclusions estrictes se segueixen dels exemples de la Secció 4.2 i de les condicions necessàries de la Secció 5.1. Per exemple, $a^+ \cup b^+$ no pertany a IC ni a EC (Lema 5.1.3 de la pàgina 187) però, clarament, sí pertany a ICC_c i a ECC_c .

A més a més, el llenguatge ab^+a no pertany a ECC (Lema 5.1.1 de la pàgina 185) i el llenguatge $\{a^n b^n c^n \mid n \geq 1\}$ no pertany a ICC (Lema 5.1.5 de la pàgina 189), però ambdós llenguatges pertanyen a TC_c . Aquest fet és obvi en el cas del primer llenguatge (de fet, es pot observar que $ab^+a \in ICC_c$) i, en el cas del segon llenguatge, ho podem provar mitjançant la següent gramàtica: $G = (\{a, b, c\}, \{abc\}, \{(a, bc)\}, \varphi), \varphi(a^n, b^n, c^n) = \{(a, bc)\}, n \geq 1$.

Hem obtingut la inclusió estricta de tots els casos, excepte aquells de la forma $X_c \subset X$, per a $X \in \{ICC, ECC, TC\}$, els quals poden ser provats de la següent manera: a partir de la tesi de Turing-Church (o mitjançant demostració directa) tenim que $X_c \subseteq RE$. Per altra banda, si en l'Exemple 4.2.6 de la pàgina 177 comencem per un llenguatge L que no pertany a RE , aleshores

$L_{ex}(G_L)$ és no-recursivament enumerable. Una afirmació similar és també certa en el cas de llenguatges generats internament: per a cada $L \subseteq V^*$ i $c, d \notin V$, construïm la gramàtica

$$\begin{aligned} G_L &= (V \cup \{c, d\}, \{cc\}, \{(\lambda, a) \mid a \in V\} \cup \{(d, d)\}, \varphi), \\ \varphi(cc) &= \{(\lambda, a) \mid a \in V\}, \\ \varphi(cx) &= \{(d, d)\}, x \in L. \end{aligned}$$

Obtenim

$$\begin{aligned} L_{in}(G_L) &= \{cc\}V^* \\ &\cup \{cd^m cx_1 dx_2 d \dots x_n dx_{n+1} \mid n \geq 1, x_1 \dots x_i \in L, 1 \leq i \leq n\}. \end{aligned}$$

A partir de cc podem generar totes les cadenes de $\{cc\}V^*$; després d'usar el context (d, d) , el context (λ, a) ja no pot ser aplicat.

Si L no pertany a una família de llenguatges F tancada sota derivada a la dreta i a l'esquerra (les famílies de la jerarquia de Chomsky, inclosa la família de llenguatges recursius, tenen aquestes propietats de clausura), aleshores $L_{in}(G_L) \notin F$: $L = \partial_{cdc}^e(\partial_d^d(L_{in}(G_L)))$. $\square$

Lema 5.2.3 Cada família IC, ICC_c, ICC és incomparable amb cada família EC, ECC_c, ECC .

Demostració. El llenguatge $a^*ba^*ba^*$ pertany a IC (és generat per $G = (\{a, b\}, \{bb\}, \{(\lambda, a)\})$), però no pertany a ECC (Lema 5.1.7 de la pàgina

192); mentre que $\{x \mid i(x) \mid x \in \{a, b\}^*\}$ pertany a EC (és generat per $G = (\{a, b\}, \{\lambda\}, \{(a, a), (b, b)\})$), però no pertany a ICC (Lema 5.1.9 de la pàgina 196). $\square$

5.3 Relacions amb la jerarquia de Chomsky

A partir de les definicions i del fet que totes les famílies de llenguatges contextuais contenen llenguatges infinits, tenim que

Lema 5.3.1 $FIN \subset IC \cap EC$.

Si denotem per LIN_1 la família de llenguatges generada per gramàtiques lineals amb només un símbol terminal (és a dir, $LIN_1 = \{L \in LIN \mid Var_{LIN}(L) = 1\}$), aleshores

Lema 5.3.2 $EC = LIN_1$.

Demostració. Si $G = (V, B, C)$ és una gramàtica contextual sense selecció, aleshores podem construir la gramàtica lineal

$$G' = (\{S\}, V, S, \{S \rightarrow uSv \mid (u, v) \in C\} \cup \{S \rightarrow w \mid w \in B\}).$$

La igualtat $L_{ex}(G) = L(G)$ és òbvia, ja que $EC \subseteq LIN_1$.

Inversament, si $G = (\{S\}, V, S, P)$ és una gramàtica lineal, podem construir la gramàtica contextual $G' = (V, B, C)$ amb

$$B = \{w \in V^* \mid S \rightarrow w \in P\},$$

$$C = \{(u, v) \mid u, v \in V^*, S \rightarrow uSv \in P\}.$$

Clarament, $L_{ex}(G') = L(G)$, en aquest cas $LIN_1 \subseteq EC$. □

El següent resultat és bastant sorprenent, i mostra que *reescriure* (en el cas de gramàtiques regulars) pot ser simulat mitjançant *adjunció* de contextos.

Lema 5.3.3 $REG \subset ICC_c$.

Demostració. Sigui L un llenguatge regular i sigui $M = (K, V, q_0, F, \delta)$ l'autòmat finit determinístic mínim que reconeix L .

Per a cada $w \in V^*$, definim la funció $\rho_w : K \longrightarrow K$ mitjançant

$$\rho_w(q) = q' \text{ si } \delta(q, w) = q', \quad q \in K.$$

Òbviament, si $x_1, x_2 \in V^*$ són tals que $\rho_{x_1} = \rho_{x_2}$, aleshores per a cada $u, v \in V^*$, ux_1v pertany a L si, i només si, ux_2v pertany a L .

El conjunt de funcions de K a K és finit, ja que el conjunt de funcions ρ_w com hem vist és finit. Sigui n_0 aquest nombre finit. Construïm la gramàtica

contextual $G = (V, B, C, \varphi)$ amb

$$\begin{aligned} B &= \{w \in L \mid |w| \leq n_0 - 1\}, \\ C &= \{(u, v) \mid 1 \leq |uv| \leq n_0\}, \\ \varphi(w) &= \begin{cases} \{(u, v) \mid |uvw| \leq n_0 \text{ i } \rho_w = \rho_{uvw}\}, & \text{per a } |w| \leq n_0 - 1, \\ \emptyset, & \text{en qualsevol altre cas.} \end{cases} \end{aligned}$$

A partir de la definició de les funcions ρ_w i la definició de B, C, φ , se segueix immediatament que $L_{in}(G) \subseteq L$.

Suposem que la inclusió inversa no és certa, i sigui $x \in L - L_{in}(G)$ una cadena de longitud mínima amb aquesta propietat. Aleshores $x \notin B$, en aquest cas $|x| \geq n_0$. Sigui $x = zz'$ amb $|z| = n_0$ i $z' \in V^*$. Si $z = a_1 a_2 \dots a_{n_0}$, aleshores té $n_0 + 1$ prefixos, que anomenarem $\lambda, a_1, a_1 a_2, \dots, a_1 \dots a_{n_0}$. Existeixen solament n_0 funcions diferents ρ_w . No obstant, existeixen dos prefixos u_1, u_2 de z tal que $u_1 \neq u_2$ i $\rho_{u_1} = \rho_{u_2}$. Sense perdre generalitat podem suposar que $|u_1| < |u_2|$. Substituint u_2 per u_1 obtenim una cadena x' que pertany també a L . Com sigui que $|x'| < |x|$ i x era de longitud mínima en $L - L_{in}(G)$, obtenim $x' \in L_{in}(G)$. No obstant, $|u_2| - |u_1| \leq |u_2| \leq n_0$, en aquest cas $u_2 = u_1 u_3$, aleshores $(\lambda, u_3) \in C$ i $(\lambda, u_3) \in \varphi(u_1)$. Això implica que $x' \Rightarrow_{in} x$, i per tant $x \in L_{in}(G)$, que és una contradicció. En conclusió, $L \subseteq L_{in}(G)$.

Les inclusions estrictes se segueixen de manera òbvia (ICC_c conté llenguatges no-lliures del context – veure els Exemples 4.2.4 de la pàg. 176, 4.2.8 de la pàg. 178 i 4.2.9 de la pàg. 179, de la secció 4.2). $\square$

Notem que en la demostració anterior només els contextos de la forma (λ, v) són necessaris, i que la funció φ té valors no buits només per a un nombre finit de cadenes. A més a més, de fet, hem provat que

Corol·lari 5.3.1 $REG \subseteq ICC(FIN)$.

Lema 5.3.4 $CF \subseteq TC_c$.

Demostració. Com a conseqüència dels lemes de bombeig (veure, p.e., el Teorema 3.8.4 de la pàgina 154, tot llenguatge lliure del context té la propietat d'IBS; d'acord al Lema 5.1.2 de la pàgina 185, aleshores pertany a TC . Ja que la relació de pertanyença és decidible per a llenguatges lliures del context, la funció φ definida com en la demostració del Lema 5.1.2 de la pàgina 185 és computable i, de fet en aquest cas, tot llenguatge lliure del context pertany a TC_c . La inclusió estricta se segueix dels exemples de la Secció 4.2 (p.e., l'Exemple 4.2.4 de la pàgina 176). $\square$

Lema 5.3.5 $IC \subseteq CS$.

Demostració. A partir d'una gramàtica contextual G , és fàcil de construir una gramàtica de longitud creixent G' tal que $L_{in}(G) = L(G)$. $\square$

Lema 5.3.6 REG és incomparable amb cadascuna de les famílies IC, EC, ECC_c, ECC . LIN i CF són incomparables amb cadascuna de les famílies $IC, ICC_c, ICC, ECC_c, ECC$. CS és incomparable amb cadascuna de les famílies ICC_c, ICC ,

ECC_c , ECC , TC_c , TC . RE és incomparable amb cadascuna de les famílies ICC , ECC , i TC .

Demostració. Hem vist que $REG - ECC \neq \emptyset$ (ab^+a és un llenguatge que prova aquesta relació) i que $LIN - ICC \neq \emptyset$ (el llenguatge del Lema 5.1.9 de la pàgina 196 és un exemple de llenguatge lineal que no pertany a ICC). Per altra banda, EC conté llenguatges no-regulars ($\{a^n b^n \mid n \geq 1\}$ n'és un exemple), IC conté llenguatges no-lliures del context (veure Exemple 4.2.4 de la pàgina 176). Començant la construcció en l'Exemple 4.2.6 de la pàgina 177 a partir d'un llenguatge recursiu no-lliure del context, obtenim una gramàtica contextual externa que genera un llenguatge que no pertany a CS ; de manera similar per a gramàtiques contextuales internes (veure la demostració del Lema 5.2.2 de la pàgina 198). De la mateixa manera podem provar que ICC i ECC contenen llenguatges que no són recursivament enumerables. $\square$

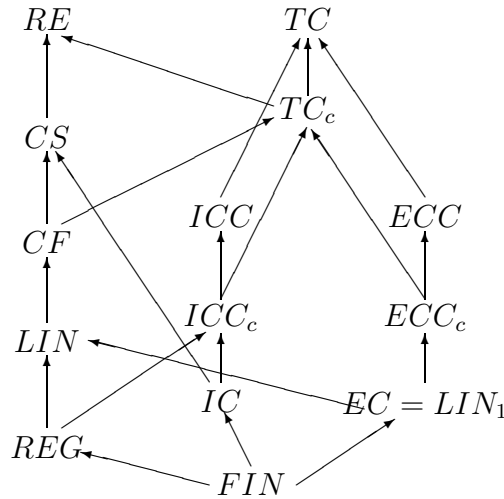
Sintetitzant els lemes previs tenim

Teorema 5.3.1 Les relacions que mostra el diagrama de la Figura 5.3.1 de la pàgina 205, on una fletxa d'una família F_1 a una altra família F_2 indica la inclusió estricta $F_1 \subset F_2$ i cada dues famílies no connectades són incomparables.

Ja hem apuntat al Capítol 3 que els llenguatges lliures del context i els llenguatges regulars definits sobre vocabularis d'una-lletra coincideixen. Els llenguatges d'una-lletra també tenen un comportament especial respecte a les famílies de llenguatges contextuales.

Per exemple, hem vist que ECC_c conté llenguatges no-regulars sobre $V = \{a\}$ (Exemple 4.2.5 de la pàgina 177). Pel fet que tot llenguatge regular d'un símbol té la propietat d'EBS, se segueix que cadascun d'aquests llenguatges pertany a la família ECC_c . Per altra banda, les restriccions de EC i IC a vocabularis d'una-lletra coincideixen, i existeixen llenguatges d'una-lletra regulars que no pertanyen a EC ni a IC : prenguem $L = \{a^2\} \cup \{a^{3n} \mid n \geq 1\}$. Si $L = L_{ex}(G) (= L_{in}(G))$ per algun $G = (\{a\}, B, C)$ aleshores, cada context $(u, v), |uv| > 0$, usat per a produir cadenes a^{3n} (existeix tal context, ja que L és infinit), ha de tenir $uv = a^{3i}, i \geq 1$. Afegint aquest context a a^2 obtenim la cadena paràsita $a^{3i+2}, i \geq 1$.

Figura 5.3.1 Relacions entre famílies de llenguatges contextuals i entre aquestes i les famílies de llenguatges de la jerarquia de Chomsky.



A més a més, tenim

Teorema 5.3.2 Un llenguatge d'una-lletra és regular si, i només si, pertany a la família ICC.

Demostració. Pel fet que $REG \subset ICC_c$ (Lema 5.3.3 de la pàgina 201), només hem de provar una de les implicacions.

Considerem una gramàtica contextual G presentada en forma modular, $G = (\{a\}, B, (S_1, C_1), \dots, (S_n, C_n))$, amb $S_i \subseteq a^*$, $1 \leq i \leq n$. Si a^{p_i} és la cadena més curta pertanyent a S_i , $1 \leq i \leq n$, aleshores els contextos de C_i poden ser aplicats a tota cadena a^m amb $m \geq p_i$. A més a més, a partir de cada S_i podem esborrar totes les cadenes diferents de a^{p_i} i, per tant, podem treballar amb $G' = (\{a\}, B, (\{a^{p_1}\}, C_1), \dots, (\{a^{p_n}\}, C_n))$ per les quals tenim $L_{in}(G) = L_{in}(G')$.

Denotem

$$p_0 = \max\{|x| \mid x \in B\},$$

$$p = \max\{p_i \mid 0 \leq i \leq n\},$$

$$q = \max\{|uv| \mid (u, v) \in C_i, 1 \leq i \leq n\},$$

$$M = L_{in}(G) \cap \{a^m \mid m \leq p\},$$

$$N = L_{in}(G) \cap \{a^m \mid p < n \leq p + q\}.$$

Construïm la gramàtica lineal per la dreta $G'' = (\{S, X\}, \{a\}, S, P)$, amb el conjunt P format per les següents regles:

$$\begin{aligned} S &\rightarrow x, \text{ per a } x \in M \cup N, \\ S &\rightarrow xX, \text{ per a } x \in N, \\ X &\rightarrow uvX, \text{ per a } (u, v) \in \bigcup_{i=1}^n C_i, \\ X &\rightarrow uv, \text{ per a } (u, v) \in \bigcup_{i=1}^n C_i. \end{aligned}$$

Obtenim $L(G'') = L_{in}(G')$. La inclusió $\subseteq$ és òbvia a partir de la construcció de la gramàtica G'' . Inversament, sigui $x = a^m$ una cadena pertanyent a $L_{in}(G)$. Si $m \leq p + q$, aleshores $S \rightarrow x \in P$ i, per tant, $x \in L(G'')$. Suposem que $m > p + q$; se segueix que $a^m \notin B$. Existeix una derivació de x a partir d'una cadena $y \in A$ i usant els contextos $(u_1, v_1), \dots, (u_k, v_k)$. Pel fet que $|u_i v_i| \leq q, 1 \leq i \leq k$, se segueix que existeix j tal que $p < |yu_1 v_1 \dots u_j v_j| \leq p + q$. La cadena $z = yu_1 v_1 \dots u_j v_j$ pertany a N i, per tant, la regla $S \rightarrow zX$ pertany a P . Les regles $X \rightarrow u_r v_r X, r = j + 1, \dots, k - 1, X \rightarrow u_k v_k$ també pertanyen a P . A més a més, podem trobar una derivació de la cadena x en la gramàtica G'' . Per tant, tenim la inclusió $L_{in}(G') \subseteq L(G'')$. $\square$

Notes bibliogràfiques. Els resultats bàsics relatius a la capacitat generativa de les gramàtiques contextuais externes poden ser trobats a (Păun, 1974b) i els relatius a gramàtiques contextuais internes a (Nguyen i Păun, 1980). Una síntesi d'ambdues apareix a (Păun, 1982a). El Lema 5.1.7 de la pàgina 192 és

de (Păun, 1974b). La inclusió $REG \subseteq ICC$ està demostrada a (Istrail, 1978b; Istrail, 1989; Vicolov, 1991); la demostració donada aquí (Lema 5.3.3 de la pàgina 201), més curta que les mencionades prèviament, és de (Ehrenfeucht et al., 1996b). Tot el material d'aquest capítol, incloent la síntesi del diagrama de la Figura 5.3.1 de la pàgina 205, pot ser trobat també a (Ehrenfeucht et al., 1996a), però sense tots els detalls de les demostracions. Una visió més extensa es pot trobar a (Păun, 1996a).

Capítol 6

Propietats de clausura de les famílies de gramàtiques contextuais bàsiques

Dins les famílies de llenguatges contextuais bàsics, les propietats de clausura de les famílies X_c són les mateixes que les de les famílies X , per a $X \in \{ICC, ECC, TC\}$ i, per tant, discutirem aquí només les famílies IC , EC , ICC , ECC , i TC . Exceptuant TC , aquestes famílies tenen propietats de clausura pobres (aquest és un resultat esperat, si ens fixem en el fet que les gramàtiques contextuais no usen símbols auxiliars o no-terminals en les seves derivacions).

Ens fixarem en algunes de les operacions definides a la Secció 3.2 del Capítol 3, i presentarem una successió de resultats, establerts en un seguit de lemes, els quals seran agrupats i sintetitzats en un teorema-taula final (justificarem els resultats negatius usant exemples i contraexemples de les seccions

prévies).

6.1 Operació d'unió

Si definim l'operació d'unió segons la Definició 3.2.1 de la pàgina 105, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada* sota l'operació d' α ”, per a $\alpha = \text{unió}$, aleshores

Lema 6.1.1 Les famílies ECC , TC són tancades sota l'operació d'unió.

Demostració. Considerem dos llenguatges arbitraris $L_1, L_2 \in ECC$. Aleshores, pel Lema 5.1.1 de la pàgina 185, aquests dos llenguatges tenen la propietat d' EBS . És fàcil de veure que també la seva unió $L_1 \cup L_2$ té també aquesta propietat. D'acord amb el Lema 5.1.1 de la pàgina 185, tenim que $L_1 \cup L_2 \in ECC$ i, per tant, tenim la clausura de la família ECC sota l'operació d'unió.

La clausura de TC sota unió és conseqüència de la seva clausura sota substitució (Lema 6.8.1 de la pàgina 221). El llenguatge $L_1 = \{a, b\}$ pertany a TC . Aplicant la substitució s definida per $s(a) = L$, $s(b) = L'$, on L, L' són llenguatges arbitraris pertanyents a TC , tenim que $s(L_1) = L \cup L'$, amb $L \cup L' \in TC$. Per tant, tenim la clausura de la família TC sota l'operació d'unió. $\square$

Lema 6.1.2 Les famílies IC, EC, ICC, TC són no-tancades sota l'operació d'unió.

Demostració. Els llenguatges a^+ i b^+ pertanyen ambdós a IC i a EC , però la seva unió, $a^+ \cup b^+$ no pertany ni a IC ni a EC (Lema 5.1.3 de la pàgina 187). Per tant, queda provada la no-clausura d'aquestes famílies sota l'operació d'unió.

Per a ICC considerem el llenguatge $L_1 = a^+, L_2 = \{a^n b^n \mid n \geq 1\}$. És fàcil de veure que ambdós llenguatges pertanyen a ICC , però la seva unió, $a^+ \cup \{a^n b^n \mid n \geq 1\}$, no pertany a ICC : la raó és que qualsevol context $(a^i, a^j) \in \varphi(a^k)$ – necessitem aquest context de manera que $i + j > 0, k \geq 0$, per tal de generar cadenes pertanyents a L_1 de longitud arbitrària – pot ser usat per a derivar $a^n b^n$, per a un valor d' n suficientment gran i d'aquesta manera, produir una cadena paràcita: $a^n b^n \Rightarrow_{in} a^{n+i+j} b^n \notin L_2 \in ICC$. $\square$

6.2 Operació d'intersecció amb conjunts regulars

Si definim l'operació d'intersecció segons la Definició 3.2.2 de la pàgina 105, per a les famílies de gramàtiques contextuais (IC, EC, ICC, ECC, TC) i segons la Definició 3.2.14 de la pàgina 110 de “família tancada sota l'operació α ”, per a $\alpha = \text{intersecció (amb conjunts regulars)}$, aleshores

Lema 6.2.1 Les famílies IC , EC , ICC , ECC , TC són no-tancades sota l'operació d'intersecció amb conjunts regulars.

Demostració. Ja que V^* pertany a cadascuna de les famílies IC , EC , ECC , però existeixen llenguatges regulars que no pertanyen a aquestes famílies, tenim demostrat el lema per a IC , EC , ECC .

Per tal de resoldre també el cas de la família ICC (la qual inclou REG – Lema 5.3.3 de la pàgina 201–, i no és tancada sota l'operació de morfismes; per la qual cosa no podem aplicar-li els arguments anteriors), considerem la gramàtica $G = (\{a, b, c\}, \{\lambda\}, \{(a, bc)\})$ i el conjunt regular

$$R = a^+b^+c^+.$$

Aleshores tenim que

$$L_{in}(G) \cap R = \{a^n b^n c^n \mid n \geq 1\},$$

un llenguatge que no pertany a ICC (pel Lema 5.1.5 de la pàgina 189), malgrat el fet que $L_{in}(G) \in IC$.

A més a més, la construcció de l'Exemple 4.2.4 de la pàgina 176, prova el següent resultat:

Observació 6.2.1 Tot llenguatge $L \subseteq V^*$ pot ésser escrit de la forma $L = h(L' \cap V^*)$, on $L' \in ECC \subseteq TC$ i h és el morfisme que esborra $c \notin V$ i deixa invariant els símbols de V .

La família TC és tancada sota morfismes arbitraris (i, per tant, també és tancada sota morfismes d'esborrament). Aleshores, la clausura d'aquesta família sota l'operació d'intersecció amb conjunts regulars implicaria que tot llenguatge fos un element de TC , la qual cosa no és certa (existeixen llenguatges – recursius – que no pertanyen a TC). En conseqüència, obtenim la no-clausura d'aquesta família sota intersecció amb conjunts regulars. $\square$

6.3 Operació d'intersecció

Si definim l'operació d'intersecció segons la Definició 3.2.2 de la pàgina 105, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada* sota l'operació d' α ”, per a $\alpha = \text{intersecció (arbitrària)}$, aleshores

Lema 6.3.1 Les famílies IC , EC , ICC , ECC , TC són no-tancades sota l'operació d'intersecció.

Demostració. Els llenguatges $L_1 = \{a, b\}^+ \{a\}$ i $L_2 = \{a\} \{a, b\}^+$ pertanyen a EC , però la seva intersecció $L_1 \cap L_2 = \{a\} \{a, b\}^* \{a\}$ no pertany a ECC , ja que $ab^*a \subseteq \text{Min}_0(L_1 \cap L_2)$ i, per tant, $\text{Min}_0(L_1 \cap L_2)$ és infinit (Lema 5.1.7 de la pàgina 192).

Per altra banda, considerem els llenguatges

$$L_3 = \{x \in \{a, b, c, d, e\}^* \mid |x|_a = |x|_b = |x|_c\},$$

$$L_4 = \{x \in \{a, b, c, d, e\}^* \mid |x|_c = |x|_d = |x|_e\}.$$

Ambdós llenguatges pertanyen a la família IC . Si afegim els contextos (λ, d) i (λ, e) a la gramàtica de l'Exemple 4.2.4 de la pàgina 176, obtenim una gramàtica G' tal que $L_{in}(G') = L_3$. Similarment per a L_4 .

No obstant, el llenguatge $L_3 \cap L_4 = \{x \in \{a, b, c, d, e\}^* \mid |x|_a = |x|_b = |x|_c = |x|_d = |x|_e\}$ no pertany a ICC (aquest llenguatge no té la propietat d' IBS (Lema 5.1.2 de la pàgina 185), ja que per a cadenes $a^n b^n c^n d^n e^n$, i per a un valor d' n suficientment gran, no podem suprimir dues subcadena no-buides acotades de tal manera que obtinguem una cadena pertanyent a $L_3 \cap L_4$). Per tant, tampoc pertany a TC .

La no-clausura de les famílies ICC, TC se segueix de la seva no-clausura sota l'operació d'intersecció amb conjunts regulars (Lema 6.2.1 anterior), ja que ambdues famílies contenen llenguatges regulars. $\square$

Corol·lari 6.3.1 Les famílies ICC i TC són no-tancades sota l'operació d'intersecció; i, a més a més, TC és no-tancada sota l'operació de complementarietat.

Demostració. Totes aquestes famílies contenen llenguatges regulars. A més a més, TC és tancada sota l'operació d'unió (Lema 6.1.1 de la pàgina 210); a

partir de les lleis de De Morgan, obtenim la no-clausura de TC sota l'operació de complementarietat. $\square$

6.4 Operació de complementarietat

Si definim l'operació de complementarietat segons la Definició 3.2.4 de la pàgina 106, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada* sota l'operació d' α ”, per a $\alpha = \text{complementarietat}$, aleshores

Lema 6.4.1 Les famílies IC , EC , ICC , ECC , TC són no-tancades sota l'operació de complementarietat.

Demostració. Per a les famílies ECC , TC l'assertió ve donada per la seva clausura sota l'operació d'unió (Lema 6.1.1 de la pàgina 210), la seva no-clausura sota l'operació d'intersecció (Lema 6.3.1 de la pàgina 213) i les lleis de De Morgan.

Per a la gramàtica $G = (\{a, b\}, \{ab, ba\}, \{(\lambda, a), (\lambda, b), (a, \lambda), (b, \lambda)\})$ tenim que

$$L_{in}(G) = L_{ex}(G) = \{a, b\}^* - (a^* \cup b^*).$$

Però, el complementari d'aquest llenguatge és $a^* \cup b^*$, un llenguatge que no pertany a $IC \cup EC$ (Lema 5.1.3 de la pàgina 187).

Per a la família ICC , considerem el llenguatge $\{a^n b^n \mid n \geq 1\}$. Aquest llenguatge pertany a ICC però el seu complementari, $\{a, b\}^* - L$, no. En efecte, si assumim que $\{a, b\}^* - L = L_{in}(G)$ per a alguna gramàtica $G = (\{a, b\}, A, C, \varphi)$, ja que $a^+ \in L_{in}(G)$, existeix un context (a^i, a^j) amb $i + j \geq 1$ tal que $(a^i, a^j) \in \varphi(a^k)$ per a algun $k \geq 0$. Aleshores podem derivar $a^{k+1}b^{k+i+j+1} \in \{a, b\}^* - L$ obtenint $a^{k+i+j+1}b^{k+i+j+1}$, que pertany a L , i no pertany a $\{a, b\}^* - L$. $\square$

Corol·lari 6.4.1 Les famílies ICC i TC són no-tancades sota l'operació d'intersecció; i, a més a més, TC és no-tancada sota l'operació de complementarietat.

Demostració. Totes aquestes famílies contenen llenguatges regulars. A més a més, TC és tancada sota unió (Lema 6.1.1 de la pàgina 210); a partir de les lleis de De Morgan obtenim la no-clausura de TC sota l'operació de complementarietat. $\square$

6.5 Operació de concatenació

Si definim l'operació de concatenació segons la Definició 3.2.5 de la pàgina 107, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada sota l'operació d' α* ”, per a $\alpha = \text{concatenació}$, aleshores

Lema 6.5.1 La família TC és tancada sota l'operació de concatenació.

Demostració. La clausura de TC sota l'operació de concatenació és conseqüència de la seva clausura sota l'operació de substitució (Lema 6.8.1 de la pàgina 221). El llenguatge $L_2 = \{ab\}$ pertany a TC . Aplicant la substitució s definida per $s(a) = L, s(b) = L'$, on L, L' són llenguatges arbitraris pertanyents a TC , tenim que $s(L_2) = LL'$ $\square$

Lema 6.5.2 Les famílies IC, EC, ICC, ECC són no-tancades sota l'operació de concatenació.

Demostració. Pel Lema 5.1.3 de la pàgina 187 sabem que $a^+b^+ \notin IC$. Pel Lema 5.1.2 de la pàgina 185 sabem que $ab^+a \notin ECC$. Per tant, aquestes famílies són no-tancades sota l'operació de concatenació (a^+, b^+ pertanyen a $IC \cap EC$, ab^+ i $\{a\}$ pertanyen a ECC).

Els llenguatges $L_1 = a^+b$, $L_2 = \{a^nba^n \mid n \geq 1\}$ pertanyen a ICC , però la seva concatenació no. Per tal de generar cadenes de la forma aba^nba^n , per a un valor d' n suficientment gran, necessitem un context de la forma (a^i, a^i) amb $i \geq 1$ tal que $(a^i, a^i) \in \varphi(a^kba^l)$ per a algun $k, l \geq 0$. Aquest context pot ésser usat per a derivar $a^mba^mba^m$, per a m suficientment gran (totes aquestes cadenes pertanyen a L_1L_2), produint una cadena $a^{m+i}ba^{m+i}ba^m$, que no pertany a L_1L_2 . $\square$

6.6 Operació de Kleene +

Si definim l'operació de Kleene + segons la Definició 3.2.10 de la pàgina 108, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família tancada sota l'operació d' α ”, per a $\alpha = \text{Kleene } +$, aleshores

Lema 6.6.1 La família TC és tancada sota l'operació de Kleene +.

Demostració. La clausura de TC sota l'operació de Kleene + és conseqüència de la seva clausura sota l'operació de substitució (Lema 6.8.1 de la pàgina 221). El llenguatge $L_3 = a^+$ pertany a TC . Aplicant la substitució s definida per $s(a) = L$, $s(b) = L'$, on L, L' són llenguatges arbitraris pertanyents a TC , tenim que $s(L_3) = L^+$. $\square$

Lema 6.6.2 Les famílies IC , EC , ICC , ECC són no-tancades sota l'operació de Kleene +.

Demostració. El llenguatge $L = \{a^n b^n \mid n \geq 1\}$ pertany a EC , però L^+ no pertany a $ECC - \text{Min}_0(L^+)$ és un conjunt infinit (Lema 5.1.7 de la pàgina 192).

Considerem ara la gramàtica contextual sense selecció $G = (\{a, b, c\}, \{c\}, \{(a, b)\})$. Les cadenes de $L_{in}(G)$ són de la forma xcy amb $xy \in D_{\{a, b\}}$. El llenguatge $L_{in}(G)$ és un subconjunt de $L_{in}(G)^+$. Per tal de generar cadenes de la

forma $a^n cb^n c$, per a un valor d' n arbitrari, necessitem un context $(a^i, b^i), i \geq 1$. Usant aquest context, a partir d'una cadena $acbacb \in L_{in}(G)^+$ podem produir $a^{i+1}cbacb^{i+1}$, que no pertany a $L_{in}(G)^+$ (ja que conté exactament dues ocurrencies de c i, per tant, hauríem d'escriure aquesta cadena com la concatenació de dues cadenes pertanyents a $L_{in}(G)$; la qual cosa és impossible). Per tant, aquest llenguatge no pertany a IC .

Finalment, considerem el llenguatge $L = \{a^n ba^n b \mid n \geq 1\}$. Aquest llenguatge pertany a ICC , però $L^+ \notin ICC$. Qualsevol gramàtica que generi L^+ ha de contenir un context $(a^i, a^i), i \geq 1$, de manera que $(a^i, a^i) \in \varphi(a^k ba^l)$ per a algun $k, l \geq 0$ (necessitem aquests contextos per a generar cadenes pertanyents a L de longitud arbitrària). Usant aquests contextos podem derivar $a^m ba^m ba^m ba^m$, per a m suficientment gran (totes aquestes cadenes pertanyen a L^+) obtenint $a^m ba^{m+i} ba^{m+i} ba^m$, que no pertany a L^+ . $\square$

6.7 Operació de morfismes (λ -lliures)

Si definim l'operació de morfismes segons la Definició 3.2.12 de la pàgina 109, per a les famílies de gramàtiques contextuais (IC, EC, ICC, ECC, TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada sota* l'operació d' α ”, per a $\alpha = \text{morfismes}$, aleshores

Lema 6.7.1 Les famílies IC, EC, ECC, TC són tancades sota l'operació de morfismes.

Demostració. La clausura de IC, EC, ECC, TC sota l'operació de morfismes és una conseqüència immediata de la seva clausura sota l'operació de substitució finita (Lema 6.9 de la pàgina 223).

La clausura de TC sota l'operació de morfismes és una conseqüència de la seva clausura sota substitucions arbitràries – Lema 6.8.1 de la pàgina 221 ($FIN \subseteq TC$). $\square$

Lema 6.7.2 La família ICC és no-tancada sota l'operació de morfismes.

Demostració. Considerem la gramàtica:

$$G = (\{a, b, c\}, \{abaca\}, \{(a, \lambda), (a, a)\}, \varphi),$$

amb $\varphi(c) = \{(a, a)\}$, $\varphi(b) = \{(a, \lambda)\}$. És fàcil de veure que

$$L_{in}(G) = \{a^n b a^m c a^m \mid n, m \geq 1\}.$$

Considerem també el morfisme $h : \{a, b, c\}^* \longrightarrow \{a, b\}^*$ definit per $h(a) = a$, $h(b) = b = h(c)$. Tenim que

$$h(L_{in}(G)) = \{a^n b a^m b a^m \mid n, m \geq 1\},$$

un llenguatge que hem vist, en la demostració del Lema 6.5.2 de la pàgina 217, no pertany a ICC . $\square$

6.8 Operació de substitució

Si definim l'operació de substitució segons la Definició 3.2.12 de la pàgina 109, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada sota* l'operació d' α ”, per a $\alpha = \text{substitució}$, aleshores

Lema 6.8.1 La família TC és tancada sota l'operació de substitució (arbitrària).

Demostració. Prenem $L \subseteq V^*$, $L \in TC$, i $s : V^* \longrightarrow 2^{U^*}$ una substitució tal que $s(a) \in TC$ per a cada $a \in V$. Assumim que $V = \{a_1, \dots, a_n\}$. D'acord amb el Lema 5.1.2 de la pàgina 185, tots el llenguatge L i tots els llenguatges $s(a_i)$, $1 \leq i \leq n$ tenen la propietat d' IBS . Siguin $p_0, p_1, \dots, p_n$ les constants associades, respectivament, i definim

$$p = p_0 \cdot \max\{p_i \mid 1 \leq i \leq n\}.$$

El llenguatge $s(L)$ té la propietat d' IBS , respecte a la constant p . En efecte, prenem una cadena $x \in s(L)$, $|x| > p$. Considerem $y \in L$ tal que $x \in s(y)$. Si existeix un símbol a_i que apareix en y el qual és substituït en x per una cadena z tal que $|z| > p_i$, aleshores $z = z_1uz_2vz_3$ per a $1 \leq |uv| \leq p_i$ i $z' = z_1z_2z_3 \in s(a_i)$. Se segueix que z pot ésser reemplaçada en x per z' i aquesta és exactament la propietat d' IBS : la cadena obtinguda pertany a $s(L)$ i $|uv| \leq p$. Si tots els símbols a_i de y són reemplaçats a x per cadenes

z_i amb $|z_i| \leq p_i$, aleshores $|y| > p_0$. Se segueix que $y = y_1 u y_2 v y_3$ per a $1 \leq |uv| \leq p_0$, $y' = y_1 y_2 y_3 \in L$. Siguin y'_1, u', y'_2, v', y'_3 les subcadenaes de x corresponents a y_1, u, y_2, v, y_3 en y . Aleshores $y'_1 y'_2 y'_3 \in s(y') \subseteq s(L)$ i $|u'v'| \leq |uv| \cdot \max\{p_i \mid 1 \leq i \leq n\} \leq p_0 \cdot \max\{p_i \mid 1 \leq i \leq n\} \leq p$. La propietat d'*IBS* es verifica altra vegada.

A més a més, d'acord amb el Lema 5.1.2 de la pàgina 185, el llenguatge $s(L)$ pertany a *TC*. $\square$

Lema 6.8.2 Les famílies *IC*, *EC*, *ICC*, *ECC* són no-tancades sota l'operació de substitució.

Demostració. En el cas de les famílies *EC*, *IC*, *ICC* la no clausura sota l'operació de substitució és una conseqüència de la no-clausura d'aquestes famílies sota l'operació d'unió (Lema 6.1.2 de la pàgina 210). Si prenem el llenguatge $L_0 = \{a, b\}$ i la substitució $s(a) = L_1, s(b) = L_2$, aleshores tenim que $L_1 \cup L_2 = s(L_0)$. En aquest cas, la clausura sota l'operació de substitució implicaria la clausura sota l'operació d'unió.

En el cas de la família *ECC*, la no-clausura sota l'operació de substitució és una conseqüència de la no-clausura de *ECC* sota concatenació (Lema 6.5.2 de la pàgina 217). Si prenem el llenguatge $L_0 = \{ab\}$ i la substitució $s(a) = L_1, s(b) = L_2$, aleshores tenim que $s(L_0) = L_1 L_2$. En aquest cas, la clausura sota l'operació de substitució implicaria la clausura sota l'operació de concatenació. $\square$

6.9 Operació de substitució finita

Si definim l'operació de substitució finita segons la Definició 3.2.12 de la pàgina 109, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada sota l'operació d' α* ”, per a $\alpha = \text{substitució (finita)}$, aleshores

Lema 6.9.1 Les famílies IC , EC , ECC , TC són tancades sota l'operació de substitució finita.

Demostració. Respecte a les famílies IC , EC , ECC , clarament, si $L \subseteq V^*$ té la propietat d' EBS (respecte a una constant p) i $s : V^* \rightarrow 2^{U^*}$ és una substitució finita, aleshores també $s(L)$ té la propietat d' EBS (respecte a la constant $p' = p \cdot \max\{|s(a)| \mid a \in V\}$). D'accord amb el Lema 5.1.1 de la pàgina 185, $s(L) \in ECC$.

Considerem ara una gramàtica contextual $G = (V, B, C)$ sense selecció i una substitució finita $s : V^* \rightarrow 2^{U^*}$. Construïm la gramàtica $G' = (U, s(B), s(C))$, on

$$s(C) = \{(u', v') \mid u' \in s(u), v' \in s(v), (u, v) \in C\}.$$

Aleshores és fàcil provar que $L_{ex}(G') = s(L_{ex}(G))$ i $L_{in}(G') = 0s(L_{in}(G))$.

La clausura de TC sota l'operació de substitució finita és una conseqüència de la seva clausura sota l'operació de substitucions arbitràries – Lema 6.8.1 de la pàgina 221 ($FIN \subseteq TC$). □

Lema 6.9.2 La família ICC és no-tancada sota l'operació de substitució finita.

Demostració. La no-clausura de la família ICC sota l'operació de substitució finita és una conseqüència de la no clausura de ICC sota l'operació de morfismes (Lema 6.7.2 de la pàgina 220). $\square$

6.10 Operació de morfismes inversos

Si definim l'operació de morfismes inversos segons la Definició 3.2.12 de la pàgina 109, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada sota* l'operació d' α ”, per a $\alpha = \text{morfismes inversos}$, aleshores

Lema 6.10.1 Les famílies IC , EC , ICC , ECC , TC són no-tancades sota l'operació de morfismes inversos.

Demostració. Considerem el llenguatge $L = \{c\}\{bc, ba\}^* \in EC$ i el morfisme $h : \{a, b, c\}^* \longrightarrow \{a, b, c\}^*$ definit per $h(a) = abc$, $h(b) = ab$, $h(c) = cb$. Tenim que $h^{-1}(L) = \{c\}\{b, c\}^*\{a\}$. La inclusió $\supseteq$ és òbvia, ja que $\{cb\}\{ab, cb\}^*\{abc\} \subseteq L$. Inversament, si $y \in h^{-1}(L)$ aleshores $h(y) \in L$. Per tant, y no pot contenir subcadena aa, ab, ac , ja que sinó $h(y)$ contindria subcadena $abcabc, abcab, abccb$ i, per tant, subcadena ca, cc ; la qual cosa és impossi-

ble. A més a més si y conté una ocurrència del símbol a , és a dir $|y|_a = 1$, aleshores aquest símbol a ha de ser l'últim símbol d' y . Cada cadena pertanyent a L té una longitud imparell. Si $|y|_a = 0$, aleshores $h(y)$ és de longitud parell, per la qual cosa no pot pertànyer a L . En conseqüència, y acaba amb el símbol a , pel que hem vist. Clarament, y ha de començar amb c , en aquest cas $y = cxa, x \in \{b, c\}^*$. Per tant, $\text{Min}_0(h^{-1}(L))$ és infinit i, en conseqüència, $h^{-1}(L) \notin ECC$ (Lema 5.1.7 de la pàgina 192).

Per a la gramàtica

$$G = (\{a, b, c, d\}, \{bacada\}, \{(b, cad)\}, \varphi),$$

$$\varphi(b^i ac^i) = \{(b, cad)\}, i \geq 1,$$

obtenim

$$L_{in}(G) = \{b^m ac^{m_1} (ad)^{n_1} c^{m_2} (ad)^{n_2} \dots c^{m_r} (ad)^{n_r} a \mid$$

$$r \geq 1, m \geq 1, 1 \leq m_r \leq n_r, n_i, m_i \geq 1, 1 \leq i \leq r-1,$$

$$\text{i } m = \sum_{i=1}^r n_i = \sum_{i=1}^r m_i\}.$$

Prenem també el morfisme $h : \{a, b, c, d\}^* \longrightarrow \{a, b, c, d\}^*$ definit per $h(a) = a, h(b) = b, h(c) = c, h(d) = da$. Les cadenes $h(x)$ no contenen la subcadena dc , que és $h^{-1}(y); y \in L_{in}(G)$ està només definida per a $y = b^m ac^m (ad)^m a$ (en aquest cas amb $r = 1$ en l'anterior escriptura de $L_{in}(G)$). A més a més, $h^{-1}(L_{in}(G)) = \{b^m ac^m ad^m \mid m \geq 1\}$, és un llenguatge que no pertany a TC (no té la propietat d' IBS ; Lema 5.1.2 de la pàgina 185).

Només ens queda parlar del cas de la família IC . Considerem la gramàtica $G = (\{a, b, c, d\}, \{dd\}, \{(\lambda, abc)\})$, així com el morfisme $h : \{a, b, c, d\}^* \longrightarrow \{a, b, c, d\}^*$ definit per $h(a) = abc, h(b) = da, h(c) = bca, h(d) = bcd$. Examinem les cadenes x , tal que $h(x) \in L_{in}(G)$. Les cadenes en $L_{in}(G)$ comencen per a o per d . Si $h(x)$ comença per a , aleshores comença per abc . Assumim que tenim $h(x) = (abc)^i z$ per a algun $i \geq 0$, i aquest és el prefix maximal de $h(x)$ d'aquesta forma. La cadena z no pot començar ni per b ni per c (aquests símbols han de tenir la corresponent ocurrència d' a en la seva banda esquerra). Les subcadenaes aa, bb, cc, ac, ba no són possibles a $h(x)$. Si z comença per a , aleshores novament hem de tenir abc , contradient la maximalitat d' i explicitada anteriorment. Per tant, z ha de començar per d (d'aquesta manera, per a $i = 0$, tenim en compte també el cas en què la cadena $h(x)$ comença amb una ocurrència de d) i continua amb a (la imatge de b). Suposem que tenim $j \geq 0$ cadenes bca , ja que $h(x) = (abc)^i da(bca)^j w$. Prenem el valor més gran de j amb aquesta propietat. La cadena w només pot començar per b i després de b només podem tenir c ; després de c no podem tenir ni a novament (això contradiu l'elecció de j) ni b (no tenim la corresponent ocurrència d' a en la seva banda dreta). Per tant, $h(x) = (abc)^i da(bca)^j bcd u$. L'única possibilitat és tenir $u = (ab)^k, k \geq 0$. En conclusió, $h^{-1}(L_{in}(G)) = \{a^i bc^j da^k \mid i, j, k \geq 0\}$. Aquest llenguatge no pertany a IC (qualsevol context introduint el símbol c pot ésser usat per a introduir el símbol c abans de l'ocurrència de b i, per tant, produeix una cadena paràsita). $\square$

6.11 Operació de barreja

Si definim l'operació de barreja segons la Definició 3.2.11 de la pàgina 108, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada sota* l'operació d' α ”, per a $\alpha = \text{barreja}$, aleshores

Lema 6.11.1 Les famílies IC, EC, ICC, ECC, TC són no-tancades sota l'operació barreja.

Demostració. Pels llenguatges $L_1 = ba^*$, $L_2 = a^*b$, ambdós pertanyents a EC , tenim que

$$L_1 \sqcup L_2 = \{a^n ba^m ba^p \mid n, m, p \geq 0, n + m \geq 1, m + p \geq 1, n + m + p \geq 2\}.$$

Aquest llenguatge té un conjunt infinit $\text{Min}_0(L_1 \sqcup L_2)$ i, per tant, no pertany a ECC (Lema 5.1.7 de la pàgina 192).

Considerem ara el llenguatge $L_1 = \{dd\}$ i la gramàtica $G = (\{a, b, c\}, \{abc\}, \{(a, bc)\})$. El llenguatge $L_1 \sqcup L_{in}(G)$ no pertany a TC (les cadenes de la forma $a^n db^n dc^n, n \geq 1$, pertanyen a aquest llenguatge i cap context pot ser esborrat de les cadenes de manera que la cadena obtinguda pertanyi al mateix llenguatge. Per tant, $L_1 \sqcup L_{in}(G)$ no té la propietat d'IBS). $\square$

6.12 Operació d'imatge mirall

Si definim l'operació d'imatge mirall segons la Definició 3.2.6 de la pàgina 107, per a les famílies de gramàtiques contextuais (IC , EC , ICC , ECC , TC) i segons la Definició 3.2.14 de la pàgina 110 de “família *tancada sota* l'operació d' α ”, per a $\alpha = \text{imatge mirall}$, aleshores

Lema 6.12.1 Les famílies IC , EC , ICC , ECC , TC són tancades sota l'operació d'imatge mirall.

Demostració. Òbvia (prenem una gramàtica i considerem la imatge mirall de tots els seus components, de la manera usual). $\square$

Resumint aquests resultats tenim

Taula 6.12.1 La taula mostra les propietats de clausura de les famílies bàsiques de gramàtiques contextuais, on S indica la clausura i N indica la no-clausura de la corresponent família de llenguatges contextuais sota la corresponent operació.

Operació	IC	EC	ICC	ECC	TC
Unió	N	N	N	S	S
Intersecció amb conjunts regulars	N	N	N	N	N
Intersecció	N	N	N	N	N
Complementarietat	N	N	N	N	N
Concatenació	N	N	N	N	S
Kleene +	N	N	N	N	S
Morfismes (λ -lliures)	S	S	N	S	S
Substitució	N	N	N	N	S
Substitució finita	S	S	N	S	S
Morfismes inversos	N	N	N	N	N
Barreja	N	N	N	N	N
Imatge mirall	S	S	S	S	S

Segons les definicions 3.6.6 donades a la pàgina 145, podem concloure que ICC és un anti-AFL, però les altres famílies no són ni AFL (cap d'elles és tancada sota intersecció amb conjunts regulars), ni anti-AFL (totes són tancades sota morfismes arbitraris).

6.13 Altres operacions de clausura: operació de prolongació a la dreta

A la vista d'aquestes propietats de clausura tan pobres respecte a les “clàssiques” operacions amb llenguatges, seria d'interès imaginar noves operacions específiques, algunes d'elles relacionades amb la manera de treballar de les gramàtiques contextuais, amb l'esperança d'obtenir propietats de clausura més productives. Una possibilitat podria ser la següent operació:

Definició 6.13.1 L'operació de *prolongació a la dreta* (també anomenada “operació dòmino”) d'un llenguatge $L \subseteq V^*$ respecte d'un altre llenguatge $L_2 \subseteq V^*$, denotada per $pdret(L_1, L_2)$, és el llenguatge més petit $L \subseteq V^*$ que compleix les següents propietats:

- (i) $L_1 \subseteq L$,
- (ii) si $x \in L, x = x_1x_2, x_2 \in V^+, i x_2a \in L_2, a \in V$, aleshores $xa \in L$.

Teorema 6.13.1 Les famílies IC, EC, ICC no són tancades sota l'operació de prolongació a la dreta, però ECC i TC sí que ho són.

Demostració. La clausura de ECC i TC sota l'operació de prolongació a la dreta es deriva dels Lemes 5.1.1 de la pàgina 185 i 5.1.2 de la pàgina 185: si L_1 té la propietat d'*EBS* o d'*IBS*, independentment del tipus de L_2 , el llenguatge $pdret(L_1, L_2)$ té la mateixa propietat.

Pel fet que $pdret(\{a, b\}, \{aa, bb\}) = a^+ \cup b^+$, se segueix que IC i EC no estan tancades sota aquesta operació.

Considerem ara la gramàtica $G = (\{a, b\}, \{babab\}, \{(\{aba\}, \{(a, a)\})\})$.
Tenim

$pdret(L_{in}(G), \{aa, ba\}) = \{ba^n ba^n ba^m \mid n \geq 1, m \geq 0\}$. Aquest llenguatge no pertany a la família ICC : per tal de generar cadenes $ba^n ba^n b$ amb un nombre arbitràriament llarg de n 's necessitem $(a^i, a^i) \in \varphi(a^j ba^k)$ per a $i \geq 1, j, k \geq 0$; aleshores $ba^{j+1} ba^{j+1} ba^k \Rightarrow_{in} ba^{j+1} ba^{j+i+1} ba^{k+i}$, que no pertany al nostre llenguatge. La qual cosa és una contradicció. $\square$

Un problema natural que apareix en aquest marc de treball és investigar la família més petita que conté una família de llenguatges contextuais donada i estigui tancada sota certes operacions (quin és, per exemple, el AFL més petit que conté ICC ?).

Notes bibliogràfiques. Tots els resultats d'aquest capítol apareixen a (Păun, 1982a) i, en general, estan basats en (Păun, 1974b; Păun, 1975b) i (Nguyen i Păun, 1980). Algunes parts de les demostracions presentades aquí, però, apareixen de manera diferent. També es poden consultar els capítols corresponents al respecte a (Ehrenfeucht et al., 1996a) o bé a (Păun, 1996a) (on les demostracions, però, no són tan detallades).

Capítol 7

Propietats de decidibilitat de les famílies de gramàtiques contextuais bàsiques

Considerarem els problemes de decidibilitat bàsics en teoria de llenguatges: buidor, finitud, equivalència, inclusió i pertanyença, amb les definicions usuals (veure la Secció 3.7 de la pàgina 150). Els resultats, establerts en una sèrie de lemes, seran agrupats en un teorema-taula final.

Abans, però, veurem un lema previ que usarem en la demostració d'altres lemes.

Lema 7.0.1 No podem decidir quan la intersecció de dos llenguatges d' F , sent F alguna de les famílies de llenguatges contextuais, diferent de IC , és buida, finita, o infinita.

És clar que, “la intersecció de dos llenguatges d’ F ” significa “la intersecció de dos llenguatges generats per dues gramàtiques arbitràriament donades corresponents a la família F ” (usarem aquesta curta formulació més endavant).

Demostració. Normalment, per tal d’obtenir resultats de no-decidibilitat, reduïm els problemes considerats al denominat *problema de correspondència de Post* (abreujadament, *PCP*). Més exactament, per a dues n -tuples de cadenes no-buides sobre $V = \{a, b\}$, $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, considerem els llenguatges

$$L(z) = \{ba^{i_k} \dots ba^{i_1} cz_{i_1} \dots z_{i_k} \mid k \geq 1, 1 \leq i_j \leq n, 1 \leq j \leq k\},$$

per a $z \in \{x, y\}$. Aleshores $L(x) \cap L(y) \neq \emptyset$ si, i només si, el problema de correspondència de Post per a (x, y) , abreujadament $PCP(x, y)$, té una solució; la qual cosa, com sabem, és indecidible. A més a més, $L(x) \cap L(y)$ és infinit si, i només si, és no-buit.

Els llenguatges $L(x), L(y)$ estan en $EC \cap ICC$. Per a

$$G(z) = (\{a, b, c\}, \{ba^i cz_i \mid 1 \leq i \leq n\}, \{(ba^i, z_i) \mid 1 \leq i \leq n\}),$$

$$G'(z) = (\{a, b, c\}, \{ba^i cz_i \mid 1 \leq i \leq n\}, \{(\{c\}, \{(ba^i, z_i) \mid 1 \leq i \leq n\})\}),$$

tenim que $L_{ex}(G(z)) = L_{in}(G'(z)) = L(z)$, $z \in \{x, y\}$. □

7.1 Propietat d'equivalència

Lema 7.1.1 El problema de l'equivalència és indecidible per a famílies ICC_c , ICC , ECC_c , ECC , TC_c , TC .

Demostració. Per a dues n -tuples x, y de cadenes no-buides sobre $\{a, b\}$, construïm la gramàtica

$$\begin{aligned} G &= (\{a, b, c\}, \{ba^i cy_i \mid 1 \leq i \leq n\}, \{(ba^i, y_i) \mid 1 \leq i \leq n\}, \varphi), \\ \varphi(z) &= \{(ba^i, y_i) \mid ba^i zy_i \notin L(x)\}, z \in L(y). \end{aligned}$$

Clarament, $L_{ex}(G) = L(y)$ si, i només si, $L(y) \cap L(x) = \emptyset$, la qual cosa és indecidible com hem vist en el Lema 7.0.1 anterior. La funció φ és computable.

Considerem ara les gramàtiques

$$\begin{aligned} G'(y) &= (\{a, b, c, d\}, \{dcd\}, \{(ba^i, y_i) \mid 1 \leq i \leq n\}, \varphi'), \\ \varphi'(c) &= \{(ba^i, y_i) \mid 1 \leq i \leq n\}, \\ G''(y) &= (\{a, b, c, d\}, \{dcd\}, \{(ba^i, y_i) \mid 1 \leq i \leq n\} \cup \{(d, d)\}, \varphi''), \\ \varphi''(c) &= \{(ba^i, y_i) \mid 1 \leq i \leq n\}, \\ \varphi''(dzd) &= \{(d, d)\}, z \in L(x). \end{aligned}$$

Obtenim $L_{in}(G'(y)) = \{d\}L(y)\{d\}$ mentre que $L_{in}(G''(y)) = L_{in}(G'(y))$ si, i només si, el context (d, d) no pot ésser mai usat, i.e., si, i només si, $L(x) \cap L(y) = \emptyset$; que és indecidible. Les funcions φ', φ'' són computables. $\square$

7.2 Propietat d'inclusió

Corol·lari 7.2.1 El problema de la inclusió és indecidible per a famílies ICC_c , ICC , ECC_c , ECC , TC_c , TC .

Demostració. La demostració d'aquest corol·lari és conseqüència del Lema 7.1.1 anterior. $\square$

7.3 Propietat de pertanyença

Lema 7.3.1 El problema de la pertanyença és decidible per a gramàtiques corresponents a les famílies IC , ICC_c , EC , ECC_c , TC_c .

Demostració. És suficient provar l'afirmació per a gramàtiques contextuais totals. Prenem $G = (V, B, C, \varphi)$, amb una funció $\varphi : V^* \times V^* \times V^* \longrightarrow 2^C$ computable, i construïm els conjunts:

$$H_0(G) = B,$$

$$H_i(G) = H_{i-1}(G) \cup \{x \in V^* \mid w \Longrightarrow x \text{ per a algun } w \in H_{i-1}(G)\}, i \geq 1.$$

Tenim que $x \in L(G)$ si, i només si, $x \in H_{|x|}(G)$. De fet, suposant que $(\lambda, \lambda) \notin C$, cada derivació per a x , si n'hi ha alguna, ha de tenir com a màxim $|x|$ passos.

Els conjunts $H_i(G)$ poden ser algorísmicament construïts i són finits; conseqüentment, podem decidir si $x \in H_{|x|}(G)$ o no. $\square$

Lema 7.3.2 El problema de la pertanyença és indecidible per a gramàtiques amb funció de selecció arbitrària (ICC, ECC, TC).

Demostració. Les famílies ICC, ECC, TC contenen llenguatges que no són recursivament enumerables (Lema 5.3.6 de la pàgina 203). $\square$

7.4 Propietat de buidor

Lema 7.4.1 El problema de la buidor és decidible per a totes les classes de gramàtiques contextuais.

Demostració. Si $G = (V, B, C, \varphi)$ és una gramàtica contextual total, aleshores $L(G) \neq \emptyset$ si, i només si, $B \neq \emptyset$. $\square$

7.5 Propietat de finitud

Lema 7.5.1 La finitud és decidible per a gramàtiques corresponents a famílies IC, ICC_c, EC .

Demostració. Si $G = (V, B, C)$ és una gramàtica sense selecció, aleshores $L_{in}(G)$ i $L_{ex}(G)$ són infinits si, i només si, $B \neq \emptyset$ i $C \neq \emptyset$ (vam suposar que el context buit, (λ, λ) , no és present).

Si $G = (V, B, C, \varphi)$, amb una funció $\varphi : V^* \longrightarrow 2^C$ computable, aleshores $L_{in}(G)$ és infinit si, i només si, existeix $(u, v) \in C, (u, v) \in \varphi(x)$ per algun $x \in Sub(B)$ (demostració del Lema 5.1.3 de la pàgina 187). Ja que $Sub(B)$ és finit (recordem que $Sub(B)$ és el conjunt de subcadena obtingut a partir de cadenes de B), l'existència d'aquest context és decidible (és essencial en aquesta demostració suposar que el context buit (λ, λ) no és present). $\square$

Lema 7.5.2 El problema de la finitud és indecidible per a famílies $ICC, ECC_c, ECC, TC_c, TC$.

Demostració. Prenem dues n -tuples de cadenes no-buides sobre $\{a, b\}$ i considerem la gramàtica $G(y)$ com abans, tal que $L_{ex}(G(y)) = L(y)$. Construïm els conjunts $K_i(G(y))$ com en la demostració del Lema 5.1.7 de la pàgina 192, $K_i(G(y)) = \{w \in L_{ex}(G(y)) \mid \text{existeix una derivació de } w \text{ en la gramàtica } G(y) \text{ usant com a màxim } i \text{ contextos}\}, i \geq 1$.

Aquests conjunts són finits i poden ser efectivament construïts. A més a més, el llenguatge $L(y)$ és lineal, i, per tant, té un problema de pertanyença resoluble. Clarament, tenim que

$$L_{ex}(G(y)) = \bigcup_{i=0}^{\infty} K_i(G(y)).$$

Construïm la gramàtica contextual

$$G = (\{a, b, c\}, \{ba^i cy_i \mid 1 \leq i \leq n\}, \{(ba^i, y_i) \mid 1 \leq i \leq n\}, \varphi),$$

$$\varphi(w) = \{(ba^i, y_i) \mid 1 \leq i \leq n\}, \text{ si existeix } r \geq 0 \text{ tal que}$$

$$w \in K_r(G(y)) \text{ i } K_r(G(y)) \cap L(x) = \emptyset.$$

Si $K_r(G(y)) \cap L(x) \neq \emptyset$, aleshores $K_p(G(y)) \cap L(x) \neq \emptyset$ per a tot $p \geq r$ (tenim que $K_r(G(y)) \subseteq K_p(G(y))$ per a $r \leq p$). Per tant, si existeix r tal que $K_r(G(y)) \cap L(x) \neq \emptyset$, aleshores $L_{ex}(G)$ és finit. Inversament, $L_{ex}(G) = L(y)$ quan $(\cup_{i=0}^{\infty} K_r(G(y))) \cap L(x) = \emptyset$. Conseqüentment, $L_{ex}(G)$ és infinit si, i només si, $L(y) \cap L(x) = \emptyset$, la qual cosa és indecidible, com hem vist en el Lema 7.0.1 de la pàgina 233.

Notem que la funció considerada anteriorment φ és computable.

Per a gramàtiques $G = (V, B, C, \varphi)$ amb no-computable φ no podem decidir si $L_{in}(G)$ és o no és infinit sense decidir abans si $\varphi(x) \neq \emptyset$ o no, per a com a mínim una cadena $x \in Sub(B)$: prenem $G_{x,y} = (\{a\}, \{\lambda\}, \{(\lambda, a)\}, \varphi)$, amb $\varphi(\lambda) = \{(\lambda, a)\}$ si, i només si, el problema de correspondència de Post per a (x, y) (abreujadament, $PCP(x, y)$) té una solució. En qualsevol altre cas, $\varphi(z) = \emptyset$ per a tot z . $\square$

Resumint aquests resultats, tenim

Taula 7.5.1 La taula mostra les propietats de decidibilitat de les famílies bàsiques de gramàtiques contextuais, on D indica que el problema és decidible,

I indica que el problema és indecidible, mentre que els signe d'interrogació ? indica que el problema resta obert (per a la corresponent família de llenguatges contextuais relatius al corresponent problema).

Problema	IC	ICC_c	ICC	EC	ECC_c	ECC	TC_c	TC
Buidor	D	D	D	D	D	D	D	D
Finitud	D	D	I	D	I	I	I	I
Equivalència	?	I	I	?	I	I	I	I
Inclusió	?	I	I	?	I	I	I	I
Pertanyença	D	D	I	D	D	I	D	I

7.6 Altres propietats de decidibilitat: l'ambigüitat

Algunes altres qüestions de decidibilitat poden ser formulades sobre gramàtiques contextuais.

Lema 7.6.1 El problema de l'ambigüitat és indecidible per a totes les famílies $ICC_c, ICC, EC, ECC_c, ECC, TC_c, TC$.

Demostració. A (Greibach, 1963) està demostrat que l'ambigüitat de les gramàtiques lineals mínimes $G = (\{S\}, \{a, b, c\}, S, P)$, amb

$$P = \{S \rightarrow ba^i Sx_i \mid 1 \leq i \leq n\} \cup \{S \rightarrow ba^i Sy_i \mid 1 \leq i \leq n\} \cup \{S \rightarrow c\},$$

és indecidible, on x, y són dues n -tuples de cadenes no-buides sobre $\{a, b\}$. Construïm les gramàtiques

$$G' = (\{a, b, c\}, \{c\}\{(ba^i, x_i), (ba^i, y_i) \mid 1 \leq i \leq n\}),$$

$$G'' = (\{a, b, c\}, \{c\}, \{(\{c\}, \{(ba^i, x_i), (ba^i, y_i) \mid 1 \leq i \leq n\})\}).$$

Tenim que $L(G) = L_{ex}(G') = L_{in}(G'')$ i cada derivació en G es correspon amb una derivació externa en G' i amb una derivació interna en G'' ; i inversament. Per tant, l'ambigüitat de G', G'' també és indecidible. $\square$

Tots els problemes que siguin indecidibles per a gramàtiques lliures del context són també indecidibles per a gramàtiques contextuais totals, ja que cada gramàtica lliure del context pot ser simulada mitjançant una gramàtica contextual total amb funció de selecció computable (Lema 5.3.4 de la pàgina 203). D'aquesta manera obtenim la no-decidibilitat de l'equivalència, la inclusió, i l'ambigüitat, però també de la regularitat.

A més a més, per a una gramàtica $G = (V, B, C, \varphi)$, és també indecidible quan $L_{ex}(G) \in EC$ o no. Considerem la gramàtica

$$G = (\{a, b, c\}, \{c\}, \{((L(x) - L(y)) \cup \{c\}, \{(ba^i, x_i) \mid 1 \leq i \leq n\}),$$

$$(L(x) \cap L(y), \{(ba^i, x_i) \mid 1 \leq i \leq n\} \cup \{(c, c)\})\}).$$

Aleshores $L_{ex}(G) = L(x) \in EC$ si, i només si, el context (c, c) no és usat mai, en aquest cas $L(x) \cap L(y) = \emptyset$, la qual cosa és indecidible (Lema 7.0.1 de la pàgina 233). En efecte, si el context (c, c) és usat, la derivació no pot continuar, la cadena obtinguda és de la forma czc per a $z \in L(x) \cap L(y)$. La cadena z

pot ser arbitràriament llarga (si $L(x) \cap L(y) \neq \emptyset$, aleshores $L(x) \cap L(y)$ és infinita), i, per tant, necessitem un context per a introduir el símbol c . En una gramàtica sense selecció, aquest context pot ser emprat un nombre arbitrari de vegades, i, per tant, es produeixen cadenes que no pertanyen a $L_{ex}(G)$. Conseqüentment, $L_{ex}(G) \notin EC$, la qual cosa és indecidible.

Notes bibliogràfiques. A l'igual que en el Capítol 6, tots els resultats d'aquest capítol apareixen a (Păun, 1982a) i, en general, estan basats en (Păun, 1974b; Păun, 1975b) i (Nguyen i Păun, 1980). Algunes parts de les demostracions presentades aquí, però, apareixen de manera diferent. També es poden consultar els capítols corresponents al respecte a (Ehrenfeucht et al., 1996a) o bé a (Păun, 1996a) (on les demostracions, però, no són tan detallades).

Capítol 8

Altres propietats rellevants de les famílies de gramàtiques contextuais bàsiques

8.1 La semilinealitat i propietats relacionades

Ja hem esmentat que si un llenguatge L és semilineal, aleshores L té la propietat de progressió aritmètica infinita (IAP) (veure Teoremes 3.8.2 i 3.8.3 i Corol·lari 3.8.1 de la pàgina 153, i la Definició 5.1.7 de la pàgina 190), la qual cosa, al mateix temps, implica que L té la propietat d'increment de longitud acotat (BLI). La implicació inversa no és certa. Ho hem vist a la Secció 5.1 per les propietats de BLI i d'IAP. Fàcilment, podem provar-ho per a la propietat d'IAP “versus” semilinealitat: pel llenguatge

$$L = \{a^{2^n}b^m \mid n, m \geq 0\}$$

tenim $length(L) = \mathbb{N} - \{0\}$ (on el conjunt longitud, $length(L) = \{|x| \mid x \in L\}$ i, per tant, té la propietat d'IAP), però

$$\Psi_{\{a,b\}}(L) = \{(2^n, m) \mid n, m \geq 0\},$$

no és un conjunt semilineal (veure la Definició 3.8.1 de la pàgina 153).

També hem vist a la Secció 5.1 que els llenguatges contextuais de tots els tipus considerats fins ara tenen la propietat de BLI, mentre que els llenguatges de $ICC \cup EC$ tenen també la propietat d'IAP. La igualtat $EC = LIN_1$ (Lema 5.3.2 de la pàgina 200) implica que tots els llenguatges de EC són semilineals. Això també és cert per a IC (cal recordar que aquesta família conté llenguatges no lliures del context i, per tant, en aquest cas, la semilinealitat no és una propietat trivial):

Teorema 8.1.1 La família IC conté només llenguatges semilineals.

Demostració. Si $G = (V, C, C)$ és una gramàtica sense selecció, aleshores $L_{in}(G)$ és lletra-equivalent (veure la Definició 3.8.1 de la pàgina 153) amb el llenguatge generat per la gramàtica lineal a la dreta

$$G' = (\{X\}, V, X, \{X \rightarrow uvX \mid (u, v) \in C\} \cup \{X \rightarrow w \mid w \in B\}),$$

i, per tant, és semilineal. □

No obstant, a partir de l'Exemple 4.2.5 de la pàgina 177 tenim

Teorema 8.1.2 La família ECC_c conté llenguatges no-semilineals.

La família de conjunts semilineals (de vectors) és tancada sota l'operació d'intersecció, veure, p.e., (Ginsburg, 1966), però la família de tots els llenguatges semilineals no és tancada sota l'operació d'intersecció; de fet, aquesta família és no tancada sota l'operació d'intersecció amb llenguatges regulars. Per exemple, considerem el llenguatge

$$L = \{a^{2^n}b \mid n \geq 1\} \cup ba^+.$$

Clarament, tenim

$$\Psi_{\{a,b\}}(L) = \{(n, 1) \mid n \geq 1\},$$

i, en aquest cas, L és semilineal; interseccionant L amb el llenguatge regular a^+b obtenim $\{a^{2^n}b \mid n \geq 1\}$, que és no-semilineal.

La semilinealitat de llenguatges com l'anterior L ó $\{a^{2^n}ba^m \mid n, m \geq 0\}$ mostra que malgrat ser una propietat forta, implicant les propietats d'IAP i de BLI, la semilinealitat no és suficientment selectiva per a refutar com a “no-suau” (veurem el significat d'aquest terme més endavant) un mecanisme generatiu donat. Si $\{a^{2^n} \mid n \geq 1\}$ efectivament no té relevància pels llenguatges naturals, aleshores també succeeix el mateix amb $\{a^{2^n}ba^m \mid n, m \geq 0\}$, malgrat la seva semilinealitat.

Segons la discussió anterior, no podem obtenir directament una demostració de la no-semilinealitat dels llenguatges de ICC mitjançant l'ús dels Exemples de la Secció 4.2. És més, tenim

Teorema 8.1.3 La família ICC_c conté llenguatges no-semilineals.

Demostració. Considerem la gramàtica contextual (presentada en forma modular)

$$G = (V, B, (S_1, C_1), (S_2, C_2), (S_3, C_3)),$$

amb

$$\begin{aligned} V &= \{a, b, c, d, e\}, \\ B &= \{dcbaabbabbcd\}, \\ S_1 &= \{dc(baa)^*(baab)^+\}, \quad C_1 = \{(\lambda, a)\}, \\ S_2 &= \{(abb)^+cd\}, \quad C_2 = \{(b, \lambda)\}, \\ S_3 &= \{cba(abb)^+abbc\}, \quad C_3 = \{(e, e)\}. \end{aligned}$$

Per a una cadena $z = dcba(abb)^nabbcd$ o $z = decba(abb)^nabbced$, denotem l'enter n com a $\gamma(x)$. Hauríem d'interseccionar $\Psi_V(L_{in}(G))$ amb $M = \{(n_1, n_2, 2, 2, 2) \mid n_1, n_2 \geq 1\}$. A més a més, només estem interessats en cadenes de la forma $decba(abb)^nabbced$: el símbol e només pot ser introduït usant la producció (S_3, C_3) i, per tant, només per a derivar cadenes de la forma $dcba(abb)^nabbcd$.

Comencem per l'axioma $w_0 = dcbaabbabbcd$, amb $\gamma(w_0) = 1$.

Prenem una cadena $z = dcba(abb)^nabbcd$ amb $n \geq 1$. Si usem (S_3, C_3) , aleshores la derivació només pot continuar usant aquesta producció de G . Podem usar també (S_1, C_1) , la qual cosa ens condueix a $dcbaabab(abb)^{n-1}abbcd$, i després a $dcbaabaab(abb)^{n-1}abbcd$. La derivació ha de continuar a (S_1, C_1) , pas a pas, d'esquerra a dreta, duplicant cada ocurrència d' a de subcadenes bab i introduint dues ocurrències d' a entre símbols b de subcadenes bb . D'aquesta

manera obtenim finalment una cadena $z' = dcbaa(baa)^nbbcd$ (cada ocurrència de b de la subcadena $(abb)^n$ de z té la subcadena aa en la seva banda dreta). Només quan aquesta reescriptura d'esquerra a dreta s'hagi acabat, podem aplicar (S_2, C_2) . Ara, de dreta a esquerra de la cadena, cada ocurrència de b de la subcadena $(baa)^n$ de z' és duplicada i cada subcadena aa és reemplaçada per $abba$. De manera provisional obtenim una cadena $z'' = dcba(abb)^{2n}abbc$ (per a cada a de la subcadena $(baa)^n$ de z' obtenim una subcadena abb , amb l'ocurrència d' a en primera posició de $(baa)^n$ de z' , tenint en compte l'ocurrència d' a després d' $(abb)^{2n}$ a z''). Conseqüentment, hem obtingut una cadena de la mateixa forma que z , però amb el doble de subcadena abb en la posició distingida de l'escriptura prèvia. Aquest procés pot continuar. En aquest cas, a partir de l'axioma w_0 amb $\gamma(w_0) = 1$ obtenim una seqüència de cadenes, $w_0, w_1, \dots$, de manera que

$$\gamma(w_{i+1}) = 2 \cdot \gamma(w_i),$$

la qual cosa implica

$$\gamma(w_i) = 2^i, \quad i \geq 0.$$

Per aquestes cadenes w_i tenim $|w_i|_a = 2^i + 2$, $|w_i|_b = 2 \cdot 2^i + 3$, $|w_i|_c = |w_i|_d = 2$.

En conseqüència, pel conjunt semilineal M especificat anteriorment tenim

$$\Psi_V(L_{in}(G)) \cap M = \{(2^i + 2, 2^{i+1} + 3, 2, 2, 2) \mid i \geq 0\},$$

que no és semilineal i, per tant, tampoc $\Psi_V(L_{in}(G))$ és semilineal. $\square$

Els Teoremes 8.1.2 i 8.1.3 mostren que, si prenem la semilinealitat com a condició de “suaument sensible al context”, aleshores els llenguatges de $ECC_c \cup$

ICC_c no són suaument sensibles al context, malgrat tenir la propietat “més natural” de BLI. En les seccions següents, també discutirem al voltant de la semilinealitat de certes famílies restringides de llenguatges contextuais.

8.2 La complexitat descriptiva

Si tenim un llenguatge infinit L , generat per una gramàtica G d’una classe X donada, en general, és fàcil de trobar un nombre infinit de gramàtiques de tipus X que generin el llenguatge L . En el cas de les gramàtiques contextuais ho podem aconseguir introduint com a nou axioma de G una cadena d’ L . Aquest procediment incrementa artificialment el nombre d’axiomes i produeix una gramàtica ambigua (els nous axiomes ja van ser generats per G i, per tant, ara poden ser produïts de dues maneres diferents).

Per altra banda, des d’un punt de vista pràctic i d’importàcia, sembla clar el fet de treballar amb gramàtiques que siguin el més simple possibles. D’alguna manera, la gramàtica “més petita” G que generi un llenguatge L pot ser considerada com la gramàtica d’ L “verdadera”, la seva descripció “estàndard”. A més a més, “l’economia de la descripció”, veure (Friedman et al., 1987), és un criteri molt important per a comparar diversos formalismes usats en la modelació dels llenguatges naturals.

Aquestes observacions ens condueixen cap a la important noció de *mesura* de la complexitat de les gramàtiques (anàlogament, dels llenguatges). En la

Secció 3.13 de la pàgina 3.13 vam presentar algunes d'aquestes mesures per a gramàtiques lliures del context (Var , $Prod$, $Symb$), definint per a elles algunes propietats específiques.

També poden ser considerades diverses mesures pel cas de les gramàtiques contextuais. Començarem per definir tres mesures bàsiques pel cas de gramàtiques contextuais totals (la particularització a altres classes de gramàtiques contextuais és òbvia). Totes les gramàtiques amb selecció considerades en aquesta secció tenen la funció de selecció computable.

Definició 8.2.1 Sigui $G = (V, B, C, \varphi)$ una gramàtica contextual total. Definim

$$Ax(G) = \text{card}(B),$$

$$Con(G) = \text{card}(C),$$

$$Phi(G) = \text{card}\{C' \subseteq C \mid \text{existeixen } x_1, x_2, x_3 \in V^* \\ \text{tal que } C' = \varphi(x_1, x_2, x_3)\}.$$

La mesura Ax avalua el nombre d'axiomes, Con és el nombre de contextos, i Phi és el nombre de classes d'equivalència definides a $V^* \times V^* \times V^*$ per la funció φ . És clar que, Ax i Con tenen sentit també per a gramàtiques sense selecció, però Phi és específic de les gramàtiques contextuais amb selecció.

Extendrem aquestes mesures, de manera natural, als llenguatges:

$$M_X(L) = \min\{M(G) \mid G \in X, L = L_\alpha(G)\},$$

per a $M \in \{Ax, Con, Phi\}$, $X \in \{EC, IC, ECC, ICC, TC\}$ i un apropiat $\alpha \in \{in, ex, -\}$.

Òbviament, si $L \in F_1 \subseteq F_2$ (i les gramàtiques que generen els llenguatges d' F_1 són casos particulars de gramàtiques que generen llenguatges d' F_2), aleshores $M_{F_2}(L) \leq M_{F_1}(L)$, per a totes les mesures M que poden ser definides per ambdós tipus de gramàtiques.

Les mesures anteriors indueixen infinites jerarquies de llenguatges en les corresponents famílies. Específicament, tenim

Teorema 8.2.1 Totes les mesures M_X , per a M i X definits de la manera anterior, estan connectades.

Demostració. Considerem primer el llenguatge finit

$$L_n = \{(a^i b^i)^3 \mid 1 \leq i \leq n\},$$

per a $n \geq 1$. Pel fet que $L_n = L_{ex}(G_n) = L_{in}(G_n)$ per a $G_n = (\{a, b\}, L_n, \emptyset)$, tenim que $Ax_X(L_n) \leq n$ per a tot $X \in \{EC, IC, ECC, ICC, TC\}$.

Ja que cap cadena d' L_n pot ser obtinguda a partir d'una altra cadena de L_n mitjançant l'adjunció d'un context no-buit, se segueix que si una gramàtica contextual total G genera L_n , aleshores no és possible cap pas de derivació a G i, per tant, el conjunt d'axiomes de G ha de ser igual a L_n . A més a més, $Ax_X(L_n) = n$ per a tot X com abans. Per tant, hem provat el teorema per a la mesura Ax .

Considerem ara els llenguatges L_n^+ per a L_n com abans, $n \geq 1$. Per a la gramàtica (sense selecció) $G'_n = (\{a, b\}, L_n, \{(\lambda, x) \mid x \in L_n\})$ tenim que $L_{ex}(G'_n) = L_n^+$ i, per tant, $Con_X(L_n^+) \leq n$ per a tot $X \in \{EC, ECC, TC\}$. Per a la gramàtica $G''_n = (\{a, b\}, L_n, \{(\lambda, x) \mid x \in L_n\}, \varphi)$, amb $\varphi(w) = \{(\lambda, x) \mid x \in L_n\}$ per a tot $w \in L_n$, tenim que $L_{in}(G''_n) = L_n^+$ i, per tant, $Con_{ICC}(L_n^+) \leq n$.

Si una gramàtica contextual total G genera el llenguatge L_n^+ , pel fet que cap cadena d' L_n pot ser obtinguda a partir d'una altra cadena d' L_n mitjançant l'adjunció d'un context, i pel fet que totes les cadenes x^i són d' L_n^+ , per a $x \in L_n, i \geq 1$, la gramàtica G ha de contenir un context (u, v) tal que uv tingui com a mínim una subcadena igual a x . Aquest context no pot ser usat per a generar y^i per a $x \neq y$ i un valor arbitrari d' i i, per tant, $Con(G) \geq n$. Se segueix que $Con_X(L_n^+) = n$ per a tot $X \in \{EC, IC, ECC, ICC, TC\}$.

El llenguatge L_n^+ anterior no pertany a IC ; per aquesta família podem considerar el llenguatge $L'_n = \{a_1, \dots, a_n\}^*$. És fàcil de veure que $Con_{IC}(L'_n) = n$.

Per a la mesura *Phi* considerem els llenguatges

$$L''_n = \{((a^i b^i)^3 c)^m \mid m \geq 1, 1 \leq i \leq n\}, \quad n \geq 1.$$

Per a la gramàtica contextual

$$G'''_n = (\{a, b, c\}, \{(a^i b^i)^4 c \mid 1 \leq i \leq n\}, \{(\lambda, (a^i b^i)^3 c) \mid 1 \leq i \leq n\}, \varphi),$$

amb

$$\varphi(x) = \{(\lambda, (a^i b^i)^3 c)\}, \text{ per a } x \in \{(a^i b^i)^3 c\}^+, 1 \leq i \leq n,$$

$$\varphi(x) = \{(\lambda, (ab)^3 c)\}, \text{ per a } x \notin L_n'',$$

tenim que $L_{ex}(G) = L_n''$. Pel fet que $Phi(G_n''') = n$ tenim que $Con_X(L_n'') \leq n$ per a $X \in \{ECC, TC\}$. Si posem

$$\varphi(x) = \emptyset, \text{ per a } x \notin L_n'',$$

aleshores $L_{in}(G_n''') = L_n''$ i, per tant, $Phi_{ICC}(L_n'') \leq n + 1$.

Si una gramàtica contextual total G genera el llenguatge L_n'' , aleshores per a cada i , $1 \leq i \leq n$, existeix un context (u, v) tal que uv té una subcadena $a^i b^i c$ i $(u, v) \in \varphi(x_1, x_2, x_3)$, per algun x_1, x_2, x_3 tal que $x_1 u x_2 v x_3 \in L_n''$. Això vol dir que també $x_1 x_2 x_3$ conté una subcadena $a^i b^i c$. Si y_1, y_2, y_3 són cadenes de $\{a, b, c\}^*$ tal que $y_1 y_2 y_3$ conté una subcadena $a^j b^j c$, aleshores no podem tenir $(u, v) \in \varphi(y_1, y_2, y_3)$ per a (u, v) com abans. D'aquesta manera hem distingit n subconjunts $C_1, \dots, C_n$ de C pels quals $\varphi^{-1}(C_i) \neq \emptyset, 1 \leq i \leq n$. Això implica que $Phi_X(L_n'') = n$ per a $X \in \{ECC, TC\}$. Si G és una gramàtica contextual interna, aleshores existeix com a mínim una cadena z de $\{a, b, c\}^*$ tal que $\varphi(z) = \emptyset$, ja que sinó podem obtenir cadenes paràsites adjuntant contextos a subcadenaes d'elements d' L_n'' diferents a les associades als contextos (u, v) com abans. Això vol dir que $Phi_{ICC}(L_n'') = n + 1$. En aquesta situació, el teorema és provat també per a la mesura Phi . $\square$

Considerarem ara els problemes o qüestions de decidibilitat Q1 – Q5 es-

pecificats a la Secció 3.13 de la pàgina 163. Ens limitarem a les gramàtiques contextuais amb selecció usades en el mode extern.

Òbviamment, donada una gramàtica G és una feina trivial el computar $Ax(G)$ i $Con(G)$ mitjançant un simple recompte de les corresponents components de G . Aquest, però, no és el cas de la mesura Phi .

Teorema 8.2.2 No podem computar $Phi(G)$ per a una gramàtica contextual arbitrària G amb una funció de selecció computable (és a dir, el problema Q1 té una resposta negativa per a la mesura Phi).

Demostració. Considerem dues n -tuples de cadenes no-buides sobre $\{a, b\}$, $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n)$. Construïm les gramàtiques (sense selecció)

$$G''(z) = (\{a, b, c\}, \{c\}, \{(ba^i, z_i) \mid 1 \leq i \leq n\}),$$

per a $z \in \{x, y\}$ (cal observar la relació/diferència entre aquestes gramàtiques i $G(z), G'(z), z \in \{x, y\}$ considerades al Lema 7.0.1 de la pàgina 233).

Òbviamment, $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) \neq \{c\}$ si, i només si, el problema de correspondència de Post per a x, y té una solució; la qual cosa sabem és un problema indecidible.

Considerem ara la gramàtica

$$G = (\{a, b, c\}, \{c\}, \{(ba^i, x_i) \mid 1 \leq i \leq n\} \cup \{(c, c)\}, \varphi),$$

amb

$$\begin{aligned}\varphi(w) &= \{(ba^i, x_i) \mid 1 \leq i \leq n\} \cup \{(c, c)\}, \text{ per a } w \in L_{ex}(G''(x)) \cap L_{ex}(G''(y)), \\ \varphi(w) &= \{(ba^i, x_i) \mid 1 \leq i \leq n\}, \text{ per a } w \in (L_{ex}(G''(x)) - L_{ex}(G''(y))) \cup \{c\}, \\ \varphi(w) &= \emptyset, \text{ per a } w \notin L_{ex}(G''(x)).\end{aligned}$$

Els llenguatges $L_{ex}(G''(x)), L_{ex}(G''(y))$ són lineals i, per tant, la funció φ és computable.

Si $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) = \{c\}$, aleshores el de φ consisteix en dos conjunts, $\emptyset$ i $\{(ba^i, x_i) \mid 1 \leq i \leq n\}$; quan $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) \neq \{c\}$, aleshores el rang de φ també inclou el conjunt $\{(ba^i, x_i) \mid 1 \leq i \leq n\} \cup \{(c, c)\}$. A més a més, $Phi(G) = 2$ si, i només si, $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) = \{c\}$, és a dir, no existeix cap algorisme per a computar $Phi(G)$. $\square$

Pel que fa als altres problemes o qüestions, Q2 – Q5, existeix una diferència entre la mesura Ax i les mesures Con, Phi .

Teorema 8.2.3 Tots els problemes o qüestions Q2 – Q5 tenen una resposta positiva per a la mesura Ax per a gramàtiques contextuais externes amb funció de selecció computable.

Per a la demostració d'aquest resultat, necessitem el següent resultat auxiliar.

Definició 8.2.2 Una gramàtica contextual $G = (V, B, C, \varphi)$ s'anomena *re-*

duïda si, per a cada $x \in B$, tenim que $L_{ex}(G) \neq L_{ex}(G_x)$ on $G_x = (V, B - \{x\}, C, \varphi)$. És clar que la definició pot ser donada per a qualsevol tipus de gramàtica contextual, però nosaltres treballarem només amb gramàtiques amb selecció usades en mode extern.

Lema 8.2.1 És decidible si una gramàtica contextual externa amb funció de selecció computable és o no és reduïda. Donada una gramàtica contextual arbitrària amb funció de selecció computable, podem construir una gramàtica reduïda equivalent.

Demostració. Sigui $G = (V, B, C, \varphi)$ una gramàtica contextual amb funció φ computable. El conjunt $Min_0(L_{ex}(G))$ pot ser construït (és un subconjunt del conjunt finit B , cada cadena d' $L_{ex}(G) - B$ té una subcadena a B i, per tant, $Min_0(L_{ex}(G)) = Min_0(B)$). Si $Min_0(L_{ex}(G)) = B$, aleshores G és reduïda.

En el cas invers, escrivim les cadenes de $B - Min_0(L_{ex}(G))$ en ordre creixent de les seves longituds,

$$z_1, z_2, \dots, z_k, \quad |z_i| \leq |z_{i+1}|, \quad 1 \leq i \leq k-1, k \geq 1.$$

Denotem per $G' = (V, Min_0(B), C, \varphi)$. Si $z_1 \in L_{ex}(G')$ (i això pot ser algorísmicament verificat), aleshores podem esborrar z_1 de B sense modificar el llenguatge generat. En qualsevol altre cas, sent la cadena més curta de $B - Min_0(L_{ex}(G))$, la cadena z_1 no pot ser obtinguda a partir de cap altra cadena de $B - Min_0(L_{ex}(G))$ i, per tant, no pot ser esborrada del conjunt axiomàtic de la gramàtica G . Construïm la gramàtica $G_1 = (V, Min_0(B) \cup \{z_1\}, C, \varphi)$ i con-

tinuem l'argument amb aquesta gramàtica i la seqüència de cadenes $z_2, \dots, z_k$. Després de discutir el cas de la cadena z_k obtenim una gramàtica G'' que és equivalent a G i reduïda. $\square$

Demostració. (del Teorema 8.2.3.) Si $L \in ECC$, aleshores $Ax_{ECC}(L) = Min_0(L)$. Considerem un llenguatge $L \in ECC$ generat per una gramàtica $G = (V, B, C, \varphi)$ amb una funció φ computable. D'acord amb el lema previ, podem suposar que G és reduïda.

Construïm una nova gramàtica de la forma $G' = (V, Min_0(L), C', \varphi')$, de la següent manera. Per a cada $x \in B - Min_0(L)$ existeix $z \in Min_0(L)$ tal que $x = uzv$, per algun $u, v \in V^*$. El conjunt B és finit i, per tant, el conjunt de tots els contextos d'aquest tipus és també finit. Denotem per C_0 aquest conjunt. Aleshores $C' = C \cup C_0$.

Definim també

$$\begin{aligned}\varphi'(x) &= \varphi(x) \cup \{(u, v) \in C_0 \mid uxv \in B - Min_0(L)\}, \text{ per a } x \in Min_0(L), \\ \varphi'(x) &= \varphi(x), \text{ per a } x \notin Min_0(L).\end{aligned}$$

És fàcil de veure que $L_{ex}(G) = L_{ex}(G')$.

A més a més, hem trobat una gramàtica per a L que és mínima respecte a la mesura Ax , amb la qual cosa el problema Q5 és resoluble. Juntament amb l'òbvia resposta positiva a Q1, obtenim la resposta positiva a Q2, que, a la vegada, implica la resposta positiva als problemes Q3 i Q4. $\square$

Observació 8.2.1 En la demostració prèvia, l'ús de gramàtiques contextuais amb selecció era essencial: per a la gramàtica $G = (\{a, b\}, \{b, b^2\}, \{(\lambda, a)\})$ tenim $L_{ex}(G) = \{ba^n, b^2a^n \mid n \geq 0\}$. Tota gramàtica contextual amb selecció $G' = (\{a, b\}, B, C)$ ha de tenir $\{b, b^2\} \subseteq B$. No obstant, $Min_0(L_{ex}(G)) = \{b\}$, és a dir $Ax_{EC}(L_{ex}(G)) < Ax(G')$ per a cada gramàtica contextual G' sense selecció.

A diferència del Teorema 8.2.3 de la pàgina 254, per a les mesures *Con* i *Phi* tenim els següents resultats.

Teorema 8.2.4 Tots els problemes o qüestions Q2 – Q5 tenen una resposta negativa per a la mesura *Con*, per a gramàtiques contextuais externes amb funció de selecció computable.

Demostració. En el Lema 7.5.2 de la pàgina 238 vam veure que el problema de la finitud és indecidible per a gramàtiques contextuais amb funció de selecció computable. Pel fet que $Con_{ECC}(L) = 0$ si, i només si, el llenguatge L és finit, se segueix que no podem decidir si $Con_{ECC}(L_{ex}(G)) = 0$ o no, per a una gramàtica contextual arbitrària G amb funció de selecció computable.

Considerem ara el llenguatge L_n^+ (sobre el vocabulari $\{a, b\}$) usat en la demostració del Teorema 8.2.1 de la pàgina 250: $Con_{ECC}(L_n^+) = n$. Construïm el llenguatge

$$L'_n = L_n^+ \cup L_{ex}(G),$$

per a una gramàtica contextual arbitrària $G = (\{c, d, e\}, B, C, \varphi)$ (amb φ com-

putable). Aleshores $Con_{ECC}(L'_n) = n$ si, i només si, $Con_{ECC}(L_{ex}(G)) = 0$ (en general, si $L, L' \in ECC$, aleshores $Con_{ECC}(L \cup L') \leq Con_{ECC}(L) + Con_{ECC}(L')$ – veure la demostració del Lema 6.1.1 de la pàgina 210, que inclou la construcció de la demostració del Lema 5.1.1 de la pàgina 185. Per altra banda, si els dos llenguatges L, L' tenen vocabularis disjunts, aleshores $Con_{ECC}(L \cup L') = Con_{ECC}(L) + Con_{ECC}(L')$). Pel fet que no podem decidir si $Con_{ECC}(L_{ex}(G)) = 0$ o no, per una arbitrària G , no podem decidir si $Con_{ECC}(L'_n) = n$ o no, per a $n \geq 0$. Això prova que el problema Q3 té una resposta negativa per a Con .

La resposta positiva a Q1 i la resposta negativa a Q3 impliquen la resposta negativa a Q2 i a Q5.

Per al problema Q4, considerem altra vegada les gramàtiques $G''(x), G''(y)$ de la demostració del Teorema 8.2.2 de la pàgina 253, associada amb dues n -tuples de cadenes no-buides sobre $\{a, b\}$, $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n)$. Construïm la gramàtica

$$G = (\{a, b, c\}, \{c\}, \{(a, \lambda), (b, \lambda)\}, \varphi),$$

amb

$$\varphi(z) = \begin{cases} \{(a, \lambda), (b, \lambda)\}, & \text{si } K_{|z|}(G''(y)) \cap L_{ex}(G''(x)) = \{c\}, \\ \{(a, \lambda)\}, & \text{en qualsevol altre cas.} \end{cases}$$

per a tot $z \in \{a, b\}^*$, on $K_j(G''(y))$ són els conjunts definits en la demostració del Lema 5.1.7 de la pàgina 192 (ja que els conjunts $K_j(G''(y))$ són finits i

$L_{ex}(G''(x))$ és un llenguatge lineal, la funció φ definida anteriorment és computable).

És fàcil de veure que si $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) = \{c\}$, aleshores per a tot $i \geq 0$ tenim que $K_i(G''(y)) \cap L_{ex}(G''(x)) = \{c\}$ i, per tant, $\varphi(z) = \{(a, \lambda), (b, \lambda)\}$ per a tot $z \in \{a, b\}^*$; a més a més a i b apareixen un nombre infinit de vegades en les cadenes d' $L_{ex}(G)$. Conseqüentment, $Con_{ECC}(L_{ex}(G)) = 2$.

Per altra banda, si $L_{ex}(G'(x)) \cap L_{ex}(G'(y)) \neq \{c\}$, aleshores existeix un nombre natural n_0 tal que $K_i(G'(y)) \cap L_{ex}(G'(x)) \neq \emptyset$ per a tot $i \geq n_0$. A més a més, només les cadenes de longitud com a màxim n_0 poden contenir el símbol b . Una gramàtica contextual amb el conjunt axiomàtic $B = \{w \in L_{ex}(G) \mid |w| \leq n_0\}$ i el conjunt de contextos $C = \{(a, \lambda)\}$ pot generar el nostre llenguatge, és a dir $Con_{ECC}(L_{ex}(G)) = 1$. No obstant, $Con(G) = 2$. Conseqüentment, no podem decidir si $Con(G) = Con_{ECC}(L_{ex}(G))$ o no. també Q4 té una resposta negativa. $\square$

Per tal de tractar els problemes Q2 – Q5 per a la mesura Φ necessitem un resultat auxiliar addicional, que és també d'interès intrínsec.

Lema 8.2.2 És indecidible si una gramàtica contextual externa arbitrària amb funció de selecció computable genera o no un llenguatge pertanyent a la família EC.

Demostració. Considerem altra vegada la gramàtica contextual amb selecció

G de la demostració del Teorema 8.2.2 de la pàgina 253. Si $L_{ex}(G_x) \cap L_{ex}(G_y) = \{c\}$, aleshores la definició de la funció φ es redueix a

$$\begin{aligned}\varphi(w) &= \{(ba^i, x_i) \mid 1 \leq i \leq n\} \cup \{(c, c)\}, \text{ per a } w \in L_{ex}(G''(x)), \\ \varphi(w) &= \emptyset, \text{ per a } w \notin L_{ex}(G''(x)).\end{aligned}$$

A més a més, $L_{ex}(G) = L_{ex}(G''(x))$, és a dir, $L_{ex}(G)$ és un llenguatge pertanyent a EC .

Inversament, si $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) \neq \{c\}$, aleshores $L_{ex}(G''(x)) \cap L_{ex}(G''(y))$ és un llenguatge infinit i, per a un nombre infinit de cadenes w , tenim que $\varphi(w) = \{(ba^i, x_i) \mid 1 \leq i \leq n\} \cup \{(c, c)\}$. A més a més, qualsevol gramàtica per al llenguatge $L_{ex}(G)$ ha de contenir un context (u, v) tal que $|uv|_c > 0$. Aquest context no pot ser aplicat a totes les cadenes de $\{a, b, c\}^*$ i, per tant, necessitem una funció de selecció. Per tant, el llenguatge $L_{ex}(G)$ no pertany a EC . Per tant, el problema de si $L_{ex}(G) \in EC$ o no hi pertany és indecidible. $\square$

Teorema 8.2.5 Els problemes Q2 i Q3 tenen una resposta negativa per a la mesura Φ , per a gramàtiques contextuais externes arbitràries amb funció de selecció computable.

Demostració. Considerem la gramàtica G de la demostració del Teorema 8.2.3 de la pàgina 254. A partir del lema anterior, sabem que és indecidible si $L_{ex}(G) \in EC$ o no. Pel fet que, per a $L \in ECC$ tenim $\Phi_{ECC}(L) = 1$ si, i només si, $L \in EC$, se segueix que no podem decidir si $\Phi_{ECC}(L_{ex}(G)) =$

1 o no. Per a $n > 1$ procedim d'igual manera que en la demostració del teorema 8.2.4 de la pàgina 257: a partir de la unió d' $L_{ex}(G)$ anterior amb un llenguatge L_{n-1} sobre el vocabulari $\{d, e\}$ tal que $Phi_{ECC}(L_{n-1}) = n - 1$ tenim $Phi_{ECC}(L_{ex}(G) \cup L_{n-1}) = n - 1 + Phi_{ECC}(L_{ex}(G))$, que és igual a n si, i només si, $Phi_{ECC}(L_{ex}(G)) = 1$; això és indecidible. Hem obtingut la resposta negativa al problema Q3.

La resposta negativa a Q3 implica una resposta negativa a Q2, independentment de la resposta dels altres problemes. $\square$

Teorema 8.2.6 El problema Q4 té una resposta negativa per a la mesura Phi , per a gramàtiques contextuais externes amb una funció de selecció computable.

Demostració. Per a la gramàtica G de la demostració del Teorema 8.2.3 de la pàgina 254 sabem que $Phi(G) = 2$ o $Phi(G) = 3$ i no podem decidir quin és el valor actual de $Phi(G)$ per a una arbitrària G . Si $Phi(G) = 2$, això vol dir que $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) = \{c\}$ i $L_{ex}(G) = L_{ex}(G''(x)) \in EC$. Això implica que $Phi_{ECC}(L_{ex}(G)) = 1$. Inversament, si $Phi_{ECC}(L_{ex}(G)) = 1$, això implica que la funció φ és uniformement definida a $\{a, b, c\}^*$, és a dir $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) = \{c\}$.

Conseqüentment, $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) = \{c\}$ si, i només si, $Phi(G) = 2$, la qual cosa succeeix si, i només si, $Phi_{ECC}(L_{ex}(G)) = 1$.

Per altra banda, $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) \neq \{c\}$ si, i només si, $Phi(G) = 3$. Provarem que també tenim $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) \neq \{c\}$ si, i només si,

$\text{Phi}_{ECC}(L_{ex}(G)) = 3$. A més a més, $\text{Phi}(G) = \text{Phi}_{ECC}(L_{ex}(G))$ si, i només si, $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) \neq \{c\}$, la qual cosa és indecidible, i obtindríem la resposta negativa al problema Q4.

El fet que $\text{Phi}_{ECC}(L_{ex}(G)) = 3$ impliqui $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) \neq \{c\}$ és obvi. Suposem que $L_{ex}(G''(x)) \cap L_{ex}(G''(y)) \neq \{c\}$. Aleshores $L_{ex}(G''(x)) \cap L_{ex}(G''(y))$ és un conjunt infinit i existeix com a mínim un context (u, v) que conté el símbol c que és usat un nombre infinit de vegades quan està produint el llenguatge $L_{ex}(G)$. Se segueix que $\text{Phi}_{ECC}(L_{ex}(G)) > 1$. A més a més, la funció de selecció ha de ser definida de manera diferent en els conjunts $(L_{ex}(G''(x)) \cap L_{ex}(G''(y))) - \{c\}$ i $(L_{ex}(G''(x)) - L_{ex}(G''(y))) \cup \{c\}$. Per a w en aquests conjunts hem de tenir $\varphi(w) \neq \emptyset$. No obstant, per a $w \notin L_{ex}(G''(x))$ no podem tenir $\varphi(w) = \varphi(z)$ per algun $z \in (L_{ex}(G''(x)) \cap L_{ex}(G''(y))) - \{c\}$ o $z \in (L_{ex}(G''(x)) - L_{ex}(G''(y))) \cup \{c\}$, ja que les cadenes de la forma czc pertanyen a $L_{ex}(G)$ però no pertanyen a $L_{ex}(G''(x))$. Si $\varphi(czc)$ conté un context (ba^i, x_i) , aleshores $ba^iczcx_i \in L_{ex}(G)$, la qual cosa no és possible. Per tant, la funció φ ha de ser definida de manera diferent en el complementari d' $L_{ex}(G''(x))$, és a dir $\text{Phi}_{ECC}(L_{ex}(G)) \geq 3$, la qual cosa conclou la demostració del teorema.

□

El problema Q5 resta *obert* per a la mesura Phi .

Moltes altres mesures poder ser també considerades. Per exemple,

Definició 8.2.3 Donada una gramàtica contextual total $G = (V, B, C, \varphi)$ po-

dem definir:

$$\begin{aligned}
 MAx(G) &= \max\{|w| \mid w \in B\}, \\
 TAx(G) &= \sum_{w \in B} |w|, \\
 MCon(G) &= \max\{|uv| \mid (u, v) \in C\}, \\
 TCon(G) &= \sum_{(u, v) \in C} |uv|, \\
 MPhi(G) &= \max\{\text{card}(\varphi(x)) \mid x \in V^*\}.
 \end{aligned}$$

Aquestes mesures indiquen la longitud màxima d'un axioma, la longitud total dels axiomes, la longitud màxima de les cadenes d'un context, la longitud total de les cadenes de tots els contextos, el nombre màxim de contextos associats a una cadena, respectivament (cal recordar que els paràmetres $MAx(G)$ i $MCon(G)$ apareixen diverses vegades en les demostracions dels lemes de la Secció 5.1).

Extendrem aquestes mesures, de manera natural, a llenguatges de les famílies EC, ECC, IC , i ICC . Per definició, per a llenguatges $L \in EC$, $MPhi_{ECC}(L) = 0$ i, per a $L \in IC$, $MPhi_{ICC}(L) = 0$.

Com era d'esperar, tenim

Teorema 8.2.7 Totes les mesures $MAx, TAx, MCon, TCon, MPhi$ estan connectades en el cas de les gramàtiques contextuais internes i externes.

Demostració. Considerem els llenguatges

$$L_n = \{a^n\}^*,$$

generats per les gramàtiques $G_n = (\{a\}, \{a^n\}, \{(\lambda, a^n)\})$, $n \geq 1$, en els modes extern i intern. A més a més, $M_F(L_n) \leq n$, $F \in \{EC, IC\}$, per a tot $M \in \{MAx, TAx, MCon, TCon\}$. És fàcil de veure que, tota gramàtica contextual amb selecció G que generi aquest llenguatge, ha de tenir a^n com a axioma (i, per tant, $MAx(G) \geq n, TAx(G) \geq n$); i cap context (a^i, a^j) pot tenir $0 \neq i + j < n$ (i, per tant, $MCon(G) \geq n, TCon(G) \geq n$).

Per a les mesures $MPhi$ i les famílies EC, ECC considerem els llenguatges

$$L'_n = \{c\}\{a^ib \mid 1 \leq i \leq n\}^*, \quad n \geq 2$$

(per a $n = 1$ l'assertió és òbvia), generats en el mode extern per la gramàtica contextual sense selecció $G'_n = (\{a, b, c\}, \{c\}, \{(\lambda, a^ib) \mid 1 \leq i \leq n\})$, $n \geq 2$. Suposem que $MPhi_{ECC}(L'_n) < n$ i sigui $G = (\{a, b, c\}, B, C, \varphi)$ una gramàtica tal que $L'_n = L_{ex}(G)$ i $MPhi(G) < n$. Clarament, $c \in B$ i per a cada $(u, v) \in C$ tal que $(u, v) \in \varphi(x)$ per algun $x \in L'_n$ hem de tenir $u = \lambda, v \in \{a^ib \mid 1 \leq i \leq n\}^*$. Denotem per

$$b_r(L'_n) = \{x \in L'_n \mid |x|_b \leq r\},$$

per a $r \geq 0$. Considerem també els conjunts $K_i(G)$ definits com en la demostració del Lema 5.1.7 de la pàgina 192 (cal recordar que $K_i(G)$ conté totes les cadenes d' $L_{ex}(G)$ que poden ser generades per G usant com a màxim i contextos). Pel fet que $|uv|_b > 0$ per a tot $(u, v) \in C$ el qual és efectivament

usat en una derivació, tenim $b_r(L'_n) \subseteq K_r(G)$ per a tot $r \geq 1$. No obstant, tenim

$$\begin{aligned} \text{card}(b_r(L'_n)) &= \sum_{i=0}^r n^i, \\ \text{card}(K_r(G)) &\leq \sum_{i=0}^r \text{card}(B) \cdot (n-1)^i = \text{card}(B) \cdot \sum_{i=0}^r (n-1)^i. \end{aligned}$$

Pel fet que $n \geq 2$, per a un valor d' r suficientment gran, tenim

$$\text{card}(b_r(L'_n)) > \text{card}(K_r(G)),$$

la qual cosa contradiu la inclusió $b_r(L'_n) \subseteq K_r(G)$.

Conseqüentment, $M\text{Phi}_{ECC}(L'_n) = n$.

Pel cas intern prenem la gramàtica

$$\begin{aligned} G_n &= (\{a, b, c, d, e\}, \{eabeab\}, C_n, \varphi_n), \quad n \geq 1, \\ C_n &= \{(ab, ab)\} \cup \{(c^i dc^i, c^i dc^i) \mid 1 \leq i \leq n\}, \\ \varphi_n(b^s ea^t) &= \{(ab, ab)\}, \quad s, t \geq 1, \\ \varphi_n(ea^s b^t e) &= \{(c^i dc^i, c^i dc^i) \mid 1 \leq i \leq n\}, \quad s, t \geq 1. \end{aligned}$$

Mitjançant la definició de φ_n , tenim $M\text{Phi}_{ICC}(L_{in}(G_n)) \leq n$.

Considerem una gramàtica arbitrària $G = (\{a, b, c, d, e\}, B, C, \varphi)$ tal que $L_{in}(G) = L_{in}(G_n)$. Denotem

$$M = \{c^i dc^i ea^m b^m ec^i dc^i a^m b^m \in L_{in}(G_n) \mid m \geq 1, 1 \leq i \leq n\}.$$

Cap derivació $x \Rightarrow_{in} y$, amb $x, y \in M$ és possible:

- existeixen quatre subcadenaes c^i i, per tant, el nombre d'ocurrències del símbol c no pot ser incrementat;
- si existeix una derivació que incrementa el nombre d'ocurrències d' a i de b , aleshores hem de tenir $(a^t b^t, a^t b^t) \in \varphi(b^r e c^i d c^i a^r)$, per a algun $t \geq 1, r \geq 1$; aleshores
$$c^i d c^i e a^{r+1} b b^r e c^i d c^i a^r a b^{r+1} \Rightarrow_{in} c^i d c^i e a^{r+1} b a^t b^t b^r e c^i d c^i a^r a^t b^t a b^{r+1},$$
i la cadena obtinguda no pertany a $L_{in}(G_n)$.

Conseqüentment, per tal de generar cadenaes d' M amb un valor d' m arbitràriament gran (p.e., amb $m > MAx(G)$), necessitem les derivacions $x \Rightarrow_{in} y$, per a $x \notin M, y \in M$. Si $y = c^i d c^i e a^m b^m e c^i d c^i a^m b^m$, aleshores l'única possibilitat és tenir $x = e a^m b^m e a^m b^m$ (cada cadena d' $L_{in}(G)$ conté dues ocurrències del símbol e). Aleshores, el context usat és $(c^i d c^i, c^i d c^i)$, associat a $e a^m b^m e$. Això implica que $\{(c^i d c^i, c^i d c^i) \mid 1 \leq i \leq n\} \in \varphi(e a^m b^m e)$, per a tot $m > MAx(G)$; és a dir, $MPhi(G) \leq n$. Per tant, $MPhi_{ICC}(L_{in}(G_n)) = n$. $\square$

No discutirem aquí les qüestions de decidibilitat Q1 – Q5 per les noves mesures introduïdes, però considerarem un altre problema important, que és el de la millora simultània de dues, o més, mesures. En aquest cas, tenint dues mesures M_1, M_2 de la complexitat descriptiva de llenguatges d'una família F donada (generats per gramàtiques d'una classe donada X), seria molt útil trobar gramàtiques que fossin al mateix temps mínimes des dels punts de vista d'ambdues mesures (en qualsevol altre cas, apareix una qüestió delicada: ¿quina de les dues mesures M_1, M_2 és més important/expressiva

?). Desafortunadament, com era d'esperar, apareix aquí un principi general de treball: en molts casos, quan millorem un criteri, n'empitjorem un altre. En termes de gramàtiques i llenguatges és possible que, per a un llenguatge L donat, la gramàtica més econòmica per a L , des del punt de vista d' M_1 , no sigui la gramàtica més econòmica des del punt de vista d' M_2 , i inversament.

Definició 8.2.4 Formalment, per a una mesura $M : X \longrightarrow N$ de gramàtiques d'una classe X que generen els llenguatges d'una família F , definim (el conjunt de gramàtiques mínimes respecte a M que generen L):

$$M^{-1}(L) = \{G \in X \mid L(G) = L, M(G) = M_X(L)\}.$$

Dues mesures M_1, M_2 s'anomenen *incompatibles* si existeix un llenguatge $L \in F$ tal que

$$M_1^{-1}(L) \cap M_2^{-1}(L) = \emptyset$$

(les dues mesures no poden ser simultàniament minimitzades per al llenguatge L). En qualsevol altre cas, M_1 i M_2 s'anomenen *compatibles*.

Teorema 8.2.8 Cada dues d'aquestes mesures Ax, MAx, TAx són compatibles per a gramàtiques contextuais externes.

Demostració. Això és una conseqüència directa de (la demostració del) Teorema 8.2.3 de la pàgina 254: donat $L \in ECC$ podem construir $G = (V, B, C, \varphi)$ amb $B = Min_0(L)$ i per aquesta gramàtica G tenim $M(G) = M_{ECC}(L)$, per a tot $M \in \{Ax, MAx, LAx\}$. A més a més $Ax^{-1}(L) \cap MAx^{-1}(L) \cap TAx^{-1}(L) \neq \emptyset$, per a tot $L \in ECC$. $\square$

Per altra banda tenim

Teorema 8.2.9 Les mesures de tots els parells de sota són incompatibles per a gramàtiques contextuais externes:

- (i) (M_1, M_2) , amb $M_1 \in \{Ax, MAx, TAx\}$, $M_2 \in \{Con, MCon, TCon, Phi, MPhi\}$;
- (ii) (M_1, M_2) , amb $M_1 \in \{MCon, TCon\}$, $M_2 \in \{Con, Phi, MPhi\}$;
- (iii) (Con, Phi) , $(MPhi, Phi)$;
- (iv) $(MPhi, Con)$.

Demostració. (i) Prenem el llenguatge $L = \{a, aa\}$. Clarament, $Min_0(L) = \{a\}$ i, per tant, $Ax_{ECC}(L) = MAx_{ECC}(L) = TAx_{ECC}(L) = 1$. No obstant, qualsevol gramàtica G tal que $L_{ex}(G) = L$ i $Ax(G) = 1$, ha de contenir com a mínim un context; és a dir, $M(G) \geq 1$, per a tot $M \in \{Con, MCon, TCon, MPhi\}$, i $Phi(G) \geq 2$ (el context (u, v) , amb $uv = a$, usat per a generar aa a partir d' a pot ser aplicat a a , però no a aa). Per altra banda, $M_{ECC}(L) = 0$, per a $M \in \{Con, MCon, TCon\}$, i $Phi_{ECC}(L) = 1$, $MPhi_{ECC}(L) = 0$: tenim que $L = L_{ex}(G)$ per a $G = (\{a\}, \{a, aa\}, \emptyset)$.

(ii) Considerem el llenguatge

$$L = \{ba^{7n}, ba^{7n+1}, ba^{7n+2}, ba^{7n+4} \mid n \geq 1\}.$$

És generat per la gramàtica $G = (\{a, b\}, \{ba\}, \{(\lambda, a), (\lambda, a^3)\}, \varphi)$, amb

$$\varphi(ba^{7n+1}) = \{(\lambda, a), (\lambda, a^3)\}, n \geq 0,$$

$$\varphi(ba^{7n+4}) = \{(\lambda, a^3)\}, n \geq 0,$$

$$\varphi(ba^{7n}) = \{(\lambda, a)\}, n \geq 1,$$

$$\varphi(x) = \emptyset, \text{ en qualsevol altre cas.}$$

A més a més, $MCon_{ECC}(L) \leq 3$, $TCon_{ECC}(L) \leq 4$.

Per altra banda, $Con_{ECC}(L) = 1$, $Phi_{ECC}(L) = 1$, $MPhi_{ECC}(L) = 0$, ja que $L = L_{ex}(G')$ per a $G' = (\{a, b\}, \{ba, ba^2, ba^4, ba^7\}, \{(\lambda, a^7)\})$.

Considerem ara una gramàtica $G'' = (\{a, b\}, A, C, \varphi)$ tal que $L = L_{ex}(G'')$ i tenint $Con(G'') = 1$. Sigui $C = \{(u, v)\}$. Clarament, $u = \lambda$ i $v = a^r$. Distingirem set casos:

1. $r = 7k + 1$; aleshores $a^{7n+2} \Rightarrow_{ex} ua^{7n+2}v = a^{7(n+k)+3} \notin L$;
2. $r = 7k + 2$; aleshores $a^{7n+1} \Rightarrow_{ex} ua^{7n+1}v = a^{7(n+k)+3} \notin L$;
3. $r = 7k + 3$; aleshores $a^{7n+2} \Rightarrow_{ex} ua^{7n+2}v = a^{7(n+k)+5} \notin L$;
4. $r = 7k + 4$; aleshores $a^{7n+1} \Rightarrow_{ex} ua^{7n+1}v = a^{7(n+k)+5} \notin L$;
5. $r = 7k + 5$; aleshores $a^{7n+1} \Rightarrow_{ex} ua^{7n+1}v = a^{7(n+k)+6} \notin L$;
6. $r = 7k + 6$; aleshores $a^{7n} \Rightarrow_{ex} ua^{7n}v = a^{7(n+k)+6} \notin L$;
7. $r = 7k$.

Conseqüentment, $MCon(G'') \geq 7, TCon(G'') \geq 7$ (la qual cosa implica que $MCon, TCon$ són incompatibles amb Con).

Si $Phi(G'') = 1$ o $MPhi(G'') = 0$, aleshores G'' és una gramàtica contextual simple i, novament, no podem tenir a G'' contextos de la forma (λ, a^r) amb $r \neq 7k$. Conseqüentment, també Phi i $MPhi$ són incompatibles amb $MCon$ i $TCon$.

(iii) Considerem el llenguatge

$$L = \{a^n b a^n b^i \mid n \geq i, 0 \leq i \leq 2\}.$$

Tenim $L = L_{ex}(G)$ per a la gramàtica $G = (\{a, b\}, \{aba\}, \{(a, a), (\lambda, b), (\lambda, b^2)\}, \varphi)$, amb

$$\varphi(a^r b a^r) = \{(a, a), (\lambda, b), (\lambda, b^2)\}, r \geq 1,$$

$$\varphi(x) = \emptyset, \text{ en qualsevol altre cas.}$$

A més a més, $Phi_{ECC}(L) \leq 2$ (de fet, $Phi_{ECC}(L) = 2$, ja que $L \notin EC$).

Prenem ara una gramàtica $G' = (\{a, b\}, B, C, \varphi)$ amb $L = L_{ex}(G')$ i $Phi(G') = 2$. Clarament, $\varphi(a^n b a^n b^2) = \emptyset$. Per tant, l'altre conjunt de la forma $\varphi^{-1}(C''), C'' \subseteq C$ és $\varphi^{-1}(C)$; és a dir, per a $x \in \{a, b\}^*$, o bé $\varphi(x) = \emptyset$ o bé $\varphi(x) = C$. Com que cada cadena $a^n b a^n$ pertany a L , un context $(a^i, a^i), i \geq 1$, pertany a C . A partir d'aquesta cadena hem d'obtenir també $a^n b a^n b$ i, per tant, també un context $(a^i, a^i b)$ és present a C . Suposem ara que $a^n b a^n b^2$ és obtinguda d' $a^n b a^n b$ usant el context (λ, b) ; és a dir, $(\lambda, b) \in \varphi(a^n b a^n b)$. No

obstant, $(a^i, a^i) \notin \varphi(a^n b a^n b)$, la qual cosa implica que $\text{Phi}(G') \geq 3$, que és una contradicció. Conseqüentment, $a^n b a^n b^2$ és obtinguda a partir de $a^n b a^n$ usant el context (λ, b^2) . Aleshores $\text{Con}(G') \geq 3$ i $\text{MPhi}(G') \geq 3$.

No obstant, $\text{Con}_{ECC}(L) \leq 2$, $\text{MPhi}_{ECC}(L) \leq 2$, ja que $L = L_{ex}(G'')$ per a $G'' = (\{a, b\}, \{aba\}, \{(a, a), (\lambda, b)\}, \varphi)$, amb

$$\varphi(a^n b a^n) = \{(a, a), (\lambda, b)\}, n \geq 1,$$

$$\varphi(a^n b a^n b) = \{(\lambda, b)\}, n \geq 1,$$

$$\varphi(x) = \emptyset, \text{ en qualsevol altre cas.}$$

(iv) Prenem el llenguatge

$$L = \{a^n b a^n, a^n b a^n b, a^n b a^n b^3, a^n b a^n b^6 \mid n \geq 1\}.$$

Pot ser generat per la gramàtica $G = (\{a, b\}, \{aba\}, \{(a, a), (\lambda, b), (\lambda, b^2), (\lambda, b^3)\}, \varphi)$, amb

$$\varphi(a^n b a^n) = \{(a, a), (\lambda, b)\}, n \geq 1,$$

$$\varphi(a^n b a^n b) = \{(\lambda, b^2)\}, n \geq 1,$$

$$\varphi(a^n b a^n b^3) = \{(\lambda, b^3)\}, n \geq 1,$$

$$\varphi(x) = \emptyset, \text{ en qualsevol altre cas.}$$

Per tant, $\text{MPhi}_{ECC}(L) \leq 2$.

Considerem una gramàtica $G' = (\{a, b\}, B, C, \varphi)$ tal que $L = L_{ex}(G')$ i $\text{MPhi}(G') = 2$. És fàcil de veure que existeixen a C els contextos (a^i, a^i) , $i \geq$

1, i (λ, b) , i ambdós poden ser aplicats a cadenes $a^n b a^n$. Si alguna cadena $a^n b a^n b^3$ és obtinguda a partir de $a^n b a^n$ mitjançant l'addició del context (λ, b^3) , se segueix que $\text{card}(\varphi(a^n b a^n)) \geq 3$, la qual cosa implica que $M\text{Phi}(G') > 2$. Conseqüentment, existeix un context (λ, b^2) a C , tal que $(\lambda, b^2) \in \varphi(a^n b a^n b)$. Similarment, ha d'existir un context (λ, b^3) a C ; és a dir, $\text{Con}(G') \geq 4$.

No obstant, $\text{Con}_{ECC}(L) \leq 3$, ja que $L = L_{ex}(G'')$ per a $G'' = (\{a, b\}, \{aba\}, \{(a, a), (\lambda, b), (\lambda, b^3)\}, \varphi)$, amb

$$\varphi(a^n b a^n) = \{(a, a), (\lambda, b), (\lambda, b^3)\}, n \geq 1,$$

$$\varphi(a^n b a^n b^3) = \{(\lambda, b^3)\}, n \geq 1,$$

$$\varphi(x) = \emptyset, \text{ en qualsevol altre cas.}$$

Això completa la demostració de la incompatibilitat de Con i $M\text{Phi}$. □

El problema de la incompatibilitat de les mesures $M\text{Con}$, $T\text{Con}$ per a gramàtiques contextuais externes resta encara *per resoldre*. Pel que fa a les gramàtiques contextuais internes, a (Georgescu, 1996) és provat que, per aquestes gramàtiques, cada dues d'aquestes mesures Ax , MAx , TAx són compatibles. Però cada parell (M_1, M_2) és incompatible, per a $M_1 \in \{Ax, MAx, TAx\}$, $M_2 \in \{\text{Con}, M\text{Con}, T\text{Con}, \text{Phi}, M\text{Phi}\}$, així com per a $M_1 \in \{\text{Con}, \text{Phi}\}$, $M_2 \in \{M\text{Con}, T\text{Con}\}$. Per als altres parells (M_1, M_2) la qüestió està encara oberta.

8.3 L'ambigüitat

Per a qualsevol classe de gramàtiques, un natural (i important) problema és el de l'ambigüitat sintàctica: donada una gramàtica, existeixen cadenes en el llenguatge generat que tinguin dues descripcions sintàctiques? “Descripcions sintàctiques” significa, intuïtivament, “derivacions diferents”. Per a gramàtiques de Chomsky lliures del context la noció és clara: una derivació (diferent) és descrita mitjançant un arbre de derivació i, per tant, una gramàtica lliure del context, és ambigua quan existeix una cadena que pot ser produïda per derivacions que tinguin diferents arbres de derivació.

A la vista del Lema 5.3.2 de la pàgina 200, la demostració del qual ens proporciona un pas directe entre una gramàtica contextual externa sense selecció i una gramàtica lineal (mínima), i a la inversa, també pel cas de les gramàtiques contextuais externes podem definir fàcilment l'ambigüitat: descrivim una derivació per a l'axioma i per a la seqüència de contextos usada en la derivació. L'anomenarem *seqüència de control*. Si existeix una cadena a $L_{ex}(G)$ que pot ser generada a G mitjançant dues derivacions, descrites mitjançant dues seqüències de control diferents, aleshores direm que G és *ambigua*. Com és usual, si totes les gramàtiques que generen un llenguatge donat L són ambigües, aleshores L s'anomena *inherentment ambigu*; en qualsevol altre cas L és no-ambigu. És clar que aquestes nocions han de ser definides respecte a una classe de gramàtiques donada. Els llenguatges inherentment ambigus, respecte a unes classes de gramàtiques determinades, poden ser no-ambigus

respecte a altres classes de gramàtiques. Això succeeix també en el nostre cas.

Teorema 8.3.1 Existeixen llenguatges inherentment ambigus respecte a les gramàtiques contextuals externes sense selecció, però tots els llenguatges de ECC són no-ambigus respecte a les gramàtiques contextuals externes amb selecció.

Demostració. La primera afirmació es demostra mitjançant el llenguatge

$$L = \{a^n b a^m \mid n \geq m \geq 0\}.$$

Tota gramàtica $G = (\{a, b\}, B, C)$ tal que $L = L_{ex}(G) = L$ ha de contenir un context $(a^i, a^i), i \geq 1$, necessari per tal de generar les cadenes de la forma $a^n b a^n$, amb n arbitràriament gran; així com un context $(a^j, \lambda), j \geq 1$, necessari per tal de generar les cadenes $a^n b$, amb n arbitràriament gran. Prenem un axioma $a^p b a^q$ de A . La cadena $a^{p+i+j} b a^{q+i}$ té dues derivacions a G ,

$$\begin{aligned} a^p b a^q &\Rightarrow_{ex} a^{p+i} b a^{q+i} \Rightarrow_{ex} a^{p+i+j} b a^{q+i}, \\ a^p b a^q &\Rightarrow_{ex} a^{p+j} b a^q \Rightarrow_{ex} a^{p+i+j} b a^{q+i}, \end{aligned}$$

i, per tant, G és ambigua i L és inherentment ambigua.

Considerem ara un gramàtica contextual amb selecció, $G = (V, B, C, \varphi)$ i considerem el següent *procediment* (el qual no és un algorisme):

Fixem un ordre de les cadenes a V^* , per exemple, el lexicogràfic, $V^* = \{x_1, x_2, \dots\}$. El conjunt $E_0 = \emptyset$. Per a $i = 1, 2, \dots$ considerem la cadena x_i .

Si existeixen $(u, v), (u', v') \in C$, $(u, v) \neq (u', v')$, $(u, v) \in \varphi(x_i)$, $(u', v') \in \varphi(x_i)$, i $ux_iv = u'x_iv$, aleshores esborrem un dels contextos $(u, v), (u', v')$ de $\varphi(x_i)$. Continuem d'aquesta manera fins obtenir un conjunt $\varphi'(x_i)$ tal que $ux_iv \neq u'x_iv'$, per a tot $(u, v), (u', v') \in \varphi'(x_i)$. Per a $j > i$ posem $\varphi'(x_j) = \varphi(x_j)$. Definim

$$\text{següent}(x_i) = \{ux_iv \in L \mid (u, v) \in \varphi'(x_i)\}.$$

Si existeix $(u, v) \in \varphi'(x_i)$ tal que $ux_iv \in \text{següent}(x_i) \cap L_{ex}(G_{i-1})$, on $G_{i-1} = (V, E_{i-1}, C, \varphi')$, aleshores esborrem (u, v) de $\varphi'(x_i)$. Continuem d'aquesta manera fins a obtenir un conjunt $\varphi''(x_i)$ tal que $ux_iv \notin L_{ex}(G_{i-1})$, per a tot $(u, v) \in \varphi''(x_i)$. Per a $j > i$ posarem $\varphi''(x_j) = \varphi(x_j)$.

Definim $E_i = E_{i-1} \cup \{x_i\}$ i continuem el procediment.

Obtenim una gramàtica $G' = (V, B, C, \varphi''')$ tal que $L_{ex}(G) = L_{ex}(G')$, però per a cada $x \in L_{ex}(G')$ existeix exactament una derivació a G' . A més a més, G' és no-ambigua i, per tant, $L_{ex}(G)$ és no-ambigu. $\square$

Més inesperadament, tenim

Teorema 8.3.2 L'ambigüitat d'una gramàtica contextual externa, amb o sense selecció, és indecidible.

Demostració. Clarament, és suficient provar l'afirmació per a gramàtiques amb selecció.

Siguin $(x_1, \dots, x_n)$ i $(y_1, \dots, y_n)$ dues n -tuples de cadenes no-buides sobre

$\{a, b\}$ i considerem la gramàtica lineal

$$\begin{aligned} G &= (\{S\}, \{a, b, c\}, S, P), \\ P &= \{S \rightarrow ba^i Sx_i \mid 1 \leq i \leq n\} \\ &\cup \{S \rightarrow ba^i Sy_i \mid 1 \leq i \leq n\} \\ &\cup \{S \rightarrow c\}. \end{aligned}$$

A (Greibach, 1963) és provat que l'ambigüitat d'aquesta gramàtica és indecidible (la qüestió es redueix a una variant del problema de correspondència de Post, el qual és també indecidible). Com en la demostració del Lema 5.3.2 de la pàgina 200, podem associar a G la gramàtica contextual

$$G' = (\{a, b, c\}, \{c\}, \{(ba^i, x_i), (ba^i, y_i) \mid 1 \leq i \leq n\}),$$

que és ambigua si, i només si, G és ambigua. Conseqüentment, l'ambigüitat de G' és indecidible. $\square$

Per a gramàtiques contextuais internes, la definició de l'ambigüitat no és tan simple, ja que no queda clar quina informació hem de considerar per tal de descriure una *derivació* diferent en un determinat cas. Una derivació queda identificada amb precisió mitjançant l'axioma, els contextos usats i els *emplaçaments* on els contextos són adjuntats (si és així, també els selectors queden identificats). Però no tota aquesta informació és essencial des del punt de vista sintàctic. Per exemple, l'ordre d'ús dels contextos sembla ser no rellevant. Per tant, podem definir diferents nivells d'ambigüitat, dependent de la informació usada per a descriure una derivació. Discutirem aquí cinc variants.

Donada una gramàtica contextual $G = (V, B, C, \varphi)$ i una derivació interna

$$\delta : w_1 \Rightarrow_{in} w_2 \Rightarrow_{in} \dots \Rightarrow_{in} w_m, m \geq 1,$$

tal que

$$w_1 \in B,$$

$$w_i = x_{1,i}x_{2,i}x_{3,i}, \quad x_{1,i}, x_{2,i}, x_{3,i} \in V^*,$$

$$w_{i+1} = x_{1,i}u_i x_{2,i}v_i x_{3,i}, \text{ per a}$$

$$(u_i, v_i) \in \varphi(x_{2,i}), 1 \leq i \leq m-1,$$

la seqüència

$$w_1, (u_1, v_1), \dots, (u_{m-1}, v_{m-1})$$

s'anomena *la seqüència de control* associada a δ .

És clar que la seqüència de control no identifica la derivació, ja que els contextos poden ser adjuntats en posicions diferents. Per tant, anomenarem a la seqüència

$$w_1, x_{1,1}(u_1)x_{2,1}(v_1)x_{3,1}, x_{1,2}(u_2)x_{2,2}(v_2)x_{3,2}, \dots, \\ x_{1,m-1}(u_{m-1})x_{2,m-1}(v_{m-1})x_{3,m-1}$$

una *descripció* de δ (els parèntesis identifiquen d'una manera precisa els contextos usats i l'emplaçament on els adjuntem).

Un cas intermedi és especificar només els contextos i els seus selectors, i no pas la posició on apareixen. Aleshores, anomenarem la seqüència

$$w_1, ((u_1, v_1), x_{2,1}), ((u_2, v_2), x_{2,2}), \dots, ((u_{m-1}, v_{m-1}), x_{2,m-1})$$

una *seqüència de control completa*.

Si tenim en compte només els contextos usats (i els selectors), i no pas l'ordre d'ús dels mateixos, aleshores obtenim la *seqüència de control no ordenada* i la *seqüència de control no ordenada completa*, respectivament. Poden ser donades formalment com a seqüències d'enters, especificant el nombre de vegades que cada context – i selector – és usat.

Una gramàtica contextual G s'anomena 1-ambigua si existeixen dues derivacions a G que tenen diferents seqüències de control no ordenades que generen la mateixa cadena.

Reemplaçant en aquesta definició les paraules “seqüències de control no ordenades” per “seqüències de control no ordenades completes”, “seqüències de control”, “seqüències de control completes”, i “descripcions”, respectivament, direm que la gramàtica G és 2-, 3-, 4-, 5-ambigua, respectivament.

Per a una millor claredat, escriurem, de manera més concisa:

1-ambigüitat = dues seqüències de control no ordenades,

2-ambigüitat = dues seqüències de control no ordenades completes ,

3-ambigüitat = dues seqüències de control,

4-ambigüitat = dues seqüències de control completes,

5-ambigüitat = dues descripcions.

Una gramàtica que és no t -ambigua, per a algun $t = 1, 2, 3, 4, 5$, s'anomena t -no-ambigua. Un llenguatge $L \in ICC(F)$ és *inherentment t -ambigu* si cada gramàtica G (amb F selecció) que genera L és t -ambigua. Un llenguatge pel qual existeix una t -no-ambigua gramàtica s'anomena *t -no-ambigu*.

És clar que, en molts casos, donat un llenguatge t -no-ambigu, podem també trobar gramàtiques t -ambigües que el generin. Per exemple, $L = \{a^n b a^n \mid n \geq 1\}$ és 5-no-ambigu, ja que pot ser generat per la gramàtica

$$G = (\{a, b\}, \{aba\}, (\{b\}, \{(a, a)\}))$$

(existeix un únic axioma i un únic context, el qual pot ser adjuntat en un únic emplaçament). No obstant, $L = L_{in}(G')$, també, per a

$$G' = (\{a, b\}, \{aba\}, (\{b, ab\}, \{(a, a)\})).$$

La gramàtica G' és no 3-ambigua (en general, les gramàtiques amb només un axioma i només un context són no 3-ambigües), però és 5-ambigua: la cadena $a^2 b a^2$ pot ser obtinguda a partir d' aba de dues maneres:

$$aba \Rightarrow_{in} a(a)b(a)a, \quad aba \Rightarrow_{in} (a)ab(a)a.$$

Començarem investigant l'existència de llenguatges inherentment ambigus.

Tot llenguatge inherentment t -ambigu és inherentment s -ambigu, per a $(t, s) \in \{(1, 2), (1, 3), (2, 4), (3, 4), (4, 5)\}$. L'invers no és (sempre) cert:

Teorema 8.3.3 Existeixen llenguatges inherentment 5-ambigus (respecte a gramàtiques contextuais amb selecció arbitrària) que són 4-no-ambigus (fins i tot respecte a gramàtiques contextuais amb selecció finita).

Demostració. Considerem el llenguatge

$$L = \{a^n b a^m \mid n, m \geq 1\}.$$

Pot ser generat per la gramàtica

$$G = (\{a, b\}, \{aba\}, (\{a\}, \{(\lambda, a)\})).$$

Clarament, aquesta gramàtica no és 4-ambigua, ja que L no és inherentment 4-ambigu (fins i tot respecte a les classes de gramàtiques contextuais més restringides, aquelles que tenen selecció finita).

No obstant, aquest llenguatge és inherentment 5-ambigu. En aquest cas, considerem una gramàtica $G = (V, B, C, \varphi)$ que generi L . Cada axioma de B pot ser de la forma $a^i b a^j, i, j \geq 1$ i, per tant, els contextos de C usats de fet en la derivació de cadenes d' L han de ser de la forma (u, v) , amb $u, v \in a^*$.

Suposem que existeix un context (u, v) i una cadena $a^r, r \geq 0$, tal que $(u, v) \in \varphi(a^r)$. Aleshores la gramàtica és ambigua: la cadena $a^{r+1}ba$ pertany a L i, per tant, existeix una derivació $w \Rightarrow_{in}^* a^{r+1}ba$ a G , per a algun $w \in B$. Aquesta derivació pot ser continuada usant el context (u, v) en dos

emplaçaments:

$$\begin{aligned} w &\Longrightarrow_{in}^* a^r aba \Longrightarrow_{in} ua^r vaba \\ w &\Longrightarrow_{in}^* aa^r ba \Longrightarrow_{in} aua^r vba. \end{aligned}$$

Ja que $u, v \in a^*$, hem obtingut la mateixa cadena i, per tant, G és 5-ambigua.

A més a més, tot context (u, v) usat en una derivació ha de ser associat amb cadenes de la forma $a^r ba^s$, $r, s \geq 0$ (és a dir, $(u, v) \in \varphi(a^r ba^s)$). Pel fet que totes les cadenes de la forma $a^k ba$ pertanyen a L , existeix un context de la forma (a^i, λ) , $i \geq 1$. Ja que totes les cadenes de la forma aba^k pertanyen a L , existeix un context de la forma (λ, a^j) , $j \geq 1$. Considerem dues cadenes $a^r ba^s, a^p ba^q$ tal que $(a^i, \lambda) \in \varphi(a^r ba^s)$ i $(\lambda, a^j) \in \varphi(a^p ba^q)$. Prenem la cadena $a^{r+p} ba^{s+q}$. Ja que aquesta cadena pertany a L , existeix una derivació $w \Longrightarrow_{in}^* a^{r+p} ba^{s+q}$ a G , per a algun $w \in B$. Aquesta derivació pot ser continuada usant, en primer lloc, (a^i, λ) i després (λ, a^j) , o a l'inversa:

$$\begin{aligned} w &\Longrightarrow_{in}^* a^p a^r ba^s a^q \Longrightarrow_{in} a^p a^i a^r ba^s a^q = a^{i+r} a^p ba^q a^s \\ &\Longrightarrow_{in} a^{i+r} a^p ba^q a^j a^s = a^{i+r+p} ba^{q+j+s}, \\ w &\Longrightarrow_{in}^* a^r a^p ba^q a^s \Longrightarrow_{in} a^r a^p ba^q a^j a^s = a^p a^r ba^s a^{q+j} \\ &\Longrightarrow_{in} a^p a^i a^r ba^s a^{q+j} = a^{p+i+r} ba^{s+q+j}. \end{aligned}$$

Hem obtingut la mateixa cadena i, per tant, novament, la gramàtica és 5-ambigua.

Observem que en l'argumentació donada anteriorment el tipus de selector

no juga cap paper i, per tant, G pot ser de qualsevol tipus. $\square$

A més a més, tenim

Teorema 8.3.4 Existeixen llenguatges inherentment 4-ambigus (respecte a gramàtiques contextuais amb selecció arbitrària) que són 2- i 3- no-ambigus (fins i tot respecte a gramàtiques amb selecció finita).

Demostració. Prenem el llenguatge

$$L = \{a^n b a^n a^m c a^m \mid n, m \geq 0\}.$$

Pot ser generat per la gramàtica

$$G = (\{a, b, c\}, \{bc\}, (\{b, c\}, \{(a, a)\})).$$

Aquesta gramàtica no és ni 2- ni 3-ambigua (conté només un context) i, per tant, L és 2- i 3-no-ambigua.

Considerem ara una gramàtica arbitrària $G' = (\{a, b, c\}, B, C, \varphi)$ que generi L . Segons la forma que tinguin les cadenes d' L , és fàcil de veure que existeixen contextos de la forma $(a^i, a^i), (a^j, a^j), i, j \geq 1$, tal que $(a^i, a^i) \in \varphi(a^k b a^l)$, per a algun $k, l \geq 0$, i $(a^j, a^j) \in \varphi(a^p c a^q)$, per a algun $p, q \geq 0$. Quan generem una cadena $a^n b a^n a^m c a^m$, amb valors d' n, m suficientment grans, hem d'usar els dos contextos mencionats anteriorment; ja que els seus selectors poden ser disjunts (no coincidents), l'ordre d'ús d'aquests contextos pot ser canviat. D'aquesta manera obtenim dues derivacions les seqüències completes ordenades de les

quals són diferents (és possible tenir $i = j$ i, per tant, el mateix context; però $b \neq c$ i, per tant, els selectors són diferents). Això vol dir que G' és 4-ambigua, la qual cosa implica que el llenguatge L és inherentment 4-ambigu. $\square$

Teorema 8.3.5 Existeixen llenguatges inherentment 3-ambigus (respecte a gramàtiques contextuais amb selecció arbitrària) que són 1- i 2- no ambigües.

Demostració. Considerem el llenguatge

$$L = \{a^n b^m \mid n, m \geq 1\}$$

i una gramàtica arbitrària $G = (\{a, b\}, B, C, \varphi)$ que generi L . Tota cadena $a^k b$ pertany a L ; aquesta cadena només pot ser generada a partir d'una cadena $a^p b$, $p \geq 1$ i usant contextos (u, v) amb $u, v \in a^*$. Si $(u, v) \in \varphi(a^t)$, aleshores v pot ser diferent de λ ; si $(u, v) \in \varphi(a^t b)$, aleshores $v = \lambda$. Similarment, ja que cada cadena ab^k pertany a L , tenim axiomes de la forma ab^p , $p \geq 1$, i contextos (u', v') amb $u', v' \in b^*$. Si $(u', v') \in \varphi(b^t)$, aleshores u' pot ser no-buit, però si $(u', v') \in \varphi(ab^t)$, aleshores $u' = \lambda$.

Poden aparèixer diferents casos combinant les possibilitats discutides anteriorment, pels contextos (u, v) , (u', v') i els seus selectors. Prenent una z suficientment llarga de L (aquesta és una cadena a la qual (u, v) i (u', v') poden ser adjuntats). Aleshores, després d'afegir (u, v) a z , també podem afegir-li (u', v') , i a la inversa. Després d'afegir (u', v') podem també afegir (u, v) . Les cadenes obtingudes són les mateixes ((u, v) contribueix al nombre d'ocurrències del símbol a , (u', v') contribueix al nombre d'ocurrències del símbol b), però

les dues derivacions tenen diferents descripcions (l'ordre de $(u, v), (u', v')$ és diferent en una i altra derivació). Conseqüentment, G és 3-ambigua.

A continuació veurem un dels múltiples possibles casos suggerits anteriorment: prendrem $(a^i, \lambda) \in \varphi(a^t b)$ i $(\lambda, b^j) \in \varphi(ab^s)$. Aleshores, per a $z = a^t b^s$ tenim

$$\begin{aligned} w &\Rightarrow_{in}^* a^t b^s \Rightarrow_{in} a^i a^t a^s \Rightarrow_{in} a^i a^t b^s b^j, \\ w &\Rightarrow_{in}^* a^t b^s \Rightarrow_{in} a^t b^s b^j \Rightarrow_{in} a^i a^t b^s b^j, w \in B. \end{aligned}$$

Les cadenes obtingudes són les mateixes, però les derivacions són diferents.

El tipus de selecció no juga cap paper en l'argument anterior.

L'anterior llenguatge L pot ser generat per la gramàtica

$$G = (\{a, b\}, \{ab\}, (\{a\}, \{(a, \lambda)\}), (\{b\}, \{(b, \lambda)\})).$$

Per tal de generar una cadena $a^n b^m$ hem d'usar el context (a, λ) i el selector a , exactament, $n - 1$ vegades; i el context (b, λ) i el selector b , exactament, $m - 1$ vegades. Conseqüentment, la gramàtica G és 2-no-ambigua (i, per tant, també 1-no-ambigua). $\square$

No se sap si existeixen o no llenguatges que siguin inherentment 2-ambigus i 1- o 3-no-ambigus, o llenguatges inherentment 1-ambigus. Aquest últim *problema no resol*t és particularment interessant. Podem *conjecturar* que aquests llenguatges existeixen. Una resposta parcial (que suporta aquesta conjectura) és el següent resultat.

Teorema 8.3.6 Existeixen llenguatges inherentment 1-ambigus respecte a gramàtiques contextuals internes sense selecció.

Demostració. Considerem el llenguatge

$$L = \{x \in \{a, b\}^* \mid |x|_a = |x|_b\}.$$

Prenem una gramàtica (sense selecció) $G = (V, B, C)$ que generi L . Tota cadena de la forma $a^n b^n, n \geq 1$, pertany a L . Per tal de generar aquestes cadenes hem d'usar un context d'una de les formes següents:

$$\begin{aligned} 1) \quad & (a^i, b^i), (a^i b^{i_1}, b^{i_2}), \quad i = i_1 + i_2, i_1, i_2 \geq 0, \text{ o} \\ & (a^{i_1}, a^{i_2} b^i), \quad i = i_1 + i_2, i_1, i_2 \geq 0, \quad i \geq 1. \end{aligned}$$

Similarment, per tal de generar les cadenes $b^n a^n$, amb $n \geq 1$ arbitràriament gran, el conjunt C ha de contenir un context d'una de les formes següents:

$$\begin{aligned} 2) \quad & (b^j, a^j), (b^j a^{j_1}, a^{j_2}), \quad j = j_1 + j_2, j_1, j_2 \geq 0, \text{ o} \\ & (b^{j_1}, b^{j_2} a^j), \quad j = j_1 + j_2, j_1, j_2 \geq 0, \quad j \geq 1. \end{aligned}$$

Per a cadascun dels $3 \times 3 = 9$ possibles casos, considerem la cadena

$$w = a^{ij} b^{ij} a^{ij} b^{ij}.$$

Pot ser escrit com a

$$w = (\underbrace{a^i \dots a^i}_{j \text{ vegades}} \underbrace{b^i \dots b^i}_{j \text{ vegades}})^2,$$

i, per tant, w pot ser obtinguda començant per λ (que pertany a L) i adjuntant només contextos del tipus 1 anterior. Podem escriure també

$$w = (a^i)^j (b^{ij} a^{ij}) (b^i)^j.$$

La cadena $w' = a^{ij} b^{ij}$ pertany a L i, per tant, w pot ser obtinguda a partir de w' usant només contextos del tipus 2 anterior, ja que

$$b^{ij} a^{ij} = \underbrace{b^j \dots b^j}_{i \text{ vegades}} \underbrace{a^j \dots a^j}_{i \text{ vegades}}.$$

Una cadena $a^i \dots a^i b^i \dots b^i$ pot ser obtinguda mitjançant, en primer lloc, produint l'extern $a^i b^i$ – i aquest és un context del tipus 1 – i després insertant $a^i b^i$ al mig, i així successivament. El mateix per a les cadenes $b^j \dots b^j a^j \dots a^j$.

Conseqüentment, G és 1-ambigua.

Ens falta provar que L pot ser generat mitjançant una gramàtica contextual interna sense selecció. Això és obvi: $L = L_{in}(G)$, per a $G = (\{a, b\}, \{\lambda\}, \{(a, b), (b, a)\})$. $\square$

Observem que els llenguatges usats en les demostracions dels Teoremes 8.3.3 de la pàgina 279 i 8.3.5 de la pàgina 280 són regulars en el sentit de Chomsky, però els llenguatges dels Teoremes 8.3.4 de la pàgina 282 i 8.3.6 de la pàgina 285 són lliures del context i no lineals.

El fet que hi hagi llenguatges regulars que siguin inherentment t -ambigus, per a diversos valors de t , contrasta amb el fet que els llenguatges regulars

siguin no-ambigus quan són generats mitjançant gramàtiques regulars. Aquest resultat pot ser interpretat en el sentit que la noció de t -ambigüitat, per a $t = 3, 4, 5$, és massa restrictiva. La mateixa conclusió pot ser deduïda també a partir del següent resultat, de nou contrastant amb la situació existent en els llenguatges lliures del context: tot llenguatge lliure del context $L \subseteq V^*$, fins i tot si és inherentment ambigu, és un subconjunt d'un llenguatge no-ambigu, és a dir, un conjunt regular, V^* .

Teorema 8.3.7 Existeixen llenguatges (regulars) L que són inherentment 3-ambigus i de manera que tot superllenguatge $L', L \subseteq L'$, és també inherentment 3-ambigu.

Demostració. Prenem el llenguatge

$$L = \{a^n \mid n \geq 1\} \cup \{b^n \mid n \geq 1\} \cup \{a^n b^m \mid n, m \geq 1\}.$$

Considerem un llenguatge $L' \subseteq \{a, b\}^*$ tal que $L \subseteq L'$ i prenem $G = (\{a, b\}, B, C, \varphi)$ que generi L' . Ja que $a^n \in L'$, per a tot $n \geq 1$, existeix un context $(a^{i_1}, a^{i_2}) \in \varphi(a^{k_1})$, per a algun $i_1 + i_2 \geq 1, k_1 \geq 0$. Pel fet que $b^n \in L'$, per a tot $n \geq 1$, existeix un context $(b^{j_1}, b^{j_2}) \in \varphi(b^{k_2})$, per a algun $j_1 + j_2 \geq 1, k_2 \geq 0$. La cadena $a^{k_1} b^{k_2}$ pertany a L' i, per tant, existeix una derivació $w \Rightarrow_{in}^* a^{k_1} b^{k_2}$ a G , per a algun $w \in B$. Aquesta derivació pot ser continuada amb (a^{i_1}, a^{i_2}) i després amb (b^{j_1}, b^{j_2}) , o a l'inversa. En ambdós casos obtenim la mateixa cadena, $a^{k_1+i_1+i_2} b^{k_2+j_1+j_2}$, la qual cosa significa que G és 3-ambigua. Per tant, L' (i també L) és inherentment 3-ambigu. $\square$

És clar que l'anterior resultat no és cert per a 1- i 2-ambigüitat: V^* és 1- i 2-no-ambigu, ja que pot ser generada per la gramàtica $G = (V, \{\lambda\}, (\{\lambda\}, \{(\lambda, a) \mid a \in V\}))$.

Com era d'esperar (aquest és també el cas per a gramàtiques lliures del context i per a gramàtiques contextuais externes – veure el Teorema 8.3.2 de la pàgina 275); l'important problema de trobar un algorisme capaç de dir-nos si una gramàtica contextual arbitrària és o no és ambigua no té solució.

Teorema 8.3.8 L'ambigüitat (de qualsevol tipus t , $t = 1, 2, 3, 4, 5$) de les gramàtiques contextuais internes (fins i tot amb selecció finita) és indecidible.

Demostració. Prenem dues n -tuples de cadenes no-buides sobre el vocabulari $\{a, b\}$ donades arbitràriament, és a dir, $(x_1, \dots, x_n), (y_1, \dots, y_n)$, i construïm la gramàtica contextual

$$G = (\{a, b, c, d, e\}, \{cccdc, ceccc\}, (S_1, C_1), (S_2, C_2), (S_3, C_3), (S_4, C_4)),$$

$$(S_1, C_1) = (\{cccdc\}, \{(x_i, a^i b) \mid 1 \leq i \leq n\}),$$

$$(S_2, C_2) = (\{ceccc\}, \{(y_i, a^i b) \mid 1 \leq i \leq n\}),$$

$$(S_3, C_3) = (\{ccdca\}, \{(e, \lambda)\}),$$

$$(S_4, C_4) = (\{acecc, bcecc\}, \{(\lambda, d)\}).$$

A l'axioma $cccdc$ només podem afegir-li contextos de C_1 . Després d'usar com a mínim un d'aquests contextos, podem usar el context de C_3 (la cadena

$ccdc$ és present). Per tant, tenim

$$\begin{aligned} ccdc &\Rightarrow_{in}^* x_{i_1} \dots x_{i_k} ccdca^{i_k}b \dots a^{i_1}b \\ &\Rightarrow_{in} x_{i_1} \dots x_{i_k} ceccdca^{i_k}b \dots a^{i_1}b = w_1, \end{aligned}$$

per a $k \geq 1, 1 \leq i_j \leq n, 1 \leq j \leq k$. L'ús del parell (S_3, C_3) pot ser iterat:

$$w_1 \Rightarrow_{in}^* x_{i_1} \dots x_{i_k} ce^r ccdca^{i_k}b \dots a^{i_1}b = w_2,$$

per a $r \geq 2$. Després d'usar dues vegades la producció (S_3, C_3) , no pot ser usada cap altra producció $(S_i, C_i), i \neq 3$. En una cadena de la forma w_1 (després d'un ús de (S_3, C_3)) podem també usar (S_4, C_4) :

$$w_1 \Rightarrow_{in}^* x_{i_1} \dots x_{i_k} ceccd^s ca^{i_k}b \dots a^{i_1}b = w_3,$$

per a $s \geq 2$. Per a aquesta cadena no podem usar les produccions $(S_i, C_i), i \neq 4$. Finalment, per a una cadena de la forma w_1 , no podem usar produccions del tipus $(S_1, C_1), (S_2, C_2)$.

Similarment, començant per $ceccc$ podem obtenir (usant com a mínim una vegada la producció (S_2, C_2) i després (S_4, C_4))

$$\begin{aligned} ceccc &\Rightarrow_{in}^* y_{j_1} \dots y_{j_l} ceccca^{j_l}b \dots a^{j_1}b \\ &\Rightarrow_{in} y_{j_1} \dots y_{j_l} ceccdca^{j_l}b \dots a^{j_1}b = w_4, \end{aligned}$$

per a $l \geq 1, 1 \leq j_p \leq n, 1 \leq p \leq l$. L'ús de (S_4, C_4) pot ser iterat:

$$w_4 \Rightarrow_{in}^* y_{j_1} \dots y_{j_l} ceccd^{s'} ca^{j_l}b \dots a^{j_1}b = w_5,$$

per a $s' \geq 2$. Per a aquesta cadena no podem usar les produccions $(S_i, C_i), i \neq$

4. Per a una cadena del tipus w_4 podem usar també (S_3, C_3) :

$$w_4 \Rightarrow_{in}^* y_{j_1} \dots y_{j_l} c e^{r'} c d c a^{j_l} b \dots a^{j_1} b = w_6,$$

per a $r' \geq 2$. Cap de les produccions $(S_1, C_1), (S_2, C_2)$ pot ser aplicada a una cadena de la forma w_4 .

Ara, comparant la forma de les cadenes dels sis tipus, $w_1 - w_6$, veiem que les úniques igualtats possibles $z_1 = z_2$ apareixen en els següents casos:

(A) z_1 de la forma w_1 i z_2 de la forma w_4 . Això vol dir

$$z_1 = x_{i_1} \dots x_{i_k} c e c c d c a^{i_k} b \dots a^{i_1} b,$$

$$z_2 = y_{j_1} \dots y_{j_l} c e c c d c a^{j_l} b \dots a^{j_1} b,$$

i $z_1 = z_2$ implica que $k = l, i_1 = j_1, \dots, i_k = j_k$ (ja que sinó no podem tenir $a^{i_k} b \dots a^{i_1} b = a^{j_l} b \dots a^{j_1} b$) i $x_{i_1} \dots x_{i_k} = y_{i_1} \dots y_{i_k}$ i, per tant, existeix una solució al problema de correspondència de Post respecte a les n -tuples $(x_1, \dots, x_n), (y_1, \dots, y_n)$.

(B) z_1 de la forma w_2 i z_2 de la forma w_6 . Això vol dir

$$z_1 = x_{i_1} \dots x_{i_k} c e^r c c d c a^{i_k} b \dots a^{i_1} b,$$

$$z_2 = y_{j_1} \dots y_{j_l} c e^{r'} c c d c a^{j_l} b \dots a^{j_1} b,$$

i $z_1 = z_2$ implica que $k = l, r = r', i_1 = j_1, \dots, i_k = j_k$ (ja que sinó no podem tenir $a^{i_k} b \dots a^{i_1} b = a^{j_l} b \dots a^{j_1} b$) i $x_{i_1} \dots x_{i_k} = y_{i_1} \dots y_{i_k}$. Novament, el problema

de la correspondència de Post respecte a les n -tuples $(x_1, \dots, x_n), (y_1, \dots, y_n)$ té una solució.

(C) z_1 de la forma w_3 i z_2 de la forma w_5 . Això vol dir

$$\begin{aligned} z_1 &= x_{i_1} \dots x_{i_k} ceccd^s ca^{i_k} b \dots a^{i_1} b, \\ z_2 &= y_{j_1} \dots y_{j_l} ceccd^{s'} ca^{j_l} b \dots a^{j_1} b, \end{aligned}$$

i $z_1 = z_2$ implica $k = l, s = s', i_1 = j_1, \dots, i_k = j_k$ (ja que sinó no podem tenir $a^{i_k} b \dots a^{i_1} b = a^{j_l} b \dots a^{j_1} b$) i $x_{i_1} \dots x_{i_k} = y_{j_1} \dots y_{j_l}$. Novament, el problema de la correspondència de Post respecte a les n -tuples $(x_1, \dots, x_n), (y_1, \dots, y_n)$ té una solució.

En tots els casos, l'ambigüitat de G significa 1-ambigüitat (ja que comencem a partir d'axiomes diferents per a generar les cadenes considerades w_1, w_2) la qual cosa implica l'existència d'una solució per al problema de la correspondència de Post per a les n -tuples x, y de cadenes considerades.

Inversament, si el problema de correspondència de Post per a $(x_1, \dots, x_n), (y_1, \dots, y_n)$ té una solució i, per tant, existeixen $i_1, \dots, i_k$ tal que $x_{i_1} \dots x_{i_k} = y_{i_1} \dots y_{i_k}$ aleshores, clarament, G és 1-ambigua. Per exemple, podem trobar una cadena

$$z_1 = a_{i_1} \dots x_{i_k} ceccdca^{i_k} b \dots a^{i_1} b$$

de la forma w_1 i una cadena

$$z_2 = y_{i_1} \dots y_{i_k} ceccdca^{i_k} b \dots a^{i_1} b$$

de la forma w_4 tal que $z_1 = z_2$.

L'ambigüïtat (de qualsevol tipus) de G és equivalent amb l'existència d'una solució per al problema de correspondència de Post, el qual sabem és un problema indecidible. $\square$

Val la pena mencionar que si considerem llenguatges sobre un vocabulari d'una-lletra, aleshores el resultat anterior ja no és cert: la t -ambigüïtat és decidible per a tot t . Específicament, tenim

Teorema 8.3.9 Es compleix que

1. Tot llenguatge $L \subseteq a^*$ és 4-no-ambigu.
2. Tot llenguatge $L \subseteq a^*$ és 5-ambigu si, i només si, és infinit.
3. Una gramàtica $G = (\{a\}, B, C, \varphi)$ és 2-ambigua si, i només si, és t -ambigua, per a qualsevol $t \in \{3, 4\}$.
4. És decidible si una gramàtica arbitrària $G = (\{a\}, B, C, \varphi)$ és o no és 1-ambigua o 2-ambigua.

No entrarem aquí en detalls de la demostració. Els punts 1 i 2 són conseqüència del fet que els llenguatges d'una-lletra regulars coincideixen amb els llenguatges d'una-lletra de la família *ICC* (Teorema 5.3.2 de la pàgina 206): això implica que un context és suficient (4-no-ambigüïtat); però, per a cadenes suficientment llargues qualsevol context pot ser usat en diversos emplaçaments

(5-ambigüitat). Aleshores, una gramàtica G amb el vocabulari $\{a\}$ és 1-no-ambigua si, i només si, conté només un context; i 2-ambigua si, i només si, o bé G és 1-ambigua o bé existeix un context associat a dos selectors diferents. L'equivalència entre 2-ambigüitat i 3-, 4-ambigüitat és fàcil de ser obtinguda – poden trobar-se més detalls al respecte a (Ilie, 1996c).

Considerarem ara un nou tipus d'ambigüitat: l'*ambigüitat intrínseca*. La idea és il·lustrada mitjançant el següent exemple: considerem la frase

(1) *Pere veié Maria ballant.*

Qui ballava, en Pere o la Maria ? La frase és sintàcticament ambigua. Suposem ara que aquesta frase és generada per una gramàtica contextual. Això vol dir que cada pas de la derivació es correspon també a una frase correcta. Les possibles subfrases correctes de l'anterior frase són

(2) *Pere veié.*

(3) *Pere veié Maria.*

(4) *Pere veié Maria ballant.*

La frase

(5) **Pere veié ballant.*

no és acceptable. Això vol dir que en la nostra gramàtica podem produir la cadena (potser és un axioma) *Pere veié* i tenim els contextos $(\lambda, \textit{Maria})$, $(\lambda, \textit{ballant})$.

Usant aquests contextos només podem produir la següent derivació

$$Pere\ veié \implies Pere\ veié\ Maria \implies Pere\ veié\ Maria\ ballant.$$

La derivació

$$Pere\ veié \implies Pere\ veié\ ballant \implies Pere\ veié\ Maria\ ballant.$$

no és possible, $(\lambda, ballant)$ no és acceptat per $Pere\ veié$. L'ambigüitat d'(1) no apareix !

No obstant, si considerem els contextos $(\lambda, Maria)$, $(\lambda, Maria\ ballant)$, començant per $Pere\ veié$ obtenim les mateixes tres frases correctes (2), (3), (4) com abans.

En ambdós casos, usem només una cadena inicial i dos contextos, i aquest és el nombre òptim d'axiomes i de contextos per a aquest exemple. La tinença de dues descripcions òptimes del mateix llenguatge ens indica alguna ambigüitat intrínseca del llenguatge, d'un tipus diferent a la de l'ambigüitat considerada anteriorment, que és la usada normalment en teoria de llenguatges formals.

Per això, sembla natural fer-nos el següent raonament: considerar una manera, intuïtivament adequada, de definir les gramàtiques *òptimes* per a un llenguatge (mínim, segons els punts de vista donats). Si un llenguatge pot ser generat mitjançant dues gramàtiques òptimes *diferents*, aleshores se'l considera intrínsecament ambigu.

Apareixen dos problemes. El problema de més fàcil solució és el relacionat amb la definició de gramàtiques *diferents*: per a les gramàtiques contextuais aquest fet és clar; per a les gramàtiques de Chomsky pot significar “diferent mòdul reanomenant els símbols no-terminals”. No és tan fàcil, però, resoldre el problema relacionat amb la definició d’una gramàtica *òptima*. Ens endinsem ja en l’esmunyedissa àrea de la complexitat descriptiva, on poden ser considerades moltes mesures, en general incompatibles les unes amb les altres – veure la Secció 8.2.

Podem seguir, com a mínim, dues estratègies en aquest cas: o bé considerar només un paràmetre de complexitat (i, per tant, definir l’ambigüitat intrínseca respecte a aquest paràmetre), o bé prendre diversos paràmetres, ordenats mitjançant una relació de prioritat.

A continuació, considerarem breument la segona estratègia pel cas de les gramàtiques contextuais internes, és a dir, considerarem els paràmetres Ax i Con de la Secció 8.2.

La complexitat combinada $AxCon$ d’un llenguatge L serà

$$AxCon(L) = (Ax_{ICC}(L), \min\{Con(G) \mid L = L_{in}(G), Ax(G) = Ax_{ICC}(L)\}).$$

Clarament, $AxCon(L) = (Ax_{ICC}(L), m)$ per a algun $m \geq Con_{ICC}(L)$, ja que escollim una gramàtica que és òptima respecte a Con d’entre totes les gramàtiques Ax òptimes.

Cada gramàtica G tal que $AxCon(L) = (Ax(G), Con(G)), L = L_{in}(G)$,

s'anomena *AxCon-òptima* per a L . Denotem per $AxCon^{-1}(L)$ el conjunt de totes les gramàtiques d'aquest tipus.

Exemple 8.3.1 El llenguatge $L = a^+ \cup b^+$ pot ser generat mitjançant

$$G = (\{a, b\}, \{a, b\}, (\{a\}, \{(\lambda, a)\}), (\{b\}, \{(\lambda, b)\})),$$

que és *AxCon-òptima*, ja que és fàcil de veure que $Ax_{ICC}(L) = 2$, $Con_{ICC}(L) = 2$ i, per tant, $AxCon(L) = (2, 2)$. Per això, $G \in AxCon^{-1}(L)$. Per altra banda, és fàcil de veure que si $G \in AxCon^{-1}(L)$, aleshores ha de contenir els axiomes a, b , un dels contextos $(\lambda, a), (a, \lambda)$ i un dels contextos $(\lambda, b), (b, \lambda)$. Per tant, $AxCon^{-1}(L)$ conté, exactament, quatre gramàtiques diferents.

Un llenguatge L és *intrínsecament AxCon ambigu* si $\text{card}(AxCon^{-1}(L)) \geq 2$. En general, si

$$\text{card}(AxCon^{-1}(L)) = k, \quad k \geq 1,$$

aleshores direm que L és *intrínsecament AxCon ambigu de grau k* . Si el conjunt $AxCon^{-1}(L)$ és infinit, aleshores direm que L té un grau d'ambigüïtat infinit.

Sorprenentment, tenim

Teorema 8.3.10 Tot llenguatge finit és *AxCon* no-ambigu; mentre que, tot llenguatge contextual intern infinit generat per una gramàtica amb F selecció, tal que F és tancada sota l'operació d'intersecció amb llenguatges regulars, té un grau infinit d'*AxCon* ambigüïtat.

Demostració. Prenem un llenguatge $L \subseteq V^*$. Aquest llenguatge és finit, aleshores tota gramàtica $G = (V, B, C, \varphi)$ tal que $L = L_{in}(G)$ ha de o bé tenir $C = \emptyset$, o bé $\varphi^{-1}((u, v)) \cap Sub(L) = \emptyset$, per a tots $(u, v) \in C$, ja que sinó generaria un llenguatge infinit: quan $x \Rightarrow_{in} y, x \in L, x = x_1x_2x_3, y = x_1ux_2vx_3$, per a $(u, v) \in \varphi(x_2)$, aleshores $y \Rightarrow_{in} z$, també, per a $z = x_1uux_2vux_3$ i, per tant, $x_1u^n x_2v^n x_3 \in L$ per a tot $n \geq 1$ (el context (λ, λ) , si és que apareix, pot ser esborrat sense canviar el llenguatge generat). A més a més, $Ax_{ICC}(L) = \text{card}(L)$, $Con_{ICC}(L) = 0$. Existeix, exactament, una gramàtica G' amb $(Ax(G'), Con(G')) = (\text{card}(L), 0)$ i, per tant, L és no-ambigua.

Si L és infinit, aleshores $L = L_{ICC}(G)$, per a alguna gramàtica $G = (V, B, C, \varphi)$ tal que existeixi $(u, v) \in C, uv \neq \lambda$, i $w \in Sub(L), (u, v) \in \varphi(w)$. Prenem $z \in L, z = x_1x_2x_3$, amb $x_2 = w$. Aleshores $x_1u^n x_2v^n x_3 \in L$, per a tot $n \geq 1$.

Considerem la gramàtica $G' = (V, B, C, \varphi')$ amb

$$\begin{aligned}\varphi'(x) &= \varphi(x), \text{ per a } x \neq u^i w v^j, i, j \geq 1, \\ \varphi'(u^i w v^j) &= \varphi(u^i w v^j) - \{(u, v)\}, \text{ per a } i, j \geq 1.\end{aligned}$$

Tenim que $L_{in}(G) = L_{in}(G')$: l'ús de (u, v) a G associat amb els selectors $u^i w v^j, i, j \geq 1$ (si és que aquest és el cas), poden ser simulats a G' mitjançant l'ús de (u, v) associat amb w .

La gramàtica G' és del mateix tipus que G : si $C' \subseteq C$ tal que $(u, v) \notin C'$, aleshores $\varphi'^{-1}(C') = \varphi(C)$. Si $(u, v) \in C'$, aleshores $\varphi'^{-1}(C') = \varphi^{-1}(C') \cap$

$(V^* - u^+wv^+)$ i $\varphi'^{-1}(C' - \{(u, v)\}) = \varphi^{-1}(C') \cap u^+wv^+$. Per tant, també G' té F selecció.

Ara, considerem les gramàtiques

$$G_k = (V, A, C, \varphi_k), \quad k \geq 1,$$

on

$$\varphi_k(x) = \varphi(x), \quad \text{per a } x \neq u^i w v^i, i \geq 1,$$

$$\varphi_k(u^i w v^i) = \varphi'(u^i w v^i) \cup \{(u, v)\}, \quad \text{per a } 1 \leq i \leq k.$$

Tenim $L_{in}(G_k) = L_{in}(G') = L_{in}(G)$, $Ax(G_k) = Ax(G)$, $Con(G_k) = Con(G)$. Si G és una $AxCon$ -gramàtica òptima per a L , aleshores G_k és també $AxCon$ òptima. La definició de φ_k difereix de la de φ' anterior en un nombre finit d'arguments i, per tant, també $G_k, k \geq 1$, són gramàtiques amb F selecció.

Cada dues gramàtiques $G_r, G_s, s \neq r$, són diferents i, per tant, L té un grau d'ambigüitat intrínseca infinit. $\square$

El resultat previ pot ser interpretat en el sentit que la noció d'ambigüitat intrínseca expressada és encara massa dèbil. Altres mesures haurien de ser tingudes en consideració.

No obstant, aquesta noció d'ambigüitat intrínseca encaixa amb el que succeix en el cas dels llenguatges regulars: els llenguatges regulars són intuïtivament no-ambigus (d'alguna manera, són també sintàcticament no-ambigus) i,

significativament, poden ser descrits mitjançant un únic autòmat finit determinista mínim (quant al nombre d'estats).

A més a més, els resultats anteriors també encaixen amb la situació que apareix en els llenguatges naturals, dels quals se sap són altament ambigus (veure arguments a (Miquel-Vergés, 1995a)). Anteriorment hem provat que els llenguatges contextuais interns són altament ambigus (des dels punts de vista introduïts aquí). Això podria significar que les gramàtiques contextuais internes són un model *adequat* dels llenguatges naturals.

Notes bibliogràfiques. El Teorema 8.1.3 de la pàgina 245 és de (Ehrenfeucht et al., 1995a). La Secció 8.2 està basada en (Păun, 1975b; Păun, 1991; Păun, 1996a). Resultats relacionats poden ser trobats a (Georgescu, 1994; Georgescu, 1996). Per exemple, (Georgescu, 1994) considera diverses mesures de complexitat sintàctica per a gramàtiques contextuais amb contextos esborrats (principalment les qüestions relacionades amb la connectitud), mentre que a (Georgescu, 1996) es poden trobar resultats relacionats amb les mesures de decidibilitat Ax , MAx , TAx , Con , $MCon$, $TCon$, Phi per a gramàtiques contextuais internes (com és usual, totes les mesures excepte Phi són computables per a gramàtiques, però per a totes les mesures de tots els altres problemes de computabilitat i de decidibilitat tenen respostes negatives).

A (Miquel-Vergés, 1995a) es mostren alguns dels arguments que fan que els llenguatges naturals, en general, siguin altament ambigus; i la seva coincidència, en aquest aspecte concret, amb els llenguatges contextuais. Ja més

específicament, l'ambigüïtat de les gramàtiques contextuals és investigada a (Păun, 1982a). La primera afirmació del Teorema 8.3.1 de la pàgina 274 està basada en (Gross, 1964). Les variants d'ambigüïtat per a gramàtiques contextuals internes considerades en la Secció 8.3 van ser introduïdes a (Martín-Vide et al., 1996c); tots els resultats presentats aquí són de 8.3, excepte el Teorema 8.3.7 de la pàgina 287, que és de (Ilie, 1996c).

Una versió resumida del que s'ha presentat en aquest capítol pot ser trobada també a (Ehrenfeucht et al., 1996a) i, especialment, a (Păun, 1996a).

Capítol 9

Gramàtiques contextuais amb selecció restringida

Com hem definit a la Secció 4.1, les gramàtiques contextuais tenen una definició infinitària: la funció de selecció φ en la presentació funcional és definida per a totes les cadenes de V^* (anàlogament, els selectors en la presentació modular poden ser conjunts infinits). No obstant, la noció d'una “gramàtica” suposa algun tipus de mecanisme finit que descrigui la sintaxi d'un llenguatge. Per tant, es fan necessàries algunes restriccions en la definició general d'una gramàtica contextual (en la funció φ i en els selectors). La idea més natural és la d'imposar certa regularitat en la definició de la selecció. Al final de la Secció 4.1 ja vam considerar breument el cas en què els selectors són llenguatges d'una família donada F , mentre que en seccions precedents aquestes restriccions han aparegut en diversos teoremes. Veurem ara un estudi sistemàtic d'aquestes variants.

Són particularment interessants els casos en què els selectors són llenguatges finits o regulars (el seu problema de pertanyença pot ser en aquests casos resolt en temps real – veure la Taula 3.7.1 de la pàgina 151). Començarem, però, per considerar gramàtiques (internes i externes) amb selectors d'un tipus donat de la jerarquia de Chomsky.

9.1 Definicions i resultats bàsics

Definició 9.1.1 Sigui F una família de llenguatges donada. Una gramàtica contextual *amb F selecció* és una gramàtica $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$, amb $S_i \in F$, per a tot $1 \leq i \leq n$.

Denotem per $ICC(F)$, $ECC(F)$ les famílies de llenguatges generades en el mode intern i en el mode extern, respectivament, per gramàtiques contextuais amb F selecció. Considerarem que F és una de les famílies FIN, REG, CF, CS, RE , amb especial èmfasi quan $F \in \{FIN, REG\}$.

Molts dels exemples usats en capítols anteriors són gramàtiques amb F selecció per a $F \in \{FIN, REG\}$. Per altra banda, $ICC = ICC(ARB)$ i $ECC = ECC(ARB)$; i, per tant, $ICC(F) \subseteq ICC$ i $ECC(F) \subseteq ECC$, per a totes les famílies F . A més a més, totes les condicions necessàries per a les famílies ICC, ECC són també vàlides per a $ICC(F), ECC(F)$, respectivament. Algunes d'aquestes condicions tenen formes característiques pel cas

de les famílies $ICC(F)$, $ECC(F)$ per a un F particular. Per exemple,

Lema 9.1.1 Si $L \subseteq V^*$, $L \in ICC(FIN)$, aleshores existeixen tres constants p, q, r tal que tota $z \in L$, $|z| > p$, pot ser escrita de la forma $z = uvwxy$, amb $u, v, w, x, y \in V^*$, $0 < |vx| \leq q$, $|w| \leq r$, i $uv^iwx^iy \in L$, per a tot $i \geq 0$.

Demostració. És la mateixa que la demostració del Lema 5.1.5 de la pàgina 189, prenent com a $r = \max\{|w| \mid w \in S, (S, C) \in P\}$, per a una gramàtica $G = (V, B, P)$ amb selecció finita. $\square$

Les següents relacions són fàcilment demostrables (en el cas de RE podem usar la tesi de Turing-Church).

Lema 9.1.2 $ICC(F) \subseteq F$, $ECC(F) \subseteq F$, per a $F \in \{CS, RE\}$.

Lema 9.1.3 $IC \subseteq ICC(REG)$, $EC \subseteq ECC(REG)$.

Com era d'esperar, tenim la següent jerarquia:

Lema 9.1.4 $ICC(FIN) \subset ICC(REG) \subset ICC(CF) \subset ICC(CS) \subset ICC(RE)$.

Demostració. Les inclusions són òbvies, hem de provar només que són estrictes.

(1) Per a la gramàtica

$$G_1 = (\{a, b\}, \{abab\}, (ab^+a, \{(a, a)\}), (ba^+b, \{(b, b)\})),$$

tenim $L_{in}(G_1) = \{a^n b^m a^n b^m \mid n, m \geq 1\}$. Aquest llenguatge no té la propietat expressada en el Lema 9.1.1 de la pàgina 303 i, per tant, $L_{in}(G_1) \notin ICC(FIN)$, i $ICC(REG) - ICC(FIN) \neq \emptyset$.

(2) Considerem la gramàtica contextual G de l'Exemple 4.2.8 de la pàgina 178. Denotem-la per G_2 . És fàcil de veure que té una CF -selecció i, per tant, $L_{in}(G_2) \in ICC(CF)$. Suposem que $L_{in}(G_2) = L_{in}(G)$ per a alguna $G = (\{a, b, c\}, B, (S_1, C_1), \dots, (S_n, C_n))$, amb conjunts regulars $S_i, 1 \leq i \leq n$.

Si $w = bc^i a^n b^n caca^n b^n c^i$ és un axioma de B i $w \Rightarrow_{in} w'$, aleshores hem de tenir $w' = bc^j a^{n'} b^{n'} caca^{n'} b^{n'} c^j$ amb $n' = n$.

Si $n' > n$, aleshores existeixen $(a^{n'-n} b^{n'-n}, a^{n'-n} b^{n'-n}) \in \varphi(b^n caca^n)$ i podem obtenir també la derivació $bc^i a^{n+1} b b^n caca^n a b^{n+1} c^i \Rightarrow_{in} ba^i a^{n+1} ba^{n'-n} b^{n'-n} b^n caca^n a^{n'-n} b^{n'-n} a b^{n+1} c^i$, que no pertany a $L_{in}(G_2)$; la qual cosa és una contradicció.

A més a més, començant per cadenes de B de la forma $ba^n ba^m caca^n ba^m, n, m \geq 1$, obtenim un nombre infinit de cadenes de la forma anterior w, w' . Considerem un pas de derivació en el cas en què les primeres ocurrències noves de c són afegides a aquesta cadena, $ba^n b^m caca^n b^m \Rightarrow_{in} bc^i a^{n'} b^{m'} caca^{n'} b^{m'} c^i, i \geq 1$. Clarament, la primera cadena $b^{m'}$ ha de ser igual a b^m i l'última cadena $a^{n'}$ ha de ser igual a a^n i, per tant, $n' = n, m' = m$. Com que $n' = m'$, tenim també que $n = m$. Existeix un context (c^i, c^i) tal que $(c^i, c^i) \in \varphi(a^n b^m caca^n b^m)$ per a un valor arbitrari de $n, m \geq 1$. Si (c^i, c^i) està associat amb algun S_j que

conté un nombre infinit de cadenes $a^n b^m c a c a^n b^m$, aleshores $S_j \cap \{a^n b^m c a c a^n b^m \mid n, m \geq 1\}$ és un subconjunt infinit de $\{a^n b^n c a c a^n b^n \mid n \geq 1\}$. Això no és possible quan S_j és regular; la qual cosa és una contradicció. Conseqüentment, $L_{in}(G_2) \notin ICC(REG)$ i $ICC(CF) - ICC(REG) \neq \emptyset$.

(3) Per a la gramàtica

$$G_3 = (\{a, b, c\}, \{cbac\}, (\{cb\}, \{(\lambda, a)\}), (\{b\}, \{(a, \lambda)\}), \\ (\{caba^{2^n}c \mid n \geq 1\}, \{(c, c)\})),$$

obtenim

$$L_{in}(G_3) = \{cba^n c \mid n \geq 1\} \\ \cup \{ca^n ba^m c \mid n, m \geq 1\} \\ \cup \{c^i a^n ba^{2^m} c^i \mid n, m \geq 1, i \geq 2\}.$$

Després d'usar el context (c, c) , el context (λ, a) ja no pot ser aplicat.

Suposem que $L_{in}(G_3) = L_{in}(G)$ per a algun $G = (\{a, b, c\}, A, (S_1, C_1), \dots, (S_n, C_n))$, amb conjunts lliures del context $S_i, 1 \leq i \leq n$. Considerem les cadenes de la forma $w = c^2 aba^{2^m} c^2, m \geq 1$. Ja que el nombre de contextos és finit, per a un valor d' m gran no podem tenir $z \Rightarrow_{in} w$ per a algun $z = c^2 aba^{2^p} c^2$ i, per tant, w ha de ser produït a partir d'una cadena $z' = caba^q c$. A més a més, existeix $(c, a^i c) \in \varphi(caba^{2^m-i}), i \geq 0$, o $(c, c) \in \varphi(caba^{2^m} c)$. En el primer cas podem tenir també $caba^{2^m-i} a^p c \Rightarrow_{in} ccaba^{2^m} a^p cc$, per a tot $p \geq 1$, produint cadenes que no pertanyen a $L_{in}(G_3)$. En el segon cas,

ja que un nombre arbitrari de cadenes $caba^{2^m}c$ han de pertànyer a S_j , per a algun llenguatge lliure del context S_j , hem de tenir també a S_j cadenes $caba^p c$ amb $p \neq 2^k, k \geq 1$ (interseccionant S_j amb $caba^+c$, després esborrant amb una derivada per l'esquerra el prefix cab i amb una derivada a la dreta el sufix c , hem d'obtenir un subllenguatge regular de a^+). Aleshores $caba^p c \Rightarrow_{in} ccaba^p cc$ i, per tant, produïm de nou una cadena que no pertany a $L_{in}(G_3)$. En conclusió, $L_{in}(G_3) \notin ICC(CF)$ i $ICC(CS) - ICC(CF) \neq \emptyset$.

(4) Per a un llenguatge $L \subseteq V^+, L \in RE - CS$ i dos símbols c, d que no pertanyen a V , construïm la gramàtica (similar a G_L en la demostració del Lema 5.2.2 de la pàgina 198)

$$G_L = (V \cup \{c, d\}, \{ccac \mid a \in V\}, (\{cc\}, \{(\lambda, a) \mid a \in V\}), (\{c\}L\{c\}, \{(d, d)\})).$$

Obtenim

$$L_{in}(G_L) = \{cc\}V^+\{c\} \cup \{cd^i c x c d^i \mid i \geq 1, x \in L\}.$$

Després d'utilitzar el context (d, d) , el context (λ, a) no pot ser usat.

Ja que $L = \partial_{cdc}^e(\partial_{cd}^d(L_{in}(G_L)))$, tenim $L_{in}(G_L) \notin CS$ i, per tant, $L_{in}(G_L) \notin ICC(CS)$ (Lema 9.1.2 de la pàgina 303). Aleshores, $ICC(RE) - ICC(CS) \neq \emptyset$. $\square$

Una jerarquia similar existeix pel cas extern.

Lema 9.1.5 $ECC(FIN) \subset ECC(REG) \subset ECC(CF) \subset ECC(CS) \subset ECC(RE)$.

Demostració. La família $ECC(FIN)$ conté només llenguatges finits i, per tant, la primera inclusió és òbviament pròpia.

Considerem ara la gramàtica contextual

$$G = (\{a, b, c\}, \{\lambda\}, (\{a, b\}^*, \{(\lambda, a), (\lambda, b)\}), (L_0, \{(c, c)\})),$$

per a

$$L_0 = \{a^n b^n \mid n \geq 1\}.$$

Tenim $L_{ex}(G) = \{a, b\}^* \cup \{c\} L_0 \{c\}$. Aquest llenguatge no pertany a $ECC(REG)$. Suposem el contrari i considerem una gramàtica amb selecció regular $G' = (\{a, b, c\}, B, (S_1, C_1), \dots, (S_n, C_n))$ tal que $L_{ex}(G') = L_{ex}(G)$. Una cadena de la forma cxc no pot ser derivada i, per tant, necessitem un context $(u, v) = (ca^i, b^j c)$ tal que $\varphi^{-1}((u, v)) \cap L_{ex}(G)$ sigui infinit (un nombre infinit de cadenes de la forma cxc han de ser generades usant aquest context) i regular ($\varphi^{-1}((u, v))$ és la unió de tots els selectors S_i tal que $(u, v) \in C_i, 1 \leq i \leq n$; en la intersecció anterior, podem reemplaçar $L_{ex}(G)$ per $\{a, b\}^*$, ja que només podem aplicar el context (u, v) a aquestes cadenes sense córrer el perill de produir una cadena paràsita). No obstant, només cadenes de la forma $ca^n b^m c$ amb $n = m$ pertanyen a $L_{ex}(G)$ i, per tant, $\varphi^{-1}((u, v)) \cap L_{ex}(G)$ consisteix en un nombre infinit de cadenes de la forma $a^{n-i} b^{n-j}$. Aquest llenguatge no és regular, la qual cosa és una contradicció.

Podem aplicar el mateix argument al llenguatge generat per la gramàtica

$$G = (\{a, b, c, d\}, \{\lambda\}, (\{a, b, c\}^*, \{(\lambda, a), (\lambda, b), (\lambda, c)\}), (L'_0, \{(d, d)\})),$$

per a

$$L'_0 = \{a^n b^n c^n \mid n \geq 1\},$$

per tal d'obtenir la inclusió estricta de la inclusió $ECC(CF) \subset ECC(CS)$.

Si l'anterior L_0 fos un llenguatge pertanyent a $RE - CS$ aleshores, fent ús del fet que la família CS és tancada sota l'operació d'intersecció i sota l'operació de morfismes restringits (veure la Taula 6.12.1 de la pàgina 228), a partir de $L_{ex}(G) \in CS$ obtindríem que $L \in CS$, la qual cosa és una contradicció. Per tant, també l'última inclusió és pròpia. $\square$

En el Corol·lari 5.3.1 de la pàgina 203 vam veure que $REG \subseteq ICC(FIN)$. Una altra interessant relació entre una família de llenguatges contextuais (i, per tant, generada usant l'operació d'*adjunció*) i una família de llenguatges en la jerarquia de Chomsky (i, per tant, generat usant *reescriptura*) és la següent.

Lema 9.1.6 $ECC(REG) \subseteq LIN$.

Demostració. Considerem una gramàtica $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$ amb selectors regulars $S_i, 1 \leq i \leq n$. Prenem un autòmat finit (no necessàriament determinista) $M_i = (K_i, V, q_{0,i}, F_i, \delta_i)$ tal que $S_i = L(M_i), 1 \leq i \leq n$. Suposem, sense perdre cap generalitat, que $K_i \cap K_j = \emptyset$ per a tot $i \neq j$, i construïm la gramàtica lineal $G' = (N, V, S, P)$ amb

$$N = \{S\} \cup \{Q \mid Q \subseteq \bigcup_{i=1}^n (K_i \times K_i)\},$$

(els conjunts Q són considerats no-terminals) i P contenint les següents regles:

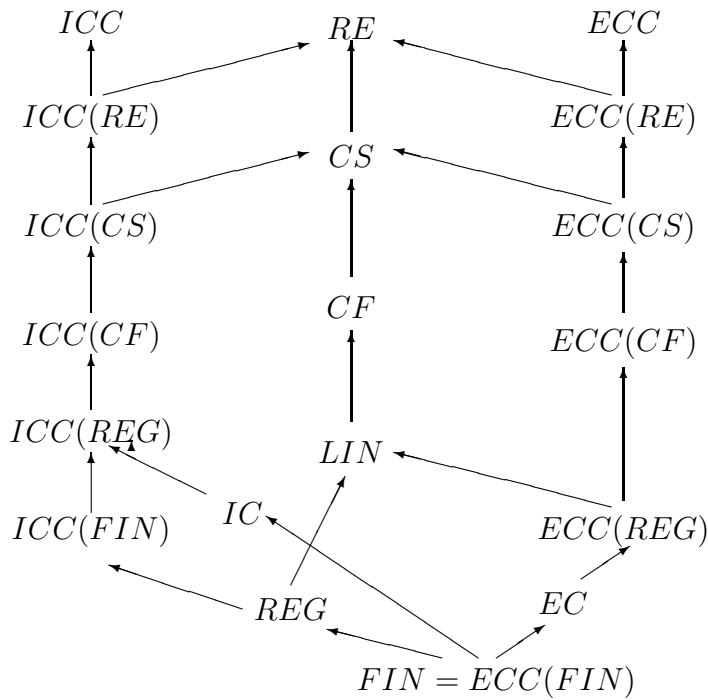
1. $S \rightarrow uQv$,
per a $(u, v) \in C_i, Q = \{(q_{0,i}, q_f)\}$, per a algun $q_f \in F_i, 1 \leq i \leq n$.
2. $Q \rightarrow uQ'v$, on
 - $(u, v) \in C_i$ per a algun $i, 1 \leq i \leq n$, i $Q \in N$,
 - si $(q_1, q_2) \in Q, q_1, q_2 \in K_j$ per a algun $j, 1 \leq j \leq n$, aleshores Q' conté un parell $(q'_1, q'_2), q'_1, q'_2 \in K_j$, tal que $q'_1 \in \delta_j(q_1, u), q_2 \in \delta_j(q'_2, v)$,
 - a més a més dels parells (q'_1, q'_2) definits anteriorment, afegim a Q' un parell $(q_{0,i}, q_f), q_f \in F_i$, si encara no és a Q' .
3. $Q \rightarrow x$, on
 - $x \in B, Q \in N$,
 - per a tot $(q, q') \in Q$ tal que $q, q' \in K_i$ per a algun $i, 1 \leq i \leq n$, tenim $q' \in \delta_i(q, x)$.
4. $S \rightarrow x, x \in B$.

A partir de l'anterior exposició és fàcil de veure que $L_{ex}(G) = L(G')$ i, per tant, $L_{ex}(G) \in LIN$. $\square$

Resumint aquests lemes, així com els resultats obtinguts en les Seccions 5.2 i 5.3, els quals són també vàlids per a gramàtiques amb selecció restringida, obtenim

Teorema 9.1.1 Es compleixen les relacions que la figura 9.1.1 mostra, on una fletxa d'una família F_1 a una altra família F_2 indica la inclusió estricta $F_1 \subset F_2$, i les famílies no relacionades són incomparables.

Figura 9.1.1 Relacions entre les famílies de gramàtiques contextuais amb selecció restringida.



Demostració. Per a demostrar les incomparabilitats, usem el fet que $REG - ECC \neq \emptyset$ (ab^+a és un llenguatge que prova aquesta relació) i que $L = \{x \mid i(x) \mid x \in \{a,b\}^*\} \in EC - ICC$, (com vam demostrar al Lema 5.1.9 de la pàgina 196). $\square$

Pel fet que aquesta relació és important i a causa de les similituds i de les diferències entre gramàtiques contextuais externes i gramàtiques lineals mereix una discussió més a fons, presentarem ara una forma normal per a gramàtiques contextuais externes amb selecció regular, la qual té com a conseqüència més immediata una nova demostració de la inclusió $ECC(REG) \subseteq LIN$.

El procés de derivació en una gramàtica contextual introdueix símbols al final de la cadena en qüestió; en una gramàtica lineal procedim en la direcció contrària, introduint primer els símbols extrems i després movent-nos en cada pas cap al centre. Per tant, podríem tractar d'“invertir una derivació externa δ en una gramàtica contextual,

$$\delta : w \Rightarrow_{ex} u_1 w v_1 \Rightarrow_{ex} u_2 u_1 w v_1 v_2 \Rightarrow_{ex} \dots \Rightarrow_{ex} u_n \dots u_1 w v_1 \dots v_n,$$

i associar-li una derivació equivalent en una gramàtica lineal,

$$S \Rightarrow u_n \bar{S}_n v_n \Rightarrow \dots \Rightarrow u_n \dots u_1 \bar{S}_1 v_1 \dots v_n \Rightarrow u_n \dots u_1 w v_1 \dots v_n,$$

on $\bar{S}_i, 1 \leq i \leq n$, són símbols associats amb els selectors S_i que apareixen en produccions usades a δ (i.e., per a tot $2 \leq i \leq n, u_{i-1} \dots u_1 w v_1 \dots v_{i-1} \in S_i$ i $(u_i, v_i) \in C_i$). No obstant, aquesta idea pot no funcionar en general si és que no modifiquem el llenguatge generat, ja que una regla $\bar{S}_{i+1} \rightarrow u_i \bar{S}_i v_i$ suposa que cada cadena $u_i x v_i$, amb $x \in S_i$, pertany a S_{i+1} , la qual cosa pot no ser certa en general.

Això ens porta a considerar la següent propietat. Sigui $G = (V, B, P)$ una gramàtica contextual amb les produccions a P de la forma $(S_i, (u_i, v_i))$ (és a

dir, amb només un context en cada producció. És fàcil de veure que, per a cada gramàtica amb conjunts de contextos arbitraris en les seves produccions, podem trobar una gramàtica equivalent amb produccions iguals a les anteriors – el conjunt de contextos és finit).

Definició 9.1.2 Siguin $\pi_1 = (S_1, (u_1, v_1)), \pi_2 = (S_2, (u_2, v_2))$ dues produccions de P . Direm que el parell ordenat (π_1, π_2) és *encaixant* (“matching”) si, i només si, per a cada $x \in S_1$ tenim $u_1 x v_1 \in S_2$. Una derivació a G és *encaixant* (“matching”) si, i només si, cada dues produccions consecutives π_i, π_{i+1} usades formen un parell encaixant. Una gramàtica contextual G és en la *forma normal encaixant* (“matching normal form”) si per a cada $x \in L_{ex}(G)$ existeix una derivació encaixant a G .

A cada derivació encaixant podem associar-li una derivació lineal equivalent, la qual cosa ens duu al resultat que mostrarem a continuació. Notem que els conjunts de selecció S_i de les produccions poden ser llenguatges arbitraris. Pel fet que les gramàtiques contextuais amb selecció arbitrària poden generar arbitràriament llenguatges complexos (fins i tot llenguatges no-recursivament enumerables), el següent lema és bastant sorprenent.

Lema 9.1.7 Si G és una gramàtica contextual en la forma normal encaixant, aleshores $L_{ex}(G)$ és un llenguatge lineal.

Demostració. Sigui $G = (V, B, P)$ amb $B = \{w_1, \dots, w_k\}$, i $P = \{\pi_1, \dots, \pi_n\}$, on, per a cada $1 \leq i \leq n$, $\pi_i = (S_i, (u_i, v_i))$. A cada conjunt S_i li associem un

no-terminal $\overline{S}_i$. Aleshores sigui G' la gramàtica lineal

$$G' = (\{\overline{S}_i \mid 1 \leq i \leq n\} \cup \{S\}, V, S, P'),$$

on P' conté les següents produccions:

1. $S \rightarrow w_i$, per a $1 \leq i \leq k$,
 $S \rightarrow u_i \overline{S}_i v_i$, per a $1 \leq i \leq n$,
2. $\overline{S}_i \rightarrow w_j$, per a cada $1 \leq i \leq n$, i $1 \leq j \leq k$, tal que $w_j \in S_i$,
3. $\overline{S}_j \rightarrow u_i \overline{S}_i v_i$, per a cada parell encaixant (π_i, π_j) , on $1 \leq i, j \leq n$.

Provarem ara que $L_{ex}(G) = L(G')$.

($\subseteq$) Sigui $z \in L_{ex}(G)$ i considerem una derivació encaixant per a z in G ,

$$\begin{aligned} w_i &\Longrightarrow_{ex} u_{j_1} w_i v_{j_1} \Longrightarrow_{ex} u_{j_2} u_{j_1} w_i v_{j_1} v_{j_2} \Longrightarrow_{ex} \dots \\ &\Longrightarrow_{ex} u_{j_m} \dots u_{j_1} w_i v_{j_1} \dots v_{j_m} = z, \end{aligned}$$

que usa les produccions $\pi_{j_1} = (S_{j_1}, (u_{j_1}, v_{j_1})), \dots, \pi_{j_m} = (S_{j_m}, (u_{j_m}, v_{j_m}))$ de manera que $w_i \in S_{j_1}$, $u_{j_s} \dots u_{j_1} w_i v_{j_1} \dots v_{j_s} \in S_{j_{s+1}}$, $1 \leq s \leq m-1$. Els parells $(\pi_{j_s}, \pi_{j_{s+1}})$, $1 \leq s \leq m-1$, són encaixants i, per tant, les regles $\overline{S}_{j_{s+1}} \rightarrow u_{j_{s+1}} \overline{S}_{j_s} v_{j_{s+1}}$, $1 \leq s \leq m-1$ pertanyen a P' ; P' conté també la regla $S \rightarrow u_{j_m} \overline{S}_{j_m} v_{j_m}$ i, pel fet que $w_i \in S_{j_1}$, la regla $\overline{S}_{j_1} \rightarrow w_i$. Conseqüentment, tenim la següent derivació a G' :

$$\begin{aligned} S &\Longrightarrow u_{j_m} \overline{S}_{j_m} v_{j_m} \Longrightarrow u_{j_m} u_{j_{m-1}} \overline{S}_{j_{m-1}} v_{j_{m-1}} v_{j_m} \Longrightarrow \dots \\ &\Longrightarrow u_{j_m} \dots u_{j_1} \overline{S}_{j_1} v_{j_1} \dots v_{j_m} \Longrightarrow u_{j_m} \dots u_{j_1} w_i v_{j_1} \dots v_{j_m} = z. \end{aligned}$$

Per tant, $z \in L(G')$.

($\supseteq$) Prenem $z \in L(G')$ produïda per una derivació a G'

$$\begin{aligned} S &\Longrightarrow u_{j_m} \bar{S}_{j_m} v_{j_m} \Longrightarrow u_{j_m} u_{j_{m-1}} \bar{S}_{j_{m-1}} v_{j_{m-1}} v_{j_m} \Longrightarrow \dots \\ &\Longrightarrow u_{j_m} \dots u_{j_1} \bar{S}_{j_1} v_{j_1} \dots v_{j_m} \Longrightarrow u_{j_m} \dots u_{j_1} w_i v_{j_1} \dots v_{j_m} = z. \end{aligned}$$

Les regles usades són $S \rightarrow u_{j_m} \bar{S}_{j_m} v_{j_m}$, $\bar{S}_{j_1} \rightarrow w_i$, i $\bar{S}_{j_{s+1}} \rightarrow u_{j_s} \bar{S}_{j_s} v_{j_s}$, $1 \leq s \leq m-1$. La regla $\bar{S}_{j_1} \rightarrow w_i$ és introduïda a P' només quan $w_i \in S_{j_1}$ i $\bar{S}_{j_{s+1}} \rightarrow u_{j_s} \bar{S}_{j_s} v_{j_s}$ correspon a un parell encaixant $(\pi_{j_s}, \pi_{j_{s+1}})$, $1 \leq s \leq m-1$. Conseqüentment, $w_i \in S_{j_1}$, $u_{j_1} w_i v_{j_1} \in S_{j_2}, \dots, u_{j_s} \dots u_{j_1} w_i v_{j_1} \dots v_{j_s} \in S_{j_{s+1}}, \dots$, i $u_{j_{m-1}} \dots u_{j_1} w_i v_{j_1} \dots v_{j_{m-1}} \in S_{j_m}$. Per tant, tenim la següent derivació externa a G :

$$\begin{aligned} w_i &\Longrightarrow_{ex} u_{j_1} w_i v_{j_1} \Longrightarrow_{ex} u_{j_2} u_{j_1} w_i v_{j_1} v_{j_2} \Longrightarrow_{ex} \dots \\ &\Longrightarrow_{ex} u_{j_m} \dots u_{j_1} w_i v_{j_1} \dots v_{j_m} = z. \end{aligned}$$

Per tant, $z \in L_{ex}(G)$. □

Notem que si els selectors que apareixen en les produccions de G no són recursius, aleshores la construcció de la demostració del lema anterior no és efectiva.

A la vista del lema anterior, si una gramàtica G genera un llenguatge $L_{ex}(G)$ que no és lineal, aleshores no podem construir cap gramàtica G' que sigui equivalent amb G i en la forma normal encaixant. No obstant, per a totes les gramàtiques amb selectors regulars aquest resultat és sempre cert.

Donarem una demostració d'aquesta afirmació després d'establir un seguit de lemes preliminars (sempre que sigui possible, seran indicats per a gramàtiques contextuais arbitràries en comptes de només per a gramàtiques contextuais amb selecció regular).

Definició 9.1.3 Donada una gramàtica contextual $G = (V, B, P)$, una gramàtica $G' = (V, B, P')$ amb $P \subseteq P'$ s'anomena una *extensió* de G .

Lema 9.1.8 Si $(S_1, (u, v))$ és una producció d'una gramàtica contextual $G = (V, B, P)$ i $S_2 \subseteq S_1$, aleshores $L_{ex}(G) = L_{ex}(G')$, on $G' = (V, A, P \cup \{(S_2, (u, v))\})$.

Demostració. És òbvia, ja que l'ús de la producció $(S_2, (u, v))$ pot ser reemplaçat per l'ús de $(S_1, (u, v))$. $\square$

Per a dues produccions $\pi_1 = (S_1, (u_1, v_1))$, $\pi_2 = (S_2, (u_2, v_2))$ d'una gramàtica contextual $G = (V, B, P)$, denotem

$$\mu[\pi_1, \pi_2] = S_1 \cap u_1 \setminus S_2 / v_1.$$

Lema 9.1.9 Per a π_1, π_2 com abans, el parell $(\pi_{1,2}, \pi_2)$, amb $\pi_{1,2} = (\mu[\pi_1, \pi_2], (u, v))$ és encaixant.

Demostració. Per la definició de $\mu[\pi_1, \pi_2]$ tenim que $u_1 x v_1 \in S_2$ per a tot $x \in \mu[\pi_1, \pi_2]$ i, per tant, es compleix el lema. $\square$

Lema 9.1.10 Si $G = (V, B, P)$ és una gramàtica contextual, i $\pi_1, \pi_2 \in P$, amb $\pi_1 = (S_1, (u_1, v_1))$, aleshores $G' = (V, B, P \cup \{\pi_{1,2}\})$, amb $\pi_{1,2} = (\mu[\pi_1, \pi_2], (u_1, v_1))$,

és una gramàtica contextual externament equivalent amb G .

Demostració. Per definició, $\mu[\pi_1, \pi_2] \subseteq S_1$ i, per tant, el lema se segueix del Lema 9.1.8. $\square$

Procedirem ara a establir aquells resultats preliminars (necessaris per a la demostració del teorema de la forma normal encaixant) que són certs només per a gramàtiques contextuais amb selecció regular.

Definició 9.1.4 Un autòmat finit total determinista sense estats finals s'anomena un *sistema de transició*. En aquest cas, aquesta construcció és una 4-tupla $M = (K, V, s_0, \delta)$, on V és un vocabulari, K és un conjunt d'estats, $s_0 \in K$ és l'estat inicial, i $\delta : K \times V \longrightarrow K$ és una funció de transició.

Dos sistemes de transició $M_i = (K_i, V_i, s_{0,i}, \delta_i), i = 1, 2$, s'anomenen *iguals* si, i només si, $V_1 = V_2$ i existeix una bijecció $\varphi : K_1 \longrightarrow K_2$ tal que

1. $\varphi(s_{0,1}) = s_{0,2}$,
2. $\delta_1(s, a) = s'$ sii $\delta_2(\varphi(s), a) = \varphi(s')$, per a tot $a \in V_1$ i tot $s, s' \in K_1$.

Sigui $\mathcal{T}$ un conjunt finit de sistemes de transició sobre el mateix vocabulari V . La *clausura* de $\mathcal{T}$, denotada $cl(\mathcal{T})$, és el conjunt més petit $\mathcal{T}'$ de sistemes de transició que conté $\mathcal{T}$ i és tancat sota les dues següents operacions.

1. *Prenent subsistemes* determinats mitjançant l'ajustament de qualsevol estat a ser inicial.

En aquest cas, per a $M = (K, V, s_0, \delta)$ i $s \in K$, definim $sub(M, s) = (K', V, s, \delta')$, on $K' = \{s' \in K \mid \text{existeix } u \in V^* \text{ tal que } \delta(s, u) = s'\}$ (el conjunt de tots els estats assolibles a partir d' s) i $\delta' : K' \times V \longrightarrow K'$, definida per $\delta'(s', a) = \delta(s', a)$, per a tot $s' \in K'$ i all $a \in V$ (la restricció de δ a $K' \times V$).

2. El producte directe.

Per a dos sistemes de transició $M_i = (K_i, V, s_{0,i}, \delta_i), i = 1, 2$, el producte directe de M_1, M_2 és el sistema de transició $M_1 \otimes M_2 = (K_{1,2}, V, s_{0,1}, s_{0,2}), \delta_{1,2}$, on $K_{1,2} = \{(s_1, s_2) \mid \text{existeix } u \in V^* \text{ tal que } \delta_1(s_{0,1}, u) = s_1, \text{ i } \delta_2(s_{0,2}, u) = s_2\}$ (el conjunt de parells de $K_1 \times K_2$ assolibles a partir de $s_{0,1}, s_{0,2}$ en els dos sistemes de transició mitjançant la mateixa cadena de V^*), $\delta_{1,2}((s_1, s_2), a) = (\delta_1(s_1, a), \delta_2(s_2, a))$, per a $(s_1, s_2) \in K_{1,2}, a \in V$.

A continuació veurem algunes propietats d'aquestes operacions.

Lema 9.1.11 El producte directe és commutatiu i associatiu

Demostració. Se segueix directament de la definició d'igualtat de dos sistemes de transició isomòrfics □

Lema 9.1.12 Si $M_i = (K_i, V, s_{0,i}, \delta_i), i = 1, 2$, són dos sistemes de transició i $s_1 \in K_1, s_2 \in K_2$, aleshores

$$sub(M_1, s_1) \otimes sub(M_2, s_2) = sub(M_1 \otimes M_2, (s_1, s_2)).$$

Demostració. Se segueix directament de la definició de la igualtat de sistemes de transició i del fet que cada M_i és determinista i total. $\square$

Corol·lari 9.1.1 Tot $M \in cl(\mathcal{T})$ és de la forma

$$M = sub(M_{i_1} \otimes M_{i_2} \otimes \cdots \otimes M_{i_m}, (s_{i_1}, s_{i_2}, \dots, s_{i_m})),$$

on $m \geq 1$, $M_{i_j} \in \mathcal{T}$, i s_{i_j} és un estat de M_{i_j} , per a cada $1 \leq j \leq m$.

Lema 9.1.13 Donat un sistema de transició $M = (K, V, s_0, \delta)$, només existeix un nombre finit d'autòmats deterministes els sistemes de transició subjacents dels quals són iguals a M .

Demostració. M és el sistema de transició subjacent d'un autòmat finit M' si, i només si, $M' = (K, V, s_0, F, \delta)$ amb $F \subseteq K$. En aquest cas, existeixen $2^{card(K)}$ d'aquests autòmats (hem inclòs el cas $F = \emptyset$). $\square$

El següent resultat juga un paper crucial per a la demostració del teorema de la forma normal encaixant.

Lema 9.1.14 Si $\mathcal{T}$ és un conjunt finit de sistemes de transició, aleshores $cl(\mathcal{T})$ és finit.

Demostració. Sigui $\mathcal{T} = \{M_1, \dots, M_n\}$, on, per a cada $1 \leq i \leq n$, $M_i = (K_i, V, s_{0,i}, \delta_i)$ i sigui $M \in cl(\mathcal{T})$. D'acord amb l'anterior corol·lari, M és un subsistema del producte de $M_0 = M_{i_1} \otimes M_{i_2} \otimes \cdots \otimes M_{i_m}$ de $m \geq 1$ elements de $\mathcal{T}$.

Sigui m_0 el més petit d'aquests m , per al M donat. Considerem l'estat inicial d' M , $(s_{i_1, j_1}, s_{i_2, j_2}, \dots, s_{i_{m_0}, j_{m_0}})$, on $s_{i_k, j_k} \in K_{i_k}$ per a cada $1 \leq k \leq m_0$.

Observació 9.1.1 Per a cap k, l tal que $1 \leq k < l \leq m_0$ tenim que $i_k = i_l$ i $j_k = j_l$.

Demostració. (de l'observació) Pel fet que tots els $M_i \in \mathcal{T}$ són deterministes i totals, si $i_k = i_l, j_k = j_l$, aleshores

$$\delta_{i_k}(s_{i_k, j_k}, u) = \delta_{i_l}(s_{i_l, j_l}, u),$$

per a tot $u \in V^*$. A més a més, podem definir un isomorfisme φ de K_M , el conjunt d'estats d' M , a un conjunt K'_M de $(m_0 - 1)$ -tuples, mitjançant

$$\begin{aligned} \varphi(s_{i_1, j_1}, \dots, s_{i_k, j_k}, \dots, s_{i_{l-1}, j_{l-1}}, s_{i_l, j_l}, s_{i_{l+1}, j_{l+1}}, \dots, s_{i_{m_0}, j_{m_0}}) = \\ = (s_{i_1, j_1}, \dots, s_{i_k, j_k}, \dots, s_{i_{l-1}, j_{l-1}}, s_{i_{l+1}, j_{l+1}}, \dots, s_{i_{m_0}, j_{m_0}}) \end{aligned}$$

(ens saltem l ja que $s_{i_k, j_k} = s_{i_l, j_l}$).

Conseqüentment, M és igual a un sistema de transició M' el qual és un subsistema del producte de $m_0 - 1$ elements de $\mathcal{T}$, contradient el fet que m_0 és mínim. $\square$

Aquesta observació implica que tot M_i de $\mathcal{T}$ pot aparèixer en el producte M_0 com a màxim $\text{card}(K_i)$ vegades i, per tant, la constant m_0 està acotada per

$$\sum_{i=1}^n \text{card}(K_i).$$

Per tant, cada $M \in cl(\mathcal{T})$ és un subsistema del producte $M_{i_1} \otimes \cdots \otimes M_{i_m}$ amb $m \leq m_0$, la qual cosa implica que $cl(\mathcal{T})$ és finit. $\square$

Ens fixarem ara en les operacions usades per a definir $cl(\mathcal{T})$ en la construcció del conjunt $\mu[\pi_1, \pi_2]$ com en els Lemes 9.1.9 de la pàgina 315 i Lema 9.1.10 de la pàgina 315.

Lema 9.1.15 Sguin $\pi_1 = (S_1, (u_1, v_1))$, $\pi_2 = (S_2, (u_2, v_2))$ produccions en una gramàtica contextual $G = (V, A, P)$ amb selecció regular i siguin M_1, M_2 els sistemes de transició subjacents dels autòmats finits deterministes totals M'_1, M'_2 que reconeixen S_1 i S_2 , respectivament. Existeix un autòmat M' que accepta $\mu[\pi_1, \pi_2]$ tal que el seu sistema de transició subjacent M és una subsistema de $M_1 \otimes M_2$.

Demostració. Sigui $M'_i = (K_i, V, s_{0,i}, F_i, \delta_i)$, per a $i = 1, 2$; per tant, $M_i = (K_i, V, s_{0,i}, \delta_i)$. Sigui $M' = (K, s_0, F, \delta_i)$ l'autòmat finit determinista que reconeix $\mu[\pi_1, \pi_2] = S_1 \cap u_1 \setminus S_2 / v_1$ on:

$$K = \{(s_1, s_2) \mid \text{existeix } u \in V^* \text{ tal que} \\ \delta_1(s_{0,1}, u) = s_1, \delta_2(s_{0,2}, u_1 u) = s_2\}$$

(tots els parells d'estats assolibles per la mateixa cadena d'entrada a M_1 a partir de $s_{0,1}$ i a M_2 a partir de $\delta_2(s_{0,2}, u_1)$),

$$s_0 = (s_{0,1}, \delta_2(s_{0,2}, u_1)),$$

$$F = \{(s_1, s_2) \mid s_1 \in F_1, \delta(s_2, v_1) \in F_2\},$$

$$\delta((s_1, s_2), a) = (\delta_1(s_1, a), \delta_2(s_2, a)), \quad (s_1, s_2) \in K, a \in V.$$

Clarament $L(M') = \mu[\pi_1, \pi_2]$.

Si ignorem els estats finals de M' , obtenim un sistema de transició M que és clarament igual a $M_1 \otimes \text{sub}(M_2, \delta_2(s_{0,2}, u_1)) = \text{sub}(M_1 \otimes M_2, (s_{0,1}, \delta_2(s_{0,2}, u_1)))$. Per tant, es compleix el lema. $\square$

Per a les produccions $\pi_i = (S_i, (u_i, v_i))$, $i = 1, 2$, en una gramàtica contextual amb selecció regular, escrivim

$$\pi_1 \subseteq \pi_2 \text{ si, i només si } S_1 \subseteq S_2 \text{ i } (u_1, v_1) = (u_2, v_2).$$

Ara estem en condicions de provar el teorema de la forma normal encaixant.

Teorema 9.1.2 Per a qualsevol gramàtica contextual G amb selecció regular, existeix una gramàtica contextual G' amb selecció regular i en la forma normal encaixant externament equivalent. A més a més, G' és una extensió de G .

Demostració. Sigui $G = (V, B, P)$ una gramàtica contextual amb selecció regular. Construïm una seqüència $G_1, G_2, \dots, G_n$ de gramàtiques contextuais $G_i = (V, B, P_i)$, $1 \leq i \leq n$, amb selecció regular, de la següent manera:

1. $P_1 = P$,
2. (C1) Si existeixen $\pi_1 = (S_1, (u_1, v_1))$, $\pi_2 = (S_2, (u_2, v_2))$ a P_i
tal que per a cada $\pi_3 \in P_i$ tenim $\pi_3 \neq (\mu[\pi_1, \pi_2], (u_1, v_1))$, aleshores
$$P_{i+1} = P_i \cup \{(\mu[\pi_1, \pi_2], (u_1, v_1))\},$$

anar a 2;

3. (C2) En tots els casos en que $n = i$, parar.

La gramàtica que busquem és $G' = G_n$ donada mitjançant aquesta construcció.

Observació 9.1.2 (Correcció de la definició) Existeix un i tal que (C2) es compleix (i, per tant, el procediment acaba).

Demostració. Sigui $\mathcal{T}$ el conjunt de tots els sistemes de transició que s'amaguen sota els autòmats finits deterministes M , per a les produccions $(S, (u, v))$ de P . Aquest és un conjunt finit. Pel Lema 9.1.15 de la pàgina 320, per a cada producció $(S', (u', v'))$ de $P_i, 1 \leq i \leq n$, el sistema de transició adjacent M de M' pertany a $cl(\mathcal{T})$. Pel Lema 9.1.14 de la pàgina 318, aquest conjunt és finit. Pel Lema 9.1.13 de la pàgina 318, el conjunt d'autòmats finits que correspon a sistemes de transició a $cl(\mathcal{T})$ també és finit. Per tant tota P_i és un subconjunt d'un conjunt finit fixat de produccions, i.e, la construcció s'acaba. $\square$

Observació 9.1.3 Per a cada producció $\pi_3 \in P_n$ existeix una producció $\pi_1 \in P$ tal que $\pi_3 \subseteq \pi_1$.

Demostració. Provarem per inducció sobre i que per a cada $\pi_3 \in P_i$ existeix $\pi_1 \in P$ tal que $\pi_3 \subseteq \pi_1$.

Per a $i = 1$ l'afirmació es compleix trivialment, $\pi_3 \subseteq \pi_3 \in P_1 = P$.

Suposem que l'afirmació és verdadera per a tot $i \leq j$, per a algun $j \geq 1$, i considerem P_{j+1} .

Tot $\pi_3 \in P_{j+1}$ pertany a P_j i aleshores, per la hipòtesi d'inducció, $\pi_3 \subseteq \pi_1 \in P$; o bé $\pi_3 \in P_{j+1} - P_j$. En aquest últim cas $\pi_3 = (\mu[\pi_4, \pi_2], (u_1, v_1))$ per a algun $\pi_4, \pi_2 \in P_j$. A partir de la construcció de $\mu[\pi_4, \pi_2]$ se segueix que $\pi_3 \subseteq \pi_4$, i a partir de la hipòtesi d'inducció tenim que $\pi_4 \subseteq \pi_1$ per a $\pi_1 \in P$. Conseqüentment, $\pi_3 \subseteq \pi_1 \in P$, la qual cosa conclou la demostració de l'observació. $\square$

Observació 9.1.4 $L_{ex}(G_n) = L_{ex}(G)$.

Demostració. A partir de la construcció, per a tot $1 \leq i \leq n-1$, $P_{i+1} = P_i \cup \{(\mu[\pi_1, \pi_2], u_1, v_1)\}$ per a algun $\pi_1, \pi_2 \in P_i$ i, per tant, pel Lema 9.1.10 de la pàgina 315, tenim que $L_{ex}(G_i) = L_{ex}(G_{i+1})$. Això implica que $L_{ex}(G_i) = L_{ex}(G_1) = L_{ex}(G)$, per a tot $i = 1, 2, \dots, n$. $\square$

Observació 9.1.5 Per a cada $z \in L_{ex}(G_n)$ existeix una derivació encaixant a G_n que produeix z .

Demostració. Ja que $z \in L_{ex}(G_n) = L_{ex}(G)$, existeix una derivació per a z a G ,

$$w_j \xRightarrow{\pi_1}_{ex} u_1 w_j v_1 \xRightarrow{\pi_2}_{ex} \dots \xRightarrow{\pi_m}_{ex} u_m \dots u_1 w_j v_1 \dots v_m = z,$$

usant les produccions $\pi_1, \pi_2, \dots, \pi_m$, on $\pi_k = (S_k, (u_k, v_k))$, per a $1 \leq k \leq m$, $w_j \in B \cap S_1$, i $u_{k-1} \dots u_1 w_j v_1 \dots v_{k-1} \in S_k$, per a tot $2 \leq k \leq m$.

Ara, començant per les produccions $\pi_1, \pi_2, \dots, \pi_m$, construïm (a l'inrevés, començant per m) una seqüència de produccions $\pi'_1, \dots, \pi'_m$ a P_n , de la següent

manera:

1. $\pi'_m = \pi_m$,
2. per a cada $2 \leq i \leq m$, π'_{i-1} és una producció $\pi_{0,i-1}$ de P_n tal que

$$\pi_{0,i-1} = (\mu[\pi_{i-1}, \pi'_i], (u_{i-1}, v_{i-1})).$$

Des del moment en què, en la construcció de P_n , la condició (C2) és satisfeta, quan P_n és obtingut; existeix una producció $\pi_{0,i-1}$ com la d'abans a P_n i, per això, π'_{i-1} pot ser construïda.

Ara, començant per w_j i usant les produccions $\pi'_1, \pi'_2, \dots, \pi'_m$, en aquest ordre, obtenim una derivació per a z a G_n . Segons la definició de $\pi'_1, \dots, \pi'_m$, aquesta és una derivació encaixant ($\mu[\pi_{i-1}, \pi'_i]$ assegura que $u_{i-1}xv_{i-1} \in S'_i$, per a $\pi'_i = (S'_i, (u_i, v_i))$ per a tot $x \in \mu[\pi_{i-1}, \pi'_i]$).

Conseqüentment,

- (i) G' existeix — per l'Observació 9.1.2;
- (ii) G' és una extensió de G — per construcció;
- (iii) G' és externament equivalent amb G — per l'Observació 9.1.4;
- (iv) G' és en la forma encaixant — per l'Observació 9.1.5.

Això conclou la demostració. □

Corol·lari 9.1.2 $EC(REG) \subseteq LIN$.

La igualtat $FIN = ECC(FIN)$ fa que les gramàtiques contextuais amb

FIN-selecció treballant en el mode extern no siguin interessants. Aquest resultat mostra, de fet, que la selecció finita no està definida apropiadament per a aquestes gramàtiques. Una variant molt més natural és deixar que el context que és adjuntat depengui d'un prefix finit, d'un sufix finit de la cadena en qüestió.

Definició 9.1.5 Una *gramàtica contextual amb selecció acotada* (externa) és una construcció

$$G = (V, B, (S_1, C_1), \dots, (S_n, C_n)), n \geq 1,$$

on V és un vocabulari, B és un conjunt finit de V^* , S_i són conjunts finits de $V^* \times V^*$, i C_i són conjunts finits de $V^* \times V^*$, $1 \leq i \leq n$.

Per a aquesta gramàtica definim la relació $\Rightarrow_{ex}$ a V^* mitjançant

$$\begin{aligned} x \Rightarrow_{ex} y \quad \text{sii} \quad x = x_1 x_2 x_3, y = u x v, \text{ on} \\ x_1, x_2, x_3 \in V^*, (x_1, x_3) \in S_i, (u, v) \in C_i, \\ \text{per a algun } 1 \leq i \leq n. \end{aligned}$$

Per tant, el context $(u, v) \in C_i$ només és adjuntat a cadenes que tinguin un parell d'un prefix i d'un sufix a S_i . El llenguatge generat, $L_{ex}(G)$, es defineix de la manera usual. Denotem

$$k = \max\{|x| \mid (x, y) \in S_i \text{ o } (y, x) \in S_i, 1 \leq i \leq n\}.$$

S'anomena la (selectiva) *profunditat* de G .

Denotem per $ECC(k)$, $k \geq 0$, la família de llenguatges generats externament per gramàtiques amb selecció acotada, com abans, de profunditat com a màxim k . Denotem també $ECC(\infty) = \bigcup_{k \geq 0} ECC(k)$.

Teorema 9.1.3 $EC = ECC(0) \subset ECC(1) \subset \dots \subset ECC(\infty) \subset ECC(REG)$.

Demostració. Només necessitem provar que les incusions són estrictes.

Pel fet que $a^+ \cup b^+$ pot ser generada per $G = (\{a, b\}, \{a, b\}, (\{(\lambda, a)\}, \{(\lambda, a)\}), (\{(\lambda, b)\}, \{(\lambda, b)\}))$, tenim que $ECC(1) - ECC(0) \neq \emptyset$.

De $k \geq 2$ obtenim que $ECC(k) - ECC(k-1) \neq \emptyset$ usant el llenguatge

$$L_k = a^k b (aa^{k-1})^+ \cup a^k b (ba^{k-1})^+.$$

En aquest cas, $L_k = L_{ex}(G_k)$, per a

$$G_k = (\{a, b\}, \{a^k ba^k, a^k bba^{k-1}\}, (\{(\lambda, a^k)\}, \{(\lambda, a^k)\}), (\{(\lambda, ba^{k-1})\}, \{(\lambda, ba^{k-1})\})).$$

Suposem que $L_k = L_{ex}(G)$ per a alguna $G = (\{a, b\}, B, (S_1, C_1), \dots, (S_n, C_n))$, amb $S_i \subseteq \{a, b, \lambda\}^{k-1} \times \{a, b, \lambda\}^{k-1}$. De la forma de les cadenes de L_k , per a tot $(x_1, x_2) \in S_i$ el qual és usat en una derivació de G hem de tenir $x_1 = a^p$, $x_2 = a^q$, per a algun $p, q \leq k-1$. Pel fet que $a^k b (ba^{k-1})^s$ pertany a L_k per a tot $s \geq 1$, existeix un context $(\lambda, (ba^{k-1})^j)$, $j \geq 1$. Pel fet que els selectors usats en les derivacions són de la forma (a^p, a^q) discutida abans, podem derivar $a^k ba^k \xRightarrow{ex} a^k ba^k (ba^{k-1})^j$, que no pertany a L_k ; la qual cosa és una contradicció.

A més a més, el llenguatge

$$L = \{a^n ba^n, ba^n ba^n b, a^n bba^n, bba^n bba^n bb \mid n \geq 1\}$$

pertany a $ECC(REG) - ECC(\infty)$. En aquest cas, per a la gramàtica

$$G = (\{a, b\}, \{aba, abba\}, (a^+ ba^+, \{(a, a), (b, b)\}), (a^+ bba^+, \{(a, a), (bb, bb)\}))$$

tenim que $L_{ex}(G) = L$. No obstant, $L \notin ECC(\infty)$. Aquest fet pot ser vist de la següent manera. Les cadenes $ba^n ba^n b, bba^n bba^n bb$, amb n arbitràriament gran, només poden ser generades afegint un context $(ba^{i_1}, a^{i_2}b), (bba^{i_3}, a^{i_4}bb)$ a alguna cadena $a^{n_1}ba^{n_2}, a^{n_3}bba^{n_4}$, respectivament. Els selectors associats a aquests contextos són de la forma $(a^{m_1}, a^{m_2}), (a^{m_3}, a^{m_4})$, respectivament, per a determinats nombres enters acotats m_1, m_2, m_3, m_4 . D'aquesta manera podem produir cadenes de les formes $ba^n bba^n b$ i $bba^n ba^n bb$, les quals no pertanyen a L ; la qual cosa és una contradicció. $\square$

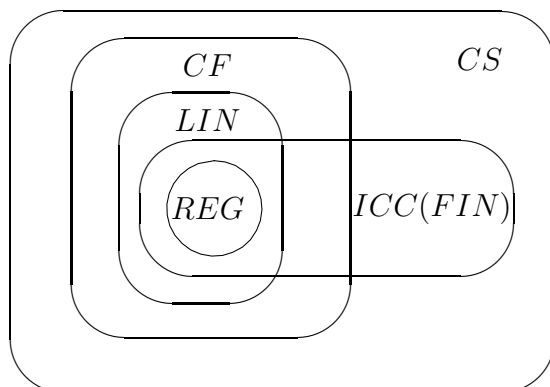
Les famílies més interessants $ICC(F), ECC(F)$ són les més inferiors de les branques verticals d'inclusions de la dreta i de l'esquerra en la Figura 9.1.1 de la pàgina 310. És a dir, $ICC(FIN)$ i $ECC(REG)$. Ens centrarem, a les dues pròximes seccions, en aquests dos casos particulars.

9.2 Gramàtiques contextuais internes amb selecció finita

Coneixem ja l'emplaçament de la família $ICC(FIN)$ en la jerarquia de Chomsky – veure la Figura 9.2.1 que mostra d'una altra manera la informació con-

tinguda en la Figura 9.1.1 de la pàgina 310.

Figura 9.2.1 La família $ICC(FIN)$ en la jerarquia de Chomsky.



En la Secció 8.1 (Teorema 8.1.3 de la pàgina 245) vam veure que la família $ICC(REG)$ conté llenguatges no-semilineals. Aquest resultat pot ser intensificat.

Teorema 9.2.1 La família $ICC(FIN)$ conté llenguatges no-semilineals.

Demostració. Considerem la gramàtica $G = (V, B, P)$, on

$$V = \{a, b, c, d, e, f, g\},$$

$$B = \{dcabaababcd\},$$

i el conjunt P que conté les produccions següents:

regles creuades:

d'esquerra a dreta: $ED : (\{abaab\}, \{(b, a)\}),$

de dreta a esquerra: $DE : (\{abbab\}, \{(b, a)\}),$

regles per als extrems:

per a l'extrem esquerre: $E1 : (\{dcbbab\}, \{(\lambda, a)\}),$

$E2 : (\{dcb\}, \{(\lambda, aa)\}),$

$E3 : (\{dc\}, \{(\lambda, a)\}),$

per a l'extrem dret: $D1 : (\{aabacd\}, \{(b, \lambda)\}),$

$D2 : (\{bbaaba\}, \{(\lambda, b)\}),$

$D3 : (\{abba\}, \{(\lambda, bb)\}),$

regles d'acabament:

$A1 : (\{baabac\}, \{(\lambda, e)\}),$

$A2 : (\{baabaced\}, \{(g, \lambda)\}),$

$A3 : (\{bagb\}, \{(g, \lambda)\}),$

$A4 : (\{cgb\}, \{(f, \lambda)\}).$

Provarem que $L_{in}(G)$ no és semilineal. Per a fer-ho, considerem el conjunt semilineal

$$R = \{(n_1, n_2, 2, 2, 1, 1, n_3) \mid n_1, n_2, n_3 \geq 1\}.$$

Pel fet que la família de conjunts semilineals de vectors és tancada sota l'operació d'intersecció, per tal de provar que $L_{in}(G)$ no és semilineal, és suficient amb veure que el conjunt

$$N = \Psi_V(L_{in}(G)) \cap R$$

no és semilineal.

Denotem

$$W(N) = \{w \in L_{in}(G) \mid \Psi_V(w) \in N\}.$$

Per obtenir una cadena de $W(N)$, la derivació ha de seguir els passos.

(1) Començar amb una cadena de la forma

$$dcabaab(ab)^n cd$$

(per a l'axioma tenim $n = 1$).

(2) Aplicar la regla d'encreuament d'esquerra a dreta, ED, de la manera següent:

$$\begin{aligned} dcabaab(ab)^n cd &\Rightarrow_{in} dc(ba)^1 baabaab(ab)^{n-1} cd \\ &\Rightarrow_{in} dc(ba)^3 baabaab(ab)^{n-2} cd \Rightarrow_{in} dc(ba)^5 baabaab(ab)^{n-3} cd \\ &\Rightarrow_{in} \dots \Rightarrow_{in} dc(ba)^{2(n-2)-1} baabaab(ab)^2 cd \\ &\Rightarrow_{in} dc(ba)^{2(n-1)-1} baabaab(ab)^1 cd \\ &\Rightarrow_{in} dc(ba)^{2n-1} baabaabcd \Rightarrow_{in} dc(ba)^{2n+1} baabacd. \end{aligned}$$

(3) Escollir, de manera no-determinista, la variant (3.1) o (3.2).

(3.1) Aplicar les regles per l'extrem dret, D1, D2, D3:

$$\begin{aligned} dc(ba)^{2n+1} baabacd &\Rightarrow_{in} dc(ba)^{2n+1} bbaabacd \\ &\Rightarrow_{in} dc(ba)^{2n} babbaabacd \Rightarrow_{in} dc(ba)^{2n} babbabb(ab)^2 cd \end{aligned}$$

i anar al pas (4).

(3.2) Anar al pas (7), per a acabar la derivació i obtenir una cadena de $W(N)$.

(4) Aplicar la regla d'encreuament de dreta a esquerra, DE, de manera.

$$\begin{aligned}
 dc(ba)^{2n}babbabb(ab)^2cd &\Longrightarrow_{in} dc(ba)^{2n-1}babbabb(ab)^4cd \\
 &\Longrightarrow_{in} dc(ba)^{2n-2}babbabb(ab)^6cd \Longrightarrow_{in} dc(ba)^{2n-3}babbabb(ab)^8cd \\
 &\Longrightarrow_{in} \dots \Longrightarrow_{in} dc(ba)^2babbabb(ab)^{2(2n-1)}cd \\
 &\Longrightarrow_{in} dcbababbabb(ab)^{2(2n)}cd \Longrightarrow_{in} dcbabbabb(ab)^{2(2n+1)}cd \\
 &\Longrightarrow_{in} dcbbabb(ab)^{2(2n+2)}cd.
 \end{aligned}$$

(5) Aplicar les regles per l'extrem esquerre, E1, E2, E3:

$$\begin{aligned}
 dcbbabb(ab)^{2(2n+2)}cd &\Longrightarrow_{in} dcbbabab(ab)^{2(2n+2)}cd \\
 &\Longrightarrow_{in} dcbaababab(ab)^{2(2n+2)}cd \Longrightarrow_{in} dcabaab(ab)^{2(2n+3)}cd.
 \end{aligned}$$

(6) Amb la cadena obtinguda, anar al pas (1).

(7) Acabar la derivació aplicant la regla d'acabament, A1, A2, A3, i A4 (la cadena prové del pas (3.2) i, per tant, té la forma mostrada a continuació):

$$\begin{aligned}
dc(ba)^m baabacd &\Rightarrow_{in} dc(ba)^m baabaced \\
&\Rightarrow_{in} dc(ba)^{m-1} bagbaabaced \Rightarrow_{in} dc(ba)^{m-2} bagbagbaabaced \\
&\Rightarrow_{in} dc(ba)^{m-3} bagba(gba)^2 abaced \\
&\Rightarrow_{in} dc(ba)^{m-4} bagba(gab)^3 abaced \\
&\Rightarrow_{in} \dots \Rightarrow_{in} dc(ba)^1 bagba(gba)^{m-2} abaced \\
&\Rightarrow_{in} dcbagba(gba)^{m-1} abaced \\
&\Rightarrow_{in} dcgba(gba)^m abaced \Rightarrow_{in} dfc(gba)^{m+1} abaced.
\end{aligned}$$

Obtenim una cadena de $W(N)$.

Suposant que hem començat amb la cadena $dcabaababcd$ i que hem realitzat la seqüència de passos:

$$\underbrace{1, 2, 3.1, 4, 5, 6, 1, 2, 3.1, 4, 5, 6, \dots, 1, 2, 3.1, 4, 5, 6, 1, 2, 3.2, 7,}_{r \geq 1 \text{ vegades}}$$

aleshores, el valor d' m al final del pas (7) és

$$m = 2^{2r+3} - 11.$$

Per tant, si tenim en compte que qualsevol seqüència de passos que produeix una cadena de $W(N)$ és de la forma descrita anteriorment, obtenim

$$\begin{aligned}
N &= \{(2^{2r+3} - 8, 2^{2r+3} - 9, 2, 2, 1, 1, 2^{2r+3} - 10) \mid r \geq 1\} \\
&\cup \{(6, 5, 2, 2, 1, 1, 4)\},
\end{aligned}$$

que, òbviamment, és no semilineal.

En aquest punt, ens quedaria per provar que, per tal d'obtenir una cadena a $W(N)$, hem de seguir derivacions com les descrites anteriorment. Aquest fet pot ser obtingut examinant la forma de les produccions de P i la forma de les cadenes obtingudes després de cada pas de derivació (els detall d'aquesta demostració poden ser trobats a (Ilie, 1996b)). $\square$

Recordem també, del Capítol 6 – veure, per exemple, la Taula 6.12.1 de la pàgina 228 – que ICC és un anti-AFL, mentre que IC és tancada sota les operacions de morfismes, substitució finita i imatge mirall. També per a $ICC(FIN)$ tenim

Lema 9.2.1 $ICC(FIN)$ és un anti-AFL.

Demostració. Moltes de les demostracions de les propietats de clausura negatives indicades al Capítol 6 per a la família ICC es converteixen en afirmatives per a $ICC(FIN)$, directament o mitjançant senzilles modificacions. Apareix alguna dificultat en el cas dels morfismes inversos i, per tant, considerarem aquest cas en detall.

Considerem la gramàtica contextual $G = (\{a, b, c, d\}, \{ccbc dc\}, (\{ccbc dc\}, \{(ab, ab)\}), (\{cc\}, \{(\lambda, ba)\}), (\{c\}, \{(d, a)\}))$, i el morfisme $h : \{a, b\}^* \longrightarrow \{a, b, c, d\}^*$ definit per $h(a) = ab$, $h(b) = cdc$. Tota cadena de $L_{in}(G)$ conté exactament quatre ocurrences del símbol c i, per tant, totes les cadenes x tal que $h(x) \in L_{in}(G)$ han de contenir exactament dues ocurrences de b . Per tant, $h(x)$ és de la forma

$x_1cdc x_2cdc x_3$, amb $x_1, x_2, x_3 \in \{a, b\}^*$. Això vol dir que el context (d, a) és usat només una vegada i, per tant, $h(x) = (ab)^n cdc(ab)^m cdc(ab)^n$, $n \geq 0, m \geq 1$. La derivació d'aquesta cadena té lloc de la següent manera:

$$\begin{aligned} ccbcdc &\Longrightarrow_{in}^* (ab)^n ccbcdc(ab)^n \Longrightarrow_{in}^* (ab)^n cc(ba)^{m-1} bcdc(ab)^n \\ &\Longrightarrow_{in}^* (ab)^n cdca(ba)^{m-1} bcdc(ab)^n = (ab)^n cdc(ab)^m cdc(ab)^n. \end{aligned}$$

Conseqüentment, $h^{-1}(L_{in}(G)) = \{a^n ba^m ba^n \mid n \geq 0, m \geq 1\}$, un llenguatge que no pertany a la família $ICC(FIN)$ (ni tampoc a ICC). $\square$

A la vista d'aquestes propietats de clausura (negatives), és d'interès el considerar la compleció de $ICC(FIN)$ respecte a certes operacions AFL. El resultat és espectacular: el trio complet – veure la Definició 3.6.6 de la pàgina 145 – més petit que conté $ICC(FIN)$ és RE . De fet, un resultat encara més puixant també es compleix (que no involucra el tancament sota l'operació d'intersecció amb conjunts regulars).

Teorema 9.2.2 Tot llenguatge recursivament enumerable, L , pot ser escrit de la forma $L = h_1(h_2^{-1}(L'))$, on $L' \in ICC(FIN)$, h_1 és una codificació dèbil (veure la corresponent Definició 3.2.12 de la pàgina 109) i h_2 és un morfisme.

Demostració. Prenem $L \subseteq V_T^*$, $L \in RE$ i una gramàtica de tipus-0 $G = (V_N, V_T, S, P)$ per a L . Considerem tres nous símbols, $[,], \vdash$, i construïm la gramàtica contextual $G' = (V, B, P')$, amb

$$\begin{aligned} V &= V_N \cup V_T \cup \{[,], \vdash\}, \\ B &= \{S\}, \end{aligned}$$

i el conjunt P' que consisteix en les següents produccions:

1. $(\{u\}, \{([,]v)\})$, per a cada $u \rightarrow v \in P$,
2. $(\{\alpha[u]\}, \{(\vdash, \alpha)\})$, per a $\alpha \in V_N \cup V_T, u \rightarrow v \in P$,
3. $(\{\alpha \vdash \beta\}, \{(\vdash, \alpha)\})$, per a $\alpha, \beta \in V_N \cup V_T$.

Considerem un nou símbol, b_w , associat a cada cadena w en el conjunt

$$U = \{[u] \mid u \rightarrow v \in P\} \cup \{\vdash \alpha \mid \alpha \in V_N \cup V_T\}.$$

Denotem per W el conjunt de tots aquests símbols b_w i definim la codificació dèbil $h_1 : (W \cup V_T)^* \longrightarrow V_T^*$ mitjançant

$$h_1(b_w) = \lambda, w \in U, \quad h_1(a) = a, a \in V_T,$$

així com el morfisme $h_2 : (W \cup V_T)^* \longrightarrow (V_N \cup V_T \cup \{[,], \vdash\})^*$ mitjançant

$$h_2(b_w) = w, w \in U, \quad h_2(a) = a, a \in V_T.$$

La intuïció que s'amaga darrera de la construcció de G' (i de h_1, h_2) és la següent. Els símbols entre $[\]$ i el símbol immediat a la dreta de $\vdash$ són considerats *morts*; tots els altres símbols d'una cadena de $L_{in}(G)$, exceptuant $[,], \vdash$, estan *vius* (els símbols $[,], \vdash$ s'anomenen també *assassins*). Els símbols vius d'una cadena $z \in L_{in}(G')$ es corresponen amb una forma sentencial produïda per G . Simulem una derivació a G amb els símbols vius de z , usant produccions del grup 1 de P' . D'aquesta manera, obtenim nous símbols vius i morts.

La seva posició pot ser canviada usant produccions dels grups 2, 3 anteriors (els símbols vius poden ser desplaçats cap a la dreta, creuant els obstacles $w \in U$, dels símbols morts). El morfisme invers h_2^{-1} només pot ser aplicat quan és obtinguda una cadena de $(U \cup V_T)^*$. Això implica que tots els símbols no-terminals estan morts, la derivació corresponent a G és una derivació terminal. La codificació h_1 esborra tots els símbols b_w i, per tant, obtenim una cadena de V_T^* .

Usant aquestes explicacions podem provar la igualtat $L = h_1(h_2^{-1}(L_{in}(G')))$.

No entrarem en detalls, només exemplificarem l'anterior construcció començant per una gramàtica per al llenguatge

$$L = \{a^n b^n c^n \mid n \geq 1\}.$$

Prenem $G = (\{S, A\}, \{a, b, c\}, S, P)$, amb

$$P = \{S \rightarrow abc, S \rightarrow aSAc, cA \rightarrow Ac, bA \rightarrow bb\}.$$

La gramàtica contextual corresponent és $G = (V, B, P')$, amb

$$V = \{S, B, a, b, c, [,], \vdash\},$$

$$A = \{S\},$$

$$P' = \{(S, ([,]abc)), (S, ([,]aSBc)), (cB, ([,]Bc)), (bB, ([,]bb))\}$$

$$\cup \{(\alpha[S], (\vdash, \alpha)), (\alpha[cB], (\vdash, \alpha)), (\alpha[bB], (\vdash, \alpha)) \mid \alpha \in \{S, B, a, b, c\}\}$$

$$\cup \{(\alpha \vdash \beta, (\vdash, \alpha)) \mid \alpha, \beta \in \{S, B, a, b, c\}\}.$$

Per tal de simplificar l'escriptura, hem escrit $(x, (u, v))$ en comptes de $(\{x\}, \{(u, v)\})$.

Considerem una derivació a G , per exemple, per a la cadena $a^2b^2c^2$,

$$S \Longrightarrow aSAc \Longrightarrow aabcAc \Longrightarrow aabAcc \Longrightarrow aabbcc.$$

A continuació veurem com G' procedeix per tal de simular aquesta derivació per tal de produir una cadena per a la qual h_2^{-1} sigui definit:

$$\begin{aligned} S &\Longrightarrow_{in} [S]\underline{a}SAc \\ &\Longrightarrow_{in} [S]\underline{a}[S]\underline{abc}Ac \\ &\Longrightarrow_{in} [S]\underline{a}[S]\underline{ab}[cA]\underline{Acc} \\ &\Longrightarrow_{in} [S]\underline{a}[S]\underline{a} \vdash b[cA]\underline{bAcc} \\ &\Longrightarrow_{in} [S]\underline{a}[S]\underline{a} \vdash b[cA][bA]\underline{bbcc} = z. \end{aligned}$$

Tots els símbols subratllats estan vius, els altres estan morts. Mitjançant h_2^{-1} , cada bloc $[S], \vdash b, [cA], [bA]$ és reemplaçat pel símbol associat b_w , que posteriorment és esborrat per h_1 . A més a més, $h_1(h_2^{-1}(z)) = aabbcc$. $\square$

L'anterior construcció pot ser modificada per tal d'obtenir

Corol·lari 9.2.1 Tot llenguatge $L \in RE$ pot ser escrit de la forma $L = L_1 \setminus L_2$, per a $L_1 \in REG, L_2 \in ICC(FIN)$.

Demostració. Començant per una gramàtica de tipus-0 $G = (V_N, V_T, S, P)$, construïm la gramàtica G' com abans i, després, procedim de la manera següent:

- afegim el nou símbol $\#$ al vocabulari de G' ,
- posem $B = \{S\#$ en comptes de $B = \{S\}$,
- afegim les produccions de la forma

$$4. \quad (\{\alpha\# \}, \{(\vdash, \alpha)\}), \alpha \in V_T.$$

Denotem per G'' la gramàtica contextual obtinguda.

Obtenim $L = (R\#) \setminus L_{in}(G'')$, on R és el llenguatge regular següent:

$$R = (\{[u] \mid u \rightarrow v \in P\} \cup \{\vdash \alpha \mid \alpha \in V_N \cup V_T\})^*.$$

En aquest cas, les produccions de la forma 4 poden “transportar cap a la dreta” els símbols terminals vius, passant pel marcador $\#$, preservant la seva ordenació. El quocient per $R\#$ selecciona del conjunt de cadenes generades per G'' només aquelles cadenes de la forma $w_1\#w_2$ amb $w_1 \in R$ i, per tant, que consisteixen en blocs $[u], \vdash \alpha$; és a dir, símbols morts. Per tant, les regles de la forma 4 més el quocient per $R\#$ fan el mateix treball que h_1, h_2^{-1} en la demostració del Teorema 9.2.2 de la pàgina 334. $\square$

A (Latteux et al., 1985) és provat que tot llenguatge recursivament enumerable pot ser escrit com el quocient de dos llenguatges lineals. El resultat del Corol·lari 9.2.1 no és una conseqüència del resultat de (Latteux et al., 1985) (o a la inversa), ja que $ICC(FIN)$ i LIN són famílies incomparables (veure el Lema 5.3.6 de la pàgina 203).

Corol·lari 9.2.2 Tot llenguatge $L \in RE$ pot ser escrit de la forma $L = g(L')$, per a $L' \in ICC(FIN)$ i g una *gsm* (veure la Definició 3.6.5 de la pàgina 142).

Demostració. Una *gsm* pot simular simultàniament el treball de h_1 i de h_2^{-1} en la demostració del Teorema 9.2.2 de la pàgina 334. $\square$

Una interessant conseqüència d'aquestes representacions és

Teorema 9.2.3 La família $ICC(FIN)$ és incomparable amb tota família F tal que $LIN \subseteq F \subset RE$ i F sigui tancada

- 1) sota l'operació de quocient per l'esquerre amb conjunts regulars, o
- 2) sota codificacions dèbils i morfismes inversos.

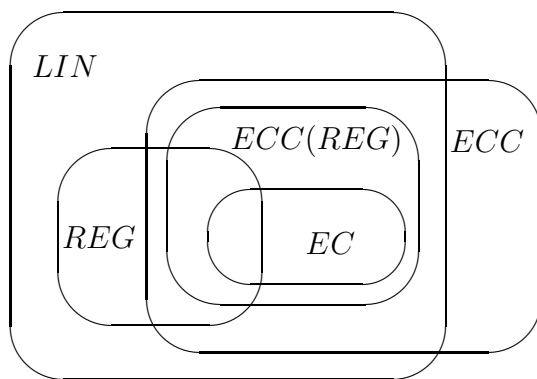
Demostració. Sabem que $LIN - ICC \neq \emptyset$ (veure el Lema 5.3.6 de la pàgina 203) i, per tant, $F - ICC(FIN) \neq \emptyset$ per a tot F com abans. Inversament, si $ICC(FIN) \subseteq F$, pel fet que F té les propietats de clausura mencionades, se segueix del Teorema 9.2.2 de la pàgina 334 i dels seus Corol·laris 9.2.1 i 9.2.2, que $RE \subseteq F$, la qual cosa contradiu la inclusió estricta $F \subset RE$. Per tant, F és incomparable amb $ICC(FIN)$. $\square$

Per exemple, una família F que compleix les condicions del Teorema 9.2.3 de la pàgina 339 és MAT^λ , de llenguatges generats per gramàtiques matricials amb regles independents del context arbitràries (però sense control d'aparició; veure la Secció 3.4.2). Aquesta família és un semi-AFL total i està estrictament

inclosa a RE . D'acord amb el teorema previ, $ICC(FIN) - MAT^\lambda \neq \emptyset$. Ja que $ICC(FIN) \subseteq CS$, obtenim $CS - MAT^\lambda \neq \emptyset$, una qüestió que havia restat per resoldre durant molt de temps ((Salomaa, 1973), (Dassow i Păun, 1989)) i només recentment ha estat resolt – veure (Hauschild i Jantzen, 1994) i (Dassow et al., 1996a)). Una altra important família F com abans és $ET0L$, de llenguatges generats per sistemes de Lindenmayer taulats extesos sense interaccions (veure la Secció 3.5) i, per tant, $ICC(FIN) - ET0L \neq \emptyset$.

9.3 Gramàtiques contextuais externes amb selecció regular

Figura 9.3.1 Aquesta figura mostra una representació gràfica de les relacions entre $ECC(REG)$ i d'altres famílies de la Figura 9.1.1 de la pàgina 310.



En la Taula 6.12.1 de la pàgina 228, vèiem que $ECC(REG)$ no és tancada sota les operacions de concatenació, clausura de Kleene, intersecció amb conjunts regulars i morfismes inversos. A més a més, tenim

Lema 9.3.1 La família $ECC(REG)$ és no-tancada sota les operacions amb morfismes.

Demostració. Per a la gramàtica

$$G = (\{a, b, c, d\}, \{b\}, (d^*bd^*, \{(d, \lambda), (\lambda, d)\}), (\{b\}, \{(a, \lambda)\}), \\ (a^+ba^*, \{(a, \lambda), (a, a)\}), (a^+ba^+, \{(\lambda, c)\})),$$

obtenim

$$L_{ex}(G) = d^*bd^* \cup \{a^nba^m, a^nba^mc \mid n > m \geq 1\}.$$

Considerem el morfisme $h : \{a, b, c, d\}^* \longrightarrow \{a, b, c\}^*$ definida per $h(a) = a$, $h(b) = b$, $h(c) = c$, $h(d) = a$.

El llenguatge

$$h(L_{ex}(G)) = a^*ba^* \cup \{a^nba^mc \mid n > m \geq 1\}$$

no pertany a $ECC(REG)$. Suposem el contrari, i prenem $G' = (\{a, b, c\}, B, (S_1, C_1), \dots, (S_r, C_r))$ tal que $L_{ex}(G') = h(L_{ex}(G))$. Tota cadena de $L_{ex}(G')$ conté el símbol b i, per tant, totes les cadenes de B contenen una ocurrència del símbol b . A partir de cadenes de la forma $a^i ba^m c$ només podem produir cadenes de la forma $a^n ba^m c$. Conseqüentment, per a un valor de m suficientment

gran, les cadenes $a^n b a^m c$ són produïdes per derivacions que comencen a partir d'axiomes de la forma $a^i b a^j, i, j \geq 0$. Considerem el pas en què el símbol c és introduït, $a^{n_1} b a^{m_1} \Rightarrow_{ex} a^{n_2} a^{n_1} b a^{m_1} a^{m_2} c$. Això vol dir que existeix una producció (S_j, C_j) a G' amb $a^{n_1} b a^{m_1} \in S_j$, $(a^{n_2}, a^{m_2} c) \in C_j$. Ja que $m = m_1 + m_2$ pot ser arbitràriament gran i m_2 és acotat, se segueix que m_1 pot ser arbitràriament gran. Usant un lema de bombeig per al llenguatge regular S_j , trobem una constant $s \geq 1$ tal que $a^{n_1} b a^{m_1+ts} \in S_j$ per a tot $t \geq 0$. Totes les cadenes $a^{n_1} b a^{m_1+ts}$ pertanyen a $h(L_{ex}(G))$ i, per tant, podem realitzar la derivació $a^{n_1} b a^{m_1+ts} \Rightarrow_{ex} a^{n_1+n_2} b a^{m_1+m_2+ts} c$. Si t és suficientment gran, aleshores podem tenir $n_1 + n_2 < m_1 + m_2 + ts$; aquesta cadena no pertany a $h(L_{ex}(G))$; la qual cosa és una contradicció. La igualtat $h(L_{ex}(G)) = L_{ex}(G')$ no és certa, $h(L_{ex}(G)) \notin ECC(REG)$. $\square$

La família $ECC(REG)$ és tancada sota l'operació d'unió. Aquest resultat es basa essencialment en el fet que treballem amb un conjunt finit d'axiomes, no amb un únic axioma. Per exemple, el llenguatge $\{a, b\}$ no pot ser generat per una gramàtica contextual amb només un axioma. Això suggereix considerar la família més petita de llenguatges generada per gramàtiques amb selecció regular; és a dir, gramàtiques amb només un axioma. Denotem aquesta família mitjançant $ECC_1(REG)$. D'acord amb les observacions anteriors, aquesta família és un anti-AFL.

Investigarem ara la compleció d'aquesta família respecte a la unió, la intersecció amb conjunts regulars i morfismes. Més exactament, considerarem

llenguatges de la forma

$$h((L_1 \cup L_2 \cup \dots \cup L_n) \cap R),$$

on h és un morfisme, R és un llenguatge regular, i $L_1, \dots, L_n$ són llenguatges de $ECC_1(REG)$. Considerem els següents casos (la notació per a cadascun és donada entre parèntesis):

(A1)	$n = 1$ i $L_1 \in EC_1(REG)$	(Au = axioma únic)
(A2)	$n = 1$ i $L_1 \in ECC(REG)$	(Fa = conjunt finit d'axiomes)
(A3)	n arbitrari, $L_i \in EC_1(REG)$,	(Uf = unió finita)
	$1 \leq i \leq n$	

(B1)	$R = V^*$	(Si = compressió simple)
(B2)	R arbitrari	(Re = compressió regular)

(C1)	h és la identitat	(Id = identitat)
(C2)	h és una codificació	(Co = codificació)
(C3)	h és una codificació dèbil	(Cd = codificació dèbil)
(C4)	h és no-esborrador	(Mn = morfisme no-esborrador)
(C5)	h és arbitrari	(Ma = morfisme arbitrari)

Conseqüentment, obtenim $3 \times 2 \times 5 = 30$ casos diferents, que ens donen 30 famílies de llenguatges. Identificarem aquestes famílies per la tripla de les abreviacions corresponents als casos involucrats A, B, C . Aleshores, mit-

jançant $\mathcal{L}(X_1, X_2, X_3)$, on (X_1, X_2, X_3) és un element de

$$\{Au, Fa, Uf\} \times \{Si, Re\} \times \{Id, Co, Cd, Mn, Ma\}.$$

Per exemple, segons les definicions, tenim $ECC_1(REG) = \mathcal{L}(Au, Si, Id)$, i $ECC(REG) = \mathcal{L}(Fa, Si, Id)$.

Totes les 30 famílies considerades anteriorment estan incloses en LIN , ja que $ECC_1(REG) \subseteq ECC(REG) \subseteq LIN$ i LIN és tancat sota les operacions d'unió, morfismes arbitraris i intersecció amb conjunts regulars.

No obstant, moltes d'aquestes famílies són iguals a la família de llenguatges lineals (aquesta és una de les raons per escollir les operacions d'unió, morfismes i intersecció amb llenguatges regulars com a operacions bàsiques). Com a conseqüència investigarem les interrelacions entre aquestes famílies i entre aquestes i LIN .

És fàcil de veure que $REG \subseteq \mathcal{L}(Au, Re, Id)$ $FIN \subseteq \mathcal{L}(Fa, Si, Id)$, i que totes les famílies anteriors contenen llenguatges no-regulars. A més a més, sabem que

$$L_1 = ab^+a \notin \mathcal{L}(Fa, Si, Id),$$

$$L_2 = \{a, b\} \notin \mathcal{L}(Au, Si, Id),$$

$$L_3 = a^*ba^* \cup \{a^nba^mc \mid n > m \geq 1\} \in \mathcal{L}(Au, Si, Co) - \mathcal{L}(Fa, Si, Id).$$

Aquests fets, així com les següents variants o formes més puixants de les mateixes seran molt útils per a crear el diagrama o figura de les relacions entre

totes les famílies.

$$L_1 \in \mathcal{L}(Au, Re, Id) - \mathcal{L}(Uf, Si, Ma),$$

$$L_2 \in \mathcal{L}(Fa, Si, Id) - \mathcal{L}(Au, Si, Ma),$$

$$L_3 \in \mathcal{L}(Uf, Si, Id).$$

Quan considerem una tripla (X_1, X_2, X_3) de $\{Au, Fa, Uf\} \times \{Si, Re\} \times \{Id, Co, Cd, Mn, Ma\}$ moltes vegades succeeix que un tret X_i és tan fort que és possible de reemplaçar un altre tret X_j per un altre més dèbil, X'_j , de manera que encara obtenim una família de llenguatges que no és més petita que la original. Els següents lemes presenten algunes situacions d'aquest tipus.

Lema 9.3.2 Per a tot $X \in \{Co, Cd, Mn, Ma\}$ tenim $\mathcal{L}(Uf, Si, X) \subseteq \mathcal{L}(Fa, Si, X)$.

Demostració. Per a un llenguatge $L \in \mathcal{L}(Uf, Si, X)$, considerem les gramàtiques $G_i = (V_i, \{w_i\}, P_i)$, $1 \leq i \leq n$, i un morfisme $h : (\bigcup_{i=1}^n V_i)^* \longrightarrow U^*$ tal que $L = h(\bigcup_{i=1}^n L_{ex}(G_i))$.

Per a cada i , $1 \leq i \leq n$, let $[V_i, i] = \{[a, i] \mid a \in V_i\}$ i definim les codificacions $g_i : V_i \longrightarrow [V_i, i]$ mitjançant $g_i(a) = [a, i]$, per a tot $a \in V_i$. Per a $(u, v) \in V_i^* \times V_i^*$ escriurem $g_i((u, v)) = (g_i(u), g_i(v))$ i extenem aquesta definició de manera natural a conjunts de contextos.

Considerem la gramàtica $G = (V, B, P)$ amb

$$\begin{aligned} V &= \bigcup_{i=1}^n [V_i, i], \\ B &= \{g_1(w_1), \dots, g_n(w_n)\}, \\ P &= \{(g_i(S), g_i(C)) \mid (S, C) \in P_i, 1 \leq i \leq n\}. \end{aligned}$$

Tenim $L_{ex}(G) = \bigcup_{i=1}^n g_i(L_{ex}(G_i))$. Això és una conseqüència del fet que g_i son codificacions una-a-una; això assegura que cada derivació a G que comença de $g_i(w_i)$ usarà només les produccions $(g_i(S), g_i(C))$ per a $(S, C) \in P_i$.

Considerem ara $h' : (\bigcup_{i=1}^n [V_i, i])^* \longrightarrow U^*$ definit per

$$h'([a, i]) = h(a), \text{ per a tot } a \in V_i \text{ i } 1 \leq i \leq n.$$

Clarament tenim $h'([a, i]) = h(g_i^{-1}([a, i]))$, per a tot $a \in V_i$ i tot $1 \leq i \leq n$.

La igualtat $h'(L_{ex}(G)) = L$ pot ser provada fàcilment. Pel fet que les funcions h, h' són del mateix tipus (g_i^{-1} és una codificació), se segueix que $L \in \mathcal{L}(Fa, Si, X)$. $\square$

Lema 9.3.3 $\mathcal{L}(Fa, Re, X) \subseteq \mathcal{L}(Au, Re, X)$, per a tot $X \in \{Id, Co, Cd, Mn, Ma\}$.

Demostració. Sigui $G = (V, B, P)$ una gramàtica amb $B = \{w_1, \dots, w_k\}$, sigui R un llenguatge regular, i sigui $h : V^* \longrightarrow U^*$ un morfisme. Construïm la gramàtica $G' = (V, \{\lambda\}, P')$ amb

$$P' = P \cup \{(\{\lambda\}, \{(\lambda, w_i)\}) \mid 1 \leq i \leq k\}.$$

Òbviament, $L_{ex}(G') = L_{ex}(G) \cup \{\lambda\}$.

Si $\lambda \in L_{ex}(G)$ (això significa, de fet, que algunes cadenes w_i són buides), aleshores deixem R invariable. Clarament, $h(L_{ex}(G) \cap R) = h(L_{ex}(G') \cap R)$.

Si $\lambda \notin L_{ex}(G)$, aleshores prenem $R' = R - \{\lambda\}$ i obtenim $h(L_{ex}(G) \cap R) = h(L_{ex}(G') \cap R')$ (la cadena buida λ és eliminada per la intersecció).

Conseqüentment, $h(L_{ex}(G) \cap R) \in \mathcal{L}(Au, Re, X)$, on X depèn de la forma de h . □

Lema 9.3.4 $\mathcal{L}(X, Si, Ma) \subseteq \mathcal{L}(X, Si, Cd)$, per a $X \in \{Au, Fa\}$.

Demostració. Prenem una gramàtica $G = (V, \{w_1, \dots, w_k\}, P)$ i un morfisme $h : V^* \longrightarrow U^*$. Considerem els nous símbols $[\text{ i }]$, així com una còpia a' per a cada $a \in V$. Denotem $V' = \{a' \mid a \in V\}$ i definim el morfisme $g : V^* \longrightarrow (U \cup V' \cup \{[,]\})^*$ mitjançant

$$g(a) = [a'h(a)], \text{ per a tot } a \in V.$$

Construïm la gramàtica $G_h = (V_h, B_h, P_h)$, amb

$$V_h = U \cup V' \cup \{[,]\},$$

$$B_h = \{g(w_1), \dots, g(w_k)\},$$

$$P_h = \{(g(S), g(C)) \mid (S, C) \in P\}.$$

Considerem també la codificació dèbil $h' : V_h \longrightarrow U \cup \{\lambda\}$, definida per

$$\begin{aligned} h'(a) &= a, \text{ per a tot } a \in U, \\ h'(a') &= h'(\lambda) = h'(\epsilon) = \lambda, \text{ per a tot } a \in V. \end{aligned}$$

Aleshores es pot provar que $h(L_{ex}(G)) = h'(L_{ex}(G_h))$ (l'efecte de h és “amagat” a G_h mitjançant el morfisme g ; la codificació h' és usada per a esborrar els símbols auxiliars). $\square$

Lema 9.3.5 $\mathcal{L}(X, Y, Mn) \subseteq \mathcal{L}(X, Y, Co)$, per a tot $X \in \{Au, Fa\}, Y \in \{Si, Re\}$.

Demostració. Prenem una gramàtica $G = (V, \{w_1, \dots, w_k\}, P)$ amb $k \geq 1$ ($k = 1$ correspon al cas $X = Au$), un morfisme no-esborrador $h : V^* \longrightarrow U^*$ i un llenguatge regular $R \subseteq V^*$. Considerem novament els símbols còpia a' per a tot $a \in V$, el vocabulari

$$V_h = U \cup V' \times U \cup U \times V' \cup V' \times U \times V',$$

i el morfisme $g : V^* \longrightarrow V_h^*$ tal que per a cada $a \in V$

$$g(a) = \begin{cases} (a', b, a'), & \text{si } h(a) = b \in U, \\ (a', b_1)b_2 \dots b_{n-1}(b_n, a'), & \text{si } h(a) = b_1 \dots b_n, \text{ amb } n \geq 2 \\ & \text{i } b_i \in U, \text{ per a tot } 1 \leq i \leq n. \end{cases}$$

Construïm ara la gramàtica $G_h = (V_h, A_h, P_h)$, amb

$$\begin{aligned} B_h &= \{g(w_1), \dots, g(w_k)\}, \\ P_h &= \{(g(S), g(C)) \mid (S, C) \in P\}. \end{aligned}$$

Sigui $R_h = g(R)$ i considerem la codificació $h' : V_h \longrightarrow U$ definida per

$$h'(a) = a, \text{ per a tot } a \in U,$$

$$h'((a', b)) = h'((b, a')) = h'((a', b, a')) = b, \text{ per a tot } a \in V \text{ i } b \in U.$$

La igualtat

$$h(L_{ex}(G) \cap R) = h'(L_{ex}(G_h) \cap R_h),$$

es compleix (com en la demostració del Lema 9.3.4 de la pàgina 347, l'efecte del morfisme h és introduït per g a G_h , el bloc de lletres que corresponen a un símbol $a \in V$ és marcat per les primeres ocurrencies d' a en els símbols (a', b) , (b, a') , (a', b, a') ; aquests parells i triples de símbols es corresponen amb símbols de U mitjançant la codificació h').

Quan $R = V^*$, tenim el cas $Y = Si$. □

A continuació demostrarem que dotze d'aquestes famílies de llenguatges considerades anteriorment són iguals a LIN . Clarament, és suficient amb provar que LIN està inclòs en aquestes famílies, ja que sabem que totes les famílies que considerem estan incloses a LIN (ja que $ECC_1(REG) \subseteq ECC(REG) \subseteq LIN$ i LIN és tancat sota les operacions d'unió, morfismes arbitraris, i intersecció amb conjunts regulars).

La noció de *separació regular* és crucial en les demostracions. Es defineix de la manera següent.

Per a una gramàtica lliure del context $G = (V_N, V_T, S, P)$ i per a cada $X \in V_N$, considerem el llenguatge $L(G_X)$, on $G_X = (V_N, V_T, X, P)$ (i, per

tant, $L(G_X)$ és el conjunt de cadenes generades per G començant pel símbol no-terminal X).

Definició 9.3.1 Una gramàtica lliure del context $G = (V_N, V_T, S, P)$ s'anomena *regularment separable* si per a cada símbol auxiliar o no-terminal $X \in V_N$ existeix un llenguatge regular $R_X \subseteq V_T^*$ tal que:

1. $R_X \cap R_Y = \emptyset$, per a tot $X, Y \in V_N$ tal que $X \neq Y$, i
2. $L(G_X) \subseteq R_X$, per a tot $X \in V_N$.

Definició 9.3.2 Un llenguatge lineal L és *separable* si existeixen un llenguatge finit F i una gramàtica lineal $G = (N, T, S, P)$ que és regularment separable tal que

1. $L = F \cup L(G)$,
2. $F \cap R_X = \emptyset$, per a tot $X \in N$.

Definició 9.3.3 Una gramàtica lineal $G = (N, T, S, P)$ es diu que té la *3nt-propietat* (de “three disjoint nonterminal sets”) si només conté regles lineals per la dreta i per l'esquerra, no conté cap regla de cadena i cap λ -regla i $N = N_e \cup N_d \cup N_f$, on N_e, N_d, N_f són conjunts disjunts dos a dos, tal que

$$N_e = \{B \mid B \rightarrow uX \in P, u \in T^+, X \in N\},$$

$$N_d = \{C \mid C \rightarrow Xv \in P, v \in T^+, X \in N\},$$

$$N_f = \{D \mid D \rightarrow w \in P, w \in T^+\}.$$

Els símbols no-terminals de N_e s'anomenen *empenyedors esquerrans* (“left-pushers”), els de N_d s'anomenen *empenyedors dretans* (“right-pushers”), i els símbols no-terminals de N_f s'anomenen *finals*.

Lema 9.3.6 Per a tot llenguatge lineal L existeix un llenguatge finit F i una gramàtica lineal G amb la 3nt-propietat, tal que $L = F \cup L(G)$.

Demostració. Sigui $L \subseteq V^*$ un llenguatge lineal. Suposem $V = \{a_1, \dots, a_n\}$. Tenim

$$L = \{x \in L \mid |x| \leq 1\} \cup \bigcup_{i=1}^n \{a_i\}(\partial_{a_i}^l(L) - \{\lambda\}).$$

Tot llenguatge $L_i = \partial_{a_i}^e(L) - \{\lambda\}$ és lineal (i λ -lliure). Sigui $G_i = (N_i, V, S_i, P_i)$ una gramàtica lineal per a L_i , $1 \leq i \leq n$. Sense perdre generalitat podem suposar que G_i no conté cap regla de cadena i cap λ -regla. A més a més, reemplaçant cada regla $p : X \rightarrow uYv$, $u, v \in V^+$, per $X \rightarrow uY_p, Y_p \rightarrow Yv$, per a Y_p un nou no-terminal, associat a aquesta regla p , obtenim una gramàtica equivalent. Per tant, podem suposar que totes les regles no-terminals de P_i són o bé lineals per l'esquerra o bé lineals per la dreta. Suposem que tots els conjunts N_i , $1 \leq i \leq n$ són disjunts dos a dos i construïm la gramàtica $G_0 = (N, V, S, P)$, amb

$$N = \{S\} \cup \bigcup_{i=1}^n N_i, \quad S \text{ és un símbol nou,}$$

$$P = \{S \rightarrow a_i S_i \mid 1 \leq i \leq n\} \cup \bigcup_{i=1}^n P_i.$$

És fàcil de veure que $L = F \cup L(G_0)$, per a $F = \{x \in L \mid |x| \leq 1\}$ i que S només apareix en regles “empenyedores per l'esquerra”.

Reemplacem ara cada regla de la forma $X \rightarrow uY$ de P per

$$X^e \rightarrow uY^e, X^e \rightarrow uY^d, X^e \rightarrow uY^f,$$

cada regla de la forma $X \rightarrow Yv$ per

$$X^d \rightarrow Y^e v, X^d \rightarrow Y^d v, X^d \rightarrow Y^f v,$$

i cada regla $X \rightarrow w$ per

$$X^f \rightarrow w.$$

Denotem per P' el conjunt de totes les regles obtingudes d'aquesta manera i sigui $G = (\{X^e, X^d, X^f \mid X \in N\}, T, S^e, P')$. És fàcil de veure que $L(G_0) = L(G)$ (les derivacions en les dues gramàtiques són idèntiques mòdul els superíndexs e, d, f). A més a més, G té la 3nt-propietat: considerem els conjunts (òbviament disjunts): $N_\alpha = \{X^\alpha \mid X \in N\}$, $\alpha \in \{e, d, f\}$. $\square$

Lema 9.3.7 Tot llenguatge lineal és una codificació de llenguatges lineals separables.

Demostració. Sigui $L \subseteq V^*$ un llenguatge lineal. D'acord amb el lema previ, existeixen un conjunt finit $F \subseteq V \cup \{\lambda\}$ i una gramàtica $G_0 = (N, V, S, P)$ amb la 3nt-propietat, tal que $L = F \cup L(G_0)$ i per a cada $x \in L(G_0)$ tenim $|x| \geq 2$. Siguin N_e, N_d, N_f els conjunts de símbols no-terminals de les tres categories considerades anteriorment.

Construïm la gramàtica $G = (N', V', S_0, P')$, amb

$$N' = \{B_i \mid B \in N_e \cup N_d, 0 \leq i \leq 2\} \cup N_f,$$

$$V' = V \times N,$$

i P' és obtingut de la manera següent. Per a cada $X \in N'$, definim el morfisme $h_X : V^* \rightarrow V'^*$ mitjançant $h_X(a) = (a, X), a \in V$. Aleshores

- per a cada producció empenyedora per l'esquerra $B \rightarrow uX$ a P i per a cada $i, 0 \leq i \leq 2$, introduïm a P' les regles

$$B_i \rightarrow h_{B_i}(u)X_i, \text{ si } X \in N_e,$$

$$B_i \rightarrow h_{B_i}(u)X_{i+1(mod\ 3)}, \text{ si } X \in N_d,$$

$$B_i \rightarrow h_{B_i}(u)X, \text{ si } X \in N_f;$$

- per a cada producció empenyedora per la dreta $C \rightarrow Xv$ a P i per a cada $i, 0 \leq i \leq 2$, introduïm a P' les regles

$$C_i \rightarrow X_{i+1(mod\ 3)}h_{C_i}(v), \text{ si } X \in N_e,$$

$$C_i \rightarrow X_ih_{C_i}(v), \text{ si } X \in N_d,$$

$$C_i \rightarrow Xh_{C_i}(v), \text{ si } X \in N_f;$$

- per a cada producció final $D \rightarrow w$ a P introduïm $D \rightarrow h_D(w)$.

Considerem també la codificació $h' : V \cup V' \rightarrow V$ definida per $h'((a, X)) = a$, per a tot $a \in V$ i $X \in N$, i $h'(a) = a$, per a $a \in V$. A partir de la construcció

de G , tenim $h'(L(G)) = L(G_0)$ (les derivacions de les dues gramàtiques coincideixen, mòdul els subíndexs 0, 1, 2 de no-terminals i els emparellaments de terminals i no-terminals; quan l'emparellament és esborrat per la codificació h' , obtenim la mateixa cadena). Com que $h'(F) = F$, tenim $L = h'(F \cup L(G))$; per tant, ens queda per demostrar que G és una gramàtica separable i que $F \cap R_X = \emptyset$, per a tot $X \in N'$. Aquesta última propietat és òbvia, ja que $F \subseteq V \cup \{\lambda\}$ i $R_X \subseteq V'^+$ per a tot $X \in N'$.

Per tal de provar que G és separable, examinem la forma de les cadenes de $L(G_X)$, per a cada symbol $X \in N'$. Tota cadena de $L(G)$ és de la forma $(a_1, Z_1) (a_2, Z_2) \dots (a_m, Z_m)$. Si aquesta cadena és generada començant a partir d' X , aleshores $Z_1 \in \{X_0, X_1, X_2\}$ si X és un empenyedador per l'esquerra; i $Z_m \in \{X_0, X_1, X_2\}$ si X és un empenyedador per la dreta. Conseqüentment, el parell (Z_1, Z_m) identificarà, conjuntament amb el tipus d' X , les cadenes produïdes per X a G .

Taula 9.3.1 Aquesta taula conté totes les classes possibles de parells (Z_1, Z_2) . Denotem Z_1 per F_2 “la 2a. component del primer parell”, i Z_m per L_2 “la 2a. component de l'últim parell”.

	Tipus de X	F_2	L_2
1	empenyedor-esquerra	$X_i, 0 \leq i \leq 2$	$\forall D \in N_f$
2	empenyedor-esquerra	$X_i, 0 \leq i \leq 2$	$\forall B_j, B \in N_e, 0 \leq j \leq 2$
3	empenyedor-esquerra	$X_i, 0 \leq i \leq 2$	$\forall C_{i+1(mod\ 3)}, C \in N_d$
4	empenyedor-dreta	$\forall D \in N_f$	$X_i, 0 \leq i \leq 2$
5	empenyedor-dreta	$\forall C_j, C \in N_d, 0 \leq j \leq 2$	$X_i, 0 \leq i \leq 2$
6	empenyedor-dreta	$\forall B_{i+1(mod\ 3)}, B \in N_e$	$X_i, 0 \leq i \leq 2$
7	final	D	C

Els conjunts R_X, R_{X_i} poden ser ara definits de la següent manera:

- si $X \in N_e$, i $0 \leq i \leq 2$, aleshores

$$R_{X_i} = \{(a, X_i) \mid a \in V\} V'^* \{(a, L_2) \mid a \in V, L_2 \text{ com en les files } 1, 2, 3 \text{ de la taula}\};$$

- si $X \in N_d$, i $0 \leq i \leq 2$, aleshores

$$R_{X_i} = \{(a, F_2) \mid a \in V, F_2 \text{ com en les files } 4, 5, 6 \text{ de la taula}\} V'^* \{(a, X_i) \mid a \in V\};$$

- si $X \in N_f$, aleshores

$$R_X = \{(a, X) \mid a \in V\} V'^* \{(a, X) \mid a \in V\} \cup \{(a, X) \mid a \in V\}.$$

Observació 9.3.1 Els conjunts R_{X_i} i R_X són disjunts dos a dos.

Demostració. La gramàtica G_0 té la 3nt-propietat així com també la gramàtica G i, per tant, un no-terminal és o bé un empenyedador-esquerra, o bé un empenyedador-dreta, o bé un final. Els parells (F_2, L_2) en la classificació prèvia (i, per tant, en la definició dels conjunts R_{X_i}, R_X) són diferents d'una fila de la taula respecte a una altra fila. Això és clar per a tots els parells de files, degut a la 3nt-propietat, excepte pel cas de les files 3 i 6, on F_2 és un empenyedador-esquerra i L_2 és un empenyedador-dreta. No obstant, en la fila 3 tenim $F_2 = X_i$ i $L_2 = Y_{i+1(mod\ 3)}$ per a tot $0 \leq i \leq 2$ i, per tant, L_2 té el subíndex més gran que

i (mòdul 3), mentre que en la fila 6 tenim $F_2 = Y_{i+1(mod\ 3)}, L_2 = X_i, 0 \leq i \leq 2$, és a dir, F_2 té un subíndex més gran (mòdul 3). Conseqüentment, també en aquest cas tenim conjunts disjunts de cadenes. $\square$

Observació 9.3.2 Per a cada $X \in N', L(G_X) \subseteq R_X$.

Demostració. Analitzem les possibles formes de les derivacions a G , que comencin per un símbol $X \in N'$.

Cas 1. X és un empenyedador-esquerra, $X = B_i$, per a $0 \leq i \leq 2$ i $B \in N_e$.

Si una derivació no usa regles drete-empenyedores, aleshores és d'una de les formes

$$\begin{aligned} B_i &\Longrightarrow h_{B_i}(u_1)B_i^{(1)} \Longrightarrow h_{B_i}(u_1)h_{B_i^{(1)}}(u_2)B_i^{(2)} \\ &\Longrightarrow \emptyset ts \Longrightarrow h_{B_i}(u_1) \dots h_{B_i^{(n)}}(u_{n+1})D \\ &\Longrightarrow h_{B_i}(u_1) \dots h_{B_i^{(n)}}(u_{n+1})h_D(w) = (a_1, B_i)z(a_s, D). \end{aligned}$$

Conseqüentment, $F_2 = B_i$ i $L_2 = D$, que correspon a la fila 1 en la Taula 9.3.1.

Si una derivació usa una regla drete-empenyedora, aleshores considerem el primer pas on un no-terminal drete-empenyedor és introduït:

$$\begin{aligned} B_i &\Longrightarrow h_{B_i}(u_1)B_i^{(1)} \Longrightarrow \dots \Longrightarrow h_{B_i}(u_1) \dots h_{B_i^{(n)}}(u_{n+1})C_{i+1(mod\ 3)} \\ &\Longrightarrow h_{B_i}(u_1) \dots h_{B_i^{(n)}}(u_{n+1})C_{i+1(mod\ 3)}^{(1)}h_{C_{i+1(mod\ 3)}}(v_1) \\ &\Longrightarrow \dots \Longrightarrow h_{B_i}(u_1) \dots h_D(w) \dots h_{C_{i+1(mod\ 3)}}(v_1) \\ &= (a_1, B_i)z(a_s, C_{i+1(mod\ 3)}). \end{aligned}$$

Tenim $F_2 = B_i$ i $L_2 = C_{i+1(mod\ 3)}$, que correspon a la fila 3 de la Taula 9.3.1.

Cas 2. X és un empenyedors-dreta, $X = C_i$, per a $0 \leq i \leq 2$ i $C \in N_d$.

Quan una derivació no usa regles empenyedors per l'esquerra, aleshores obtenim una cadena que correspon a la fila 4 de la taula. Quan apareixen també empenyedors per l'esquerra, obtenim una cadena que correspon a la fila 6 de la Taula 9.3.1 (l'argument és similar a l'usat en el Cas 1).

Cas 3. X és un no-terminal final, $X = D$ per a algun $D \in N_f$.

Tenim només derivacions d'un sol pas que ens donen cadenes corresponents a la fila 7 de la Taula 9.3.1.

Per tant, G és separable, la qual cosa completa la demostració. $\square$

Cal notar que cap cadena terminal generada per l'anterior gramàtica G es correspon amb les files 2 o 5 de la classificació de la Taula 9.3.1 (només les cadenes no-terminals poden tenir F_2 i L_2 empenyedors-esquerrans o empenyedors-dretans).

Lema 9.3.8 Si L és un llenguatge lineal separable, aleshores $L \in \mathcal{L}(\text{Fa}, \text{Re}, \text{Id})$.

Demostració. Sigui F un conjunt finit i $G_0 = (N, V, S, P_0)$ una gramàtica lineal regularment separable tal que $L = F \cup L(G_0)$ i $F, R_X, X \in N$, com en les Definicions 9.3.1 de la pàgina 350 i 9.3.2 de la pàgina 350.

Construïm la gramàtica contextual $G = (V, B, P)$, amb

$$B = \{w \in V^* \mid X \rightarrow w \in P\} \cup F,$$

i P que conté totes les produccions

$$(R_Y, \{(u, v)\}), \text{ per a } X \rightarrow uYv \in P, \text{ amb } u, v \in V^*, \text{ tal que } uv \neq \lambda,$$

(per a cada producció, o bé u o bé v és buit).

Considerem també el llenguatge regular $R = R_S \cup F$. Aleshores es verifica que $L = L_{ex}(G) \cap R$. $\square$

Combinant totes les relacions que ja hem demostrat anteriorment, tenim el següent resultat.

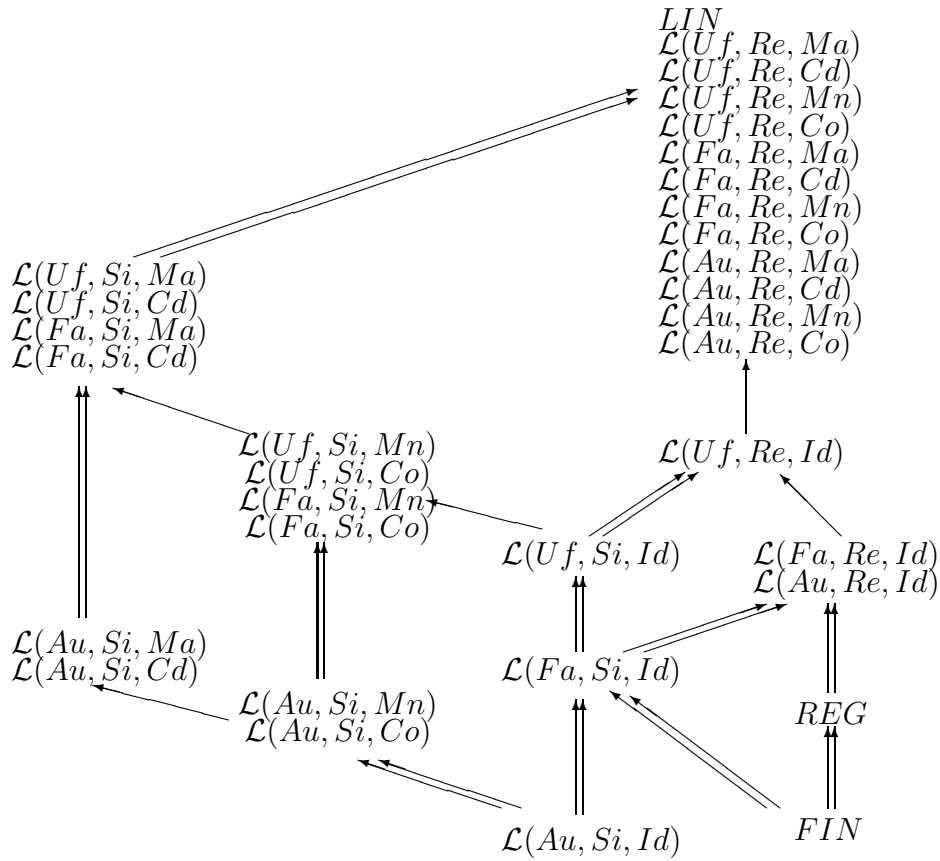
Teorema 9.3.1 Les relacions que el diagrama de la Figura 9.3.2 de la pàgina 359 mostra són certes. Una fletxa doble d'una família F_1 a una família F_2 indica que F_1 està estrictament inclosa a F_2 , i una fletxa simple indica que $F_1 \subseteq F_2$ i que no coneixem si la inclusió és pròpia o no. Dues famílies no relacionades en aquest diagrama no són necessàriament incomparables, amb l'excepció dels parells que es relacionen amb una de les famílies FIN, REG : aquestes dues famílies són incomparables amb totes les famílies amb les quals no estiguin relacionades en el diagrama. Les famílies que estan escrites en un mateix node del diagrama són iguals.

D'acord amb (Baker i Book, 1974) (veure també (Turakainen, 1978)), tot llenguatge recursivament enumerable L pot ser representat com $h(L_1 \cap L_2)$,

on L_1, L_2 són llenguatges lineals i h és una codificació dèbil. A la vista dels resultats previs, els llenguatges lineals L_1, L_2 poden ser escrits com $L_i = h_i(L'_i \cap R_i)$, $i = 1, 2$, amb R_i llenguatges regulars, h_i codificacions, i L'_i in $\mathcal{L}(Au, Si, Id)$. Per tant, tenim

$$L = h(h_1(L'_1 \cap R_1) \cap h_2(L'_2 \cap R_2)).$$

Figura 9.3.2 Aquesta representació pot ser millorada de la manera següent.



Lema 9.3.9 Si L_1, L_2 són llenguatges lineals, aleshores podem representar-los de la forma $L_i = h_i(L'_i \cap R_i), i = 1, 2$, on $L'_i \in \mathcal{L}(Fa, Si, Id)$, cada R_i és un llenguatge regular, i cada h_i és una codificació, tal que

- (1) $h_1 = h_2$;
- (2) $h_1(L'_1 \cap R_1) \cap h_2(L'_2 \cap R_2) = h_1(L'_1 \cap R_1 \cap L'_2 \cap R_2)$.

Demostració. Comencem a a partir de dues gramàtiques lineals $G_i = (N_i, V, S_i, P_i), i = 1, 2$, amb la 3nt-propietat, i a partir de dos conjunts finits F_i tal que $L_i = F_i \cup L(G_i), i = 1, 2$. Aleshores, modifiquem les construccions de les demostracions del Lema 9.3.7 de la pàgina 352 i del Lema 9.3.8 de la pàgina 357 de la següent manera.

Considerem la demostració del Lema 9.3.7.

En la construcció de la gramàtica separable per a $L(G_1)$, denotada per $G'_1 = (N'_1, V', S_1, P'_1)$, usem ara

$$V' = V \times N'_1 \times N'_2$$

i reemplacem els morfismes $h_X : V^* \longrightarrow (V \times N'_1)^*$, per $X \in N'_1$, per les substitucions finites $s_X : V^* \longrightarrow 2^{V'^*}$, definides per

$$s_X(a) = \{(a, X, Y) \mid Y \in N'_2\}, \text{ per a tot } a \in V.$$

Introduïm ara un tercer component en tots els símbols terminals, on qualsevol símbol de N'_2 pot aparèixer.

En la construcció de la gramàtica separable per a $L(G_2)$, denotada per $G'_2 = (N'_2, V', S_2, P'_2)$, reemplacem els morfismes $h_X : V^* \longrightarrow (V \times N'_2)^*$, per a $X \in N'_2$, per la substitució finita $s'_X : V^* \longrightarrow 2^{V'^*}$, definida per

$$s'_X(a) = \{(a, Y, X) \mid Y \in N'_1\}, \text{ per a tot } a \in V.$$

Posem ara X en el tercer component, no en el segon com en $\varphi_X(a)$, i col·loquem en el segon component qualsevol símbol de N'_1 .)

Usem les substitucions s_X i s'_X en la construcció de les produccions de P'_1, P'_2 com en la demostració del Lema 9.3.7 i després definim la codificació $g : V \cup V' \longrightarrow V$ mitjançant $g((a, X, Y)) = a$, per a $a \in V, X \in N'_1, Y \in N'_2$, i $g(a) = a$, per a $a \in V$.

Les igualtats $g(L(G'_i)) = L(G_i), i = 1, 2$, se segueixen de la construcció. A més a més, G'_1 i G'_2 són separables, ja que no hem modificat els símbols no-terminals en la construcció de la demostració del Lema 9.3.7, i el segon component de símbols generats a G'_1 i el tercer component de símbols generats a G'_2 poden ser usats per a classificar les cadenes com en la demostració del Lema 9.3.7. Observem també que la codificació g és la mateixa per ambdós llenguatges i que usa la definició $g(a) = a$, per a $a \in V$, només per a cadenes dels conjunts finits F_1, F_2 (mentre que totes les cadenes generades per les gramàtiques G'_1, G'_2 són, com a mínim, de longitud 2).

Ara, la demostració del Lema 9.3.8 de la pàgina 357 no canvia els símbols en el vocabulari; és a dir, treballem amb el mateix vocabulari $V \times N'_1 \times N'_2$

quan definim els llenguatges L'_1, L'_2 a $\mathcal{L}(Fa, Si, Id)$ i els llenguatges regulars R_1, R_2 tal que $L(G'_i) = L'_i \cap R_i$, per a $i = 1, 2$.

Per tal de completar la demostració hem de veure que

$$g(L'_1 \cap R_1) \cap g(L'_2 \cap R_2) = g(L'_1 \cap R_1 \cap L'_2 \cap R_2).$$

La inclusió $\supseteq$ és òbvia.

Inversament, prenem $x \in g(L'_1 \cap R_1) \cap g(L'_2 \cap R_2)$. Aleshores $x \in g(L'_1 \cap R_1)$ i $x \in g(L'_2 \cap R_2)$ i, per tant, $x = g(y_1)$ per a algun $y_1 \in L'_1 \cap R_1$, i $x = g(y_2)$ per a algun $y_2 \in L'_2 \cap R_2$. Recordem de la demostració del Lema 9.3.8 que $R_1 = R_{S_1} \cup F_1$, i $R_2 = R_{S_2} \cup F_2$, on R_{S_1}, R_{S_2} són els conjunts regulars que inclouen els llenguatges generats quan comencem per S_1, S_2 a G'_1, G'_2 , respectivament.

Si $|x| \leq 1$, aleshores $y_1 \in F_1$ i $y_2 \in F_2$ i, per tant, $y_1 = y_2$ i $x \in g(L'_1 \cap R_1 \cap L'_2 \cap R_2)$.

Si $|x| > 1$, aleshores $y_1 \in R_{S_1}$ i $y_2 \in R_{S_2}$. Suposem que $x = a_1 \dots a_k, a_i \in V, 1 \leq i \leq k$. D'acord amb la construcció de G'_1 , existeixen $X_1, \dots, X_k$ a N'_1 tal que totes les cadenes de la forma $(a_1, X_1, Y_1)(a_2, X_2, Y_2) \dots (a_k, X_k, Y_k)$, amb arbitrari $Y_1, \dots, Y_k$ a N'_2 , pertanyen a $L(G'_1)$; per tant, també pertanyen a $L'_1 \cap R_1$. D'acord amb la construcció de G'_2 , existeixen $Z_1, \dots, Z_k$ a N'_2 tal que totes les cadenes de la forma $(a_1, W_1, Z_1)(a_2, W_2, Z_2) \dots (a_k, W_k, Z_k)$, amb arbitrari $W_1, \dots, W_k$ a N'_1 , pertanyen a $L(G'_2)$; per tant, també pertanyen a $L'_2 \cap R_2$. Conseqüentment, la cadena $(a_1, X_1, Y_1)(a_2, X_2, Y_2) \dots (a_k, X_k, Y_k)$

pertany a la intersecció $(L'_1 \cap R_1) \cap (L'_2 \cap R_2)$, ja que la codificació d'aquesta cadena mitjançant g ens dona x , tenim $x \in g(L'_1 \cap R_1 \cap L'_2 \cap R_2)$. Per tant, la inclusió $\subseteq$ es compleix. $\square$

Teorema 9.3.2 Tot llenguatge recursivament enumerable L pot ser representat com a $L = h(L_1 \cap L_2 \cap R)$, on h és una codificació dèbil, $L_1, L_2 \in \mathcal{L}(\text{Fa}, \text{Si}, \text{Id})$ i R és un llenguatge regular.

Demostració. Com ja hem fet esment abans, usant la representació de (Baker i Book, 1974), (Turakainen, 1978), i els Lemes 9.3.7 de la pàgina 352 i 9.3.8 de la pàgina 357, podem representarn L de la forma $L = g(h_1(L_1 \cap R_1) \cap h_2(L_2 \cap R_2))$, amb $L_1, L_2 \in \mathcal{L}(\text{Fa}, \text{Si}, \text{Id})$ (fins i tot podem prendre $L_1, L_2 \in \mathcal{L}(\text{Au}, \text{Si}, \text{Id})$), R_1, R_2 conjunts regulars, g una codificació dèbil, i h_1, h_2 codificacions. D'acord amb el Lema 9.3.9, podem prendre $h_1 = h_2$ tal que $h_1(L_1 \cap R_1) \cap h_2(L_2 \cap R_2) = h_1(L_1 \cap R_1 \cap L_2 \cap R_2)$. La composició de g i h_1 és una codificació dèbil (la denotem per h) i $R_1 \cap R_2$ és un llenguatge regular (el denotem per R). Per tant, $L = h(L_1 \cap L_2 \cap R)$, la qual cosa completa la demostració. $\square$

Notes bibliogràfiques. El primer lloc on les gramàtiques contextuais (externes) amb selecció (regular) restringida van ser considerades és a (Istrail, 1978a). Resultats més recents relatius a aquestes gramàtiques apareixen a (Păun, 1978b; Păun, 1978c; Păun, 1982a); per exemple, les demostracions dels Lemes 9.1.4 de la pàgina 303 i 9.1.5 de la pàgina 306 poden ser trobats a (Păun, 1982a). La inclusió $ECC(REG) \subseteq LIN$ és demostrada a (Păun et al.,

1994a) (a (Păun, 1985a) només és provat que $ECC(REG) \subseteq CF$). La forma normal encaixant, Lemes 9.1.7 – 9.1.15, i Teorema 9.1.2 de la pàgina 321 són de (Ehrenfeucht et al., 1994) (on també és definida una forma normal “estàtica”: gramàtiques amb selectors disjunts).

Implícitament, la forma normal encaixant ja va ser introduïda a (Novotný, 1976b) sota el nom d’una *gramàtica contextual regular*: $G = (V, B, C, \varphi)$ s’anomena regular si, per a cada $x, y \in V^*$, si $\varphi(x) = \varphi(y)$ i $(u, v) \in \varphi(x)$, aleshores $\varphi(uxv) = \varphi(uyv)$. La condició $\varphi(x) = \varphi(y)$ pot ser replantejada com a “ x, y són elements del mateix selector”; per tant, $\varphi(uxv) = \varphi(uyv)$ significa que $uxv \in S'$ per a tot x pertanyent a un selector S i, per tant, les produccions $(S, (u, v)), (S', (u', v'))$, per a algun $(u', v') \in \varphi(uxv)$, són encaixants. A (Novotný, 1976b) és demostrat que aquestes gramàtiques només generen (en el mode extern) llenguatges lineals.

Les gramàtiques contextuais externes amb selecció acotada (Definició 9.1.5 de la pàgina 325 i Teorema 9.1.3 de la pàgina 326) són considerades a (Păun, 1978b). El Teorema 9.2.1 de la pàgina 328 és de (Ilie, 1996b) i el Teorema 9.2.2 de la pàgina 334 és demostrat a (Ehrenfeucht et al., 1996b), on la demostració és donada amb tot tipus de detalls. La Secció 9.3 està basada en (Ehrenfeucht et al., 1995c).

Una visió general pot ser trobada a (Ehrenfeucht et al., 1996a) o a (Păun, 1996a).

Capítol 10

Gramàtiques contextuais amb restriccions en els components

A continuació veurem algunes idees de com modificar la definició de les gramàtiques contextuais o dels llenguatges generats per elles. Els següents capítols i seccions estan dedicats a l'estudi d'aquestes variants. Algunes d'elles es corresponen amb idees clàssiques ja explorades en la teoria de llenguatges formals, especialment pel cas de les gramàtiques lliures del context (Dassow et al., 1996b); però moltes d'altres són específiques de les gramàtiques contextuais.

En primer lloc, considerarem variants de les gramàtiques contextuais, on imposarem certes restriccions en els seus components. Veurem les gramàtiques contextuais amb ús minimal/maximal dels selectors, les gramàtiques contextuais amb contextos esborrats, les gramàtiques contextuais amb contextos unilaterals i, finalment, les gramàtiques contextuais deterministes.

10.1 Gramàtiques contextuais amb ús minimal i maximal dels selectors

Examinem sistemàticament la possibilitat de generar els llenguatges L_1, L_2, L_3 , expressats a continuació, mitjançant gramàtiques dels tipus considerats fins ara en els capítols precedents:

$$L_1 = \{a^n b^n c^n \mid n \geq 1\},$$

$$L_2 = \{xcx \mid x \in \{a, b\}^*\},$$

$$L_3 = \{a^n b^m c^n d^m \mid n, m \geq 1\}.$$

Pel Lema 5.1.5 de la pàgina 189 sabem que $L_1 \notin ICC$. Al començament de la demostració del Lema 9.1.4 de la pàgina 303 vam veure que $L'_3 = \{a^n b^m a^n b^m \mid n, m \geq 1\}$ pertany a $ICC(REG) - ICC(FIN)$. De la mateixa manera (usant el Lema 9.1.1 de la pàgina 303 per a la part negativa de l'assertió), obtenim $L_3 \in ICC(REG) - ICC(FIN)$.

Clarament (segons la Definició 5.1.8 de la pàgina 192), si, per a un llenguatge $L \subseteq V^*$, denotem

$$Min_0(L) = \{x \in L \mid PSub(x) \cap L = \emptyset\},$$

on $PSub(x)$ denota el conjunt de subcadenaes pròpies d'una cadena x (veure les definicions de la Secció 3.1), aleshores $Min_0(L_1) = L_1, Min_0(L_3) = L_3$ i, per tant, aquests llenguatges no pertanyen a ECC (Lema 5.1.7 de la pàgina 192). Per a L_2 tenim $Min_0(L_2) = \{c\}$, però $Min_1(L_2)$ és infinit: totes les cadenes

de la forma $ba^n cba^n$ pertanyen a L_2 i la seva única subcadena que pertany a $Min_0(L_2)$ és c . Per tant, totes aquestes cadenes pertanyen a $Min_1(L_2)$. Conseqüentment, tampoc $L_2 \notin ECC$.

També el llenguatge L_2 va més enllà de la competència del les gramàtiques contextuais internes: suposem que $L_2 = L_{in}(G)$, per a algun $G = (\{a, b, c\}, B, C, \varphi)$. Totes les cadenes de la forma $a^n c a^n$ pertanyen a L_2 . Per tal de generar-les necessitem un context (a^i, a^i) , $i \geq 0$, tal que $(a^i, a^i) \in \varphi(a^j c a^k)$, per a algun $j, k \geq 0$. Considerem la cadena $w = a^{j+k} b a^{j+k} c a^{j+k} b a^{j+k}$. Aquesta cadena pertany a L_2 i, per tant, podem derivar $w \Rightarrow_{in} a^{j+k} b a^{j+k+i} c a^{j+k+i} b a^{j+k}$. La cadena obtinguda no pertany a L_2 ; la qual cosa és una contradicció.

Tots aquests tres llenguatges L_1, L_2, L_3 pertanyen a la família $TC(REG)$. Per a L_3 aquest fet és una conseqüència de la relació $L_3 \in ICC(REG)$. Per a les gramàtiques

$$\begin{aligned} G_1 &= (\{a, b, c\}, \{abc\}, \{(a, bc)\}, \varphi_1), \\ \varphi_1(a^n, b^m, c^p) &= \{(a, bc)\}, \quad n, m, p \geq 1, \\ G_2 &= (\{a, b, c\}, \{c\}, \{(a, a), (b, b)\}, \varphi_2), \\ \varphi_2(\lambda, xc, y) &= \{(a, a), (b, b)\}, \quad x, y \in \{a, b\}^*, \end{aligned}$$

òbviamment tenim

$$L(G_i) = L_i, \quad i = 1, 2.$$

Clarament, aquestes dues gramàtiques tenen selecció regular, en el sentit considerat al final de la secció 4.1.

La raó per la qual les gramàtiques contextuais de formes simples (és a dir, gramàtiques internes i externes) no poden generar aquests llenguatges (amb l'excepció d' L_3 , el qual pertany a la família $ICC(REG)$) és el fet que les produccions necessàries per a generar cadenes de longitud arbitrària d'aquests llenguatges poden ser també usades en emplaçaments “equivocats”, amb la qual cosa es generen cadenes que no pertanyen a aquests llenguatges.

Una manera d'evitar aquest inconvenient és la d'imposar restriccions en els selectors usats. Aquesta actuació ens proporciona un efecte espectacular: tots els llenguatges mencionats anteriorment poden ser generats quan són usats selectors *maximals*.

10.1.1 Definicions bàsiques i el poder de l'ús minimal dels selectors

Definició 10.1.1 Considerem una gramàtica contextual

$$G = (V, B, (S_1, C_1), \dots, (S_n, C_n)).$$

Definim les següents variants de la relació de derivació interna:

$$\begin{aligned} x \Rightarrow_{ml} y \quad \text{sï} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \\ & \text{per a } x_2 \in S_i, (u, v) \in C_i, 1 \leq i \leq n, \\ & \text{i no existeixen } x'_1, x'_2, x'_3 \in V^* \text{ tal que} \\ & x = x'_1 x'_2 x'_3, x'_2 \in S_i, \text{ i} \\ & |x'_1| \geq |x_1|, |x'_3| \geq |x_3|, |x'_2| < |x_2|; \end{aligned}$$

$$\begin{aligned}
x \Longrightarrow_{mg} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \\
& \text{per a } x_2 \in S_i, (u, v) \in C_i, 1 \leq i \leq n, \\
& \text{i no existeixen } x'_1, x'_2, x'_3 \in V^* \text{ tal que} \\
& x = x'_1 x'_2 x'_3, x'_2 \in S_j, \text{ per a algun } 1 \leq j \leq n, \text{ i} \\
& |x'_1| \geq |x_1|, |x'_3| \geq |x_3|, |x'_2| < |x_2|,
\end{aligned}$$

$$\begin{aligned}
x \Longrightarrow_{Ml} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \\
& \text{per a } x_2 \in S_i, (u, v) \in C_i, 1 \leq i \leq n, \\
& \text{i no existeixen } x'_1, x'_2, x'_3 \in V^* \text{ tal que} \\
& x = x'_1 x'_2 x'_3, x'_2 \in S_i, \text{ i} \\
& |x'_1| \leq |x_1|, |x'_3| \leq |x_3|, |x'_2| > |x_2|;
\end{aligned}$$

$$\begin{aligned}
x \Longrightarrow_{Mg} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \\
& \text{per a } x_2 \in S_i, (u, v) \in C_i, 1 \leq i \leq n, \\
& \text{i no existeixen } x'_1, x'_2, x'_3 \in V^* \text{ tal que} \\
& x = x'_1 x'_2 x'_3, x'_2 \in S_j, \text{ per a } 1 \leq j \leq n, \text{ i} \\
& |x'_1| \leq |x_1|, |x'_3| \leq |x_3|, |x'_2| > |x_2|.
\end{aligned}$$

Direm que:

- $\Longrightarrow_{ml}$ és una derivació en el *mode minimal local*, és a dir, no podem usar com a selector una cadena que tingui, com a mínim, una subcadena pròpia com a selector de la mateixa producció;

- $\Rightarrow_{mg}$ és una derivació en el *mode minimal global*, és a dir, el selector és minimal respecte a totes les produccions de la gramàtica; mentre que
- $\Rightarrow_{Ml}$ és una derivació en el *mode maximal local*, és a dir, cap supercadena estricta del selector usat pot ser usada per la mateixa producció; i, finalment
- $\Rightarrow_{Mg}$ és una derivació en el *mode maximal global*, és a dir, el selector és el més gran de tots en aquesta posició, respecte a totes les produccions.

Cal notar que, en tots els casos anteriors, ens hem referit al selector seleccionat relacionant-lo amb l'emplaçament específic que ocupa en la cadena en qüestió; res ens priva, però, que sub- o super-cadenes apareguin en altres posicions de la cadena.

Per a $\alpha \in \{ml, mg, Ml, Mg\}$ definim

$$L_\alpha(G) = \{z \in V^* \mid w \Rightarrow_\alpha^* z, \text{ per a algun } w \in B\}.$$

Les famílies de llenguatges de la forma $L_\alpha(G)$, per a G una gramàtica contextual interna amb F selecció, són denotades per $ICC_\alpha(F)$, $\alpha \in \{ml, mg, Ml, Mg\}$. Nosaltres considerarem $F \in \{FIN, REG\}$. Òbviament, les produccions que tenen contextos buits, (λ, λ) , són importants, ja que poden bloquejar l'ús d'altres produccions. Per tant, a diferència d'altres vegades, en aquesta secció no adoptarem cap suposició relativa a la buidor dels contextos.

Observació 10.1.1 Podem considerar també les gramàtiques amb ús minimal o maximal dels selectors i sense selecció. En aquest cas els selectors són iguals a V^* i, per tant, la restricció minimal imposa l'ús de λ com a selector; mentre que la restricció maximal imposa una derivació externa. En el primer cas, l'ús d'un context (u, v) significa insertar la cadena uv . Per tant, obtenim casos particulars de gramàtiques d'inserció, de pes 0 (veure el capítol 14 “A dual model: inserton grammars” a (Păun, 1996a)), o bé de gramàtiques contextuais unilaterals (internes, sense selecció; veure la Secció 10.3).

Les següents relacions són òbvies.

Lema 10.1.1 $ICC_\alpha(FIN) \subseteq ICC_\alpha(REG)$, per a tot $\alpha \in \{ml, mg, Ml, Mg\}$.

Mitjançant senzilles construccions (la verificació de la minimalitat o bé de la maximalitat del repertori de selectors, els quals són membres de conjunts regulars, pot ser realitzada a nivell de les gramàtiques sensibles al context), obtenim

Lema 10.1.2 $ICC_\alpha(REG) \subseteq CS$, per a tot $\alpha \in \{ml, mg, Ml, Mg\}$.

A continuació investigarem, breument, el poder de les gramàtiques contextuais amb els selectors usats en mode minimal. Tenim

Lema 10.1.3 $ICC_{mg}(F) \subseteq ICC_{ml}(F) \subseteq ICC(F)$, $F \in \{FIN, REG\}$.

Demostració. Considerem una gramàtica $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$.

Si algun S_i conté dues cadenes x, x' tal que x és una subcadena pròpia d' x' , aleshores x' mai és usada en una derivació localment minimal i, per tant, pot ser eliminada sense modificar el llenguatge generat. De manera similar, quan $x \in S_i$ i $x' \in S_j$, per a qualsevol i, j , en el cas de derivacions globalment minimal. Formalment, podem construir les gramàtiques

$$G_g = (V, B, (S'_1, C_1), \dots, (S'_n, C_n)),$$

$$G_l = (V, B, (S''_1, C_1), \dots, (S''_n, C_n)),$$

amb

$$S'_i = S_i - \bigcup_{j=1}^n (V^+ S_j V^* \cup V^* S_j V^+), \quad 1 \leq i \leq n,$$

$$S''_i = S_i - (V^+ S_i V^* \cup V^* S_i V^+), \quad 1 \leq i \leq n.$$

Clarament, si $S_i \in F$, aleshores $S'_i \in F, S''_i \in F$ (REG és tancat sota l'operació de diferència). A més a més, $L_{mg}(G) = L_{ml}(G_g) = L_{in}(G_g)$ i, per tant, $ICC_{mg}(F) \subseteq ICC_{ml}(F)$, i $L_{ml}(G) = L_{in}(G_l)$. Aleshores $ICC_{ml}(F) \subseteq ICC(F)$. $\square$

Per a selector finits també tenim

Lema 10.1.4 $ICC(FIN) \subseteq ICC_{ml}(FIN)$.

Demostració. Considerem una gramàtica $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$

[illegible]

Corol·lari 10.1.1 $ICC(FIN) = ICC_{ml}(FIN)$.

Lema 10.1.5 $ICC(REG) - ICC_{ml}(REG) \neq \emptyset$.

$$G = (\{a, b, c\}, \{abc\}, (b^+c, \{(ab, c)\})).$$

Suposem que $L_{in}(G) = L_{ml}(G')$ per a algun $G' = (\{a, b, c\}, B, (S_1, C_1), \dots, (S_n, C_n))$. Per tal de generar cadenes $a^m b^m c^m$, per a un valor d' m arbitràriament gran, necessitem contextos de la forma $(a^i b^i, c^i)$ o $(a^i, b^i c^i)$. En aquest cas, si tenim un context $(a^i b^j, b^k c^i), i = j + k$, amb $k \geq 1$, aleshores el selector associat ha de ser algun $S \subseteq b^*$ i en les derivacions només podem

usar l'element minimal d' S , és a dir, b^s . A partir d'una cadena $a^p b^p c^p$, per a un valor de p suficientment gran, podem produir $a^{p+i} b^j b^s b^k c^i b^{p-s} c^p$. Aquesta cadena no pertany a $L_{in}(G)$; la qual cosa és una contradicció.

El selector d'un context $(a^i b^i, c^i)$ (el cas de contextos de la forma $(a^i, b^i c^i)$ és similar) ha de ser de la forma $b^j c^k, j \geq 0, k \geq 0$. Si $k = 0$, aleshores, novament, podem produir cadenes amb ocurrències barrejades de b i de c . Per tant, hem d'usar selectors de la forma $b^j c^k$ amb $k \geq 1$. El conjunt d'aquestes cadenes, que són incomparables, és finit (lema de König); si dues cadenes són comparables, aleshores la més llarga no ens és útil i podem esborrar-la sense modificar els llenguatges generats. En conclusió, podem suposar que tots els selectors dels contextos $(a^i b^i, c^i)$ són finits. Aquest fet implica que el conjunt de cadenes $a^m b^m c^m$ generades per G' és finit, ja que el conjunt de subcadenes b^p d'aquestes cadenes és finit. La qual cosa és una contradicció. $\square$

Com era d'esperar, en tots els casos, les gramàtiques contextuais amb selectors regulars són més poderoses que les gramàtiques amb selectors finits.

Lema 10.1.6 $ICC_{mg}(REG) - ICC(FIN) \neq \emptyset$.

Demostració. Considerem la gramàtica

$$G = (\{a, b\}, \{abab\}, (ab^+a, \{(a, a)\}), (ba^+b, \{(b, b)\})).$$

Obtenim

$$L_{mg}(G) = L_{in}(G) = \{a^n b^m a^n b^m \mid n, m \geq 1\}.$$

A partir de la demostració del Lema 9.1.4 de la pàgina 303 sabem que aquest llenguatge no pertany a la família $ICC(FIN)$. $\square$

Corol·lari 10.1.2 Les inclusions $ICC_\alpha(FIN) \subset ICC_\alpha(REG)$, $\alpha \in \{ml, mg\}$, són pròpies.

La primera inclusió del Lema 10.1.3 de la pàgina 371 és pròpia. A més a més,

Lema 10.1.7 $ICC(FIN) - ICC_{mg}(REG) \neq \emptyset$.

Demostració. Considerem el llenguatge

$$L = \{a^n cb^m cb^m ca^n \mid n, m \geq 1\}.$$

Aquest llenguatge pertany a la família $ICC(FIN)$, ja que és generat per

$$G = (\{a, b, c\}, \{acbcba\}, (c, \{(b, b)\}), (acbcba, \{(a, a)\}))$$

(per tal de generar una cadena $a^n cb^m cb^m ca^n$ usem, en primer lloc, $n-1$ vegades el context (a, a) ; després usem $m-1$ vegades el context (b, b)).

No obstant, $L \notin ICC_{mg}(REG)$. En aquest cas, suposem que $L = L_{mg}(G')$ per a algun $G' = (\{a, b, c\}, B, (S_1, C_1), \dots, (S_n, C_n))$. Per tal de generar cadenes $a^n cb^m cb^m ca^n$, amb un valor d' m arbitràriament gran, necessitem contextos de la forma (b^i, b^i) associats amb cadenes $b^p cb^q$, amb $p, q \geq 0$. Per tal de generar cadenes com les anteriors, per a un valor d' n arbitràriament gran,

necessitem contextos de la forma (a^j, a^j) associats amb cadenes $a^s cb^g cb^g ca^t$, amb $g \geq 1$, $s, t \geq 0$.

Considerem una cadena $w_{n,m} = a^n cb^m cb^m ca^n$, per a un valor d' n i d' m arbitràriament gran. Totes aquestes cadenes pertanyen a L i, per tant, existeix una derivació a G de la forma $w \Rightarrow_{mg}^* w_{n,m}$.

Suposem que un context (b^i, b^i) com l'anterior és usat en aquesta derivació:

$$w \Rightarrow_{mg}^* a^{n'} cb^{m'} cb^{m'} ca^{n'} \Rightarrow_{mg} a^{n'} cb^{m'+i} cb^{m'+i} ca^{n'} \Rightarrow_{mg}^* a^n cb^m cb^m ca^n.$$

El selector usat $b^p cb^q$, és una subcadena estricta de $cb^{m'+i} cb^{m'+i} c$ 1, per tant, després d'usar-lo, mai no podrem usar un context (a^j, a^j) . Per tant, en qualsevol derivació a G , si els contextos- a i els contextos- b són usats, aleshores hem d'usar primer els contextos- a i després els contextos- b . Aquest fet és impossible. La raó és que en un pas de derivació $a^n cb^m cb^m ca^n \Rightarrow_{mg} a^{n+j} cb^m cb^m ca^{n+j}$ els contextos- b no canvien la subcadena $cb^m cb^m c$. Si en un pas posterior podem usar un context- b , pel fet que aquest context usa un selector que és una subcadena estricta de $cb^m cb^m c$, se segueix que l'ús del context- a era il·legal, no era possible en el mode mg . Cap context- b pot seguir un context- a .

Conseqüentment, cada derivació només usa contextos- a o només contextos- b , la qual cosa significa que tota cadena $a^n cb^m cb^m ca^n$ de $L_{mg}(G)$ té un dels valors n, m acotats. No obstant, L conté cadenes amb n i m , ambdós, arbitràriament grans. Per tant és una contradicció. $\square$

Corol·lari 10.1.3 Les inclusions $ICC_{mg}(F) \subset ICC_{ml}(F)$, $F \in \{FIN, REG\}$

són pròpies.

Lema 10.1.8 $REG \subseteq ICC_{mg}(FIN)$.

Demostració. A partir de la demostració del Lema 5.3.3 de la pàgina 201, sabem que, donat un llenguatge regular $L \subseteq V^*$, podem construir una gramàtica contextual $G = (V, B, C, \varphi)$ tal que $L = L_{in}(G)$ i

1. $\varphi(x)$ és no-buit només per a un conjunt finit de cadenes $x \in V^*$;
2. C conté només contextos de la forma $(\lambda, v), v \in V^+$.

Considerem aquesta gramàtica G i denotem

$$p = \max\{|w| \mid \varphi(w) \neq \emptyset\}$$

(la longitud més gran d'una cadena selectora). A més a més, per a $t \geq 0$, definim

$$L_{in}(G, t) = \{w \in L_{in}(G) \mid \text{existeix una derivació a } G,$$

$$w_0 \Rightarrow_{in} w_1 \Rightarrow_{in} \dots \Rightarrow_{in} w_m = w, w_0 \in B, \text{ tal que si}$$

$$w_{i-1} = x_{i,1}x_{i,2}x_{i,3}, w_i = x_{i,1}x_{i,2}v_i x_{i,3}, \text{ per a}$$

$$(\lambda, v_i) \in \varphi(v_i), \text{ aleshores } |x_{i,1}x_{i,2}| \geq t, \text{ per a tot } 1 \leq i \leq m\}$$

(el segon membre d'un context és insertat en un lloc que està situat a més de t -símbols de l'extrem esquerre de la cadena considerada).

Clarament, $L_{in}(G) = L_{in}(G, 0)$ i $L_{in}(G, t) \subseteq L_{in}(G)$ per a tot t .

Ara, modifiquem G de la següent manera.

Per a cada $w \in V^*$ i $(\lambda, v) \in C$ tal que $(\lambda, v) \in \varphi(w)$ definim $\varphi_1(w) = \varphi(w)$ i $\varphi_1(wv) = \varphi(wv) \cup \{(\lambda, v)\}$.

A més a més, afegim a B totes les cadenes x' tal que $x \Rightarrow_{in} x'$ per a $x \in B$ i, si $(\lambda, v) \in \varphi(w)$ és el context usat en aquesta derivació, aleshores $x = x_1wx_2$ amb $|x_1w| \leq p$.

Denotem per G_1 la gramàtica (V, B_1, C, φ_1) obtinguda d'aquesta manera. Clarament, $L_{in}(G_1) = L_{in}(G)$ (tots els nous axiomes són cadenes de $L_{in}(G)$ i insertar v després de wv és el mateix que insertar v després de w).

A més a més,

$$L_{in}(G_1) = L_{in}(G_1, s),$$

per a

$$s = \min\{|v| \mid (\lambda, v) \in C\}.$$

En aquest cas, si hem de produir una cadena de $L_{in}(G_1)$ usant un selector w , si aquest selector està emplaçat a una distància menor que p a partir de l'extrem esquerre de la cadena, aleshores existeix un axioma a B_1 que introdueix de nou el context associat (λ, v) i, per tant, no necessitem més aquest pas si comencem per un altre axioma. Per a passos posteriors que usin el mateix selector podem usar el selector wv .

Repetim aquest procés, començant per G_1 . Després d'una nova iteració obtenim G_2 tal que $L_{in}(G) = L_{in}(G_2) = L_{in}(G_2, 2s)$. Després de m passos obtenim G_m tal que $L_{in}(G) = L_{in}(G_m) = L_{in}(G_m, ms)$. Ja que $s > 0$, en com a màxim p passos, obtenim una gramàtica G' per a la qual tenim $L_{in}(G) = L_{in}(G') = L_{in}(G', p) = L$.

Ara, tenint $G' = (V, B', C, \varphi')$ construïda com abans, considerem la gramàtica $G'' = (V, B', C, \varphi'')$, amb φ'' definida per

$$\varphi''(w) = \{(\lambda, v) \mid w = w_1 w_2, w_1, w_2 \in V^*, (\lambda, v) \in \varphi(w_2)\},$$

per a tot $w \in V^*, |w| = p$ (prolonguem cap a l'esquerre tots els selectors de G' , fins a obtenir selector de la mateixa longitud p).

Ja que $L_{in}(G', p) = L_{in}(G')$, tenim $L_{in}(G') = L_{in}(G'')$.

Pel fet que tots els selectors de G'' tenen la mateixa longitud, $L_{in}(G'') = L_{mg}(G'')$. Conseqüentment, $L \in ICC_{mg}(FIN)$. $\square$

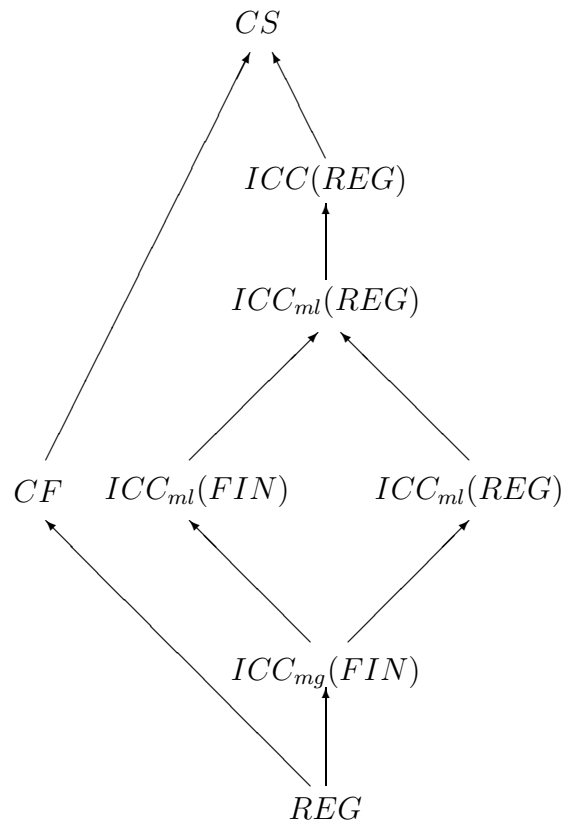
Corol·lari 10.1.4 $REG \subseteq ICC_{mg}(FIN)$.

Sintetitzant els anterior resultats, obtenim

Teorema 10.1.1 Les relacions del diagrama de la Figura 10.1.1 es compleixen. Les fletxes indiquen inclusions estrictes, mentre que les famílies no relacionades en aquest diagrama són incomparables.

Demostració. Les relacions d'aquest diagrama que no siguin conegudes a partir dels capítols previs d'aquesta tesi, se segueixen dels lemes anteriors. Per exemple, sabem que $CF - ICC(REG) \neq \emptyset$. Per altra banda, $ICC_{mg}(FIN)$ conté llenguatges no lliures del context: la gramàtica G de l'exemple 4.2.9 de la pàg. 179 té els selectors a, bc, dc i, per tant, són incomparables entre ells. Per tant, $L_{in}(G) = L_{mg}(G)$, que no pertany a CF . $\square$

Figura 10.1.1 Relacions entre les famílies de llenguatges $ICC_\alpha(F)$, per a $\alpha \in \{ml, mg, -\}$ i amb les famílies de llenguatges de la jerarquia de Chomsky.



10.1.2 El poder de les gramàtiques contextuais amb ús maximal dels selectors

La mateixa demostració que pel Lema 10.1.4 de la pàgina 372 mostra que

Lema 10.1.9 $ICC(FIN) \subseteq ICC_{Ml}(FIN)$.

Rest a encara per demostrar si la inclusió inversa és certa o no. Com era d'esperar una rèplica del corol·lari del Lema 10.1.6 de la pàgina 374 és també cert pel cas maximal.

Lema 10.1.10 $ICC_{\alpha}(REG) - ICC_{\alpha}(FIN) \neq \emptyset, \alpha \in \{Ml, Mg\}$.

Demostració. Per a la gramàtica G de la demostració del Lema 10.1.6 tenim $L_{Ml}(G) = L_{Mg}(G) = \{a^n b^m a^n b^m \mid n, m \geq 1\}$, un llenguatge que no pot ser generat per una gramàtica amb selectors finits (usant selectors finits no podem generar cadenes $a^n b^m a^n b^m$ amb n i m arbitràriament grans). Pel fet que usarem aquesta notació més endavant, denotem el llenguatge $\{a^n b^m a^n b^m \mid n, m \geq 1\}$ mitjançant L_1 . $\square$

Lema 10.1.11 $ICC(FIN) - ICC_{Mg}(REG) \neq \emptyset$.

Demostració. Considerem novament el llenguatge $\{a^n c b^m c b^m c a^n \mid n, m \geq 1\}$ de la demostració del Lema 10.1.7 de la pàgina 375. Pel fet que usarem aquest llenguatge més endavant, el denotem mitjançant L_2 . Sabem que $L \in ICC(FIN) = ICC_{ml}(FIN)$.

Suposem que $L_2 = L_{Mg}(G)$ per a alguna $G = (\{a, b, c\}, B, (S_1, C_1), \dots, (S_k, C_k))$. Per tal de generar cadenes $a^n cb^m cb^m ca^n$, per a un valor d' n i d' m arbitràriament gran, necessitem:

- contextos (a^i, a^i) associats amb selectors de la forma $a^p cb^m cb^m ca^q$, per a algun $p, q \geq 0, m \geq 1$,
- contextos (b^j, b^j) associats amb selectors de la forma $b^s cb^t$, per a algun $s, t \geq 1$ (si un dels s, t és zero, aleshores podem introduir ocurrències de b al davant de la primera ocurrència de c o després de la tercera ocurrència de c en cadenes $a^n cb^m cb^m ca^n$, per a un valor d' m suficientment gran).

Si en una derivació usem un context- a , aleshores no podem usar cap context- b en un pas posterior: mentre que la subcadena central $cb^m cb^m c$ no hagi estat canviada, o bé el context- a és encara aplicable o bé un context- a , amb un selector més llarg, és aplicable. Els contextos- b usen subcadenaes pròpies de $cb^m cb^m c$ i, per tant, no són permesos en el mode Mg .

Per tant, les derivacions a G comencen mitjançant una fase

$$w \Longrightarrow_{Mg}^* a^{n_1} cb^m cb^m ca^{n_1},$$

on només són usats contextos- b , després continua (possiblement) mitjançant una fase

$$a^{n_1} cb^m cb^m ca^{n_1} \Longrightarrow_{Mg}^* a^n cb^m cb^m ca^n,$$

on només són usats contextos- a (i la subcadena cb^mcb^mc no és modificada). Això vol dir que $w = a^{n_1}cb^{m_1}cb^{m_1}ca^{n_1}$, per a algun $m_1 \geq 1$.

Per a un valor $n \geq 1$ donat, denotem

$$M(n) = \{x \in L_{Mg}(G) \mid w \Longrightarrow_{Mg}^* x \text{ usant només contextos-}b \\ w \in B, w = a^n cb^m cb^m ca^n, \text{ per a algun } m \geq 1\}.$$

Sigui

$$n_0 = \max\{n \mid M(n) \text{ és infinit}\}.$$

Totes les cadenes de $M(n_0)$ són de la forma $a^{n_0}cb^mcb^mca^{n_0}$, $m \geq 1$. Denotem

$$M'(n_0) = \{w \in L_{Mg}(G) \mid w \Longrightarrow_{Mg} x \text{ usant contextos-}b, x \in M(n_0)\}.$$

Pel fet que $M(n_0)$ és infinit i usem un conjunt finit de contextos, el conjunt $M'(n_0)$ és també infinit. Pel fet que tot $w \in M(n_0)$ és derivat usant un context- b , se segueix que cap context- a pot ser aplicat a w , ja que sinó la derivació no és dóna en el mode Mg . No obstant, per a cada m tal que $a^{n_0}cb^mcb^mca^{n_0} \in M'(n_0)$ totes les cadenes $z = a^n cb^m cb^m ca^n$, $n \geq 1$, pertanyen a L_2 . Denotem el seu conjunt amb $M''(n_0)$. Per tal de generar aquestes cadenes, per a un valor d' n arbitrari, hem d'usar contextos- a . Pel fet que aquest context $(a^i, a^i) \in \varphi(a^p cb^m cb^m ca^q)$ no pot ser aplicat a $a^{n_0}cb^m cb^m ca^{n_0}$, se segueix que, com a mínim una de les relacions $p > n_0$, $q > n_0$, és compleix. Aquest context no pot ser aplicat als axiomes de B i tampoc no pot ser aplicat a cadenes obtingudes a partir d'axiomes usant només contextos- b . Però ja hem vist que els contextos- a poden ser usats només després dels contextos- b . A més a més, les cadenes de

$M''(n_0)$ han de ser generades a partir d'axiomes $a^{n_1}cb^{m_1}cb^{m_1}ca^{n_1}$ amb $n_1 > n_0$. Mitjançant la selecció d' n_0 , aquests axiomes només són capaços de generar un nombre finit de cadenes de la forma $a^{n_1}cb^mcb^mca^{n_1}$. El conjunt $M''(n_0)$ és infinit, el conjunt d'axiomes és finit i, per tant, $M''(n_0)$ no pot ser abraçat per cadenes generades d'aquesta manera; la qual cosa és una contradicció. La igualtat $L_2 = L_{Mg}(G)$ no és possible i, per tant, $L_2 \notin ICC_{Mg}(REG)$. $\square$

Cal notar que el tipus de selectors no juga cap paper en l'argument anterior i, per tant, $L_2 \notin ICC_{Mg}(ARB)$.

Sorprenement (i a diferència del cas de la restricció minimal), tenim

Lema 10.1.12 $ICC_{Mg}(REG) - (ICC(REG) \cup ICC_{Ml}(REG)) \neq \emptyset$.

Demostració. El llenguatge

$$L_3 = a^+ \cup \{a^n b^n \mid n \geq 1\}$$

no pertany a $ICC(REG)$ (veure la demostració del Lema 6.1.2 de la pàgina 210)). De manera similar, podem veure que $L_3 \notin ICC_{Ml}(REG)$ (els contextos usats per a generar cadenes a^n , per a un valor d' n arbitràriament gran, poden ser també usats per a generar, a partir de $a^n b^n$, cadenes de la forma $a^p b^q$ amb $p > q$, les quals no pertanyin a L). Per altra banda, tenim $L = L_{Mg}(G)$ per a

$$G = (\{a, b\}, \{a, ab\}, (a, \{(\lambda, a)\}), (a^+b, \{(a, b)\}))$$

(en el mode maximal global, el context (λ, a) no pot ser usat per a reescriure cadenes $a^n b^n$; a més, el selector ab no apareix en les cadenes de a^+). $\square$

Corol·lari 10.1.5 $ECC(REG) - ICC_{Ml}(REG) \neq \emptyset$.

Lema 10.1.13 $ECC(F) \subseteq ICC_{Mg}(F)$, $F \in \{FIN, REG\}$.

Demostració. Per a $F = FIN$ l'assertió és òbvia, ja que $ECC(FIN) = FIN$.

Considerem una gramàtica $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$ amb selectors regulars S_i , $1 \leq i \leq n$, i construïm la gramàtica $G' = (V, B, (S_1, C_1), \dots, (S_n, C_n), (S_{n+1}, \{(\lambda, \lambda)\}))$, on

$$S_{n+1} = V^* \left(\bigcup_{i=1}^n S_i \right) V^*.$$

Si $x_1 x_2 x_3 \Rightarrow_{Mg} x_1 u x_2 v x_3$ a G' , aleshores hem de tenir $x_1 = \lambda$, $x_3 = \lambda$. En qualsevol altre cas, la producció $(S_{n+1}, \{(\lambda, \lambda)\})$ ha de ser usada (no canvia res, però impedeix l'ús d'altres produccions quan $x_1 x_3 \neq \lambda$).

A més a més, $L_{Mg}(G') = L_{ex}(G)$, i $ECC(REG) \subseteq ICC_{Mg}(REG)$. $\square$

Cal notar el paper essencial del context buit en aquesta demostració.

Lema 10.1.14 Es compleix que

i) $ICC_{Mg}(FIN) - ICC_{Ml}(REG) \neq \emptyset$,

ii) $ICC_{Mg}(FIN) - ICC(REG) \neq \emptyset$.

Demostració. Considerem el llenguatge

$$L_4 = \{a^n b \mid n \geq 1\} \cup \{a^n b^n \mid n \geq 1\}.$$

Aquest llenguatge pot ser generat en el mode Mg per la gramàtica

$$G = (\{a, b\}, \{ab, a^2b^2\}, (\{ab\}, \{(a, \lambda)\}, (\{a^2b^2\}, \{(a, b)\}))).$$

No obstant, $L_4 \notin ICC_{Ml}(REG) \cup ICC(REG)$. Suposem el contrari i prenem $G' = (\{a, b\}, B, (S_1, C_1), \dots, (S_n, C_n))$ tal que $L_\alpha(G') = L_4, \alpha \in \{in, Ml\}$.

Per tal de generar totes les cadenes $a^n b, n \geq 1$, necessitem un context (a^i, a^j) , amb $i + j > 1$, o bé associat amb $a^s b$ per a algun $s \geq 0$ (aleshores $j = 0$), o bé associat amb $a^k, k \geq 0$. Per a $\alpha = in$, la contradicció és clara: les cadenes $a^{n'} b^n$ amb $n' > n$ poden ser produïdes en ambdós casos.

Suposem que G' és usada en el mode maximal local. Per tal de generar les cadenes $a^n b^n, n \geq 1$, necessitem també un context $(a^k, b^k), k \geq 1$. Aquest context no pot ser aplicat a una cadena a la qual el context (a^i, a^j) anterior pugui ser aplicat (a partir de $a^m b$ obtenim $a^{m+k} b^{k+1}$, que no pertany a L). A més a més, (a^k, b^k) i (a^i, a^j) pot ser usat independentment (els dos contextos pertanyen a conjunts C_s, C_t de G' amb $1 \leq s, t \leq n, s \neq t$). Aquest fet implica que (a^i, a^j) pot ser aplicat novament a una cadena $a^q b^q$, per a un valor de q suficientment gran, produint cadenes que no pertanyen a L . $\square$

Lema 10.1.15 $ICC_{Ml}(REG) - ICC(REG) \neq \emptyset$.

Demostració. Considerem el llenguatge $L_5 = \{x \text{ } mi(x) \mid x \in \{a, b\}^*\}$. A partir del Lema 5.1.9 de la pàgina 196 sabem que $L \notin ICC$. Per a la gramàtica

$$G = (\{a, b\}, \{aa, bb\}, (\{a, b\}^*, \{(a, a), (b, b)\}))$$

tenim $L_{MI}(G) = L_5$ i, per tant, $L \in ICC_{MI}(REG)$. $\square$

En la Secció 8.1 vam provar que $ICC(REG)$ conté llenguatges no-semilineals, després a la Secció 9.2 vam provar que també $ICC(FIN)$ conté llenguatges no-semilineals. A partir de les inclusions $ICC(FIN) \subseteq ICC_{MI}(FIN) \subset ICC_{MI}(REG)$, obtenim, d'aquesta manera, que $ICC_{MI}(FIN)$ i $ICC_{MI}(REG)$ també contenen llenguatges no-semilineals. El cas de $ICC_{Mg}(FIN)$, $ICC_{Mg}(REG)$ serà considerat a continuació.

Lema 10.1.16 La família $ICC_{Mg}(FIN)$ conté llenguatges no-semilineals.

Demostració. Considerem la gramàtica

$$G = (\{a, b, c, d\}, \{c^2(ab)^4c^2\}, P),$$

amb P consistent en les següents produccions:

1. $(\{cca\}, \{(\lambda, a)\})$,	12. $(\{ccabb\}, \{(\lambda, \lambda)\})$,
2. $(\{ccaaba\}, \{(\lambda, a)\})$,	13. $(\{abbabba\}, \{(\lambda, bb)\})$,
3. $(\{ccaabaa\}, \{(\lambda, \lambda)\})$,	14. $(\{abbabbabb, (\lambda, \lambda)\})$,
4. $(\{aaba\}, \{(\lambda, a)\})$,	15. $(\{bcc\}, \{(a, \lambda)\})$,
5. $(\{aaba\}, \{(\lambda, \lambda)\})$,	16. $(\{aaabcc\}, \{(\lambda, \lambda)\})$,
6. $(\{babcc, (\lambda, \lambda)\})$,	17. $(\{babab\}, \{(a, \lambda)\})$,
7. $(\{bcc\}, \{(b, \lambda)\})$,	18. $(\{ababab\}, \{(\lambda, \lambda)\})$,
8. $(\{bbcc\}, \{(\lambda, \lambda)\})$,	19. $(\{cab\}, \{(d, d)\})$,
9. $(\{baabb\}, \{(b, \lambda)\})$,	20. $(\{dcab\}, \{(\lambda, \lambda)\})$,
10. $(\{bbaabb\}, \{(\lambda, \lambda)\})$,	21. $(\{dab\}, \{(\lambda, d)\})$,
11. $(\{cca\}, \{(\lambda, bb)\})$,	22. $(\{dabd\}, \{(\lambda, \lambda)\})$.

La intuïció que s'amaga al darrera d'aquesta construcció és la següent: les produccions de la forma $(\{w\}, \{(\lambda, \lambda)\})$ són usades per tal de prevenir la (incorrecta) aplicació de les produccions associades (w conté, com a subcadena, selectors d'altres produccions). Començant per una cadena $c^2(ab)^nc^2$ (inicialment tenim $n = 4$), usant produccions de les formes 1 – 5 obtenim una cadena $c^2(a^2b)^nc^2$ i, per tant, amb cada ocurrència del símbol a doblada. Aleshores, usant produccions dels tipus 6 – 10, de dreta a esquerra, també doblem el nombre d'ocurrències del símbol b , obtenint $c^2(a^2b^2)^nc^2$. Usant produccions dels tipus 11 – 14, introduïm ara, d'esquerra a dreta, una cadena b^2 entre cada parell d'ocurrències del símbol a . Obtenim una cadena $c^2(ab^2ab^2)^nc^2$. Finalment, usant produccions dels tipus 15 – 17, és introduït un símbol a entre cada parell d'ocurrències del símbol b i, per tant, obtenim una cadena $c^2(abababab)^nc^2$. El

procés pot ser iterat. Aquest fet només es pot realitzar en presència del prefix c^2 i del sufix c^2 . Després d'usar la producció 19, el prefix c^2 és reemplaçat per cdc . Usant les produccions 19 – 22, una cadena $c^2(ab)^m c^2$ pot ser reemplaçada per $cdc(abd)^m c^2$. Aleshores no podem realitzar cap més pas.

Si interseccionem $\Psi_{\{a,b,c,d\}}(L_{Mg}(G))$ amb el conjunt semilineal

$$Q = \{(n, n, 4, n + 1) \mid n \geq 1\},$$

aleshores obtenim el conjunt

$$Q' = \{(4^i, 4^i, 4, 4^i + 1) \mid i \geq 1\},$$

el qual no és semilineal. La família de conjunts semilineals és tancada sota l'operació d'intersecció, la qual cosa prova que $L_{Mg}(G)$ no és semilineal.

Examinem el funcionament de G en el mode maximal global, que ens dona cadenes w tal que $\Psi_{\{a,b,c,d\}}(w) \in Q$. Aquest fet significa que una ocurrència de d ha estat introduïda per a cada parell ab 1, per tant, les produccions 19 i 20 han estat usades per a tots els emplaçaments possibles, obtenint una cadena de la forma $cdc(abd)^m c^2$, $m \geq 1$. Aquest fet només és possible si comencem per una cadena $c^2(ab)^m c^2$. Si comencem per aquesta cadena, introduïm el símbol d des de l'esquerra, però després d'això, usem una producció del tipus 7 o del tipus 15 des de la dreta. Apareixerà, o bé una subcadena b^2 , o bé una subcadena a^2 , i l'ús de produccions del tipus 19 no pot ser realitzat per a tots els m emplaçaments. Per tant, la derivació té dues fases, una d'elles quan usem les produccions del tipus 1 – 18 i l'altra quan usem produccions del tipus 19 – 21. Examinem la primera fase.

Començant amb $c^2(ab)^n c^2$ (inicialment $n = 4$), podem usar o bé la producció 1, que ens porta a $c^2 a^2 b(ab)^{n-1} c^2$, o bé la producció 11, que ens porta a $c^2 ab^3(ab)^{n-1} c^2$. Les produccions 7 i 15 no poden ser usades a causa de la producció 6, i la producció 17 no pot ser usada a causa de la producció 18. Usant una producció que introduïxi el context (λ, λ) no canviem res. Cap producció pot ser aplicada a $c^2 ab^3(ab)^{n-1} c^2$ (per a $n \geq 4$). Per tant, hem de començar amb una producció 1:

$$c^2(ab)^n c^2 \implies_{M_g} c^2 a^2 b(ab)^{n-1} c^2.$$

La producció 1 no pot ser usada novament, a causa de la producció 2. Podem continuar usant la producció 2 (només una vegada, a causa de la producció 3); les produccions 7 i 15 no poden ser usades a causa de la producció 6, i la producció 11 no pot ser usada a causa de la producció 2. A partir de $c^2 a^2 b a^2 b(ab)^{n-2} c^2$ podem continuar, d'esquerra a dreta, usant iterativament la producció 4 i produint $c^2(a^2 b)^n c^2$.

A aquesta cadena podem aplicar-li només les produccions 7 i 15. Aplicant la producció 15, obtenim $c^2(a^2 b)^{n-1} a^3 b c^2$; aleshores no és possible cap altra continuació. Usant la producció 7, obtenim $c^2(a^2 b)^{n-1} a^2 b^2 c^2$. Usant la producció 9, de dreta a esquerra, produïm $c^2(a^2 b^2)^n c^2$. Només després de doblar cada ocurrència de b podem aplicar la producció 11 (en qualsevol altra cas la producció 3 és aplicable). D'esquerra a dreta, mitjançant la producció 13 introduïm ara b^2 per a cada aa i, per tant, produïm $c^2(abbabb)^n c^2$. De dreta a esquerra, ara podem reemplaçar cada bb per bab , usant les produccions 15, 17. Produïm, d'aquesta manera, una cadena $c^2(abababab)^n c^2$.

El procés pot ser iterat.

Cap producció que introdueixi un símbol a o b pot ser usada dues vegades en la mateixa posició, a causa de les produccions que introdueixen (λ, λ) .

Conseqüentment, per tal de produir cadenes amb la imatge de Parikh a Q , hem de treballar de la manera descrita anteriorment, obtenint

$$\Psi_{\{a,b,c,d\}}(L_{Mg}(G)) \cap Q = Q',$$

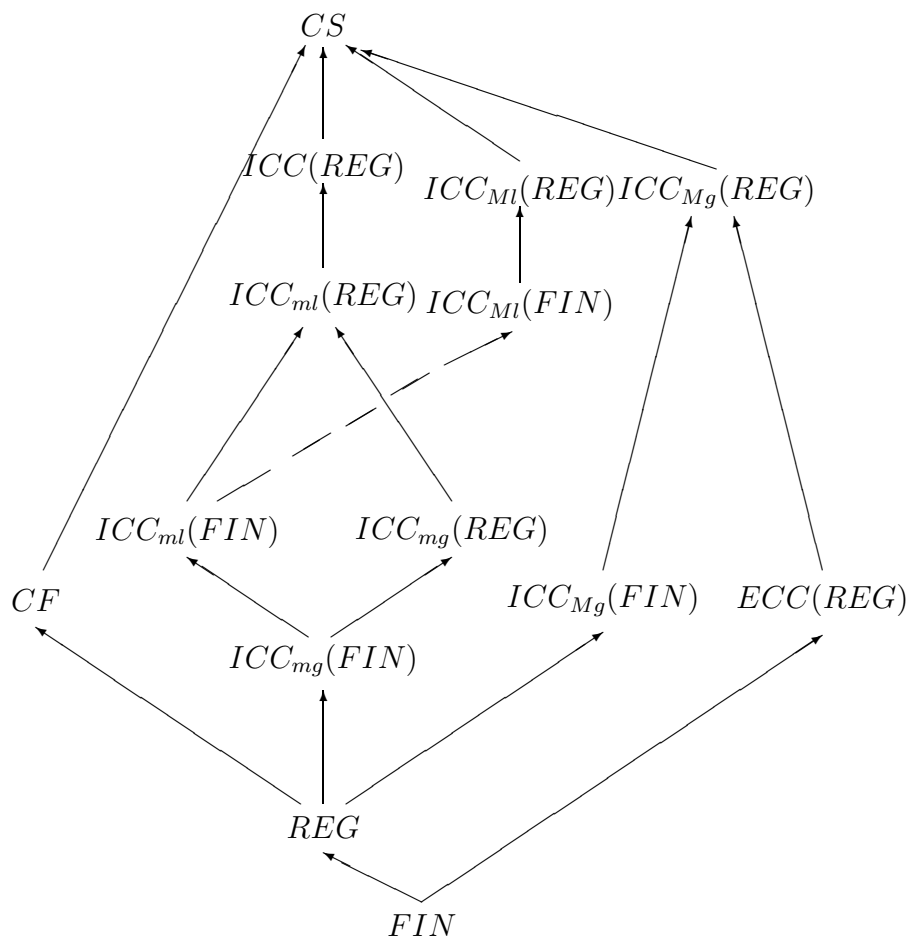
que conclou la demostració. $\square$

Corol·lari 10.1.6 $ICC_{Mg}(FIN) - CF \neq \emptyset$.

Sintetitzarem ara els resultats dels lemes previs, així com els resultats de la Figura 10.1.1 de la pàgina 380.

Teorema 10.1.2 Les relacions del diagrama de la Figura 9.2 es compleixen; les fletxes indiquen inclusions estrictes, les fletxes puntejades indiquen una inclusió que no se sap si és pròpia. Les famílies no connectades en aquest diagrama no són necessàriament incomparables.

Figura 10.1.2 Relacions entre les famílies de llenguatges $ICC_{\alpha}(F)$ i $ECC_{\alpha}(F)$, per a $\alpha \in \{ml, mg, Ml, Mg, -\}$ i amb les famílies de llenguatges de la jerarquia de Chomsky.



Moltes de les famílies d'aquest diagrama són incomparables. En primer lloc, CF i $ECC(REG)$ són incomparables amb qualsevol família de la Figura 10.1.2 que no estigui connectada amb elles: $REG - ECC(REG) \neq \emptyset$ i $ECC(REG) - ICC(REG) \neq \emptyset$, $ECC(REG) - ICC_Ml(REG) \neq \emptyset$ (veure el corol·lari del Lema 10.1.12 de la pàgina 384). Llistarem ara els parells que resten (F_1, F_2) , de famílies de llenguatges contextuals, especificant les relacions d'incomparabilitats

conegudes per a elles de la forma $L, (F_1, F_2), L'$, amb el següent significat: $L \in F_1 - F_2, L' \in F_2 - F_1$. Els llenguatges usats $L_1 - L_5$ són els considerats en els Lemes 10.1.10 de la pàgina 381, el 10.1.11 de la pàgina 381, el 10.1.12 de la pàgina 384, el 10.1.14 de la pàgina 385, i, finalment, el 10.1.15 de la pàgina 386, respectivament; mentre que L_6 és el llenguatge usat en la demostració del Lema 10.1.7 de la 375. Quan L o L' no és especificat, significa que aquest llenguatge és desconegut.

$$\begin{array}{ll}
& (ICC_{mg}(FIN), ICC_{Mg}(FIN)), \quad L_4 \\
& (ICC_{mg}(FIN), ICC_{Mg}(REG)), \quad L_3 \\
L_2, & (ICC(FIN), ICC_{Mg}(FIN)), \quad L_4 \\
L_2, & (ICC(FIN), ICC_{Mg}(REG)), \quad L_1 \\
L_6, & (ICC(FIN), ICC_{Mg}(REG)), \quad L_1 \\
L_1, & (ICC_{mg}(REG), ICC_{Ml}(FIN)), \quad L_6 \\
& (ICC_{mg}(REG), ICC_{Ml}(REG)), \quad L_5 \\
L_1, & (ICC_{mg}(REG), ICC_{Mg}(FIN)), \quad L_4 \\
& (ICC_{mg}(REG), ICC_{Mg}(REG)), \quad L_3 \\
L_1, & (ICC_{ml}(REG), ICC_{Ml}(FIN)) \\
& (ICC_{ml}(REG), ICC_{Ml}(REG)), \quad L_5 \\
L_2, & (ICC_{ml}(REG), ICC_{Mg}(FIN)), \quad L_4 \\
L_2, & (ICC_{ml}(REG), ICC_{Mg}(REG)), \quad L_3 \\
L_1, & (ICC(REG), ICC_{Ml}(FIN)),
\end{array}$$

$$\begin{array}{lll}
& (ICC(REG), ICC_{Ml}(REG)), & L_5 \\
L_1, & (ICC(REG), ICC_{Mg}(FIN)), & L_4 \\
L_2, & (ICC(REG), ICC_{Mg}(REG)), & L_3 \\
L_2, & (ICC_{Ml}(FIN), ICC_{Mg}(FIN)), & L_4 \\
L_2, & (ICC_{Ml}(FIN), ICC_{Mg}(REG)), & L_1 \\
L_1, & (ICC_{Ml}(REG), ICC_{Mg}(FIN)), & L_4 \\
L_2, & (ICC_{Ml}(REG), ICC_{Mg}(REG)), & L_3.
\end{array}$$

A la vista de la incomparabilitat de $ICC(FIN)$ i $ICC_{Mg}(F)$, $F \in \{FIN, REG\}$, i de la importància de les famílies $ICC_{Mg}(F)$ per a la modelització dels llenguatges natural (veure el Capítol 18.2), pot ser molt útil mencionar la següent rèplica del Teorema 9.2.2 de la pàgina 334.

Teorema 10.1.3 Tot llenguatge $L \in RE$ pot ser escrit de la forma $L = h_1(h_2^{-1}(L'))$, on $L' \in ICC_{Mg}(FIN)$, h_1 és una codificació dèbil i h_2 és un morfisme.

Demostració. Prenem $L \subseteq V_T^*$, $L \in RE$, i considerem una gramàtica de Chomsky de tipus-0 $G_0 = (V_N, V_T, S, P)$ per a L , en la forma normal de Kuroda (veure la Definició 3.11.4 de la pàgina 162) i, per tant, contenint regles de la forma:

1. $X \rightarrow YZ$, $X \rightarrow a$, $X \rightarrow \lambda$, per a $X, Y, Z \in V_N$, $a \in V_T$,
2. $XY \rightarrow ZU$, per a $X, Y, Z, U \in V_N$.

Prenem un nou símbol, $c \notin V_T$, i construïm la gramàtica de Chomsky $G_1 = (V_N \cup \{S'\}, V_T \cup \{c\}, S', P')$, on

$$\begin{aligned} P' = & \{S' \rightarrow Sc\} \\ & \cup \{XY \rightarrow ZU \mid XY \rightarrow ZU \in P\} \\ & \cup \{XU \rightarrow xU \mid \text{per a } X \rightarrow x \text{ una regla del tipus 1 de } P \\ & \text{i } U \in V_N \cup V_T \cup \{c\}\}. \end{aligned}$$

És fàcil de veure que $L(G_1) = L(G_0)\{c\}$.

Comencem ara el procediment de la demostració del Teorema 9.2.2 de la pàgina 334 a partir de la gramàtica G_1 i construïm la gramàtica contextual G , exactament com en la Secció 9.2, i extenen els morfismes h_1, h_2 mitjançant

$$h_1(c) = \lambda, \quad h_2(c) = c.$$

Pel fet que totes les regles de P' , amb l'excepció de $S' \rightarrow Sc$, que és usada només una vegada, tenen els membres esquerres de la mateixa longitud, la restricció maximal usant els selectors associats no té cap efecte. Pel que fa als selectors u i $\alpha[u]$, que apareixen en les produccions del tipus 1 i les del tipus 2, respectivament, qualsevol manera d'ús del primer selector, per a algun u que estigui ja mort (com és el cas del segon selector), és il·legal. Es produeixen derivacions no reeixides. El símbol c és preservat per h_2^{-1} i és esborrat per h_1 . Conseqüentment, amb els detalls de la demostració de la Secció 9.2, obtenim $L = h_1(h_2^{-1}(L_{M_g}(G)))$. La qual cosa completa la demostració. $\square$

Corol·lari 10.1.7 Tot $L \in RE$ pot ser escrit de la forma $L = g(L')$, on $L' \in ICC_{Mg}(FIN)$ i g és una màquina seqüencial generalitzada o *gsm*.

Demostració. Una màquina seqüencial generalitzada o *gsm* pot simular, a la vegada, el funcionament de h_1 i el funcionament de h_2^{-1} . $\square$

10.2 Gramàtiques contextuais amb contextos esborrats

En una gramàtica contextual adjuntem contextos, mentre que en un autòmat contextual esborrem contextos. Les dues operacions poden ser considerades d'una manera uniforme en una gramàtica capaç d'adjuntar i d'esborrar contextos. Donarem la definició només per a les gramàtiques contextuais totals, la particularització per a les altres classes és òbvia.

Definició 10.2.1 Un gramàtica contextual (total) *amb contextos esborrats* és una construcció

$$G = (V, B, C_1, C_2, \varphi, \delta),$$

on

- (V, B, C_1, φ) és una gramàtica contextual usual,
- $C_2 \subseteq V^* \times V^*$, i $\delta : V^* \times V^* \times V^* \longrightarrow 2^{C_2}$; els contextos de C_2 són esborrats i δ selecciona els contextos que seran esborrats de la següent

manera: per a $x, y \in V^*$ escriurem $x \implies y$ si, i només si, o bé

$$(1) \quad x = x_1x_2x_3, y = x_1ux_2vx_3, (u, v) \in \varphi(x_1, x_2, x_3),$$

$$\text{o bé } (2) \quad x = x_1ux_2vx_3, y = x_1x_2x_3, (u, v) \in \delta(x_1, x_2, x_3).$$

Definició 10.2.2 Una gramàtica $G = (V, B, C_1, C_2, \varphi, \delta)$ amb contextos esborrats és sense selecció quan $\varphi(x_1, x_2, x_3) = C_1$, $\delta(x_1, x_2, x_3) = C_2$, per a tot $x_1, x_2, x_3 \in V^*$. Aquesta gramàtica s'escriu de la forma $G = (V, B, C_1, C_2)$.

Per a una família X de llenguatges contextuais, per a $X \in \{EC, ECC, IC, ICC, TC\}$, denotem per X_d la corresponent família generada mitjançant gramàtiques amb contextos esborrats (“deleted”). De la manera usual, també podem definir gramàtiques contextuais amb contextos esborrats i amb F selecció, per a una família de llenguatges donada (la restricció és imposada simultàniament a φ i a δ).

És clar que, $X \subseteq X_d$, $X \in \{EC, ECC, IC, ICC, TC\}$, i s'espera que aquestes inclusions siguin pròpies. El següent teorema confirma aquestes expectatives.

Teorema 10.2.1 (i) $ECC_d - TC \neq \emptyset$;

(ii) $EC_d - ECC \neq \emptyset$;

(iii) $IC_d - TC \neq \emptyset$.

Demostració. (i) Considerem la gramàtica

$$G_1 = (\{a, b\}, \{b\}, \{(\lambda, a), (\lambda, b)\}, \{(b, \lambda)\}, \varphi, \delta),$$

$$\varphi(ba^n) = \{(\lambda, a), (\lambda, b)\}, \quad n \geq 0,$$

$$\varphi(ba^n ba^m) = \{(\lambda, a)\}, \quad n \geq 0, m \geq 0,$$

$$\varphi(ba^n ba^n) = \{(\lambda, b)\}, \quad n \geq 0,$$

$$\varphi(ba^n ba^m ba^p) = \{(\lambda, a)\}, \quad n, m, p \geq 0,$$

$$\delta(a^n ba^m ba^m) = \{(b, \lambda)\}, \quad n, m \geq 0.$$

Obtenim

$$\begin{aligned} L_{ex}(G_1) &= ba^* \cup ba^* ba^* \cup \{ba^n ba^n ba^m \mid n, m \geq 0\} \\ &\cup \{a^n ba^n ba^n \mid n \geq 0\}. \end{aligned}$$

Aquest llenguatge no pertany a la família TC , ja que no té la propietat d'IBS: no podem esborrar dues subcadena no-buides de longitud acotada d'una cadena $a^n ba^n ba^n$, amb n arbitràriament gran, de manera que la cadena resultant pertany de nou a $L_{ex}(G)$ (cal notar que la gramàtica G_1 té selecció lineal).

(ii) Prenem

$$G_2 = (\{a, b, c, d\}, \{ab\}, \{(c, \lambda), (d, d)\}, \{(\lambda, bd)\}).$$

Obtenim

$$\begin{aligned} L_{ex}(G_2) = & \{d^{i_1}c^{j_1} \dots d^{i_k}c^{j_k}abd^{i_1+\dots+i_k} \mid i_s \geq 0, j_s \geq 0, 1 \leq s \leq k, k \geq 0\} \\ & \cup \{d^{i_1}c^{j_1} \dots d^{i_k}dc^{j_k}ad^{i_1+\dots+i_k} \mid i_s \geq 0, j_s \geq 0, 1 \leq s \leq k, k \geq 0\}. \end{aligned}$$

En aquest cas, començant per ab podem afegir un nombre arbitrari de contextos (c, λ) i (d, d) . El context (λ, bd) pot ser esborrat només una vegada; és a dir, a partir de cadenes de la forma dc^iabd , i obtenim dc^ia ; a aquesta cadena podem afegir-li de nou els contextos (c, λ) and (d, d) .

Ara tenim

$$Min_0(L_{ex}(G_2)) = \{ab\} \cup \{dc^ia \mid i \geq 0\},$$

que és infinit i, per tant, $L_{ex}(G_2) \notin ECC$ (Lema 5.1.7 de la pàgina 192).

(iii) Considerem la gramàtica

$$\begin{aligned} G_3 = & (\{a, b, c, d, e, b', c', d', e'\}, \{\lambda\}, \\ & \{(a, bb'), (c'c, dd'), (e'e, \lambda)\}, \{(b', c'), (d', e')\}). \end{aligned}$$

Obtenim

$$L_{in}(G_3) \cap a^+b^+c^+d^+e^+ = \{a^n b^n c^n d^n e^n \mid n \geq 1\}.$$

En aquest cas, si b', c', d', e' no apareixen, significa que han estat esborrats per $(b', c'), (d', e')$ i, per tant, tenim una cadena w amb $|w|_{b'} = |w|_{c'}, |w|_{d'} = |w|_{e'}$. Però $|w|_{c'} = |w|_{d'}$ i, per tant, obtenim una cadena w' amb $|w'|_a = |w'|_b = |w'|_c = |w'|_d = |w'|_e$.

Inversament, tenim $\lambda \Rightarrow_{in}^* a^n(bb')^n(c'c)^n(dd')^n(e'e)^n \Rightarrow_{in}^* a^n b^n c^n d^n e^n$.

El llenguatge $L_{in}(G_3)$ no pertany a la família TC , ja que per a les cadenes de la forma $a^n b^n c^n d^n e^n$, amb n arbitràriament gran, la propietat d'IBS no es compleix (Lema 5.1.2 de la pàgina 185). $\square$

A més a més, els contextos esborrats incrementen considerablement el poder de totes les classes de gramàtiques contextuais considerades anteriorment (fins i tot, hom *conjectura* que $ECC_d(REG) \subseteq LIN$). A més, l'ús de contextos esborrats ens permet de tenir símbols en el vocabulari d'una gramàtica contextual usats només en el procés de derivació, els quals no poden aparèixer en el llenguatge generat (de manera similar als elements no-terminals o auxiliars i els símbols estesos usats en les gramàtiques de Chomsky – veure la Definició 3.3.3 de la pàgina 112 – i als sistemes de Lindenmayer – veure la Secció 3.5 –, respectivament).

Específicament,

Definició 10.2.3 Una gramàtica contextual *estesa* és una construcció

$$G = (V, V_0, B, C_1, C_2, \varphi, \delta),$$

on tots els components estan definits igual que abans i $V_0 \subseteq V$. El llenguatge $L_\alpha(G)$, $\alpha \in \{ex, in\}$, generat per aquesta gramàtica, és la intersecció de V_0^* amb $L_\alpha(G')$ i, per tant, $G' = (V, B, C_1, C_2, \varphi, \delta)$, és a dir,

$$L_\alpha(G) = \{w \in V_0^* \mid x \Rightarrow_\alpha^* w, \text{ per algun } x \in B\}.$$

Les gramàtiques contextuais esteses amb contextos esborrats (i funció δ arbitrària) tenen un espectacular gran poder de generació.

Considerem, per exemple, un llenguatge arbitrari $L \subseteq V^*$ i construïm la gramàtica

$$\begin{aligned} G_L &= (V \cup \{c\}, V, \{c\}, \{(\lambda, a), (\lambda, c)\}, \{(c, c)\}, \varphi, \delta), \\ \varphi(cx) &= \{(\lambda, a), (\lambda, c)\}, \quad x \in V^*, \\ \delta(x) &= \{(c, c)\}, \quad x \in L, \end{aligned}$$

on $c \notin V$. Òbviament, $L_{ex}(G) = L$ (observem que φ és una funció de selecció regular). A més a més, tot llenguatge pot ser generat per una gramàtica contextual externa amb contextos esborrats i símbols estesos.

10.3 Gramàtiques contextuais amb contextos unilaterals

En molts exemples i demostracions dels capítols previs hem usat contextos del tipus (u, v) amb només una de les cadenes u, v no-buïdes. Particularment, el treball d'una gramàtica contextual externa amb selecció regular és molt atractiu en el cas en què tots els contextos siguin de la forma (λ, v) ; és a dir, les cadenes són prolongades només per la banda dreta. Suposem, per exemple, que hem obtingut una cadena x en un llenguatge selector S_i . Això voldria dir que podem trobar un camí entre l'estat inicial de l'autòmat finit M_i que reconeix el llenguatge S_i fins a un estat final de M_i . Després d'afegir un context

(λ, v) a x , obtenim $y = xv$. Si M_i pot continuar l'exploració d' x com a una part d' y , és a dir, pot assolir de nou un estat final després de passar sobre v , aleshores podem afegir una nova cadena a la dreta d' y . Si y no pertany a S_i , però realitzem un salt a S_j , per a algun $1 \leq j \leq n, j \neq i$, aleshores podem imaginar-nos que hem passat directament a un estat final de l'autòmat M_j , que correspon a un reconeixement d' y a M_j (i, per tant, tots els autòmats usats en una derivació comencen a treballar al mateix temps i continuen treballant sincrònicament si els contextos són adjuntats d'acord amb la seva selecció). El procés continua d'aquesta manera. El llenguatge generat és un producte de la cooperació entre autòmats associats als seus respectius llenguatges selectors $S_1, \dots, S_n$.

Les consideracions prèvies fan de l'estudi del les gramàtiques contextuais unilaterals un cas molt natural. Denotem per $1X$ la família de llenguatges generats per gramàtiques del tipus X amb només contextos de la forma (λ, v) , i per $11X$ la família generada quan només son permesos els contextos de la dreta (λ, v) i els contextos de l'esquerra (u, λ) (però, no així, els contextos (u, v) amb u i v no-buits), per a $X \in \{EC, ECC, IC, ICC, TC\}$. La selecció restringida es defineix de la manera usual.

Com era d'esperar, el poder de les gramàtiques unilaterals és estrictament més petit que el de les gramàtiques amb contextos arbitraris, però, fins i tot en aquest cas restringit podem obtenir interessants llenguatges.

La següent reformulació de la propietat d'IBS per a gramàtiques contex-

tuals unilaterals és òbvia (veure la Definició 5.1.2 de la pàgina 184 relativa a la propietat d'IBS, i el Lema 5.1.2 de la pàgina 185 relatiu a la caracterització dels llenguatges TC):

Lema 10.3.1 Un llenguatge $L \subseteq V^*$ pertany a la família $1TC$ si, i només si, existeixen les constants p, q tal que tot $x \in L, |x| > p$, pot ser escrit de la forma $x = x_1ux_2$, on $x_1, x_2, u \in V^*$, $x_1x_2 \in L$, i $0 < |u| \leq q$.

Teorema 10.3.1 Es compleix

1. $EC - 11TC \neq \emptyset$, $IC - 11TC \neq \emptyset$;
2. $1EC \subset 11EC \subset (REG \cap EC)$;
3. $1IC = 11IC \subset (CF \cap IC)$, $1IC - LIN \neq \emptyset$;
4. $1ECC(REG) \subset REG$, $11ECC(REG) \subset LIN$, però $1ECC - RE \neq \emptyset$,
 $11ECC(REG) - REG \neq \emptyset$;
5. $1ICC \subset 11ICC$, $IC - 11ICC \neq \emptyset$;
6. $11TC = 1TC$, $1TC(REG) - CF \neq \emptyset$.
7. $REG \subset 1ICC(FIN)$.

Demostració. (1) El llenguatge $\{a^nba^n \mid n \geq 1\}$ pertany a EC , i el llenguatge $\{x \in \{a, b, c\}^* \mid |x|_a = |x|_b = |x|_c\}$ pertany a IC , però cap d'ells

pertany a $11TC$ (aplicant l'anterior versió de la propietat d'IBS per a les cadenes $a^n b a^n, a^n b^n c^n$, respectivament, per a les quals no podem esborrar cap subcadena acotada sense que obtinguem cadenes paràsites).

(2) El llenguatge $a^+ b a^+$ pertany a $11EC - 1EC$ (tota cadena conté una ocurrència de b i, per tant, necessitem els contextos de la dreta i de l'esquerra per tal d'incrementar el nombre d'ocurrències d' a a la dreta i a l'esquerra de b).

Per a $G = (V, B, C)$ amb els contextos pertanyents a C de la forma $(\lambda, v), (u, \lambda)$ tenim $L_{ex}(G) = \{u \mid (u, \lambda) \in C\}^* B \{v \mid (\lambda, v) \in C\}^*$ i, per tant, $L_{ex}(G) \in REG$.

(3) Per a $G = (V, B, C)$ amb C contenint els contextos de la dreta i de l'esquerra, construïm $G' = (V, B, C')$, $C' = \{(\lambda, v) \mid (\lambda, v) \in C, \text{ o } (v, \lambda) \in C\}$, i després construïm la gramàtica lliure del context $G'' = (\{S, X\}, V, S, \{S \rightarrow X a_1 X a_2 \dots X a_k X \mid a_1 a_2 \dots a_k \in B, a_i \in V, 1 \leq i \leq k, k \geq 1\} \cup \{S \rightarrow X \mid \lambda \in B\} \cup \{X \rightarrow \lambda\} \cup \{X \rightarrow X a_1 X a_2 \dots X a_k X \mid (\lambda, a_1 a_2 \dots a_k) \in C', a_i \in V, 1 \leq i \leq k, k \geq 1\}$. Tenim $L_{in}(G) = L_{in}(G') = L(G'')$ degut al mode intern de derivació (sense selecció).

Per altra banda, per a $G = (\{a, b\}, \{\lambda\}, \{(\lambda, ab)\})$, $L_{in}(G)$ és el llenguatge de Dyck sobre $\{a, b\}$, el qual se sap és no-lineal.

(4) La inclusió $11ECC(REG) \subseteq LIN$ és una conseqüència del Lema 9.1.6 de la pàgina 308 i de la inclusió $1ECC(REG) \subseteq REG$ que se segueix com un

cas particular de la demostració d'aquest lema.

La gramàtica G_L de l'Exemple 4.2.6 de la pàgina 177 és unilateral i, per tant, $1ECC - RE \neq \emptyset$.

A més a més, la gramàtica amb selecció regular $G = (\{a, b\}, \{ab\}, \{(a, \lambda), (\lambda, b)\}, \varphi)$ amb

$$\varphi(x) = \begin{cases} (a, \lambda), & |x| \text{ parell,} \\ (\lambda, b), & |x| \text{ imparell,} \end{cases}$$

genera el llenguatge no-regular

$$L_{ex}(G) = \{a^n b^n, a^{n+1} b^n \mid n \geq 1\}.$$

(5) Considerem la gramàtica contextual interna $G = (\{a, b, c, d\}, \{abcd\}, \{(a, c), (b, d)\})$. Tota cadena de la forma $a^n b^m c^n d^m$, amb n, m arbitràriament gran, pertany a $L_{in}(G)$, però no podem esborrar cap subcadena acotada d'aquesta cadena sense que obtinguem una cadena paràsita i, per tant, cap derivació d'una gramàtica amb només contextos unilaterals pot donar-nos aquesta cadena. Per tant, $IC - 11ICC \neq \emptyset$.

Prenem ara la gramàtica $G = (\{a, b, c\}, \{c\}, \{(a, \lambda), (\lambda, ab)\}, \varphi)$, amb $\varphi(xcy) = \{(a, \lambda), (\lambda, ab)\}$ per a tot $x, y \in \{a, b\}^*$. Obtenim $L_{in}(G) = a^* \{c\} D_{\{a, b\}}$. Aquest llenguatge no pertany a $1ICC$. En aquest cas, suposem que $L_{in}(G) = L_{in}(G')$ per a algun $G' = (\{a, b, c\}, B, C, \varphi')$ amb només contextos dretans a C . Ja que tota cadena $a^n c$ pertany a $L_{in}(G')$, hem de tenir un context (λ, a^i) a C' tal que $i \geq 1$ i $(\lambda, a^i) \in \varphi'(a^j)$, $j \geq 0$. Aleshores podem obtenir

$ca^{j+1}b^{j+1} \Rightarrow_{in} ca^{j+1+i}b^{j+1}$, que no pertany a $L_{in}(G)$; la qual cosa és una contradicció (cal notar que l'anterior gramàtica G té selecció regular i, per tant, hem obtingut $11ICC(REG) - 1ICC \neq \emptyset$.)

(6) Tot llenguatge de la família $11TC$ té la propietat del Lema 10.3.1 de la pàgina 403 i, per tant, pertany a $1TC$.

Considerem la gramàtica $G = (\{a, b\}, \{ababa\}, \{(\lambda, a)\}, \varphi)$, amb

$\varphi(\lambda, a^i, ba^jba^k) = \{(\lambda, a)\}$, si i, j, k són tots parells o tots imparells,

$\varphi(a^i, ba^j, ba^k) = \{(\lambda, a)\}$, si o bé i és imparell i j, k són parells,

o i és parell i j, k són imparells,

$\varphi(a^i ba^j, ba^k, \lambda) = \{(\lambda, a)\}$, si o bé i, j són imparells i k és parell,

o i, j són parells i k és imparell.

Aleshores obtenim

$$L(G) = \{a^n ba^n ba^n, a^{n+1} ba^n ba^n, a^{n+1} ba^{n+1} ba^n \mid n \geq 1\},$$

que, òbviament, no és lliure del context. La gramàtica G anterior té selecció regular, en el sentit definit al final de la Secció ?? per a gramàtiques contextuais totals.

(7) És una conseqüència de la demostració del Lema 5.3.3 de la pàg. 201, on els contextos de la forma (λ, v) i els selectors finits són suficient. $\square$

Per a gramàtiques contextuais unilaterals obtenim rèpliques als principals resultats de les Seccions 9.2 i 9.3.

Teorema 10.3.2 Tot llenguatge recursivament enumerable $L \subseteq V^*$ pot ser escrit de la forma $L = (R \setminus L') \cap V^*$, per a $L' \in 1ICC(CF)$ i $R \in REG$.

Demostració. D'acord a (Penttonen, 1974), una *transformació gramatical* és una tripla $\tau = (N, T, P)$, on N, T són vocabularis disjunts i P és un conjunt finit de regles de reescriptura sobre $N \cup T$. Per a un llenguatge $L \subseteq N^*$, definim

$$\tau(L) = \{x \in T^* \mid y \Rightarrow^* x, \text{ per a algun } y \in L\}.$$

El Teorema 3 de (Penttonen, 1974) diu que tot llenguatge $L \in CS, L \subseteq V^*$, pot ser escrit de la forma $L = \tau(L_0)$, per a $L_0 \in REG, L_0 \subseteq N^*$, i $\tau = (N, V, P)$ amb P que conté produccions de la forma

$$A \rightarrow B, AB \rightarrow AC, A \rightarrow a, \text{ per a } A, B, C \in N, a \in V.$$

Prenem ara un llenguatge $L \in RE, L \subseteq V^*$. Existeixen dos nous símbols, $b, c \notin V$, i un llenguatge $L' \subseteq b^*cL, L' \in CS$, tal que per a cada $w \in L$ existeix una cadena $b^i cw$ in $L', i \geq 0$ (aquesta és una lleugera modificació del Teorema 3.9.1 de la pàgina 159). Denotem $V' = V \cup \{b, c\}$. Per a aquest llenguatge L' , considerem $L_0 \in REG$ i τ com abans, $\tau = (N_0, V', P), L_0 \subseteq N_0^*, N_0 \cap V' = \emptyset$ i $L' = \tau(L_0)$. Prenem una gramàtica de Chomsky $G_0 = (N_1, N_0, X_0, P_0)$, amb $N_1 \cap N_0 = \emptyset, N_1 \cap V' = \emptyset$, que generi L_0 . Sense perdre cap tipus de generalitat podem suposar que $N_1 = N'_1 \cup \{X_f\}$ i P_0 conté regles de les formes

$$\text{a) } X_1 \rightarrow X_2 A, \text{ per a } X_1 \in N'_1, X_2 \in N_1, A \in N_0,$$

b) $X_f \rightarrow \lambda$

(aquesta gramàtica existeix sempre per a un llenguatge regular: prenem una gramàtica regular per l'esquerra per a L_0 i reemplacem cada regla terminal $X \rightarrow A$ per $X \rightarrow X_f A$ i, després, afegim la regla $X_f \rightarrow \lambda$).

Construïm la gramàtica contextual $G = (W, \{X_0\}, P')$ amb

$$W = N_1 \cup N_0 \cup V' \cup \{[, \# \},$$

i el conjunt P' que conté les següents produccions:

- (1) $(\{X_1\}, \{(\lambda,]X_2A)\})$, per a $X_1 \rightarrow X_2A \in P_0, X_1 \in N'_1, X_2 \in N_1, A \in N_0$,
- (2) $(\{X_f\}, \{(\lambda,])\})$,
- (3) $(\{A\}, \{(\lambda,]\alpha)\})$, per a $A \rightarrow \alpha \in P, A \in N_0, \alpha \in N_0 \cup V'$,
- (4) $(\{AB\}, \{(\lambda,]AC)\})$, per a $AB \rightarrow AC \in P, A, B, C \in N_0$,
- (5) $(\{\alpha x^{[x]} \mid x \in (N_0 \cup V')^+\}, \{(\lambda,]\alpha)\})$, per a $\alpha \in N_0 \cup V'$,
- (6) $(\{X_f\}\{x^{[x]} \mid x \in (N_0 \cup V')^+\}^+ b^* c, \{(\lambda, \#)\})$.

Considerem també el llenguatge regular

$$R = \{X \mid X \in N'_1\}^+ \{X_f\} (N_0 \cup V' \cup \{]\})^* \{\#\}.$$

Aleshores tenim $L = (R \setminus L_{in}(G)) \cap V^*$.

La intuïció que s'amaga al darrera de la construcció prèvia és la següent. El símbol $]$ és un *assassí*. Cada ocurrència de $]$ assassina un símbol α de $N_0 \cup N_1 \cup V'$ d'acord amb les següents regles:

1. si $x = x_1\alpha]x_2$, per a $\alpha \in N_0 \cup N_1 \cup V'$, aleshores l'ocurrència especificada de α és assassinada per l'especificada ocurrència de $]$;
2. si $x = x_1\alpha x_2]^{x_2}]x_3$, per a $\alpha \in N_0 \cup N_1 \cup V'$ i $x_2 \in (N_0 \cup N_1 \cup V')^*$, aleshores α és assassinat per l'ocurrència de $]$ que hi ha davant d' x_3

(en la regla 2, tots els símbols d' x_2 estan morts, assassinats per les $|x_2|$ ocurrències de $]$ en la banda dreta de x_2).

Les produccions del tipus 1, 2 produeixen una cadena de $L(G_0)$, barrejada amb símbols morts i assassins. Les produccions dels tipus 3, 4 simulen les corresponents regles de τ . Les produccions del tipus 5 mouen símbols vius d'esquerra a dreta, a través de símbols morts i d'assassins. Aquest mecanisme és útil per a preparar AB per a produccions del tipus 4, i per a transportar cap a la dreta còpies vives de símbols terminals. Per tal d'obtenir una cadena de RV^* hem d'usar exactament una vegada una producció del tipus 6. Aquesta producció verifica que els parells $X], X \in N_1$ apareguin a l'esquerra dels símbols usats en la simulació del funcionament de τ i que, tots els símbols a l'esquerra de $\#$, siguin: o bé assassins, o bé símbols morts (i, per tant, la derivació a τ és terminal), o bé ocurrències vives de b i de c .

La igualtat $L = (R \setminus L_{in}(G)) \cap V^*$ pot ser demostrada de manera similar a les demostracions de les igualtats corresponents en el Teorema 9.2.2 de la pàgina 334 i els seus corol·laris. Les següents afirmacions són útils per a aquest objectiu:

1. Per a tota derivació reeixida (és a dir, una derivació que ens dóna una cadena de RV^*) a G , $X_0 \Rightarrow_{in}^* z_1 X_f] z_2 \# w$, existeix una derivació $X_0 \Rightarrow_{in}^* z_1 X_f] w_1 \Rightarrow_{in}^* z_1 X_f] z_2 \# w$, amb produccions dels tipus 1, 2 usades abans de qualsevol altra producció dels altres tipus.
2. Si en una derivació com l'anterior usem més produccions dels tipus 1, 2 per a derivar el prefix $z_1 X_f]$ a $z_1 X_f] z_2 \# w$, aleshores obtenim una derivació que ja no és més reeixida.
3. Si en un pas de derivació $u \Rightarrow_{in} v$ usem una producció del tipus 3, $(\{A\}, \{(\lambda,]\alpha)\})$, per a A viu a u , aleshores A estarà mort a v i α estarà viu a v ; si A està mort a u , aleshores A i α , ambdós, estaran morts a v .
4. Si en un pas de derivació $u \Rightarrow_{in} v$ usem una producció del tipus 4, $(\{AB\}, \{(\lambda,]AC)\})$, per a A, B vius a u , aleshores AB estan morts a v i AC estarà viu a v ; si A, B estan morts a u , aleshores tots els símbols corresponents A, B, C estan morts a v .
5. Si en un pas de derivació $u \Rightarrow_{in} v$ usem una producció del tipus 4, $(\{AB\}, \{(\lambda,]AC)\})$, per a A viu i B mort a u , aleshores la nova ocurrència d' A estarà viva a v ; però l'antiga ocurrència d' A i les ocurrències de B i C estaran mortes a v (en la demostració d'aquesta afirmació és essencial usar el fet que les regles no lliures del context de P tenen el mateix símbol en la primera posició de les seves bandes esquerra i dreta).
6. Si en un pas de derivació $u \Rightarrow_{in} v$ usem una producció del tipus 5, que adjunta $] \alpha$ a algun $\alpha x]^{|x|}$, i α està viu a u , aleshores aquesta ocurrència

de α és morta a v i la recentment introduïda està viva; si α està morta a u , aleshores resta morta a v i l'ocurrència recentment introduïda de α a v també estarà morta.

Denotem per $viu(x)$ la cadena obtinguda esborrant de x tots els símbols morts.

Conseqüentment, si $u \Rightarrow_{in} v$ a G mitjançant una producció π dels tipus 3, 4, aleshores $viu(u) \Rightarrow viu(v)$ a τ usant la regla de P associada a π . A més a més, per a $u \Rightarrow_{in} v$ usant una producció del tipus 5, tenim $viu(u) = viu(v)$.

Les produccions del tipus 6 només poden ser usades una vegada; i, a la dreta del símbol $\#$ introduït d'aquesta manera, només hem de tenir símbols de V . Pel fet que el selector d'una producció del tipus 6 és acotat pel prefix $X_f]$ i un sufix $b^i c$, se segueix que la cadena a la qual apliquem aquesta regla és de la forma $z_1 X_f] u_1 b^i c u_2$ tal que $viu(u_1) = \lambda$ i $u_2 \in V^*$. Per tant, obtenim $z_1 X_f] u_1 b^i c \# u_2$. D'acord amb les afirmacions anteriors, si continuem reescrivint aquesta cadena, aleshores cap nou símbol viu pot ser introduït a l'esquerra de $\#$. Aquest fet prova que $(R \setminus L_{in}(G)) \cap V^* \subseteq L$. La inclusió inversa se segueix fàcilment. $\square$

Corol·lari 10.3.1 La família $1ICC(CF)$ és incomparable amb qualsevol família F tal que $LIN \subseteq F \subset RE$, que sigui tancada sota l'operació de quocient per l'esquerra amb llenguatges regulars i l'operació d'intersecció amb llenguatges regulars.

Novament MAT^λ i $ET0L$ són importants famílies de llenguatges que satisfan les condicions del resultat previ i, per tant, $1ICC(CF) - MAT^\lambda \neq \emptyset$, $1ICC(CF) - ET0L \neq \emptyset$.

També són certes les rèpliques als resultat de la Secció 9.3 pel cas de gramàtiques unilaterals. Denotem per $\mathcal{L}_1(X, Y, Z)$ la família de llenguatges generada per gramàtiques unilaterals associades a les famílies $\mathcal{L}(X, Y, Z)$ com en la Secció 9.3.

Les següents relacions són, o bé òbvies o bé són casos particulars dels resultats obtinguts en les seccions anteriors.

Lema 10.3.2 $\mathcal{L}_1(X, Y, Z) \subseteq REG = \mathcal{L}_1(X, Re, Z)$, per a tots els valors de X, Y, Z .

A més a més tenim

Lema 10.3.3 Es compleix

$$\text{i) } \mathcal{L}_1(Uf, Si, Ma) \subseteq \mathcal{L}_1(Fa, Si, Id) \text{ i}$$

$$\text{ii) } \mathcal{L}_1(Au, Si, Ma) \subseteq \mathcal{L}_1(Au, Si, Id).$$

Demostració. Prenem un llenguatge $L = \mathcal{L}_1(Uf, Si, Ma)$, $L = h(\bigcup_{i=1}^n L_{ex}(G_i))$, on $h : V^* \longrightarrow U^*$ és un morfisme i $G_i = (V, \{w_i\}, P_i)$, per a cada $1 \leq i \leq n$.

El llenguatge L és regular. Construïm la gramàtica $G = (U, A, P)$, amb

$$A = \{h(w_1), \dots, h(w_n)\},$$

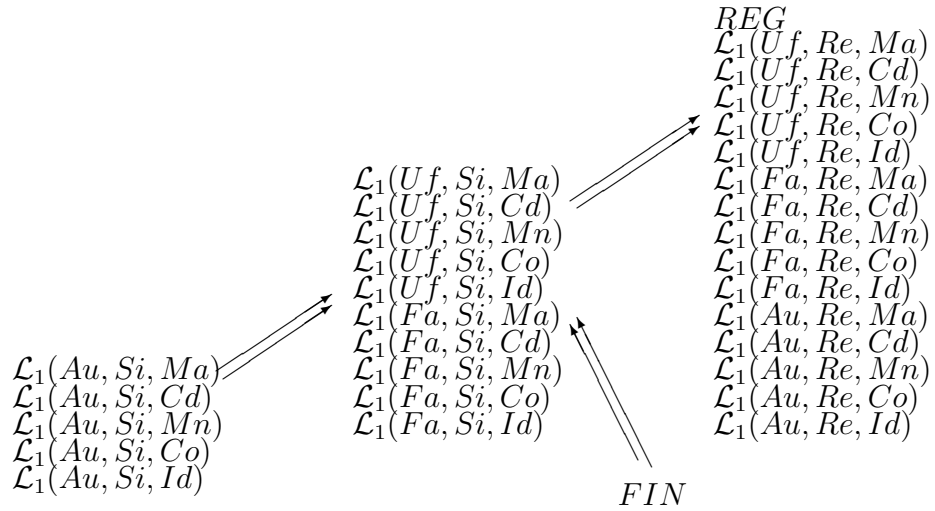
$$P = \{(L/\{h(v)\}, \{(\lambda, h(v))\}) \mid \text{existeix } (S, \{(\lambda, v)\}) \in P_i,$$

per a algun $1 \leq i \leq n\}$.

Òbviament, P és un conjunt finit.

Mitjançant un argument estàndard podem veure que $L = L_{ex}(G)$. $\square$

Figura 10.3.1 Conseqüentment, per a gramàtiques unilaterals obtenim la situació que mostra el diagrama, amb només tres famílies diferents. Les inclusions són pròpies, ja que el llenguatge regular $L_1 = ab^+a$ no pertany a $\mathcal{L}_1(X, Y, Z)$ per a qualsevol X, Y, Z amb $Y \neq Re$ i, qualsevol llenguatge finit pertany a $\mathcal{L}_1(Fa, Si, Id)$. Això implica que $L_2 = \{a, b\}$, considerat també al començament de la Secció 9.3, pertany a $\mathcal{L}_1(Fa, Si, Id) - \mathcal{L}_1(Au, Si, Id)$.



10.4 Gramàtiques contextuais deterministes

Definició 10.4.1 Una gramàtica contextual amb selecció

$$G = (V, B, C_1, C_2, \varphi, \delta)$$

(possiblement amb contextos esborrats), s'anomena *determinista* si

$$\text{card}(\varphi(x)) \leq 1 \text{ i } \text{card}(\delta(x)) \leq 1, \text{ per a tot } x \in V^*.$$

Afegim la lletra D , al davant dels noms de les famílies considerades anteriorment, per a denotar les corresponents famílies de llenguatges generats per gramàtiques deterministes; és a dir, $DECC$, $DICC$, DTC , $DECC_d$, $DEC(REG)$, etc.

Òbviament, $DX \subseteq X$ per a qualsevol família de llenguatges contextuais. Provarem que aquesta inclusió és pròpia per a tot X . No obstant, la suma de factors, determinisme més contextos esborrats, és encara un mecanisme més poderós.

Teorema 10.4.1 Es compleix

- (i) $DECC_d - TC \neq \emptyset$;
- (ii) $DICC_d(REG) - ICC \neq \emptyset$.

Demostració. (i) Per a la gramàtica

$$\begin{aligned} G_1 &= (\{a, b\}, \{ab\}, \{(\lambda, b)\}, \{(a, \lambda)\}, \varphi, \delta), \\ \varphi(ab^i) &= \{(\lambda, b)\}, \quad i \geq 1, \\ \delta(b^{2^n}) &= \{(a, \lambda)\}, \quad n \geq 1, \end{aligned}$$

obtenim

$$L_{ex}(G_1) = \{ab^n \mid n \geq 1\} \cup \{b^{2^n} \mid n \geq 1\},$$

que no pertany a TC , ja que no té la propietat d'IBS (les cadenes b^{2^n} no tenen cap descomposició $x_1ux_2vx_3$ amb uv acotat i $x_1x_2x_3$ pertanyent a $L_{ex}(G_1)$).

(ii) Considerem ara la gramàtica

$$\begin{aligned} G_2 &= (\{a, b, c\}, \{bcaabaab\}, \{(b, b), (b, \lambda)\}, \{(ca, a)\}, \varphi, \delta), \\ \varphi(caa) &= \{(b, b)\}, \\ \varphi(baa) &= \{(\lambda, b)\}, \\ \delta(ab^i a) &= \{(ca, a)\}, \quad i \geq 1. \end{aligned}$$

Començant per $bcaabaab$ podem afegir un nombre arbitrari de contextos (b, b) i (λ, b) , obtenint en aquest cas concret, cadenes de la forma $b^n caab^n aab^m$, $n, m \geq 1$. El context (ca, a) només pot ser esborrat una vegada, ja que c apareix només una vegada i, per tant, obtenim $b^n ab^n ab^m$, $n, m \geq 1$. Cap altra context pot ser afegit a aquesta cadena i, per tant,

$$L_{in}(G_2) = \{b^n caab^n aab^m, b^n ab^n ab^m \mid n, m \geq 1\}.$$

Aquest llenguatge no pertany a *ICC*. En aquest cas, suposem que $L_{in}(G_2) = L_{in}(G)$ per a algun $G = (\{a, b, c\}, B, C, \varphi')$. Les cadenes de la forma $b^n ab^n ab^m$, amb n, m arbitràriament grans, pertanyen a $L_{in}(G_2)$ i, per tant, han de ser generats a G començant amb cadenes $b^i ab^i ab^j$, $i, j \geq 1$, de B i usant contextos $(b^k, b^k) \in \varphi'(b^s ab^t)$, $k \geq 1$, $s, t \geq 0$, per a incrementar el nombre d'ocurrències de b en els primers dos blocs de símbols b . Aquest context també pot ser usat en una derivació de la forma $b^n ab^n ab^m \Rightarrow_{in} b^n ab^{n+k} ab^{m+k}$ (que facilita que n, m siguin suficientment grans) i, per tant, obtenim cadenes paràsites; la qual cosa és una contradicció. $\square$

Per a gramàtiques contextuais externes, el determinisme disminueix significativament el poder de generació.

Teorema 10.4.2 Tot llenguatge de *DECC* és magre.

Demostració. Prenem $G = (V, B, C, \varphi)$ i denotem $k = \text{card}(B)$.

Si $(\lambda, \lambda) \in C$, aleshores podem esborrar-lo sense afectar el llenguatge generat i, per tant, podem suposar que, per a tot $(u, v) \in C$, tenim $uv \neq \lambda$.

Per a cada $x \in B$ existeix una única seqüència, possiblement finita, de la forma

$$S(x) = x, u_1 x v_1, u_2 u_1 x v_1 v_2, \dots,$$

tal que $(u_1, v_1) \in \varphi(x)$, $(u_i, v_i) \in \varphi(u_{i-1} \dots u_1 x v_1 \dots v_{i-1})$, $i \geq 2$ (les derivacions es tenen lloc d'una manera determinista, única). Ja que $(\lambda, \lambda) \notin C$, tenim

$|x| < |u_1 x v_1| < |u_2 u_1 x v_1 v_2| < \dots$. Conseqüentment, per a tot enter positiu n , existeixen com a màxim k cadenes de longitud n a $L_{ex}(G)$ (com a màxim una d'aquestes seqüències $S(x), x \in B$). En conclusió, $L_{ex}(G)$ és magre (amb la constant de magnitud k). $\square$

Corol·lari 10.4.1 $1EC - DECC \neq \emptyset$.

Demostració. El llenguatge $\{a, b\}^*$ és no magre, però pertany a $1EC$. $\square$

Els llenguatges de $DICC$ no són necessàriament magres. Per exemple, considerem la gramàtica (unilateral i amb selecció finita) $G = (\{a, b\}, \{ab\}, \{(\lambda, a), (\lambda, b)\}, \varphi)$ amb $\varphi(a) = \{(\lambda, a)\}, \varphi(b) = \{(\lambda, b)\}$. Aleshores $L_{in}(G) = a^+b^+$, que no és magre (per a cada n , el llenguatge $L_{in}(G)$ conté $n - 1$ cadenes de longitud n).

Fins i tot combinat amb els contextos esborrats (que incrementen el poder), la introducció del determinisme imposa severes restriccions en la capacitat generativa de les gramàtiques contextuais externes.

Teorema 10.4.3 $EC - DECC_d \neq \emptyset$.

Demostració. El llenguatge $L = \{cx \mid x \in \{a, b\}^*\}$ pertany, clarament, a la família EC . Suposem que $L = L_{ex}(G)$ per a alguna gramàtica determinista $G = (\{a, b, c\}, B, C_1, C_2, \varphi, \delta)$. Tot context de $C_1 \cup C_2$ ha de ser de la forma (λ, v) (el símbol c ha de ser present en la primera posició de cada cadena i,

per tant, ni podem afegir ni podem esborrar símbols de l'extrem esquerra de de les cadenes de $L_{ex}(G)$.

Prenem una cadena $x \in B$ i examinem les cadenes de $L_{ex}(G_x)$ per a $G_x = (\{a, b, c\}, \{x\}, C_1, C_2, \varphi, \delta)$. Si només usem els contextos de C_1 , aleshores obtenim la seqüència (possiblement única) $S(x) = x, xv_1, xv_1v_2, \dots$, amb $(\lambda, v_1) \in \varphi(x)$, $(\lambda, v_i) \in \varphi(xv_1 \dots v_{i-1})$, $i \geq 2$. Si comencem per una cadena z de $S(x)$ i esborrem contextos aleshores obtenim, o bé cadenes que ja estaven a $S(x)$, o bé cadenes de la forma $w = w'y$, $w' \in S(x) \cup \{\lambda\}$, $y \in \{a, b\}^*$, per a les quals existeix una cadena $w'w''$ a $S(x)$ (un prefix d'una cadena de $S(x)$). Incrementant novament la longitud, mitjançant l'ús de contextos de C_1 , podem obtenir una nova seqüència de cadenes, possiblement no inclosa a $S(x)$.

Considerem ara totes les seqüències $S(z)$ on, o bé $z = x$, o bé z és una cadena com l'anterior w ; obtingudes a partir d'una cadena de $S(x)$ mitjançant alguns esborraments, però no pertanyents de nou a $S(x)$. Per a cada longitud n , existeix com a màxim una d'aquestes cadenes z (esborrant d'una cadena de $S(x)$ alguns contextos de C_2 , obtenim, o bé una cadena de $S(x)$, o bé una cadena de longitud diferent a les longituds de les cadenes de $S(x)$, però que és un prefix d'una cadena de $S(x)$). Continuem prenent una cadena d'algun $S(z)$, per a $z \neq x$, i examinem les cadenes obtingudes esborrant-hi contextos. Aquestes cadenes són, o bé de $S(z)$ i, per tant, podem ignorar-les (les cadenes que poden ser derivades a partir d'ella estan ja entre el conjunt de cadenes que poden ser derivades de cadenes de $S(z)$), o bé no pertanyen a $S(z)$. En

aquest segon cas, aquestes cadenes són, o bé de longitud intermèdia respecte a les longituds de les cadenes de $S(z)$, ja que hem vist abans que n'hi ha com a màxim una de cada longitud, o bé són més curtes que $|z|$. En aquest últim cas, la nova cadena és, o bé pertanyent a $S(x)$ i, per tant, sense cap interès, o bé té la seva longitud diferent de les longituds de les cadenes de $S(x)$ (ja que $z = xv_1v_2 \dots v_sy'$, una cadena més curta que z i obtinguda a partir de z esborrant contextos és de la forma $xv_1 \dots v_jy''$ i, per tant, de la forma de z ; és a dir, pertany a $S(x)$ o bé és “intermèdia”).

Aquest argument pot ser iterat: tota cadena obtinguda esborrant d'alguna cadena d'una seqüència $S(z)$, no importa com sigui obtinguda z , és, o bé igual a una cadena de $S(z)$, o bé igual a una cadena d'una seqüència a partir de la qual és obtinguda z , o bé igual a una cadena d'una seqüència prèvia (a partir de la qual és obtinguda alguna z' , tal que z és obtinguda de z'), etc.. O bé té la longitud diferent a les longituds de les cadenes en totes aquestes seqüències.

Això implica que, per a cada longitud n , existeix com a màxim una cadena z , $|z| = n$, a la qual podem afegir-li contextos per tal d'obtenir cadenes de longitud més gran i que sigui minimal; en el sentit que, esborrant contextos de z (si és que és possible) obtenim cadenes ja presents en altres seqüències. Conseqüentment, començant per x podem tenir, per a cada n , com a màxim n seqüències que hagin començat amb cadenes de longitud més curta que n . Tota seqüència conté cadenes amb longitud estrictament creixent. A més a més, tenim en total com a màxim n cadenes de longitud n a $L_{ex}(G_x)$ 1, per

tant, $L_{ex}(G)$ conté com a màxim $n \cdot \text{card}(B)$ cadenes de longitud n .

No obstant, el llenguatge L conté 2^{n-1} cadenes de longitud n per a cada $n \geq 1$, per tant, $L_{en}(G)$ no pot ser igual a L . $\square$

A la vista de la relació $DEEC_d - TC \neq \emptyset$ demostrada en el Teorema 10.4.1 de la pàgina 414, el teorema previ prova que els dos trets, el determinisme i l'ús de contextos esborrats, tenen influències bastant contradictòries en el poder generatiu de les gramàtiques contextuais (externes): podem produir llenguatges que no pertanyen a TC però no podem abraçar EC . Aquest fet és reforçat pel següent resultat referit a gramàtiques contextuais internes.

Teorema 10.4.4 $ICC - DTC \neq \emptyset$.

Demostració. Considerem la gramàtica

$$\begin{aligned} G &= (\{a, b, c\}, \{abab\}, C, \varphi), \\ C &= \{(ab, ab), (c, c)\}, \\ \varphi(b^n a^n) &= C, \quad n \geq 1. \end{aligned}$$

El llenguatge $L_{in}(G)$ no pertany a la família DTC . En aquest cas, suposem que $L_{in}(G) = L(G')$, per a alguna gramàtica (determinista) $G' = (\{a, b, c\}, B', C', \varphi')$ amb $\text{card}(\varphi'(x_1, x_2, x_3)) \leq 1$ per a tot $x_1, x_2, x_3 \in \{a, b, c\}^*$. Sense cap pèrdua de generalitat podem suposar que $(\lambda, \lambda) \notin C'$ (aquest context pot ser esborrat sense modificar el llenguatge generat).

Examinem les cadenes $a^n b^n a^n b^n$ and $a^n c b^n a^n c b^n, n \geq 1$. Totes aquestes cadenes pertanyen a $L_{in}(G) = L(G')$. Si algun $z \in L(G')$ produeix aquesta cadena en una derivació directa, $z \Rightarrow a^n b^n a^n b^n$ o $z \Rightarrow a^n c b^n a^n c b^n$: aleshores, l'única possibilitat és tenir $z = a^m b^m a^m b^m$ i els contextos usats són

$$(a^{n-m} b^{n-m}, a^{n-m} b^{n-m}) \in \varphi'(a^m, b^m a^m, b^m)$$

en el primer cas i

$$(a^{n-m} c b^{n-m}, a^{n-m} c b^{n-m}) \in \varphi'(a^m, b^m a^m, b^m)$$

en el segon cas.

A més a més, ja que G' és determinista, per a cada cadena $z = a^m b^m a^m b^m$ existeix com a màxim una derivació $z \Rightarrow a^n b^n a^n b^n, z \Rightarrow a^n c b^n a^n c b^n$ com abans (posteriors derivacions $z \Rightarrow w$ són possibles, però per a descomposicions $z = x_1 x_2 x_3$ diferents a les usades en les derivacions anteriors).

Denotem

$$k = \max\{n \mid a^n b^n a^n b^n \in B', \text{ o } a^n c b^n a^n c b^n \in B'\}.$$

Tenim $k \geq 1$, ja que, com a mínim, $abab$ pertany a B' (aquesta és la cadena més curta de $L(G')$).

Considerem ara els conjunts

$$M_1 = \{a^n b^n a^n b^n \mid 1 \leq n \leq 2k+1\},$$

$$M_2 = \{a^n c b^n a^n c b^n \mid 1 \leq n \leq 2k+1\}.$$

Clarament,

$$M_1 \cup M_2 \subseteq L(G')$$

i, a partir de la discussió anterior, per a cada $y \in (M_1 \cup M_2) - B'$ existeix $x \in M_1$ tal que $x \implies y$. No obstant,

$$\text{card}(M_1 \cup M_2) = 4k + 2,$$

$$\text{card}(B') \leq 2k,$$

i, per tant, hem de tenir, com a mínim, $2k + 2$ derivacions de la forma $x \implies y$ anterior, per a $x \in M_1, y \in (M_1 \cup M_2) - B'$. Pel fet que $\text{card}(M_1) = 2k + 1$, existeix $x \in M_1$ pel qual són possibles dues derivacions $x \implies y_1, x \implies y_2$, per a $y_1 \neq y_2, y_1, y_2 \in (M_1 \cup M_2) - B'$. Aquest fet contradiu el determinisme de G' i, per tant, la igualtat $L_{in}(G) = L(G')$ és impossible i $L_{in}(G) \notin DTC$. $\square$

Corol·lari 10.4.2 Les inclusions $DICC \subset ICC$, i $DTC \subset TC$ són pròpies.

No obstant, les famílies deterministes (fins i tot sense usar contextos esborrats) no són “massa petites”.

Teorema 10.4.5 $1DECC - CF \neq \emptyset$, $DICC(REG) - CF \neq \emptyset$.

Demostració. Per a la gramàtica

$$G = (\{a\}, \{a^3\}, \{(\lambda, a), (\lambda, a^2)\}, \varphi),$$

$$\varphi(a^n) = \begin{cases} \{(\lambda, a^2)\}, & n = 2^i - 1, i \geq 1, \\ \{(\lambda, a)\}, & \text{en qualsevol altre cas,} \end{cases}$$

obtenim $L_{ex}(G) = a^+ - \{a^{2^n} \mid n \geq 0\}$, que no és lliure del context.

Pel cas intern tenim

$$G = (\{a, b\}, \{abababa\}, \{(a, a)\}, \varphi),$$

$$\varphi(ba^i b) = \{(a, a)\}, i \geq 1,$$

i prenem $L_{in}(G) = \{a^n b a^m b a^n b a^m \mid n, m \geq 1\}$, que és clarament no lliure del context. $\square$

Malgrat la capacitat generativa de les gramàtiques contextuais internes amb selecció disminueix quan imposem la restricció determinista, podem trobar llenguatges “fortament no-deterministes” a *DICC*. Considerarem dos exemples.

1. Per a tot $V = \{a_1, \dots, a_k\}, k \geq 1$, el llenguatge V^* pertany a $1DICC(FIN)$, ja que és generat per

$$G = (V, \{x \in V^* \mid |x| \leq k\}, \{(\lambda, a_i) \mid 1 \leq i \leq k\}, \varphi),$$

$$\varphi(x) = \{(\lambda, a_i)\}, x \in V^*, |x| = i, 1 \leq i \leq k.$$

En aquest cas, cada $z \in V^*, |z| \leq k$, és un axioma de G . Tota cadena de longitud com a mínim k pot ser prolongada cap a la dreta mitjançant qualsevol símbol de V , depenent d'un sufix diferent de la cadena. Per tant, $V^* \subseteq L_{in}(G)$. La inclusió inversa és trivial.

2. El llenguatge $L = \{xc \, mi(x) \mid x \in \{a, b\}^*\}$ pertany a $DICC(CF)$. En

aquest cas, considerem la gramàtica

$$G' = (\{a, b, c\}, \{uc\ mi(u) \mid u \in \{a, b\}^*, |u| \leq 2\}, \{(a, a), (b, b), (ba, ab)\}, \varphi),$$

$$\varphi(c) = \{(ba, ab)\},$$

$$\varphi(xc\ mi(x)) = \begin{cases} \{(a, a)\}, & |x| = 2k + 1, k \geq 0, \\ \{(b, b)\}, & |x| = 2k, k \geq 1. \end{cases}$$

En aquest cas, prenem una cadena $z = wc\ mi(w) \in L$ i escollim l'ocurrència més a l'esquerra del símbol b en una posició imparell, o del símbol a en una posició parell; si aquest símbol existeix, el tret imparell/parell de les posicions es compta a partir del c central cap a l'esquerra. Per tant, escrivim $z = b^j(ab)^i bxc\ mi(x)b(ba)^i b^j$, per a $i \geq 0, j \in \{0, 1\}, |x| = 2k, k \geq 1$, o $z = a^j(ba)^i axc\ mi(x)a(ab)^i a^j$, per a $i \geq 0, j \in \{0, 1\}, |x| = 2k + 1, k \geq 0$. Aleshores $z' = b^j(ab)^i xc\ mi(x)(ba)^i b^j$ té tots els símbols b a $b^j(ab)^i$ en posicions imparells i tots els símbols a en posicions parells. Per tant, si $xc\ mi(x)$ pertany a $L_{in}(G')$, aleshores podem afegir-li els contextos $(a, a), (b, b)$, en les posicions extremes, per tal de construir la cadena z' . Si tenim z' , afegim les ocurrències de b adjacents a $xc\ mi(x)$ i obtenim la cadena z . Procedim de manera similar per a l'altra forma de z . D'aquesta manera, reduïm el problema de la construcció de z al problema de la construcció d'una cadena estrictament més curta. Si continuem d'aquesta manera, eventualment, o bé assolim un axioma, o bé una escriptura de z com abans, amb $x = \lambda$. Per tant, de la forma $(ba)^i c(ab)^i, i \geq 1$ (amb totes les ocurrències de b en posicions parells i totes les ocurrències d' a en posicions imparells). Aquesta cadena pot ser produïda usant el context (ba, ab) . Concloem que $z \in L_{in}(G')$ en tots els casos. És a dir, $L \subseteq L_{in}(G')$.

La inclusió inversa és òbvia.

Notes bibliogràfiques. Les gramàtiques contextuais amb ún minimal i maximal dels selectors van ser introduïdes a (Martín-Vide et al., 1995a) o també eren considerades gramàtiques amb ús dispers dels selectors. Molts dels resultats de la Secció 10.1 estan basats en (Martín-Vide et al., 1995a). El Lema 10.1.8 de la pàgina 377 és de (Păun, 1996a). El Lema 10.1.16 de la pàgina 387 és de (Marcus et al., 1996b), el Teorema 10.1.3 de la pàgina 394 és de (Marcus et al., 1995b), on s'observa també que les gramàtiques amb ús maximal dels selectors inclouen les tres construccions no lliures del context bàsiques dels llenguatges naturals.

Les gramàtiques contextuais amb contextos esborrats i amb contextos unilaterals van ser introduïdes a (Păun et al., 1994a). Els resultats de la Secció 10.2 i els del Teorema 10.3.1 de la pàgina 403 són de (Păun et al., 1994a). El Teorema 10.3.2 de la pàgina 407 és de (Ehrenfeucht et al., 1995b), el diagrama de la Figura 10.3.1 de la pàgina 413 (traduït al català) és de (Ehrenfeucht et al., 1995c).

Les gramàtiques contextuais deterministes van ser considerades, per primera vegada, a (Păun et al., 1994a). El Teorema 10.4.4 de la pàgina 420 és de (Ehrenfeucht et al., 1995a).

Capítol 11

Gramàtiques contextuais amb modificacions en les seves relacions de derivació

En les quatre variants anteriors de gramàtiques contextuais hem imposat restriccions en els components de la gramàtica. A partir d'ara considerarem un seguit de modificacions en les relacions de derivació definides respecte a una gramàtica contextual. Considerarem les gramàtiques cobtextuais: amb derivació màxima a l'esquerra, amb derivació paral·lela, amb derivació bloquejada, amb derivació de primera profunditat i, finalment, amb derivació marcada.

11.1 Gramàtiques contextuais amb derivació màxima a l'esquerra

Definició 11.1.1 Donada una gramàtica contextual

$$G = (V, B, (S_1, C_1), \dots, (S_n, C_n)),$$

la *derivació més a l'esquerra* (“leftmost derivation”) respecte a G és defineix per

$$\begin{aligned} x \Rightarrow_{left} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \\ & \text{per a algun } x_1, x_2, x_3 \in V^*, x_2 \in S_i, (u, v) \in C_i, 1 \leq i \leq n, \\ & \text{tal que no existeix cap descomposició } x = x'_1 x'_2 x'_3 \text{ amb} \\ & |x'_1| < |x_1| \text{ i } x'_2 \in S_j, \text{ per a algun } 1 \leq j \leq n. \end{aligned}$$

Amb paraules, adjuntem un context en l'emplaçament més a l'esquerra possible, prenent en consideració totes les possibles descomposicions de la cadena en qüestió i totes les produccions de la gramàtica. Una condició més forta és usar un selector com a *prefix* de la cadena derivada:

$$\begin{aligned} x \Rightarrow_{pr} y \quad \text{sii} \quad & x = x_1 x_2, y = u x_1 v x_2, \text{ on} \\ & x_1, x_2 \in V^*, x_1 \in S_i, (u, v) \in C_i, \text{ per a algun } 1 \leq i \leq n. \end{aligned}$$

Per a aquests dos nous tipus de derivacions podem imposar les restriccions d'ús minimal o maximal en l'ús dels selectors. Considerarem aquí només la

restricció maximal global:

$$\begin{aligned}
 x \Rightarrow_{lfM} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \text{ on } x_1, x_2, x_3 \in V^*, \\
 & x_2 \in S_i, (u, v) \in C_i, \text{ per a algun } 1 \leq i \leq n, \text{ i} \\
 (1) \quad & \text{no existeix cap descomposició } x = x'_1 x'_2 x'_3 \text{ amb} \\
 & |x'_1| < |x_1| \text{ i } x'_2 \in S_j, \text{ per a algun } 1 \leq j \leq n, \\
 (2) \quad & \text{no existeix cap descomposició } x = x_1 x'_2 x'_3, \text{ amb} \\
 & |x'_2| > |x_2| \text{ i } x'_2 \in S_j, \text{ per a algun } 1 \leq j \leq n;
 \end{aligned}$$

$$\begin{aligned}
 x \Rightarrow_{prM} y \quad \text{sii} \quad & x = x_1 x_2, y = u x_1 v x_2, \text{ on } x_1, x_2 \in V^*, \\
 & x_1 \in S_i, (u, v) \in C_i, \text{ per a algun } 1 \leq i \leq n, \text{ i} \\
 & \text{no existeix cap descomposició } x = x'_1 x'_2 \text{ amb} \\
 & |x'_1| > |x_1| \text{ i } x'_1 \in S_j, \text{ per a algun } 1 \leq j \leq n.
 \end{aligned}$$

El llenguatge generat per una gramàtica G , en el mode $\alpha \in \{lf, pr, lfM, prM\}$ (és a dir, usant la relació de derivació $\Rightarrow_\alpha$, respectivament), és denotat per $L_\alpha(G)$. La família d'aquests llenguatges generats per gramàtiques amb F selecció és denotada per $ICC_\alpha(F)$. Si $F = ARB$, aleshores escriurem simplement ICC_α .

A continuació veurem dos exemples.

$$G_1 = (\{a, b, c, d\}, \{cd\}, (\{cd\}, \{(c, d), (a, b)\}), (\{c^2 d^2\}, \{(c, d)\})).$$

Començant per cd , apliquem la primera producció fins que usem el context (c, d) : $cd \Rightarrow_{lf}^* a^i cdb^i \Rightarrow_{lf} a^i ccddb^i, i \geq 0$. A partir d'aquest moment en enda-

vant la primera producció no pot ser usada, ja que la descomposició $(a^i)(ccdd)(b^i)$ força la derivació amb la segona producció: $a^i ccddb^i \Rightarrow_{lf}^* a^i c^j d^j b^i, j \geq 2$. Clarament, les derivacions més a l'esquerra i la més a l'esquerra maximal coincideixen. A més a més,

$$L_{lf}(G_1) = L_{lfM}(G_1) = \{a^n c^m d^m b^n \mid n \geq 0, m \geq 1\}.$$

Aquest és un llenguatge lineal no-regular. Considerem també la gramàtica

$$G_2 = (\{a, b\}, \{aba\}, (a^+ba, \{(a, a)\}), (a^+b, \{(b, b)\})).$$

Començant per aba podem usar qualsevol de les dues produccions. Si usem la primera, aleshores obtenim a^2ba^2 . En general, amb una cadena a^nba^n , podem usar ambdues produccions; la primera producció, afegeix un parell més de símbols a i, per tant, obtenim una cadena de la mateixa forma. Si usem la segona producció, aleshores obtenim $ba^nbb a^n$ i, a partir d'aquell moment, ja no podem aplicar la primera producció. Conseqüentment, tenim

$$L_{lf}(G_2) = \{b^m a^n b^{m+1} a^n \mid n \geq 1, m \geq 0\},$$

un llenguatge que no és lliure del context.

És fàcil de veure que, en el mode maximal a la dreta, la segona producció no és mai usada i, per tant,

$$L_{lfM}(G_2) = \{a^nba^n \mid n \geq 1\}.$$

Investigarem ara la mesura de les famílies $ICC_\alpha(F)$, per a $\alpha \in \{lf, pr, lfM, prM\}$ i $F \in \{FIN, REG\}$, comparant-les amb altres famílies de llenguatges contextuais i amb famílies de la jerarquia de Chomsky.

A partir del primer exemple considerat anteriorment obtenim

Lema 11.1.1 $ICC_\alpha(FIN) - REG \neq \emptyset$, $\alpha \in \{lf, lfM\}$.

No obstant,

Lema 11.1.2 $ICC_\alpha(FIN) \subseteq REG$, $\alpha \in \{pr, prM\}$.

Demostració. Prenem una gramàtica $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$, amb selectors finits. Denotem

$$k = \max\{|x| \mid x \in S_i, 1 \leq i \leq n\},$$

i construïm les gramàtiques lineals per l'esquerra

$$G_\alpha = (N, V, S, P_\alpha), \quad \alpha \in \{pr, prM\},$$

amb

$$N = \{S\} \cup \{[x] \mid x \in V^*, 0 \leq |x| \leq k\},$$

$$P_\alpha = \{S \rightarrow [x] \mid x \in B, |x| \leq k\}$$

$$\cup \{S \rightarrow [x]y \mid xy \in B, |x| = k, y \in V^+\}$$

$$\cup \{[x] \rightarrow [y] \mid x \Rightarrow_\alpha y, [x] \in N, |y| \leq k\}$$

$$\cup \{[x] \rightarrow [x']y \mid x \Rightarrow_\alpha x'y, [x] \in N, |x'| = k, y \in V^+\}$$

$$\cup \{[x] \rightarrow x \mid [x] \in N\}.$$

És fàcil de veure que $L_\alpha(G) = L(G_\alpha)$ (els símbols no-terminals d' N i les regles de P_α controlen la derivació, de manera que a G_α , les derivacions- α a G són simulades exactament). $\square$

Per altra banda, tenim

Lema 11.1.3 $ICC_\alpha(REG) - CF \neq \emptyset$, per a tot α .

Demostració. Considerem la següent variant del segon exemple anterior

$$G = (\{a, b\}, \{aba\}, (a^+ba, \{(a, a), (b, \lambda)\}), (b^+a^+b, \{(b, b)\})).$$

A una cadena a^nba^n (inicialment tenim $n = 1$), en el mode més a l'esquerra i prefix, només podem aplicar-li la primera producció. Usant el primer context obtenim una cadena de la mateixa forma i, per tant, totes les cadenes a^nba^n , $n \geq 1$, poden ser generades. Si usem el segon context, aleshores obtenim ba^nba^n i, a partir d'aquest moment, la primera producció ja no pot ser aplicada; el selector de la segona producció apareix, estrictament, a l'esquerra de l'emplaçament on és present el selector de la primera producció. Usant la segona producció podem generar $b^ma^nb^ma^n$ per a tot $m \geq 1$. Totes les derivacions més a l'esquerra i prefix són maximals. Conseqüentment,

$$L_\alpha(G) = \{a^nba^n \mid n \geq 1\} \cup \{b^ma^nb^ma^n \mid m, n \geq 1\},$$

$\alpha \in \{lf, pr, lfM, prM\}$. Clarament, aquest llenguatge no és lliure del context. $\square$

Una de les inclusions del Lema 11.1.2 de la pàgina 431 és pròpia, l'altra no. De fet, el següent resultat, més puixant, (amb REG en comptes de FIN) és també cert:

Lema 11.1.4 $REG - ICC_{prM}(REG) \neq \emptyset$.

Demostració. Considerem el llenguatge regular

$$L = ba^+ba^*$$

i suposem que $L = L_{prM}(G)$ per a alguna gramàtica $G = (\{a, b\}, B, (S_1, C_1), \dots, (S_n, C_n))$. Tota cadena de B és de la forma $ba^i ba^j$ i, per tant, tots els contextos de G són de la forma $(\lambda, a^k), k \geq 0$, i només podem usar en les derivacions cadenes-selectors de les formes $ba^r, ba^r ba^s, r, s \geq 0$. Sense perdre cap tipus de generalitat, podem suposar que tot selector S_i és, o bé un subconjunt de ba^* (l'anomenarem un selector del tipus I), o bé un subconjunt de $ba^* ba^*$ (que anomenarem selector del tipus II). Si és necessari, reemplacem cada producció (S_i, C_i) per $(S_i \cap ba^*, C_i)$ i $(S_i \cap ba^* ba^*, C_i)$. Pel fet que treballem en el mode maximal, no pot ser usada cap producció del tipus I després d'usar una producció del tipus II (si una cadena-selector del tipus II és present, aleshores el context afegit no el modifica). Per a generar qualsevol cadena $ba^p ba^q$, amb $p > \max\{r \mid ba^r ba^s \in B\}$ i amb $q > \max\{s \mid ab^r ba^s \in B\}$, hem d'usar ambdues produccions del tipus I i del tipus II. Segons l'observació anterior, hem d'aplicar primer les del tipus I.

Per a cada p , les cadenes ba^pba^q pertanyen a L per a tot q i, per tant, per a cada p existeix i_p tal que $ba^pba^s \in S_{i_p}$, per a algun $s \geq 0$. Per a un p donat, denotem $s(p) = \min\{s \mid ba^pba^s \in S_i, \text{ per a algun } 1 \leq i \leq n\}$. A més a més, per a una cadena ba^pba^q , només podem aplicar-li una producció del tipus I si $q < s(p)$ (ja que, sinó, una producció del tipus II és aplicable). Això significa que $L_{prM}(G)$ conté cadenes ba^pba^q , per a un valor de p arbitràriament gran, només si q és acotat per $s(p)$; no obstant, L conté cadenes ba^pba^q amb q no acotat per cap funció respecte a p . La qual cosa és una contradicció. $\square$

Corol·lari 11.1.1 $ICC_\alpha(FIN) - ICC_{prM}(REG) \neq \emptyset$, amb $\alpha \in \{lf, lfM, pr\}$.

Demostració. L'anterior llenguatge L pot ser generat, en els modes lf , lfM , mitjançant la gramàtica

$$G = (\{a, b\}, \{bab\}, (\{ab\}, \{(a, \lambda), (\lambda, a)\})),$$

i, el mode pr , mitjançant la gramàtica

$$G' = (\{a, b\}, \{bab\}, (\{b, bab\}, \{(\lambda, a)\}))$$

(generem primer $baba^q$ de bab i, després, $ba^pba^q, p \geq 1, q \geq 0$). $\square$

Lema 11.1.5 $REG \subseteq ICC_{pr}(FIN)$.

Demostració. Examinem, novament, la demostració del Lema 5.3.3 de la pàgina 201, on, començant per un autòmat finit M que reconeix un llenguatge regular $L \subseteq V^*$, construïm una gramàtica contextual $G = (V, B, C, \varphi)$, amb

selecció finita, tal que $L_{in}(G) = L$. Com ja vam apuntar a la Secció 5.3, si només prenem a C els contextos de la forma (λ, v) , aleshores el llenguatge generat és el mateix.

A més a més, $L_{pr}(G) = L$.

La inclusió $\subseteq$ és òbvia, ja que $L_{pr}(G) \subseteq L_{in}(G) = L$.

Inversament, suposem que $L - L_{pr}(G) \neq \emptyset$ i prenem $x \in L - L_{pr}(G)$ que sigui la de longitud mínima que compleixi aquesta propietat. Seguim novament l'argument de la demostració del Lema 5.3.3 per a la inclusió $L \subseteq L_{in}(G)$. La derivació $x' \Rightarrow_{in} x$ trobada és, de fet, una derivació $x' \Rightarrow_{pr} x$; ja que $x' = u_1 x''$ i $x = u_1 u_3 x''$, per a $(\lambda, u_3) \in \varphi(u_1)$. Aquest fet implica que $x' \in L_{pr}(G)$ (tenim $|x'| < |x|$) i, per tant, $x \in L_{pr}(G)$; una contradicció de la suposició inicial. $\square$

Com sembla intuïtivament, la derivació prefix és més restrictiva que la derivació més a l'esquerra. La primera pot ser simulada per la segona.

Lema 11.1.6 Es compleix

$$ICC_{pr}(F) \subseteq ICC_{lf}(F), \quad ICC_{prM}(F) \subseteq ICC_{lfM}(F), \quad F \in \{FIN, REG\}.$$

Demostració. Prenem una gramàtica $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$ i construïm la gramàtica

$$G' = (V, B, (S_1, C_1), \dots, (S_n, C_n), (\{\lambda\}, \{(\lambda, \lambda)\})).$$

Òbviament, G' és del mateix tipus que G .

Tenim $L_{pr}(G) = L_{lf}(G')$, $L_{prM}(G) = L_{lfM}(G')$.

($\subseteq$) Si $x \Rightarrow_{pr} y$ a G aleshores, clarament, $x \Rightarrow_{lf} y$ a G' ; i si $x \Rightarrow_{prM} y$ a G , aleshores, a causa de λ , el possible nou selector de G' , és més curt que qualsevol $y \in S_i$, $1 \leq i \leq n$. Se segueix que $x \Rightarrow_{lfM} y$ a G' .

Ja que les dues gramàtiques tenen el mateix conjunt d'axiomes, se segueixen les inclusions.

($\supseteq$) L'ús de la producció $(\{\lambda\}, \{(\lambda, \lambda)\})$ no canvia res. Prenem una derivació $x \Rightarrow_{lf} y$ a G' usant una de les produccions (S_i, C_i) , $1 \leq i \leq n$. Escrivim $x = x_1x_2x_3$, $y = x_1ux_2vx_3$, per a algun $x_2 \in S_{i_0}$, $(u, v) \in C_{i_0}$. Si $x_1 \neq \lambda$, aleshores la derivació $x \Rightarrow_{lf} y$ no era més a l'esquerra $(\{\lambda\}, \{(\lambda, \lambda)\})$ poden ser aplicades; la qual cosa és una contradicció. Per tant, $x_1 = \lambda$; és a dir, $x \Rightarrow_{lf} y$ és, de fet, una derivació prefix a G . Si és maximal a G' , aleshores també és maximal a G i, per tant, tenim la inclusió. $\square$

Cal notar l'important paper que juga el context buit en aquesta demostració (similar al cas de les gramàtiques amb ús maximal de selectors, veure, e.g., les demostracions del Lema 10.1.13 de la pàgina 385 i, principalment, del Lema 10.1.16 de la pàgina 387).

Les inclusions anteriors són pròpies. A més a més,

Lema 11.1.7 $ICC_\alpha(FIN) - ICC_\beta(REG) \neq \emptyset$, per a $\alpha \in \{lf, lfM\}, \beta \in \{pr, prM\}$.

Demostració. El llenguatge

$$L = \{ba^nca^n \mid n \geq 0\}$$

pot ser generat en el mode lf , lfM per la gramàtica

$$G = (\{a, b, c\}, \{bc\}, (\{c\}, \{(a, a)\})),$$

però no pot ser generat en el mode prefix: tots els contextos usats han de ser de la forma (λ, v) , però nosaltres necessitem contextos $(a^i, a^i), i \geq 1$, per tal de generar les dues subcadenaes a^n de cadenaes ba^nca^n a L . $\square$

El corol·lari del Lema 11.1.4 de la pàgina 433 pot suggerir que la derivació prefix maximal és dèbil. El següent lema suggereix el contrari (pel cas de selectors regulars).

Lema 11.1.8 $ICC_{prM}(REG) - ICC_\alpha(REG) \neq \emptyset$, per a $\alpha \in \{lf, pr\}$.

Demostració. Considerem el llenguatge $L = \{xmi(x) \mid x \in \{a, b\}^*\}$. Per a la gramàtica

$$G = (\{a, b\}, \{\lambda\}, (V^*, \{(a, a), (b, b)\}))$$

tenim, clarament, $L_{prM}(G) = L$ (la restricció maximal, més restriccions prefix, assegura l'ús de contextos en el mode extern).

Suposem que $L = L_{pr}(G')$ per a algun $G' = (\{a, b\}, B, (S_1, C_1), \dots, (S_n, C_n))$. Ja que $a^m b b a^m \in L$ per a tot m , necessitem un context $(a^i, a^i), i \geq 1$, tal que $(a^i, a^i) \in \varphi(a^r b b a^s)$ per a algun $r, s \geq 0$. Considerem la cadena $z = a^r b b a^s a^i b b a^r$. Aquesta cadena pertany a L per a tot r, s i, per tant, $z \Rightarrow_{pr} a^i a^r b b a^s a^i a^s b b a^r$ és una correcta derivació prefix. No obstant, la cadena produïda no pertany a L ; la qual cosa és una contradicció.

L'argument és el mateix en el cas de derivacions més a l'esquerra. $\square$

A partir dels Lemes 11.1.2 i 11.1.3 obtenim que la inclusió $ICC_\alpha(FIN) \subseteq ICC_\alpha(REG)$ és pròpia per a $\alpha \in \{pr, prM\}$. De manera similar, tenim

Lema 11.1.9 $ICC_\alpha(REG) - ICC_\beta(FIN) \neq \emptyset$, per a $\alpha, \beta \in \{lf, lfM\}$.

Demostració. És fàcil de veure que el llenguatge $L = \{a^n b a^n \mid n \geq 1\} \cup \{b^m a^n b^m a^n \mid m, n \geq 1\}$ de la demostració del Lema 11.1.3 no pot ser generat per una gramàtica amb selectors finits: per tal de generar cadenes $b^m a^n b^m a^n$ amb, per a un valor d' n i d' m arbitràriament gran, necessitem contextos de la forma $(b^i, b^i), (a^j, a^j), i, j \geq 1$, com a mínim un d'ells associat amb cadenes-selectors de longitud arbitrària. $\square$

Passarem ara a comparar les famílies $ICC_\alpha(F), \alpha \in \{lf, pr, lfM, prM\}$, amb $ICC(F)$.

Lema 11.1.10 $ICC_\alpha(FIN) - ICC(REG) \neq \emptyset$, per a $\alpha \in \{lf, lfM\}$.

Demostració. El llenguatge $L = \{bba^m, a^{n+1}ba^nba^{m+1} \mid m \geq 1, n \geq 0\}$, pot ser generat en la manera més a l'esquerra per la gramàtica

$$G = (\{a, b\}, \{bb\}, (\{bb\}, \{(\lambda, a), (a, a)\}), (\{ab\}, \{(a, a)\})).$$

Es pot veure que $L_{lf}(G) = L_{lfM}(G)$ (en el moment que el context (a, a) no és usat, hem d'aplicar la primera producció; després d'això, només podem usar la segona producció).

Aquest llenguatge no pertany a ICC : el context $(a^i, a^i) \in \varphi(a^jba^k), i \geq 1$, per a algun $j, k \geq 0$, necessari per a generar cadenes $a^{n+1}ba^nba$, per a un valor d' n arbitràriament gran, pot ser usat per a derivar $a^{j+1}ba^jba^k \Rightarrow_{in} a^{j+1}ba^{j+i}ba^{k+i}$, obtenint una cadena que no pertany al llenguatge.

Conseqüentment, $ICC_{lf}(FIN)$ i $ICC_{lfM}(FIN)$ contenen llenguatges que no pertanyen a $ICC(REG)$. $\square$

A la vista dels Lemes 11.1.2 i 5.3.3, $ICC_{pr}(FIN)$ i $ICC_{prM}(FIN)$ estan inclosos a $ICC(FIN)$. Més sorprenent és el fet que, per a selectors regulars, aquesta relació no és certa.

Lema 11.1.11 $ICC_{\alpha}(REG) - ICC(REG) \neq \emptyset$, amb $\alpha \in \{pr, prM\}$.

Demostració. Considerem la gramàtica

$$G' = (\{a, b\}, \{bb\}, (\{bb\}, \{(\lambda, a), (a, a)\}), (a^+b, \{(a, a)\})).$$

Només és una modificació de la gramàtica G considerada en la demostració anterior. Obtenim $L_{pr}(G') = L_{prM}(G') = L_{lf}(G)$, i sabem que aquest llenguatge

no pertany a $ICC(REG)$. □

Lema 11.1.12 $LIN - ICC_\alpha \neq \emptyset$, amb $\alpha \in \{lf, lfM\}$.

Demostració. Considerem el llenguatge lineal $L = \{a^m b a^n c a^n \mid n, m \geq 1\}$ i suposem que $L = L_{lf}(G)$ per a alguna $G = (\{a, b, c\}, B, (S_1, C_1), \dots, (S_n, C_n))$. Per tal de generar cadenes $ab^t c a^t$, per a un valor de t arbitràriament gran, necessitem un context $(u, v) = (a^r, a^r) \in C_i, r \geq 1$, per a algun S_i que contingui $a^g c a^h, g, h \geq 0$. Per tal de generar $a^m b a c a$, per a un valor d' m arbitràriament gran, necessitem un context $(u', v') = (a^p, a^q) \in C_j, p + q \geq 1$, per a algun S_j que contingui $a^s, s \geq 0$, o un context $(u', v') = (a^p, \lambda) \in C_j, p \geq 1$, per a algun S_j que contingui una cadena $a^s b z, s \geq 0$. Considerem aquestes cadenes a S_j amb el valor d' s més petit. A qualsevol cadena de la forma $a^f b a^l c a^l, l \geq 1, f \geq s$, respectivament $a^f b z y, f \geq s$, només podem afegir-li el context (u', v') . A més a més, no poden ser generades cadenes $a^m b a^n c a^n$ amb $m \geq s$, per a un valor d' n arbitràriament gran. Per tant, L no pot ser igual a $L_{lf}(G)$; és a dir, $L \notin ICC_{lf}$ (el tipus de selector a G no juga cap paper en aquest argument).

L'argument és el mateix pel cas de derivació més a l'esquerra maximal.

□

Lema 11.1.13 $ICC(FIN) - ICC_\alpha(REG) \neq \emptyset$, amb $\alpha \in \{lf, pr, lfM, prM\}$.

Demostració. Hem de provar només que el llenguatge L considerat anteriorment pertany a $ICC(FIN)$. Ho podem fer considerant la gramàtica

$$G = (\{a, b, c\}, \{abaca\}, (\{b\}, \{(a, \lambda)\}), (\{c\}, \{(a, a)\})).$$

Aleshores, tenim $L_{in}(G) = L$. □

Lema 11.1.14 $ICC_\alpha(REG) \subset CS$, per a tot α .

Demostració. Una senzilla simulació d'una gramàtica contextual treballant en el mode α , per a una gramàtica creixent per l'esquerra prova les inclusions.

Les inclusions són pròpies, ja que tot llenguatge d'una família $ICC_\alpha(REG)$ té la propietat d'IBS. □

Sintetitzant els lemes anteriors, obtenim

Teorema 11.1.1 Les relacions que mostra el diagrama de la Figura 11.1.1 es compleixen. Una fletxa indica inclusions estrictes. Totes les famílies no relacionades en el diagrama són incomparables, excepte pel cas dels set parells

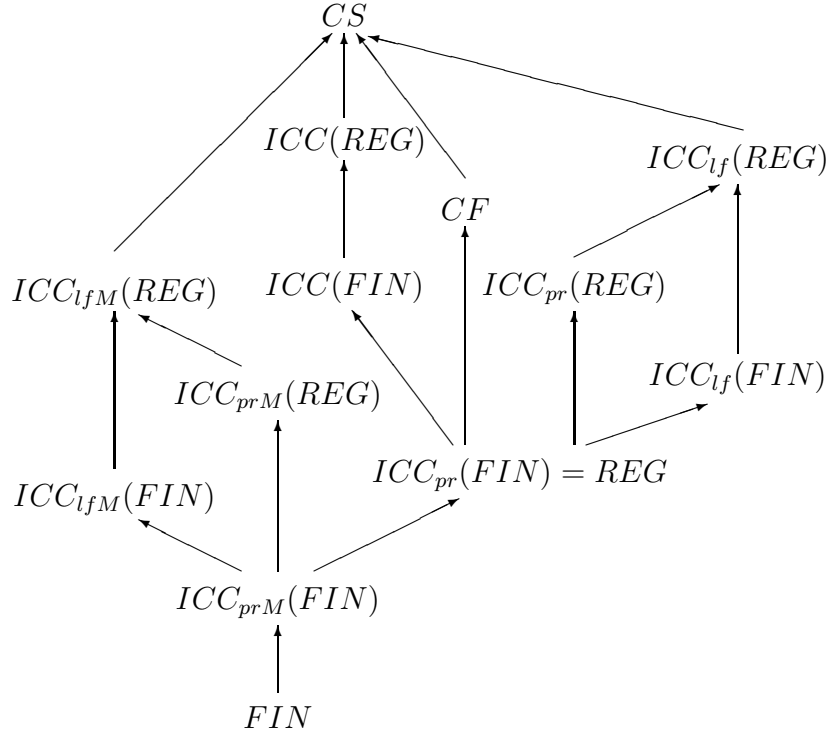
següents, pels quals la seva incomparabilitat resta encara per descobrir:

$$\begin{aligned}
 & (ICC_{lfM}(FIN), REG), \\
 & (ICC_{lfM}(REG), REG), \\
 & (ICC_{lfM}(REG), ICC_{pr}(REG)), \\
 & (ICC_{lfM}(REG), ICC_{lf}(FIN)), \\
 & (ICC_{lfM}(REG), ICC_{lf}(REG)), \\
 & (CF, ICC_{lf}(FIN)), \\
 & (CF, ICC_{lfM}(FIN)).
 \end{aligned}$$

Per a tots aquests set parells (F_1, F_2) sabem que $F_1 - F_2 \neq \emptyset$ (veure els Lemes 11.1.1, 11.1.7, 11.1.7, 11.1.8, 11.1.8, 11.1.12 i 11.1.12, respectivament).

A (Păun, 1996a), Păun conjectura amb el fet que $ICC_{lf}(FIN) \subseteq CF$ i $ICC_{lfM}(FIN) \subseteq CF$. Resta, però, per demostrar.

Figura 11.1.1 Relació entre famílies de derivació màxima a l'esquerra i la jerarquia de Chomsky



Observació 11.1.1 Tancarem aquesta secció amb l'observació que els llenguatges de *reduplicació* i de *concordances creuades*, M_6 i M_9 en la Secció 18.2, poden ser generats mitjançant gramàtiques amb alguna de les restriccions considerades anteriorment.

És clar que, $M_9 \notin ICC_\alpha(REG)$ per a $\alpha \in \{pr, prM\}$: no podem usar cap context de la forma (b^i, d^i) , $i \geq 1$, però necessitem aquests contextos per incrementar les subcadenaes b^m, d^m in $a^n b^m c^n d^m$. Per altra banda, $M_9 \in ICC_\alpha(REG)$ per a $\alpha \in \{lf, lfM\}$: per a

$$G = (\{a, b, c, d\}, \{abcd\}, (b^+c, \{(b, d), (a, c)\}), (aab^+c, \{(a, c)\}))$$

tenim $M_9 = L_{lf}(G) = L_{lfM}(G)$.

El llenguatge M_6 pertany a les quatre famílies: prenem la gramàtica

$$G' = (\{a, b, c\}, \{c\}, (V^*\{c\}, \{(a, a), (b, b)\})).$$

Òbviament, $L_\alpha(G') = L$ per a tot $\alpha \in \{lf, pr, lfM, prM\}$ (les restriccions de més-a-l'esquerra i de prefix forcen als contextos a ser adjuntats en l'única plaça de la dreta).

11.2 Gramàtiques contextuais amb derivació paral·lela

En el cas de la derivació directa usual $\Rightarrow_{in}$ usem un parell selector-context, sense cap restricció. En el mode maximal de derivació, el selector no ha de ser una subcadena d'un altre selector aplicable, però encara no és considerada cap altra restricció referent al selector; en el sentit que, la part que queda de la cadena derivada que no està involucrada en la derivació, no està restringida. El mateix succeeix pel cas de les derivacions màximes a l'esquerra en la secció prèvia. En les derivacions paral·leles el conjunt de la cadena és

Definició 11.2.1 Sigui

$$G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$$

una gramàtica contextual. Definim la següent relació a V^* :

$$x \Rightarrow_p y \quad \text{sii} \quad x = x_1 x_2 \dots x_r, y = u_1 x_1 v_1 u_2 x_2 v_2 \dots u_r x_r v_r, \\ \text{per a } (u_j, v_j) \in C_{i_j}, x_j \in S_{i_j}, 1 \leq j \leq r, 1 \leq i_j \leq n, r \geq 1.$$

Denotem per $L_p(G)$ el llenguatge generat d'aquesta manera i per $CIS_p(F)$ la família d'aquests llenguatges generada per gramàtiques amb F selecció. Quan la selecció no és restringida usarem la notació CIS_p .

Els següents exemples mostren que la derivació paral·lela és molt poderosa:

Exemple 11.2.1

$$G_1 = (\{a, b, c\}, \{abc, a^2 b^2 c^2\}, (a^+ b^+, \{(a, b)\}), (bc^+, \{(\lambda, c)\})), \\ G_2 = (\{a\}, \{a^k\}, (\{a\}, \{(\lambda, a^{k-1})\})), \quad k \geq 1.$$

Tenim

$$L_p(G_1) = \{a^n b^n c^n \mid n \geq 1\}$$

(cada cadena que pot ser derivada a partir de la cadena $a^2 b^2 c^2$ conté, exactament, una subcadena ab i una subcadena bc i, per tant, només poden ser fetes derivacions amb, com a màxim, dos contextos afegits. Quan són afegits exactament dos contextos, el nombre d'ocurrències de símbols a, b, c s'incrementa sincrònicament) i

$$L_p(G_2) = \{a^{k^n} \mid n \geq 1\}$$

(afegim $k - 1$ nous símbols a cada símbol d'una cadena a^m i, per tant, obtenim una cadena de longitud $m + m \cdot (k - 1) = m \cdot k$).

Notem que ambdues gramàtiques són deterministes i que G_2 és unilateral.

A (Păun, 1996a) poden ser trobats els resultats relatius al poder de generació, qüestions de decidibilitat i algunes variants obtingudes de les gramàtiques contextuais amb derivació paral·lela.

11.3 Gramàtiques contextuais amb derivació bloquejada

En una derivació respecte a una gramàtica contextual, totes les cadenes intermèdies són acceptades en el llenguatge generat. Aquest fet és diferent del que succeeix en el cas de les gramàtiques de Chomsky, on només les cadenes terminals són acceptades; és a dir, cadenes per a les quals no és possible cap altre pas de derivació posterior.

Definició 11.3.1 Seguint aquest model, sembla natural considerar també el següent llenguatge associat a una gramàtica contextual $G = (V, B, C, \varphi)$:

$$B_\alpha(G) = \{x \in L_\alpha(G) \mid \text{no existeix } y \in V^* \text{ tal que } x \Longrightarrow_\alpha y\},$$

on $\alpha \in \{ex, in, p\}$. Direm que $B_\alpha(G)$ és *el llenguatge bloquejant* de la gramàtica G en el mode de derivació α . Afegim la lletra B al davant de les notacions usades anteriorment per a les famílies de llenguatges contextuais per tal de denotar les corresponents famílies de llenguatges bloquejats; d'aquesta manera,

obtenim $BEC(F)$, $BICC(F)$, $BICC_p(F)$, etc. Quan $F = ARB$ ometem la seva escriptura. A continuació investigarem breument aquestes famílies.

En primer lloc, considerarem un exemple molt simple:

Exemple 11.3.1

$$G = (\{a\}, \{a^2\}, \{(\lambda, a), (\lambda, \lambda)\}, \varphi),$$

$$\varphi(a^2) = \{(\lambda, a), (\lambda, \lambda)\}.$$

Pel fet que només les cadenes de longitud parell poden ser desenvolupades, i $L_p(G) = \{a^n \mid n \geq 2\}$, obtenim

$$B_p(G) = \{a^{2n+1} \mid n \geq 1\}.$$

Teorema 11.3.1 $BICC = FIN$.

Demostració. Si $L \subseteq V^*$ és un llenguatge finit, aleshores $L = B_{in}(G)$ per a $G = (V, L, \{(\lambda, \lambda)\}, \varphi)$, amb $\varphi(x) = \emptyset$ per a tot $x \in V^*$ (cap pas de derivació pot ser fet a G) i, per tant, $FIN \subseteq BICC(FIN)$.

Inversament, per a una gramàtica $G = (V, B, C, \varphi)$, si tenim un pas de derivació $x = x_1x_2x_3 \Rightarrow_{in} x_1ux_2vx_3 = y$, amb $(u, v) \in \varphi(x_2)$, aleshores també tenim $y \Rightarrow_{in} x_1u^2x_2v^2x_3$. A més a més, $B_{in}(G) \subseteq B$; és a dir, $B_{in}(G) = \{x \in B \mid \text{cap derivació } x \Rightarrow_{in} y \text{ és possible}\}$; la qual cosa implica que $BICC \subseteq FIN$. □

En conclusió, les famílies $BICC(F)$ no tenen cap més interès. La família $BICC_p(FIN)$ conté no només els llenguatges finits (aquest fet pot ser observat de la mateixa manera que en la primera part de l'anterior demostració), sinó també tots els llenguatges co-finitos (sempre i quan siguin λ -lliures).

Teorema 11.3.2 Si $L \subseteq V^+$ tal que $V^* - L$ és finit, aleshores $L \in BICC_p(FIN)$.

Demostració. Per a L com en el teorema, definim $k = \max\{|x| \mid x \in V^* - L\}$ i construïm la gramàtica $G = (V, B, C, \varphi)$, amb

$$\begin{aligned} B &= \{x \in L \mid |x| \leq k\} \cup \{\lambda\}, \\ C &= \{(\lambda, x) \mid k + 1 \leq |x| \leq 2k + 1\}, \\ \varphi(\lambda) &= C, \\ \varphi(x) &= \emptyset, x \in V^+. \end{aligned}$$

Cap cadena de V^* , diferent de λ , pot ser reescrita, però a partir de λ podem introduir totes les cadenes de V^* de longitud com a mínim $k + 1$: cadascuna d'aquestes cadenes x pot ser escrita com a $x = x_1x_2 \dots x_r, r \geq 1$, amb $|x_i| = k + 1, 1 \leq i \leq r - 1, k + 1 \leq |x_r| \leq 2k + 1$ i, per tant, $\lambda \implies_p x_1x_2 \dots x_r = x$ (els contextos necessaris estan a C). $\square$

En el que queda d'aquesta secció, investigarem les famílies $BECC(F)$ de llenguatges $B_{ex}(G)$. És a dir, el conjunt de cadenes que bloquegen la derivació, en gramàtiques contextuais externes amb selecció, restringida a una família F que conté llenguatges infinits (per a gramàtiques externes, $F = FIN$ ens dona, òbviament, llenguatges finits i, per tant, aquest cas no té cap interès).

Lema 11.3.1 Tot llenguatge de la família $BECC$ té un complementari infinit.

Demostració. Prenem $L \subseteq V^*, L \in BECC$. Si L és finit, aleshores $V^* - L$ és trivialment infinit. Si L és infinit, aleshores considerem una gramàtica $G = (V, B, C, \varphi)$ tal que $L = B_{ex}(G)$. Per a cada $x \in L - B$ existeix un pas de derivació $y \Rightarrow_{ex} x$ i, per tant, $y \in V^* - L$. Ja que $|x| - |y|$ està acotat (per la longitud maximal dels contextos de C) se segueix que $V^* - L$ conté cadenes y com abans de longitud arbitrària. Per tant, $V^* - L$ és infinit. $\square$

Teorema 11.3.3 Les famílies $ECC(F)$ són incomparables amb totes les famílies $BECC(F')$, per a F, F' qualsevol família que conté REG .

Demostració. A partir del lema previ trobem que $a^* \notin BECC$, però clarament pertany a $ECC(REG)$.

Inversament, la gramàtica (amb selecció regular)

$$G = (\{a, b\}, \{a\}, \{(\lambda, a), (b, b)\}, \varphi),$$

$$\varphi(a^i) = \{(\lambda, a), (b, b)\}, i \geq 1,$$

genera el llenguatge $B_{ex}(G) = ba^+b$, que no pertany a EC . $\square$

Mitjançant una senzilla modificació de la construcció en el Lema 11.3.1 de la pàgina 449 (prenem un autòmat finit M_1 que reconegui el conjunt de cadenes, pel qual la derivació en la gramàtica inicial G no pot ser continuada i introduïm a la gramàtica lineal G' les regles $S \rightarrow \{(s_{0,1}, s_f)\}$, on $s_{1,0}$ és l'estat inicial de M_1 i s_f és un estat final de M_1) obtenim

Teorema 11.3.4 $BECC(REG) \subset LIN$.

L'anterior Lema 11.3.1 pot ser usat per a obtenir, d'una manera immediata, propietats de clausura negatives de les famílies $BECC(F)$.

Teorema 11.3.5 Les famílies $BECC(F)$, sent F qualsevol família de llenguatges que conté REG , són no-tancades sota les operacions d'unió, concatenació, Kleene $+$, morfismes λ -lliures i morfismes inversos. Si F és tancada sota l'operació d'intersecció amb llenguatges regulars, aleshores també $BECC(F)$ és tancada sota aquesta operació.

Demostració. Unió. Els llenguatges $L_1 = \{a^{2n} \mid n \geq 2\}$ i $L_2 = \{a^{2n+1} \mid n \geq 2\}$ pertanyen a $BECC(REG)$ (per exemple, $L_1 = B_{ex}(G)$, per a $G = (\{a\}, \{a^3\}, \{(\lambda, a), (\lambda, a^2)\}, \varphi)$ amb $\varphi(a^{2n+1}) = \{(\lambda, a), (\lambda, a^2)\}$), però $L_1 \cup L_2 = a^* - \{\lambda, a, a^2, a^3, a^4\}$ no pertany a $BECC$.

Concatenació. Per a L_1 , com abans, obtenim $L_1\{a, a^2\} = a^* - \{\lambda, a, a^2, a^3, a^4\} \notin BECC$.

Kleene $*$. Per a $L = \{a^2, a^3\}$ tenim $L^+ = a^* - \{\lambda, a\} \notin BECC$.

Morfismes. $L = \{a^n b^2 \mid n \geq 1\} = B_{ex}(G)$ per a $G = (\{a, b\}, \{a\}, \{(\lambda, a), (\lambda, b^2)\}, \varphi)$ amb $\varphi(a^i) = \{(\lambda, a), (\lambda, b^2)\}$. Amb els morfisme h definit per $h(a) = h(b) = a$ obtenim $h(L) = a^* - \{\lambda, a, a^2\} \notin BECC$.

Morfismes inversos. Per a $L = \{a^{2n} \mid n \geq 3\} \in BECC(REG)$ i h , definit per $h(a) = aa$, obtenim $h^{-1}(L) = a^* - \{\lambda, a, a^2\} \notin BECC$.

Intersecció amb llenguatges regulars. Per a $G = (V, B, C, \varphi)$ i $R \subseteq V^*$ regular, construïm $G' = (V, B, C \cup \{(\lambda, \lambda)\}, \varphi')$ amb

$$\varphi'(x) = \begin{cases} \varphi(x), & x \in R, \\ \varphi(x) \cup \{(\lambda, \lambda)\}, & x \in V^* - R. \end{cases}$$

Pel fet que tota cadena de $V^* - R$ pot ser desenvolupada indefinidament usant el context (λ, λ) , tenim $B_{ex}(G) \cap R = B_{ex}(G') \in BECC(F)$, que ens assegura que F és tancada sota l'operació d'intersecció amb llenguatges regulars. $\square$

11.4 Gramàtiques contextuais amb derivació de primera profunditat

Els autòmats de pila, els quals sabem són equivalents a les gramàtiques lliures del context (veure la Definició 3.6.3 de la pàgina 138 i el Teorema 3.6.2 de la pàgina 138), tenen una memòria del tipus LIFO: “last-in-first-out”; és a dir, l'últim d'entrar és el primer de sortir. Algunes vegades és important usar també les regles d'una gramàtica lliure del context d'aquesta manera: re-escrivint, en cada pas, un del últims símbols no-terminals introduïts. Aquest fet s'anomena *derivació de primera profunditat* (“depth-first derivation”); la qual, combinada amb l'estratègia FIFO (“first-in-first-out”; és a dir, primer d'entrar,

primer en sortir), s'anomena una *derivació de primera amplada* (“breadth-first derivation”).

Aquest fet també suggereix una variant natural de les gramàtiques contextuais: usar els selectors de manera que, com a mínim un dels membres del context introduït en el pas precedent, és una subcadena del selector usat en aquell moment.

Aquesta derivació té també una interessant motivació des d'un punt de vista pràctic. Sembla lògic pensar que volem generar una cadena de la manera més econòmica possible; saltar d'un emplaçament de la cadena activa a un altre emplaçament diferent pot ser molt costós (imaginem que l'operació és realitzada per un ordinador; moure's al voltant de la cadena implica una despesa de temps). Aleshores sembla interessant buscar derivacions d'alguna manera localitzada, que treballin en cada pas en el veïnatge de l'emplaçament on havíem treballat en el pas previ.

Aquestes gramàtiques tenen també una motivació *a posteriori*: són capaces d'abastar dues de les principals construccions de no-llibertat-del-context en els llenguatges naturals: la reduplicació i les concordances creuades. Un altre element, interessant des del punt de vista teòric, és el fet que el context buit juga un paper crucial en aquestes gramàtiques, ja que si l'usem podem canviar lliurement l'emplaçament de treball, sense modificar la cadena activa.

Definició 11.4.1 Donada una gramàtica contextual

$$G = (V, B, (S_1, C_1), \dots, (S_n, C_n)),$$

definim una *derivació de profunditat primera* (“depth-first derivation”) a G com una derivació

$$w_1 \Rightarrow_{df} w_2 \Rightarrow_{df} \dots \Rightarrow_{df} w_m, \quad m \geq 1,$$

on

- (i) $w_1 \in B, w_1 \Rightarrow_{in} w_2$ en la significació usual,
- (ii) per a cada $i = 2, 3, \dots, m$, si $w_{i-1} = x_1 x_2 x_3, w_i = x_1 u x_2 v x_3$ ((u, v) és el context adjuntat a w_{i-1} per tal d’obtenir w_i), aleshores $w_i = y_1 y_2 y_3, w_{i+1} = y_1 u' y_2 v' y_3$, tal que $y_2 \in S_j, (u', v') \in C_j$, per a algun $j, 1 \leq j \leq n$, i y_2 conté, com a mínim, una de les cadenes u, v com a subcadena. És a dir, com a mínim una de les següent afirmacions és certa:

- $|y_1| \leq |x_1|, |y_1 y_2| \geq |x_1 u|,$
- $|y_3| \leq |x_3|, |y_2 y_3| \geq |v x_3|.$

El conjunt de totes les cadenes $w_m, m \geq 1$, generat per derivacions de profunditat primera en la gramàtica G , més totes les cadenes de B , constitueix el llenguatge generat per G en la manera de profunditat primera i és denotat per $L_{df}(G)$.

Les famílies de llenguatges $L_{df}(G)$, per a G gramàtiques contextuais amb F selecció i usant, possiblement, el context buit (λ, λ) , són denotades per $ICC_{df}^\lambda(F)$. Usem la notació $ICC_{df}(F)$ pel cas quan el context buit no és permès.

Ara considerarem només gramàtiques amb FIN i REG selecció.

En primer lloc examinarem alguns exemples.

Exemple 11.4.1 Considerem la gramàtica

$$G_1 = (\{a, b\}, \{ab, aabb\}, (\{abb, aab\}, \{(a, b)\})).$$

A partir de $aabb$ podem realitzar, una de les dues derivacions següents

$$aabb \Rightarrow_{df} a\underline{a}abb\underline{b}, \quad aabb \Rightarrow_{df} \underline{a}aabb\underline{b}.$$

La cadena obtinguda és la mateixa en ambdós casos, a^3b^3 , però els últims símbols introduïts – els que estan subratllats – són diferents en ambdós casos i, per tant, la continuació és diferent d'un cas a l'altre. Més concretament, per la primera, només podem usar la cadena-selector aab ; mentre que, per la segona, només podem usar la cadena-selector abb . Obtenim,

$$a\underline{a}abb\underline{b} \Rightarrow_{df} a\underline{a}aabb\underline{b}b, \quad \underline{a}aabb\underline{b} \Rightarrow_{df} aa\underline{a}abb\underline{b}b.$$

El procés pot ser iterat i, per tant, obtenim

$$L_{df}(G_1) = \{a^n b^n \mid n \geq 1\}.$$

Cal fer esment al fet que G_1 té selecció finita i que per a la gramàtica

$$G'_1 = (\{a, b\}, \{ab\}, (\{ab\}, \{(a, b)\}))$$

tenim $L_{in}(G'_1) = L_{df}(G_1)$, però $L_{df}(G'_1) = \{ab, aabb\}$: després de $ab \Rightarrow_{df} \underline{aabb}$ no és possible cap més pas de derivació en el mode de primera profunditat.

Si, en canvi, afegim la producció $(\{b\}, \{(\lambda, \lambda)\})$ a G'_1 , aleshores obtenim

$$G''_1 = (\{a, b\}, \{ab\}, (\{ab\}, \{(a, b)\}), (\{b\}, \{(\lambda, \lambda)\})),$$

i de nou tenim $L_{df}(G''_1) = \{a^n b^n \mid n \geq 1\}$. Una derivació de primera profunditat per a aquesta gramàtica és:

$$ab \Rightarrow_{df} \underline{aabb} \Rightarrow_{df} aab\underline{\lambda b \lambda} \Rightarrow_{df} a\underline{a}abb\underline{b}.$$

Observem que el segon pas “oblida” els símbols introduïts prèviament i que el tercer pas fa ús de cadenes “fantasmes”, que identifiquen ab amb $ab\underline{\lambda}$.

Considerem també

$$G_2 = (\{a, b, c, d\}, \{abcd\}, (ab^+c, \{(a, c)\}), (bc^+d, \{(b, d)\})).$$

Novament, a partir de l'axioma, podem realitzar dos passos de derivació diferents:

$$abcd \Rightarrow_{df} \underline{a}abcc\underline{d} \quad (1)$$

$$abcd \Rightarrow_{df} ab\underline{b}cdd\underline{d} \quad (2)$$

A la primera cadena només podem aplicar-li la segona producció de G_2 (el símbol actiu c és inclòs pel selector), mentre que a la segona cadena només podem aplicar-li la primera producció. Obtenim, respectivament,

$$\underline{a}abcc\underline{d} \Longrightarrow_{df} aab\underline{b}cc\underline{d},$$

$$a\underline{b}b\underline{c}d\underline{d} \Longrightarrow_{df} \underline{a}abb\underline{c}c\underline{d}d.$$

Els papers s'han intercanviat: a la primera cadena ara només podem aplicar-li la primera producció i obtenim una cadena de la forma a (1). A la segona cadena només podem aplicar-li la segona producció i obtenim una cadena de la forma a (2). El procés pot ser iterat, a cada pas el nombre d'ocurrències d' a i de b (i, per tant, també de c i de d) sent el mateix o fiferenciant-se en una sola ocurrència. A més a més,

$$L_{df}(G_2) = \{a^n b^n c^n d^n, a^{n+1} b^n c^{n+1} d^n, a^n b^{n+1} c^n d^{n+1} \mid n \geq 1\}.$$

Exemple 11.4.2 Considerem ara les gramàtiques

$$G_3 = (\{a, b, c\}, \{c\}, (\{c\}V^*, \{(a, a), (b, b)\})),$$

$$G_4 = (\{a, b, c, d\}, \{abcd\}, (a^+ b^+ c^+, \{(a, c)\}), (b^+ c^+ d^+, \{(b, d)\})).$$

Es pot verificar que

$$L_{df}(G_3) = \{xcx \mid x \in \{a, b\}^*\},$$

$$L_{df}(G_4) = \{a^n b^m c^n d^m \mid n, m \geq 1\}.$$

A continuació mostarem algunes curtes derivacions de cadascuna d'aquestes

gramàtiques:

$$\begin{aligned}
 c &\Rightarrow_{df} \underline{a}c\underline{a} \Rightarrow_{df} a\underline{a}c\underline{a} \Rightarrow_{df} aab\underline{c}aab \Rightarrow_{df} aab\underline{a}caab\underline{a}, \\
 abcd &\Rightarrow_{df} \underline{a}abc\underline{c}d \Rightarrow_{df} a\underline{a}abcc\underline{c}d \Rightarrow_{df} \\
 &\Rightarrow_{df} a\underline{a}aabcc\underline{c}cd \Rightarrow_{df} aaaab\underline{b}cccc\underline{d}d.
 \end{aligned}$$

Observem que, en el cas de G_3 , la restricció de primera profunditat força l'adjunció del nou context en l'única posició de la dreta, assegurant, d'aquesta manera, la igualtat de les dues parts de la cadena obtinguda.

A continuació investigarem breument el poder de les gramàtiques contextuais amb derivació de primera profunditat. Directament de les definicions, tenim

Lema 11.4.1 Es compleix

$$ICC_{df}(FIN) \subseteq ICC_{df}(REG);$$

$$ICC_{df}^{\lambda}(FIN) \subseteq ICC_{df}^{\lambda}(REG);$$

$$ICC_{df}(F) \subseteq ICC_{df}^{\lambda}(F), \text{ amb } F \in \{FIN, REG\};$$

$$FIN \subseteq ICC_{df}(F), \text{ per a tot } F.$$

El llenguatge $L_{df}(G_2)$ anterior no té la propietat de bombeig del Lema 5.1.5 de la pàgina 189 i, per tant, no pertany a ICC .

De fet, un resultat més puixant és també cert:

Lema 11.4.2 $ICC_{df}(FIN) - ICC \neq \emptyset$.

Demostració. Considerem la gramàtica

$$G = (\{a, b, c\}, \{a, c\}, (S_1, C_1), (S_2, C_2), (S_3, C_3)),$$

$$S_1 = \{a\}, \quad C_1 = \{(\lambda, a)\},$$

$$S_2 = \{c\}, \quad C_2 = \{(a^2c^2, c^2b^2)\},$$

$$S_3 = \{a^2c^5, c^5b^2\}, \quad C_3 = \{(a^2, b^2)\}.$$

Començant per axioma a i usant la primera producció, podem generar totes les cadenes $a^n, n \geq 1$. A una cadena a^n no podem aplicar-li cap de les altres produccions. Començant per c , i usant (S_2, C_2) , obtenim $\underline{a^2c^2c^2b^2}$. La segona producció no pot ser aplicada de nou. Tampoc és aplicable la segona producció després d'usar (S_3, C_3) , ja que no són introduïts nous símbols c . Ni tampoc la primera. També després d'usar (S_3, C_3) , la primera producció no és aplicable: sempre introduïm dues ocurrencies d' a de S_1 que poden copsar només un de les produccions. Conseqüentment, només pot ser usat (S_3, C_3) . Com en el primer exemple considerat anteriorment, es pot veure que podem usar alternativament les cadenes-selectors a^2c^5 i c^5b^2 , a derivacions de la forma

$$\underline{a^2c^2c^2b^2} \Rightarrow_{df} \underline{a^2}a^2c^5\underline{b^2}b^2 \Rightarrow_{df} a^4\underline{a^2}c^5b^2\underline{b^2}b^2 \Rightarrow_{df} \dots$$

En conclusió,

$$L_{df}(G) = \{c\} \cup \{a^n \mid n \geq 1\} \cup \{a^{2n}c^5b^{2n} \mid n \geq 1\}.$$

Aquest llenguatge no pertany a la família ICC : per tal d'obtenir cadenes a^n , per a un valor d' n arbitràriament gran, no necessitem un context $(a^i, a^j), i+j \geq$

1, associat a cadenes-selectors $a^k, k \geq 0$. Aquest context pot ser aplicat a $a^{2n}c^5b^{2n}$, per a un valor d' n suficientment gran, produint $a^{2n+i+j}c^5b^{2n}$, que no pertany a $L_{df}(G)$. Una contradicció que finalitza la demostració. $\square$

Com ja havíem insinuat, l'ús del context buit és molt important:

Lema 11.4.3 $ICC(F) \subseteq ICC_{df}^\lambda(F)$, per a $F \in \{FIN, REG\}$.

Demostració. Prenem una gramàtica contextual $G = (V, B, (S_1, C_1), \dots, (S_n, C_n))$ i construïm la gramàtica

$$G' = (V, B, (S_1, C_1), \dots, (S_n, C_n), (S_{n+1}, \{(\lambda, \lambda)\})),$$

on

$$S_{n+1} = V \cup \{u \mid (u, v) \in C_i, \text{ o } (v, u) \in C_i, \text{ per a algun } 1 \leq i \leq n\}.$$

Obtenim $L_{in}(G) = L_{df}(G')$. Ja que S_{n+1} és finit, les dues gramàtiques tenen selecció del mateix tipus.

La inclusió $L_{df}(G') \subseteq L_{in}(G)$ és òbvia: G' té les mateixes produccions que G , més el context buit, que no genera res.

No obstant, aquesta última producció permet a G' de simular cada derivació a G en una manera de primera profunditat. Suposem que tenim una cadena activa de la forma $x_1 \underline{u} x_2 \underline{v} x_3$, (com és usual, hem subratllat l'últim context introduït), i suposem que volem ara usar un selector que implica u ni v . Prenem un cas extrem, i suposem que volem usar com a selector el primer símbol de

x_1 . Escrivim $x_1 = a_1 a_2 \dots a_k, a_i \in V, 1 \leq i \leq k$ i, per tant, la cadena, en el seu conjunt, és $a_1 a_2 \dots a_k \underline{u} x_2 \underline{v} x_3$. Usant la producció $(S_{n+1}, \{(\lambda, \lambda)\})$ obtenim

$$\begin{aligned} a_1 \dots a_{k-1} a_k \underline{u} x_2 \underline{v} x_3 &\Longrightarrow_{df} a_1 \dots a_{k-1} a_k \underline{\lambda} u \underline{\lambda} x_2 v x_3 \\ &\Longrightarrow_{df} a_1 \dots a_{k-1} \underline{\lambda} a_k \underline{\lambda} u x_2 v x_3 \\ &\Longrightarrow_{df} a_1 \dots a_{k-2} \underline{\lambda} a_{k-1} \underline{\lambda} a_k u x_2 v x_3 \Longrightarrow_{df} \\ &\dots \Longrightarrow_{df} a_1 \underline{\lambda} a_2 \underline{\lambda} a_3 \dots a_k u x_2 v x_3, \end{aligned}$$

i, en aquest moment, a_1 és activat per la presència de la recentment introduïda cadena buida.

Conseqüentment, també $L_{in}(G) \subseteq L_{df}(G')$ i, per tant, hem completat la demostració. $\square$

Si no usem el context buit, la situació canvia:

Lema 11.4.4 $ICC(FIN) - ICC_{df}(REG) \neq \emptyset$.

Demostració. Per a la gramàtica

$$G = (\{a, b, c\}, \{cc\}, (\{c\}, \{(a, b)\})),$$

tenim

$$L_{in}(G) = \{a^n c b^n a^m c b^m \mid n, m \geq 1\}.$$

No obstant, $L_{in}(G) \notin ICC_{df}(REG)$. Suposem el contrari, i prenem $G' = (\{a, b, c\}, B, (S_1, C_1), \dots, (S_p, C_p))$ tal que $L_{in}(G) = L_{df}(G')$. Pel fet que $B \subseteq L_{df}(G')$, les cadenes de B han de ser de la forma $a^i c b^i a^j c b^j, i, j \geq 0$. Si un

context és adjuntat a aquesta cadena, aleshores o bé el nombre d'ocurrències d' a i de b al voltant del primer c , o bé el nombre d'ocurrències d' a i de b al voltant del segon c són incrementades simultàniament. Però aquest no és el cas de les dues primeres ocurrències de a, b i les segones ocurrències de a, b al mateix temps.

Suposem que incrementem les primeres subcadenaes d' a i de b ; l'altre cas és simètric. A més a més, tenim

$$a^i c b^i a^j c b^j \implies_{df} a^{i_1} \underline{a}^k a^{i_2} c b^{i_3} \underline{b}^k b^{i_4} a^j c b^j,$$

per a algun $i_1 + i_2 = i_3 + i_4 = i, k \geq 1$. La derivació ha de continuar usant un selector que impliqui, com a mínim, un dels símbols subratllats a^k, b^k . Cap context $(u, v) \neq (\lambda, \lambda)$ pot ser usat de manera que contribueixi al sufix $a^j c b^j$ sense destruir la igualtat del nombre d'ocurrències d' a i de b al voltant del segon c . Conseqüentment, el llenguatge $L_{df}(G')$ conté cadenes de la forma $a^n c b^n a^m c b^m$ amb, com a mínim, un dels n, m més petit que $\max\{s \mid a^s c b^s a^t c b^t \in B \text{ o } a^t c b^t a^s c b^s \in B\}$. No obstant, $L_{in}(G)$ conté aquestes cadenes, per a un valor d' n i d' m arbitràriament gran. La qual cosa és una contradicció que mostra que la igualtat $L_{in}(G) = L_{df}(G')$ és impossible. $\square$

Lema 11.4.5 $ICC_{df}^\lambda(REG) \subseteq CS$.

Demostració. Mitjançant la senzilla simulació d'una gramàtica contextual treballant en el mode de profunditat primera, usant una gramàtica de longitud creixent. $\square$

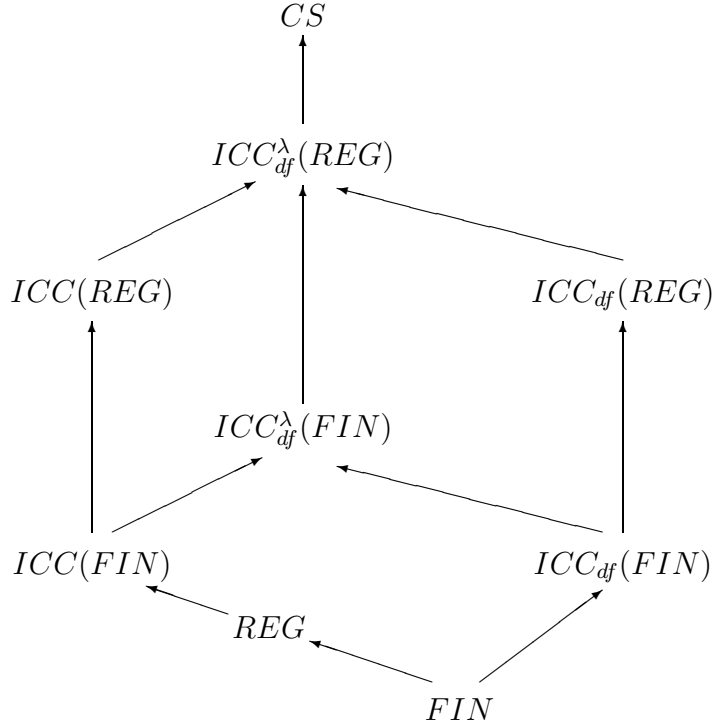
Lema 11.4.6 $ICC_{df}(REG) - ICC_{df}^{\lambda}(FIN) \neq \lambda$.

Demostració. Clarament, el llenguatge $L_{df}(G_4)$ considerat al començament d'aquesta secció no pot ser generat per una gramàtica amb selectors finits, fins i tot usant el context buit. $\square$

Sintetitzant els lemes previs, tenim

Teorema 11.4.1 Les relacions que mostra el diagrama de la Figura 11.4.1 es compleixen, on les fletxes indiquen inclusions pròpies. Cada dues famílies no relacionades en aquest diagrama són incomparables, amb l'excepció del parell que involucra la família REG , per la qual només sabem que $ICC_{df}(FIN) - REG \neq \emptyset$.

Figura 11.4.1 Relacions entre els llenguatges contextuais amb derivació de primera profunditat internes (amb FIN i REG selecció), entre elles i amb relació amb els llenguatges de la jerarquia de Chomsky.



La família $ICC_{df}(REG)$ és incomparable amb CF : el llenguatge $L_{in}(G)$ de la demostració del Lema 10.22 és lliure del context, mentre que els llenguatges $L_{df}(G_i)$, $i = 3, 4$, són no lliures del context. també $ICC_{df}^{\lambda}(FIN)$ conté llenguatges no lliures del context, ja que $ICC(FIN) - CF \neq \emptyset$. No sabem si hi ha o no llenguatges lliures del context que no pertanyin a $ICC_{df}^{\lambda}(FIN)$ o que no pertanyin a $ICC_{df}^{\lambda}(REG)$.

11.5 Gramàtiques contextuais amb derivació marcada

Aquesta variant de derivació en gramàtiques contextuais és, en alguns aspectes, similar a la derivació bloquejada: la derivació té lloc de forma usual i acaba amb l'adjunció d'un context d'un conjunt especial. Aquest conjunt especial pot ser vist, per altra banda, com simètric al conjunt axiomàtic (si tenim un conjunt de cadenes inicials, les quals són correctes – en el sentit de ben formades – per definició, és natural tenir també un conjunt de cadenes “finals”, que tanqui el procés de derivació). Per altra banda, es correspon amb la situació en què ens trobem en algunes circumstàncies, quan l'últim pas d'una “derivació” és distingit (podem recordar, per exemple, el cas dels sistemes d'acció simples, on l'última acció elemental és un conjunt donat d'accions finals).

Definició 11.5.1 Formalment, una gramàtica contextual *amb derivacions marcades* (“with marked derivations”) és una 6-tupla

$$G = (V, B, C_1, C_2, \varphi, \psi)$$

on

- (V, B, C_1, φ) és una gramàtica contextual usual,
- C_2 és un conjunt finit de contextos sobre V i
- $\psi : V^* \longrightarrow 2^{C_2}$.

Els contextos de C_2 són els terminals i ψ és la funció que selecciona els contextos finals.

Per a una gramàtica G com abans, denotem per G' la gramàtica usual subjacent (V, B, C_1, φ) i per $\Rightarrow_\alpha^\psi$, per a $\alpha \in \{ex, in\}$. En un pas de derivació, un context de C_2 és adjuntat (en el mode extern o intern, respectivament), tan aviat com és seleccionat per la funció ψ . Aleshores, el llenguatge generat per G , en el mode $\alpha \in \{ex, in\}$, és

$$L_\alpha = \{x \in V^* \mid w \Rightarrow_\alpha^* y \Rightarrow_\alpha^\psi x, \text{ per a } w \in B, y \in L_\alpha(G')\}.$$

Cal notar que un axioma $w \in B$ pertany a L_α si, i només si, $(\lambda, \lambda) \in C_2$ i $(\lambda, \lambda) \in \psi(w)$.

Denotem per $ECC_{mk}(F), ICC_{mk}(F)$ les famílies de llenguatges generades en el mode extern i intern, respectivament, per gramàtiques amb F selecció, usant derivacions marcades; per a les famílies corresponents a gramàtiques sense selecció escriurem EC_{mk}, IC_{mk} , respectivament, i, com és usual, no especificarem $F = ARB$.

Com era d'esperar, la possibilitat de parar la derivació en un moment determinat, és un tret molt poderós.

Teorema 11.5.1 Qualsevol llenguatge pertany a la família ECC_{mk} .

Demostració. Per a un llenguatge arbitrari $L \subseteq V^*$, definim la gramàtica

$$G = (V, \{\lambda\}, \{(\lambda, a) \mid a \in V\}, \{(\lambda, \lambda)\}, \varphi, \psi),$$

$$\varphi(x) = \{(\lambda, a) \mid a \in V\}, \text{ per a } x \in V^*,$$

$$\psi(x) = \{(\lambda, \lambda)\}, \text{ per a } x \in L.$$

La igualtat $L_{ex}(G) = L$ és òbvia (una derivació pot ser finalitzada si, i només si, produeix una cadena d' L). $\square$

Corol·lari 11.5.1 $ECC_{mk} - TC \neq \emptyset$.

Teorema 11.5.2 Es compleix

$$(i) \ EC_{mk} - ECC \neq \emptyset;$$

$$(ii) \ ECC_{mk}(REG) \subset LIN.$$

Demostració. La demostració completa i detallada d'aquest teorema es pot trobar a (Păun, 1996a). $\square$

Corol·lari 11.5.2 $ECC_{mk}(REG) \subset TC$.

Demostració. Aquest fet és una conseqüència de la inclusió $CF \subseteq TC$ (Lema 5.3.4 de la pàgina 203). $\square$

Per altra banda, tenim

Teorema 11.5.3 Es compleix

$$ECC_{mk}(CF) - TC \neq \emptyset,$$

$$IC_{mk} - TC \neq \emptyset.$$

Demostració. Per a la gramàtica

$$G = (\{a, b\}, \{b\}, \{(a, a), (\lambda, b), (\lambda, a)\}, \{(\lambda, \lambda)\}, \varphi, \psi),$$

$$\varphi(a^i b a^j) = \{(a, a), (\lambda, b)\}, \quad i, j \geq 0,$$

$$\varphi(a^i b a^j b a^k) = \{(\lambda, a)\}, \quad i, j, k \geq 0,$$

$$\psi(a^i b a^j b a^j) = \{(\lambda, \lambda)\}, \quad i, j \geq 1,$$

tenim

$$L_{ex}(G) = \{a^n b a^n b a^n \mid n \geq 1\}.$$

Aquest llenguatge no té la propietat d'IBS i, per tant, no pertany a TC (Lema 5.1.2 de la pàgina 185).

Considerem ara la gramàtica

$$G' = (\{a, b, c, d\}, \{abc\}, \{(a, bc)\}, \{(d, d)\}).$$

És fàcil de veure que el llenguatge $L_{in}(G)$ és igual a $L_{in}(G') \sqcup \{dd\}$, on $G' = (\{a, b, c\}, \{abc\}, \{(a, bc)\})$. En la demostració del Lema 6.11.1 de la pàgina 227 vam veure que aquest llenguatge no pertany a TC . $\square$

Una rèplica al Lema 5.3.2 de la pàgina 200 és també certa per a la família EC_{mk} . Denotem per LIN_2 la família de llenguatges lineals L tal que $Var_{LIN}(L) \leq 2$.

Teorema 11.5.4 $EC_{mk} \subset LIN_2$.

Demostració. Per a una gramàtica $G = (V, B, C_1, C_2)$ construïm la gramàtica lineal $G' = (\{S, X\}, V, S, \{S \rightarrow uXv \mid (u, v) \in C_2\} \cup \{X \rightarrow uXv \mid (u, v) \in C_1\} \cup \{X \rightarrow w \mid w \in B\})$. La igualtat $L_{ex}(G) = L(G')$ és òbvia i, per tant, $EC_{mk} \subseteq LIN_2$. La inclusió és pròpia: el llenguatge

$$L = \{a^n b^m c b^m a^n \mid n, m \geq 1\}$$

pot ser generat per la gramàtica lineal $G = (\{S, X\}, \{a, b, c\}, S, \{S \rightarrow aSa, S \rightarrow abXba, X \rightarrow bXb, X \rightarrow c\})$, però cap gramàtica contextual sense selecció pot generar aquest llenguatge (necessitem un context que introdueixi ocurrències d' a i un altre context que introdueixi ocurrències de b , ambdós usats un nombre arbitrari de vegades i, per tant, no són contextos finals; aquests contextos poden ser usats en qualsevol ordre i, per tant, obtenim cadenes paràsites). $\square$

Notes bibliogràfiques. Les derivacions més a l'esquerra van ser introduïdes a (Păun et al., 1994b) mentre que les derivacions prefix i més a l'esquerra maximals van ser introduïdes a (Păun et al., 1996b). Els resultats d'aquesta secció estan basats en aquests dos articles i en la informació afegida per Păun a (Păun, 1996a).

La derivació paral·lela és introduïda a (Păun et al., 1995a) i els resultats mostrats es basen en (Păun et al., 1995a) i en (Păun, 1996a). Les gramàtiques

de profunditat primera estan basades en (Martín-Vide et al., 1995b). Una variant més restrictiva de les derivacions de profunditat primera és introduïda a (Ilie, 1996a), on prova que els llenguatges que s'obtenen són analitzables en temps polinomial (una de les característiques que, com veurem, sembla ha de tenir qualsevol llenguatge que vulgui ser un model “adequat” pel llenguatge natural).

Finalment, les gramàtiques contextuais externes amb derivacions marcades són introduïdes a (Novotný, 1974) (i, posteriorment, investigades a (Păun, 1982a)) sota el nom de ν -gramàtiques contextuais.

Capítol 12

Gramàtiques contextuais regulades

En la Secció 3.4 vam considerar algunes restriccions introduïdes en la derivació de les gramàtiques de Chomsky lliures del context. Són les denominades gramàtiques (lliures del context) *amb mecanismes de control, o amb derivacions controlades, o regulades*.

La idea que s'amaga darrera de tots aquests mecanismes de control en el cas de les gramàtiques de Chomsky lliures del context és la següent: donada una gramàtica lliure del context, la restringim mitjançant el control de l'aplicació de les seves regles, de manera que algunes derivacions que són possibles en el procés usual de derivació de les gramàtiques lliures del context, són ara evitades. En aquest cas, el conjunt de cadenes generades per les restriccions és un subconjunt del llenguatge lliure del context generat de la manera

usual. Aquests subconjunts generats poden ser llenguatges no-lliures del context. Aquests mecanismes són més poderosos que les gramàtiques lliures del context.

Alguns exemples d'aquestes gramàtiques (lliures del context) regulades són: les gramàtiques (lliures del context) *matricials*, *programades*, *controlades*, *de context aleatori*, *condicionals*, etc.

En les variants de les gramàtiques contextuais discutides fins ara, l'ordre d'ús de les produccions d'una gramàtica contextual era lliure. No obstant, podem restringir aquesta llibertat, controlant la derivació en l'estil de la re-escriptura regulada per a les gramàtiques de Chomsky.

Així, doncs, en aquest capítol extendrem a les gramàtiques contextuais algunes restriccions ja investigades per a les derivacions en les gramàtiques lliures del context (veure la Secció 3.4 al respecte). La selecció definida per la funció de selecció pot ser considerada com un cas similar a la restricció en gramàtiques lliures del context condicionals (veure la Definició 3.4.4 de la pàgina 128), on un conjunt regular és associat a cada regla i una regla només pot ser aplicada a cadenes del llenguatge associat. Per tant, per als mecanismes de control definits més endavant, per a gramàtiques amb selecció, tenim un cert tipus de doble regulació de la derivació. Com era d'esperar, d'aquesta manera el poder de generació s'incrementa en termes generals.

Veurem els casos de: les gramàtiques contextuais matricials, les gramàtiques

contextuals programades i, finalment, les gramàtiques contextuals controlades regularment.

12.1 Gramàtiques contextuals matricials

Seguint la Definició 3.4.2 de la pàgina 123, aplicada al cas de les gramàtiques contextuals, definim:

Definició 12.1.1 i) Una *gramàtica contextual matricial* és una quintupla ordenada

$$G = (V, B, C, \varphi, M),$$

on

- (V, B, C, φ) és una gramàtica contextual usual i
- M és un conjunt finit de seqüències de les formes $[w, (u_1, v_1), \dots, (u_n, v_n)]$, $n \geq 0$, i $[(u_1, v_1), \dots, (u_n, v_n)]$, $n \geq 1$; on $w \in B$, i $(u_i, v_i) \in C$, $1 \leq i \leq n$. Aquesta seqüència s'anomena una *matriu*.

ii) Un pas de derivació en una gramàtica contextual matricial G consisteix en l'ús de tots els contextos d'una matriu, en l'ordre de la seva aparició; per a les matrius que comencen amb un axioma, aquest axioma és la cadena derivada. Per exemple, en el cas extern, un pas de la derivació $x \Rightarrow_{ex} y$ a G és, o bé una derivació $x \Rightarrow_{ex} u_1 x v_1 \Rightarrow_{ex} u_2 u_1 x v_1 v_2 \Rightarrow_{ex} \dots \Rightarrow_{ex}$

$u_n \dots u_2 u_1 x v_1 v_2 \dots v_n$, per a una matriu $[x, (u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)]$ a G i, per tant, amb $x \in B$; o bé una derivació com l'anterior on $[(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)]$ és una matriu a G . És clar que els contextos adjuntats també acompleixen la restricció imposada per la funció de selecció: $(u_1, v_1) \in \varphi(x)$, $(u_i, v_i) \in \varphi(u_{i-1} \dots u_1 x v_1 \dots v_{i-1})$, per a cada $i = 2, \dots, n$.

La derivació interna és definida de manera similar, implicant l'adjunció interna de contextos.

Quan usem la presentació modular de les gramàtiques, una gramàtica contextual matricial s'escriu de la forma

$$G = (V, B, (S_1, C_1), \dots, (S_n, C_n), M),$$

amb M i la derivació, definides com abans.

Denotem per MX la família de llenguatges generats per gramàtiques contextuais matricials corresponents a una família contextual usual X ; d'aquesta manera obtenim les famílies MEC , MIC , $MECC(REG)$, $MICC(FIN)$, etc.

Com veurem, en general els trets matricials no incrementen el poder de les gramàtiques contextuais externes però, en canvi, sí incrementen considerablement el poder de les gramàtiques contextuais internes.

Lema 12.1.1 Tot llenguatge de la família $MECC$ té la propietat d'*EBS*.

Demostració. La demostració és òbvia. Els contextos de l'última matriu usada en una derivació poden ser esborrats d'una cadena suficientment llarga. Obtenim una cadena del mateix llenguatge, que difereix en longitud respecte a la cadena considerada només en un valor acotat. $\square$

Teorema 12.1.1 Si F és una família de llenguatges que és tancada sota les operacions d'intersecció i de derivada per l'esquerra i per la dreta, aleshores $MECC(F) = ECC(F)$.

Demostració. Les inclusions $ECC(F) \subseteq MECC(F)$ són òbvies a partir de les definicions, per a tot F .

Considerem ara una gramàtica contextual matricial amb F selecció, $G = (V, B, C, \varphi, M)$. Assumim que $M = M_1 \cup M_2$, amb

$$\begin{aligned} M_1 &= \{[w_i, (u_{i,1}, v_{i,1}), \dots, (u_{i,n_i}, v_{i,n_i})] \mid w_i \in B, (u_{i,j}, v_{i,j}) \in C, \\ &\quad n_i \geq 0, 1 \leq j \leq n_i, 1 \leq i \leq m_1\}, \\ M_2 &= \{[(u_{i,1}, v_{i,1}), \dots, (u_{i,n_i}, v_{i,n_i})] \mid (u_{i,j}, v_{i,j}) \in C, n_i \geq 1, \\ &\quad 1 \leq j \leq n_i, 1 \leq i \leq m_2\}. \end{aligned}$$

Construïm la gramàtica contextual usual $G' = (V, B', C', \varphi')$ amb

$$\begin{aligned} B' &= \{u_{i,n_i} \dots u_{i,1} w_i v_{i,1} \dots v_{i,n_i} \mid [w_i, (u_{i,1}, v_{i,1}), \dots, (u_{i,n_i}, v_{i,n_i})] \in M_1, \\ &\quad u_{i,n_i} \dots u_{i,1} w_i v_{i,1} \dots v_{i,n_i} \in L_{ex}(G)\}, \\ C' &= \{(u_{i,n_i} \dots u_{i,1}, v_{i,1} \dots v_{i,n_i}) \mid [(u_{i,1}, v_{i,1}), \dots, (u_{i,n_i}, v_{i,n_i})] \in M_2\}, \end{aligned}$$

i la funció φ' és definida de la manera següent: per a $x \in V^*$ tenim

$$\begin{aligned} (u_{i,n_i} \dots u_{i,1}, v_{i,1} \dots v_{i,n_i}) &\in \varphi'(x) \text{ si, i només si,} \\ (u_{i,1}, v_{i,1}) &\in \varphi(x), (u_{i,j}, v_{i,j}) \in \varphi(u_{i,j-1} \dots u_{i,1} x v_{i,1} \dots v_{i,j-1}), \\ &\text{per a tot } j = 2, \dots, n_i, \end{aligned}$$

per a tots els contextos $(u_{i,n_i} \dots u_{i,1}, v_{i,1} \dots v_{i,n_i}) \in C'$.

A partir de la construcció veiem fàcilment que $L_{ex}(G) = L_{ex}(G')$.

Per altra banda, la gramàtica G' és amb F selecció quan G és amb F selecció: a partir de la definició de φ' , per a cada context $(u_{i,n_i} \dots u_{i,1}, v_{i,1} \dots v_{i,n_i}) \in C'$ tenim

$$\begin{aligned} \varphi'^{-1}((u_{i,n_i} \dots u_{i,1}, v_{i,1} \dots v_{i,n_i})) &= \\ \varphi^{-1}((u_{i,1}, v_{i,1})) \cap & \\ \partial_{u_{i,1}}^e(\partial_{v_{i,1}}^d(\varphi^{-1}(u_{i,2}, v_{i,2}))) \cap & \\ \partial_{u_{i,2}u_{i,1}}^e(\partial_{v_{i,1}v_{i,2}}^d(\varphi^{-1}(u_{i,3}, v_{i,3}))) \cap & \\ \dots \dots \dots & \\ \partial_{u_{i,n_i-1} \dots u_{i,1}}^e(\partial_{v_{i,1} \dots v_{i,n_i-1}}^d(\varphi^{-1}(u_{i,n_i}, v_{i,n_i}))). & \end{aligned}$$

Per a una família de llenguatges tancada sota les operacions de derivada per l'esquerra i per la dreta i sota l'operació d'intersecció, el tipus de selecció en les dues gramàtiques és el mateix. $\square$

Corol·lari 12.1.1 $MECC(F) = ECC(F)$, per a $F \in \{FIN, REG, CS, RE, ARB\}$.

Per a gramàtiques amb selecció lliure del context (CF) aquest resultat no es compleix.

Teorema 12.1.2 $MECC(CF) - ECC(CF) \neq \emptyset$.

Demostració. En la demostració del Lema 9.1.5 de la pàgina 306 vam veure que el llenguatge

$$L = \{a, b, c\}^* \cup \{da^n b^n c^n d \mid n \geq 1\}$$

no pertany a la família $ECC(CF)$. No obstant, podem generar aquest llenguatge mitjançant la següent gramàtica contextual matricial:

$$\begin{aligned} G &= (\{a, b, c, d\}, \{\lambda\}, \{(\lambda, a), (\lambda, b), (\lambda, c), (\lambda, d), (d, \lambda)\}, \varphi, M), \\ \varphi(x) &= \{(\lambda, a), (\lambda, b), (\lambda, c)\}, \quad x \in \{a, b, c\}^*, \\ \varphi(a^n b^n c^m) &= \{(d, \lambda)\}, \quad n, m \geq 1, \\ \varphi(da^n b^m c^m) &= \{(\lambda, d)\}, \quad n, m \geq 1, \\ M &= \{[\lambda], [(\lambda, a)], [(\lambda, b)], [(\lambda, c)], [(d, \lambda), (\lambda, d)]\}. \end{aligned}$$

(L'ús d'una matriu que introdueixi ocurrencies del símbol d és només possible per a cadenes de la forma $a^n b^n c^n$ i la cadena resultant és $da^n b^n c^n d$.) $\square$

El poder de les gramàtiques contextual internes matricials és significativament més gran que el de les gramàtiques contextuais usals.

Teorema 12.1.3 $MIC - TC \neq \emptyset$.

Demostració. Considerem la gramàtica

$$G = (\{a, b, c, d, e\}, \{\lambda\}, \{(\lambda, a), (\lambda, b), (\lambda, c), (\lambda, d), (\lambda, e)\}, M),$$

$$M = \{[\lambda], (\lambda, a), [(\lambda, b), (\lambda, c), (\lambda, d), (\lambda, e)]\}.$$

Tota derivació de la gramàtica G ha d'usar tots els contextos. Conseqüentment, $L_{in}(G) = \{x \in \{a, b, c, d, e\}^* \mid |x|_a = |x|_b = |x|_c = |x|_d = |x|_e\}$. Aquest llenguatge no té la propietat d'IBS (cap context pot ser esborrat d'una cadena $a^n b^n c^n d^n e^n$ amb n arbitràriament gran) i, per tant, no pertany a la família TC (veure el Lema 5.1.2 de la pàgina 185). $\square$

Mitjançant una senzilla construcció es pot provar la inclusió del pròxim re-orema; la seva inclusió pròpia se segueix del fet que els llenguatges en qualsevol família $MICC(F)$ tenen la propietat de BLI.

Teorema 12.1.4 $MICC(F) \subset CS$, per a tot $F \subseteq CS$.

Observació 12.1.1 Tancarem aquesta exposició mencionant el fet que el llenguatge M_6 , amb

$$M_6 = \{xcx \mid x \in \{a, b\}^*\} \text{ (reduplicació marcada),}$$

no pot ser generat per gramàtiques contextuais matricials de cap tipus.

Però M_3 , amb

$$M_3 = \{a^n b^n c^n \mid n \geq 1\} \text{ (concordança triple),}$$

és generat, en el mode intern, per la gramàtica contextual matricial (amb selecció finita)

$$G = (\{a, b, c\}, \{abc\}, \{(\lambda, a), (\lambda, b), (\lambda, c)\}, \varphi, M),$$

$$\varphi(a) = \{(\lambda, a)\},$$

$$\varphi(b) = \{(\lambda, b)\},$$

$$\varphi(c) = \{(\lambda, c)\},$$

$$M = \{[abc], [(\lambda, a), (\lambda, b), (\lambda, c)]\}$$

(l'ús d'una matriu força l'increment simultani del nombre d'ocurrències de tots els símbols a, b, c , mentre que la funció de selecció assegura l'ús dels contextos en els emplaçaments adequats).

12.2 Gramàtiques contextuais programades

Seguint la Definició 3.4.3 de la pàgina 126, aplicada al cas de les gramàtiques contextuais, definim:

Definició 12.2.1 i) Una *gramàtica contextual programada* és una quintupla ordenada

$$G = (V, B, C, \varphi, U),$$

on

- (V, B, C, φ) és una gramàtica contextual usual (que denotarem G') i

- U és un conjunt de parells de la forma $[w, (u, v,)]$, $[(u_1, v_1), (u_2, v_2)]$; on $w \in B$, i $(u, v), (u_1, v_1), (u_2, v_2)$ són contextos de C .

ii) Els parells d' U defineixen un graf sobre el conjunt $B \cup C$, que controla la seqüenciació dels contextos usats en les derivacions. Específicament, havent definit de la manera usual les relacions

$$\Longrightarrow_{\alpha}, \quad \alpha \in \{ex, in\},$$

per a una derivació

$$\delta : w \Longrightarrow_{\alpha} w_1 \Longrightarrow_{\alpha} \dots \Longrightarrow_{\alpha} w_m$$

a G' , que usa l'axioma $w \in B$ i els contextos $(u_1, v_1), \dots, (u_m, v_m)$ (el pas $w_{i-1} \Longrightarrow_{\alpha} w_i$ és realitzat usant el context (u_i, v_i)) direm que

$$q(\delta) = w, (u_1, v_1), \dots, (u_m, v_m)$$

és la *seqüència de control* associada a δ .

iii) Aleshores, el llenguatge $L_{\alpha}(G)$ generat per G en mode α és defineix com

$$L_{\alpha}(G) = \{x \in V^* \mid \text{existeix una derivació } \delta : w \Longrightarrow_{\alpha}^* x \text{ a } G'\}$$

amb $q(\delta) = w, (u_1, v_1), \dots, (u_m, v_m)$, tal que

$$[w, (u_1, v_1)] \in U, [(u_i, v_i), (u_{i+1}, v_{i+1})] \in U, 1 \leq i \leq m-1\}.$$

Cal notar algunes diferències essencials entre els mecanismes de control matricials i programats en les derivacions:

- i) en una gramàtica programada, tota part inicial d'una derivació produeix una cadena del llenguatge generat, mentre que la derivació en una matriu ha de ser completada;
- ii) per altra banda, en les gramàtiques matricials, després d'usar una matriu, podem usar-ne qualsevol altra; mentre que en les gramàtiques programades tots els parells de contextos usats consecutivament són prescrits.

La família de llenguatges generada per les gramàtiques contextuais programades, corresponent a una família contextual usual X , és denotada per PX (el mateix estil que per a les gramàtiques matricials).

Teorema 12.2.1 Es compleix que

- (i) $ECC(F) \subseteq PECC(F) \subseteq ECC$, per a tot F ; i, per tant, $ECC = PECC$.
- (ii) $PEC - EC \neq \emptyset$.

Demostració. (i) La primera inclusió és òbvia. Pel fet que tot llenguatge de $PECC$ té la propietat d'EBS, també se segueix la segona inclusió.

(ii) El llenguatge $a^+ \cup b^+$ no pertany a la família EC (Lema 5.1.3 de la pàgina 187), però, en canvi, pot ser generat per la gramàtica programada

$$G = (\{a, b\}, \{a, b\}, \{(\lambda, a), (\lambda, b)\}, U),$$

$$U = \{[a, (\lambda, a)], [(\lambda, a), (\lambda, a)], [b, (\lambda, b)], [(\lambda, b), (\lambda, b)]\}$$

(els contextos no poden ser mesclats, per restriccions imposades per U). $\square$

Teorema 12.2.2 Es compleix que

(i) $ICC(F) \subset PICC(F) \subset TC$, per a tot F .

(ii) $PIC - MICC \neq \emptyset$.

Demostració. (i) La primera inclusió és òbvia; si ens fixem en el punt (ii), és pròpia. Tot llenguatge de la família $PICC$ té la propietat d'IBS i, per tant, pertany a TC . També aquesta inclusió és pròpia: considerem el llenguatge

$$L = \{a^n b^n a^m b^m \mid n \geq 1, m = n \text{ o } m = n + 1\}.$$

És fàcil de veure que aquest llenguatge pertany a TC (per exemple, L té la propietat d'IBS). No obstant, aquest llenguatge no pot ser generat per una gramàtica contextual programada en el mode intern: per tal de derivar una cadena $a^n b^n a^m b^m$, amb qualsevol m comparat amb n , necessitem un context (a^i, b^i) , $i \geq 1$, associat amb un selector $a^j b^k$ per a algun $j, k \geq 0$. Ni la funció de selecció, ni la seqüència de restricció, poden imposar l'adjunció d'aquest context a una subcadena $a^j b^k$ en el prefix $a^n b^n$; o a una subcadena d'aquesta mateixa forma en el sufix $a^m b^m$ de la cadena derivada. Per tant, podem produir cadenes que no pertanyen a L .

(ii) El llenguatge $a^+ \cup \{a^n b^n \mid n \geq 1\}$ no pertany a $MICC$ (per la mateixa idea de la demostració del lema 6.1.2 de la pàgina 210: una matriu usada per

a derivar a^n pot ser també usada per a derivar $a^n b^n$, produint cadenes $a^m b^n$ amb $m > n$), però pot ser generat per la gramàtica contextual programada

$$G = (\{a, b\}, \{a, ab\}, \{(\lambda, a), (a, b)\}, U),$$

$$U = \{[a, (\lambda, a)], [(\lambda, a), (\lambda, a)], [ab, (a, b)], [(a, b), (a, b)]\}$$

$((\lambda, a))$ no pot ser usat per a derivar cadenes de la forma $a^n b^n$. $\square$

Corol·lari 12.2.1 Les famílies $MICC(F_1)$ i $PICC(F_2)$ són incomparables, per a tot F_1, F_2 .

Demostració. Pel Teorema 12.1.3 de la pàgina 477 sabem que $MIC - TC \neq \emptyset$, però pel punt (i) del lema anterior tenim que $PICC \subseteq TC$. Per tant, $MIC - PICC \neq \emptyset$. Inversament, $PIC - MICC \neq \emptyset$, pel punt (ii) del lema anterior. $\square$

Observació 12.2.1 Notem que el llenguatge M_3 , amb

$$M_3 = \{a^n b^n c^n \mid n \geq 1\} \text{ (concordança triple),}$$

no pot ser generat per una gramàtica contextual programada en el mode intern, sense tenir en compte el seu tipus de selecció.

12.3 Gramàtiques contextuais controlades

Seguint la Definició 3.4.1 de la pàgina 121, aplicada al cas de les gramàtiques contextuais, definim:

Definició 12.3.1 i) Una *gramàtica contextual controlada (regularment)* és una 6-tupla:

$$G = (V, B, C, \varphi, Lab, K),$$

on

- $G' = (V, B, C, \varphi)$ és una gramàtica contextual usual,
- Lab és un conjunt d'etiquetes associades amb elements de $B \cup C$ i K és un llenguatge regular sobre Lab .

ii) Denotem per e la funció d'etiquetament,

$$e : B \cup C \longrightarrow Lab.$$

Per a una derivació a G'

$$\delta : w \Longrightarrow_{\alpha} w_1 \Longrightarrow_{\alpha} \dots \Longrightarrow_{\alpha} w_m,$$

usant l'axioma $w \in B$ i els contextos $(u_1, v_1), \dots, (u_m, v_m)$ (el pas $w_{i-1} \Longrightarrow_{\alpha} w_i$ és realitzat usant el context (u_i, v_i)), denotem

$$e(\delta) = e(w)e((u_1, v_1)) \dots e((u_m, v_m)).$$

Anomenarem a aquesta cadena la *cadena de control* de la derivació δ (comparem-la amb la seqüència de control $q(\delta)$ considerada en la Secció 8.3).

iii) Aleshores, el llenguatge $L_\alpha(G)$ generat per G en el mode α es defineix com

$$L_\alpha(G) = \{x \in V^* \mid \text{existeix una derivació } \delta : w \Longrightarrow_\alpha^* x \\ \text{a } G', \text{ amb } e(\delta) \in K\}.$$

Òbviament, a partir d'aquesta definició observem que, de fet, només les cadenes de $K \cap e(A)e(C)^*$ poden controlar les derivacions de la gramàtica G .

Denotem per KX la família de llenguatges generada per gramàtiques contextuais controlades regularment associades a la família contextual usual X .

Aquestes gramàtiques són generalitzacions de les gramàtiques contextuais matricials i de les gramàtiques contextuais programades.

Teorema 12.3.1 Es compleix que

i) $MECC(F) \subset KECC(F)$, $PECC(F) \subset KECC(F)$, per a tot F i

ii) $MICC(F) \subset KICC(F)$, $PICC(F) \subset KICC(F)$, per a tot F .

Demostració. Si tenim una gramàtica matricial $G = (V, B, C, \varphi, M)$, construïm la gramàtica contextual regularment controlada $G' = (V, B, C, \varphi, Lab, e(M)^*)$, on Lab és un conjunt d'etiquetes associat de manera una-a-una (bijectivament) amb elements de $B \cup C$ i $e(M)$ és el conjunt de cadenes d'etiquetes associat de manera natural amb matrius d' M . Les igualtats $L_\alpha(G) = L_\alpha(G')$, $\alpha \in \{ex, in\}$ són òbvies.

Prenem ara una gramàtica programada $G = (V, B, C, \varphi, U)$. Sigui Lab un conjunt d'etiquetes associat de manera una-a-una (bijectivament) amb elements de $B \cup C$, i sigui $e(U)$ el conjunt de cadenes de longitud 2 formades per elements de Lab associats amb els parells de U . Considerem el llenguatge

$$K = pdret(e(B), e(U)),$$

on $pdret$ és l'operació de prolongació a la dreta considerada en la Definició 6.13.1 de la pàgina 230. El llenguatge K és regular; aquest resultat se segueix del fet que l'operació de $pdret$, respecte a llenguatges finits, preserva els llenguatges regulars; o bé podem obtenir-lo com a conseqüència del fet que el llenguatge $pdret(e(B), e(U))$ pot ser generat per una gramàtica contextual unilateral amb selecció regular en mode extern (pel Teorema 10.3.1 (4) de la pàgina 403 sabem que aquestes gramàtiques només generen llenguatges regulars).

Novament és fàcil de veure que $L_\alpha(G) = L_\alpha(G'), \alpha \in \{ex, in\}$, on $G' = (V, B, C, \varphi, Lab, K)$.

Hem obtingut les inclusions. La seva inclusió estricta per al cas intern se segueix del Corol·lari 12.2.1 del Teorema 12.2.1 de la pàgina 483. Pel Teorema 12.2.1 de la pàgina 481 sabem que $PECC \subseteq ECC$. En el pròxim teorema provarem que tots els llenguatges lineals estan a KEC . Pel fet que existeixen llenguatges lineals que no pertanyen a ECC , també les inclusions $PECC(F) \subset KECC(F)$ són pròpies. Pel Lema 12.1.1 de la pàgina 474 i el teorema que veurem posteriorment, també obtenim les inclusions estrictes $MECC(F) \subset$

$KECC(F)$, per a tot F . □

Teorema 12.3.2 $KEC = LIN$.

Demostració. Prenem una gramàtica contextual regularment controlada $G = (V, B, C, Lab, K)$. Assumim, sense perdre generalitat, que $K \subseteq e(B)e(C)^*$ (si no és així, aleshores considerem $K' = K \cap e(B)e(C)^*$ i $L_{ex}(G) = L_{ex}(G')$ per a $G' = (V, B, C, Lab, K')$). Això significa que per a tota cadena x de K existeix una derivació δ respecte a (V, B, C) tal que $x = e(\delta)$, que ens dona una cadena de $L_{ex}(G)$. Sigui $G_K = (N, Lab, S, P)$ una “gramàtica regular per l’esquerra” per al llenguatge K ; és a dir, una gramàtica lineal per l’esquerra que conté regles de les formes $X \rightarrow a$, $X \rightarrow Ya$, per a X, Y símbols no-terminals i a un símbol terminal. Construïm la gramàtica lineal $G' = (N, V, S, P')$ amb

$$P' = \{X \rightarrow uYv \mid X \rightarrow Ye((u, v)) \in P, (u, v) \in C\} \\ \cup \{X \rightarrow w \mid X \rightarrow e(w) \in P, w \in B\}.$$

La gramàtica G' afegeix els contextos als extrems de les cadenes i, per tant, les cadenes són construïdes començant pel centre. Degut al fet que la gramàtica G_K és lineal per l’esquerra, genera les cadenes de K començant per la dreta. D’acord amb aquesta cadena de control, obtenim una derivació a G' que construeix les cadenes a partir dels extrems, en l’ordre invers que una derivació a G . Conseqüentment, $L_{ex}(G) = L(G')$; és a dir, $KEC \subseteq LIN$.

Inversament, si tenim una gramàtica lineal $G = (N, V, S, P)$ construïm la gramàtica contextual controlada $G' = (V, B, C, Lab, K)$ on

$$B = \{w \in V^* \mid X \rightarrow w \in P\},$$

$$C = \{(u, v) \mid X \rightarrow uYv \in P\},$$

Lab és un conjunt d'etiquetes associat de manera una-a-una (bijectiva) a elements de $B \cup C$ i K és el llenguatge generat per la gramàtica $G'' = (N, Lab, S, P')$, amb

$$P' = \{X \rightarrow e(w) \mid B \rightarrow w \in P\}$$

$$\cup \{X \rightarrow Ye((u, v)) \mid X \rightarrow uYv \in P\}.$$

Com abans, es pot observar que $L_{ex}(G') = L(G)$. □

Teorema 12.3.3 Tot llenguatge de KIC és lletra-equivalent (veure la Definició 3.8.1 de la pàgina 153) a un llenguatge regular.

Demostració. Per a una gramàtica contextual controlada $G = (V, B, C, Lab, K)$ amb $K \subseteq e(B)e(C)^*$ (ja hem fet esment que no perdem cap generalitat en aquest cas), definim el morfisme $h : Lab^* \longrightarrow V^*$ by $h(e(w)) = w, w \in B$, $h(e((u, v))) = uv, (u, v) \in C$. Aleshores, òbviament, tenim $\Psi_V(L_{in}(G)) = \Psi_V(h(K))$. El llenguatge K és regular i, per tant, l'afirmació del teorema és certa. □

Corol·lari 12.3.1 Un llenguatge sobre un vocabulari d'una-lletra és regular si, i només si, pertany a la família KIC .

Demostració. Pel Teorema 12.3.2 de la pàg. 487 tot llenguatge d'una-lletra regular pertany a KEC ; per tant, en el cas de vocabularis d'una-lletra KEC i KIC coincideixen. Per tant, tot llenguatge regular d'una-lletra pertany a KIC . La implicació inversa se segueix del lema anterior. $\square$

12.4 Altres gramàtiques contextuais regulades

Per a gramàtiques contextuais amb selecció també podem definir gramàtiques contextuais matricials, programades, i regularment controlades amb comprovació de les aparicions: els contextos d'un conjunt especificat són omesos si no són aplicables.

Les gramàtiques contextual amb derivació controlada mitjançant altres mecanismes reguladors, també poden ser considerades (per exemple, gramàtiques de contextos aleatoris, on cada context no només pot ser usat depenent d'una cadena selectora sinó que també la presència o absència de certs símbols en la cadena derivada, en la seva totalitat). L'estudi d'aquestes gramàtiques resta encara per estudiar.

Notes bibliogràfiques. Els mecanismes de control matricials, programats, i de control regular van ser considerades en primer lloc per a gramàtiques contextuais externes sense selecció a (Păun, 1977a). A (Păun, 1982a) van ser investigades també per a gramàtiques contextual internes, però, novament,

considerant només el cas de gramàtiques sense selecció.

Pel cas de les gramàtiques de Chomsky lliures del context es pot trobar una acurada introducció als mecanismes de control de les derivacions a (Dassow et al., 1996b); mentre que, pel cas concret de les gramàtiques contextuais, es pot consultar (Ehrenfeucht et al., 1996a) (on només és considerat el cas de gramàtiques contextuais sense selecció) o bé (Păun, 1996a) (on sí es considera el cas de gramàtiques contextuais amb selecció).

Capítol 13

Altres tipus de gramàtiques contextuals

A més a més de les variants estudiades fins ara relacionades amb les gramàtiques contextuals, moltes altres variants han estat introduïdes. A continuació exposarem, molt breument, algunes d'aquestes altres variants (donarem només les definicions bàsiques, sense entrar en detalls).

13.1 Gramàtiques contextuals difuses

Ja vam veure a la Secció 2.2 que l'experiència amb llenguatges naturals mostra que qualsevol cadena selecciona un tipus de context i qualsevol context selecciona un tipus de cadena. Aquesta selecció es fa en base al nivell de gramaticalitat i la idea principal que guia aquest procés és el fet que, malgrat

la vaga natura de la propietat de bona formació, l'aproximació del conjunt difús de totes les cadenes ben formades d'un subconjunt concret de cadenes ben formades pot ser sempre millorat o reformat.

De fet, no hi cap evidència que l'estatus de les cadenes ben formades es reflecteixi de la millor manera mitjançant el concepte de *conjunt difús o borros* (“fuzzy set”) considerat per Zadeh, (Zadeh, 1965). Sembla ser que el concepte de *conjunt bast* (“rough set”), introduït per Pawlak, (Pawlak, 1982), és un millor candidat al respecte; aquesta conjectura, però, encara no ha estat verificada (al respecte es pot consultar (Miquel-Vergés, 1995a; Miquel-Vergés, 1996a)).

La teoria de conjunts difusos o borrosos (“fuzzy sets”), va ésser introduïda dins l'àmbit matemàtic, el 1965, per Lofti Zadeh a (Zadeh, 1965). L'origen d'aquesta concepció es troba en la crisi, iniciada el 1920, de la teoria clàssica. Segons aquesta, un element pertany a un conjunt donat només en el cas de posseir la propietat que defineix el conjunt en qüestió. En aquest cas, s'assigna a aquest element el valor 1 ; és a dir, “pertany al conjunt”. En cas contrari, es diu que el seu valor és 0 ; és a dir, queda exclòs d'ell. D'aquesta manera, la teoria clàssica només admet en la seva anàlisi elements clarament definits, desestimant aquells la pertinença dels quals sigui imprecisa, en no poder determinar quins valors posseeix.

No obstant, en la teoria de conjunts difusos, un element x subceptible de ser caracteritzat per una propietat, pertany al conjunt A definit per la mateixa,

encara que el seu valor sigui diferent de 0 o de 1; és a dir, encara que el seu grau de pertinença, denotat per $\mu_A(x)$, no pugui fixar-se amb precisió, en quedar aquest emmarcat en l'interval $[0,1]$.

El model de conjunts bastos (“rough sets”) és un nou tipus de model que té algunes semblances amb la teoria de conjunts borrosos, però es basa més en fets empírics, reals, no en “construccions adaptades”.

El principi anterior mostra que aquestes noves concepcions de la teoria de conjunts serveix de model matemàtic de les denominades “ciències humanes o humanitats”, ja que aquestes treballen amb elements que, quasi sempre, tenen una significació o una utilització poc definides. La Lingüística, en ajustar-se plenament a aquestes característiques, ofereix un àmbit adequat per a realitzar estudis matemàtics en l'àrea de la Semàntica.

L'aplicació dels conjunts difusos o borrosos a les gramàtiques contextuais es deu a I. Bucurescu i A. Pascu (Bucurescu i Pascu, 1987). Definirem aquí tan sols el cas de les gramàtiques contextual externes Q -difuses sense selecció. Els altres casos poden definir-se anàlogament, tenint en compte les corresponents definicions del tipus de gramàtiques contextuais considerades.

Definició 13.1.1 Una *gramàtica contextual externa Q -difusa (o Q -borrosa)* sense selecció és una quàdrupla

$$G = (V, B, V, Q)$$

on,

- V, B i C tenen la mateixa significació que en una gramàtica contextual externa sense selecció usual i
- B i C són conjunts Q -difusos (o Q borrosos) a V^* , sent Q una estructura algebraica.

Denotarem aquesta gramàtica per QG .

Quan Q és el reticle $Q = ([0, 1], \max, \min)$, la gramàtica QG s'anomena una *gramàtica contextual externa difusa*.

Donada una cadena $y \in V^*$ de la forma $y = u_n \dots u_1 x v_1 \dots v_n$, on $x \in B$ amb el grau de pertinença $\mu_B(x)$ i $(u_i, v_i) \in C$ amb el grau de pertinença $\mu_C((u_i, v_i))$, per a $1 \leq i \leq n$, aleshores una *generació contextual* associada a y és la cadena

$$(x, \mu_B(x)), ((u_1, v_1), \mu_C((u_1, v_1))), \dots, ((u_n, v_n), \mu_C((u_n, v_n))).$$

El *grau de generació contextual* d'una cadena $y \in V^*$, per a una gramàtica contextual externa Q -difusa QG , es defineix com

$$\mu_{QG}(y) = \bigoplus_{QG \in \mathcal{G}} \mu_B(x) \odot \bigodot_{i=1}^n \mu_C((u_i, v_i)), \quad n \geq 0$$

on $\mathcal{G}$ denota el conjunt de les generacions contextuais de la cadena $y \in V^*$ i el símbol $\odot$ i $\oplus$ denoten la regla additiva i la regla multiplicativa, respectivament, de l'estructura algebraica Q .

El llenguatge Q -difús $L(QG)$ generat per la gramàtica QG és un conjunt Q -difús a V^* caracteritzat per la funció de pertinença $\mu_{QG}(y)$ definida anteriorment.

La família de llenguatges generats per gramàtiques contextuais externes Q -difuses sense selecció es denota per QEC .

13.2 Gramàtiques contextuais d'infinits contextos

Quan interpretem les gramàtiques contextuais com a models generatius per a construccions del llenguatge natural, podria ser el cas que ens trobéssim en situacions en què tinguéssim disponibles un nombre arbitrari de contextos (possiblement infinits contextos).

Definició 13.2.1 Una *gramàtica contextual amb un nombre arbitrari de contextos (possiblement infinits)* és una quàdrupla

$$G = (V, B, C, \varphi)$$

on,

- $G = (V, B, C, \varphi)$ és una gramàtica contextual usual i
- C té amb un nombre de contextos arbitrari, possiblement infinit. No obstant, per tal de capturar la idea de “el més finit possible”, hem d'imposar algunes restriccions a la funció φ . Existeixen dues possibilitats bàsiques:

- (i) requerir que el conjunt $\{x \in V^* \mid \varphi(x) \neq \emptyset\}$ sigui finit i
- (ii) requerir que, per a cada $x \in V^*$, el conjunt $\varphi(x)$ sigui finit.

La primera variant sembla ser massa productiva: per a un llenguatge arbitrari $L \subseteq V^*$, considerem la gramàtica

$$G = (V \cup \{c\}, \{c\}, \{(\lambda, xc) \mid x \in L\}, \varphi),$$

$$\varphi(c) = \{(\lambda, xc)\}, x \in L.$$

Obtenim

$$L_{ex}(G) = \{c\} \cup \{c\}L\{c\},$$

$$L_{in}(G) = \{c\} \cup \{cx_1cx_2 \dots cx_kc \mid k \geq 1, x_i \in L, 1 \leq i \leq k\}.$$

D'aquesta manera podem generar llenguatges de complexitat arbitrària.

Les famílies obtingudes de llenguatges generats en el modes extern i intern es denoten per ECC_∞ i ICC_∞ , respectivament.

13.3 Gramàtiques contextuais basades en patrons

Un *patró*, en el sentit donat per (Angluin, 1980) (veure detalls i també referències a (Ehrenfeucht et al., 1995a)), es defineix de la següent manera: donat un vocabulari V i un conjunt finit $\Delta = \{X_1, \dots, X_n\}$, de *variables* (quan

$n = 1$, escrivim solament $\Delta = \{X\}$), qualsevol cadena de $(V \cup \Delta)^*$ s'anomena un *patró*. Cal notar que fins i tot les cadenes de V^* són considerades com a patrons.

Un patró identifica totes les cadenes “de la seva forma”; per exemple, $aaXbX$ descriu totes les cadenes que comencen amb dues ocurrències del símbol a i, després, continua amb dues cadenes idèntiques separades per una ocurrència de b . Si el vocabulari amb el qual treballem consisteix en els símbols a i b , aleshores la cadena repetida és qualsevol definida sobre $\{a, b\}$.

Formalment, donat un patró x , podem associar-li un llenguatge de la manera següent. Denotem per $H_{NE}(V \cup \Delta, V)$ el conjunt de tots els morfismes $h : (V \cup \Delta)^* \longrightarrow V^*$ que tenen les següents propietats:

$$\begin{aligned} h(a) &= a, \text{ per a tot } a \in V, \\ h(X) &\in V^+, \text{ per a tot } X \in \Delta. \end{aligned}$$

Aleshores, definim

$$L_{NE}(x) = \{h(x) \mid h \in H_{NE}(V \cup \Delta, V)\}$$

(el subíndex NE significa “no-esborrador”: una variable no pot ser reemplaçada per λ).

Direm que els elements de $L_{NE}(x)$ són *interpretacions* del patró x . Cal notar que, quan $x \in V^*$, tenim $L_{NE}(x) = \{x\}$.

Els patrons poden ser usats, d'una manera òbvia, com a mitjà per a identi-

ficar l'emplaçament on un context pot ser insertat, la qual cosa sembla ser productiva per a incloure certes construccions dels llenguatges naturals i de programació, on apareixen concordances a llarga distància (la cadena intermèdia pot ser identificada mitjançant una variable d'un patró). Aquest fet ens porta a la següent definició.

Definició 13.3.1 Una *gramàtica contextual amb patrons* és una quàdrupla

$$G = (V, \Delta, A, P),$$

on

- V i Δ són vocabularis disjunts,
- A és un subconjunt finit de V^* , i
- P és un subconjunt finit de $(V \cup \Delta)^+ \times (V^* \times V^*)$.

Els elements de Δ s'anomenen *variables*, V és el vocabulari *terminal*, A és el conjunt de cadenes inicials (*axiomes*) i P és el conjunt de *produccions*. En una producció $(x, (u, v))$, x és el selector patró i (u, v) és el context.

Podem considerar cinc tipus de relacions de derivació respecte a una gramàtica contextual amb patrons com l'anterior: per a $x, y \in V^*$, definim

$$x \Rightarrow_f y \quad \text{sii} \quad x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \text{ per a } x_1, x_3 \in V^*, x_2 \in L_{NE}(z), \\ \text{on } (z, (u, v)) \in P,$$

$$\begin{aligned}
x \Longrightarrow_{ml} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \text{ per a } x_1, x_3 \in V^*, x_2 \in L_{NE}(z), \\
& \text{on } (z, (u, v)) \in P, \text{ i no existeixen } x'_1, x'_2, x'_3 \text{ tal que} \\
& x = x'_1 x'_2 x'_3, |x'_1| \geq |x_1|, |x'_3| \geq |x_3|, \\
& \text{i } |x'_2| < |x_2|, x'_2 \in L_{NE}(z),
\end{aligned}$$

$$\begin{aligned}
x \Longrightarrow_{mg} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \text{ per a } x_1, x_3 \in V^*, x_2 \in L_{NE}(z), \\
& \text{on } (z, (u, v)) \in P, \text{ i no existeixen } x'_1, x'_2, x'_3 \text{ tal que} \\
& x = x'_1 x'_2 x'_3, |x'_1| \geq |x_1|, |x'_3| \geq |x_3|, \text{ i } |x'_2| < |x_2|, \\
& x'_2 \in L_{NE}(z'), \text{ per a algun } (z', (u', v')) \in P,
\end{aligned}$$

$$\begin{aligned}
x \Longrightarrow_{Ml} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \text{ per a } x_1, x_3 \in V^*, x_2 \in L_{NE}(z), \\
& \text{on } (z, (u, v)) \in P, \text{ i no existeix } x'_1, x'_2, x'_3 \text{ tal que} \\
& x = x'_1 x'_2 x'_3, |x'_1| \leq |x_1|, |x'_3| \leq |x_3|, \\
& \text{i } |x'_2| > |x_2|, x'_2 \in L_{NE}(z),
\end{aligned}$$

$$\begin{aligned}
x \Longrightarrow_{Mg} y \quad \text{sii} \quad & x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \text{ per a } x_1, x_3 \in V^*, x_2 \in L_{NE}(z), \\
& \text{on } (z, (u, v)) \in P, \text{ i no existeix } x'_1, x'_2, x'_3 \text{ tal que} \\
& x = x'_1 x'_2 x'_3, |x'_1| \leq |x_1|, |x'_3| \leq |x_3|, \text{ i } |x'_2| > |x_2|, \\
& x'_2 \in L_{NE}(z'), \text{ per a algun } (z', (u', v')) \in P.
\end{aligned}$$

La derivació $\Longrightarrow_f$ es diu que s'ha realitzat *en el mode lliure*, $\Longrightarrow_{ml}$ és una derivació *en el mode minimal local* (no existeix cap subcadena pròpia del selector usat que sigui una interpretació del mateix patró selector), i $\Longrightarrow_{Ml}$ és una derivació *en el mode maximal local* (cap supercadena pròpia del selec-

tor usat és una interpretació del mateix patró selector); $\Rightarrow_{mg}$ i $\Rightarrow_{Mg}$ són derivacions en el modes minimal i maximal *global*, respectivament (la minimalitat/maximalitat d'un selector es defineix respecte a tots els patrons selectors de la gramàtica).

Denotem per $P_{atrons}IC_\alpha$, $\alpha \in \{f, ml, mg, Ml, Mg\}$, les famílies de llenguatges $L_\alpha(G)$ com abans.

13.4 Gramàtiques contextuais de dos nivells

A la Secció 13 hem vist una possible idea per tal d'extendre el poder de les gramàtiques contextuais considerant un nombre infinit de contextos.

Investigarem ara una aproximació relacionada amb aquesta idea: els contextos que usem no són donats “a priori” sinó que també són produïts de manera contextual, mitjançant l'adjunció de contextos. Més específicament, ja que un context és un parell de cadenes i aquestes cadenes se suposa que creixen d'una manera interrelacionada, usem parells de contextos (per tant, quàdruples de cadenes) per tal de produir contextos. D'aquesta manera obtenim una gramàtica contextual amb dos nivells: un d'ells genera contextos i l'altre usa aquests contextos en la manera clàssica. Ambdós nivells usen funcions de selecció del tipus usual en les gramàtiques contextuais (aquestes gramàtiques contextuais de dos nivells també poden ser vistes com a un tipus de sistemes de gramàtiques, en el sentit de (Csuha-j-Varju et al., 1994), compost per dues gramàtiques que

cooperen d'acord amb el protocol descrit).

Una idea natural que se'ns apareix és la d'iterar l'operació i considerar tres nivells, quatre nivells, etc. La qüestió és saber quin grau de “naturalitat” té. Per tal de construir, d'una manera contextual, quàdruples, hem d'usar quatre contextos; és a dir, 8-tuples. Ara bé, és al nostre abast el controlar aquest procés, el qual involucra 8 adjuncions simultànies ? D'acord amb (Miller, 1956), la resposta és negativa. Per tant, malgrat des d'un punt de vista matemàtic són possibles gramàtiques contextuais de n -nivells, amb $n \geq 3$, no semblen ser lingüísticament adequats.

A més a més, només considerarem el cas de les gramàtiques contextuais amb dos nivells, amb selecció finita o regular, o sense selecció. Aquestes variants considerades en cadascun dels dos nivells; usant sempre la derivació interna. Aleshores

Definició 13.4.1 Una *gramàtica contextual de dos nivells* amb selecció és una construcció

$$\gamma = (V, B, C, D, \delta, \varphi),$$

on

- V és un vocabulari,
- $B \subseteq V^*$ és un llenguatge finit (d'axiomes),
- $C \subseteq V^* \times V^*$ és un conjunt finit (de contextos),

- $D \subseteq (V^* \times V^*) \times (V^* \times V^*)$ és un conjunt finit (de *contextos dobles*),
- $\delta : V^* \times V^* \longrightarrow 2^D$, i $\varphi : V^* \longrightarrow 2^{K(C,D,\delta)}$.

Definim $K(C, D, \delta)$ de la manera següent: definim la relació $\Longrightarrow$ a $V^* \times V^*$ mitjançant

$$\begin{aligned} (x, y) \Longrightarrow (x', y') \quad \text{sii} \quad & x = x_1 x_2 x_3, y = y_1 y_2 y_3, \\ & x' = x_1 u x_2 v x_3, y' = y_1 u' y_2 v' y_3, \\ & \text{for } ((u, v), (u', v')) \in \delta(x_2, y_2), \\ & x_1, x_2, x_3, y_1, y_2, y_3 \in V^*. \end{aligned}$$

En paraules, el context (x, y) és derivat dins de (x', y') mitjançant l'adjunció simultània d'un context (u, v) a x i d'un context (u', v') a y , sent seleccionats per la funció δ d'acord a les subcadenaes x_2, y_2 .

Prenem una gramàtica $\gamma = (V, B, C, D, \delta, \varphi)$. Per a cada $D' \subseteq D$ definim el llenguatge

$$\delta^{-1}(D') = \{x_1 \# x_2 \mid x_1, x_2 \in V^*, \delta(x_1, x_2) = D'\},$$

on $\#$ és un símbol que no pertany a V . Direm que δ és del tipus F , per a una família donada F de llenguatges, si $\delta^{-1}(D') \in F$ per a tot $D' \subseteq D$.

El tipus de φ és defineix com pel cas de les gramàtiques contextuais usuals, d'acord amb els llenguatges $\varphi^{-1}(C') = \{x \in V^* \mid \varphi(x) = C'\}$, per a $C' \subseteq K(C, D, \delta)$.

Una gramàtica contextual de dos nivells $\gamma = (V, B, C, D, \delta, \varphi)$ té (F_1, F_2) selecció, per a dues famílies donades de llenguatges F_1, F_2 , si φ és de tipus F_1 i δ és de tipus F_2 . Les corresponents famílies de llenguatges són denotades per $TLC(F_1, F_2)$ (de “two-level contextual language”).

13.5 *n*-gramàtiques contextuais

En les gramàtiques contextuais investigades fins ara els contextos eren entesos de la manera usual, com a parells de cadenes. Una natural generalització és tractar amb n -tuples de cadenes, per a un valor d' n donat arbitrari.

Definició 13.5.1 Sigui $n \geq 1$ un número natural. Una *n-gramàtica contextual* és una quàdrupla

$$G = (V, B, C, \varphi),$$

on

- V és un vocabulari,
- B és un llenguatge finit sobre V ,
- C és un conjunt finit d' n -tuples $(u_1, \dots, u_n)$, de cadenes sobre V i
- $\varphi : (V^*)^{n+1} \longrightarrow 2^C$.

Els element de B s'anomenen *axiomes*, els de C s'anomenen *n-contextos* i φ és la *funció de selecció*.

Per a $x, y \in V^*$ escrivim $x \implies y$ si, i només si, $x = x_1x_2 \dots x_{n+1}, y = x_1u_1x_2u_2 \dots x_nu_nx_{n+1}$ per a algun $x_1, x_2, \dots, x_{n+1} \in V^*$ i $(u_1, u_2, \dots, u_n) \in \varphi(x_1, x_2, \dots, x_{n+1})$. Com és usual, $\implies^*$ denota la clausura reflexiva i transitiva de la relació $\implies$. Aleshores, el llenguatge generat per G és

$$L(G) = \{x \in V^* \mid w \implies^* x, w \in B\}.$$

Quan la funció φ està definida per $\varphi(x_1, \dots, x_{n+1}) = C$, direm que G és una gramàtica *sense selecció*; la funció φ és omesa i la gramàtica s'escriu de la forma $G = (V, B, C)$.

Denotem per CC_n la família de llenguatges generats per n -gramàtiques contextuais, $n \geq 1$, i per C_n la família de llenguatges generats per n -gramàtiques contextuais sense selecció. A més a més, denotem per $CC_\infty = \bigcup_{n \geq 1} CC_n, C_\infty = \bigcup_{n \geq 1} C_n$. Quan només usem funcions de selecció computables, afegim el subíndex c , obtenint $CC_n^c, n \geq 1$ i CC_∞^c .

Observem que treballem com en el cas de les gramàtiques contextuais totals, separant el conjunt de la cadena en subcadenaes, d'acord amb les quals el context adjuntat és seleccionat. Conseqüentment, per a $n = 2$ obtenim una gramàtica contextual total; és a dir, $CC_2 = TC$ i $CC_1 = 1TC$. En el cas sense selecció, per a $n = 2$ obtenim gramàtiques contextuais internes sense selecció; és a dir, $C_2 = IC$ i $C_1 = 1IC$.

Notes bibliogràfiques. El concepte de conjunt difús o borros “fuzzy set” va ser introduït a (Zadeh, 1965) i el de conjunt bast “rough set” a (Pawlak, 1982). Una introducció a les seves possibles aplicacions lingüístiques es pot trobar a (Miquel-Vergés, 1995a; Miquel-Vergés, 1996a). L’aplicació dels conjunts difusos a les gramàtiques contextuais va ser introduïda a (Bucurescu i Pascu, 1987).

Les gramàtiques contextuais amb un nombre infinit de contextos van ser considerades a (Păun, 1988).

Les gramàtiques contextuais amb selectors patrons van ser introduïdes, i investigades superficialment, a (Marcus et al., 1996a); posteriorment van ser ampliades amb nous resultats a (Păun, 1996a).

Les gramàtiques contextuais que poden produir conjunts de contextos, i no pas conjunts de cadenes, van ser suggerides per Marcus en una introducció a (Păun i Rozenberg, 1996), en la variant sense selecció i sense usar immediatament els contextos per a generar un llenguatge. El gruix de resultats obtinguts relatiu a aquesta variant es poden trobar a (Martín-Vide et al., 1996b).

Les *n*-gramàtiques contextuais amb selecció van ser introduïdes a (Nguyen, 1980d). La variant sense selecció és investigada a (Păun, 1982a). Les *n*-gramàtiques contextuais amb selecció restringida van ser introduïdes a (Nguyen, 1980c) i investigades a (Nguyen i Păun, 1982a). Altres variants (i resultats) poden ser trobats a (Nguyen, 1981c), (Nguyen, 1981b) i (Nguyen i Păun,

1982a).

També es pot consultar (Păun i Rozenberg, 1996) i (Păun, 1996a), on es troba informació general al respecte (no totes les variants hi són reflectides, però); en el primer cas, molt resumida i en el segon cas, molt més detallada.

Capítol 14

Autòmats Contextuals

Una important activitat dins la teoria de llenguatges formals és la de generar cadenes o frases que siguin correctes respecte a una gramàtica donada, que és el que hem estat fent en els capítols precedents pel cas de les gramàtiques contextuais. Ara bé, és tan o més important, *analitzar* cadenes o frases donades respecte a la seva correctitud. Aquesta activitat pot ser duta a terme o bé construint *algorismes d'anàlisi* associats a classes de gramàtiques, o bé mitjançant l'ús d'*autòmats* equivalents a classes de gramàtiques. Hem fet esment a la Secció 3.6 dels autòmats equivalents a les famílies de les classes bàsiques de gramàtiques/llenguatges en la jerarquia de Chomsky. Podem intentar seguir una aproximació similar pel cas de les gramàtiques contextuais.

Una imatge general d'un autòmat contextual (*AC*) seria la d'un *autòmat de*

memòria linealment limitada, *LBA* (veure la Definició 3.6.2 de la pàgina 137). És a dir, una màquina de Turing que, en comptes de posseir una cinta potencialment infinita en la qual computar, aquesta cinta està restringida al tros que conté la cadena d'entrada i dues cel·les més que contenen els marcadors d'inici i final de cadena. Aquesta cinta és explorada per un capçal de lectura/escriptura, d'acord amb el corresponent estat d'una memòria finita.

Per tal de correspondre a “l'estil de treballar contextual”, aquest mecanisme hauria de ser capaç d'executar l'operació inversa a l'operació d'inserció d'un context (operació característica de les gramàtiques contextuais); és a dir, l'operació d'*esborrament d'un context*.

Fins ara, només s'han donat dues caracteritzacions dels llenguatges contextuais usant autòmats: la primera, donada per Gh. Păun a (Păun, 1982a), està basada en els denominats *autòmats de contracció* (“contraction automata”); i la més recent, donada per P. Jančar i al. a (Jančar et al., 1994), basada en els denominats *autòmats de reiniciació* (“restarting automata”). No obstant, en ambdós casos, només havien estat caracteritzades algunes variants de les gramàtiques/llenguatges contextuais.

Així, doncs, el nostre objectiu serà el de trobar un marc de treball més general, introduint una família d'*autòmats contextuais*. Existeixen diverses possibles variants d'autòmats contextuais: algunes caracteritzant famílies de llenguatges contextuais ja existents i, d'altres, produint noves famílies de llenguatges.

14.1 Autòmats contextuais vs. autòmats d'esborrament

Motivat per l'anàlisi del llenguatge (natural i també formal), P. Jančar i al. van introduir a (Jančar et al., 1996) diverses variants d'autòmats d'esborrament amb operació de reiniciació i també una nova classe de gramàtiques contextuais amb selecció regular. L'objectiu principal consisteix en comparar classes de llenguatges reconeguts per diverses classes d'autòmats d'esborrament amb operació de reiniciació amb classes de llenguatges generats per gramàtiques contextuais. Veurem que els llenguatges generats per gramàtiques contextuais amb funció de selecció regular poden ser caracteritzats mitjançant autòmats d'esborrament normals amb operació de reiniciació.

La motivació prové, parcialment, de l'àrea de l'anàlisi del llenguatge natural i s'assembla a la motivació de les gramàtiques contextuais (Marcus, 1969a). Una altra part de la motivació prové de l'àrea de les gramàtiques com a verificadors dels llenguatges naturals i dels llenguatges artificials (Jančar et al., 1994).

14.1.1 Definicions bàsiques

Les definicions són presentades informalment; els detalls tècnics formals poden ser afegits en la manera estàndard de la teoria d'autòmats.

Definició 14.1.1 Un *autòmat d'esborrament*, abrev. *D-autòmat* (de “deleting automaton”), és un autòmat de memòria linealment limitada el qual no pot reescriure les cel·les d'entrada, només pot esborrar-les (i.e. eliminar-les completament).

La definició original de (Jančar et al., 1993) difereix una mica de (Jančar et al., 1996) – en aquest últim cas, l'autòmat no té “finestreta d'espiera” que és, per altra banda, usada comunment en analitzadors per a llenguatges formals i naturals. El model general és, en aquest aspecte, massa ample. Sembla natural permetre al capçal de moure's només cap a la dreta (amb alguna espiera acotada) i introduir una operació de *reiniciació*, on reiniciació significa reprendre el treball des del començament (en una cadena d'entrada més curta). Per tal de simplificar la descripció, aquest autòmat és definit com un autòmat que usa instruccions més complexes. Una instrucció d'aquest autòmat realitza un seqüència (amb la longitud limitada per la mesura de l'espiera) d'operacions:

- moure's a la dreta,
- esborrar i, probablement,
- acabada amb una operació de reiniciació.

Un *autòmat d'esborrament amb operació de reiniciació M*, abrev. *DR-autòmat* (de “deleting automaton with restart operation”), amb espiera acotada, és un mecanisme amb una unitat de control d'estats finits i un capçal que es mou damunt d'una doble llista lineal d'ítems units. El primer ítem està reservat

per un símbol especial ϕ i, l'últim, per un altre símbol especial $\$$, i qualsevol altre ítem conté un símbol d'un vocabulari finit V (que no conté ni ϕ ni $\$$). El capçal té una “finestreta espiera” de longitud fixa k , per algun $k \geq 0$; això vol dir que, a més a més de l'ítem corresponent, M pot també explorar els següents k ítems veïns de la dreta (o, simplement, fins el final de la cadena quan la distància a $\$$ sigui menor que k).

D'una manera simple, aquest autòmat pot ser descrit de la següent manera: l'autòmat té una unitat de control finita, un capçal amb una “finestreta espiera” adjunta a una cinta, i treballa mitjançant certs cicles. En un cicle, mou el capçal d'esquerra a dreta a través de la cadena d'entrada de la cinta, en certs punts “amputa” alguns símbols de l'espiera, i, finalment, el cicle acaba mitjançant una “reiniciació” – i.e. reiniciant la unitat de control en el seu estat inicial i emplaçant el capçal en l'extrem esquerre de la cadena (més curta). La computació consisteix en una seqüència de cicles i finalitza parant en un estat d'acceptació o de rebuig. De fet, per tal de fer la mencionada “amputació” més natural, l'autòmat d'esborrament amb operació de reiniciació usa una doble llista d'ítems units (cel·les) en comptes de la cinta usual.

A (Jančar et al., 1994) es defienixen els *autòmats de reiniciació*, abrev. *R-autòmats* (de “restarting automata”), un tipus especial d'autòmats d'esborrament amb operació de reiniciació, els quals tenen l'operació de reiniciació estretament relacionada amb la d'esborrament, on immediatament després d'esborrar alguns símbols de l'espiera l'autòmat reinicia, i.e. cada instrucció

amb alguna operació d'esborrament acaba amb una reiniciació. A (Jančar et al., 1995a) són definits els anomenats *autòmats de reiniciació normals*, abrev. *norm- R -autòmats* (de “normal restarting automata”). Posteriorment, els mateixos investigadors van generalitzar aquesta propietat als *DR*-autòmats. Els *DR*-autòmats normals (*R*-autòmats normals, respectivament) poden, en un cicle, esborrar de la cadena corresponent com a màxim dues subcadenaes contigües de longitud acotada per la mesura de l'espiera (en el cas dels *R*-autòmats normals ambdues subcadenaes són esborrades en l'última instrucció acabada amb reiniciació).

Quan considerem un *DR*-autòmat determinista M , algunes vegades és convenient suposar-lo en la *forma cíclica forta*; veure (Jančar et al., 1994) per a *R*-autòmats. Això vol dir que les cadenaes de longitud menor que k , sent k la longitud de l'espiera, són immediatament (en el primer cicle) acceptades o rebutjades i que M realitza com a mínim dos cicles (com a mínim un d'ells amb reiniciació) per a qualsevol cadena que sigui més llarga. Per a un *DR*-autòmat no-determinista M , podem suposar-lo en la *forma cíclica dèbil* – veure (Jančar et al., 1994) pel cas dels *R*-autòmats –, i.e, qualsevol cadena de $L(M)$ de longitud com a mínim k només pot ser acceptada mitjançant la realització de com a mínim dos cicles. D'aquesta manera, un *R*-autòmat en la forma cíclica dèbil pot rebutjar immediatament una cadena “llarga”, mentre que un *DR*-autòmat en la forma cíclica forta només pot rebutjar immediatament cadenaes “curtes”. A més a més, qualsevol computació d'un *DR*-autòmat en la forma cíclica (forta) és finita (és a dir, acaba en un estat final).

Un DR -autòmat M s'anomena *DR-autòmat extern*, abrev. *ext-DR-autòmat*, si és en la forma cíclica dèbil i, a més a més, cada cicle de qualsevol computació d' M té la següent forma: la primera instrucció usada en el cicle pot amputar algun prefix de la cadena considerada; i, després, només pot seguir una seqüència d'instruccions d'operacions de moure's a la dreta, amb una excepció – l'última instrucció del cicle (que acaba amb l'operació de reiniciació) pot amputar algun sufix de la cadena explorada. Per tant, M només pot acceptar immediatament cadenes curtes, les cadenes més llargues són rebutjades o escurçades mitjançant l'esborrament de prefixos i/o sufixos curts (limitats per la mesura de l'espiera) de la cadena en un cicle. Òbviament l'*ext-DR*-autòmat és normal.

Definició 14.1.2 Una *configuració d'un DR-autòmat* M és una tripla (u, s, v) , on u és el contingut de la llista de treball des del marcador de l'extrem esquerre ϕ fins a la posició del capçal, s és l'estat actual i v és el contingut de la llista de treball des de l'ítem explorat fins al marcador de l'extrem dret $\$$. En la *configuració inicial* d'una *cadena d'entrada* w , la unitat de control està en un estat fix i inicial, s_0 , i el capçal està adjunt a aquell ítem que conté el marcador de l'extrem esquerre ϕ , i.e. $(\lambda, s_0, \phi w \$)$ (explorant també els primers k símbols de la cadena d'entrada w).

El capçal de l'autòmat pot realitzar operacions dels tres tipus següents:

- 1) *operació de moviment cap a la dreta*, abrev. MOV_D , que consisteix en moure's cap a l'ítem precedent a l'ítem explorat;

- 2) *operació d'esborrament*, abrev. *ESB*, que consisteix en esborrar de l'ítem que és explorat amb un posterior moviment cap a la dreta; i finalment,
- 3) *operació de reiniciació*, abrev. *RIN*, que consisteix en emplaçar el capçal en el primer ítem de la llista (que conté ϕ).

La *computació* d' M està controlada per un conjunt finit d'*instruccions* del tipus següent:

$$(s, au) \rightarrow (s', \omega).$$

El membre de l'esquerra d'una instrucció determina quan és aplicable – s significa l'estat actual de la unitat de control, a significa el símbol que és explorat pel capçal, i u significa el contingut de la finestreta espiera (sent u una cadena de longitud k o més petita si acaba amb \$). El membre dret d'una instrucció descriu l'activitat a realitzar, ω és una seqüència d'operacions MOV_D , ESB de longitud més petita o igual a la longitud de au , a vegades acabada amb l'operació de RIN . En les instruccions amb ω acabat amb RIN el nou estat s' és l'estat inicial d' M . Les operacions de la seqüència ω són realitzades d'esquerra a dreta.

Per exemple:

- Si $\omega = ESB\ MOV_D\ ESB$, aleshores,

- 1) M canvia l'estat actual per s' , esborra a i emplaça el capçal en el símbol precedent al símbol a esborrat (ESB),

- 2) mou el capçal cap a la dreta (MOV_D) i, finalment,
 - 3) esborra el símbol explorat i emplaça el capçal en el símbol precedent a l'ítem esborrat (ESB).
- Si $\omega = MOV_D ESB RIN$, aleshores,
 - 1) M canvia l'estat actual per s' ($= s_0$ degut a RIN), mou el capçal cap a la dreta al primer símbol d' u (MOV_D),
 - 2) esborra el símbol explorat i emplaça el capçal en el símbol precedent a l'ítem esborrat (ESB) i, finalment,
 - 3) emplaça el capçal en el marcador de l'extrem esquerre ϕ (RIN).

Els estats de la unitat de control finita estan dividits en dues classes: els *estats que no són finals* (existeix com a mínim una instrucció que és aplicable quan la unitat es troba en aquest estat) i els *estats finals* (qualsevol computació acaba per donar entrada a un estat d'aquest tipus). Els estats finals estan, a la vegada, dividits en *estats d'acceptació* i en *estats de rebuig*.

En general, un DR -autòmat és *no-determinista*, i.e. poden existir dues o més instruccions amb el mateix membre esquerre (s, au). Si no es dona aquest fet aleshores direm que l'autòmat és *determinista*.

Definició 14.1.3 Una cadena d'entrada w és *acceptada per un DR -autòmat M* si existeix una computació que comença en la configuració inicial de w (acotada pels marcadors ϕ i $\$$) de la llista i acaba en una *configuració final* on

la unitat de control és en un dels estats finals. $L(M)$ denota el llenguatge que consisteix en totes les cadenes acceptades per M i, en aquest cas, direm que M reconeix el llenguatge $L(M)$.

Definició 14.1.4 Qualsevol computació d'un DR -autòmat consisteix en certs *cicles*: en un cicle, l'autòmat usa instruccions amb operacions de moure's cap a la dreta i d'esborrar, fins arribar a un estat final o fins que una operació RIN és executada – en aquest cas la computació es reprèn en la configuració inicial.

D'acord amb la definició, es pot donar el cas que la llista de treball (cadena actual) no sigui modificada durant un cicle fins que l'operació RIN hagi estat realitzada. Aleshores, la computació es reprèn amb la mateixa configuració inicial. Però aquest fet pot ser evitat usant DR -autòmats del “tipus esborrador”, que definirem a continuació.

Donat un DR -autòmat M , podem suposar sempre que és del tipus *esborrador* – és a dir, en qualsevol cicle, o bé M esborra com a mínim un símbol, o bé accepta, o bé rebutja. És fàcil de comprovar si un DR -autòmat M és del tipus esborrador o no i, en cas que no ho sigui, construir un autòmat equivalent que sigui del tipus esborrador. Si M no és del tipus esborrador aleshores podem construir un autòmat M' equivalent del tipus esborrador de manera que per a cada estat d' M existeixi a M' un indicador afegit. Aquest indicador ens mostra si M ha esborrat com a mínim un símbol en el cicle actual. La configuració inicial està lliure d'indicadors. L'indicador en un estat és col·locat per cada instrucció que conté l'operació ESB en altres instruccions el valor de

l'indicador és conservat. Una instrucció acabada amb una operació de *RIN* pot ser executada només quan l'indicador està emplaçat en l'estat actual o bé quan en aquesta instrucció com a mínim una operació d'*ESB* precedeix a l'operació de *RIN*. En aquesta situació, l'autòmat M' modificat, no té més del doble d'estats que M . Qualsevol computació d'un *DR*-autòmat del tipus esborrador és finita (és a dir, acaba en un estat final).

14.1.2 Relació entre els autòmats d'esborrament amb operació de reiniciació i els autòmats contextuais

Els autòmats d'esborrament amb operació de reiniciació normals, o *DR*-autòmats normals, estan estretament relacionats amb algunes classes de gramàtiques contextuais. Una gramàtica contextual treballa en una direcció contrària – començant amb una cadena d'un conjunt finit base, i generant cadenes mitjançant l'adjunció d'un context o “parell de subcadena” en determinats llocs de la cadena actual. La inserció està controlada per una funció de selecció. En aquest cas, una classe natural de gramàtiques contextuais és introduïda quan la possibilitat d'una inserció pot ser decidida mitjançant un autòmat finit – són les anomenades gramàtiques contextuais amb selecció regular (veure la Definició 9.1.1 de la pàgina 302).

Les gramàtiques contextuais amb selecció regular són més generals que algunes classes de gramàtiques contextuais estudiades en la teoria de gramàtiques contextuais (c.f. (Păun, 1985a; Păun, 1994)). La classe de gramàtiques

contextuals de (Jančar et al., 1995b) que caracteritza la classe de llenguatges reconeguts per R -autòmats normals és també una subclasse d'aquestes gramàtiques. A més a més, per qualsevol gramàtica contextual G amb selecció regular podem construir un DR -autòmat normal M de manera que aquest M no només reconeix les mateixes cadenes, sinó que M també pot seguir qualsevol derivació de G (en l'ordre invers). En aquesta situació, DR -autòmats normals poden ser usats per a l'anàlisi de llenguatges generats mitjançant alguna classe de gramàtiques contextuals.

En primer lloc, veurem que qualsevol llenguatge regular pot ser reconegut mitjançant un DR -autòmat d'un tipus especial. Usarem aquest resultat en demostracions posteriors.

Lema 14.1.1 Per a qualsevol llenguatge regular R podem construir un DR -autòmat M que satisfaci les condicions següents:

- (i) $L(M) = R$,
- (ii) M executa com a màxim una instrucció que amputa com a màxim una subcadena contigua en cada cicle,
- (iii) M és en la forma cíclica dèbil,
- (iv) M rebutja tota cadena que no pertany a R en un cicle.

Demostració. Qualsevol llenguatge regular pot ser reconegut mitjançant algun autòmat finit (o d'estats finits) determinista (veure el Teorema 3.6.4 i

l'Observació 3.6.3 de la pàgina 142). Aquest autòmat finit determinista o bé accepta o bé rebutja la cadena d'entrada. A més a més, si afegim un marcador a dreta i a esquerra de la cadena d'entrada, podem reconèixer qualsevol llenguatge regular només modificant una mica l'autòmat finit determinista. Aquest autòmat té dos estats finals, un per l'acceptació i un altre pel rebuig. Durant la computació de qualsevol cadena, aquest autòmat entra en l'estat final només una vegada, quan està explorant el marcador de la dreta. Però, degut al lema de bombeig per a llenguatges regulars (veure el Teorema 3.8.6 de la pàgina 154), qualsevol cadena del llenguatge regular considerat que sigui més llarga que alguna constant, pot ser reduïda a una cadena més curta de manera que preservi la seva pertanyença a aquest llenguatge. A més a més, aquesta reducció pot ser feta a una distància limitada a partir del marcador de la dreta. Aquesta és la idea principal de la següent construcció.

Sigui R un llenguatge regular. Sigui A un autòmat finit determinista amb dos estats finals, com hem fet esment anteriorment. Construïm un DR -autòmat M que satisfaci les propietats (i), $\dots$, (iv).

L'autòmat M es mou a través de la cadena d'entrada simulant l'autòmat A i, simultàniament, explora el prefix de la part no llegida de la llista d'entrada en l'espiera.

L'autòmat M simula amputa la subcadena au de $xauay$ – el contingut de l'espiera – i es reinicia si l'autòmat A simulat pot continuar a $xauay$ de manera que A entri en el mateix estat s quan explori el primer i el segon símbol

a ; i després entri dins l'estat d'acceptació explorant el marcador de la dreta \$.

A més, M rebutja en un cicle qualsevol cadena que és rebutjada per l'autòmat simulat A i accepta en un cicle qualsevol cadena més curta que l'espiera, la qual és acceptada per l'autòmat simulat A .

A continuació hauríem de donar resposta a una qüestió: quina és la mesura de l'espiera de l'autòmat M ?

L'autòmat A té la garantia d'entrar en el mateix estat amb el mateix símbol explorat durant una computació de qualsevol subcadena de la cadena d'entrada, la longitud de la qual sigui més gran que $n = |K| * |V|$, on K és el conjunt d'estats d' A i V és el vocabulari d'entrada. És clar que la distància a una d'aquestes subcadenaes a partir del marcador de la dreta és com a màxim igual a n . Això ens dóna la mesura de l'espiera la qual és igual a $2 * n + 1$.

És fàcil de veure que aquesta construcció que hem descrit satisfà les propietats (i), ..., (iv). □

Qualsevol cadena generada per una gramàtica contextual amb selecció regular $G = (V, B, C, \mathcal{P}, \mathcal{I}, \mathcal{S}, \varphi)$ es deriva a partir d'una cadena més curta pertanyent al conjunt base (o axiomàtic) B . Voldríem construir un autòmat que reconegués el llenguatge generat per la gramàtica contextual G .

Per tal de reconèixer els elements de la base B (o axiomes), per a qualsevol DR -autòmat M denotem per $L_1(M)$, i.e. $L_1(M)$ és el conjunt de totes les

cadena acceptada per M en només un cicle,

$$L_1(M) = \{ u \mid u \in L(M) \wedge \neg \exists v \in L(M) : u \longrightarrow_M v \}$$

A més a més, l'autòmat simulat M hauria de satisfer la següent propietat:

$$\text{per a qualsevol cadena } x, y \quad x \Rightarrow_G^* y \quad \text{sii} \quad y \longrightarrow_M^* x.$$

En particular, un pas de la derivació en G hauria de correspondre a un cicle del DR -autòmat corresponent.

En aquestes condicions, necessitem que el DR -autòmat corresponent sigui normal i en forma cíclica dèbil. La normalitat se segueix per la manera de derivar de les gramàtiques contextuais – com a màxim dues subcadena contigües són insertades en una cadena determinada en un pas de la derivació de G . La forma cíclica dèbil assegura que només cadenes curtes siguin acceptades immediatament. Cadenes més llargues han de ser escurçades o rebutjades.

Teorema 14.1.1 Per a qualsevol $norm$ - DR -autòmat podem construir un $norm$ - DR -autòmat equivalent en la forma cíclica dèbil

Demostració. Sigui M un DR -autòmat. Construïrem el DR -autòmat N equivalent en la forma cíclica dèbil.

Sigui $R = L_1(M)$ el conjunt de totes les cadenes w que poden ser acceptades per M en un cicle. És fàcil de veure que R és un llenguatge regular i que $R \subseteq L(M)$. A més a més, degut al Lema 14.1.1 de la pàgina 518, podem

construir un DR -autòmat M_R que reconegui el llenguatge R i que compleixi els següents requeriments: M_R executi com a màxim una instrucció que amputi com a màxim una subcadena contigua en cada cicle, M_R és en la forma cíclica dèbil, i M_R rebutja tota cadena que no pertany a R immediatament en un cicle.

El DR -autòmat requerit N simula el DR -autòmat M_R i el DR -autòmat original M . Al començament de cada cicle el DR -autòmat N d'una manera no-determinista canvia a M i a M_R . Ara discutirem ambdós casos:

- N canvia a M al començament del cicle:
 - Si M redueix la cadena d'entrada i continua amb el següent cicle aleshores N treballa de la mateixa manera.
 - Si M accepta la cadena d'entrada aleshores N la rebutja.
 - Si M rebutja la cadena d'entrada aleshores N també la rebutja.
- N canvia a M_R al començament del cicle:
 - Si M_R redueix la cadena d'entrada i continua amb el següent cicle aleshores N treballa de la mateixa manera.
 - Si M_R accepta la cadena d'entrada aleshores N també l'accepta.
 - Si M_R rebutja la cadena d'entrada aleshores N també la rebutja.

El DR -autòmat N és en la forma cíclica dèbil pel fet que pot acceptar imme-

diatament en un cicle les cadenes d'entrada que puguin ser acceptades immediatament en un cicle pel DR -autòmat M_R .

El DR -autòmat N pot reduir la cadena d'entrada i continuar en el següent cicle si, i només si, o bé M o bé M_R ho poden fer també. Si w és acceptada per N , aleshores existeix una seqüència de cadenes $w_1, w_2, \dots, w_n$, per algun $n \geq 1$ tal que $w = w_1 \rightarrow_N w_2 \rightarrow_N w_3 \rightarrow_N \dots \rightarrow_N w_n$, on w_n és acceptada per N en un cicle. Aleshores, existeix algun i_0 tal que $w = w_1 \rightarrow_M w_2 \rightarrow_M w_3 \rightarrow_M \dots \rightarrow_M w_{i_0}$, $i_0 < k$ i $w_{i_0} \rightarrow_{M_R} w_{i_0+1}$ o $i_0 = k$. D'acord amb el Lema 14.1.1 $w_{i_0} \in L(M_R)$ i, per tant, $w_{i_0} \in L(M)$. Concloem que tota cadena acceptada per N és també acceptada per M .

A més a més, quan el DR -autòmat M assoleix l'estat d'acceptació, aleshores tenim la garantia de què w – la cadena continguda en la llista d'entrada – pertany a $R = L(M_R)$. També existeix alguna computació d'acceptació de M_R en w . Això implica que w ha de ser acceptada per alguna computació d' N .

L'autòmat N canvia entre dos autòmats al començament de cada cicle. Pel fet que ambdós autòmats siguin *norm-DR*-autòmats, l'autòmat N és també un *norm-DR*-autòmat. $\square$

Corol·lari 14.1.1 Per a qualsevol *norm-R*-autòmat existeix un *norm-R*-autòmat equivalent en la forma cíclica dèbil.

Demostració. Usant la construcció de la demostració del teorema anterior, si M és *norm-R*-autòmat, aleshores l'autòmat N construït és *norm-R*-autòmat

en la forma cíclica dèbil. □

La següent observació amb una demostració òbvia (i per tant omesa) mostra que podem “unir” dos cicles d’un *DR*-autòmat amb instruccions comunes usades en ambdós cicles.

Observació 14.1.1 Si $i_1, \dots, i_k, i, i_{k+1}, \dots, i_m$ és una seqüència d’instruccions usades en un cicle d’ M i $j_1, \dots, j_l, i, j_{l+1}, \dots, j_n$ és una seqüència d’instruccions usades en un altre cicle d’ M , aleshores M pot executar un cicle usant les instruccions $i_1, \dots, i_k, i, j_{l+1}, \dots, j_n$.

Aquesta observació serà usada en la demostració del següent teorema.

Teorema 14.1.2 Per qualsevol *norm-DR*-autòmat M en la forma cíclica dèbil podem construir una gramàtica contextual $G = (V, B, C, \mathcal{P}, \mathcal{I}, \mathcal{S}, f)$ amb selecció regular de manera que

- $B = L_1(M)$,
- $u \Rightarrow_G v \iff v \longrightarrow_M u$.

Demostració. Inspeccionarem qualsevol computació d’ M de la cadena d’entrada w durant un cicle on no rebutgi. M és normal – i.e. en cada cicle de qualsevol computació d’ M són esborrades com a màxim dues subcadenaes contigües de la cadena de manera que o bé dues subcadenaes (una d’elles pot ser buida) són esborrades en una instrucció, o bé M esborra dues subcadenaes

en dues instruccions diferents (una subcadena en cada). Poden succeir dues coses:

- (1) M accepta w .
- (2) $xuyvz \longrightarrow_M xyz$, on $uv \neq \lambda$.

Pel fet que M és en la forma cíclica dèbil la situació descrita a (1) pot succeir només per conjunts finits de cadenes. A més a més, $B = L_1(M)$.

El conjunt de contextos C es defineix com el conjunt de tots els parells (u, v) tals que $xuyvz \longrightarrow_M xyz$ per algun x, y, z .

La reducció a (2) pot ser donada per, com a màxim, dues instruccions. A més a més, per a cada instrucció i tal que i esborra dues subcadenaes u i v ($uv \neq \lambda$) podem definir els següents conjunts:

- $P_i^{u,v}$ és el conjunt de tots els prefixos x tal que $xuyvz \longrightarrow_M xyz$ per algun y, z usant la instrucció i .
- $I_i^{u,v}$ és el conjunt de tots els infixos y tal que $xuyvz \longrightarrow_M xyz$ per algun x, z usant la instrucció i .
- $S_i^{u,v}$ és el conjunt de tots els sufixos z tal que $xuyvz \longrightarrow_M xyz$ per algun x, y usant la instrucció i .

Similarment, per a qualsvol parell d'instruccions i, j tal que i esborra una

simple subcadena no-buida u i j esborra una simple subcadena no-buida v podem definir els següents conjunts:

- $P_{i,j}^{u,v}$ és el conjunt de tots els prefixos x tal que $xuyvz \rightarrow_M xyz$ per algun y, z usant les instruccions i, j .
- $I_{i,j}^{u,v}$ és el conjunt de tots els infixos y tal que $xuyvz \rightarrow_M xyz$ per algun x, z usant les instruccions i, j .
- $S_{i,j}^{u,v}$ és el conjunt de tots els sufixos z tal que $xuyvz \rightarrow_M xyz$ per algun x, y usant les instruccions i, j .

És fàcil de veure que $P_i^{u,v}$, $I_i^{u,v}$, $S_i^{u,v}$, $P_{i,j}^{u,v}$, $I_{i,j}^{u,v}$, $S_{i,j}^{u,v}$ són llenguatges regulars i, aleshores, definirem

$$\mathcal{P} = \{P_i^{u,v} \mid i \text{ és una instrucció que esborra dues subcadenaes} \} \\ \cup \{P_{i,j}^{u,v} \mid i, j \text{ són instruccions cadascuna de les quals esborra una subcadena} \}.$$

Els conjunts I i S poden ser definits de manera anàloga.

La funció de selecció f és definida de la següent manera: $(u, v) \in f(P, I, S)$ si, i només si, $P = P_i^{u,v}$, $I = I_i^{u,v}$, $S = S_i^{u,v}$ o $P = P_{i,j}^{u,v}$, $I = I_{i,j}^{u,v}$, $S = S_{i,j}^{u,v}$ per algunes instruccions i, j .

Ens falta provar que $xyz \Rightarrow_G xuyvz$ si, i només si, $xuyvz \rightarrow_M xyz$. Veurem que aquesta equivalència té lloc quan els contextos són amputats per dues instruccions. La demostració seria semblant a la del cas quan ambdós

contextos són amputats per una instrucció i l'obviarem.

Primer, suposem que $xuyvz \longrightarrow_M xyz$ succeeix i els contextos u i v són amputats mitjançant instruccions i i j , respectivament. Això implica que $x \in P_{i,j}^{u,v}$, $y \in I_{i,j}^{u,v}$, i $z \in S_{i,j}^{u,v}$. D'acord amb la definició de la funció de selecció, el parell de contextos (u, v) pertany a $f(P_{i,j}^{u,v}, I_{i,j}^{u,v}, S_{i,j}^{u,v})$. A més a més, $xyz \Rightarrow_G xuyvz$.

La demostració de la implicació inversa és una mica més complicada. Assumim que $xyz \Rightarrow_G xuyvz$. Degut a la definició de la funció de selecció, la següent afirmació és certa per algunes instruccions i i j del DR -autòmat M : $x \in P_{i,j}^{u,v}$, $y \in I_{i,j}^{u,v}$, $z \in S_{i,j}^{u,v}$ i $(u, v) \in f(P_{i,j}^{u,v}, I_{i,j}^{u,v}, S_{i,j}^{u,v})$. Això implica que

$$(1) \quad xuy_1vz_1 \longrightarrow_M xy_1z_1,$$

$$(2) \quad x_2uyvz_2 \longrightarrow_M x_2yz_2,$$

$$(3) \quad x_3uy_3vz \longrightarrow_M x_3y_3z$$

per algun $y_1, z_1, x_2, z_2, x_3, y_3$. Però nosaltres volem provar que $xuyvz \longrightarrow_M xyz$. A més a més, necessitem comprovar que podem unir aquests tres cicles en un de sol. Els cicles (1) i (2) usen les mateixes instruccions per amputar el context u . A més, degut al fet de l'Observació 14.1.1 de la pàgina 524, podem unir els cicles (1) i (2). Obtenim el següent cicle:

$$(4) \quad xuyvz_2 \longrightarrow_M xyz_2.$$

Els cicles (3) i (4) usen la mateixa instrucció j quan amputen el context v . A més, podem usar per segona vegada l'Observació 14.1.1 per tal d'unir els cicles (3) i (4). Obtenim el cicle requerit: $xuyvz \longrightarrow_M xyz$. $\square$

Teorema 14.1.3 Donat un *norm-DR*-autòmat podem construir una gramàtica contextual amb selecció regular equivalent.

Demostració. Aquesta afirmació és una conseqüència directa del Teorema 14.1.1 i del Teorema 14.1.2. $\square$

A continuació veurem la direcció inversa.

Teorema 14.1.4 Donada una gramàtica contextual $G = (V, B, C, \mathcal{P}, \mathcal{I}, \mathcal{S}, f)$ amb selecció regular podem construir un *norm-DR*-autòmat M en la forma cíclica dèbil de manera que

- $B = L_1(M)$,
- $u \longrightarrow_M v \iff v \Rightarrow_G u$.

Demostració. Sigui $G = (V, B, C, \mathcal{P}, \mathcal{I}, \mathcal{S}, f)$ una gramàtica contextual amb selecció regular. Qualsevol llenguatge regular pertanyent a $\mathcal{P} \cup \mathcal{I} \cup \mathcal{S}$ pot ser acceptat per un autòmat finit. Sigui k la longitud màxima de qualsevol cadena de B . Aleshores $(k + 1)$ serà la mesura de l'espiera del *norm-DR*-autòmat M construït. Durant qualsevol dels seus cicles, M treballa de la següent manera:

Al començament de cada cicle, M verifica si observa el marcador de la dreta (en la seva espiera) i si la cadena d'entrada pertany a B i, si és així, aleshores M accepta. En qualsevol altre cas M selecciona un llenguatge regular $P \in \mathcal{P}$ com a prefix, $I \in \mathcal{I}$ com a infix i $S \in \mathcal{S}$ com a sufix, i un parell de contextos $(u, v) \in C$ tal que $(u, v) \in f(P, I, S)$. Aleshores M es mou a través de la cadena d'entrada tot simulant un autòmat finit A_P que reconeix un llenguatge regular P . Si l'autòmat simulat A_P entra en un estat d'acceptació, aleshores M , o bé continua la simulació d' A_P , o bé tracta d'amputar el context u . Si u pot ser amputat aleshores M el marca i comença la simulació d' A_I – l'autòmat finit que reconeix el llenguatge I – immediatament després d' u . De la mateixa manera, M tracta d'amputar el segon context v i començar la simulació d' A_S – l'autòmat finit que reconeix el llenguatge S . La cadena d'entrada és rebutjada o bé si un dels autòmats finits simulat no pot continuar i està en un estat de no-acceptació, o bé si A_P o A_I és en un estat d'acceptació però u o v , respectivament, no poden ser amputades.

Sembla clar que la cadena d'entrada w pot ser acceptada en un cicle si, i només si, pertany a B . A més a més, w pot ser reduïda a la cadena més curta de la forma xyz si, i només si, $w = xuyvz$, on $x \in P \in \mathcal{P}$, $y \in I \in \mathcal{I}$, i $z \in S \in \mathcal{S}$ i $(u, v) \in f(P, I, S)$.

M amputa com a màxim dues subcadenaes d'una longitud limitada a partir de la cadena d'entrada. A més, M és un *norm-DR*-autòmat. tenim també que M és en la forma cíclica dèbil ja que la cadena d'entrada w és acceptada en

un cicle si, i només si, $w \in B$ i la mesura de l'espiera d' M és més gran que la longitud de qualsevol cadena del conjunt base B . $\square$

14.2 Autòmats contextuais vs. autòmats de contracció

S.H. von Solms va definir a (von Solms, 1975) un nou autòmat, denominat *autòmat de contracció de complexitat* (n, k) , $b \geq 0$, $k \geq 1$, de manera que diverses classes de llenguatges formals conegudes, que fins aquell moment només havien estat caracteritzades mitjançant gramàtiques, podien ser també caracteritzades mitjançant aquest nou autòmat.

Gh. Păun presenta en l'últim capítol de (Păun, 1982a) alguns resultats sobre autòmats contextuais similar als *autòmats de contracció* mencionats anteriorment. No obstant, només dues classes de llenguatges contextuais són caracteritzats: la classe de llenguatges contextuais simples, caracteritzada mitjançant *autòmats contextuais simples*, i la classe de llenguatges ν -contextuais programats (gramàtiques contextuais externes amb derivacions marcades), caracteritzada mitjançant *autòmats contextuais (dreta, esquerra)-constants*.

Desafortunadament, la definició de nous tipus d'autòmats que caracteritzin altres tipus de llenguatges contextuais no és encara considerada.

14.2.1 Definicions bàsiques

A l'igual que en la secció anterior, les definicions són presentades informalment; els detalls tècnics formals poden ser afegits de manera estàndard de la teoria d'autòmats.

Definició 14.2.1 Un *autòmat de contracció de complexitat (n,k)* , abrev. (n,k) -*C*-autòmat, (de “contraction automaton of complexity (n,k) ”) és un autòmat de memòria linealment limitada amb algunes restriccions en les seves capacitats de lectura/escriptura. Considerem un autòmat de memòria linealment limitada – veure la Definició 3.6.2 de la pàgina 137 – $M = (K, V, \Gamma, \delta, s_0, F)$ amb les següents restriccions i canvis:

- (i) $F = \{s_f\}$,
- (ii) K conté un estat distingit s_c , i
- (iii) Γ conté dos símbols distingits, el símbol estrella $*$ i el símbol final Ω .

Direm que dos símbols de la cinta A i B són *símbols veïns* si estan separats per $t \geq 0$ símbols estrella. Sigui $n \geq 0$ i $k \geq 1$ dues constants.

Suposem que M sempre comença en un estat inicial s_0 explorant el marcadore de l'extrem dret $\$$. A partir d'aquí M es mou cap a la dreta, en modus de només lectura, fins a trobar el marcadore de l'extrem esquerra ϕ . En cada estadi durant un moviment de només-lectura cap a l'esquerra, l'estat d' M depèn

només de $\$$ i d'un subconjunt E de $(\Gamma - V)$, els símbols del qual apareixen entre el capçal de lectura i el marcador de la dreta $\$,$ i amb $|E| \leq n$ ($|E|$ és el nombre d'elements d' E). En aquesta situació, quan M arriba al marcador de l'extrem esquerre ϕ , és sabedor de $\$$ i d' $m \leq n$ símbols diferents de $(\Gamma - V)$ que apareixen en la seva cinta.

Quan arriba a ϕ , M canvia d'estat i comença a moure's cap a la dreta. Durant un moviment cap a la dreta, M pot realitzar $p \leq k$ reemplaçaments. Bàsicament M pot relitzar només dos *tipus* de reemplaçaments:

- una *operació d'escriptura* que reemplaça un simple símbol de la cinta per un altre símbol de la cinta,
- o una *operació estrella* que reemplaça dos símbols veïns, el de l'esquerra pel símbol estrella, i el de la dreta per un altre símbol de la cinta.

El símbol estrella no pot ser mai reemplaçat per cap altre símbol. Durant un moviment cap a la dreta, M pot, sota aquestes condicions, realitzar i operacions d'escriptura i j operacions estrella, amb $i + j \leq k$.

Com ja hem dit, M canvia d'estat i de direcció quan ϕ és assolit. El nou estat $s' \in K$ depèn de l'estat $s \in K$ en el qual ϕ és assolit, i, s depèn, com hem mencionat prèviament, de $\$$ i de $m \leq n$ símbols diferents de $(\Gamma - V)$ que apareixen en la seva cinta. Aquest nou estat s' determina un *element de reemplaçament*, que a la vegada determina

- (i) el nombre total de reemplaçaments p , amb $p \leq k$, que poden ser realitzats durant el següent moviment cap a la dreta;
- (ii) els símbols i els símbols veïns que poden ser reemplaçats durant aquests p reemplaçaments; i
- (iii) els símbols que poden ser usats com a reemplaçaments en les operacions d'escriptura i d'estrella.

Després de l'últim reemplaçament, i.e., el reemplaçament p , M entra en un estat especial, l'*estat còpia* s_c . En aquest estat M només pot copiar símbols i moure's cap a la dreta. Si \$ és assolit en l'estat s_c , M entra al seu estat inicial s_0 , i comença un altre moviment de lectura cap a l'esquerra. M entra al seu estat final s_f i accepta, si el símbol final Ω , juntament amb $r \geq 0$ símbols estrella, són els únics símbols de la cinta.

Un autòmat, com l'anteriorment descrit, s'anomena un autòmat de contracció de complexitat (n, k) , abrev. (n, k) - C -autòmat (de “contraction automaton of complexity (n, k) ”), i és representat mitjançant

$$M = (K, V, \Gamma, \delta, s_0, s_c, s_f, \Omega, *).$$

Formalment, un (n, k) - C -autòmat, $n \geq 0, k \geq 1$, és una 9-tupla $M = (K, V, \Gamma, \delta, s_0, s_c, s_f, \Omega, *)$ amb les següents característiques:

- (i) $K = K_E \cup K_D \cup K_C \cup \{s_c, s_f\}$ és un conjunt finit no buit d'estats. K_L, K_R, K_C i $\{s_c, s_f\}$ són mútuament disjunts (en aquesta situació E, D, C a K_E, K_D, K_C correspon a “esquerre”, “dreta”, i “còpia”, respectivament);
- (ii) Γ és un conjunt finit, no buit de símbols de la cinta;
- (iii) V , un subconjunt de Γ , és el conjunt de símbols d'entrada. V conté dos símbols especials, ϕ i $\$$, els marcadors dels extrems esquerra i dret, respectivament;
- (iv) δ és una funció, la *funció pròxim moviment*, de $K \times \Gamma$ a subconjunts finits de $(K \times \Gamma \times \{D, E\})$, D per “dreta” i E per “esquerra” (no considerarem formalment com treballa la funció δ ; el nostre objectiu és només contemplar algunes idees bàsiques que puguin ser aplicades als autòmats contextuais);
- (v) $s_0 \in K, s_c \in K$ i $s_f \in K$ són l'estat inicial, l'estat còpia, i l'estat final, respectivament;
- (vi) $\Omega \in (\Gamma - V)$ és el símbol final;
- (v) $*$ $\in (\Gamma - V)$ és el símbol estrella.

Definició 14.2.2 Una *configuració* d'un (n, k) - C -autòmat $M = (K, V, \Gamma, \delta, s_0, s_c, s_f, \Omega, *)$ és un parell

$$(s, \phi w_1 \underline{A} w_2 \$), \quad s \in K, \quad w_1, w_2 \in \Gamma^*, \quad A \in \Gamma$$

La configuració significa que M es troba en un estat s , explorant el símbol de la cinta A , amb $\$w_1$ a l'esquerra i $w_2\$$ a la dreta d' A .

Definició 14.2.3 El llenguatge acceptat per un (n, k) - C -autòmat M és el conjunt

$$L(M) = \{w \in (V - \{\$, \$\})^* \mid (s_0, \$w\$) \vdash_M^* (s_f, \$ *^i \Omega *^j \$)\},$$

on $i + j + 1 = |w|$ i $*^t = * * \dots * (t \text{ vegades})$.

Definició 14.2.4 Un *cicle* d'un (n, k) - C -autòmat M és un moviment des del seu marcador de l'extrem dret $\$$ fins al marcador de l'extrem esquerre $\$$, i retornar altra vegada a $\$$. Sense perdre cap tipus de generalitat podem assumir que, durant un cicle, M canvia la seva cinta en com a mínim una posició.

14.2.2 Relació entre els (n, k) - C -autòmat i els autòmats contextuals

És clar que si un (n, k) - C -autòmat M , com el definit anteriorment, ha escrit un símbol estrella, aquesta posició en la cinta on el símbol estrella ha estat escrit esdevé “inactiu” en el sentit que mai cap altre símbol podrà ser escrit en aquesta posició. En comptes d'escriure un símbol estrella, M podria, senzillament, esborrar aquesta posició de la seva cinta i escurçar la cinta una cel·la. Direm que M realitza una operació de contracció. En aquesta situació, doncs, una operació estrella és reemplaçada per una operació de contracció, i, pel fet que M realitza operacions de contracció, la seva cinta és escurçada. M accepta

si la seva cinta té longitud 3 i consisteix en les cadenes $\phi\Omega\$,$ on Ω és el símbol final. Aquesta definició equivalent justifica el nom d'“autòmat de contracció”.

Pel fet que el símbol estrella $*$ desapareix de la la definició, qualsevol element $(s', *, D)$ de $\delta(s, A)$ esdevé (s', D) , i

$$(s, \phi Y_1 Y_2 \dots Y_p \underline{A} B Y_{p+1} \dots Y_r \$) \vdash_M^* (s', \phi Y_1 Y_2 \dots Y_p * \underline{B} Y_{p+1} \dots Y_r \$)$$

esdevé

$$(s, \phi Y_1 Y_2 \dots Y_p \underline{A} B Y_{p+1} \dots Y_r \$) \vdash_M^* (s', \phi Y_1 Y_2 \dots Y_p \underline{B} Y_{p+1} \dots Y_r \$).$$

En aquest cas, el llenguatge acceptat pel (n, k) - C -autòmat M és el conjunt $L(M)$ de manera que

$$L(M) = \{w \in (V - \{\phi, \$\})^* \mid (s_0, \phi w \$) \vdash_M^* (s_f, \phi \Omega \$)\}.$$

Aquesta maquinària està relacinada amb la nostra imatge d'un autòmat contextual associada a la d'un *autòmat de memòria linealment limitada*, i.e. una simple màquina de Turing amb, en comptes de tenir una cinta potencialment infinita en la qual computar, està restringida a la part de la cinta que conté la cadena d'entrada w a més de dues cel·les addicionals que contenen els marcadors dels extrems (ϕ i $\$$). Aquesta cinta és explorada per un capçal de lectura/escriptura, segons l'estat actiu de la memòria finita. Per tal de correspondre a l'estil contextual de treballar, aquesta maquinària és

capaç d'executar l'operació inversa a la de la inserció d'un context, és a dir, la d'*esborrar un context* (parell de cadenes (u, v) , on u és la primera component del context i v és la segona component del context).

En el cas dels autòmats contextuals, l'operació de contracció està associada a l'operació d'esborrament. En un estat inicial s_0 , començant per una cadena w acotada pels marcadors $\$, \$$, aquesta cadena w és acceptada per l'autòmat contextual M si, i només si, l'autòmat és capaç d'esborrar tots els símbols excepte $\$, \$$ i després arribar a un estat final s_f . D'una manera planera, el treball d'un autòmat contextual consisteix en cicles: començar amb el capçal de lectura/escriptura en l'extrem esquerre de la cadena, anar cap a la dreta, explorant parts de la cadena i esborrant dues subcadenaes de la mateixa, aleshores anar cap a l'esquerre, deixant la cadena invariable, fins arribar de nou a l'extrem esquerre de la cadena. (El treball de l'autòmat és exactament d'aquest tipus quan esborra només en presència dels marcadors dels extrems, però és molt menys concret en el cas en què no totes les operacions d'esborraments involucrin els marcadors $\$, \$$.)

El llenguatge associat a l'autòmat contextual M , de manera usual, seria:

$$L(M) = \{w \in (V - \{\$, \$\})^* \mid (s_0, \underline{\$}w\$) \vdash_M^* (s_f, \underline{\$}), s_f \in F\},$$

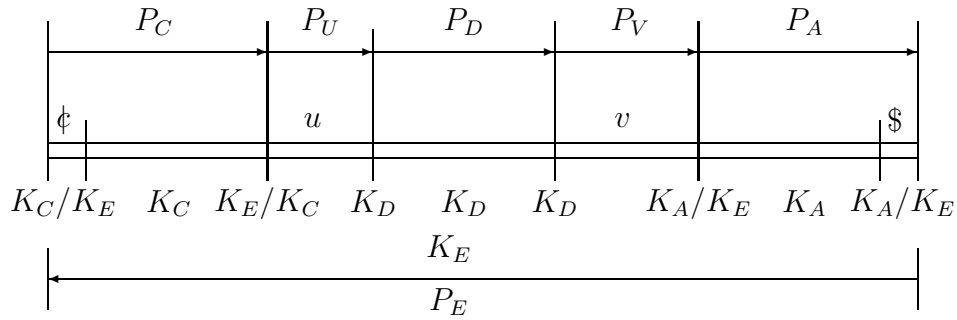
El treball d'un autòmat contextual és intuïtivament descrit a la Figura 14.2.1, on s'indiquen els conjunts de moviments (expressats com regles de re-escriptura $P = \cup\{P_i\}_{i=\{C,U,D,V,A,E\}}$) usats en diversos segments de la cadena

considerada:

- començament-de-la-cadena,
- esborrament-de-la-cadena-**u**,
- moviment-cap-a-la-dreta,
- esborrament-de-la-cadena-**v**,
- acabament-de-la-cadena,
- moviment-cap-a-l'esquerre,

respectivament; i els conjunts d'estats ($K = \bigcup \{K_i\}_{i=\{C,U,D,V,A,E\}}$) assolibles en diversos moments de l'anàlisi.

Figura 14.2.1 Descripció intuïtiva del funcionament d'un autòmat contextual basat en automats de contracció.



Més explícitament, la intuïció que hi ha al darrera d'aquesta figura és la següent: els estats de K_C i les regles de P_C corresponen a l'exploració del començament de la cadena. El treball comença en l'estat s_0 o en un altre estat de K_C . És possible no usar cap regla de P_C , pel fet que el treball de P_U pot començar o bé en un estat de K_C o en un estat de K_E . Les regles de P_U executen l'acció d'esborrament d'una subcadena de la cadena explorada (associada al primer terme d'un context). Aquesta cadena pot ser buida: les regles $s\lambda\lambda \rightarrow \lambda s'$ poden aparèixer a P_U . S'entra aleshores en un estat de K_D que determina el moviment del capçal de l'autòmat cap a la dreta, explorant de nou una subcadena de la cadena considerada. Cal notar que assolir un estat de K_D és un fet obligatori; P_V i, per tant, també els altres conjunts de moviments, no poden treballar a P_U si no introdueix un estat a K_D . A més a més, a com hem apuntat abans, aquest fet pot succeir, si és necessari, mitjançant una regla de canvi-d'estat $s \rightarrow s'$ a P_U . Aleshores, mitjançant regles de P_V s'esborra una cadena més, essent possible la cadena buida (amb la qual cosa es completa l'esborrament d'un context), continuant amb regles de P_A , que involucren estats de K_A i explora un sufix de la cadena considerada. El prefix explorat per P_C i el sufix explorat per P_A poden consistir en només un símbol, els marcadors ϕ , $\$$, respectivament. Aquest possible únic símbol és suficient per a analitzar a partir d'un estat de K_L a un altre de K_C , mitjançant una regla de P_C , o bé mitjançant una regla de K_D , usant una regla de P_D . Similarment, el marcador $\$$ fa possible la introducció del símbol s de K_E . Mitjançant regles de P_E es va des de la dreta cap a l'esquerra, en la presència

d'estats de K_L . Cal notar que durant les operacions de moviment (regles de $P_C \cup P_D \cup P_A \cup P_E$) els símbols explorats no són mai canviats, només l'estat actual pot ser modificat i que els marcadors $\phi, \$$ tampoc són mai modificats (fins i tot les regles d'esborrament de $P_U \cup P_V$ els deixen invariants).

Aquesta idea intuïtiva ens condueix cap a la definició següent:

Definició 14.2.5 Un *autòmat contextual* és una construcció

$$M = (K, V, \phi, \$, s_0, F, P),$$

on K és un conjunt finit d'estats, V és un vocabulari, $\phi, \$$ són dos símbols especials que no pertanyen a V (*marcadors dels extrems*), $s_0 \in K$ és l'estat inicial, $F \subseteq K$ és el conjunt d'estats finals i P és un conjunt finit de *produccions* o *moviments*, expressats com a regles de reescriptura. Hi ha quatre tipus d'estats amb $K = K_C \cup K_D \cup K_A \cup K_E$, on els quatre conjunts K_C, K_D, K_A, K_E són disjunts dos a dos (i $s_0 \in K_C$ i $F \subseteq K_D$) i sis tipus de moviments amb $P = P_C \cup P_U \cup P_D \cup P_V \cup P_A \cup P_E$, de les següents formes

- P_C conté regles de la forma $sa \rightarrow as'$, amb $s \in K_E \cup K_C, s' \in K_C, a \in V \cup \{\phi\}$;
- P_U conté regles de la forma $s\alpha x \rightarrow \alpha s'$, amb $s \in K_E \cup K_C, s' \in K_D, x \in V^*, \alpha \in \{\phi, \lambda\}$;
- P_D conté regles de la forma $sa \rightarrow as'$, amb $s \in K_D, s' \in K_D, a \in V$;

- P_V conté regles de la forma $sx\alpha \rightarrow s'\alpha$, amb $s \in K_D, s' \in K_A \cup K_E, x \in V^*, \alpha \in \{\$, \lambda\}$;
- P_A conté regles de la forma $sa \rightarrow as'$, amb $s \in K_A, s' \in K_A \cup K_E, a \in V \cup \{\$\}$;
- P_E conté regles de la forma $as \rightarrow s'a$, amb $s \in K_E, s' \in K_E, a \in V \cup \{\$, \lambda\}$.

L'autòmat contextual definim la relació de reescriptura $\Rightarrow$ entre cadenes sobre $V \cup K \cup \{\$, \lambda\}$ de la manera usual (un autòmat contextual pot ser vist com un sistema de reescriptura, en el sentit general). Aleshores, podem definir dos llenguatges associats a un autòmat M :

$$L_0(M) = \{w \in V^* \mid s_0\phi w\$ \Rightarrow^* \phi s_f\$, \text{ per algun } s_f \in F\},$$

$$L_t(M) = \{w \in V^* \mid s\phi w\$ \Rightarrow^* \phi s_f\$, \text{ per algun } s \in K_C, s_f \in F\}.$$

El llenguatge $L_0(M)$ fa ús de l'estat inicial d' M , el llenguatge $L_t(M)$ està definit usant com a estat inicial qualsevol element de K_C (el subíndex t de $L_t(M)$ vol significar “total”). En ambdós casos es comença per una cadena acotada pels marcadors $\phi, \$$; aquesta cadena és acceptada si, i només si, l'autòmat és capaç d'esborrar tots els símbols excepte $\phi, \$$ i asoleix un estat final.

En aquesta situació, podem distingir diversos tipus d'autòmats contextuais:

- autòmats *marcats* = amb totes les regles de $P_U \cup P_V$ tenint $\alpha \neq \lambda$;
- autòmats *no-marcats* = amb totes les regles de $P_U \cup P_V$ tenint $\alpha = \lambda$;
- autòmats *c-constant* = tenint totes les regles $sa \rightarrow as'$ de P_C amb $s = s'$;
- autòmats *d-constant* = tenint totes les regles $sa \rightarrow as'$ de P_D amb $s = s'$;
- autòmats *d-no-cíclic* = cada reescriptura $sw \Rightarrow^* ws'$ que usa regles de P_D té $s \neq s'$;
- autòmats *a-constant* = tenint totes les regles $sa \rightarrow as'$ de P_A amb $s = s'$;
- autòmats *e-constant* = tenint totes les regles $as \rightarrow s'a$ de P_E amb $s = s'$;
- autòmats *simples* = tenint $\text{card}(K_E) = 1$.

Clarament, un autòmat simple és també *e-constant*. A més a més, per a autòmats marcats les regles de P_C i P_A no són usades i, en aquest cas concret, podem ignorar-les; això vol dir que els autòmats marcats poden ser també considerats *c-* i *a-constant*.

14.3 Caracterització de llenguatges contextuais usant famílies de *DR*-autòmats

La definició de gramàtica contextual amb selecció regular considerada pot ser extesa a famílies generals de llenguatges de manera similar com es mostra a

(Păun et al., 1994b) i (Păun et al., 1995a). P. Jančar i al. no consideren casos més generals pel fet que ells estan principalment interessats en autòmats que reconeguin classes similars de llenguatges, i en comparara diferents tipus de gramàtiques contextuais amb selecció regular amb diversos tipus d'autòmats d'esborrament. Alguns tipus d'aquests autòmats semblen ser un bon marc teòric per a l'anàlisi de l'ordre lliure de les paraules en llengües naturals (el txec, per exemple, en el seu cas). Per aquestes raons la restricció a gramàtiques contextuais amb selecció *regular* sembla ser suficient.

Basant-nos en la Secció 14.1 definirem, en primer lloc, una classe de *DR*-autòmats que reconeixen, exactament, la família $EC(REG)$. Cal notar la similitud entre una derivació segons una gramàtica $RCG(\{\lambda\}, REG, \{\lambda\}) = EC(REG)$ i un cicle en el cas dels *DR*-autòmats definits.

Recordem una de les Definicions de 14.1.1 donada a la pàgina 509. Un *DR*-autòmat M s'anomena *extern* (denotat per *ext-DR*-autòmat) si:

- és en la forma cíclica dèbil i, a més a més,
- cada cicle de qualsevol computació d' M té la següent forma: la primera instrucció usada en el cicle pot amputar algun prefix de la cadena considerada; i, després, només pot seguir una seqüència d'instruccions d'operacions de moure's a la dreta, amb una excepció – l'última instrucció del cicle (que acaba amb l'operació de reiniciació) pot amputar algun sufix de la cadena explorada. Per tant, M només pot acceptar immediatament

cadena curta, les cadenes més llargues són rebutjades o escurçades mitjançant l'esborrament de prefixos i/o sufixos curts (limitats per la mesura de l'espiera) de la cadena en un cicle. Òbviament l'*ext-DR*-autòmat és normal.

Aleshores

Teorema 14.3.1 Un llenguatge L és generat per una gramàtica contextual externa amb selecció regular si, i només si, L és reconegut per un *ext-DR*-autòmat. És a dir,

$$\mathcal{L}(EC(REG)) = \mathcal{L}(ext-DR)$$

Demostració. a) $EC(REG) = RCG(\{\lambda\}, REG, \{\lambda\})$ implica que per a qualsevol llenguatge L pertanyent a $EC(REG)$ existeix una gramàtica contextual amb selecció regular $G = (V, B, C, \{\{\lambda\}\}, \mathcal{I}, \{\{\lambda\}\}, f)$ tal que $L = L(G)$. Usant la construcció de la demostració del Teorema 14.1.4 de la pàgina 528 obtenim un *norm-DR*-autòmat M que reconeix $L(G)$. És fàcil de verificar que aquest autòmat M construït és extern.

b) Sigui M un *ext-DR*-autòmat. Usant la construcció de la demostració del Teorema 14.1.2 de la pàgina 524 obtenim una gramàtica contextual amb selecció regular $G = (V, B, C, \{\{\lambda\}\}, \mathcal{I}, \{\{\lambda\}\}, f)$ tal que $L(G) = L(M)$. $\square$

De manera similar podem mostrar construccions d'autòmats corresponents a altres classes de gramàtiques contextuais amb selecció regular. Per exem-

ple, gramàtiques contextuais amb prefix regular, infix finit, sufix acotat i amb selecció regular.

Definició 14.3.1 Una gramàtica contextual amb prefix regular, infix finit, sufix acotat, i amb selecció regular, abrev. *RFBRCG* (de “regular-prefix finite-infix bounded-suffix contextual grammar with regular selection”), és una gramàtica contextual amb selecció regular $G = (V, B, C, \mathcal{P}, \mathcal{I}, \mathcal{S}, \varphi)$ de manera que $\mathcal{I} \subset_{fin} FIN$ i $\mathcal{S} \subset_{fin} \{K \cdot V^* \mid K \in FIN\}$ ($A \subset_{fin} B$ significa que A és un subconjunt finit de B).

Pel fet de ser els infixos finits i els sufixos acotats els *DR*-autòmats corresponents són els *R*-autòmats normals.

Teorema 14.3.2 Un llenguatge L està generat per una gramàtica contextual amb prefix regular, infix finit, sufix acotat i amb selecció regular si, i només si, L és reconegut per un *norm-R*-autòmat. És a dir,

$$\mathcal{L}(RFBRCG) = \mathcal{L}(norm-R)$$

Demostració. a) Sigui $G = (V, B, C, \mathcal{P}, \mathcal{I}, \mathcal{S}, f)$ una gramàtica arbitrària pertanyent a *RFBRCG*. Construïrem un *norm-R*-autòmat M que reconegui el llenguatge $L(G)$.

Usant la construcció de la demostració del Teorema 14.1.4 de la pàgina 528 obtindrem un *norm-DR*-autòmat M' que reconeix $L(G)$. Escollirem la mesura de l'espiera d' M de manera que pugui esborrar ambdues parts d'un context en

una instrucció mitjançant la verificació simultània d'un prefix finit de la resta de la cadena seguit del context dret. Això és possible pel fet que, per verificar si el sufix de la cadena específica pertany a S , és suficient amb verificar-ho per algun prefix finit de la cadena.

Sigui

$$\mathcal{S} = \{F_1 \cdot V^*, F_2 \cdot V^*, \dots, F_n \cdot V^*\}, \text{ on } F_i \in FIN, \text{ per a } 1 \leq i \leq n, \\ \text{i sigui } k_S = \max_{1 \leq i \leq n} \{|u|; u \in F_i\}.$$

La mesura de l'espiera d' M serà $k_M = \max(k_S + k_C, k_B)$, on k_C és la longitud màxima de qualsevol context juntament amb el seu infix selector a G , i.e.

$$k_C = \max\{|uyv|; u, v, y \in V^*, (u, v) \in C : (u, v) \in f(x, y, z) \text{ per algun } x, z \in V^*\}$$

i k_B és la longitud màxima d'una cadena del conjunt base B .

Més detalls de com modificar el *norm-DR*-autòmat M' en el *norm-R*-autòmat M són obvis i els obviarem.

b) La inclusió inversa pot ser demostrada de la següent manera. Sigui M un *norm-R*-autòmat amb espiera k en la forma cíclica dèbil. Descriurem com construir una gramàtica de prefix regular, infix finit, sufix acotat, amb selecció regular $G = (V, B, C, \mathcal{P}, \mathcal{I}, \mathcal{S}, f)$ que generi el llenguatge $L(M)$. Usant la construcció de la demostració del Teorema 14.1.2 de la pàgina 524, construïm una

gramàtica contextual amb selecció regular $G = (V, B, C, \mathcal{P}, \mathcal{I}, \mathcal{S}, f)$ de manera que $L(G) = L(M)$. Clarament $\mathcal{P} \subset_{fin} REG$. Pel fet que M és un autòmat de reiniciació, $\mathcal{I} \subset_{fin} FIN$ i M reinicia immediatament després d'esborrar els contextos – i.e. només prefixos finits del sufix de la cadena són verificats. En aquesta situació $\mathcal{S} \subset_{fin} \{K \cdot V^* \mid K \subset V^{k*}\}$, on V^{k*} és el conjunt de totes les cadenes de V^* de longitud com a màxim k .

L'autòmat M és en la forma cíclica dèbil, és a dir, el conjunt de cadenes acceptades per M sense reiniciació és finit i aquest serà el conjunt base B de la gramàtica G . Per tant, la gramàtica contextual construïda és amb prefix regular, infix finit, sufix acotat, amb selecció regular.

Per tant, la família de gramàtiques $RFBRG$ genera, exactament, la classe de llenguatges que és reconeguda per autòmats amb reiniciació normals. $\square$

14.4 Caracterització de famílies de llenguatges contextuais usant famílies d'autòmats basades en autòmats de contracció

Per tal de denotar les famílies de llenguatges reconegudes mitjançant autòmats contextuais identificarem per I quan considerem el cas dels llenguatges del tipus “inicial” $L_0(M)$, i per T quan considerem llenguatges del tipus “total” $L_t(M)$; per M indicarem l'ús d'autòmats “marcats”, i per U l'ús d'autòmats “no marcats”; mitjançant C, D, A, E indicarem els autòmats que

són “ c -, d -, a -, e -constants”, respectivament; mentre que N indicarà “ d -no-cíclic” i S que l’autòmat és “simple” (l’absència d’algun dels trets C, D, A, E, N, S és indicada mitjançant l’absència de la corresponent lletra).

Per exemple, $TUCDAS$ és la família de tots els llenguatges de la forma $L_t(M)$, reconeguts per autòmats contextuals no-marcats que són c -, d - i a -constants i simples. Mentre que IMS és la família de llenguatges de la forma $L_0(M)$ per a M un autòmat contextual simple marcat (el qual no és, no obstant, d -constant). Clarament, un autòmat d -no-cíclic és no d -constant i només pot explorar cadenes de longitud acotada quan usi regles de P_D .

14.4.1 Resultats preliminars

A partir de les anteriors observacions tenim

Lema 14.4.1 $XS \subseteq XD$ per a qualsevol família X , i $XY \subseteq XCA$ per a qualsevol família XY . A més a més, qualsevol família XZY està inclosa dins la corresponent família XY , on $Z \in \{C, D, A, E\}$ i XY és qualsevol combinació possible dels altres trets.

A més a més tenim

Lema 14.4.2 $TX \subseteq IX$ per a qualsevol família X .

Demostració. Si $M = (K, V, \phi, \$, s_0, F, P)$ és un autòmat contextual, construïm l'autòmat $M' = (K \cup \{s'_0\}, V, \phi, \$, s'_0, F, P')$, on s'_0 és un nou estat i P' s'obté a partir de P afegint les regles

$$\{s'_0\phi \rightarrow \phi s \mid s \in K_C\}$$

a P_C , i les regles

$$\{s'_0\alpha x \rightarrow \alpha s' \mid s\alpha x \rightarrow \alpha s' \in P_U, x \in V^*, \alpha \in \{\phi, \lambda\}\}$$

a P_U .

Usant les primeres noves regles podem començar a s'_0 i produir immediatament qualsevol estat de K_C usant les regles del segon conjunt podem simular l'ús de qualsevol regla de P_U en presència de ϕ . Aquestes operacions són possibles només al començament del treball de l'autòmat. Després d'això, M' usa exactament els mateixos moviments que M . En conseqüència, $L_0(M') = L_t(M)$. $\square$

Lema 14.4.3 Tota família de llenguatges reconeguda per autòmats contextuals està inclosa a CS .

Demostració. Donat un autòmat contextual M , és fàcil de construir un autòmat de memòria linealment limitada M' que treballi exactament com M , marcant d'alguna manera els símbols que són esborrats per les regles d' M a partir de P_U i de P_V (en la següent exploració d'aquests símbols, l'autòmat M' només ha

d'ignorar-los, deixant-los invariables i deixant l'estat constant). Independentment de quin és el tipus d' M , tenim $L_0(M) = L(M')$ i, per tant, el llenguatge $L_0(M)$ és sensible al context. A partir del Lema 14.4.2, obtenim que també $L_t(M)$ és sensible al context.

14.4.2 Resultats de caracterització

A continuació presentarem alguns resultats relacionats amb algunes classes d'autòmats contextuais definits, de la manera que hem vist prèviament, els quals reconeixen, exactament, algunes classes de llenguatges contextuais:

Teorema 14.4.1 Autòmats contextuais que reconeixen, exactament, les famílies de llenguatges EC i IC .

- (1) Un llenguatge L és generat per una gramàtica contextual externa sense selecció si, i només si, L és reconegut per un autòmat contextual “total” que sigui marcat, d -constant i simple. És a dir,

$$\mathcal{L}(TMDS) = \mathcal{L}(EC).$$

- (2) Un llenguatge L és generat per una gramàtica contextual interna sense selecció si, i només si, L és reconegut per un autòmat contextual “total” que sigui no-marcant, c -, d - i a -constant i simple. És a dir,

$$\mathcal{L}(TUCDAS) = \mathcal{L}(IC).$$

Demostració. b) A (1) considerem una gramàtica contextual sense selecció $G = (V, B, C)$ i construïm l'autòmat contextual

$$M = (K, V, \clubsuit, \$, s_0, \{s_f\}, P),$$

amb

$$K_C = \{s_0\},$$

$$K_A = \emptyset,$$

$$K_D = \{s_{(u,v)} \mid (u, v) \in C\} \cup \{s_f\},$$

$$K_E = \{s_1\},$$

i P consistint en les següents regles:

$$P_U = \{s_0 \clubsuit u \rightarrow \clubsuit s_{(u,v)} \mid (u, v) \in C\} \cup$$

$$\cup \{s_0 \clubsuit w \rightarrow \clubsuit s_f \mid w \in B\},$$

$$P_D = \{s_{(u,v)} a \rightarrow a s_{(u,v)} \mid (u, v) \in C\},$$

$$P_V = \{s_{(u,v)} v \$ \rightarrow s_1 \$ \mid (u, v) \in C\},$$

$$P_E = \{a s_1 \rightarrow s_1 a \mid a \in V\} \cup \{\clubsuit s_1 \rightarrow s_0 \clubsuit\}.$$

És fàcil de veure que $L_{ex}(G) = L_t(M)$ i que M és marcat, simple i d -constant. Per tant, tenim la inclusió $EC \subseteq TMDS$.

Quan la gramàtica G és usada en el mode intern (2), construïm M de la mateixa manera que abans amb el conjunt de regles reemplaçat per

$$\begin{aligned}
 P_C &= \{s_0a \rightarrow as_0 \mid a \in V \cup \{\phi\}\}, \\
 P_U &= \{s_0u \rightarrow s_{(u,v)} \mid (u,v) \in C\} \cup \\
 &\quad \cup \{s_0w \rightarrow s_f \mid w \in B\}, \\
 P_D &= \{s_{(u,v)}a \rightarrow as_{(u,v)} \mid (u,v) \in C\}, \\
 P_V &= \{s_{(u,v)}v \rightarrow s_1 \mid (u,v) \in C\}, \\
 P_E &= \{as_1 \rightarrow s_1a \mid a \in V\} \cup \{\phi s_1 \rightarrow s_0\phi\}.
 \end{aligned}$$

Obtenim un autòmat M' pel qual tenim $L_{in}(G) = L_t(M')$ (Cal notar que després d'esborrar u i v corresponents a un context (u,v) , introduïm l'estat s_1 que va cap a l'esquerre, reintroduïm s_0 , que pot anar cap a qualsevol posició de la cadena activa cap a la dreta; per tant, no existeix cap restricció en l'emplaçament d'on el següent context és esborrat). L'autòmat M' és simple, c -, d -, a -constant i no-marcad. Per tant, tenim la inclusió $IC \subseteq TUCDAS$.

a) Inversament, considerem un autòmat contextual $M = (K, V, \phi, \$, s_0, F, P)$ que sigui (1) marcat, d -constant i simple, o (2) no-marcad, c -, d - i a -constant

i simple. Sigui $K_E = \{s_1\}$. Construïm la gramàtica $G = (V, B, C)$, amb

$$B = \{w \in V^* \mid \phi s_1 uv\$ \implies \phi sv\$ \implies \phi s_f \$, \text{ per a algun}$$

$$u, v \in V^*, w = uv, s \in K, s_f \in F\},$$

$$C = \{(u, v) \mid \text{existeixen } s, s', s'' \in K \text{ tal que}$$

$$s\alpha u \rightarrow \alpha s' \in P_U, s'v\beta \rightarrow s''\beta \in P_V,$$

$$u, v \in V^*, \alpha \in \{\phi, \lambda\}, \beta \in \{\$, \lambda\}\}.$$

És fàcil de veure que en el cas (1) tenim $L_t(M) = L_{ex}(G)$ i en el cas (2) tenim $L_t(M) = L_{in}(G)$, la qual cosa completa la demostració. $\square$

Teorema 14.4.2 Autòmats contextuais que reconeixen, exactament, les famílies de llenguatges llenguatges $ECC(REG)$ i $ICC(REG)$.

- (1) Un llenguatge L és generat per una gramàtica contextual externa amb selecció regular si, i només si, L és reconegut per un autòmat “total” que sigui marcat i simple. És a dir,

$$= \mathcal{L}(TMS) = \mathcal{L}(ECC(REG)).$$

- (2) Un llenguatge L és generat per una gramàtica contextual interna amb selecció regular si, i només si, L és reconegut per un autòmat contextual “total” que sigui no-marcant, c - i a -constant i simple. És a dir,

$$\mathcal{L}(TUCAS) = \mathcal{L}(ICC(REG)).$$

Demostració. a) La diferència entre aquesta demostració i la del teorema anterior és que en aquest cas no tenim autòmats d -constant. Això vol dir que si usem les regles de P_D , podem verificar si la cadena explorada, després d'usar una regla de P_U i abans d'usar una regla de P_V , és o no és un element del selector associat al context esborrat. Pel fet que aquest selector és un llenguatge regular, aquesta verificació és possible. Inversament, l'exploració realitzada per regles de P_D correspon al treball d'un autòmat finit, començant en l'estat inicial després d'usar una regla de P_U i finalitzant en un estat final quan usa una regla de P_V ; associant aquest llenguatge regular amb el context esborrat (u, v) , obtenim una gramàtica contextual amb selecció regular.

Considerem primer el cas extern (1).

Considerem una gramàtica $G = (V, B, P)$ amb $P = \{(S_1, C_1), \dots, (S_n, C_n)\}$, on S_i són llenguatges regulars. Considerem un autòmat finit $M_i = (K_i, V, s_{i,0}, F_i, P_i)$ per a cada $S_i, 1 \leq i \leq n$. Sense perdre cap generalitat podem assumir que els conjunts $K_i, 1 \leq i \leq n$, són mútuament disjunts. Construïm l'autòmat contextual

$$M = (K, V, \phi, \$, s_0, \{s_f\}, P),$$

amb

$$K_C = \{s_0\},$$

$$K_A = \emptyset,$$

$$K_D = \{[s_{(u,v)}, s] \mid (u, v) \in C, s \in K_i\} \cup \{s_f\},$$

$$K_E = \{s_1\},$$

i P consistint en les següents regles:

$$\begin{aligned}
P_U &= \{s_0\phi u \rightarrow \phi[s_{(u,v)}, s_{i,0}] \mid (u,v) \in C_i, 1 \leq i \leq n\} \cup \\
&\quad \cup \{s_0\phi w \rightarrow \phi s_f \mid w \in B\}, \\
P_D &= \{[s_{(u,v)}, s]a \rightarrow a[s_{(u,v)}, s'] \mid (u,v) \in C_i, 1 \leq i \leq n, \text{ i} \\
&\quad sa \rightarrow as' \in P_i\}, \\
P_V &= \{[s_{(u,v)}, s]v\$ \rightarrow s_1\$ \mid (u,v) \in C_i, 1 \leq i \leq n, \text{ i } s \in F_i\}, \\
P_E &= \{as_1 \rightarrow s_1a \mid a \in V\} \cup \{\phi s_1 \rightarrow s_0\phi\}.
\end{aligned}$$

És fàcil de veure que $L_{ex}(G) = L_t(M)$ i que M és marcat, simple i d -constant. Per tant, tenim la inclusió $ECC(REG) \subseteq TMDS$.

Les modificacions pel cas intern (2) són òbvies.

b) La demostració d'aquesta part està basada en el fet que: per a cada context (u,v) prenem el selector $S_{(u,v)} = \{w \in V^* \mid s\alpha u w v \beta \implies \alpha s_1 w v \beta \implies^* \alpha w s_2 v \beta \implies \alpha w s_3 \beta\}$ respecte a l'autòmat $M, \alpha \in \{\phi, \lambda\}, \beta \in \{\$, \lambda\}$. $\square$

Teorema 14.4.3 Un llenguatge L és generat per una gramàtica contextual total amb selecció regular si, i només si, L és reconegut per un autòmat contextual “total” que sigui no-marcad i simple. És a dir,

$$\mathcal{L}(TUS) = \mathcal{L}(TC(REG)).$$

Demostració. És òbvia a partir del Teorema 14.4.2 anterior i de la definició de gramàtica contextual total (Definició 4.1.3 de la pàgina 172). $\square$

Teorema 14.4.4 Autòmats contextuais que reconeixen, exactament, les famílies de llenguatges $ECC(FIN)$ i $ICC(FIN)$.

- (1) Un llenguatge L és generat per una gramàtica contextual externa amb selecció finita si, i només si, L és reconegut per un autòmat contextual “total” que sigui marcat, d -no-cíclic i simple. És a dir,

$$\mathcal{L}(TMNS) = \mathcal{L}(ECC(FIN)).$$

- (2) Un llenguatge L és generat per una gramàtica contextual interna amb selecció finita si, i només si L és reconegut per un autòmat contextual “textual” que sigui no-marcad, c - i a -constant, d -no-cíclic i simple. És a dir,

$$\mathcal{L}(TUNS) = \mathcal{L}(ICC(FIN)).$$

Demostració. La demostració del Teorema 14.4.2 de la pàgina 553 pot ser fàcilment particularitzada pel cas de d -no-cíclics autòmats. $\square$

14.4.3 Derivacions marcades

En totes les caracteritzacions anteriors hem treballat amb autòmats contextuais simples. La raó és clara: quan K_L no està restringida a només un element, aleshores podem “recordar”, mitjançant els estats de les regles de P_L , quin era el context esborrat prèviament. D’aquesta manera, el context esborrat en el pròxim pas depèn del context esborrat un pas abans. Veurem ara que

aquests autòmats caracteritzen els llenguatges generats per les denominades gramàtiques contextuais programades (els contextos de les quals estan connectats mitjançant un graf, similar al cas de les gramàtiques lliures del context programades). Semblantment, quan tractem amb el llenguatge $L_0(M)$, podem controlar el primer context que és esborrat. Aquestes gramàtiques s'anomenen *amb derivacions marcades*. Aquesta restricció (així com la restricció programada) és sabut que incrementa el poder de les gramàtiques contextuais.

Les gramàtiques contextuais amb derivacions marcades – veure la Secció 11.5 – són en alguns aspectes similars a les gramàtiques amb derivació bloquejada, que són considerades a (Păun et al., 1995a): la derivació té lloc de manera usual i acaba amb l'adjunció d'un context d'un conjunt especial. Aquest conjunt especial pot ser vist, per altra banda, com simètric al conjunt axiomàtic (si tenim un conjunt de cadenes inicials, que són correctes per definició, és natural el tenir també un conjunt de cadenes “finals”, que tanquen el procés de derivació), i, per altra banda, es correspon amb la situació que apareix en diverses circumstàncies, quan l'últim pas d'una “derivació” és distingit (cal recordar, per exemple, el cas dels sistemes d'accions simples, on l'última acció elemental és un element d'un conjunt d'accions finals donat).

Ja hem fet esment al fet que els autòmats contextuais poden especificar el primer context que és esborrat si és considerat el llenguatge $L_0(M)$ associat a ells. Això ens proporciona caracteritzacions de les famílies $EC_{mk}, IC_{mk}, ECC_{mk}, ICC_{mk}$, $F \in \{FIN, REG\}$, similar a les dels Teoremes 14.4.1 de la

pàgina 550, 14.4.2 de la pàgina 553 i 14.4.4 de la pàgina 556. En aquesta situació, obtenim les següents igualtats (les demostracions són idèntiques a les demostracions donades pels teoremes indicats anteriorment, començant per esborrar un context del conjunt C_2 d'una gramàtica contextual amb derivació marcada).

Teorema 14.4.5 Tenim

- (i) $IMDS = EC_{mk}$,
- (ii) $IUCDAS = IC_{mk}$,
- (iii) $IMS = ECC_{mk}(REG)$,
- (iv) $IUCAS = ICC_{mk}(REG)$.

14.4.4 Llenguatges contextuais programats

Els autòmats contextuais capaços de recordar, d'un cicle a un altre, el context esborrat poden caracteritzar famílies de llenguatges generades per gramàtiques contextuais programades, considerades a la Secció 12.2. Això és cert per a autòmats contextuais que siguin no simples, i només d -constants: un estat de K_L , els estats d'anada cap a l'esquerra, poden ser associats amb cada context. És clar que, les restriccions imposades pel parell del conjunt U d'una gramàtica contextual programada són ara verificades en l'ordre invers: per a $[(u_1, v_1), (u_2, v_2)] \in U$ esborrem en primer lloc el context (u_2, v_2) , emmagatzem

aquesta informació en un estat $s_{(u_2, v_2)}$ de K_L , després esborrem (u_1, v_1) , tal com és imposat per $s_{(u_2, v_2)}$, i així successivament. Les construccions de les demostracions dels Teoremes 14.4.1, 14.4.2, i 14.4.3 són usades pels conjunts de regles P_C, P_U, P_D, P_V, P_A . A més a més, obtenim

Teorema 14.4.6 Tenim

$$(i) \quad TMDE = PEC,$$

$$(ii) \quad TUCDAE = PIC,$$

$$(iii) \quad TME = PECC(REG),$$

$$(iv) \quad TUCAE = PICC(REG),$$

$$(v) \quad TUCNAE = PICC(FIN).$$

14.4.5 Altres caracteritzacions

El sentit comú ens diu que, usant aquests tipus d'autòmats contextuais, podem construir *a priori* caracteritzar qualsevol família de llenguatges contextuais. És, però, una investigació que resta per fer.

Notes bibliogràfiques. Els autòmats considerats en aquest capítol són nous, però relacionats amb cert tipus d'autòmats introduïts per Păun a (Păun,

1982a), seguint el model dels denominats *autòmats de contracció* (“contraction automata”) de von Solms a (von Solms, 1975); així com en els denominats *autòmats de reiniciació* (“restarting automata”) de Jančar i al. que apareixen a (Jančar et al., 1994). La caracterització d’una família particular de llenguatges contextuais apareix també a (Jančar et al., 1995a), usant aquests autòmats de reiniciació. En l’article (Jančar et al., 1996), es consideren les gramàtiques contextuais totals amb els selectors de la forma (S_L, S_M, S_R) , sent S_L, S_M, S_R llenguatges regulars, que seleccionen x_1, x_2, x_3 , respectivament, en un pas de derivació $x_1x_2x_3 \implies x_1ux_2vx_3$. Els llenguatges generats d’aquesta manera són caracteritzats per autòmats d’esborrament amb operació de reiniciació. Casos particulars d’aquests autòmats poden reconèixer, exactament, famílies de llenguatges contextuais interns o externs.

Alguns resultats del mostrats en aquest capítol ja apareixen a (Miquel-Vergés, 1996b) o bé a (Martín-Vide et al., 1996a), però altres són originals.

Capítol 15

Caracterització de llenguatges associats a l'acció dramàtica (teatre i còmic) mitjançant gramàtiques contextuals

15.1 Llenguatges associats al teatre

Els primers treballs que poden fer pensar en una teoria matemàtica del teatre tenen molt pocs anys. Podem fixar exactament les coordenades de la seva aparició. Durant els anys 1966–67 Solomon Marcus – matemàtic, professor de Lingüística matemàtica a la Facultat de Matemàtiques de la Universitat de Bucarest – imparteix un curs de “Poètica matemàtica” (en romanès) a la Facultat de Llengua i Literatura Romanesa de la Universitat de Bucarest (Marcus, 1967f). No és, però, fins a finals de 1967 quan verdaderament s’enceta

aquesta temàtica amb la publicació a la revista *T. A. Information* d'un article del mateix Solomon Marcus titulat "Modèles mathématiques dans l'étude du drame" (Marcus, 1967e). És un article molt curt, gairebé una nota, de dues pàgines – de fet és un abstracte d'una lectura relacionada amb l'estratègia dels personatges teatrals impatida a l'"Association pour la traduction automatique et la linguistique appliquée" de París – on l'autor planteja, a grans trets, allò que podria ser un aplicació dels models matemàtics en la descripció i explicació de l'obra dramàtica. Cal pensar, doncs, que eren tan sols les insinuacions d'uns camins epistemològics possibles.

Fou el departament d'anàlisi matemàtica de la Universitat de Bucarest on es produïren les primeres experiències de tractament teòric i interdisciplinaris en sentit ampli de l'objecte teatre com a espectacle. Des d'aleshores estudiem el teatre en un espai de dimensions ben diferents a com s'havia estudiat. Entre les dimensions noves hi havia la matemàtica pura i aplicada; és a dir, estructures algebraïques, teoria de la informació, cibernètica, lògica matemàtica, etc. Podem considerar, doncs, que l'aportació més important de la lingüística matemàtica, en aquest terreny, ha estat el fixar, per al drama, unes noves dimensions en el seu espai, desconegudes fins en aquest moment. El cert, però, és que d'aquestes noves dimensions encara no estem, avui, en condicions de veure gaires fruits.

Per una part el llenguatge matemàtic en les seves diferents formes, conjuntístic, probabilístic, etc. és el model més adequat de llenguatge amb el qual

és possible de presentar un aparell teòric. És a dir: a una teoria científica li exigim, d'antuvi, que sigui presentada com el llenguatge capaç de descriure un univers d'objectes, en el nostre cas una obra, o una sèrie d'obres teatrals. I aquest llenguatge ve condicionat per una sèrie d'exigències epistemològiques, d'univocitat, coherència lògica i total explicitud. Naturalment, el llenguatge que té en grau més alt aquests trets de "bon llenguatge científic és el llenguatge formal, el llenguatge matemàtic.

És per això que, en algun sentit, tot metallenguatge amb pretensions de llenguatge científic – i ho és el llenguatge que cal emprar per a elaborar una autèntica semiòtica teatral – tendeix, d'alguna manera, a la matematització. I aquesta matematització podem pensar que té dos vessants, un d'anàlisi i un altre de síntesi, un de delimitació i un altre de creació, un de tancat i un altre d'obert. De fet són els dos costats de la llei matemàtica, l'un com a resum de fets, com a expressió de regularitat i l'altre com a enunciat de futures prediccions.

En aquest sentit, podem pensar que, la matematització ens pot fornir un metallenguatge clar, senzill i econòmic per expressar unes regularitats observades pels crítics i teòrics de la literatura que havien formulat en el metallenguatge usual en la crítica literària i que té en grau reduït les característiques que enumeràvem més amunt. Tanmateix, tot i sent important, aquest tipus de matematització podem pensar que no és el fonamental. Ho és molt més aquell que anomenàvem sintètic. És a dir, aquell que ens permet de formular

la “màquina teatral”.

Anomenarem màquina teatral el metallenguatge, formal, que constitueix una teoria del teatre. Aquell sistema capaç d’enumerar frases – teoremes – que són precisament les obres de teatre. Aquell sistema que produeix, genera, uns objectes com les obres dramàtiques. Naturalment, aquests sistemes són molt complexos i avui amb prou feines podem albirar quins tipus d’estructures els donen suport. Tanmateix, cal pensar que a l’horitzó hi ha la màquina teatral, com a sistema lògic, com a sistema simulador que funciona en una situació de caixa negra.

Sota la llum d’aquestes coordenades epistemològiques que acabem de proposar, com es desenvolupa la teoria matemàtica del teatre? Ja hem deixat ben clar que no hi ha més de tres dècades d’investigacions. Afegirem que, fins avui, tan sols s’ha treballat en la primera direcció de les dues que assenyalàvem, i encara de forma fragmentària, sense intents globalitzadors. És a dir, s’ha emprat el llenguatge matemàtic com a metallenguatge per expressar certs fets, certes regularitats, que, evidentment s’havien expressat o, en tot cas, s’haurien pogut expressar, com deiem, en el llenguatge usual de la teoria literària. En aquest sentit, des del punt de vista epistemològic, l’ús del llenguatge matemàtic ja ha estat eficaç perquè ha fet guanyar en claredat, economia, simplicitat, coherència, universalitat, etc. Un enunciat del metallenguatge sovint equival a tota una pàgina d’un llenguatge literari poc clar.

En aquesta línia, unes vegades podem aplicar el càlcul de probabilitats i

altres la teoria de conjunts, la lògica d'enunciats o la teoria de la informació, amb la finalitat, sempre, d'elaborar uns microsistemes formals que serveixin per a descriure certes zones de l'àmbit teòric teatral però que són lluny encara, en conjunt, de constituir el suport teòric de l'aparell teatral. Lluny de fornir uns bons models d'allò que hem anomenat, potser de manera massa optimista, màquina teatral com a sistema generador d'obres-espectacle.

Centrat el marc general podem descriure ara algunes d'aquestes aproximacions.

D'allò que no hi ha dubte és del caràcter combinatori de l'art dramàtic. Aquest caràcter combinatori fa possible l'aplicació al teatre d'uns models matemàtics que porten aparellats unes mètriques determinades. És natural també que fos precisament en el domini de la combinatòria on trobéssim els primers precedents de tractament matemàtic del teatre. S'observà que entre les nombroses situacions dramàtiques possibles, en partir d'uns elements bàsics, tan sols se n'utilitzen unes quantes.

Ja Carlo Gozzi¹ al segle XVIII s'interessà pel drama com a combinatòria. Més recentment G. Polti en la seva obra *Les 36 situacions dramàtiques* (1934)

¹Gozzi, Carlo (Venècia 1720–1806) Escriptor italià. Promotor, juntament amb el seu germà Gasparo, de l'Accademia Letteraria dei Granelleschi (1747–62), de tendència conservadora. Autor de *Marfisa bizzarra* (1774), poema en dotze cants, de *Memorie inutili* (1797), autobiografia, i de deu *Fiabe* (1761–65) per al teatre (*Amore delle tre melarance*, *Turandot*, *L'augellin belverde*, etc), algunes de les quals han estat modernament convertides en llibrets d'òpera.

planteja la mateixa problemàtica. Tanmateix, en aquest sentit, l'obra més important és la de Souriau (Souriau, 1950) que podem considerar marca el vertader inici d'una modelització sistemàtica del teatre.

Per a Souriau les situacions són tipus d'esdeveniments. Intenta identificar les funcions dramàtiques fonamentals i després mitjançant procediments combinatoris obté les 210.141 situacions possibles. Souriau planteja la situació dramàtica a partir del conjunt de relacions possibles establertes entre personatges. És obvi que posa en evidència els aspectes estructural i funcional en l'estudi dramàtic. D'alguna manera, la seva obra és en bona mesura comparable a la de Saussure en lingüística i, sobretot, a la de V. Propp que fonamentà la teoria de la narració.

L'analogia entre aquests tres autors, Saussure, Propp i Souriau, recolza el fet evident que els tres introdueixen un conjunt de nocions de base que donen suport a tot el nou sistema. Després, els seus seguidors ho desenvoluparan. En el cas de Saussure, de la lingüística, Trubetzkoy, Hjelmslev, Harris, Chomsky, etc. En el cas de Propp, teoria de la literatura, Bremon, Barthes, Todorov, etc. I en el cas de Souriau, teoria del teatre, P. Ginester, S. Marcus, M. Dinu, etc.

Souriau parlava sempre de l'àlgebra del teatre. I és que el drama és fàcilment concebible com una àlgebra en el sentit de combinatòria – veure (Dinu, 1984). Uns elements, unes operacions, unes propietats i unes estructures. En la revisió de l'obra de Souriau que realitzà Marcus (1968–69) en

un seminari de lingüística matemàtica trobem el punt exacte que representa l'origen de les investigacions matemàtiques en el domini del teatre.

La naturalesa combinatòria del drama és evident. Imaginem un espectador que assisteix a un espectacle teatral; és obvi que observarà com, a l'escenari, entren i surten actors, els pot reconèixer successivament i, a més a més, pot distingir entre dos personatges diferents. Aquest conjunt d'informació que un espectador ideal té al seu abast podem considerar-lo com l'*estructura escènica d'una peça*.

Evidentment, aquest espectador ideal es troba davant la impossibilitat de fer distincions en l'estructura espacial de la peça; per exemple, en el decorat i en el vestit o en el desplaçament dels personatges, moviment del cos dels actors, etc., i fins i tot l'estructura temporal dels sons, és a dir, allò que podem escoltar en l'escena, sons, entonacions, paraules, etc. A un nivell merament intuitiu resulta força difícil fer distincions, d'establir unitats en aquestes estructures que acabem d'assenyalar, si més no per a l'espectador ideal.

L'estructura escènica d'una peça és, doncs, el conjunt d'informació que està en condicions de rebre i interpretar un espectador ideal situat com a observador, i que fa ús tan sols de la intuïció, davant d'una obra dramàtica.

Una bona forma de representar aquesta estructura escènica d'una peça de teatre és la representada per Marcus com a matriu binària que té un nombre de fileres m igual al nombre de personatges i un nombre n de columnes igual

al d'escenes (Marcus, 1967f). Així, esquemàticament, donats els m personatges $p_1, p_2, \dots, p_m$ i les n escenes $S_1, S_2, \dots, S_n$ la matriu representativa de l'estructura escènica es formularia com:

$$PERSONATGES \left\{ \begin{array}{c} \overbrace{\left(\begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \ddots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{array} \right)}^{ESCENES} \end{array} \right.$$

El símbol a_{ij} , intersecció de la fila i i de la columna j és 1 si el personatge p_i apareix a l'escena S_j . En tot altre cas a_{ij} és zero (i.e., la seva “funció característica”).

$$a_{ij} = \begin{cases} 0 & \text{si } p_i \notin S_j \\ 1 & \text{si } p_i \in S_j \end{cases}$$

Naturalment, aquest mètode es recolza, ja ho hem dit, en la pura intuïció d'un espectador ideal que tendirà a segmentar l'obra en unes unitats naturals. Aquestes unitats són les *escenes*. Per a aquest espectador, i per a nosaltres, una escena vindria definida operacionalment com el segment d'espectacle que té lloc entre dos canvis successius de la configuració dels personatges. Considerarem el drama com una seqüència d'escenes.

Com a escena acceptarem el sentit, més o menys tradicional, com el segment de l'obra dramàtica que té una longitud tal que la configuració dels personatges presents no canvia. Si ens mantenim dins dels límits d'aquesta acceptació, considerant l'aparició o desaparició de com a mínim un personatge com a la condició necessària i suficient per a afirmar que s'ha produït un canvi en la situació escènica, aleshores, despulladaa del seu contingut verbal (que conté les rèpliques dels personatges i les indicacions escenogràfiques) l'obra de teatre se'ns presenta com una cadena de situacions relacionada cadascuna d'elles con una configuració de personatges. Del caràcter biunívoc d'aquesta relació es desprèn que cada vegada que els mateixos personatges es troben en escena es tracta de la mateixa situació escènica.

Seguint el model de Marcus a (Marcus, 1970b, capítol VIII), Păun defineix a (Păun, 1976b) alguns llenguatges que modelen i sintetitzen les tendències recursives de l'acció dramàtica a nivell de les configuracions dels personatges.

Considerem una obra de teatre O . Denotarem els seus *personatges* com $p_1, p_2, \dots, p_n$ i les seves *escenes* com $S_1, S_2, \dots, S_r$. Qualsevol subconjunt del conjunt $V_1 = \{p_1, \dots, p_n\}$ es denomina *configuració de personatges*. Denotarem amb $C(S_i)$ la configuració de personatges que té lloc en l'escena S_i i amb V_2 el conjunt de totes les configuracions de personatges de l'obra O , és a dir, $V_2 = \{C(S_i) \mid 1 \leq i \leq r\}$. Amb $p \in S_k$ o $C \in S_k$, on $p \in V_1$ i $C \in V_2$, significarem la presència del personatge p , o la presència de la configuració C en l'escena S_k .

Definirem els llenguatges L_i , $1 \leq i \leq 6$ sobre V_1 o V_2 usant el següent model: sigui L_i $1 \leq i \leq 6$ un conjunt finit de cadenes de la forma $a_{k1}a_{k2}$, per a a_{k1}, a_{k2} pertanyents a V_1 o a V_2 .

A partir dels llenguatges anterior L_i , considerarem els conjunts

$$L'_i = pdret(L_i, L_i),$$

amb $1 \leq i \leq 6$ i on $pdret$ és l'operació “dòmino o prolongació a la dreta-- veure la Definició 6.13.1 de la pàgina 230.

Aleshores,

$$L'_i = \{a_{k1}a_{k2} \dots a_{kr} \mid r \geq 2, a_{kj}a_{k,j+1} \in L_i, 1 \leq j \leq r-1\},$$

amb $1 \leq i \leq 6$; i r és el nombre total d'escenes, k és el nombre total de personatges i j és l'escena específica. Brainerd va provar a (Brainerd, 1972) que L'_i , $1 \leq i \leq 6$ és un llenguatge regular. Cada llenguatge L'_i modela una característica de l'acció dramàtica.

- Sigui $L_1 = \{pp' \mid \exists S_k, S_{k+1} \in O; p \in S_k, p' \in S_{k+1}\}$.

L_1 és el conjunt de tots els parells de personatges que estan presents a l'escenari en escenes contigües.

El llenguatge L'_1 és el llenguatge definit a (Marcus, 1970b).

- $L_2 = \{pp' \mid \exists S_k, S_{k+1} \in O; p \in S_k, p \in S_{k+1}, p' \notin S_k, p' \in S_{k+1}\}$.

L_2 és el conjunt de tots els parells de personatges que estan presents

a l'escenari i un, i solament un d'ells, estava també present en l'escena anterior.

La infinitud d' L'_2 implica l'existència d'un conjunt de parells de la forma $(p_{i1}, p_{i2}), (p_{i2}, p_{i3}), \dots, (p_{ik}, p_{i1})$. Per tant, si L'_2 és infinit implica que la continuïtat de les relacions entre personatges no és accidental.

- $L_3 = \{pp' \mid \exists S_k, S_{k+1} \in O; p \in S_k, p \notin S_{k+1}, p' \notin S_k, p' \in S_{k+1}\}$.

L_3 és el conjunt de tots els parells de personatges que s'eviten l'un a l'altre, però això succeeix en escenes consecutives, en el moment precís del canvi.

La infinitud d' L'_3 implica l'existència d'un conjunt de parells de la forma $(p_{i1}, p_{i2}), (p_{i2}, p_{i3}), \dots, (p_{ik}, p_{i1})$. Per tant, si L'_3 és infinit les no-coincidències entre personatges no són accidentals.

- $L_4 = \{CC' \mid \exists S_k, S_{k+1} \in O; C = S_k, C' = S_{k+1}\}$.

L_4 és el conjunt de configuracions de personatges que estan presents a l'escenari en escenes contigües.

L'_4 modela l'evolució de l'acció en connexió amb la configuració dels personatges. Si en l'obra existeixen dues escenes amb la mateixa configuració de personatges, aleshores L'_4 és un llenguatge infinit; si no és així, L'_4 és finit. D'aquesta manera, queda subratllada la importància de les configuracions i dels personatges d'aquestes configuracions que es repeteixen en l'obra.

- $L_5 = \{CC' \mid CC' \in L_4, C \cap C' \neq \emptyset\}$.

L_5 és el conjunt de configuracions de personatges que tenen en comú algun personatge.

L'_5 modela la continuïtat de l'acció al nivell de les configuracions.

- $L_6 = \{CC' \mid CC' \in L_4, C \cap C' = \emptyset\}$.

L_6 és el conjunt de configuracions de personatges que no tenen en comú cap personatge.

L'_6 modela la continuïtat de l'acció al nivell de les configuracions.

Pel fet que els drames tràgics estan basats en una forta confrontació entre personatges, podem suposar que per a aquest tipus de drames L'_2 i L'_5 són especialment representatius. Per altra part, els drames còmics, les comèdies, estan basats en el fet que alguns personatges no són conscients d'algunes coses que els ha succeït amb altres personatges; per això, L'_3 i L'_6 són molt complicats, mentre que L'_2 i L'_5 són més simples.

Els llenguatges L'_i , $1 \leq i \leq 6$, són llenguatges contextuais externs unilaterals programats i, per tant, llenguatges contextuais externs unilaterals amb selecció – veure la Secció 12.2, en general, i la Definició 12.2.1 de la pàgina 479. Les següents gramàtiques generen L'_i .

$$G_i = (V_1, L_i, \{(\lambda, a_i) \mid a_i \in V_1\}, \varphi) \quad 1 \leq i \leq 3,$$

$$G_i = (V_2, L_i, \{(\lambda, a_i) \mid a_i \in V_2\}, \varphi) \quad 4 \leq i \leq 6,$$

on

$$\begin{aligned}\varphi(\lambda) &= \emptyset, \\ \varphi(xa_i) &= \{(\lambda, a_j) \mid a_i a_j \in L_i\} \text{ per a } x \in V_{1,2}^*, a_i \in V_{1,2}.\end{aligned}$$

Fàcilment podem convertir-les en gramàtiques contextuais externes unilaterals amb selecció programades si considerem

$$U = \{[a_i a_j, (\lambda, a_k)], [(\lambda, a_k), (\lambda, a_r)] \mid a_i a_j, a_k a_r \in\}.$$

També, per a cada personatge p podem considerar els números

$$n_k(p) = \text{card}\{p_i p_j \in L_k \mid p_i = p \vee p_j = p\} \text{ per a } k = 1, 2, 3, \text{ i}$$

$$I_m(p) = \frac{n_m(p)}{n_1(p)} \text{ per a } m = 2, 3.$$

I_2 mesura, o bé la contribució de p a la continuïtat de l'acció dramàtica o bé la seva neutralitat; per altra part, I_3 mesura o bé la duplicitat o bé la ignorància de p .

Si considerem les següents mesures de la complexitat – veure la Definició 8.2.1 de la pàgina 249 – de la gramàtica $G = (V, B, C, \varphi)$:

- $Bas(G) = \text{card}(L)$;
- $Con(G) = \text{card}(C)$;
- $Fi(G) = \text{card}\{M \in \mathcal{P}(C) \mid \exists y \in V^* \text{ tal que } \varphi(y) = M\}.$

Per a un llenguatge L definim

$$K(L) = \min\{K(G) \mid L = L(G) \text{ per a } K \in Bas, Com, Fi\}.$$

Les mesures Bas , Con , Fi poseeixen significacions concretes per a aquests llenguatges. Per exemple, $Bas(L'_i) = card(L_i)$, ja que $L_i \subseteq L'_i$.

Per a una obra O podem definir un índex de continuïtat de l'acció dramàtica, un índex de la continuïtat de les relacions entre personatges $IR(O)$; i un índex de la intensitat del conflicte entre personatges $IC(O)$. Aleshores

$$IR(O) = \frac{Bas(L'_2)}{Bas(L'_1)} \qquad IC(O) = \frac{Bas(L'_3)}{Bas(L'_1)}$$

Les mesures IR i IC poden ser computades també per a cada acte de l'obra, per a poder observar com evolucionen els conflictes.

Exemple 15.1.1 *Cimera internacional contra el terrorisme.* Suposem que viatgem en el temps i ens situem en els dies previs al conflicte libio-americà per les suposades implicacions del règim de Gaddafi en el terrorisme mundial. Es celebra una cimera internacional per tal d'evitar el conflicte armat on hi participen Bush (b), Gaddafi (g), Mitterrand (m) i Reagan (r);



i les escenes S_1, S_2, S_3, S_4 definides per:

S_1 : Bush, Reagan i Mitterrand es troben per a parlar sobre el terrorisme. Bush i Reagan volen castigar Gaddafi envaint Líbia. Mitterrand proposa convocar una reunió amb tots els implicats. Bush i Reagan accepten. Per tant, $C(S_1) = \{b, m, r\}$.

S_2 : Mitterrand comunica a Gaddafi la intenció de convocar una reunió sobre el terrorisme. Gaddafi refusa l'oferta i proclama la seva intenció de lluitar fins a la mort. Per tant, $C(S_2) = \{g, m\}$.

S_3 : Mitterrand comunica a Bush i Regan la resposta de Gaddafi. Aquests manifesten a Mitterrand la seva definitiva intenció d'envair Líbia. Per tant, $C(S_3) = \{b, m, r\}$.

S_4 : Mitterrand, desolant, reflexiona sobre la condició humana i les seves mancances. Per tant, $C(S_4) = \{m\}$.

En tal situació,

$$V_1 = \{b, g, m, r\}$$

i,

$$V_2 = \{C(S_i) \mid 1 \leq i \leq 4\} = \{\{b, m, r\}, \{g, m\}, \{b, m, r\}, \{m\}\}.$$

La matriu binària associada a l'obra dramàtica serà:

$$\begin{pmatrix} & \mathbf{S_1} & \mathbf{S_2} & \mathbf{S_3} & \mathbf{S_4} \\ \mathbf{b} & 1 & 0 & 1 & 0 \\ \mathbf{g} & 0 & 1 & 0 & 0 \\ \mathbf{m} & 1 & 1 & 1 & 1 \\ \mathbf{r} & 1 & 0 & 1 & 0 \end{pmatrix}$$

- L_1 són tots els parells de personatges que estan presents a l'escenari en escenes contigües.

$$L_1 = \{pp' \mid \exists S_k, S_{k+1} \in O; p \in S_k, p' \in S_{k+1}\} =$$

$$= \{bg, bm, gb, gm, gr, mb, mg, mr, rg\}.$$

$$L'_1 = L_1 \cup \{bgbg \dots\}; \text{card}(L'_1) = \infty.$$

- L_2 són tots els parells de personatges que estan presents a l'escenari i un, i solament un d'ells, estava també present en l'escena anterior.

$$L_2 = \{pp' \mid \exists S_k, S_{k+1} \in O; p \in S_k, p' \in S_{k+1}, p' \notin S_k\} = \{mb, mg, mr\}.$$

$$L'_2 = L_2; \text{card}(L'_2) = 3.$$

- L_3 són tots els parells de personatges que s'eviten l'un a l'altre, però això succeeix en escenes consecutives, en el moment precís del canvi.

$$L_3 = \{pp' \mid \exists S_k, S_{k+1} \in O; p \in S_k, p' \notin S_{k+1}, p' \notin S_k\} = \{bg, gb, gr, rg\}.$$

$$L'_3 = L_3 \cup \{bgb, bgbg \dots, grg, grgr \dots\}; \text{card}(L'_3) = \infty.$$

L'_3 és el llenguatge dels personatges que s'eludeixen l'un a l'altre. La infinitud d' L'_3 ve donada per l'existència d'un conjunt de parells de la forma $(b, g), (g, b)$ o bé $(g, r), (r, g)$. Per tant, si L'_3 és infinit, les no-coincidències entre els personatges (*Bush, Gaddafi*) i (*Gaddafi, Reagan*) no són accidentals.

- L_4 són tots els parells de configuracions de personatges presents a l'escenari en escenes contigües.

$$L_4 = \{CC' \mid \exists S_k, S_{k+1} \in O; C = S_k, C' = S_{k+1}\} =$$

$$= \{\{b, m, r\}\{g, m\}, \{g, m\}\{b, m, r\}, \{b, m, r\}\{m\}\}.$$

$$L'_4 = L_4 \cup \{\{b, m, r\}\{g, m\}\{b, m, r\} \dots\}; \text{card}(L'_4) = \infty.$$

L'_4 modela l'evolució de l'acció en connexió amb la configuració dels personatges. En l'obra existeixen dues escenes amb la mateixa configuració ($S_1 = S_3 = \{b, m, r\}$); això fa que L'_4 sigui un llenguatge infinit. D'aquesta manera, queda subratllada la importància de la configuració i dels personatges de la configuració $\{Bush, Mitterrand, Reagan\}$ que es repeteix en l'obra.

- L_5 són tots els parells de configuracions de personatges que tenen en comú algun personatge.

$$\begin{aligned} L_5 &= \{CC' \mid CC' \in L_4, C \cap C' \neq \emptyset\} = \\ &= \{\{b, m, r\}\{g, m\}, \{g, m\}\{b, m, r\}, \{b, m, r\}\{m\}\}. \end{aligned}$$

$$L'_5 = L_5 \cup \{\{b, m, r\}\{g, m\}\{b, m, r\} \dots\}, \text{card}(L'_5) = \infty.$$

L'_5 modela la continuïtat de l'acció al nivell de les configuracions. És el personatge de *Mitterrand* qui contribueix a aquesta continuïtat.

- L_6 són tots els parells de configuracions de personatges que no tenen en comú cap personatge.

$$L_6 = \{CC' \mid CC' \in L_4, C \cap C' = \emptyset\} = \{\emptyset\}.$$

$$L'_6 = \emptyset, \text{card}(L'_6) = 0.$$

No hi ha cap trencament en la continuïtat de l'obra dramàtica.

També, per a cada personatge p podem considerar els nombres

$$n_k(p) = \text{card}\{p_i p_j \in L_k \mid p_i = p \vee p_j = p\}, \text{ per a } 1 \leq k \leq 3.$$

$$\begin{aligned} n_1(b) &= 4, & n_1(g) &= 6, & n_1(m) &= 6, & n_1(r) &= 4; \\ n_2(b) &= 1, & n_2(g) &= 1, & n_2(m) &= 3, & n_2(r) &= 1; \\ n_3(b) &= 2, & n_3(g) &= 4, & n_3(m) &= 0, & n_3(r) &= 2. \end{aligned}$$

$$I_m(p) = \frac{n_m(p)}{n_1(p)} \text{ per a } k = 2, 3.$$

$$\begin{aligned} I_2(b) &= 0.25, & I_2(g) &= 0.16, & I_2(m) &= 0.50, & I_2(r) &= 0.25; \\ I_3(b) &= 0.50, & I_3(g) &= 0.66, & I_3(m) &= 0.00, & I_3(r) &= 0.50. \end{aligned}$$

I_2 mesura, o bé la contribució de p a la continuïtat de l'acció dramàtica o bé la seva neutralitat. Qui més contribueix a la seva continuïtat és *Mitterrand* – ja que $I_2(m) = 0.50$ que és el valor més gran –, i qui més contribueix a la seva neutralitat és *Gaddafi* – ja que $I_2(g) = 0.16$ que és el valor més petit.

Per altra part, I_3 mesura o bé la complicitat o bé la ignorància de p pel que fa a la intensitat del conflicte. Qui més contribueix a aquesta intensitat en el conflicte és *Gaddafi* – ja que $I_3(g) = 0.66$ que és el valor més gran – i qui menys *Mitterrand* – ja que $I_3(m) = 0.00$ que és el valor més petit.

També podem definir un índex de continuïtat de l'acció dramàtica per al conjunt de l'obra $IR(O)$, un índex de la continuïtat de les relacions entre

personatges (on $Bas(L'_i) = card(L_i)$ ja que $L_i \subset L'_i$); i un índex de la intensitat del conflicte entre personatges

$$IR(O) = \frac{Bas(L'_2)}{Bas(L'_1)} = \frac{3}{10} = 0.3 \quad IC(O) = \frac{Bas(L'_3)}{Bas(L'_1)} = \frac{4}{10} = 0.4$$

Les mesures IR i IC poden ser computades també per a cada acte de l'obra, per a poder observar com evolucionen els conflictes.

15.2 Llenguatges associats al còmic

Imaginem un lector que llegeix un còmic; és obvi que observarà com, en el còmic, apareixen i desapareixen personatges, els pot reconèixer successivament i, a més a més, pot distingir entre dos personatges diferents. Aquest conjunt d'informació que un lector ideal té al seu abast podem considerar-lo com l'*estructura “vinyètica” d'un còmic*.

Evidentment, aquest lector ideal es troba davant la impossibilitat de fer distincions en l'estructura espacial simulada en el còmic; per exemple, en el decorat i en el vestit o en el desplaçament simulat dels personatges, moviment del cos simulat en els personatges, etc., i fins i tot l'estructura lingüística expressada en el text (les diferents llengües expressades). A un nivell merament intuitiu resulta força difícil fer distincions, d'establir unitats en aquestes estructures que acabem d'assenyalar, si més no per a l'espectador ideal.

L'estructura “vinyètica” d'un còmic és, doncs, el conjunt d'informació que

està en condicions de rebre i interpretar un lector ideal situat com a observador, i que fa ús tan sols de la intuïció, davant de la lectura d'un còmic.

Una bona forma de representar aquesta estructura “vinyètica” d'un còmic pot ser també, per analogia, la representada per Marcus pel cas de l'estructura escènica del teatre, en aquest cas, com a matriu binària que té un nombre de fileres m igual al nombre de personatges i un nombre n de columnes igual al de vinyetes. Així, esquemàticament, donats els m personatges $p_1, p_2, \dots, p_m$ i les n vinyetes $S_1, S_2, \dots, S_n$ la matriu representativa de l'estructura “vinyètica” es formularia com:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \ddots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

El símbol a_{ij} , intersecció de la fila i i de la columna j és 1 si el personatge p_i apareix a la vinyeta S_j . En tot altre cas a_{ij} és zero (i.e., la seva “funció característica”).

$$a_{ij} = \begin{cases} 0 & \text{si } p_i \notin S_j \\ 1 & \text{si } p_i \in S_j \end{cases}$$

Igual que en el cas de les obres dramàtiques, també podem aplicar aquest

mètode al llenguatge dels còmics. Aquest mètode es recolzaria en la segmentació d'un còmic en unes unitats naturals (que ja vindria donada pel mateix dibuixant). Aquestes unitats serien les *vinyetes*. Considerarem el còmic com una seqüència de vinyetes.

Considerem un còmic O . Denotarem els seus *personatges* com $p_1, p_2, \dots, p_n$ i les seves *vinyetes* com $S_1, S_2, \dots, S_r$. Qualsevol subconjunt del conjunt $V_1 = \{p_1, \dots, p_n\}$ es denomina *configuració de personatges*. Denotarem amb $C(S_i)$ la configuració de personatges que té lloc en la vinyeta S_i i amb V_2 el conjunt de totes les configuracions de personatges del còmic O , és a dir, $V_2 = \{C(S_i) \mid 1 \leq i \leq r\}$. Amb $p \in S_k$ o $C \in S_k$, on $p \in V_1$ i $C \in V_2$, significarem la presència del personatge p , o la presència de la configuració C en la vinyeta S_k .

En aquest punt podem considerar els mateixos llenguatges que en el cas de l'obra dramàtica.

Exemple 15.2.1 *Pàgina 6 del còmic “Asterix el gal”.* Considerem el còmic “Asterix el gal” en les seves versions catalana², romanesa³ i anglesa⁴. Per raons d'espai ens centrarem només en la pàgina 6 del còmic.

²*Asterix el gal* de Goscinny i Uderzo, Ed. Grijalbo/Dargaud S.A., Barcelona, 1983

³*Asterix eroul galilor* de Goscinny i Uderzo, Ed. Egmont România, București, 1994

⁴*Asterix the gaul* de Goscinny i Uderzo, Ed. del Prado S.A., Madrid, 1984



Versió catalana



Versió romanesa



Versió anglesa

Imaginem un lector que llegeix un còmic; és obvi que, malgrat no posseir cap tipus de coneixement sobre l'idioma en qüestió, observarà com, en el còmic, apareixen i desapareixen personatges, els pot reconèixer successivament i, a

- L_1 són tots els parells de personatges que estan presents al còmic en vinyetes contigües.

$$L_1 = \{12, 13, 14, 15, 16, 17, 21, 23, 24, 25, 27, 31, 32, 41, 42, 51, 52, 61, 62, 78, 87\}.$$

$$L'_1 = L_1 \cup \{1221 \dots, 1331 \dots, 1441 \dots\}; \text{card}(L'_1) = \infty.$$

- L_2 són tots els parells de personatges que estan presents al còmic i un, i solament un d'ells, estava també present en la vinyeta anterior.

$$L_2 = \{23, 24, 25, 26, 78\}.$$

$$L'_2 = L_2; \text{card}(L'_2) = 5.$$

- L_3 són tots els parells de personatges que s'eviten l'un a l'altre, però això succeeix en vinyetes consecutives, en el moment precís del canvi de vinyeta.

$$L_3 = \{13, 14, 15, 16, 17, 27, 31, 41, 51, 61\}.$$

$$L'_3 = L_3 \cup \{1331 \dots, 1441 \dots, 1551 \dots, 1661 \dots\}; \text{card}(L'_3) = \infty.$$

L'_3 és el llenguatge dels personatges que s'eludeixen l'un a l'altre. La infinitud d' L'_3 ve donada per l'existència d'un conjunt de parells de la forma $(1, 3), (3, 1)$ o bé $(1, 4), (4r, 1)$ o bé $(1, 5), (5, 1)$ o bé $(1, 6), (6, 1)$.

Per tant, si L'_3 és infinit, les no-coincidències entre 1, *Julius Pompilius* i els quatre *soldats romans* no són accidentals.

- L_4 són tots els parells de configuracions de personatges presents a l'escenari en escenes contigües.

$$L_4 = \{\{12\}\{23456\}, \{23456\}\{12\}, \{12\}\{7\}, \{78\}\{78\}, \{78\}\{78\}\}$$

$$L'_4 = L_4 \cup \{\{12\}\{23456\}\{12\} \dots\}; \text{card}(L'_4) = \infty.$$

L'_4 modela l'evolució de l'acció en connexió amb la configuració dels personatges. En l'obra es repeteixen configuracions o grups de configuracions ($(S_1S_2 = S_3S_4 = \{12\}\{23456\}; S_7S_8 = S_8S_9 = \{78\}\{78\})$); això fa que L'_4 sigui un llenguatge infinit. D'aquesta manera, queda subratllada la importància de la configuració i dels personatges de la configuració *Julius Pompilius*, *Caius Bonus i els quatre soldats romans* i *Astèrix i Obèlix* que es repeteixen en l'obra.

- L_5 són tots els parells de configuracions de personatges que tenen en comú algun personatge.

$$L_5 = \{\{12\}\{23456\}, \{23456\}\{12\}, \{78\}\{78\}, \{78\}\{78\}\}.$$

$$L'_5 = L_5 \cup \{\{12\}\{23456\}\{12\}, \dots\}; \text{card}(L'_5) = \infty.$$

L'_5 modela la continuïtat de l'acció al nivell de les configuracions. Observem que hi ha dos blocs de personatges que, per separat contribueixen a aquesta continuïtat. Per una part *Julius Pompilius*, *Caius Bonus i els quatre soldats romans* i, per altra part, *Astèrix i Obèlix*.

- L_6 són tots els parells de configuracions de personatges que no tenen en comú cap personatge.

$$L_6 = \{\{12\}\{7\}\}.$$

$$L'_6 = \{\{12\}\{7\}\}; \text{card}(L'_6) = 1.$$

Observem que, efectivament, hi ha un trencament en la continuïtat de l'obra dramàtica. Correspon al pas de la vinyeta S_5 a la S_6 .

Capítol 16

Caracterització de llenguatges associats als “sistemes d’accions” mitjançant gramàtiques contextuals

L’operació d’adjunció de contextos, o més generalment encara, l’operació d’adjunció de cadenes, que depèn d’una part o del total de la cadena, té també, a un nivell conceptual més general, una significació física i metafísica: si acceptem que el temps té una direcció i que, per tant, existeix (alguna) causalitat en el món, aleshores la història de qualsevol procés, natural o social, pot ser descrita com a una seqüència d’*esdeveniments elementals* que estan relacionats, d’una manera més o menys precisa, amb la influència *del passat* sobre *el futur*. Aquestes especulacions ens proporcionen un determinat model de llenguatge formal relacionat amb diferents processos, especialment de l’àrea social. És

el cas dels treballs de Marcus – relacionats amb el folklore (Marcus, 1978) –, de Nowakowska – relacionats amb els sistemes d'accions (Nowakowska, 1973) – i de Păun, relacionats amb certs processos econòmics (Păun, 1980a; Păun, 1983).

El fet d'afegir un nou esdeveniment (o diversos esdeveniments relacionats) a una seqüència d'esdeveniments és, exactament, la manera de treballar d'una gramàtica contextual com hem pogut veure al Capítol 4.

Així doncs, moltes activitats humanes i socials poden ser descomposades, de manera natural, en un conjunt finit d'accions elementals. D'aquesta manera obtenim un vocabulari de l'activitat, on qualsevol desenvolupament particular de l'activitat és una cadena o seqüència d'accions.

El conjunt de cadenes d'accions correctes és un llenguatge que caracteritza l'activitat. Generalment, l'acció següent depèn del desenvolupament previ del procés.

Definició 16.0.1 Nowakowska (Nowakowska, 1973) defineix un *llenguatge d'accions amb estructures semàntiques* com una quintupla

$$(D, L, \varphi, R, \rho),$$

on

- D és un conjunt finit d'accions;
- $L \subset D^*$ és un conjunt de cadenes d'accions;

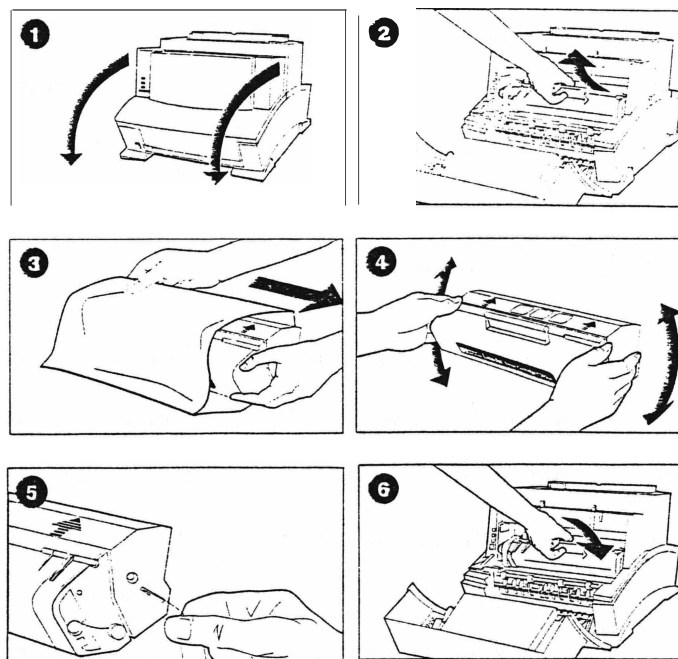
- $\varphi : D \longrightarrow \mathbb{N}$ és una funció on el valor de $\varphi(d)$ pot ser interpretat com la duració de l'acció $d \in D$;
- R és el conjunt de resultats de les accions de D i
- ρ és una relació, $\rho \subset L \times R \times \mathbb{N}$, que associa un resultat d' R amb l'acció d' L en un moment d' $\mathbb{N}$.

El llenguatge L és donat *a priori* i pot ser investigat mitjançant els models i mètodes analítics desenvolupats per Marcus a (Marcus, 1967a). D'ara endavant, ignorarem el component semàntic del model; el nostre propòsit és el de determinar la sintaxi d' L , la sintaxi d'aquells llenguatges que poden ser considerats llenguatges d'accions. En altres paraules, coneixent les regles d'una acció, volem construir una gramàtica que generi exactament les cadenes correctes d'accions; volem “la gramàtica de l'acció”.

En la nostra aproximació, les accions elementals són reduïdes a símbols simples, l'únic significat dels quals consisteix en la seva mutua relació i el seu emplaçament dins del model formal del sistema d'accions en qüestió. El nostre interès és el de construir un mecanisme de la teoria de llenguatges formals (una gramàtica o un autòmat) que simuli el sistema, de manera que el llenguatge que es correspongui amb aquest mecanisme sigui, exactament, el conjunt de cadenes d'accions correctes respecte a algunes “regles d'acció” coneegudes. Aquestes regles poden ser donades *a priori* – fins i tot abans del sistema que existeix en aquell moment –, o poden ser inferides (“apreses”) per un individu que observa l'evolució del sistema.

Seguidament considerarem alguns exemples per tal de fer més clares les definicions posteriors.

Exemple 16.0.2 *Canviar el cartutx de tóner d'una impressora làser.* Suposem que un un profà en la matèria vol escriure una guia d'instruccions d'instal·lació de cartutxos de tóner per a una impressora làser determinada. Després d'observar un gran nombre de tècnics que duen a terme aquest procés s'adona de què l'acció en qüestió es desenvolupa de la següent manera:



- 1) el tècnic obre la tapa anterior de la impressora làser (a_1),
- 2) extreu el cartutx de tóner a canviar de la impressora (a_2),

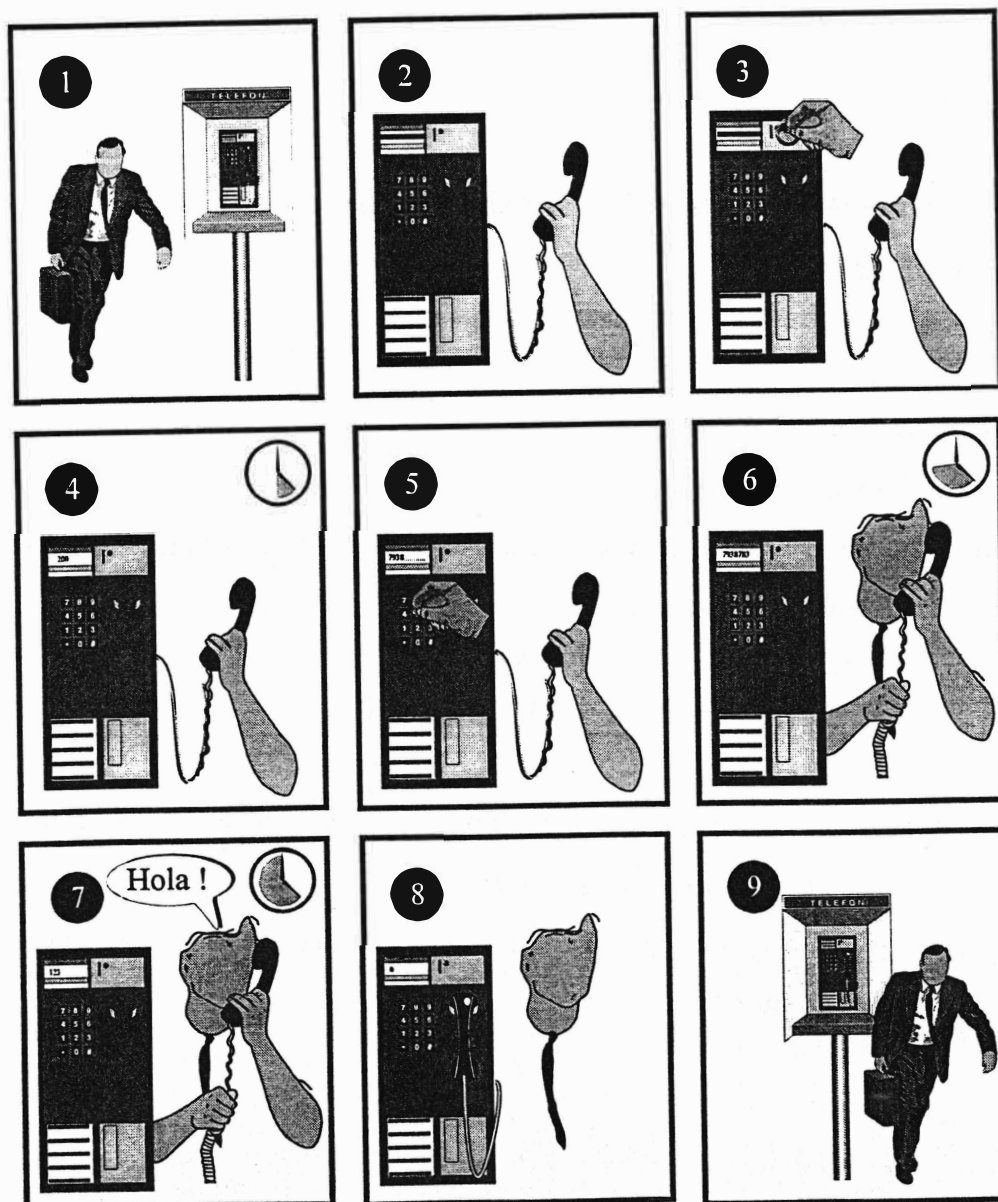
- 3) obre la bossa de protecció i aïllament del cartutx de tóner nou i n'extreu el seu contingut (a_3),
- 4) sacseja el cartutx de tóner nou durant un interval de temps n indeterminat, per tal que el tóner es distribueixi uniformement (a_4),
- 5) n'extreu la cinta plàstica que bloqueja la sortida del tóner (a_5) i, finalment,
- 6) col·loca el nou cartutx de tóner dins la impressora làser i tancar la tapa anterior (a_6).

La persona en qüestió no és completament conscient de la significació d'algunes de les accions que realitza el tècnic. Constata, però, que el procés en qüestió sempre es desenvolupa de la mateixa manera i, per tant, redueix l'acció a una seqüència d'accions de la forma

$$a_1 a_2 a_3 a_4^n a_5 a_6; \quad n \geq 1,$$

on els símbols a_i són accions sense cap significat específic. De qualsevol manera, solament estaran permesos parells consecutius de la forma $a_1 a_2$, $a_2 a_3$, $a_3 a_4$, $a_4 a_4$, $a_4 a_5$, $a_5 a_6$ i on l'acció a_4 pot ser arbitràriament repetida. Aquestes són les regles de l'acció.

Exemple 16.0.3 *L'ús de cabines telefòniques a la via pública.* Suposem que un alienígena, extraordinàriament intel·ligent, està interessat en establir el model de la conducta humana. Necessita modelar algunes accions molt simples i comença per estudiar l'ús que els humans fan de les cabines telefòniques a la via pública. Després d'observar un gran nombre de persones s'adona de què l'acció en qüestió es desenvolupa de la següent manera:



- 1) un ser humà localitza un cabina telefònica a la via pública (a_1);
- 2) despenja el microtelèfon (part del telèfon on van allotjats el micròfon i l'auricular, que se separa del telèfon i que hom manté amb la mà en parlar) (a_2);
- 3) introdueix una moneda (a_3);
- 4) espera un interval de temps n que, per ser algunes vegades molt llarg, sembla il·limitat (a_4);
- 5) pitja t tecles numèriques, corresponents al número a marcar (a_5);
- 6) espera un interval de temps m il·limitat amb l'auricular del microtelèfon enganxat a l'orella (a_6);
- 7) parla a través del micròfon del microtelèfon durant un interval de temps p (a_7);
- 8) penja el microtelèfon (a_8); i, finalment,
- 9) s'allunya de la cabina telefònica (a_9).

L'alienígena no és conscient de la significació de les accions del ser humà. Constatada, però, que l'acció en qüestió sempre es desenvolupa de la mateixa manera (no tindrem en compte telèfons defectuosos o persones que usen indegudament el telèfon), i, per tant, redueix l'acció a una seqüència de la forma

$$a_1 a_2 a_3 a_4^n a_5^t a_6^m a_7^p a_8 a_9; \quad n, m \geq 1, \quad p \geq 0, \quad t \text{ fix},$$

on els símbols a_i són gestos sense cap significació. De qualsevol manera, solament estaran permesos parells consecutius de la forma $a_1a_2, a_2a_3, a_3a_4, a_4a_4, a_4a_5, a_5a_5, a_5a_6, a_6a_6, a_6a_7, a_7a_7, a_7a_8, a_8a_9$ i on les accions a_4, a_6, a_7 , poden ser arbitràriament repetides, però l'acció a_5 necessàriament ha de repetir-se t vegades. Aquestes són les regles de l'acció.

Aleshores, seguint el model proposat per Păun a (Păun, 1979a, pàg. 35), obtenim el següent model formal del sistema d'accions:

Definició 16.0.2 Un *sistema d'accions simple*, abreujat *SAS*, és una quàdrupla

$$\sigma = \{A, A_0, A_f, \varphi\}$$

on

- A és un conjunt finit no buit, els elements del qual es denominen “accions (elementals)”;
- $A_0, A_f \subset A$, són conjunts no buits que contenen, respectivament, les “accions inicials” i les “accions finals”; i
- $\varphi : A^* \longrightarrow \mathcal{P}(A) = 2^A$ és la funció “acció següent”.

Suposarem que les accions tenen la mateixa duració encara que, d'una manera anàloga, podríem considerar el cas en què les duracions són arbitràries. No obstant, ja que mitjançant φ podem expressar la duració, només considerarem el primer cas. En aquest context, fixem-nos en què l'acció a_5 de l'Exemple

16.0.3 es repeteix exactament t vegades i, per tant, pot ser interpretat com la duració d'una acció a_5^1 , que pot reemplaçar a_5 .

El conjunt de “processos correctes (cadena o seqüència d'accions) respecte a un SAS σ com l'anterior és:

$$L(\sigma) = \{a_1 \dots a_n \mid n \geq 1, a_1 \in A_0, a_n \in A_f, a_i \in \varphi(a_1 \dots a_{i-1}), 2 \leq i \leq n\}$$

A (Păun, 1979a) també es defineixen els sistemes d'accions compostos i compostos paral·lels, els quals modelen la situació que apareix quan hem de considerar més d'un participant en l'acció.

Per tant, en qualsevol cas, comencem amb accions del conjunt A_0 i continuem segons ens imposi la funció φ , de manera que un procés és completat si acaba amb una acció del conjunt A_f .

Podem observar que, si ignorem aquesta última condició (és a dir, si la component A_f del model és ignorada), aleshores obtenim exactament una gramàtica contextual (no restringida) amb selecció on $V = A$ i $B = A_0$, que treballa en el mode extern. En la forma completa, qualsevol sistema d'accions simple es correspon amb una gramàtica contextual amb derivació marcada – veure la Secció 11.5.

Així doncs, des d'aquest punt de vista, els sistemes d'accions i les seves aplicacions – veure més avall – poden ser vistos, en qualsevol cas, com a aplicacions de gramàtiques contextuais.

Si formalitzem l'Exemple 16.0.2 mitjançant un “sistema d'accions simple”

$$\sigma_1 = \{A, A_0, A_f, \varphi\}$$

obtenim:

$$A = \{a_1, a_2, a_3, a_4, a_5, a_6\},$$

$$A_0 = \{a_1\},$$

$$A_f = \{a_6\},$$

$$\varphi : A \longrightarrow \mathcal{P}(A)$$

$$\varphi(a_i) = \{a_{i+1} \mid 1 \leq i \leq 3\},$$

$$\varphi(a_4) = \{a_4, a_5\},$$

$$\varphi(a_5) = \{a_6\},$$

$$\varphi(x) = \emptyset, \text{ en qualsevol altre cas.}$$

Aleshores

$$L(\sigma_1) = \{a_1 a_2 a_3 a_4^n a_5 a_6; \ n \geq 1\}.$$

Aquest model σ_1 de sistema d'accions simple es correspon amb la gramàtica contextual externa unilateral amb derivació marcada

$$G_1 = (V, B, C_1, C_2, \varphi, \psi),$$

amb

$$\begin{aligned}
 V &= \{a_1, a_2, a_3, a_4, a_5, a_6\}, \\
 B &= \{a_1\}, \\
 C_1 &= \{(\lambda, a_1), (\lambda, a_2), (\lambda, a_3), (\lambda, a_4), (\lambda, a_5), (\lambda, a_6)\} \\
 C_2 &= \{(\lambda, a_6)\}
 \end{aligned}$$

$$\begin{aligned}
 \varphi : V^* &\longrightarrow \mathcal{P}(C_1) = 2^{C_1} \\
 \varphi(a_1) &= \{(\lambda, a_2)\}, \\
 \varphi(a_1 a_2) &= \{(\lambda, a_3)\}, \\
 \varphi(a_1 a_2 a_3) &= \{(\lambda, a_4)\}, \\
 \varphi(a_1 a_2 a_3 a_4^n) &= \{(\lambda, a_4), (\lambda, a_5) \mid n \geq 1\}, \\
 \varphi(a_1 a_2 a_3 a_4^n a_5) &= \{(\lambda, a_6) \mid n \geq 1\}, \\
 \varphi(x) &= \emptyset, \text{ en qualsevol altre cas,}
 \end{aligned}$$

$$\begin{aligned}
 \psi : V^* &\longrightarrow \mathcal{P}(C_2) = 2^{C_2} \\
 \psi(a_1 a_2 a_3 a_4^n a_5) &= \{\lambda, a_6 \mid n \geq 1\}, \\
 \psi(x) &= \emptyset, \text{ en qualsevol altre cas.}
 \end{aligned}$$

El llenguatge generat per la gramàtica contextual G_1 és

$$L(G_1) = \{a_1 a_2 a_3 a_4^n a_5 a_6 \mid n \geq 1\}.$$

L'estudi del llenguatge generat per un sistema d'accions simple, $L(SAS)$, pot donar-nos una idea del procés modelat. No obstant, l'anterior marc de treball és massa general per tal d'obtenir resultats significants. En la Definició 16.0.2, l'acció següent depèn del conjunt de l'evolució anterior de l'acció. En l'Exemple 16.0.2, aquesta dependència és efectiva però en l'Exemple 16.0.3 només l'últim valor de t determina l'acció següent.

Per això, hem d'imposar algunes restriccions relatives a la funció “acció següent”. Un possible cas particular és el següent (Păun, 1996a, pàg. 30).

Definició 16.0.3 Un *sistema d'accions simple amb historicitat finita* és una 5-tupla

$$\sigma = (A, A_0, A_f, k, \varphi),$$

on

- A, A_0, A_f tenen la mateixa significació que en un SAS ,
- k és un nombre enter positiu i
- $\varphi : E \longrightarrow 2^A$, per a algun $E \subseteq (A \cup \{\lambda\})^k$ (i.e., φ és una funció parcial del conjunt de cadenes de longitud com a màxim k , sobre A , al conjunt de subconjunts d' A).

L'enter k s'anomena l'*ordre d'historicitat* de σ i es denota per $Ist(\sigma)$. Per a un llenguatge $L \subset A^*$ aleshores

$$Ist(L) = \inf\{Ist(\sigma) \mid L = L(\sigma)\}.$$

El *procés* identificat per σ es descriu mitjançant el llenguatge

$$\begin{aligned} L(\sigma) = \{a_1 a_2 \dots a_n \mid & n \geq 1, a_1 \in A_0, a_n \in A_f, \text{ per a cada } 2 \leq i \leq n, \\ & a_i \in \varphi(a_j a_{j+1} \dots a_{i-1}), \text{ per a algun } 1 \leq j \leq i-1 \\ & \text{tal que } a_j a_{j+1} \dots a_{i-1} \in L\}. \end{aligned}$$

Per exemple, si formalitzem l'Exemple 16.0.3 anterior mitjançant un “sistema d'accions simple amb historicitat finita”

$$\sigma_2 = \{A, A_0, A_f, t, \varphi\}$$

obtenim:

$$A = \{a_1 \dots a_9\},$$

$$A_0 = \{a_1\},$$

$$A_f = \{a_9\},$$

$$\varphi : A^t \longrightarrow \mathcal{P}(A) = 2^A,$$

$$\varphi(a_i) = \{a_{i+1}\}, \quad i = 1, 2, 3, 8,$$

$$\varphi(a_i) = \{a_i, a_{i+1}\}, \quad i = 4, 6, 7,$$

$$\varphi(a_4 a_5^i) = \{a_5\}, \quad 1 \leq i \leq t-1,$$

$$\varphi(a_5^t) = \{a_6\},$$

$$\varphi(x) = \emptyset, \text{ en qualsevol altre cas.}$$

El llenguatge generat per aquest sistema d'accions simple amb historicitat

finita de grau t és

$$L(\sigma_2) = \{a_1 a_2 a_3 a_4^n a_5^t a_6^m a_7^p a_8 a_9; \ n, m \geq 1, \ p \geq 0, \ t \text{ fix}\}.$$

Aquest sistema d'accions simple amb historicitat finita σ_2 es correspon amb la gramàtica contextual externa unilateral amb derivació marcada i amb selecció restringida G_2

$$G_2 = (V, B, C, \varphi, \psi),$$

amb

$$V = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\},$$

$$B = \{a_1\},$$

$$C_1 = \{(\lambda, a_k) \mid 2 \leq k \leq 9\},$$

$$C_2 = \{(\lambda, a_9)\},$$

$\varphi : V^* \longrightarrow 2^A$, per a algun $E \subseteq (A \cup \{\lambda\})^k$, es defineix com a

$$\begin{aligned} \varphi(a_1^{m_1} \dots a_k^{m_k} a_i) &= \{(\lambda, a_{i+1}) \mid i = 1, 2, 3, 8; 1 \leq k \leq i-1; m_{1,2,3} = 1, \\ &\quad m_{4,6} \geq 1, m_5 = t, m_7 \geq 0\}, \end{aligned}$$

$$\begin{aligned} \varphi(a_1^{m_1} \dots a_k^{m_k} a_i) &= \{(\lambda, a_i), (\lambda, a_{i+1}) \mid i = 4, 6, 7; 1 \leq k \leq i-1; m_{1,2,3} = 1, \\ &\quad m_{4,6} \geq 1, m_5 = t, m_7 \geq 0\}, \end{aligned}$$

$$\varphi(a_1^{m_1} \dots a_k^{m_k} a_4 a_5^i) = \{(\lambda, a_5) \mid 1 \leq i \leq t-1; 1 \leq k \leq i-2; m_{1,2,3} = 1, m_4 \geq 1\},$$

$$\varphi(a_1^{m_1} \dots a_k^{m_k} a_5^t) = \{(\lambda, a_6) \mid 1 \leq k \leq 4; m_{1,2,3} = 1, m_4 \geq 1, t \text{ fixe}\},$$

$$\varphi(x) = \emptyset \text{ en qualsevol altre cas,}$$

$$\begin{aligned}\psi V^* &\longrightarrow \mathcal{P}(C_2) = 2^{C_2}, \\ \psi(a_1^{m_1} \dots a_k^{m_k} a_5^t) &= \{(\lambda, a_6) \mid 1 \leq k \leq 4; m_{1,2,3} = 1, m_4 \geq 1, t \text{ fixe}\}.\end{aligned}$$

El llenguatge generat per la gramàtica contextual G_2 és

$$L(G_2) = \{a_1 a_2 a_3 a_4^n a_5^t a_6^m a_7^p a_8 a_9; \ n, m \geq 1, p \geq 0, t \text{ fix}\}$$

També podem definir aquest llenguatge de la següent manera: per a un vocabulari arbitrari V i dos llenguatges arbitraris L_1, L_2 sobre V , definim el llenguatge

$$pdret(L_1, L_2),$$

de *prolongació a la dreta* – segons la Definició 6.13.1 de la pàgina 230 –, de cadenes d' L_1 d'acord a cadenes d' L_2) com el llenguatge més petit $L \subseteq V^*$ tal que

- (i) $L_1 \subseteq L$,
- (ii) si $x \in L, x = x_1 x_2, x_1, x_2 \in V^*$, i $x_2 a \in L_2$ per a algun $a \in V$, aleshores $xa \in L$.

Aleshores tenim

$$L(\sigma) = pdret(A_0, L_\varphi) \cap A^* A_f,$$

on

$$L_\varphi = \{xa \mid x \in E, a \in \varphi(x)\}.$$

Cal notar que, com ja vam veure, la prolongació a la dreta és una operació purament contextual.

A (Păun, 1979c; Păun, 1980a) es demostra que $pdret(L_1, L_2) \in REG$ per a $L_1 \in REG, L_2 \in FIN$. A més a més, si σ és un sistema d'accions simple amb historicitat finita, aleshores $L(\sigma)$ és un llenguatge regular (i, per tant, el procés pot ser controlat o observat mitjançant un autòmat finit).

En termes de gramàtiques contextuais, la historicitat finita es correspon amb una selecció restringida en l'adjunció de contextos (més específicament, en aquest cas tenim contextos *unilaterals* – hem estudiat aquesta variant de les gramàtiques contextuais a la Secció 10.3).

L'operació de prolongació a la dreta s'anomena també *operació dòmino*. Només cal recordar el que succeeix quan tenim un seguit de fitxes del dòmino posades dempeus, en fila, una darrera l'altra i fem caure la primera fitxa. Aleshores aquesta fa caure la segona, i la segona la tercera i així successivament fins que ja no queden més fitxes dempeus. Aquest fet justifica també la seva aplicació per a la modelització de fenòmens politico-econòmics (p.e., la famosa teoria del dòmino que va ser aplicada als països de l'est d'Europa, després del daltabaix polític-econòmic de l'antiga URSS, que va ocasionar la caiguda progressiva dels règims comunistes dels seus països satèl·lits). Per exemple, podem acabar aquest apartat relatiu als sistemes d'accions estudiant el següent exemple, de màxima actualitat, aplicat a la política internacional

Exemple 16.0.4 Considerem aquest article aparegut a “L’Avui” del diumenge dia 5-1-97.

L'AVUI
 DIUMENGE
 5 DE GENYER DE 1997

• OPINIÓ •

E D I T O R I A L

Un pas insuficient

El president serbi, Slobodan Milosevic, ha fet, finalment, un pas concret. Però s'ha quedat curt. Ha reconegut, després de més quaranta dies de protestes en el carrer, la victòria de l'oposició en nou dels setze districtes de Belgrad i en tres ciutats de les catorze que aquesta reclama. Però aquest reconeixement escàs no deixa clara l'alcaldia de la capital ni de la majoria de les ciutats més importants del país. El moviment pro-democràtic serbi, unit al voltant de la coalició Zajedno, va rebutjar ahir aquesta proposta del govern i va avisar que les concentracions i les manifestacions seguiran mentre no es reconegui la victòria en els termes que reclamen.

La tímida concessió de Milosevic va arribar divendres en un moment especialment dur per al seu govern. Dijous, l'Església ortodoxa sèrbia condemnava el règim amb contundència. En un comunicat públic, el Sant Sínode, la màxima assemblea eclesiàstica del país, acusava el govern de sembrar la discòrdia i de provocar el vessament de sang únicament per mantenir-se en el poder. Els bisbes serbis afirmaven, també, que el règim de Milosevic “ha trait els territoris occidentals [els serbis de Coràcia i de Bòsnia], ha portat la nació i l'Estat a la fallida i el poble a la mendicitat”. La presa de posició de l'Església arraconava encara més el president serbi, que, a més del seu partit i el seu govern, només té al costat les forces policials i una part de l'exèrcit, la divisió interna del qual cobra més força a mesura que passen els dies. Per acabar de complicar la situació del president, divendres l'Organització per a la Seguretat i la Cooperació a Europa (OSCE) va aprovar l'informe redactat per la comissió presidida per Felipe González on s'afirma que l'oposició va guanyar les eleccions en catorze de les divuit grans ciutats sèrbies. L'OSCE va respondre al gest de Milosevic demanant-li el reconeixement “complet” del resultat de la consulta popular.

Totes aquestes dissidències internes i reaccions hostils externes han de fer que Milosevic mediti la seva situació i accepti la duresa dels fets –durs, per als seus interessos–. I això passa per reconèixer totes les demandes de l'oposició que ha fet seves l'OSCE. No hi ha, ara mateix, una altra sortida digna per al seu govern. L'enduriment de les seves posicions o la repressió policial o militar dels actes de protesta dels seus detractors portarien implícit un greu conflicte que podria conduir a greus enfrontaments civils. Milosevic hauria de reconèixer ja els resultats de les eleccions municipals i no hauria de desgastar el seu govern ni un dia més, si és que pretén preparar les legislatives que s'han de fer enguany. La seva negativa només portarà precarietat i retardarà la necessària recuperació que el seu poble demana.

- 1) El 17-11-96 es produeixen eleccions municipals a Sèrbia (a_1).

- 2) L'endemà mateix de l'escrutini, la comissió electoral confirma la victòria de l'oposició al règim de Slobodan Milosevic en 15 de les 18 ciutats més grans de Sèrbia, entre les quals hi ha la capital del país, Belgrad (a_2).
- 3) El règim del president Milosevic anul·la parcialment les eleccions municipals (a_3).
- 4) L'oposició sèrbia realitza més de quaranta dies $[d]$ de protestes en el carrer [exactament, $d = 46$] en oposició al frau electoral protagonitzat pel règim de Slobodan Milosevic [considerarem que l'acció simple a_4 es correspon amb "un dia de protesta"].
- 5) El dijous, dia 2-1-97, l'Església ortodoxa sèrbia condemna el règim amb contundència (a_5).
- 6) La presa de posició de l'Església arracona encara més el president serbi, que, a més del seu partit i el seu govern, només té al costat les forces policials i una part de l'exèrcit, la divisió interna del qual cobra més força a mesura que passen els dies (a_6).
- 7) Per acabar de complicar la situació del president, el divendres dia 3-1-97, l'Organització per a la Seguretat i la Cooperació a Europa (OSCE) aprova l'informe redactat per la comissió presidida per Felipe González on s'afirma que l'oposició va guanyar les eleccions en catorze de les divuit grans ciutats sèrbies (a_7).
- 8) El divendres, dia 3-1-97, Slobodan Milosevic reconeix, finalment, la victòria

de l'oposició en nou dels setze districtes de Belgrad i en tres ciutats de les catorze, que aquesta reclama (a_8).

- 9) Slobodan Milosevic, però, no deixa clara l'alcaldia de la capital ni de la majoria de les ciutats més importants del país (a_9).
- 10) L'OSCE respon al gest de Milosevic demanant-li el reconeixement "complet" del resultat de la consulta popular (a_{10}).
- 11) El moviment pro-democràtic serbi, unit al voltant de la coalició Zajedno, rebutja aquesta proposta del govern (a_{11}).
- 12) El moviment pro-democràtic serbi, unit al voltant de la coalició Zajedno, avisa que les concentracions i les manifestacions seguiran mentre no es reconegui la victòria en els termes que reclamen (a_{12}).
- 13) Totes aquestes dissidències internes i reaccions hostils fan que Milosevic mediti la seva situació i accepti la duresa dels fets (a_{13}).
- 14) Milosevic reconeix totes les demandes de l'oposició que ha fet seves l'OSCE (a_{14}).
- 15) Milosevic prepara les eleccions legislatives que s'han de fer enguany (a_{15}).
- 16) Milosevic no reconeix les demandes de l'oposició que ha fet seves l'OSCE (a_{16}).
- 17) El govern endureix les seves posicions o la repressió policial o militar dels actes de protesta dels seus detractors (a_{17}).

- 18) Es produeix un greu conflicte social que condueix a greus enfrontaments civils (a_{18}).
- 19) Com a conseqüència dels greus enfrontaments civils es retarda la necessària recuperació democràtica que el poble serbi reclama (a_{19}).

Observem que

- Els punts 1) – 3) corresponen a accions ja succeïdes que, malgrat no aparèixer explícitament en l'article, òbviament se sobreentenien; els punts 4) – 12) corresponen a fets expressats en l'article i els punts 13) – 19) corresponen a fets (hipotètics), expressats en l'article, que poden succeir en un futur.
- Es de suposar que, segons el que es manifesta al punt 12), les concentracions i manifestacions segueixen dia rera dia mentre no es reconegui la victòria en els termes que reclama el moviment pro-democràtic serbi i cessaran tan aviat com el règim de Slobodnan Milosevic faci públic el seu reconeixement.

Si formalitzem l'Exemple 16.0.4 anterior mitjançant un “sistema d'accions simple”

$$\sigma = \{A, A_0, A_f, \varphi\}$$

obtenim:

$$\begin{aligned} A &= \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}, a_{17}, a_{18}, a_{19}\}, \\ A_0 &= \{a_1\}, \\ A_f &= \{a_{15}, a_{19}\}, \end{aligned}$$

$$\begin{aligned}
\varphi : A &\longrightarrow \mathcal{P}(A) \\
\varphi(a_1) &= \{a_2\}, \\
\varphi(a_2) &= \{a_3, a_{15}\}, \\
\varphi(a_3) &= \{a_4\}, \\
\varphi(a_4 a_4^{d_1}) &= \{a_4 \mid 1 \leq d_1 \leq 45\}, \\
\varphi(a_1^{46}) &= \{a_5\}, \\
\varphi(a_5) &= \{a_6\}, \\
\varphi(a_6) &= \{a_7\}, \\
\varphi(a_7) &= \{a_1\}, \\
\varphi(a_7 a_4) &= \{a_8\}, \\
\varphi(a_i) &= \{a_{i+1} \mid 8 \leq i \leq 12\}, \\
\varphi(a_{13}) &= \{a_1\}, \\
\varphi(a_{13} a_4) &= \{a_{14}, a_{16}\}, \\
\varphi(a_{14}) &= \{a_{15}\}, \\
\varphi(a_{16}) &= \{a_1\}, \\
\varphi(a_{16} a_4) &= \{a_1\}, \\
\varphi(a_{16} a_4^{d_2}) &= \{a_{17} \mid d_2 \geq 1\}, \\
\varphi(a_{17}) &= \{a_1, a_{18}\}, \\
\varphi(a_{17} a_4) &= \{a_1, a_{18}\}, \\
\varphi(a_{17} a_4^{d_3}) &= \{a_{18} \mid d_3 \geq 1\}, \\
\varphi(a_{18}) &= \{a_1, a_{19}\}, \\
\varphi(a_{18} a_4) &= \{a_1, a_{19}\}, \\
\varphi(a_{18} a_4^{d_4}) &= \{a_{19} \mid d_4 \geq 1\},
\end{aligned}$$

Es correspon amb la gramàtica contextual externa unilateral amb derivació marcada

$$G_3 = (V, B, C_1, C_2, \varphi, \psi),$$

amb

$$\begin{aligned} V &= \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}, a_{17}, a_{18}, a_{19}\}, \\ B &= \{a_1\}, \end{aligned}$$

$$\begin{aligned} C_1 &= \{(\lambda, a_1), (\lambda, a_2), (\lambda, a_3), (\lambda, a_4), (\lambda, a_5), (\lambda, a_6), \\ &\quad (\lambda, a_7), (\lambda, a_8), (\lambda, a_9), (\lambda, a_{10}), (\lambda, a_{11}), (\lambda, a_{12}), (\lambda, a_{13}), \\ &\quad (\lambda, a_{14}), (\lambda, a_{15}), (\lambda, a_{16}), (\lambda, a_{17}), (\lambda, a_{18}), (\lambda, a_{19})\}, \end{aligned}$$

$$C_2 = \{(\lambda, a_{15}), (\lambda, a_{19})\}$$

$$\begin{aligned} \varphi : V^* &\longrightarrow \mathcal{P}(C_1) \\ \varphi(a_1) &= \{(\lambda, a_2)\}, \\ \varphi(a_1 a_2) &= \{(\lambda, a_3), (\lambda, a_{15})\}, \\ \varphi(a_1 a_2 a_3) &= \{(\lambda, a_4)\}, \\ \varphi(a_1 a_2 a_3 a_4 a_4^{d_1}) &= \{(\lambda, a_4) \mid 1 \leq d_1 \leq 45\}, \\ \varphi(a_1 a_2 a_3 a_4^{46}) &= \{(\lambda, a_5)\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5) &= \{(\lambda, a_6)\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6) &= \{(\lambda, a_7)\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7) &= \{(\lambda, a_4)\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4) &= \{(\lambda, a_8)\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8) &= \{(\lambda, a_9)\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9) &= \{(\lambda, a_{10})\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10}) &= \{(\lambda, a_{11})\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11}) &= \{(\lambda, a_{12})\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12}) &= \{(\lambda, a_{13})\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13}) &= \{(\lambda, a_4)\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4) &= \{(\lambda, a_{14}), (\lambda, a_{16})\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{14}) &= \{(\lambda, a_{15})\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16}) &= \{(\lambda, a_4)\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2}) &= \{(\lambda, a_{14}), (\lambda, a_{17}) \mid d_2 \geq 1\}, \\ \varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{14}) &= \{(\lambda, a_{15})\}, \end{aligned}$$

$$\begin{aligned}
\varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17}) &= \{(\lambda, a_4), (\lambda, a_{18}) \mid d_2 \geq 1\}, \\
\varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3}) &= \{(\lambda, a_{14}), (\lambda, a_{18}) \mid d_2, d_3 \geq 1\}, \\
\varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{14}) &= \{(\lambda, a_{15})\}, \\
\varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{18}) &= \{(\lambda, a_4), (\lambda, a_{19}) \mid d_2, d_3 \geq 1\}, \\
\varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{18} a_4^{d_4}) &= \{(\lambda, a_{14}), (\lambda, a_{19}) \mid d_2, d_3, d_4 \geq 1\}, \\
\varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{18} a_4^{d_4} a_{14}) &= \{(\lambda, a_{15})\}, \\
\varphi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{18} a_4^{d_4}) &= \{(\lambda, a_{19}) \mid d_2, d_3, d_4 \geq 1\}, \\
\varphi(x) &= \emptyset \text{ en qualsevol altre cas.}
\end{aligned}$$

$$\begin{aligned}
\psi : V^* &\longrightarrow \mathcal{P}(C_2) \\
\psi(a_1 a_2) &= \{(\lambda, a_{15})\}, \\
\psi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{14}) &= \{(\lambda, a_{15})\}, \\
\psi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{14}) &= \{(\lambda, a_{15})\}, \\
\psi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{14}) &= \{(\lambda, a_{15})\}, \\
\psi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{18} a_4^{d_4} a_{14}) &= \{(\lambda, a_{15})\}, \\
\psi(a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{18} a_4^{d_4}) &= \{(\lambda, a_{19})\}, \\
\psi(x) &= \emptyset \text{ en qualsevol altre cas.}
\end{aligned}$$

Aleshores, el llenguatge generat per G_3 és

$$\begin{aligned}
L(G_3) = \{ &a_1 a_2 a_{15}, \\
&a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{14} a_{15}, \\
&a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{14} a_{15}, \\
&a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{14} a_{15}, \\
&a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{18} a_4^{d_4} a_{14} a_{15}, \\
&a_1 a_2 a_3 a_4^{46} a_5 a_6 a_7 a_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_4 a_{16} a_4^{d_2} a_{17} a_4^{d_3} a_{18} a_4^{d_4} a_{19} \}.
\end{aligned}$$

Com ja hem fet esment, l'estudi del llenguatge generat per un sistema d'accions simple o bè el de la gramàtica contextual associada ens pot donar

una idea del procés modelat. En aquest cas una possible interpretació seria

- 1) $\underbrace{a_1 a_2 a_{15}}_{\substack{\text{abans del 2-1-97} \\ \text{PASSAT}}}$
- 2) $\underbrace{a_1 a_2 a_3 \mathbf{a}_4^{45}}_{\substack{\text{abans del 2-1-97} \\ \text{PASSAT}}} \underbrace{\mathbf{a}_4 a_5 a_6 a_7}_{2-1-97} \underbrace{\mathbf{a}_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13}}_{3-1-97} \underbrace{\mathbf{a}_4 a_{14} a_{15}}_{4-1-97},$
 PRESENT
- 3) $\underbrace{a_1 a_2 a_3 \mathbf{a}_4^{45}}_{\substack{\text{abans del 2-1-97} \\ \text{PASSAT}}} \underbrace{\mathbf{a}_4 a_5 a_6 a_7}_{2-1-97} \underbrace{\mathbf{a}_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13}}_{3-1-97} \underbrace{\mathbf{a}_4 a_{16}}_{4-1-97} \underbrace{\mathbf{a}_4^{\mathbf{d}_2} a_{14} a_{15}}_{\substack{\text{després del 4-1-97} \\ \text{FUTUR}}},$
 PRESENT
- 4) $\underbrace{a_1 a_2 a_3 \mathbf{a}_4^{45}}_{\substack{\text{abans del 2-1-97} \\ \text{PASSAT}}} \underbrace{\mathbf{a}_4 a_5 a_6 a_7}_{2-1-97} \underbrace{\mathbf{a}_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13}}_{3-1-97} \underbrace{\mathbf{a}_4 a_{16}}_{4-1-97} \underbrace{\mathbf{a}_4^{\mathbf{d}_2} a_{17} \mathbf{a}_4^{\mathbf{d}_3} a_{14} a_{15}}_{\substack{\text{després del 4-1-97} \\ \text{FUTUR}}},$
 PRESENT
- 5) $\underbrace{a_1 a_2 a_3 \mathbf{a}_4^{45}}_{\substack{\text{abans del 2-1-97} \\ \text{PASSAT}}} \underbrace{\mathbf{a}_4 a_5 a_6 a_7}_{2-1-97} \underbrace{\mathbf{a}_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13}}_{3-1-97} \underbrace{\mathbf{a}_4 a_{16}}_{4-1-97} \underbrace{\mathbf{a}_4^{\mathbf{d}_2} a_{17} \mathbf{a}_4^{\mathbf{d}_3} a_{18} \mathbf{a}_4^{\mathbf{d}_4} a_{14} a_{15}}_{\substack{\text{després del 4-1-97} \\ \text{FUTUR}}},$
 PRESENT
- 6) $\underbrace{a_1 a_2 a_3 \mathbf{a}_4^{45}}_{\substack{\text{abans del 2-1-97} \\ \text{PASSAT}}} \underbrace{\mathbf{a}_4 a_5 a_6 a_7}_{2-1-97} \underbrace{\mathbf{a}_4 a_8 a_9 a_{10} a_{11} a_{12} a_{13}}_{3-1-97} \underbrace{\mathbf{a}_4 a_{16}}_{4-1-97} \underbrace{\mathbf{a}_4^{\mathbf{d}_2} a_{17} \mathbf{a}_4^{\mathbf{d}_3} a_{18} \mathbf{a}_4^{\mathbf{d}_4} a_{19}}_{\substack{\text{després del 4-1-97} \\ \text{FUTUR}}},$
 PRESENT

En paraules. Quan el 17-11-96 es produeixen eleccions municipals a Sèrbia (a_1) i l'endemà mateix de l'escrutini, la comissió electoral confirma la victòria de l'oposició al règim de Slobodan Milosevic en 15 de les 18 ciutats més grans de Sèrbia, entre les quals hi ha la capital del país, Belgrad (a_2), la situació “normal” hagués estat que Milosevic s'hagués dedicat, única i exclusivament, a preparar les eleccions legislatives que s'han de fer enguany (a_{15}) 1.

No obstant, però, el règim del presentent Milosevic anul·là parcialment les eleccions municipals (a_3) i l'oposició sèrbia reacciona realitzant 45 dies de protestes en el carrer en oposició al frau electoral protagonitzat pel règim de Slobodan Milosevic (a_4^{45}).

El dijous dia 2-1-97, mentre continuaven les protestes al carrer (a_4), l'Església ortodoxa sèrbia condemnà el règim amb contundència (a_5), arraconant encara més el president serbi, que, a més del seu partit i el seu govern, només tnia al costat les forces policials i una part de l'exèrcit, la divisió interna del qual cobrava més força a mesura que passaven els dies (a_6). Per acabar de complicar la situació del president, el l'Organització per a la Seguretat i la Cooperació a Europa (OSCE) aprovà l'informe redactat per la comissió presidida per Felipe González on s'afirmava que l'oposició havia guanyat les eleccions en catorze de les divuit grans ciutats sèrbies (a_7).

Les manifestacions de protesta van continuar un dia més a_4 i el divendres, dia 3-1-97, Slobodan Milosevic reconeixia, finalment, la victòria de l'oposició en nou dels setze districtes de Belgrad i en tres ciutats de les catorze, que aquesta reclamava (a_8). Slobodan Milosevic, però, no deixava clara l'alcaldia de la capital ni de la majoria de les ciutats més importants del país (a_9). L'OSCE va respondre al gest de Milosevic demanant-li el reconeixement "complet" del resultat de la consulta popular (a_{10}). Per la seva part, el moviment pro-democràtic serbi, unit al voltant de la coalició Zajedno, va rebutjar aquesta proposta del govern (a_{11}) i avisà que les concentracions i les manifestacions

seguirien mentre no es reconegué la victòria en els termes que reclamaven (a_{12}). Totes aquestes dissidències internes i reaccions hostils fan que Milosevic mediti la seva situació i accepti la duresa dels fets (a_{13}).

En la situació actual, Milosevic només té dues possibilitats cara al futur: o bé reconeix totes les demandes de l'oposició que ha fet seves l'OSCE (a_{14}) i es dedica a preparar les eleccions legislatives que s'han de fer enguany (a_{15}) [2]. O bé no reconeix les demandes de l'oposició que ha fet seves l'OSCE (a_{16}).

En aquest últim cas, però, és de suposar que implicarà que les manifestacions continuïn durant un període de temps indefinit ($a_a^{d_2}$), fins que reconegui totes les demandes de l'oposició que ha fet seves l'OSCE (a_{14}) i es dediqui a preparar les eleccions legislatives que s'han de fer enguany (a_{15}) [3].

I si no és així, i el govern endureix les seves posicions o la repressió policial o militar dels actes de protesta dels seus detractors (a_{17}), aleshores, mentre continuïn les protestes al carrer ($a_4^{d_2}$) – i, sempre i quan no reconegui totes les demandes de l'oposició (a_{14}) i es dediqui a preparar les eleccions legislatives (a_{15}) [4] –, es produirà un greu conflicte social que conduirà a greus enfrontaments civils (a_{18}). En últim terme, continuaran les manifestacions al carrer ($a_4^{d_4}$) per tal d'aconseguir de Milosevic el seu reconeixement de totes les demandes de l'oposició (a_{14}) i la seva dedicació a preparar les eleccions legislatives d'enguany (a_{15}) [5] i, com a mal pitjor, com a conseqüència dels greus enfrontaments civils, es retardarà la necessària recuperació democràtica que el poble serbi reclama (a_{19}) [6].

Usant aquests models de sistemes d'accions, així com models més complexos, s'han investigat diversos processos particulars. És el cas de processos econòmics, investigats a (Păun, 1980a); conversacions governades mitjançant diversos protocols (Păun, 1980b); l'estratègia dels personatges, a nivell de la seva presència en l'escena, en les obres de teatre, investigada a (Păun, 1976b) (seguint l'estil de (Marcus, 1970b)); programes de rastreig, investigats a (Bălănescu i Gheorghe, 1987), etc. Com ja hem dir, tots aquests exemples poden ser considerats com a aplicacions de les gramàtiques contextuais – però no pas en el camp on les gramàtiques contextuais es van originar; és a dir, en l'estudi del llenguatge natural !

Un tòpic que ha estat extensivament investigat en la teoria de seqüències infinites i sembla ser un cas particular dels processos anteriors (amb una historicitat no acotada) és el de les “seqüències autogenerades” (o “self-generated sequences”), en el sentit de (Sloane i Plouffe, 1995). La idea bàsica és la següent: a partir d'una cadena $x \in V^+$ i una funció $f : V^+ \longrightarrow V^+$, per a un vocabulari donat V , construïm la seqüència

$$z(f, x) = x_0 x_1 x_2 \dots,$$

on $x_0 = x$ i $x_{n+1} = f(x_0 x_1 \dots x_n)$, $n \geq 0$.

Per exemple, prenem $V = \{0, 1\}$ i denotem per $\beta_2(m)$ la representació, en base 2, del número enter m . Considerem la funció $f_1 : \{0, 1\}^+ \longrightarrow \{0, 1\}^+$ definida per

$$f_1(x) = \beta_2(|x|_1)1$$

És a dir, contem el nombre d'ocurrències del dígit 1 en la cadena x , expressem aquest número en base 2, i afegim una ocurrència més d'1. Aquest és el començament de la seqüència obtinguda, quan comencem per $x_0 = 1$:

$$z(f_1, 1) = 1(1)1(11)1(110)1(1001)1(1100)1(1111)1(101100)1 \dots$$

(els parèntesis contenen les cadenes $\beta_2(|x_i|_1)$ i només s'han afegit per a més claredat).

No entrarem en detalls pel que fa a aquest cas en particular. Al respecte es pot consultar (Dassow et al., 1993) i (Păun i Salomaa, 1996) per a propietats i referències relatives a aquestes seqüències (anomenades seqüències *autollegides* “self-reading”, i (Sloane i Plouffe, 1995) per l'àrea general de les seqüències autogenerades. Només farem esment al fet que mode contextual d'incrementar les seqüències, mitjançant l'adjunció d'una cadena a l'extrem dret de la cadena en qüestió, d'acord als seus contextos. És clar que la cadena adjuntada no és de longitud acotada i, per tant, hem d'assumir que treballem amb un conjunt de contextos infinit. Aquest fet ha estat també considerat pel cas de les gramàtiques contextuais – veure la Secció 13.2.

Capítol 17

Caracterització de llenguatges amb “computació natural” mitjançant gramàtiques contextuais

La teoria de gramàtiques formals que està en la mateixa base de la sintaxi formal va fent, des de Chomsky, un ús quasibé exclusiu dels procediments de reescriptura (veure la Definició 3.3.1 de la pàgina 111). En efecte, a partir de la jerarquia de Chomsky, s’han desenvolupat diverses tècniques generatives, però totes elles més o menys pròximes a aquelles eines originals. Aquesta circumstància pot portar a creure que la reescriptura és inevitable en la teoria de llenguatges formals, en la teoria d’autòmats, en la teoria d’algorismes o en les ciències de la computació en general.

Sabem que les màquines de Turing caracteritzen els llenguatges recursivament enumerables (veure el teorema 3.6.1 de la pàgina 137), la classe més àmplia concebible de llenguatges. Doncs bé, en aquets últims anys han aparegut eines generatives (en el sentit més ampli de la paraula) que no fan ús de la reescriptura i que tenen la mateixa capacitat generativa que els dispositius clàssics. I, a més a més, resulta que aquestes noves eines semblen ser “més naturals” que la reescriptura, en el sentit de què no són simples artificis creats amb finalitats de descripció científica, sinó en certa manera operacions presents, d’una o altra forma, en la mateixa naturalesa. Si la capacitat d’aquests nous procediments no va a la saga de la dels procediments clàssics, la seva adequació per a la descripció del llenguatge natural, pel que fa a aquest nou paràmetre de naturalitat, pot ser superior. Alguns d’aquests mecanismes, presentats a (Martín-Vide, 1996), són:

- l’adjunció (i l’esborrament),
- la inserció (i l’esborrament) i
- l’empalmament.

17.1 L’adjunció (i l’esborrament)

L’operació d’adjunció és l’ingredient essencial de les gramàtiques d’adjunció d’arbres (*TAG*), presentades per primera vegada a (Joshi et al., 1975). No

obstant, les *TAG* fan ús d'arbres i de símbols auxiliars i, per tant, en realitat s'aparten poc de la reescriptura tradicional. A més a més, les *TAG* no poden caracteritzar els llenguatges recursivament enumerables (veure, p.e., una introducció al respecte a (Ortega-Robert, 1995)). Per ambdós motius, les deixarem de banda de la nostra exposició, en la mesura en què són meres variants de les gramàtiques chomskyanes i tenen una limitada capacitat generativa.

Com ja hem vist, també utilitzen l'operació d'adjunció les gramàtiques contextuais (*CG*), proposades per Marcus (Marcus, 1969a). El que caracteritza a les gramàtiques contextuais és que es tracta de gramàtiques sense símbols auxiliars que adjunten parells de cadenes (els denominats contextos) a altres cadenes prèviament obtingudes (a partir d'un axioma inicial). El fet més interessant, pel que respecta a la “computació natural”, de les gramàtiques contextuais consisteix en que, usant una operació de tall adequada, caracteritzen els llenguatges recursivament enumerables. Ens centrarem exclusivament en un tipus de gramàtiques contextuais: les gramàtiques contextuais internes amb selecció (*ICC*); i en dues de les seves variants: les gramàtiques contextuais internes amb selecció restringida (*ICC(F)*, sent *F* una família de llenguatges) – veure el Capítol 9 – i les gramàtiques contextuais internes amb selecció restringida amb contextos unilaterals (*1ICC(F)*, sent *F* una família de llenguatges) – veure la Secció 10.3. Recordarem algunes definicions i resultats obtinguts amb anterioritat.

Definició 17.1.1 (Basada en la Definició 9.1.1 de la pàgina 302.) Una *ICC* amb selecció de tipus *F*, sent *F* una família de llenguatges, que denotarem per

$ICC(F)$, és una estructura:

$$G = (V, B, (S_1, C_1), \dots, (S_n, C_n)),$$

on V és un vocabulari, B (el conjunt d'axiomes) és un subconjunt finit de V^* , $S_i \subseteq V^*$ són llenguatges d' F (llenguatges selectors), i C_i (els conjunts de contextos) són subconjunts finits de $V^* \times V^*$, $1 \leq i \leq n$.

Definició 17.1.2 (Basada en la Definició 4.1.2 de la pàgina 168.) Respecte a una gramàtica contextual com l'anterior, definim la relació de derivació interna $\Rightarrow_{in} V^*$: per a $x, y \in V^*$, escrivim

$$x \Rightarrow_{in} y \quad \text{sii} \quad x = x_1 x_2 x_3, y = x_1 u x_2 v x_3, \text{ per algun} \\ x_1, x_2, x_3 \in V^*, (u, v) \in C_i, x_2 \in S_i, 1 \leq i \leq n.$$

Així doncs, el context $(u, v) \in C_i$ s'adjunta a x_2 si x_2 és un element del llenguatge selector associat a C_i .

Denotant mitjançant $\Rightarrow_{in}^*$ la clausura reflexiva i transitiva de $\Rightarrow_{in}$, el llenguatge generat per G es defineix com:

$$L_{in}(G) = \{x \in V^* \mid w \Rightarrow_{in}^* x, \text{ per algun } w \in B\}.$$

Denotem per $ICCL(F)$ la família de llenguatges generats per $ICCG$ amb selecció del tipus F . Considerarem que F és una de les famílies de llenguatges FIN, REG, LIN, CF, CS , i RE , que designen les famílies dels llenguatges

finites, regulars, lineals, lliures o independents del context, sensibles o dependents del context i recursivament enumerables. Resumirem alguns importants resultats coneguts:

- 1) (Lema 9.1.4 de la pàgina 303.)

$$ICCL(FIN) \subset ICCL(REG) \subset ICCL(CF) \subset ICCL(CS) \subset ICCL(RE).$$

- 2) (Corol·lari 5.3.1 de la pàgina 203.)

$$REG \subset ICCL(FIN).$$

- 3) (Teorema 9.1.1 de la pàgina 310 i Lema 9.1.2 de la pàgina 303.)

LIN i CF són incomparables amb totes les famílies $ICCL(F)$, $F \in \{REG, CF, CS, RE\}$, però $ICL(CS) \subset CS$.

- 4) (Lema 9.2.1 de la pàgina 333.)

$ICCL(FIN)$ és una anti-família abSTRACTA de llenguatges (anti- AFL), és a dir, un conjunt de llenguatges que no és tancat per a cap de les sis operacions de les AFL : unió, concatenació, estrella de Kleene, morfismes, intersecció amb llenguatges regulars i morfismes inversos (veure la Definició 3.6.6 de la pàgina 145).

Aquest últim resultat desencadena la pregunta natural de trobar la AFL més petita que contingui a $ICCL(FIN)$. La resposta és certament sorprenent.

Teorema 17.1.1 (Corol·lari 9.2.1 de la pàgina 337.) Tot llenguatge $L \in RE$ pot ser escrit de la forma $L = L_1 \setminus L_2$, per a $L_1 \in REG$, $L_2 \in ICCL(FIN)$.

($L_1 \setminus L_2$ és el quocient per l'esquerra d' L_2 respecte a L_1 , definit per – veure la Definició 3.2.13 de la pàgina 110: $L_1 \setminus L_2 = \{x \in V^* \mid yx \in L_2, y \in L_1\}$.)

Aquest fet significa que l'adjunció iterada de contextos (= enganxar) juntament amb un quocient per l'esquerra per un llenguatge regular (= tallar) pot simular qualsevol màquina de Turing. Tenint en compte la dèbil capacitat generativa de les *ICCG* amb selectors finits, el resultat és del tot inesperat. Veiem que, per tal d'assolir el poder generatiu de les gramàtiques de Chomsky de tipus 0 (el poder de reconeixement de les màquines de Turing), necessitem simultàniament capacitats de percepció del context (que garantitzen les *ICCG*) i capacitats d'esborrament (que fa possible l'operació quocient).

Per altra part,

Definició 17.1.3 (Basada en la Secció 10.3.) Una *ICCG* s'anomena unilateral si tots els seus contextos són de la forma (λ, v) , sent λ la cadena buida.

La relació de derivació i el llenguatge generat es defineixen com abans. Notem per $1ICCL(F)$ la família de llenguatges generats per *ICCG* unilaterals amb selecció del tipus F .

Teorema 17.1.2 (Teorema 10.3.2 de la pàgina 407.) Tot llenguatge $L \in RE, L \subseteq V^*$ pot ser escrit de la forma $L = (L_1 \setminus L_2) \cap V^*$, per a $L_1 \in REG$ i $L_2 \in 1ICCL(CF)$.

En el Teorema 17.1.2 tenim una caracterització de la família *RE* que usa

l'operació d'adjunció d'una cadena en cada pas, depenent d'una cadena selectora situada immediatament a la seva esquerra; no obstant, fan falta cadenes selectores arbitràriament llargues, que formen un llenguatge lliure o independent del context. Si adjuntem una cadena tenint en compte també la cadena selectora de la part dreta, aleshores obtenim el resultat del Teorema 17.1.1.

Arribats a aquest punt podem intentar trobar en la natura models que computin de manera similar a aquestes gramàtiques contextuais internes unilaterals amb selecció lliure del context ($1ICCL(CF)$).

Relacionat amb l'operació de prolongació a la dreta – veure la Definició 6.13.1 de la pàgina 230 – i amb la manera de construir seqüències autollegides infinites – veure el final del Capítol 16 són els nomenats *sistemes filamentosos amb creixement apical* – i.e, amb creixement només pels àpexs o extrems superiors o puntes – definits a (Bălănescu et al., 1987), (Krithivasan i Nirmal, 1983). Aquests sistemes són variants dels $0L$ o 0-sistemes-Lindenmayer – veure la Secció 3.5).

17.2 La inserció (i l'esborrament)

L'operació bàsica en les gramàtiques contextuais és l'*adjunció* de contextos, depenent de les cadenes delimitades per les dues cadenes afegides; en les gramàtiques de Chomsky sensibles o dependents del context un símbol és reescrit per una cadena, depenent d'un context. Les gramàtiques d'inserció són un

model intermig: les cadenes són insertades en un context. Novament l'operació bàsica és la d'adjunció de cadenes, de manera semblant a les gramàtiques contextuais, no la de reescriptura, com en el cas de les gramàtiques de Chomsky, però l'operació és controlada per un context, com en el cas de les gramàtiques sensibles o depenents del context. És per això que aquest tipus de gramàtiques, introduïdes per Galiukschov a (Galiukschov, 1981) sota el nom de gramàtiques semi-contextuais, poden ser considerades – veure (Păun, 1996a, pàgines 260–288) – com a un “model dual”.

Veurem, a grans trets les definicions i resultats més importants.

Definició 17.2.1 Una *gramàtica d'inserció* és una tripleta $G = (V, B, P)$, on V és un vocabulari, B (el conjunt d'axiomes) és un subconjunt finit de V^* , i P és un conjunt finit de tripletes de la forma (u, x, v) , $u, x, v \in V^*$. Aquestes tripletes reben el nom de regles d'inserció i s'utilitzen en la derivació de la manera següent:

Definició 17.2.2 Donades $x, z \in V^*$ escriurem $w \Longrightarrow z$ sii $w = w_1 u v w_2$, $z = w_1 u x v w_2$, per a algunes $w_1, w_2 \in V^*$, $(u, x, v) \in P$.

El llenguatge generat es defineix de la manera habitual.

Definició 17.2.3 Donada una gramàtica d'inserció $G = (V, B, P)$, definim el seu pes w com:

$$w(G) = \max\{|u| \mid (u, x, v) \in P \text{ o } (v, x, u) \in P\}.$$

La família de llenguatges generats per gramàtiques d'inserció de pes com a màxim n , $n \geq 0$, es denota per INS_n . La unió de totes aquestes famílies es simbolitza mitjançant INS_∞ .

Alguns resultats bàsics són els següents:

- 1) $FIN \subset INS_0 \subset INS_1 \subset \dots \subset INS_\infty \subset CS$.
- 2) REG és incomparable amb totes les famílies INS_n , $n \geq 1$, i $REG \subset INS_\infty$.
- 3) $INS_1 \subset CF$, però CF és incomparable amb totes les famílies INS_n , $n \geq 2$, i INS_∞ ; INS_2 conté llenguatges no-semilineals.
- 4) LIN és incomparable amb totes les famílies INS_n , $n \geq 1$ i INS_∞ .
- 5) Totes les famílies INS_n , $n \geq 0$ són anti- AFL .

A l'igual que en el cas de les gramàtiques contextuais, ens trobem amb famílies de llenguatges notablement petites. No obstant:

Teorema 17.2.1 (De (Martín-Vide et al., 1996d)) Tot llenguatge $L \in RE$ pot ser escrit de la forma $L = L_1 \setminus L_2$, per a $L_1 \in REG$ i $L_2 \in INS_n$, $n \geq 7$.

Un problema pendent de resoldre és el de si es pot optimitzar aquest resultat prenent L_2 d'una família INS_n amb $n < 7$.

Novament, el que fem primer és utilitzar les operacions d'inserció, per a després tallar un prefix de la cadena obtinguda mitjançant un quocient respecte a un llenguatge regular. No obstant, es pot introduir l'operació de tall en la gramàtica d'una manera molt natural, i sorprenentment s'obté una caracterització simple de la família RE .

Definició 17.2.4 Considerem *gramàtiques d'inserció-esborrament*, $INSDEKG$, com a estructures de la forma $G = (V, B, P_I, P_D)$, on V és un vocabulari, B (el conjunt d'axiomes) és un subconjunt finit de V^* , i P_I (regles d'inserció) i P_D (regles d'esborrament) són subconjunts de $V^* \times V^* \times V^*$.

Definició 17.2.5 Donades $w, z \in V^*$, direm que $w \implies z$ si es dona un dels dos casos següents:

i) $w = w_1uvw_2$, $z = w_1uxvw_2$, per a $(u, x, v) \in P_I$, $w_1, w_2 \in V^*$, o

ii) $w = w_1uxvw_2$, $z = w_1uvw_2$, per a $(u, x, v) \in P_D$, $w_1, w_2 \in V^*$. o

Teorema 17.2.2 Tot llenguatge $L \in RE$, $L \subseteq V^*$, pot ser escrit de la forma $L = L' \cap V^*$, per a algun $L' \in INSDEL$.

Per tal de veure la capacitat de les operacions d'inserció i d'esborrament quan funcionen conjuntament, farem ús del Teorema 17.2.1. Considerem $L \subseteq V^*$ generat per una gramàtica $G = (V_N, V_T, S, P)$ en forma normal de Kuroda – veure el teorema 3.11.4 de la pàgina 162. Construïm la gramàtica d'inserció:

$$G' = (V_N \cup V_T \cup \{Y, X_1, X_2\}, P_I, P_D),$$

amb $P_I = \{(A, X_1x, \alpha_1, \alpha_2) \mid A \longrightarrow x \in P, \alpha_1, \alpha_2 \in V_N \cup V_T \cup \{Y\}\} \cup \{AB, X_2CD, \alpha_1\alpha_2 \mid AB \longrightarrow CD \in P, \alpha_1\alpha_2 \in V_N \cup V_T \cup \{Y\}\},$

$$P_D = \{(\lambda, AX_1, \lambda) \mid A \in V_N\} \cup \{(\lambda, ABX_2, \lambda) \mid A \in V_N\} \cup \{(\lambda, YY, \lambda)\}.$$

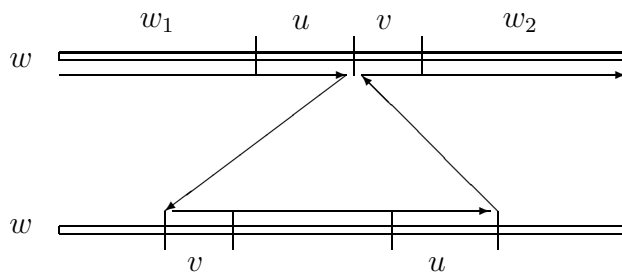
Les regles de P_I simulen les regles de P . Els símbols X_1, X_2 són marcadors dels símbols situats immediatament a la seva esquerra: X_1 marca un símbol, X_2 marca dos símbols. Degut a la part dreta dels contextos de les regles de P_I , és a dir $\alpha_1\alpha_2$, aquestes regles no es poden aplicar de manera que afectin a un símbol ja marcat. Per altra banda, els marcadors i els símbols que marquen es poden esborrar mitjançant les regles de P_D . En la part dreta de la cadena, l'ús correcte de les regles de P_I ve assegurat pel símbol auxiliar Y , les ocurrencies del qual es poden esborrar quan ja no són necessàries. En conseqüència, $L = L(G') \cap V^*$.

Si considerem la longitud de les cadenes insertades/esborrades i la longitud dels seus contextos, podem reformular l'anterior teorema dient que $L' \in INS_e^2DEL_3^0$: insertem cadenes de longitud com a màxim tres en contextos de pesos com a màxim dos, i esborrem cadenes de longitud com a màxim tres de contextos de pes zero. A (Kari et al., 1996) podem trobar millors límits. En efecte. L' pot ser element de $INS_1^2DEL_1^1$, $INS_2^1DEL_2^0$ o $INS_1^2DEL_2^0$.

Pel que fa a la presència d'aquest mecanisme en la naturalesa només cal que pensem en l'operació d'inserció com a un cas particular d'una nova operació anomenada duplicació. Com un cas de *cadena que duplica sistemes*.

Donada una cadena d'entrada i un conjunt finit de *contextos d'inserció* (u, v) , aquests contextos poden ser considerats com a “unions dèbils”, on una cadena pot ser trencada i enganxada de nou després d'afegir un nou segment. Si la cadena conté una subcadena que comença per v i acaba per u , aleshores aquesta subcadena pot ser *insertada* entre u i v , ja que les unions es preserven, ambdues dins del nou segment (ja que és una part de la cadena existent) i entre aquest segment i les parts veïnes de la cadena (tenim uv , uv als extrems de la subcadena afegida). En la Figura 17.2.1 la cadena w es mostra dues vegades, per tal d'indicar el lloc d'inserció – en la posició superior – i la subcadena per insertar – en la posició inferior. Iterant el procés, obtenim un llenguatge.

Figura 17.2.1 Anomenarem a aquesta operació *duplicació* (controlada).



Podem considerar moltes variants, depenent de la resposta a les qüestions:

- (1) On podem insertar ? i
- (2) Què podem insertar ?

En aquest cas, aquestes màquines estan estretament connectades amb idees que provenen d'un gran nombre d'àrees d'investigació: de la biologia molecular, de les operacions d'empalmament (“splicing”) i creuament (“crossover”) específiques de la recombinació de l'ADN, lingüístiques, de les **gramàtiques d'inserció**, de la teoria de llenguatges formals, de la teoria de l'aprenentatge (en el sentit que la peça original d'informació és reordenada i reproduïda, ampliant la informació existent usant “mitjans interns”), etc.

17.3 Empalmament

El cas més interessant d'ús de les operacions de tallar i enganxar per tal de caracteritzar la família *RE* és el de l'operació d'empalmar (“splicing”), per dues raons fonamentals:

- 1) L'operació d'empalmar consisteix, exacta i exclusivament, en tallar i enganxar, i
- 2) L'operació d'empalmar és una operació natural, en el sentit de què és una operació bàsica de l'ADN: podem dir que la vida computa empalmant i a nivell de les màquines de Turing.

El comportament recombinant de l'ADN és del següent tipus: els enzims de restricció tenen patrons específics que reconeixen en les parelles de seqüències de l'ADN i per on tallen, produint extrems d'una forma ben determinada;

quan dos enzims tallen dues seqüències d'ADN de tal manera que els extrems obtinguts encaixen, els fragments es poden recombinar d'una determinada manera, produint noves seqüències. Això pot succeir tant “in vivo” com “in vitro” (Head, 1987), (Adleman, 1994), (Păun, 1996b).

Definició 17.3.1 Una regla d'empalmament sobre un vocabulari V és una cadena $r = u_1\#u_2\$u_3\#u_4$, on $u_i \in V^*$, $1 \leq i \leq 4$, i $\#$, $\$$ són símbols que no pertanyen a V .

Definició 17.3.2 Donades una regla r i les cadenes $x, y, z \in V^*$, definim

$(x, y) \vdash_r z$ si i $x = x_1u_1u_2x_2$, $y = y_1u_3u_4y_2$, $z = x_1u_1u_4y_2$ per a algunes $x_1, x_2, y_1, y_2 \in V^*$.

Definició 17.3.3 Un parell $\sigma = (V, R)$, on V és un vocabulari i $R \subseteq V^*\#V^*\$V^*\#V^*$, rep el nom d'esquema H . Donat un esquema $H \sigma = (V, R)$ i un llenguatge $L \subseteq V^*$, definim:

$$\sigma(L) = \{z \in V^* \mid (x, y) \vdash_r z, \text{ per a algunes } x, y \in L, r \in R\},$$

$$\sigma^0(L) = L,$$

$$\sigma^{i+1}(L) = \sigma^i(L) \cup \sigma(\sigma^i(L)), \quad i \geq 0,$$

$$\sigma^*(L) = \bigcup_{i \geq 0} \sigma^i(L).$$

Sobre aquesta base s'ha dissenyat un nou tipus de mecanisme generatiu (Păun et al., 1996a).

Definició 17.3.4 Un sistema H extès és una estructura $\gamma = (V, T, A, R)$, on V és un vocabulari, $T \subseteq V$ és un conjunt de símbols terminals, $A \subseteq V^*$ és un conjunt d'axiomes, i $R \subseteq V^+ \# V^* \$ V^* \# V^*$, per a $\#, \$$ no perteneixents a V ; $\sigma = (V, R)$ és l'esquema H subjacent de γ .

Definició 17.3.5 El llenguatge generat per σ es defineix com:

$$L(\gamma) = \sigma^*(A) \cap T^*.$$

Donades dues famílies de llenguatges F_1, F_2 , notem mitjançant $EH(F_1, F_2)$ la família de llenguatges $L(\gamma)$, per a $\gamma = (V, T, A, R)$, amb $A \in F_1$ i $R \in F_2$ (tan A com R són llenguatges).

Teorema 17.3.1 (De (Păun, 1996c)) $EH(F_1, F_2) = RE$, per qualsevol F_1, F_2 tal que $FIN \subseteq F_1$, $REG \subseteq F_2$.

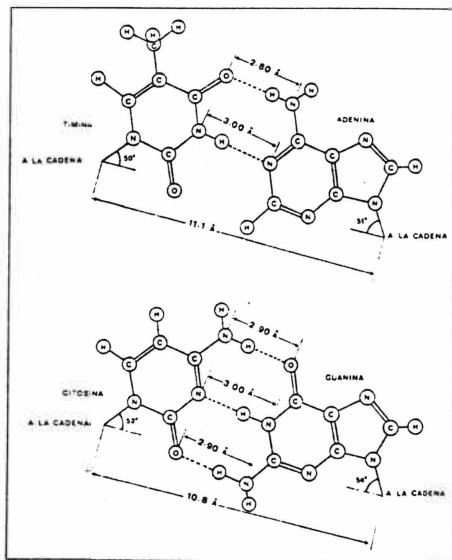
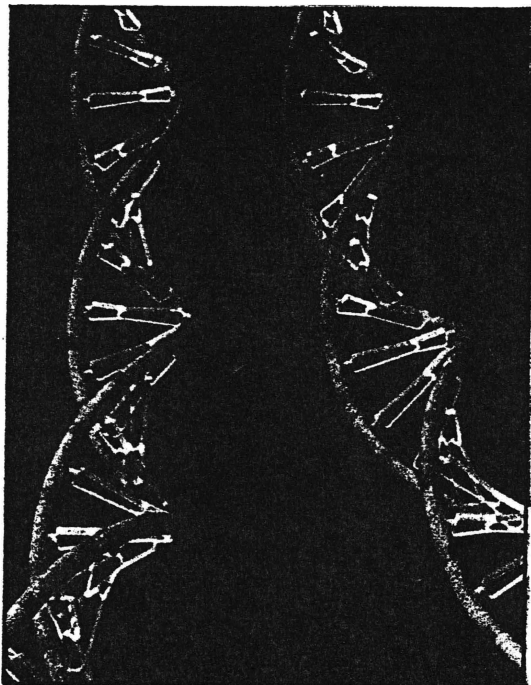
Aquest teorema no és, realment, massa atractiu, ja que utilitza llenguatges regulars infinits com a conjunts de regles d'empalmament. Afortunadament, podem prescindir de la infinitud sempre i quan imposem un mecanisme de control sobre l'ús de les regles.

Disposem, fins al moment, de diverses classes de control: contextos permissius, contextos prohibitius, metes locals, metes globals i aplicació de conveniència – veure les definicions a (Martín-Vide, 1996, pàg. 126). En tots els casos, obtenim caracteritzacions de la família RE mitjançant sistemes H extesos com a conjunts finits d'axiomes i de regles d'empalmament, aquestes últimes controlades.

D'aquests resultats es dedueix una important conseqüència per a la computació de l'ADN: existeixen sistemes H universals dels tipus mencionats, i.e, sistemes γ_u amb tots els components fixats, tal que, donat un sistema $H \gamma_0$, podem afegir un nou axioma a γ_u que codifiqui γ_0 i, aleshores, γ_u simularà γ_0 (generarà $L(\gamma_0)$). Això prova, teòricament, la possibilitat de dissenyar computadors de l'ADN universals i programables.

Ens faltaria veure com podem relacionar l'operació d'empalmar amb les gramàtiques contextuais. Podem intentar reformular les definicions donades en terme de gramàtiques contextuais.

àcid desoxiribonucleic
Dos models d'hèlixs de l'ADN
de Watson i Crick



Aparellament de les bases de l'àcid
desoxiribonucleic mitjançant ponts d'hidrogen

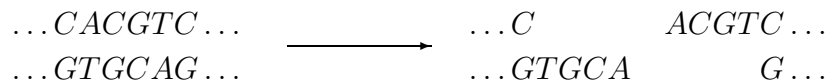
Sabem que l'operació de recombinació de les seqüències de l'ADN ¹, sota la influència d'enzims de restricció i lligament es desenvolupa de la manera següent: cada enzim reconeix una subseqüència de la forma (u, x, v) d'una seqüència d'ADN i empalma la doble tira (helicoïdal) de seqüències d'ADN en

¹L'àcid *desoxiribonucleic* o *ADN* és la molècula portadora de la informació genètica en els éssers vius. La molècula d'ADN és formada per dues llargues cadenes no ramificades de nucleòtids units covalentment per ponts 3', 5'- fosfodiester. L'esquelet d'aquestes cadenes consisteix, doncs, en grups alternats de desoxiribosa i d'àcid fosfòric units covalentment; a cada unitat de desoxiribosa va unida una base nitrogenada, que pot ésser adenina, timina, citosina o guanina. Tot i que l'ADN fou aïllat i estudiat per primera vegada ja el 1869 per F. Miescher, la seva estructura tridimensional no fou esclareïda fins pels voltants del 1950. Els estudis de difracció de raigs X, fets per Maurice H. F. Wilkins, juntament amb unes altres experiències físico-químiques, permeteren a James D. Watson i Francis H. Crick de proposar, el 1953, un model tridimensional per a l'estructura de l'ADN, el qual representà un dels avenços més importants de la història de la biologia. Segons aquest model, les dues cadenes de l'ADN van enrotllades l'una al voltant de l'altra formant una doble hèlix, a l'interior de la qual es disposen les bases aparellades mitjançant ponts d'hidrogen: l'adenina amb la timina, i la citosina amb la guanina. Així, l'ordre de bases d'una cadena imposa el de l'altra cadena complementària. Tal com postularen Watson i Crick, la informació genètica que porta l'ADN és codificada per l'ordre exacte de les bases al llarg d'una cadena i és tramesa per l'ADN segons dues vies independents: l'una anomenada *duplicació*, condueix la informació a unes noves molècules d'ADN i serveix de connexió entre generacions cel·lulars i entre la generació paterna i la filial en el cicle biològic de l'espècie; per l'altra via, la informació es transcriu a l'àcid ribonucleic o ARN, el qual la condueix als ribosomes, on té lloc la seva traducció, segons la clau genètica, de forma que dirigeix la síntesi de proteïnes. Una gran part de les proteïnes són enzims, de manera que hom pot dir que l'ADN regula les activitats metabòliques de la cèl·lula.

l'interval x , tallant una de les tires de seqüències a l'esquerra d' x i l'altra a la dreta d' x . La Figura 17.3.1 il·lustra aquest procés per a

$$(u, x, v) = \begin{pmatrix} C, & ACGT, & C \\ G & TGCA & G \end{pmatrix}.$$

Figura 17.3.1 Operació de recombinació de les seqüències de l'ADN.



Si un altre enzim talla una seqüència d'ADN d'acord a una tripla (u', x', v') amb $x = x'$, aleshores els fragments produïts pel primer enzim encaixen amb els fragments produïts pel segon enzim, poden ser enganxats (mitjançant un lligament), donant com a resultat noves seqüències d'ADN. Aquest parell d'enzims és formalitzat per una *regla d'empalmament*, de la forma $((u, x, v), (u', x, v'))$. L'aplicació d'aquesta regla s'anomena *operació d'empalmament*, definida per: si $r = ((u, x, v), (u', x, v'))$ i w_1, w_2, w_3 són cadenes sobre el vocabulari usat (quatre “lletres” en el cas de l'ADN, però l'operació pot ser definida per vocab-

ularis arbitraris), aleshores escriurem

$$\begin{aligned}
 (w_1, w_2) \vdash_r w_3 \quad \text{sii} \quad & w_1 = y_1 u x v y_2, \\
 & w_2 = z_1 u' x v' z_2, \\
 & w_3 = y_1 u x v' z_2, \\
 & \text{per algunes cadenes } y_1, y_2, z_1, z_2.
 \end{aligned}$$

S'observa novament la restricció contextual de l'acció de l'enzim: (u, v) i (u', v') controlen el lloc de tall de l'ADN. A un nivell formal, podem accentuar aquest aspecte: si la regla d'empalmament és escrita de la forma $r = ((u_1, v), (u_2, v'))$, per a $u_1 = ux, u_2 = u'x$, i escrivim

$$\begin{aligned}
 (w_1, w_2) \vdash_r w_3 \quad \text{sii} \quad & w_1 = y_1 u_1 v y_2, \\
 & w_2 = z_1 u_2 v' z_2, \\
 & w_3 = y_1 u_1 v' z_2, \\
 & \text{per a algunes cadenes } y_1, y_2, z_1, z_2,
 \end{aligned}$$

aleshores tenim una operació purament contextual equivalent a la definida anteriorment. De manera similar, podem treballar amb $r = ((u, xv), (u', xv'))$ i, per tant, en general, amb el parell de contextos, $((u_1, u_2), (u_3, u_4))$, que especifiquen el lloc on les dues cadenes que formen par de l'operació d'empalmament són tallades.

17.4 Consideracions finals

Encara que la teoria de llenguatges formals (que s'ocupa principalment de les gramàtiques de Chomsky i de les seves variants), la teoria de la computabilitat (que té a les màquines de Turing com a model estàndard d'un algorisme) i les ciències de la computació en general estan basades essencialment en l'operació de reescriptura, es poden construir models igualment potents que no usen la reescriptura, sinó operacions simples de tallar i enganxar. La reescriptura és un concepte artificial, creat pels especialistes; les operacions de tallar i enganxar, per la seva part, poden estar presents en la naturalesa de les coses. Així succeeix – segons hem vist – amb la computació de l'ADN, però també amb les operacions d'adjunció/inserció en les gramàtiques contextuais i les gramàtiques d'inserció, les quals es fonamenten en l'observació de què l'acceptació d'una cadena per un context o d'un context per una cadena, en un llenguatge donat, són fenòmens lingüístics elementals.

Ens hem ocupat aquí només de la capacitat generativa de les mencionades operacions de tallar i enganxar, i hem deixat de banda la investigació de la seva adequació descriptiva, expressivitat i eficiència computacional de les mateixes, aspectes aquests essencials a l'hora de valorar-les com a candidates superadores de la reescriptura tradicional.

Capítol 18

Caracterització dels llenguatges/llengües naturals mitjançant gramàtiques contextuals

18.1 Un model formal global de les llengües naturals: llenguatges suaument depen- dents del context

La ubicació del llenguatge natural o, com a mínim, del llenguatge natural entès com a un conjunt de cadenes, dins la jerarquia de Chomsky sembla evident: només els llenguatges no recursivament enumerables semblen estar fora de la mateixa. Assumint, doncs, que per a cada llenguatge existeix (com a mínim) una gramàtica que el genera i un autòmat que el reconeix, aleshores existeix

una gramàtica que genera exactament el conjunt de cadenes d'una llengua natural. Anàlogament a la jerarquia de llenguatges, aquesta gramàtica estarà inclosa en la jerarquia de gramàtiques de Chomsky.

És conegut que Noam Chomsky va introduir als anys cinquanta les seves gramàtiques formals com a eines per a la formalització de la sintaxi dels llenguatges naturals. Com ell afirma explícitament, (Chomsky, 1959), “el principal problema, d'immediata relevància dins la teoria del llenguatge, és el fet de determinar en quin lloc de la jerarquia de dispositius emplaçar les gramàtiques de llenguatges naturals. Seria, per exemple, summament interessant conèixer si és possible construir una gramàtica d'estructura de frase per a l'anglès.

“La jerarquia de dispositius”, en aquest context, s'entén es refereix a la jerarquia de gramàtiques de Chomsky; i aquest va ser el marc de treball usual on els investigadors tractaven de trobar “una casa” per als llenguatges naturals dins de la teoria de llenguatges formals (parafrasejant el títol de (Friedman et al., 1987)).

Des que les gramàtiques transformacionals es mostraren ineficaces per a l'explicació dels fenòmens d'ordre sintàctic (Peters i Ritchie, 1973), va cobrar molt més valor l'interès per ubicar correctament la sintaxi del llenguatge natural dins de la jerarquia de Chomsky, donada la búsqueda de gramàtiques no transformacionals per a tal representació. Més concretament, el debat se centrà en la convicció que tots els fenòmens que anteriorment rebien en tractament mitjançant transformacions podien ser realment expressats mitjançant alguna

classe de gramàtiques conegudes dins de la jerarquia. La classe de gramàtiques independents del context (*CFG*) sembla ser una eina excel·lent no només per a l'expressió lingüística, sinó també per a la implementació, donats els avençatges de complexitat que representen tals gramàtiques. En aquest sentit, el model de gramàtica d'estructura de frase generalitzada (*GPSG*, (Gazdar et al., 1985)) es va desenvolupar amb la convicció de què tots els fenòmens coneguts de la sintaxi de les llengües naturals eren expressables mitjançant conjunts de regles d'un format estrictament independent del context.

L'evidència de què la classe *CFG* no és suficient per a la representació lingüística arribaria més tard. A (Bresnan et al., 1982) es presenta el cas de dependències creuades en certes clàusules subordinades de l'holandès, mentre que Shieber (Shieber, 1985) mostra els coneguts casos del germano-suís, també relatius a dependències creuades. Finalment, destaca el treball de Culy (Culy, 1985) sobre el vocabulari del bambara, amb evidents casos de reduplicació, un llenguatge, una vegada més, depenent del context. Més controvertits són els casos de concordances múltiples per al llenguatge $\{a^n b^n c^n \mid n \geq 1\}$ o els casos de dependències creuades mitjançant l'ús de l'adverbi “respectivament”, que produeix un llenguatge depenent del context en sentit dèbil, però no en sentit fort.

Així, doncs, ben aviat s'observà que els llenguatges-llengües naturals, fins i tot l'anglès, *no eren independents del context*. Malgrat que el debat sobre aquest aspecte ha perdurat durant dues dècades, sent molts dels arguments o

demostracions que suportaven la no independència del context dels llenguatges naturals criticats i fins i tot refutats – veure, p.e., (Gazdar i Pullum, 1985), (Manaster-Ramer, 1994), (Pullum, 1985), (Gazdar i Pullum, 1982), (Friedman et al., 1987) –, sembla que, finalment, els lingüistes s’han posat d’acord en el fet que els llenguatges naturals contenen *construccions* les quals són típicament no-lliures del context ((Gazdar i Pullum, 1985), (Manaster-Ramer, 1994), (Partee et al., 1990)): reduplicació (modelada mitjançant el llenguatge formal $\{xx \mid x \in \{a,b\}^*\}$), concordances múltiples (com en el cas de $\{a^n b^n c^n \mid n \geq 1\}$), concordances creuades (il·lustrades mitjançant $\{a^n b^m c^n d^m \mid n, m \geq 1\}$). Exemples convincents d’aquestes construccions van ser observats en llengües com el bambara (Culy, 1985), mohawk (Postal, 1962), xinès-mandarí (Radzinski, 1990), així com en llengües europees com el germano-suís (Shieber, 1985), etc. Poden trobar-se arguments al respecte consultant (Higginbotham, 1984), (Langendoen, 1977), (Manaster-Ramer, 1986), (Pollard, 1984) o (Postal, 1964).

Trets similars de no independència del context també van ser observats en llenguatges de programació, veure, p.e., (Floyd, 1962), així com en moltes altres àrees on les gramàtiques formals són usades com a eines de modelització. El capítol 0.4 de (Dassow i Păun, 1989) argumenta al voltant de “set circumstàncies on les gramàtiques lliures del context no són suficients” (llenguatges naturals, llenguatges de programació, lògica, formalització de grafs en termes simbòlics, creixement biològic, modelització de processos econòmics, aproximació semiòtica formal a les narracions populars, música i arts visuals), la qual cosa ratifica el fet que “el món sembla ser no-lliure del context”.

Sembla, doncs, clara la necessitat de plantejar formalismes que sobrepassin la capacitat de les *CFG*, però que estiguin a la vegada pròpiament inclosos en la classe de gramàtiques sensibles o depenents del context (*CSG*), ja que aquesta classe de gramàtiques no és més que un sistema de reescriptura no restringida del qual s'ha tret la possibilitat d'esborrament. Així, doncs, la família de llenguatges sensibles al context és “gairebé igual” a la família de llenguatges recursivament enumerables: per a tot llenguatge recursivament enumerable L existeix un llenguatge L' associat, el qual és sensible al context, les cadenes del qual són exactament les d' L més alguna (possiblement arbitràriament llarga) prolongació o cua obtinguda a partir d'un caràcter donat (en aquest cas, sense cap estructura)¹. Per això, fins i tot si “podem assegurar (...) que els llenguatges naturals són realment sensibles al context (...) la qüestió no té un significat empíric” (Manaster-Ramer, 1994). Treballar amb gramàtiques sensibles al context és difícil (el problema de la pertanyença, malgrat ser decidible, pot resultar d'una complexitat molt gran), els símbols no-terminals no posseeixen un significat molt clar, no podem definir la semàntica de les cadenes generades en base a les derivacions (como en el cas de les gramàtiques lliures del context, en base a arbres de derivació), etc.

En aquestes circumstàncies, una idea natural va ser la de considerar gra-

¹Del Teorema 3.9.1 de la pàgina 159 inferim que tot llenguatge recursiu enumerable pot ser aproximat mitjançant un llenguatge sensible al context amb una cua. El resultat del teorema de l'espai de treball de la Secció 3.10 de la 160 ens diu que, en principi, si aquesta cua no és “massa llarga”, aleshores esborrant-la encara roman en la família de llenguatges sensibles al context.

màtiques basades en regles lliures del context usades d'una manera no-lliures del context, amb l'objectiu de generar una superfamília estricta de la lliure del context que, no obstant, conservi tantes propietats de les lliures del context com sigui possible. Una poderosa branca de la teoria de llenguatges formals es va desenvolupar seguint aquest camí: la *reescriptura regulada*.

Alguns intents relacionats amb restriccions en la derivació de gramàtiques lliures del context són les gramàtiques *matricials* i similars a matrius (les produccions lliures del context estan agrupades en “matrius”, “vectors”, i les produccions de cada grup són usades conjuntament); les gramàtiques *programades* i *controlades* (la restricció és imposada en la seqüència de regles usades); les gramàtiques *de context aleatori* (una regla és usada solament en presència o absència de certs símbols no-terminals); les gramàtiques *condicionals* (una condició global per a què una cadena pugui ser reescrita ha de ser satisfeta per tal de poder usar la producció), les gramàtiques *paral·leles* (totes les ocurrències de símbols no-terminals i terminals han de ser reescrites en un mateix pas de derivació), les gramàtiques *de valències* (cada regla té associada un valor i una derivació és acceptada només si la suma o el producte de les valències de les regles usades en aquella derivació és igual al valor preescrit), etc. Definicions, resultats, i referències bibliogràfiques poden ser trobades en aquesta tesi a la Secció 3.4, pel que fa a les gramàtiques de Chomsky lliures del context; i a les Seccions 10, 11, 13 i 12, pel cas particular de les gramàtiques contextuais.

Quasibé totes les classes de gramàtiques d'aquest tipus tenen els mateixos

aventatges (generen tots els llenguatges lliures del context, així com les construccions no-lliures del context dels llenguatges naturals) i inconvenients (el seu poder és “massa gran”, el problema de la pertinença és difícil, llenguatges del tipus $\{a^{2^n} | n \geq 1\}$, sense cap significat per als llenguatges “reals”, poden ser generats algunes vegades). Així, si les gramàtiques CF són el “límit inferior” de la representació lingüística, on és el “límit superior”?

Joshi (Joshi, 1985) planteja la descripció d’aquest límit superior com l’enumeració d’una sèrie de característiques del que ell denomina *Mildly Context Sensitive Grammars* ($MCSG$, $MCSL$ pels corresponents llenguatges). Basant-se en les propietats formals de les “gramàtiques d’adjunció d’arbres” $TAGs$, Joshi proposa que la classe de llenguatges adequada per a la descripció del llenguatge natural contingui aquells llenguatges que compleixen les tres següents propietats:

- 1) $CFL \subset MCSL \subset CSL$.

Aquesta primera propietat ens diu, simplement, que $MCSL$ conté tots els fenòmens que es troben en CFL , però, a la vegada, $MCSL$ inclou només cert tipus de dependències del context (llenguatges de còpia, o dependències creuades), però no de totes elles (per exemple, $L = \{x = \{a, b, c\}^* \mid |x|_a = |x|_b = |x|_c\}$, on el nombre d’as, bes i ces és el mateix).

- 2) $MCSL$ poden ser processats en temps polinòmic.

Aquesta segona propietat busca mantenir els aventatges de processament que suposen les CFG (processables en temps cúbic), buscant un proces-

sament que en cap cas excedeixi el temps polinòmic.

3) *MCSL* tenen creixement acotat (*constant growth property*).

Aquesta tercera propietat significa que, si les cadenes d'un llenguatge són ordenades en ordre de longitud creixent, aleshores dues cadenes consecutives no es diferencien per un nombre arbitrari d'elements (i.e, hi ha un "creixement acotat", existeix una acotació en la diferència entre la longitud de qualsevol frase del llenguatge i la longitud de la frase immediatament més llarga i pertanyent al llenguatge). De fet, qualsevol longitud donada pot ser descrita com una combinació lineal d'un conjunt finit de longituds fixades. Algunes vegades, aquesta propietat s'expressa en termes de semilinealitat, la qual cosa és una propietat molt més forta.

En una *CFG*, al final de la derivació obtenim un arbre d'estructura de frase, de manera que la cadena terminal obtinguda és un element d'un llenguatge de *CFL*. Sigui $G = (V_N, V_T, S, P)$ una gramàtica independent del context tal que generi w_i, w_{i+1} , $G \Rightarrow^n w_i$ i $G \Rightarrow^{n+1} w_{i+1}$ (i.e., G aplica a w_{i+1} un pas més que a w_i). Ambdues cadenes, w_i i w_{i+1} , pertanyen al mateix llenguatge, i la longitud de w_{i+1} és igual a la longitud de w_i més la longitud de la cadena terminal, que denotarem per a_j , de la regla que s'ha aplicat addicionalment a w_{i+1} (no comptem els símbols no terminals a p , per a $p \in V_N \times (V_N \cup V_T)^*$). Així,

$$|w_{i+1}| = |w_i| + |a_j|,$$

on $|x|$ denota la longitud d' x . D'aquesta manera, les longituds de les

cadena terminals (que són frases del llenguatge) augmenten en una constant, pertanyen al conjunt de longituds fixades per les cadenes terminals de les regles de P . Per tant, per a una cadena $w \in L(G)$ tenim

$$|w| = a_1|w_1| + a_2|w_2| + \dots + a_i|w_i| + a_m|w_m|, \quad a_i \geq 0, 1 \leq i \leq m,$$

on $w_i, 1 \leq i \leq m$, és la cadena terminal de cadascuna de les regles (aplicables i vegades), assumint que hi ha m regles. La propietat de creixement acotat, doncs, no és més que una caracterització aproximada de la intuïció lingüística de què les frases d'una llengua natural es construeixen a partir d'un conjunt de clàusules d'estructura limitada usant certes operacions de combinació lineal.

És senzill de veure que, dins de CSL , alguns llenguatges, com $\{a^n b^n c^n \mid n \geq 1\}$ o $\{a^n b^m c^n d^m \mid n, m \geq 1\}$, posseeixen aquesta propietat, mentre que $\{a^{n^2} \mid n \geq 1\}$ o $\{a^{2^n} \mid n \geq 1\}$ o $\{a^p \mid p \text{ és primer}\}$ no la tenen.

No obstant, aquestes propietats no defineixen pròpiament la classe $MCSG$ (i $MCSL$ per a llenguatges), sinó que ofereixen un conjunt de condicions necessàries per a aquesta caracterització. Basant-se en les propietats de les gramàtiques d'adjunció d'arbres (TAG), s'han suggerit tres formalismes més per a la representació del llenguatge natural, donades les propietats d'equivalència que se'n deriven: gramàtiques amb nucli (*Head Grammars* o HG (Pollard, 1984), les gramàtiques indexades (*Indexed Grammars* o IG (Aho, 1968) i les gramàtiques categorials *Categorical Grammars* o CG (Ajdukiewicz, 1935) que també són suaument sensibles al context.

Tals formalismes són diferents entre ells en el sentit que s'utilitzen objectes diferents sota regles o tipus de combinació diferents. En cada cas, doncs, s'especifica un domini de localitat diferent, motivat per la diferent aproximació que es produeix en cada formalisme als fenòmens del llenguatge natural.

Recentment, s'ha seguit una nova estratègia: incrementar el poder de les gramàtiques independents del context mitjançant *cooperació*, considerat *sistemes de gramàtiques*. D'una manera informal, es consideren construccions consistents en diverses gramàtiques usuals treballant conjuntament, de comú acord amb un protocol especificat, amb la finalitat de generar un llenguatge comú. Les variants bàsiques són: els sistemes *seqüencials* (els sistemes de distribució cooperativa *CDGS* introduïts a (Csuha-j-Varju i Dassow, 1990)), on els components treballen per torns i, en cada moment, només un d'ells està actiu; i els sistemes *paral·lels* (els sistemes de gramàtiques de comunicació en paral·lel *PCGS* introduïts a (Păun i Sântean, 1989)), els components dels quals treballen sincrònicament, cadascun d'ells en la seva pròpia frase, i cooperen mitjançant consultes o ordres.

Com Manaster Ramer menciona, "l'exposició donada per Chomsky sembla suggerir que la classe dels llenguatges naturals es pot trobar en algun lloc de la jerarquia de Chomsky. No obstant pot succeir que aquest no sigui el cas, i probablement és així. Pot succeir, per exemple, que una teoria real dels llenguatges naturals defineixi una classe de llenguatges que sigui inconmensurable amb els tipus de Chomsky, per exemple, que siguin llenguatges una mica re-

gulars, llenguatges una mica no regulars i lliures del context, llenguatges una mica no-lliures del context, etc.” (Manaster-Ramer, 1994).

Alguns tipus de gramàtiques, com és el cas de les denominades gramàtiques contextuais de Marcus o, simplement, *gramàtiques contextuais*, posseeixen aquesta possibilitat.

18.1.1 Exemples de “construccions” no-independents del context en els llenguatges/llengües naturals

En la Secció 3.8 hem mostrat un dels primers exemples, proposat per Bar-Hillel i Shamir, (Bar-Hillel i Shamir, 1964), referent a “construccions” no-lliures del context en el llenguatge natural.

En primer lloc, farem un recorregut històric i veurem les diferents propostes que han anat apareixent en el transcurs d’aquest segle.

Langendoen (Langendoen, 1977, pàg. 320-322) considera que certs patrons de prefixació i sufixació poden, en principi, conduir cap a un conjunt de paraules no-lliures del context. Si certs prefixos precisen la presència de certs sufixos, poden resultar conjunts de paraules de, per exemple, el tipus $\{a^m cb^n \mid m=n+1\}$.

No obstant, Langendoen no es mostra molt convençut d’aquest procediment, basant-se en el fet de desconèixer si l’absència de conjunts de paraules que són lliures del context és accidental, o bé se segueix a partir de certs

principis de formació de paraules prèviament postulats.

Gazdar i Pullum, (Gazdar i Pullum, 1985, pàg. 277), consideren que certs fets referenciats per Culy, (Culy, 1985), suggereixen que la qüestió plantejada per Langoenden pot ser contestada. El bambara, una llengua africana de la família del mande, sembla ser que posseeix un conjunt de paraules que és no lliure del context. En bambara, les paraules compostes s'obtenen a partir d'una base nominal *w*, obtenint paraules que posseeixen l'estructura “*w-o-w*” amb el significat de “qualsevol *w*”. Per exemple, del fet que *wulu* significa “gos”, es deriva que *wulu-o-wulu* significa “qualsevol gos”. També poden formar-se bases compostes de noms de longitud arbitrària: *wulu-filela* (“gos-vigilant”), *wulu-nyinita* (“gos-caçador”), *wulu-filela-nyita* (“gos-vigilant-caçador”), i així successivament, podem obtenir per exemple, *wulu-filela-nyita-o-wulu-filela-nyita* (“qualsevol gos-vigilant-caçador”). La reduplicació del nom s'aplica, en aquest cas, a una classe de bases nominals de longitud arbitrària.

Com Gazdar i Pullum ens expliquen a (Gazdar i Pullum, 1985, pàg. 277–278), Manaster-Ramer ha observat, en conferències no publicades, que altres llenguatges posseeixen fenòmens similars; construccions de reduplicació amb longitud arbitrària apareixen en polonès, turc i altres llenguatges.

Més interessant és la situació dels llenguatges naturals respecte a les gramàtiques lliures del context a nivell de frases, i.e., de cadenes de paraules. Sempre ha existit un acord generalitzat envers el fet que els llenguatges naturals són no-lliures del context, però els arguments que van sorgir al respecte eren

controvertits. Durant el primer període de desenvolupament de les gramàtiques generatives, la idea comú era, per exemple, que el fenomen de concordança subjecte-verb que apareix en l'anglès era no-lliure del context; no obstant, això no és així; fins i tot els llenguatges regulars poden tenir dependències entre símbols arbitraris allunyats entre sí. Gazdar i Pullum (Gazdar i Pullum, 1982, pàg. 471–504) ens donen una visió crítica de la literatura apareguda al voltant d'aquest tòpic i la seva conclusió és que “els plantejaments exposats en el camp de les gramàtiques generatives generalment no han ofert cap cosa que pugui ser presa seriosament com a argument que els llenguatges naturals no són lliures del context. Pitjor encara, fins i tot els escrits tècnics ens mostren més de vint-i-cinc anys d'esforços infructuosos per a intentar provar que no tots els llenguatges naturals són lliures del context”.

A més a més d'estudiar les fal·làcies relacionades amb la concordança subjecte-verb, analitzen cinc altres nous tipus d'arguments que podrien sustentar la no-independència del context dels llenguatges naturals:

- (1) construccions de l'anglès amb *respectively* (Bar-Hillel i Shamir, 1964), (Langendoen, 1977));
- (2) clàusules comparatives de l'anglès (Chomsky);
- (3) reduplicació i incorporació de bases nominals en mohawk (Postal, 1962);
- (4) dependències creuades de verbs i objectes en certes construccions de clàusules subordinades en holandès (Huybregts, 1976);

(5) assercions incloent expressions numèriques (J. Elster).

En posteriors investigacions (Gazdar i Pullum, 1985, pàg. 281)) també analitzen altres “esforços erronis”:

(6) clàusules *such that* de l’anglès (Higginbotham, 1984);

(7) clàusules “sluicing” de l’anglès (Langendoen, 1977) i (Postal, 1962).

Gazdar i Pullum observen que els dos últims arguments estan basats en falses pressuposicions al voltant del que és o no és gramatical en anglès.

Un nou exemple de construcció no-lliure del context en llenguatge natural va ser trobat per Shieber (Shieber, 1985) en el germano-suís (els dialectes parlats al voltant de Zürich):

“ un nombre arbitrari de sintagmes nominals (SN) poden ser seguits per un verb finit i un nombre específic de verbs no finits, sent el nombre de SN una funció de les propietats lèxiques dels verbs i exhibint la relació semàntica entre verbs i SN un patró de dependències creuades: els verbs situats a la dreta de la cadena de verbs prenen com a objectes seus SN situats a la dreta de la cadena de SN . La subcadena principal té la forma $SN^m V^n$.

Figura 18.1.1 En un cas senzill, quan $m = n = 5$, la mencionada cadena pot tenir un significat semblant a:

“ Alf veié Bob deixar Cal ajudar Don fer Ed treballar”,

però amb una ordenació de les paraules corresponent a

<i>Alf</i>	<i>Bob</i>	<i>Cal</i>	<i>Don</i>	<i>Ed</i>	<i>veié</i>	<i>deixar</i>	<i>ajudar</i>	<i>fer</i>	<i>treballar</i>
SN_1	SN_2	SN_3	SN_4	SN_5	V_1	V_2	V_3	V_4	V_5

Un patró similar va poder ser observat en l'holandès i, tal com va ser observat per Pullum i Gazdar, no sembla ser una estructura no-lliure del context. En el cas del germano-suís, a diferència de l'holandès, existeix una propietat addicional: certs verbs precisen preferentment cas datiu en comptes de cas acusatiu en els seus objectes. Aquest patró és, en general, una estructura no-lliure del context. Per tal de veure-ho, podem usar el mètode explicitat a la Secció 3.8: la intersecció entre un llenguatge CF i un llenguatge regular és un llenguatge CF . Mitjançant una adequada selecció del llenguatge regular, obtenim les clàusules on tots els SN acusatius (SN_a) precedeixen a tots els SN datius (SN_d) i, les clàusules gramaticals són, precisament, aquelles en què l'acusatiu que reclama verbs (V_a) precedeix al datiu que reclama verbs (V_d) i la quantitat dels quals concorda. Esquemàticament: $SN_a SN_d V_a V_d$, que té la forma del llenguatge $\{a^n b^m c^n d^m \mid n, m \geq 1\}$, que, com sabem, és no-lliure del context. Una demostració més detallada pot trobar-se a (Partee et al., 1990, pàg. 502–503).

Altres possibles estructures no-lliures del context en llenguatges naturals són discutides per Gazdar i Pullum. És important de remarcar la conclusió a

la qual arriben aquests autors: mai s'ha pogut demostrar que l'anglès es troba fora de la classe dels llenguatges lliures del context. No obstant, mencionen a Manaster-Ramer (comunicació personal), per suggerir que la construcció idiomàtica de l'anglès exemplificada per *RS-232 or no RS-232, this terminal isn't working* (on el patró X o $no\ X$ és essencial per a la seva acceptabilitat, i X pot prendre infinits valors), pot ser un exemple de no-independència del context en l'anglès. Savitch, referint-se a Manaster-Ramer, (Savitch, 1989), també cita aquest mateix patró, amb l'exemple de: *reduplication or no reduplication, I want a parse tree*, (Manaster-Ramer, 1983; Manaster-Ramer, 1986).

En conclusió, ens interessa usar gramàtiques independents del context i afegir alguns mecanismes per tal d'extreure'n alguns subconjunts (del llenguatge generat) amb la finalitat d'incloure-hi alguns aspectes dels llenguatges-llengües naturals i/o llenguatges de programació (i/o altres camps d'interès). L'objectiu és el d'obtenir una subfamília “suau” de llenguatges depenents del context (el terme “suau”, “mild” en anglès, s'usa en lingüística matemàtica amb un significat molt concret: semilineal i analitzable (o computable) en temps polinomial), amb el major nombre possible de propietats similars a les independents del context, però capaç d'incloure els trets de no-independència del context dels llenguatges amb els quals treballem.

18.2 Un model formal de les llengües naturals: els llenguatges contextuais amb ús maximal de selectors

En relació a la possibilitat suggerida per Manaster-Ramer, i apuntada anteriorment, pel que fa a l'existència de famílies de llenguatges que, sent suaument dependents del context, siguin incomparables amb la família de llenguatges de la jerarquia de Chomsky, veurem a continuació que una variant de les gramàtiques contextuales internes recentment introduïda, les denominades *gramàtiques contextuales internes amb ús maximal de selectors* (Martín-Vide et al., 1995a), les quals són capaces d'incloure, d'una manera molt senzilla, les tres construccions no-independents del context mencionades anteriorment; són analitzables en temps polinomial i són semilineals (i, pertant, són suaument dependents del context).

La idea és molt simple: quan un context és adjuntat a una cadena que el selecciona, aquesta cadena-selector ha de ser maximal; cap altra supercadena estricta de la mateixa pot ser usada com a selector. A la Secció 10.1 hem donat algunes definicions bàsiques i hem estudiat el poder de les gramàtiques contextuales amb ús minimal i maximal dels selectors. El que veurem a continuació és la importància d'aquestes gramàtiques per a l'estudi dels llenguatges naturals (centrant-nos en les construccions bàsiques de no-independència del context dels llenguatges-llengües naturals).

Deixarem de banda la demostració de del fet que aquests llenguatges són analitzables en temps polinomial – veure al respecte la Secció 4 de (Ilie, 1996a) – i que posseeixen la propietat de creixement acotat – veure la Secció 4 de (Marcus et al., 1996a).

Ens centrarem en la seva adequació per tal d'incloure les construccions del llenguatge/llengua natural no independents del context.

Ja hem vist que els llenguatges $\{xcx \mid x \in \{a, b\}^*\}$ (reduplicació) i $\{a^n b^n c^n \mid n \geq 1\}$ (dependències múltiples) no pertanyen a ICC , però, com ja hem apuntat diverses vegades, poden ser generats de manera molt simple mitjançant gramàtiques amb els selectors usats en el mode maximal. Per altra banda, ja hem vist que $L = \{a^n b^m c b^m a^n \mid n, m \geq 1\}$ no pertany a $ICC_{Mg}(REG)$. Per tant, les gramàtiques contextuais internes amb ús maximal (global) dels selectors engloben les construccions bàsiques no independents del context dels llenguatges naturals, però no la construcció d'imatge mirall a L . Això recorda el fet que “sabem que en els llenguatges naturals abunden les reduplicacions però escassegen les imatges mirall (Friedman et al., 1987), i sembla ser també que, semblantment, les estructures d'encreuament són més fàcils de processar que les incrustades pel centre (Manaster-Ramer, 1994).” El llenguatge L és basa en aquesta estructura imatge-mirall/incrustada pel centre i, per tant, podem interpretar la relació $L \notin ICC_{Mg}(REG)$ com un argument que prova l'adequació d'aquestes gramàtiques com a model de (certes construccions dels) llenguatges-llengües naturals.

Investigarem la possibilitat de generació d'alguns llenguatges que representen construccions que apareixen o que se suposa no apareixen en els llenguatges/llengües naturals. Específicament, considerarem els següents llenguatges:

$$M_1 = V^*, \quad V \text{ és un vocabulari donat (cap estructura),}$$

$$M_2 = \{a^n b^n \mid n \geq 1\} \text{ (concordança simple),}$$

$$M_3 = \{a^n b^n c^n \mid n \geq 1\} \text{ (concordança triple),}$$

$$M_4 = \{a^n b^n c^n d^n \mid n \geq 1\} \text{ (concordança quàdruple),}$$

$$M_5 = \{a^n b^n c^n d^n e^n \mid n \geq 1\} \text{ (concordança quintuple),}$$

$$M_6 = \{xcx \mid x \in \{a, b\}^*\} \text{ (reduplicació marcada),}$$

$$M'_6 = \{xx \mid x \in \{a, b\}^*\} \text{ (reduplicació no-marcada),}$$

$$M_7 = \{xc \, mi(x) \mid x \in \{a, b\}^*\} \text{ (imatge mirall marcada),}$$

$$M'_7 = \{x \, mi(x) \mid x \in \{a, b\}^*\} \text{ (imatge mirall no-marcada),}$$

$$M_8 = \{a^n b^m c b^m a^n \mid n, m \geq 1\} \text{ (imatge mirall marcada particular),}$$

$$M_9 = \{a^n b^m c^n d^m \mid n, m \geq 1\} \text{ (concordances creuades).}$$

Les Taules 18.2.1 i 18.2.2 presenten les possibilitats de generació d'aquest llenguatges mitjançant gramàtiques contextuais internes associades a les famílies $ICC(F)$, $ICC_\alpha(F)$, $\alpha \in \{Ml, Mg\}$, $F \in \{FIN, REG\}$: S indica Sí, N indica No, i $?$ indica que resta per esbrinar.

Taula 18.2.1 Possibilitats de generació d'aquest llenguatges mitjançant gramàtiques contextuais internes associades a les famílies $ICC(FIN)$, $ICC(REG)$, $ICC_{Ml}(FIN)$.

	$ICC(FIN)$	$ICC(REG)$	$ICC_{Ml}(FIN)$
M_1	S	S	S
M_2	S	S	S
M_3	N	N	N
M_4	N	N	N
M_5	N	N	N
M_6	N	N	N
M'_6	N	N	N
M_7	S	S	S
M'_7	N	N	N
M_8	S	S	S
M_9	N	S	N

Taula 18.2.2 Possibilitats de generació d'aquest llenguatges mitjançant gramàtiques contextuais internes associades a les famílies $ICC_{Ml}(REG)$, $ICC_{Mg}(FIN)$, $ICC_{Mg}(REG)$.

	$ICC_{Ml}(REG)$	$ICC_{Mg}(FIN)$	$ICC_{Mg}(REG)$
M_1	S	S	S
M_2	S	S	S
M_3	S	N	S
M_4	S	N	S
M_5	N	N	N
M_6	S	N	S
M'_6	N	N	?
M_7	S	S	S
M'_7	S	N	S
M_8	S	N	N
M_9	S	N	S

Només donarem algunes claus per a les demostracions d'aquestes afirmacions:

1. $G_1 = (V, \{\lambda\}, (\{\lambda\}, \{(\lambda, a) \mid a \in V\}))$ genera V^* en tots els modes.

2. $G_2 = (\{a, b\}, \{ab\}, (\{ab\}, \{(a, b)\}))$ genera M_2 en tots els modes.

3. Al començament d'aquesta secció ja hem apuntat que M_3 i M_6 no pertanyen a ICC i que aquests llenguatges i M_9 pertanyen a $ICC_{Ml}(REG)$, $ICC_{Mg}(REG)$.

Recordarem ara les gramàtiques corresponents, per tal de mostrar la seva sim-

plicitat:

$$G_3 = (\{a, b, c\}, \{abc\}, (b^+, \{(a, bc)\})),$$

$$G_6 = (\{a, b, c\}, \{c\}, (\{c\}\{a, b\}^*, \{(a, a), (b, b)\})),$$

$$G_9 = (\{a, b, c, d\}, \{abcd\}, (ab^+c, \{(a, c)\}), (bc^+d, \{(b, d)\})).$$

4. En la Secció 10.1.2 (Lema 10.1.10 de la pàgina 381) es demostra que $\{a^n b^m a^n b^m \mid n, m \geq 1\} \notin ICC_\alpha(FIN), \alpha \in \{Ml, Mg\}$. Similarment, tenim que $M_3 \notin ICC_\alpha(FIN)$.

5. L'afirmació per a M_4 pot ser provada de la mateixa manera que per la de M_3 ; p.e., per tal de produir cadenes de longitud arbitrària necessitem selectors de longituds arbitràries. Per altra banda, la gramàtica $G_4 = (\{a, b, c, d\}, \{abcd\}, (b^+c^+, \{(ab, cd)\}))$ genera el llenguatge M_4 en els modes Ml i Mg .

6. No podem esborrar cap context no-buit d'una cadena de M_5 i, per tant, una gramàtica contextual no pot generar M_5 , sigui quin sigui el mode de derivació.

Similarment, cap llenguatge, les cadenes del qual continguin més de cinc subcadenes correlades pot ser contextual.

7. Suposem que $M_6 \in ICC_\alpha(FIN), \alpha \in \{Ml, Mg\}$. Prenem una gramàtica G per a M_6 , i considerem una cadena de la forma

$$z_i = aba^2b^2 \dots a^i b^i caba^2b^2 \dots a^i b^i,$$

per a un enter i suficientment gran. Per tal de produir aquesta cadena necessitem una derivació $w = w_1 w_2 w_3 \Rightarrow_{in} w_1 u w_2 v w_3 = z_i$. És obvi que $|w_2|$ depèn d' i i, per tant, no pot ser acotat. Per tant, G no pot tenir només selectors finits.

8. Suposem que $M'_6 \in ICC(F)$, per a $F \in \{FIN, REG\}$. Prenem $G = (\{a, b\}, A, (S_1, C_1), \dots, (S_n, C_n))$ tal que $L_{in}(G) = M'_6$. Com a mínim existeix un context (u, v) a G amb $uv \neq \lambda$; totes les cadenes d' M'_6 tenen una longitud parell i, per tant, $|uv|$ ha de ser parell. Prenem x en el selector de (u, v) i considerem les cadenes $xa^i xa^i$, $i \geq |uv|$. Aleshores $uxva^i xa^i \in L_{in}(G)$ i, per tant, $uxva^i xa^i = yy$, per a algun $y \in \{a, b\}^*$. Aquest fet implica que

$$uxva^{i-|uv|/2} = a^{|uv|/2} xa^i,$$

és a dir $uxv = za^j$, $j \geq 1$. A partir de $xb^i xb^i$ obtenim, de la mateixa manera, que $uxv = z'b^k$, $k \geq 1$; la qual cosa és una contradicció.

Com en el punt 7, obtenim que $M'_6 \notin ICC_\alpha(FIN)$, $\alpha \in \{Ml, Mg\}$.

9. Suposem que $M'_6 = L_{Ml}(G)$ per a algun $G = (\{a, b\}, A, (S_1, C_1), \dots, (S_n, C_n))$ amb selectors regulars. Denotem

$$\varphi(x) = \{(u, v) \mid (u, v) \in C_i, x \in S_i, 1 \leq i \leq n\}, \quad x \in \{a, b\}^*,$$

$$\varphi^{-1}((u, v)) = \{x \in \{a, b\}^* \mid (u, v) \in \varphi(x)\}, \quad (u, v) \in C_i, 1 \leq i \leq n.$$

Totes les cadenes $a^m b a^m b$, $m \geq 1$, pertanyen a M'_6 . Prenem aquesta cadena per a un valor d' m arbitràriament gran. Si existeix un pas de derivació

$a^q \Rightarrow_{MI} a^m b a^m b$, aleshores existeix un context $(u, v) = (a^{i_1} b a^{i_2}, a^{i_3} b) \in \varphi(a^p)$, per a algun $p \leq q$. Com que $m = i_2 + p + i_3$, se segueix que p és arbitràriament gran. El conjunt $\varphi^{-1}((u, v))$ és regular (és la unió finita de conjunts regulars) i, per tant, conté un nombre infinit de cadenes de la forma a^s (apliquem un lema de bombeig a a^p a $\varphi^{-1}((u, v))$). Per tant, l'ús de (u, v) ha de ser fet per un selector maximal de la forma a^t . D'aquesta manera, podem produir una cadena $a^{j_1} b a^{j_2} b a^{j_3}$, amb j_1, j_3 acotats i un valor de j_2 arbitràriament gran. Aquesta cadena no pertany a M'_6 ; la qual cosa és una contradicció.

Per tant, en la derivació de $a^m b a^m b$ hi ha un nombre arbitrari de passos de derivació de la forma

$$a^s b a^s b \Rightarrow_{MI} a^{s+k} b a^{s+k} b,$$

amb $k \geq 1$ i $a^p b a^q \in \varphi^{-1}((a^k, a^k))$. Considerem ara una cadena

$$w = a^{i_1} b a^{i_2} b a^{i_1} b a^{i_2} b,$$

amb i_1, i_2 arbitràriament grans. Cadascuna d'aquestes cadenes pertany a M'_6 . Si l'anterior $a^p b a^q$ o qualsevol altre selector de la forma $a^r b a^{r'}$ de $\varphi^{-1}((a^k, a^k))$ és maximal a w , aleshores podem produir una cadena que no pertanyi a M'_6 . Per altra banda, $a^p b a^q$ és una subcadena de w i, per tant, els selectors inclosos a $\varphi^{-1}((a^k, a^k))$ han de contenir una cadena que és una supercadena estricta de $a^p b a^q$, per tal de prevenir la generació d'una cadena paràsita. Aquesta supracadena només pot ser de les formes $a^{j_1} b a^{i_2} b a^{j_2}$ o $a^{j_1} b a^{i_1} b a^{j_2}$. En ambdós casos, la subcadena intermèdia, $b a^{i_2} b$, $b a^{i_1} b$, respectivament, és de longitud arbitrària.

Com en el cas dels elements d'un llenguatge regular, aquestes cadenes tenen propietats de bombeig. Considerem el cas de $ba^{i_2}b$ (el segon cas és similar). Això significa que totes les cadenes de la forma

$$z = a^{j_1}ba^{i_2+rh}ba^{j_2},$$

per a algun $r \geq 1$ i tot $h \geq 0$ pertanyen a $\varphi^{-1}((u, v))$. Prenem aquesta cadena z amb h suficientment gran per tal que

$$i_2 + rh > j_1 + j_2.$$

Considerem la cadena

$$w = a^{i_2+rh-j_2}ba^{i_2+rh}ba^{j_2}.$$

Pel fet que

$$i_2 + rh = (i_2 + rh - j_2) + j_2,$$

tenim $w \in M'_6$. Ja que

$$i_2 + rh - j_2 > j_1,$$

el context (a^k, a^k) és aplicable a w , és a dir,

$$w \Longrightarrow_{ML} a^{i_2+rh-j_2+k}ba^{i_2+rh}ba^{j_2+k}.$$

La cadena obtinguda no pertany a M'_6 ; la qual cosa és una contradicció. En conclusió, M'_6 no pot pertànyer a $ICC_{Ml}(REG)$.

L'argument anterior no es compleix pel cas de la derivació global maximal i, per tant, la relació $M'_6 \in ICC_{Mg}(REG)$ resta per esbrinar.

10. Per a la gramàtica $G = (\{a, b, c\}, \{c\}, (\{c\}, \{(a, a), (b, b)\}))$ tenim $L_\alpha(G) = M_7$ per a tot α .

11. El fet que $M'_7 \notin ICC(REG)$ està demostrat en el Lema 5.1.9 de la pàgina 196. Com en el cas de M_6 , podem provar que $M'_7 \notin ICC_\alpha(FIN)$, $\alpha \in \{Ml, Mg\}$. Per altra banda, per a la gramàtica $G = (\{a, b\}, \{\lambda\}, (\{a, b\}^*, \{(a, a), (b, b)\}))$ tenim $L_{Ml}(G) = L_{Mg}(G) = M'_7$.

12. Els resultats per a M_8 són provats en els Lemes 10.1.7 de la pàgina 375 i 10.1.11 de la pàgina 381.

13. Per tal de generar cadenes $a^n b^m c^n d^m$ a M_9 amb n i m arbitràriament grans, clàrament, necessitem selectors de longitud arbitrària i, per tant, $M_9 \notin ICC_\alpha(FIN)$, $\alpha \in \{Ml, Mg\}$.

Val la pena de mencionar alguns fets relacionats amb les taules anteriors:

- El mode maximal (global o local) d'usar els selectors no incrementen necessàriament el poder de les gramàtiques contextuais, però – i aquest és un fet molt important des del punt de vista d'ús d'aquestes gramàtiques com a models de la sintaxi dels llenguatges naturals – totes les construccions bàsiques no-independents del context dels llenguatges naturals poden ser englobades d'aquesta manera (i no poden ser englobades per les gramàtiques contextuais internes usals).
- Hi ha una clara diferència entre els llenguatges “marcats” i els “no-marcats”,

sent els primers més fàcil de manejar que els altres.

- Sorprenentment, des del punt de vista matemàtic, però molt útil des del punt de vista del llenguatge natural, les gramàtiques amb un ús global maximal dels selectors no poden englobar les construccions d'imatge mirall que, per exemple, poden ser representades pel llenguatge M_8 .
- Hi ha també una clara diferència entre les gramàtiques contextuais i les gramàtiques de Chomsky, respecte als llenguatges enumerats anteriorment. Per exemple, M_7 i M'_7 són de la mateixa complexitat quan són generades mitjançant gramàtiques de Chomsky (els dos llenguatges són lineals i poden ser generats mitjançant gramàtiques quasibé idèntiques). En el marc de treball de les gramàtiques contextuais, M_7 i M'_7 són significativament diferents. Això succeeix també per M_6 i M'_6 . El cas de les gramàtiques contextuais s'aproxima a la nostra intuïció, ja que l'existència d'un marcador fa molt fàcil la verificació de la propietat que defineix la cadena en els nostres llenguatges (coneixent el “centre”, podem verificar directament la relació entre les dues meitats de la cadena).

Capítol 19

Conclusions finals

La teoria de gramàtiques formals que està en la mateixa base de la sintaxi formal ha fet, des de Chomsky, un ús quasibé exclusiu dels procediments de reescriptura. En efecte, a partir de la jerarquia de Chomsky, s'han desenvolupat diverses tècniques generatives, però totes elles més o menys pròximes a aquelles eines originals. Aquesta circumstància podria fer-nos pensar que la reescriptura és inevitable en la teoria de llenguatges formals, en la teoria d'autòmats, en la teoria d'algorismes o en les ciències de la computació en general.

Sabem que les màquines de Turing caracteritzen els llenguatges recursivament enumerables, la classe de llenguatges més àmplia que es pot concebre. Doncs bé, en aquests últims anys han aparegut eines generatives (en el sen-

tit més ampli de la paraula) que no fan ús de la reescriptura i que tenen la mateixa capacitat generativa que els dispositius clàssics. I, a més a més, resulta que aquestes noves eines semblen ser “més naturals” que la reescriptura, en el sentit de què no són simples artificis creats amb finalitats de descripció científica, sinó en certa manera operacions presents, d’una o altra forma, en la mateixa naturalesa. Si la capacitat d’aquests nous procediments no va a la saga de la dels procediments clàssics, la seva adequació per a la descripció del llenguatge natural, pel que fa a aquest nou paràmetre de naturalitat, pot ser superior. Un d’aquests mecanismes són les denominades *gramàtiques de Marcus* o *gramàtiques contextuais*.

Com Solomon Marcus, el creador d’aquestes gramàtiques (Marcus, 1969a), ha dit en nombroses ocasions, el desafiament era definir un mecanisme generatiu intrínsec, que no fes ús de símbols auxiliars, basat no en la reescriptura sinó en l’operació lingüística fonamental d’adjuntar cadenes / contextos d’acord a un procés de selecció. En la lingüística descriptiva l’associació de certes cadenes amb certs contextos (parells de cadenes), respecte a un llenguatge donat, és l’ingredient bàsic de moltes (sinó totes) les teories morfològiques. A partir d’un conjunt donat de frases correctes i incrementant-lo afegint contextos acceptats per aquestes mateixes frases o per certes subfrases de les mateixes, podem produir un llenguatge, possiblement infinit. Aquesta és, exactament, la manera de treballar de les gramàtiques contextuais.

L’estudi matemàtic de les gramàtiques contextuais es desenvolupà molt

ràpidament. No obstant, es donà una situació paradoxal: malgrat estar directament motivades per investigacions referents als llenguatges/llengües naturals, les gramàtiques contextuais fóren estudiades durant més de vint-i-cinc anys com a objectes matemàtics, con un objectiu *per se*, sense retornar a la seva aplicació als llenguatges naturals. Si bé es desenvoluparen aplicacions en connexió amb la teoria d'accions, el teatre, l'evolució dels llenguatges de programació, la biologia, etc., en canvi, mai no es va fer cap intent per tal d'investigar la rellevància d'aquestes gramàtiques dins del camp d'on sorgiren: la lingüística general, l'estudi dels llenguatges/llengües naturals.

La raó principal és, probablement, el fet que les variants considerades fins aleshores no tenien el suficient poder generatiu per a generar llenguatges com per exemple $\{xx \mid x \in \{a,b\}^*\}$ (ni tan sols el més simple $\{xcx \mid x \in \{a,b\}^*\}$), $\{a^n b^n c^n \mid n \geq 1\}$, etc., amb significació en els llenguatges/llengües naturals. Una altra raó podria ser el “divorci” tradicional entre lingüistes i matemàtics (teòrics del llenguatge formal): sembla ser que cap de les dues parts estava, ni suficientement preparada, ni disposada a treballar per tal d'enllaçar els dos vessants; cadascuna esperava veure què és el que feia l'altra.

Sortosament ambdós contratemps semblen haver-se solucionat. Fruit de la col·laboració entre un grup de lingüistes del GRLMC (“Grup de Recerca en Lingüística Matemàtica i Enginyeria del Llenguatge” de la Universitat Rovira i Virgili”) i un grup de matemàtics del *Department of Mathematics of the University of Turku & The Turku Centre for Computer Science* (de Turku,

Finlàndia) va ser introduïda una nova variant de les gramàtiques contextuals (internes), les denominades *gramàtiques contextuals internes amb ús maximal de selectors* (Martín-Vide et al., 1995a) les quals són capaces d'incloure, d'una manera molt simple, totes les construccions no independents del context mencionades anteriorment. La idea és molt senzilla: quan un context és adjuntat a una cadena que el selecciona, aquesta cadena-selector ha de ser maximal; cap supercadena estricta de la mateixa pot ser usada com a selector.

Així doncs hem aconseguit definir unes eines generatives que no fan ús de la reescriptura i que tenen la mateixa capacitat generativa que els dispositius clàssics pel que fa a la seva adequació per a la descripció del llenguatge/llengua natural i, a més a més, són “més naturals” que la reescriptura, en el sentit de què no són simples artificis creats amb finalitats de descripció científica, sinó en certa manera operacions presents, d'una o altra forma, en la mateixa naturalesa.

El nou marc de treball introduït en aquesta tesi amb la caracterització d'una família d'*autòmats contextuals*, obre les portes a la seva aplicació computacional on, entre d'altres aplicacions, estaria la de l'elaboració d'analitzadors (“parsers”) per al llenguatge/llengua natural.

Bibliografia

Abraham, S. (1965). Some questions of phrase-structure grammars. *Comput. Ling.*, 4:61–70.

Abraham, S. (1972). Compound and serial grammars. *Inform. Control*, 20:432–438.

Adleman, L. M. (1994). Molecular computation of solutions to combinatorial problems. *Science*, 226:1021–1024.

Afendras, E. A. (1969). Mathematical Models for Balkan Phonological Convergence. *A International Conference of Computational Linguistics*, Stockholm, Suècia. Coling.

Aho, A. V. (1968). Indexed grammars – an extension of context-free grammars. *Journal of the ACM*, 15:647 – 671.

Aho, A. V. i Ullman, J. D. (1972). *The theory of parsing, translation and compiling*. Prentice Hall, Inc., Englewood Cliffs., N.J.

- Ajdukiewicz, K. (1935). Die syntaktische konnexität. *Studia Philosophica*, 1:1–27.
- Amar, V. i Putzolu, G. (1964). On a family of linear grammars. *Inform. Control*, 7:283–291.
- Andraşiu, M., Dassow, J., Păun, G., i Salomaa, A. (1993). Language-theoretic problems arising from Richelieu cryptosystems. *Theoretical Computer Sci.*, 116:339–357.
- Angluin, D. (1980). Finding patterns common to a set of strings. *J. Computer System Sci.*, 21:46 – 62.
- Apostolico, A. (1996). *Handbook of Formal Languages*, volum II, capítol String editing and DNA. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Ardelean, G. (1977a). Some topological properties of a family of linear languages. *Bull. Math. Soc. Sci. Math. Roumanie*, 21(69)(3-4):227–231.
- Ardelean, G. (1977b). Topology, finite automata and the domination relation in algebraic linguistics. *Stud. Cerc. Matem.*, 29(6):577–585. En romanés.
- Baker, B. S. i Book, R. V. (1974). Reversal-bounded multipushdown machines. *J. Computer System Sci.*, 8:315 – 332.
- Bălănescu, T. i Gheorghe, M. (1987). Program tracing and languages of action. *Rev. Roum. Ling.*, 32:167–170.

- Bălănescu, T., Gheorghe, M., i Păun, G. (1987). Three variants of apical growth filamentous systems. *Intern. J. Computer Math.*, 22:227 – 238.
- Bar-Hillel, Y. i Shamir, E. (1964). *Language and Information*, capitol Finite State Languages: Formal Representations and Adequacy Problems, pages 87–98. Y. Bar-Hillel, ed., Reading, Mass., Addison-Wesley edition.
- Batog, T. (1967). *The Axiomatic Method in Phonology*. Routledge and Kegan Paul Ltd., London.
- Berstel, J. i Boasson, L. (1996). *Handbook of Formal Languages*, volum I, capitol Context-free languages. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Bodnarčuk, V. G. (1965). The metrical space of events. *Kybernetika*, 1:24–27. En rus.
- Book, R. (1976). On languages with a certain prefix property. *Math. Systems Theor.*, 10(3):229–237.
- Borscev, V. B. (1967). Okrestnostnye gramatiki. *N.T.I., Ser.*, 2(11):39–41.
- Brainerd, B. (1969). On Marcus definition of a phoneme. *Cah. Ling. Th. Appl.*, 6:15–41.
- Brainerd, B. (1971). *Introduction to the Mathematics of Language Study*. American Elsevier, New York.

- Brainerd, B. (1972). On a class of languages occurring in the study of theatre. *Discrete Mathematics*, 2:195–198.
- Brainerd, B. (1974). Etic systems and phonological theory. *Cah. Ling. Th. Appl.*, 11:25–34.
- Bresnan, J. W., Kaplan, R. M., Peters, P. S., i Zaenen, A. (1982). Cross-serial dependencies in Dutch. *Linguistic Inquiry*, 13:613–635.
- Bucurescu, I. i Pascu, A. (1987). Fuzzy contextual grammars. *Rev. Roum. Math. Pures et Appl.*, 32:497–507.
- Calude, C. (1973). Sur une classe de distances dans un semigroupe libre. *Bull. Math. Soc. Sci. Math. Roumanie*, 17(65)(2):123–133.
- Calude, C. (1976a). On metrizability of a free monoid. *Discrete Math.*, 15:307–310.
- Calude, C. (1976b). On the contextual distances. *Stud. Cerc. Matem.*, 28(1):31–35.
- Calude, C. (1976c). The Pompeiu distance between closed sets. *Simpozionul de Geometrie si Analiza Globala*, pages 281–287. Editura Academiei, en romanès.
- Calude, C. i Hromkovic, J. (1996). *Handbook of Formal Languages*, volum II, capítol Complexity: a language-theoretic point of view. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.

- Căraăuşu, A. i Păun, G. (1981). String intersection and short concatenation. *Rev. Roum. Math. Pures et Appl.*, 26(5):713–726.
- Cazimir, B. (1975a). Analyse contextuelle par rapport à deux langages. *Rev. Roum. Math. Pures et Appl.*, 20(2):201–211.
- Cazimir, B. (1975b). Opérateurs contextuels et catégories morphologiques par rapport à deux langages. *Bull. Math. Soc. Sci. Math. Roumanie*, 17(3):247–269.
- Cazimir, B. (1975c). Sur la domination Kunze-Frankova dans la linguistique algébrique. *Rev. Roum. Math. Pures et Appl.*, 20(7):749–775.
- Cazimir, B. (1976). Propositions pour un système automatique de segmentation morphématique. *Cah. Ling. Th. Appl.*, 13(2):595–603.
- Celan, E. i Marcus, S. (1973). Le diagnostic comme langage. *Cah. Ling. Th. Appl.*, 10:163–173.
- Choffrut, C. i Karhumäki, J. (1996). *Handbook of Formal Languages*, volum I, capítol Combinatorics of words. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Chomsky, N. (1956). Three models for the description of languages. *IRE Transaction on Information Theory, IT-2*, 3:113–124.
- Chomsky, N. (1957). *Syntactic Structures*. Mouton, The Hague.

- Chomsky, N. (1959). On certain formal properties of grammars. *Inform. Control*, 2:137–167.
- Ciobotaru, S. (1974). On some generalizations of regular sets and homological relations in algebraic linguistics. *Stud. Cerc. Matem.*, 26(8):1059–1073.
- Ciucar, G. (1988). On the syntactic complexity of Galiukschov semicontextual languages. *Cah. Ling. Th. Appl.*, 25:23–28.
- Clerbout, M. (1985). A solution to Păun problem p100. *Bulletin EATCS*, 26:244–246.
- Coardos, V. (1976). Formal languages and Galois connections. *Cah. Ling. Th. Appl.*, 13(2):665–668.
- Creangă, I., Reisher, C., i Simovici, D. (1974). *Introducere algebrică în informatică. Limbaje formale*. Editura Junimea, Iași. En romanès.
- Cremers, A. i Ginsburg, S. (1975). Context-free grammar forms. *J. of Computer and Syst. Sci.*, 11(1):86–117.
- Crochemore, M. (1996). *Handbook of Formal Languages*, volum II, capítol Pattern matching. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Crăciun, C. V. (1972). Allowed sequences of sets of words and contextual operators. *Cah. Ling. Th. Appl.*, 9(1):93–98.

- Crăciun, C. V. (1973). Sur quelques problèmes d'analyse algébrique contextuelle. *Recueil Linguistique de Bratislava*, IV:231 – 235.
- Csuhaj-Varju, E. i Dassow, J. (1990). On cooperating distributed grammar systems. *J. Inform. Process. Cybern., EIK*, 26:49–63.
- Csuhaj-Varju, E., Dassow, J., Kelemen, J., i Păun, G. (1994). *Grammar Systems. A Grammatical Approach to Distribution and Cooperation*. Gordon and Breach, London.
- Culik, K. (1970). N-ary grammars and the description of mappings of languages. *Kybernetika*, 6(2):99–117.
- Culik, K. i Kari, J. (1996). *Handbook of Formal Languages*, volum III, capítol Language-theoretic issues in computer graphics. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Culy, C. (1985). The Complexity of the Vocabulary of Bambara. *Linguistics and Philosophy*, 8:345–351.
- Curry, H. (1961). Some logical aspects of grammatical structure. A *Proc. of Symp. in Appl. Math.*, volum 12 of *Structure of language and its mathematical aspects*, pages 58–68. Amer. Math. Soc.
- Dascălu, I. (1975). Some extensions of Trybulec and Kunze dominations in algebraic linguistics. *Stud. Cerc. Matem.*, 27(3):281–301.
- Dassow, J. (1979). On the circular closure of languages. *Elek. Inf. und Kybern.*, 15(1/2):87–94. En romanès.

- Dassow, J. (1982). Remark on a problem of Nguyen Xuan My and Păun. *Elektr. Inform. und Kybern.*, 18:66–67.
- Dassow, J. (1985). On contextual grammar forms. *Intern. J. Computer Math.*, 17:37–52.
- Dassow, J., Marcus, S., i Păun, G. (1993). Iterative reading of numbers and “black-holes”. *Periodica Math. Hungarica*, 27:137 – 152.
- Dassow, J. i Păun, G. (1989). *Regulated Rewriting in Formal Language Theory*. Springer-Verlag, Berlin, Heidelberg.
- Dassow, J., Păun, G., i Rozenberg, G. (1996a). *Handbook of Formal Languages*, volum II, capítol Regulated rewriting and grammar systems. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Dassow, J., Păun, G., i Salomaa, A. (1996b). *Handbook of Formal Languages*, volum I, capítol Grammars with controlled derivations. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- de Luca, A. i Varricchio, G. (1996). *Handbook of Formal Languages*, volum I, capítol Unavoidable regularities. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Diekert, V. i Metivier, Y. (1996). *Handbook of Formal Languages*, volum III, capítol Partially communicative monoids and traces. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.

- Dincă, A. (1976). The metric properties of the semigroups and the languages. *Lecture Notes in Computer Science*, 45:260–264. Springer, Berlin.
- Dinu, M. (1984). *Semiotics of Drama and Theatre*, volum 10 of *Linguistic and Literary Studies in Eastern Europe*, capitol The algebra of scenic situations, pages 67–92. H. Schmid & A. van Kesteren (ed.), John Benjamins Publishing Co., Amsterdam/Philadelphia.
- Dobrušin, R. L. (1957). The elementary grammatical category. *Byulleten Objedeneniya po Problemam Mashinogo Perevoda*, 5:19–21. En rus.
- Dobrušin, R. L. (1961). Mathematical methods in linguistics. applications. *Matematicheskie Prosveshchenie*, 6:52 – 59. En rus.
- Du Feu, V. i Marcus, S. (1973). English grammatical homonymy and grammatical categories from the view point of algebraic linguistics. *Rev. Roum. Ling.*, 18(3):215–241.
- Dumitrescu, C. i Vianu, V. (1979). The topological learning space and the Bodnarčuk metric space of languages. *Found. Control Engineering*, 4(1):3–8.
- Ehrenfeucht, A., Ilie, L., Păun, G., Rozenberg, G., i Salomaa, A. (1995a). *Mathematical Linguistics and Related Topics*, capitol On the generative capacity of certain classes of contextual grammars. Gh. Păun, ed., Ed. Academiei, Bucharest.

- Ehrenfeucht, A., Mateescu, A., Păun, G., Rozenberg, G., i Salomaa, A. (1995b). On representing RE languages by one-sided internal contextual languages. Technical Report 95 / 04, Dept. of Computer Sci., Leiden Univ.
- Ehrenfeucht, A., Păun, G., i Rozenberg, G. (1994). *Mathematical Aspects of Natural and Formal Languages*, capítol Normal forms for contextual grammars, pages 79–96. Gh. Păun, ed., World Sci. Publ., Singapore.
- Ehrenfeucht, A., Păun, G., i Rozenberg, G. (1995c). The linear landscape of external contextual languages. *Acta Informatica*.
- Ehrenfeucht, A., Păun, G., i Rozenberg, G. (1996a). *Handbook of Formal Languages*, volum 3, capítol Contextual Grammars. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Ehrenfeucht, A., Păun, G., i Rozenberg, G. (1996b). On representing recursively enumerable languages by internal contextual languages. *Th. Computer Sci.*
- Eilenberg, S. (1974). *Automata, languages and machines*, volum A. Academic Press, New York.
- Engelfriet, J. i Rozenberg, G. (1996). *Handbook of Formal Languages*, volum III, capítol Graph grammars. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.

- Fertig, A. (1974). On Franclova-Trybulec domination in algebraic linguistics. *Stud. Cerc. Matem.*, 26(9):1209–1223. En romanès.
- Floyd, R. W. (1962). On the non-existence of a phrase-structure grammar for Algol-60. *Comm. of the A.C.M.*, 5(9):483–484.
- Franclova, M. (1969). The problem of replacement of words. *Prague Bulletin of Math. Ling.*, 11:3–10.
- Freund, R., Păun, G., i Rozenberg, G. (1994). Contextual array grammars. Manuscrit.
- Friedman, J., Manaster-Ramer, A., i Rounds, W. C. (1987). *Mathematics of language*, capítol Finding natural languages a home in formal language theory, pages 349 – 360. A. Manaster Ramer, ed., John Benjamins, Amsterdam, Philadelphia.
- Galiukschov, B. S. (1981). Semicontextual grammars. *Mat. logica i mat. ling.*, pages 38–50. Kalinin Univ. (in Russian).
- Gazdar, G., Klein, E., Pullum, G. K., i Sag, I. (1985). *Generalized Phrase Structure Grammar*. Basil Blackwell, Oxford.
- Gazdar, G. i Pullum, G. K. (1982). Natural languages and context-free languages. *Linguistics and Philosophy*, 4:471–504.
- Gazdar, G. i Pullum, G. K. (1985). Computationally Relevant Properties of Natural Languages and their Grammars. *New Generation Computing*, 3:273–306.

- Gécseg, F. i Steinby, M. (1996). *Handbook of Formal Languages*, volum III, capítol Tree languages. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Georgescu, G. (1994). *Mathematical Aspects of Natural and Formal Languages*, capítol Infinite hierarchies of some types of contextual languages, pages 137–150. Gh. Păun, ed., World Sci. Publ., Singapore.
- Georgescu, G. (1996). The syntactic complexity of internal contextual grammars and languages. II Intern. Congr. on Math. Ling. (ICML'96), Taragona. Pendent de publicació.
- Ginsburg, S. (1966). *The mathematical theory of context-free languages*. McGraw-Hill Book Comp., New York.
- Ginsburg, S. (1975). *Algebraic and automata-theoretic properties of formal language*. North-Holl. Publ. Comp., Amsterdam, Oxford.
- Ginsburg, S. i Greibach, S. A. (1979). Abstract families of languages. A of the Amer. Math. Soc., M., editor, *Studies in Abstract families of languages*, volum 87.
- Gladkij, A. V. (1963). Configuration characteristics of languages. *Problemy Kibernetiki*, 10:251–260. En rus.
- Gladkij, A. V. (1973). *Mathematical Models of Language*, capítol An attempt at the formal definition of case and gender of the noun, pages 159–204. F. Kiefer (ed.), Athenäum Verlag.

- Greibach, S. A. (1963). The undecidability of the ambiguity problem for minimal linear grammars. *Inform. Control*, 6:119–125.
- Gross, M. (1964). Inherent ambiguity of minimal linear grammars. *Inform. Control*, 7:366–368.
- Gross, M. i Lentin, A. (1970). *Notions sur les grammaires formelles*. Gauthier-Villars, Paris.
- Gruska, J. (1969). Some classifications of context-free languages. *Inform. Control*, 14:152–173.
- Gruska, J. (1971). Complexity and unambiguity of context-free grammars and languages. *Inform. Control*, 18:502–519.
- Gruska, J. (1973). Descriptive complexity of context-free languages. A *Proc. of Symp. Math. Found. of Computer Sci., High Tatras*, pages 71–83.
- Haines, L. H. (1964). Note on the complement of a (minimal) linear language. *Inform. Control*, 7:307–314.
- Harju, T. i Karhumäki, J. (1996). *Handbook of Formal Languages*, volum I, capítol Morphisms. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Harris, Z. S. (1951). *Methods in Structural Linguistics*. Chicago Univ. Press.
- Harrison, M. (1978). *Introduction to Formal Language Theory*. Addison-Wesley, Reading, Mass.

- Hauschild, D. i Jantzen, M. (1994). Petri nets algorithms in the theory of matrix grammars. *Acta Informatica*, 31:719 – 728.
- Havel, I. M. (1976). On the branching structure of languages. A Springer-Verlag, editor, *Proc. of Symp. Math. Found. of Computer Sci., Gdansk*, volum 45 of *Lecture notes in Computer Sciences*, Berlin.
- Havel, I. M. (1980). On branching and looping. Part I. *Th. Computer Sci.*, 10:187–220.
- Head, T. (1987). Formal language theory and DNA:an analysis of the generative capacity of specific recombinant bahaviors. *Bull. Math. Biology*, 49:737–759.
- Head, T. (1992). *Lindenmayer Systems: Impacts on Theoretical Computer Science and Developmental Biology*, capítol Splicing schemes and DNA, pages 371 – 383. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin.
- Head, T., Păun, G., i Pixton, D. (1996). *Handbook of Formal Languages*, volum II, capítol Language theory and molecular genetics. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Herman, G. i Rozenberg, G. (1975). *Developmental systems and languages*. North-Holl. Publ. Comp., Amsterdam.
- Higginbotham, J. (1984). English is not a contextfree language. *Linguistic Inquiry*, 15:119–126.

- Higman, G. H. (1952). Ordering by divisibility in abstract algebras. A *Proc. London Math. Soc.*, volum 3, pages 326–336.
- Holban, M. M. (1972). The contextual operators and the empty set in algebraic linguistics. *Stud. Cerc. Matem.*, 24(5):777 – 790. En romanès.
- Hopcroft, J. E. i Ullman, J. D. (1969). *Formal Languages and their Relation to Automata*. Addison-Wesley, Reading, Mass.
- Hopcroft, J. E. i Ullman, J. D. (1979). *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, Mass.
- Horecký, J. (1964). *Morfematická struktúra slovenčiny*. Bratislava.
- Horecký, J. (1966). A model of the grammatical gender for Western Slavic Languages. *Jazykovedný Casopis*, 17(1):3–12. En eslovac.
- Huybregts, M. A. C. (1976). Overlapping Dependencies in Dutch. *Utrecht Working Papers in Linguistics*, 1:24–65.
- Ibarra, O. (1970). Simple matrix languages. *Inform. Control*, 17:359–394.
- Ilie, L. (1994). *Mathematical Aspects of Natural and Formal Languages*, capítol On contextual grammars with parallel derivation, pages 165–172. Gh. Păun, ed., World Sci. Publ., Singapore.
- Ilie, L. (1996a). An attempt to define a class of mildly context-sensitive languages. Manuscrit.

- Ilie, L. (1996b). A non-semilinear language generated by an internal contextual grammar with finite selection. *Ann. Univ. Buc. Matem.-Inform. Series.* Pendent de publicació.
- Ilie, L. (1996c). On ambiguity in internal contextual languages. II Intern. Congr. on Math. Ling. (ICML'96), Tarragona. Pendent de publicació.
- Iosifescu, I. (1975). Semantic contextual distances. *Stud. Cerc. Matem.*, 27:545–557. En romanès.
- Istrail, S. (1978a). Contextual grammars with regulated selection. *Stud. Cerc. Matem.*, 30:287–294. En romanès.
- Istrail, S. (1978b). A problem about contextual grammars with choice. *Stud. Cerc. Matem.*, 30:135–139. En romanès.
- Istrail, S. (1980). A fixed-point approach to contextual languages. *Rev. Roum. Math. Pures et Appl.*, 25:861–869.
- Istrail, S. (1989). On a problem about marcus contextual languages. *Bull. Math. Soc. Sci. Math. Roumanie*, 33(81):335–338.
- Ito, M. (1971). Quelques remarques sur l'espace contextuel. *Mathematical Linguistics*, 57:18–28.
- Ito, M. (1972a). Sur la restriction maximale d'un langage. *Proceed. Japan Acad.*, 48:98–100.

- Ito, M. (1972b). Sur l'extension et la restriction d'un langage associé à l'espace contextuel. *Proceed. Japan Acad*, 48:94–97.
- Ito, M. (1974). Generalized contextual languages. *Cah. Ling. Th. Appl.*, 11:73–77.
- Jančar, P. (1993). Nondeterministic Forgetting Automata are Less Powerful than Deterministic Linear Bounded Automata. *Acta Math. et Inf. Univ. Ostraviensis*, 1. Ostrava.
- Jančar, P., Mráz, F., i Plátek, M. (1993). A taxonomy of forgetting automata. *A MFCS 1993*, volum 711, pages 527–536, Gdańsk, Poland. Springer.
- Jančar, P., Mráz, F., Plátek, M., Procházka, M., i Vogel, J. (1995a). On relations between restarting automata, context-free grammars and Marcus grammars. *A Developments in Language Theory Symp.*, Magdeburg.
- Jančar, P., Mráz, F., Plátek, M., Procházka, M., i Vogel, J. (1995b). Restarting Automata and Marcus Grammars. Technical Report Math/95/1, Friedrich Schiller University, Jena.
- Jančar, P., Mráz, F., Plátek, M., Procházka, M., i Vogel, J. (1996). Deleting Automata with Restart Operation and Marcus Grammars. Manuscrit.
- Jančar, P., Mráz, F., Plátek, M., i Vogel, J. (1994). Restarting Automata. Technical Report Math/94/8, Friedrich Schiller University, Jena.

- Jantzen, M. (1996). *Handbook of Formal Languages*, volum III, capítol Basic of term rewriting. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Joshi, A. (1996). *Tree-adjunct grammars*, volum III, capítol Regulated rewriting and grammar systems. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Joshi, A. i Savitch, W. (1996). *Handbook of Formal Languages*, volum II, capítol Grammatical models for natural languages. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Joshi, A. K. (1985). *Natural Language Processing: Psycholinguistic, Computational and Theoretic Perspectives*, capítol Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions ?, pages 206 – 250. Dowty, Karttunen, and Zwicky, eds., Cambridge Univ. Press, New York.
- Joshi, A. K., Kosaraju, S. R., i Yamada, H. M. (1972a). String adjunct grammars: I. local and distributed adjunction. *Inform. Control*, 21:93–116.
- Joshi, A. K., Kosaraju, S. R., i Yamada, H. M. (1972b). String adjunct grammars: II. equational representation, null symbols and linguistic relevance. *Inform. Control*, 21:235–260.
- Joshi, A. K., Levy, L., i Takahashi, M. (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences*, 10(1):136 – 163.

- Jurcă, M. (1970). On the relation between the type and the Q-class of a word. *Stud. Cerc. Matem.*, 22(7):1033–1039. En romanès.
- Jürgensen, H. (1996). *Handbook of Formal Languages*, volum I, capítol Codes. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Kanger, S. (1964). The notion of a phoneme. *Statistical Methods in Linguistics*, 3:43–48.
- Kari, L. (1991). *On Insertion and Deletion in Formal Languages*. PhD thesis, University of Turku, Finland.
- Kari, L., Păun, G., Thierrin, G., i Yu, S. (1996). At the crossroads of linguistics, DNA computing, formal language theory: characterizations of RE languages by means of insertion-deletion systems. Manuscrit.
- Karpinskaja, O. G. (1966). A method of typological description of the Slavic gender system. *Lingvisticeskie issledovanija po obščei i alavjanskoj tipologii*, pages 75–116.
- Kasai, T. (1970). On analytic languages. *Bull. Math. Soc. Sci. Math. Roumanie*, 14(62)(1):41–50.
- Keenan, E. i Möinich, U. (1993). Introduction to Mathematical Analysis of Natural Language. Technical report, Fifth European Summer School in Logic, Language and Information, Faculdade de Letras. Universidade de Lisboa.

- Kiefer, F. (1968). *Mathematical Linguistics in Eastern Europe*. American Elsevier, New York.
- Klein, W. (1971). *Parsing, Studien zur maschinelle Satzanalyse mit Abhängigkeitsgrammatiken und Transformationsgrammatiken*. Athenäum Verlag, Frankfurt-Main.
- Knuth, D. E. (1967). A characterization of paranthesis languages. *Inform. Control*, 11:269–289.
- Kortlandt, F. H. H. (1972). *Modeling the Phoneme, New Trends in East European Phonetic Theory*. Mouton, The Hague.
- Kortlandt, F. H. H. (1973). The identification of phonemic units. *La Linguistique*, 9(2):119–130.
- Kral, J. (1973). A note on grammars with regular restriction. *Kybernetika*, 9:159–161.
- Krithivasan, K. i Nirmal, N. (1983). Filamentous systems with apical growth. *Intern. J. Computer Math.*, 12:203 – 215.
- Kuich, W. (1996). *Handbook of Formal Languages*, volum I, capítol Formal power series. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Kulagina, O. (1958). On a way to define the grammatical concepts by means of set theory. *Problemy kibernetiki*, 1:203–214. En rus.

- Kunze, J. (1967). Versuch eines objektivierten Grammatikmodells II. *Zeitschrift für Phonetik, Sprach. und Komm.*, 20(5-6):415–448.
- Kunze, J. i Priess, W. (1970). Versuche eines objektivierten Grammatikmodells III. *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung*, 23(4):347–378.
- Kuroda, S. Y. (1964). Classes of languages and linear-bounded automata. *Inform. Control*, 7:207–223.
- Lambek, J. (1961). On the calculus of syntactic types. *A Structure of Language and its Mathematical Aspects, Proc. 12th Symp. Appl. Math.*, pages 166–178. Providence, R.I.
- Langendoen, D. T. (1977). *Studies in Descriptive and Historical Linguistics*, capítulo On the inadequacy of type-2 and type-3 grammars for human languages, pages 159–172. (P. J. Hopper, ed.), John Benjamins, Amsterdam.
- Latteaux, M. (1978). Une note sur la propriété de préfixe. *Math. Systems Theory*, 11:235–238.
- Latteux, M., Leguy, B., i Ratoandromanana, B. (1985). The family of one-counter languages is closed under quotient. *Acta Informatica*, 22:579 – 588.
- Lecomcev, J. K. (1966). Nekotorye voprosy metaazyk dlja opisaniya distribucii jazycovykh elementov. *Strukturnaja tipologija jazykov*, pages 225–250. Nauka, Moscú.

- Lekomceva, M. I. (1966). K opisaniju fonologiceskoi sistemy staroslawjanskogo jazyka na osnove ternarnogo principa. *Lingvisticskie isledovanija po obščei i slavjanskoj tipologii*, pages 117–123. Nauka, Moscú.
- Lindenmayer, A. (1968). Mathematical models for cellular interaction in development, i and ii. *J. Theoret. Biol.*, 18:280–315.
- Lindenmayer, A. (1971). Developmental systems without cellular interactions, their languages and grammars. *J. Theoret. Biol.*, 30:455–484.
- Lindenmayer, A. (1975). Developmental algorithms for multicellular organisms: a survey of L systems. *J. Theoret. Biol.*, 54:3–22.
- Lindenmayer, A. i Rozenberg, G., editors (1976). *Automata, languages, development. At the crossroads of biology, mathematics and computer science*. North-Holl. Publ. Comp., Amsterdam.
- Lipsky, J. M. (1974). A topology of semantic dependence. *Semiotica*, 12(2):145–170.
- Manaster-Ramer, A. (1983). The Soft Formal Underbelly of Theoretical Syntax. A *Papers from the Nineteenth Regional Meeting*, pages 254–262. Chicago Linguistic Society.
- Manaster-Ramer, A. (1986). Copying in natural languages, context-freeness, and queue grammars. A *Proc. 24th Annual Meeting of the Ass. for Comp. Linguistics*, pages 85–89.

- Manaster-Ramer, A. (1994). Uses and Misuses of Mathematics in Linguistics. A *Proc. X Congreso de Lenguajes Naturales y Lenguajes Formales*. C. Martín-Vide ed. Article no inclòs en el volum.
- Marcus, M. i Păun, G. (1990). Regulated galiukschov semicontextual grammars. *Kybernetika*, 26:316–326.
- Marcus, S. (1961). Structures linguistiques et structures topologiques. *Rev. Roum. Math. Pures et Appl.*, 6(3):501–506.
- Marcus, S. (1962). On a logical model of the part of speech. *Stud. Cerc. Matem.*, 13(1):137 – 162. En romanès.
- Marcus, S. (1963a). *Mathematical Linguistics*. Ed. Didactică și Pedagogică, București. En romanés.
- Marcus, S. (1963b). Modèles mathématiques pour la catégorie grammaticale du cas. *Rev. Roum. Math. Pures et Appl.*, 8(4):585–610.
- Marcus, S. (1963c). Typologie des langues et modèles logiques. *Acta Math. Acad. Sci. Hungaricae*, 14(3-4):269–281.
- Marcus, S. (1964). *Gramatici și automate finite*. Editura Academiei R.S.R., Bucharest. En romanès.
- Marcus, S. (1967a). *Algebraic Linguistics. Analytical Models*. Academic Press, New York.

- Marcus, S. (1967b). Catégories de Dobrušin, fermetures de Sestier et voisinage de Sakai. *Glossa*, 1(1):59–67.
- Marcus, S. (1967c). Catégories morphologiques et analyse contextuelle dans la linguistique algebrique. A *IIème Conférence Internationale sur le Traitement Automatique des Languages*, pages 1 – 8, Grenoble. Comunicació núm. 39.
- Marcus, S. (1967d). *Introduction mathématique à la linguistique structurale*. Dunod, Paris.
- Marcus, S. (1967e). Modèles mathématiques dans l'étude du drame. *T. A. Informations*, 2:86–87.
- Marcus, S. (1967f). Poetica matemàtica. Curs impartit a la Facultat de Llengua i Literatura Romanesa de la Universitat de Bucarest. Anys 1996-97. En romanès.
- Marcus, S. (1968). Opérateurs contextuels et catégories morphologiques. *Bull. Math. Soc. Sci. Math. Roumanie*, 10(3):65–72.
- Marcus, S. (1969a). Contextual grammars. *Rev. Roum. Math. Pures et Appl.*, 14:1525–1534.
- Marcus, S. (1969b). Deux types nouveaux de grammaires génératives. *Cah. Ling. Th. Appl.*, 6:69–74.
- Marcus, S. (1969c). Sur la domination au sens de Kunze dans la linguistique algebrique. *Rev. Roum. Math. Pures et Appl.*, 14(3):378–396.

- Marcus, S. (1970a). Les modèles mathématiques et l'opposition romane-slave dans la typologie du genre grammatical. A *Actes du XII^{ème} Congrès International de Linguistique et de Philologie Romance*, volum 1, pages 247–252, Bucharest. Editura Academiei.
- Marcus, S. (1970b). *Poetica matematică*. Editura Academiei RSR. En romanès.
- Marcus, S. (1973a). Contextual distances in algebraic linguistics. *Stud. Cerc. Matem.*, 25(2):223–265. En romanés.
- Marcus, S. (1973b). Distances and diameters in a semigroup. *Stud. Cerc. Matem.*, 25(3):359–378. En romanés.
- Marcus, S. (1974). Linguistics as a pilot science. A *Current trends in Linguistics*, volum 12. Ed. Th. Sebeok, Mouton, The Hague, Paris.
- Marcus, S. (1976). The metric properties of the semigroups and the languages. *Lecture Notes in Computer Science*, 45:260–264.
- Marcus, S., editor (1978). *La sémiotique formelle du folklore*. Klincksieck and The Publ. House of the Romanian Academy, Paris and Bucharest.
- Marcus, S. (1980). Textual cohesion and textual coherence. *Rev. Roum. Ling.*, 25(2):101–112.
- Marcus, S., editor (1981). *Contextual Ambiguities in Natural and Artificial Languages*, volum 1 of *Communication and Cognition Monographs*. Ghent, Belgium.

- Marcus, S., editor (1983). *Contextual Ambiguities in Natural and Artificial Languages*, volum 2 of *Communication and Cognition Monographs*. Ghent, Belgium.
- Marcus, S. (1989a). Interplay of innate and acquired in mathematical models of language learning processes. *Rev. Roum. Ling.*, 34(2):101–116.
- Marcus, S. (1989b). N. s. troubetzkoy and his followers in the study of linguistic appositions. A Vide, C. M., editor, *Actes del IVé Congrès de Llenguatges Naturals i Llenguatges Formasl*, volum 2, pages 943–961, Barcelona. PPU.
- Marcus, S. (1994). *Current issues in mathematical linguistics*, capítol The Status of Research in the Field of Analytical Algebraic Models of Language, pages 3–21. Linguistic series. C. Martín Vide (ed). North-Holland.
- Marcus, S. (1996). Natural Languages and Artificial Languages. A Semiotic, Mathematical, and Computational Perspective. Notes del Seminari impartit pel Prof. Solomon Marcus a la Universitat Rovira i Virgili de Tarragona. Tarragona, gener-abril de 1995. Pendent de publicació.
- Marcus, S., Martín-Vide, C., i Păun, G. (1995a). Contextual grammars and natural languages. Manuscrit presentat a la Intern. Conf. Math. of Natural Languages, Philadelphia, Penn.
- Marcus, S., Martín-Vide, C., i Păun, G. (1995b). On the power of internal contextual grammars with maximal use of selectors. A *Fourth meeting on mathematics of language MOL4*, Philadelphia.

- Marcus, S., Martín-Vide, C., i Păun, G. (1996a). Contextual Grammars versus Natural Languages. A *Proceedings of the International Workshop "Speech and Computer" (SPECOM'96)*, pages 28 – 33, St. Petersburg, Rússia.
- Marcus, S., Martín-Vide, C., i Păun, G. (1996b). On Internal Contextual Grammars with Maximal use of Selectors. A *8th. International Conference on Automata and Formal Languages (AFL'96)*, pages 70 – 71, Salgótarján, Hongria.
- Marcus, S., Nicolau, E., i Stati, S. (1966). *Introducere în lingvistica matematică*. Ed. Științifică, București. Traducció castellana *Introducción en la lingüística matemática*, Ed. Teide, Barcelona, 1978.
- Marcus, S. i Vasiliu, E. (1960a). Mathématique et phonologie. Théorie des graphes et consonantisme de la langue roumaine. *Rev. Roum. Math. Pures et Appl.*, 5(2):319–340.
- Marcus, S. i Vasiliu, E. (1960b). Mathématique et phonologie. Théorie des graphes et consonantisme de la langue roumaine. *Rev. Roum. Math. Pures et Appl.*, 5(3-4):681–703.
- Martín-Vide, C. (1990). ¿por qué las lenguas naturales son conjuntos infinitos no-numerables de orden 2? A Martín-Vide, C., editor, *Actes del VI Congrés de Llenguatges Naturals i Llenguatges Formals*, pages 17 – 22, Barcelona. PPU.

- Martín-Vide, C. (1993). Towards a Categorical Theory of Linguistic Trees. A
Martín-Vide, C., editor, *Actes del IX Congrés de Llenguatges Naturals i
Llenguatges Formals*, pages 133 – 140, Barcelona. PPU.
- Martín-Vide, C. (1994). Gramáticas formales (y similares) para lingüistas. A
Martín-Vide, C., editor, *Actes del X Congrés de Llenguatges Naturals i
Llenguatges Formals*, pages 71 – 91, Barcelona. PPU.
- Martín-Vide, C. (1996). Computación natural. A Martín-Vide, C., editor,
Actes del XII Congrés de Llenguatges Naturals i Llenguatges Formals,
pages 121 – 127, Barcelona. PPU.
- Martín-Vide, C., Mateescu, A., Miquel-Vergés, J., i Păun, G. (1995a). Internal
Contextual Grammars: Minimal, Maximal, and Scattered Use of Selector.
*A Proc. of The Fourth Bar-Ilan Symp. on Found. of Artificial Intelligence
BISFAI'95*, pages 132–142, Jerusalem. M. Kappel, E. Shamir, eds.
- Martín-Vide, C., Miquel-Vergés, J., i Păun, G. (1995b). Contextual grammars
with depth-first derivation. *A Proc. of the TWLT 10, Algebraic Methods
in language Processing AMILP'95, Twente, The Netherlands*.
- Martín-Vide, C., Miquel-Vergés, J., i Păun, G. (1996a). Contextual Automata.
Manuscrit.
- Martín-Vide, C., Miquel-Vergés, J., i Păun, G. (1996b). Two-level Contextual
Grammars: The Internal Case. Manuscrit.

- Martín-Vide, C., Miquel-Vergés, J., Păun, G., i Salomaa, A. (1996c). Attempting to define the ambiguity of internal contextual languages. II Intern. Congr. on Math. Ling. (ICML'96), Tarragona. Pendent de publicació.
- Martín-Vide, C., Păun, G., i Salomaa, A. (1996d). Characterizations of recursively enumerable languages by means of insertion grammars. Manuscrit.
- Mateescu, A. (1994). *Mathematical Aspects of Natural and Formal Languages*, capítol Marcus contextual grammars with shuffled contexts, pages 275–284. Gh. Păun, ed., World Sci. Publ., Singapore.
- Mateescu, A. i Salomaa, A. (1996a). *Handbook of Formal Languages*, volum I - Word, language, grammar, capítol Formal languages: an introduction and a synopsis. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Mateescu, A. i Salomaa, A. (1996b). *Handbook of Formal Languages*, volum I, capítol Language hierarchies. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Maurer, H. A., Salomaa, A., i Wood, D. (1980). Pure grammars. *Inform. Control*, 44:47–72.
- Mayoh, B. (1967). Grammatical categories. *Rev. Roum. Math. Pures et Appl.*, 12(6).
- Mazzaroli, I. G. (1969). *A Set-theoretical Approach for Codifications Endowed with a System of Relevant Oppositios*. Università di Bari, Bari, Itàlia.

Meersman, R. i Rozenberg, G. (1978). Cooperating grammar systems. A *Proc. MFCS'78 Symp.*, volum 64 of *LNCS*, pages 364–374. Springer-Verlag.

Miller, G. (1956). The magical number 7 ± 2 ; Some limits on our capacity for processing information. *Psychological Review*, 63:81 – 97.

Miquel-Vergés, J. (1994a). Gramàtiques contextuales. A Martín-Vide, C., editor, *Actes del X Congrés de Llenguatges Naturals i Llenguatges Formals*, pages 93 – 110, Barcelona. PPU.

Miquel-Vergés, J. (1994b). Introducció a les gramàtiques contextuales, report 3/94. Technical report, GRLMC, Universitat Rovira i Virgili, Tarragona. Capítol introductori a “Topics in contextual grammars: from natural languages to formal languages and back”.

Miquel-Vergés, J. (1995a). Ambigüedades contextuales en los lenguajes naturales i artificials. A Martín-Vide, C., editor, *Actes del XI Congrés de Llenguatges Naturals i Llenguatges Formals*, pages 71 – 91, Barcelona. PPU.

Miquel-Vergés, J. (1995b). Models algebraics analítics del llenguatge. Les Gramàtiques Contextuales. A Ortega i Robert, R., editor, *Actes del Ier. Taller de Ciències del Llenguatge (TCL'95), Report 4/95*, pages 51 – 60, Tarragona. GRLMC, Universitat Rovira i Virgili.

- Miquel-Vergés, J. (1996a). L'ambigüitat lingüística. II Taller de Ciències del Llenguatge (TCL'96), Universitat Rovira i Virgili, 29-31 de maig, Tarragona.
- Miquel-Vergés, J. (1996b). On contextual automata. II Intern. Congr. on Math. Ling. (ICML'96), Tarragona. Pendent de publicació.
- Mitrana, V. (1994). *Mathematical Aspects of Natural and Formal Languages*, capítol Contextual grammars: the strategy of minimal competence, pages 319–331. Gh. Păun, ed., World Sci. Publ., Singapore.
- Mitrana, V. (1995). *Mathematical Linguistics and Related Topics*, capítol Contextual insertion and deletion, pages 271–278. Gh. Păun, ed., The Publ. House of the Romanian Academy, Bucharest.
- Morse, M. i Hedlung, G. (1944). Unending chess, symbolic dynamics and a problem in semigroups. *Duke Math. J.*, 11:1–7.
- Naftanail, A. (1976). An algebraic model of the phonetic segmentation and of the accent. *Stud. Cerc. Matem.*, 28(3):293–302. En romanès.
- Naur, P. (1960). Report on the algorithmic language Algol 60. *Comm. of the A.C.M.*, 3:299–314.
- Navratil, E. (1970). Context-free grammars with regular conditions. *Kybernetika*, 2:118–126.
- Nebeský, L. (1965). Conditional replacement of words. *Prague Bulletin of Math. Ling.*, 3:3–12.

- Nebeský, L. (1966a). K jednoí matematiceskoí modeli fonemy. *Rev. Roum. Math. Pures et Appl.*, 11:4.
- Nebeský, L. (1966b). On the notion of relevant features. *The Prague Bulletin of Math. Ling.*, 6:35–43.
- Nebeský, L. (1966c). The ξ -grammar. *Prague Studies in Math. Ling.*, 2:147–154.
- Nebeský, L. (1968). Language systems. *Rev. Roum. Math. Pures et Appl.*, 13(3):371 – 375.
- Nguyen, X. M. (1980a). 3-factorization contextual grammars. Manuscrit.
- Nguyen, X. M. (1980b). Gramatici de adjuncție dependentă de context. *Stud. Cerc. Matem.*, 32:549–557. En romanès.
- Nguyen, X. M. (1980c). N-contextual grammars with syntactical invariant choice. *Found. Control Engineering*, 5:21–29.
- Nguyen, X. M. (1980d). On the n -contextual grammars. *Rev. Roum. Math. Pures et Appl.*, 25:1395–1406.
- Nguyen, X. M. (1981a). A note on n -contextual grammars. *Analele Univ. din București*, 30:49–54.
- Nguyen, X. M. (1981b). *On Some Generalizations of Contextual Grammars*. PhD thesis, Univ. of Bucharest.

- Nguyen, X. M. (1981c). Some classes of n -contextual grammars. *Rev. Roum. Math. Pures et Appl.*, 26:1221–1233.
- Nguyen, X. M. (1982). Closure properties of some classes of contextual languages. *Rev. Roum. Math. Pures et Appl.*, 27:61–72.
- Nguyen, X. M. i Păun, G. (1980). On the inner contextual grammars. *Rev. Roum. Math. Pures et Appl.*, 25:641–651.
- Nguyen, X. M. i Păun, G. (1982a). On the generative capacity of contextual schemata with syntactical invariant choice. *Elektr. Inform. und Kybern.*, 18:57–65.
- Nguyen, X. M. i Păun, G. (1982b). On the generative capacity of n -contextual grammars. *Bull. Math. Soc. Sci. Math. Roumanie*, 25(74):345–354.
- Niemi, V. (1996). *Handbook of Formal Languages*, volum II, capítol Cryptography: language-theoretic aspects. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Nijholt, A. (1996). *Handbook of Formal Languages*, volum II, capítol Parsing. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Nivat, M. (1996). *Handbook of Formal Languages*, volum III, capítol Polinomio tilings and word equations. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.

- Novotný, M. (1971). Complete characterizations of classes of chomsky by means of configurations. *Acta Facultatis Rerum Naturalium Univ. Comenianae. Mathematica*, pages 63 – 71. Número especial.
- Novotný, M. (1972). On some relations defined by languages. *Prague Studies in Math. Linguistics*, 4:157 – 170.
- Novotný, M. (1974). On a class of contextual grammars. *Cah. Ling. Th. Appl.*, 11:313–315.
- Novotný, M. (1976a). Each generalized contextual language is context-sensitive. *Rev. Roum. Math. Pures et Appl.*, 21:353–362.
- Novotný, M. (1976b). On some variants of contextual grammars. *Rev. Roum. Math. Pures et Appl.*, 21:1053–1062.
- Novotný, M. (1982). Contextual grammars vs. context-free algebras. *Czech. Math. journal*, 32:529–547.
- Novotný, M. (1985). On some constructions of grammars for linear languages. *Intern. J. Computer Math.*, 17:65–77.
- Nowakowska, M. (1973). *Language of Actions: Language of Motivations*. Mouton. The Hague.
- Ore, O. (1944). Galois connections. *Trans. Amer. Math. Soc.*, 55:493 – 513.
- Orman, G. (1970). Quelques résultats concernant les ensembles homologues. *Recueil Linguistique de Bratislava*, 4:245–249.

- Ortega-Robert, R. (1995). Gramáticas de adjunción de árboles: Tag. A Martín-Vide, C., editor, *Actes del XI Congrés de Llenguatges Naturals i Llenguatges Formals*, pages 93 – 109, Barcelona. PPU.
- Pană, D. (1975). Closure operators and the relation of dmination in algebraic linguistics. *Stud. Cerc. Matem.*, 27(4):461–47. En romanès.
- Papadimitriou, C. H. (1994). *Computational Complexity*. Addison-Wesley, Reading, Mass.
- Partee, B. H., ter Meulen, A., i Wall, R. E. (1990). *Mathematical Methods in Linguistics*. Kluwer, Dordrecht.
- Pascu, A. i Păun, G. (1976). On simple matrix languages. *Bull. Math. de la Soc. de Sci. Math. de R.S.R.*, 20(68):333–340.
- Păun, G. (1974a). Cîteva gramatici dependente de context. *Stud. Cerc. Matem.*, 26:1105–1110. En romanès.
- Păun, G. (1974b). On contextual grammars. *Stud. Cerc. Matem.*, 26:1111–1129. En romanès.
- Păun, G. (1975a). Closure operators and the relation of domination in algebraic linguistics. *Stud. Cerc. Matem.*, 27(4):461 – 470.
- Păun, G. (1975b). On the complexity of contextual grammars with choice. *Stud. Cerc. Matem.*, 27:559–569. En romanès.

- Păun, G. (1976a). Asupra proprietății lui Bar-Hillel, Perles și Shamir. *Stud. Cerc. Matem.*, 28:303–309. En romanès.
- Păun, G. (1976b). Languages associated to a dramatic work. *Cah. Ling. Th. Appl.*, 13:605–611.
- Păun, G. (1977a). Contextual grammars with restriction in derivation. *Rev. Roum. Math. Pures et Appl.*, 22:1147–1154.
- Păun, G. (1977b). The generative capacity of compound grammars. *Inform. Control*, 34:50–54.
- Păun, G. (1977c). Noi operații cu limbaje. *Stud. Cerc. Matem.*, 29:521–533. En romanès.
- Păun, G. (1977d). An operation with languages occurring in the linguistic approach to management. *R.A.I.R.O. / Theor. Computer Sci.*, 11:303–310.
- Păun, G. (1978a). Asupra unor prelungiri ale limbajelor. *Stud. Cerc. Matem.*, 30:511–520. En romanès.
- Păun, G. (1978b). An infinite hierarchy of contextual languages with choice. *Bull. Math. Soc. Sci. Math. Roumanie*, 22(70):425–430.
- Păun, G. (1978c). On some classes of contextual grammars. *Bull. Math. Soc. Sci. Math. Roumanie*, 22(70):183–189.

- Păun, G. (1978d). Operations with contextual grammars and languages. *Stud. Cerc. Matem.*, 30:425–439. En romanès.
- Păun, G. (1978e). Two infinite hierarchies defined by branching grammars. *Kybernetika*, 14(5):397–407.
- Păun, G. (1979a). A formal linguistic model of action systems. *Ars Semeiotica*, 2:33–47.
- Păun, G. (1979b). Marcus' contextual grammars and languages. A survey. *Rev. Roum. Math. Pures et Appl.*, 24:1467–1486.
- Păun, G. (1979c). On a prolongation operation of languages. *Bull. Math. Soc. Sci. Math. Roumanie*, 23(71):293 – 296.
- Păun, G. (1979d). On the family of finite index matrix languages. *J. of Computer and System Sci.*, 18:267–280.
- Păun, G. (1979e). On the generative capacity of conditional grammars. *Inform. Control*, 43:178–186.
- Păun, G. (1980a). *Generative Mechanisms for Some Economic Processes*. The Technical Publ. House, Bucharest. En romanès.
- Păun, G. (1980b). Modelling the dialogue by means of formal language theory. *Cah. Ling. Th. Appl.*, 27:59 – 65.
- Păun, G. (1980c). Rejection grammars. *Rev. Roum. Ling.*, 27:101–111.

- Păun, G. (1981). *Gramatici matriciale*. Editura științifică și enciclopedică, Bucharest. En romanès.
- Păun, G. (1982a). *Contextual Grammars*. The Publ. House of the Romanian Academy, Bucharest. En romanès.
- Păun, G. (1982b). On the strict inclusions between some families of contextual languages. *Analele Univ. Craiova*.
- Păun, G. (1983). Modelling economic processes by means of formal grammars: a survey of results at the middle of 1981. *Acta Appl. Math.*, 1:79–95.
- Păun, G. (1984). On semicontextual grammars. *Bull. Math. Soc. Sci. Math. Roumanie*, 28(76):63–68.
- Păun, G. (1985a). On some open problems about Marcus contextual languages. *Intern. J. Computer Math.*, 17:9–23.
- Păun, G. (1985b). Semiconditional contextual grammars. *Fundamenta Inform.*, 8:151–161.
- Păun, G. (1985c). Two theorems about Galiukschov semicontextual languages. *Kybernetika*, 21:360–365.
- Păun, G. (1988). A new class of contextual grammars. *Rev. Roum. Ling.*, 33:167–171.

- Păun, G. (1990, 1991). Further remarks on the syntactical complexity of Marcus contextual languages. *Ann. Univ. Buc., Ser. Matem.-Inform.*, 39, 40:72–82.
- Păun, G. (1994). Marcus contextual grammars. After 25 years. *Bulletin EATCS*, 52:263–273.
- Păun, G. (1996a). Contextual Grammars. From Natural Languages to Formal Languages and back. Pendent de publicació.
- Păun, G. (1996b). On the splicing operation. *Discrete Appl. Math.*, en premsa.
- Păun, G. (1996c). Regular extended h systems are computationally universal. *Journal of Automata, Languages, and Combinatorics*, 1:1.
- Păun, G., Rozenberg, G., i Salomaa, A. (1994a). *Developments in Language Theory*, capítol Contextual grammars: erasing, determinism, one-sided contexts, pages 370–388. G. Rozenberg, A. Salomaa, eds., World Sci. Publ., Singapore.
- Păun, G., Rozenberg, G., i Salomaa, A. (1994b). *Mathematical Aspects of Natural and Formal Languages*, capítol Marcus contextual grammars: modularity and leftmost derivation, pages 375–392. Gh. Păun, ed., World Sci. Publ., Singapore.
- Păun, G., Rozenberg, G., i Salomaa, A. (1995a). Contextual grammars: parallelism and blocking of derivation. *Fundamenta Inform.*

- Păun, G., Rozenberg, G., i Salomaa, A. (1995b). Grammars based on the shuffle operation. *J. Universal Computer Sci., JUCS*, 1:67–81.
- Păun, G., Rozenberg, G., i Salomaa, A. (1996a). Computing by splicing. *Theoretical Computer Science*, pendent d'aparició.
- Păun, G., Rozenberg, G., i Salomaa, A. (1996b). Contextual grammars: deterministic derivations and growth functions. *Rev. Roum. Math. Pures et Appl.* Pendent de publicació.
- Păun, G. i Salomaa, A. (1995). Thin and slender languages. *Discr. Appl. Math.*, 61:257–270.
- Păun, G. i Salomaa, A. (1996). Self-reading sequences. *Amer. Math. Monthly.*, pages 184 – 186.
- Păun, G. i Sântean, L. (1989). Parallel communicating grammar systems: the regular case. *Ann. Univ. Buc., Matem.-Inform. Series*, 38:145–156.
- Pawlak, Z. (1982). Rough sets. *Intern. J. Computer and Information Sciences*, pages 341 – 356.
- Penttonen, M. (1974). One-sided and two-sided contexts in formal grammars. *Inform. Control*, 25:371–392.
- Perrin, D. i Beal, M. P. (1996). *Handbook of Formal Languages*, volum II, capítol Symbolic dynamics. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.

- Peters, P. S. i Ritchie, R. W. (1973). On the generative power of transformational grammars. *Information and Sciences*, 6:49–83.
- Petrescu, A. (1976). An extension of the concept of adequate language and of that of saturated partition in algebraic linguistics. *Stud. Cerc. Matem.*, 28(3):311–320. En romanès.
- Petrescu, A. (1977). A domination relation with respect to two partitions of the vocabulary. *Stud. Cerc. Matem.*, 29(1):43–50. En romanès.
- Pin, J. E. (1996). *Handbook of Formal Languages*, volum I, capítol Syntactic monoids and groups. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Piotrovskiĭ, R. G. (1966). *Modelirovanie fonologicheskikh sistem i metody ih aravnenija*. Nauka, Moscú-Leningrad.
- Pogonovski, Y. (1981). *Tolerance Spaces with Applications to Linguistics*. Poznan.
- Pollard, C. J. (1984). *Generalized Structure Grammars, Head Grammars, and Natural Languages*. PhD thesis, Stanford University, California.
- Popescu, E. (1970). Sur quelques problèmes de linguistique algebrique posés par Revzin. *Cah. Ling. Th. Appl.*, 7(1):171–179.
- Popescu, E. (1971). On an algebraic model of part of speech. *Stud. Cerc. Matem.*, 23(10):1551–1560. En romaès.

Popescu, E. (1972). Sur la domination au sens de Trybulec dans la linguistique algebrique. *Cah. Ling. Th. Appl.*, 9(1):131–152.

Popescu, E. (1976). Sur les catégories morphologiques et les fermetures de sestier dans lalinguistique algebrique. *Rev. Roum. Math. Pures et Appl.*, 21(5):539 – 547.

Popescu, M. (1977). Sur un modèle linguistico-mathématique du changement du diagnostic. *Bull. Math. Soc. Sci. Math. Roumanie*, 21(1-2):113 – 122.

Post, E. (1946). A variant of a recursively unsolvable problem. *Bull. Amer. Math. Soc.*, 52:264–268.

Postal, P. (1962). *Some syntactic rules in Mohawk*. PhD thesis, Yale Univ., New Haven, Conn.

Postal, P. (1964). *The Structure of Language: Readings in the Philosophy of Language*, capítol Limitations of phrase structure grammars, pages 137–151. (J. A. Fodor, J. J. Katz, eds.), Prentice-Hall, Englewood Cliffs, N. J.

Prusinkiewicz, P. (1996). *Handbook of Formal Languages*, volum III, capítol Graphic models of development. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.

Păun, G. i Rozenberg, G. (1996). *Handbook of Formal Languages*, volum II, capítol Contextual grammars. G. Rozenberg, A. Salomaa, eds., Springer-

- Verlag, Berlin, Heidelberg. With a prologue (a linguistic point of view) by S. Marcus.
- Pullum, G. H. (1985). On two recent attempts to show that English is not a CFL. *Comput. Ling.*, 10:182–186.
- Pullum, G. K. i Gazdar, G. (1982). Natural Languages and Context-Free Languages. *Linguistics and Philosophy*, 4:471–504.
- Radzinski, D. (1990). Unbounded syntactic copying in Mandarin Chinese. *Linguistics and Philosophy*, 13:113–127.
- Raischi, C. (1970). Nebeský operators and characteristic contexts. *Stud. Cerc. Matem.*, 22(6):893–915. En romanès.
- Raischi, C. (1973). Linguistic systems and Galois correspondences. *Stud. Cerc. Matem.*, 25(5):667–685. En romanès.
- Restivo, A. i Giammaresi, G. (1996). *Handbook of Formal Languages*, volum III, capítol Picture, array and grid languages. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Revzin, I. I. (1962). *Models of Language*. Izdatelstvo Akademii Nauk, Moskow. En rus.
- Revzin, I. I. (1966). *Models of Language*. Methuen and Co. Ltd., Londres.
- Revzin, I. I. (1967a). *Metod modelirovaniia i tipologija alavjanskih jazykov*. Nauka, Moskva.

- Revzin, I. I. (1967b). Some means for investigating amorphous languages. *Rev. Roum. Ling.*, 12(5):417 – 421.
- Revzin, I. I. (1977). *Sovremennaja strukturnaja lingvistika*. Nauka, Moskva.
- Revzina, O. G. (1974). Typological analysis of the category of grammatical gebder. *Rev. Roum. Ling.*, 19(1):23–43. En rus.
- Riguet, J. (1948). Relations binaires, fermetures, correspondences de galois. *Bull. Soc. Math. France*, 76:114 – 155.
- Rosenkrantz, J. (1969). Programmed grammars and classes of formal languages. *J. of the A.C.M.*, 16:107–131.
- Rozenberg, G. i Salomaa, A. (1980). *The Mathematical Theory of L Systems*. Academic Press, New York.
- Rozenberg, G. i Salomaa, A. (1996a). *Handbook of Formal Languages*, volum I, capitol L systems. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Rozenberg, G. i Salomaa, A., editors (1996b). *Handbook of Formal Languages*. Springer-Verlag, Berlin, Heidelberg.
- Sakarovitch, J. (1976). *Monoides syntactiques et langages algebriques*. PhD thesis, Math. Université, Paris VII.
- Salini, Z. i Trybulec, A. (1974). Coherence of a text and its topology. *Semiotica*, 11(2):101–108.

- Salomaa, A. (1973). *Formal languages*. Academic Press, New York, London.
- Savitch, W. J. (1989). A formal model for context-free languages augmented with reduplication. *Comput. Ling.*, 15(4):250–261.
- Savitch, W. J. (1993). Why it might pay to assume that languages are infinite. *Annals of Math. and Artificial Intelligence*, 8:17 – 25.
- Schwartz, L. (1972). On Trybulec domination in algebraic linguistics. *Rev. Roum. Math. Pures et Appl.*, 17(3):425–433.
- Schwartz, L. (1975). On D-invariant and S-invariant sets in algebraic linguistics. *Stud. Cerc. Matem.*, 27(6):731–737. En romanès.
- Schwartz, L. (1976). A general notion of domination relation in algebraic linguistics. *Cah. Ling. Th. Appl.*, 13(1):305–313.
- Schwartz, L. (1982). Binary relations, Galois connections and topological closure operators. *Stud. Cerc. Matem.*, 34(1):73 – 83. En romanès.
- Schweiger, F. (1972). Einige mathematische Überlegungen zur Theorie initialer Teilwörter. *Linguistics*, 93:88–94.
- Semeniuk-Polkowska, M. (1982). Syntactic pseudo-configurations based on Trybulec domination. *Rev. Roum. Math. Pures et Appl.*, 27(6):709 – 717.
- Serrano, S. (1975). *Elementos de lingüística matemática*. Anagrama, Barcelona.

- Serrano, S. (1977). *Lógica, lingüística y matemáticas*. Anagrama, Barcelona.
- Sestier, A. (1961). Contributions à une théorie ensembliste des classifications linguistiques. A *Actes du I-er Congrès de l'AFCAL, Grenoble, 1960*, pages 293–305, Paris.
- Shieber, S. M. (1985). Evidence against the context-freeness of Natural Languages. *Linguistics and Philosophy*, 8:333–343.
- Shreider, J. A. (1970). *Conjunct-Theoretic Models of Languages*, capítulo The algebra of binary relations. S. Marcus, Nauka, Moscow. Annex. En rus.
- Shreider, J. A. (1973). On the O. S. Kulagina concept of derived relation. *Comput. Ling.*, 9:241 – 253.
- Sloane, N. J. A. i Plouffe, S. (1995). *The Encyclopedia of Integer Sequences*. Academic Press, New York, London.
- Solomon, S. Y. (1973). The n -derivative of a partition. *Recueil Linguistique de Bratislava*, 4:237 – 243.
- Souriau, E. (1950). *Les deux cent mille situations dramatiques*. Flammarion, Bibliothèque d'Esthétique, Paris.
- Squier, C. C. (1988). Semicontextual grammars: an example. *Bull. Math. Soc. Sci. Math. Roumanie*, 32(80):167–170.

- Staiger, L. (1996). *Handbook of Formal Languages*, volum I, capítol Omega-languages. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Tesnière, L. (1959). *Eléments de syntaxe structurale*. Klincksieck, Paris.
- Thomas, W. (1996). *Handbook of Formal Languages*, volum III, capítol Languages, automata and logic. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Toma-Marcu, E. (1969). On an operator related to Sestier operator. *Stud. Cerc. Matem.*, 21(3):499–508. En romanès.
- Trybulec, A. (1967a). The descomposition of the vocabulary in languages with given paradigms. *Naučno-Techničeskaja Informacija*, 12:40–44.
- Trybulec, A. (1967b). On some classes of syntactic relations. *Naucno-Techniceskaja Informacija*, 9:34–37.
- Turakainen, P. (1978). On characterization of recursively enumerable languages in terms of linear languages and vw-grammars. A *Proc. Koninklijke Nederlandse Akademie van Wetenschappen*, volum 81 of *Ser. A*, pages 145 – 153, Amsterdam.
- Uspenskij, V. A. (1957a). On the definition of case by a. n. kolmogorov. *Byulleten Obyedinenija po Problemam Mashinogo Perevoda*, 5:11 – 18. En rus.

- Uspenskij, V. A. (1957b). On the definition of parts of speech in the set-theoretic system of language. *Byulleten Obyedinenija po Problemam Mashinogo Perevoda*, 5:22 – 26. En rus.
- Valk, R. (1972). Topologische Wortmengen, topologische automaten, Zustandesendliche, stetige Abbildungen. *Der Gesellschaft für Mathematik und Datenverarbeitung*, 19.
- van Helden, W. A. (1993). *Case and Gender-Concept Formation Between Morphology and Syntax*. Rodopi, Amsterdam-Atlanta, Georgia. 2 volums.
- Vasilescu, L. (1973). Analyse distributionelle algébrique des substantifs français. *Cah. Ling. Th. Appl.*, 10(2):251 – 259.
- Vasilescu, L. (1974). Analyse distributionelle algébrique des formes verbales du français. *Cah. Ling. Th. Appl.*, 11(1):137 – 169.
- Vasilescu, L. (1975). Analyse distributionelle algébrique des formes adjectivales du français. *Rev. Roum. Ling.*, 20(3):233 – 248.
- Vasilevslîi, J. P. L. (1973). On a construction of weak closure. *Matematičeskaja Lingvistika*, pages 89–103. VINITI, Nauka, Moskva. En rus.
- Vicolov, S. (1991). Two theorems about Marcus contextual languages. *Bull. Math. Soc. Sci. Math. Roumanie*, 35(83):167–170.
- Vicolov-Dumitrescu, S. (1995). *Mathematical Linguistics and Related Topics*, capítol On total parallelism in contextual grammars, pages 350–360. Gh. Păun, ed., The Publ. House of the Romanian Academy, Bucharest.

- von Solms, S. H. (1975). The characterization by automata of certain classes of languages in the context-sensitive area. *Inform. Control*, 27:262–271.
- Šaumjan, S. K. (1968). *Problems of Theoretical Phonology*. Mouton, The Hague.
- Šreider, J. A. (1970). *Set-theoretic Models of Language*, capítol Annex: The algebra of binary realtions. S. Marcus ed., Nauka, Moskva. En rus.
- Šreider, J. A. (1973). On the o. s. kulagina concept of derived relation. *Comput. Ling.*, 9:241–253. En rus.
- Ward, M. (1942). The closure operators of a lattice. *Annals of Math.*, 43(2):191 – 196.
- Wood, D. (1996). *Handbook of Formal Languages*, volum I, capítol Grammar and L forms. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Yu, S. (1996). *Handbook of Formal Languages*, volum I, capítol Regular languages. G. Rozenberg, A. Salomaa, eds., Springer-Verlag, Berlin, Heidelberg.
- Zadeh, L. (1965). Fuzzy sets. *Inform. Control*, 8:338 – 353.
- Zelinka, B. (1966). Sur le PS-isomorphisme des langues. *Zeitschrift fur Math. Logik und Grundlagen der Math.*, 12(4):263 – 265.

Zelinka, B. (1967a). Sur les langages absolument non-adequats. *Rev. Roum. Math. Pures et Appl.*, 12(5):737 – 752.

Zelinka, B. (1967b). Un langage avec la distance contextuelle finie, mais non-bornée. *Sbornik Vědeckých Prací Vysoké Školy Strojní a Textilní v Liberci*, 6:9–12.

