



การแก้ไขการผิดไวยากรณ์โดยวิธีการแบ่งส่วนแบบย้อนรอยกองซ้อนโครงสร้างกราฟ

7.1 การแก้ไขการผิดไวยากรณ์โดยวิธีการแบ่งส่วนแบบย้อนรอยกองซ้อนโครงสร้างกราฟ

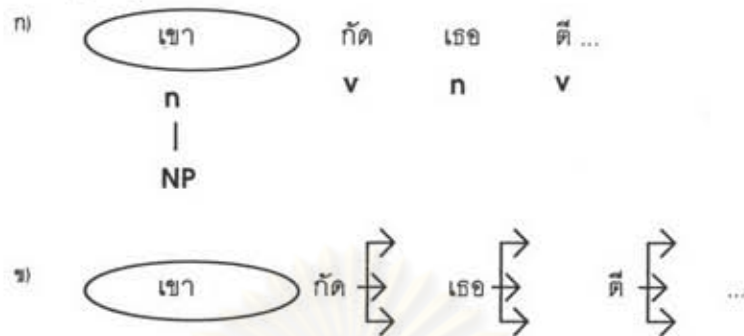
จากบทที่ผ่านมาเราจะเห็นได้อย่างชัดเจนว่า จุดที่มีการตรวจพบข้อผิดพลาดอาจจะไม่ใช่จุดที่ทำให้เกิดการผิดไวยากรณ์ที่แท้จริงในประโยคก็ได้ และถ้าไม่มีการย้อนรอยกลับมาเพื่อพิจารณาแก้ไขประโยคในจุดอื่นที่ผ่านการแบ่งส่วนมาแล้ว ก็อาจมีข้อผิดพลาดบางแห่งที่ถูกมองข้ามไป ดังตัวอย่างสุดท้ายในบทที่ 6 ตามความเป็นจริงแล้วเป็นการยากมากที่จะให้คอมพิวเตอร์กำหนดจุดที่ผิดพลาดได้อย่างถูกต้องแน่นอน ดังนั้นถ้าจะให้มีการครอบคลุมข้อผิดพลาดทั้งหมดจริงๆ เมื่อมีการตรวจพบข้อผิดพลาดแล้ว คงต้องย้อนกลับมาทำการแก้ไขในทุกจุดที่ผ่านมาทั้งหมด แต่การทำเช่นนั้นจะทำให้เสียเวลาในการทำงานในส่วนที่ไม่เกี่ยวข้องมากเกินไป และไม่เหมาะสมในทางปฏิบัติ (ทางเลือกที่พบจะมีมากเกินไป รายละเอียดอ่านได้ในบทที่ 8) ในงานที่ทำนี้จึงพยายามลดจำนวนจุดที่จะทำการแก้ไขประโยคลง ให้เหลือเฉพาะจุดที่มีความเป็นไปได้สูงกว่าจุดอื่นที่มีโอกาสเกิดข้อผิดพลาดขึ้นได้

การทำงานของวิธีการแบ่งส่วนแบบย้อนรอยกองซ้อนโครงสร้างกราฟนั้น อาศัยลักษณะการทำงานพื้นฐานของตัวแบ่งส่วนแบบแอลอาร์ ที่มีการยุบรวมองค์ประกอบย่อยในกองซ้อนให้เป็นองค์ประกอบที่ใหญ่ขึ้น ในทันทีที่เป็นไปได้ ภายใต้เงื่อนไขข้อนี้ จึงได้ตั้งสมมุติฐานขึ้นมาว่า 'โอกาสที่จะเกิดข้อผิดพลาดในองค์ประกอบที่ยุบรวมกันแล้วนั้น มีค่าน้อยกว่าโอกาสที่จะเกิดความผิดพลาดในองค์ประกอบอิสระที่ยังไม่ได้รวมกัน' โดยที่ องค์ประกอบที่ยุบรวมกัน (grouped constituent) นั้น เกิดจากขั้นตอนการลดทอนด้วยกฎการผลิตที่มีอยู่ และองค์ประกอบที่ยุบรวมกันนี้จะยุบรวมกันได้โดยสมบูรณ์ เมื่อขั้นตอนการลดทอนจบลงโดยไม่มีข้อผิดพลาดเกิดขึ้น (ตัวแบ่งส่วนสามารถทำการปรับเลื่อน หรือมีการยอมรับประโยคเกิดขึ้นแล้ว) มิฉะนั้นจะถือว่าการลดทอนที่เกิดขึ้นไม่สมบูรณ์ และกลุ่มองค์ประกอบย่อยที่นำมาลดทอนก็จะเป็นเพียง องค์ประกอบอิสระ (ungrouped element) เท่านั้น

ดังนั้นหลังจากมีการตรวจพบข้อผิดพลาดขึ้นในประโยคที่ทำการแบ่งส่วน นอกจากตัวแบ่งส่วนจะทำการแก้ไขประโยค ณ ตำแหน่งที่ตรวจพบข้อผิดพลาดแล้ว องค์ประกอบอิสระทุกตัวที่มีอยู่ในสายโครงสร้างนั้นก็จะถูกย้อนกลับไปทำการแก้ไขด้วย

จากตัวอย่างประโยคผิดไวยากรณ์ 'เขา กัด เธอ ตี ฉัน' เมื่อมีการตรวจพบข้อผิดพลาดที่ตำแหน่ง 'ตี' แล้ว ในกองซ้อนโครงสร้างกราฟขณะนั้นจะมีองค์ประกอบต่างๆ เก็บอยู่ ดังแสดงในรูปที่ 6.1 โดยที่ 'เขา' เป็นองค์ประกอบที่ยุบรวมกันแล้ว ส่วน 'กัด', 'เธอ' และ 'ตี' เป็นองค์ประกอบอิสระ

ดังนั้นเมื่อมีการตรวจพบข้อผิดพลาดขึ้น จุดที่เป็นองค์ประกอบอิสระทั้ง 3 จุด จะกลายเป็นจุดเริ่มต้นของการแก้ไขประโยค โดยแต่ละจุดจะถูกแยกออกเป็น 3 สายเพื่อจัดการกับข้อผิดพลาดทั้ง 3 แบบ



รูปที่ 7.1 ก) แสดงองค์ประกอบต่างๆ ที่มีในกองซ้อนขณะที่มีการตรวจพบข้อผิดพลาดที่ 'ตี'
ข) คำที่เป็นองค์ประกอบอิสระทั้ง 3 จุด แต่ละจุดจะถูกแยกออกเพื่อแก้ไขประโยค

จากรูปที่ 7.1 ส่วนนามวลี (NP) เป็นองค์ประกอบที่ยุบรวมกันแล้ว โดยการยุบคำนาม (n) 'เขา' ตามกฎการผลิตข้อที่ 3 คือ NP \rightarrow n อันที่จริงแล้วคำนาม 'เธอ' ควรจะถูกยุบรวมเป็นองค์ประกอบนามวลีด้วยตามกฎที่กำหนด แต่ก่อนที่การลดทอนของคำทั้งสองจะเสร็จสมบูรณ์ ได้มีการตรวจพบข้อผิดพลาดขึ้นก่อน ดังนั้นในกรณีนี้คำนาม 'เธอ' จึงยังเป็นแคงค์ประกอบอิสระเท่านั้น

เนื่องจากการแก้ไขข้อมูลแบบนี้ มีการย้อนรอยกลับไปแก้ไขคำในส่วนที่ผ่านการแจงส่วนมาแล้ว เพื่อที่เราจะได้ไม่ต้องทำการแจงส่วนใหม่ตั้งแต่ต้นประโยคในทุกจุดที่ทำการแก้ไข เราจำเป็นต้องมีการเก็บข้อมูลของสถานะเริ่มต้นก่อนการแจงส่วนคำแต่ละคำเอาไว้ (เก็บเฉพาะสถานะเริ่มต้น ก็ที่จะดึงคำต่อไปมาแจงส่วน) เมื่อมีความจำเป็นต้องกลับมาแก้ไขประโยคที่ตำแหน่งนั้นๆ ก็จะต้องดึงสถานะเริ่มต้นของแต่ละคำออกมา แล้วทำการแจงส่วนประโยคต่อไป สถานะเริ่มต้น ที่จำเป็นต้องเก็บอาจมีความแตกต่างกันไป ตามแต่ผู้เขียนโปรแกรมจะเห็นว่าจำเป็นต้องใช้ในการทำงานของตน ถ้าจะอ้างอิงตามขั้นตอนที่ใช้ในการอธิบายการแจงส่วนในวิทยานิพนธ์เล่มนี้ สถานะเริ่มต้นที่ต้องเก็บก็คือ หมายเลขของขั้นตอนก่อนหน้าที่จะดึงคำ คำใหม่มาแจงส่วน แล้วเก็บลงใน กองซ้อนพร้อมกับคำๆนั้น เมื่อมีการทำงานปรับเปลี่ยน เช่น ตัวอย่างต่อไปในรูป 7.2 จะเห็นว่า GSS(0) เก็บคำ n (เขา, 1) หมายความว่า สถานะเริ่มต้นก่อนการแจงส่วนคำนาม "เขา" อยู่ที่ขั้นตอนที่ (1) ถ้ามีการแก้ไขประโยคที่ตำแหน่งคำ "เขา" ก็จะต้องเริ่มทำงานในขั้นตอนที่ (1) หรือ คำ GSS(2) เก็บคำ v (กัด, 2) ก็คือถ้าจะแก้ไขประโยคที่ตำแหน่งคำ "กัด" จะต้องเริ่มแก้ไขตั้งแต่ขั้นตอนที่ (2) ไม่ใช่ขั้นตอนที่ (4)

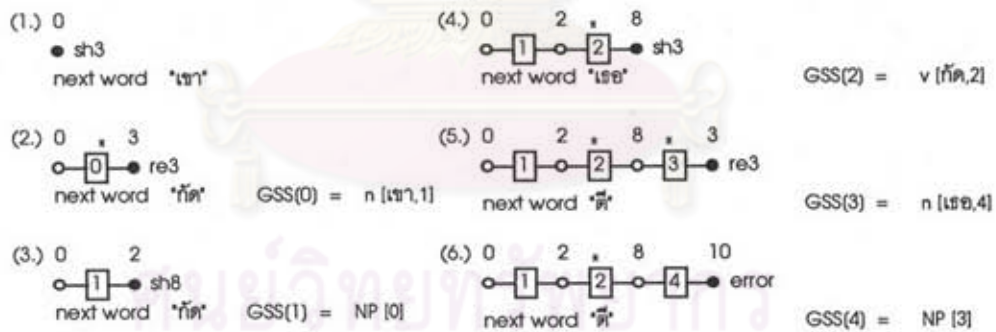
รูปที่ 7.2 ถึงรูปที่ 7.6 แสดงการทำงานของการทำงานการแจงส่วนประโยคแบบย้อนรอยกองซ้อนโครงสร้างกราฟ เมื่อทำการแจงส่วนประโยค 'เขา กัด เธอ ตี ฉัน' กองซ้อนที่มีเครื่องหมายดาว (asterisk,*) อยู่ข้างบน คือองค์ประกอบที่เป็นองค์ประกอบอิสระ และเพื่อความง่ายในการแสดงการแจงส่วนประโยค จึงขอใช้เงื่อนไขเดิมในการจำกัดการแก้ไขประโยคของตัวแจงส่วน โดยกำหนดให้มีข้อผิดพลาดได้เพียง 1 จุดเท่านั้นในประโยค อย่างไรก็ตามในการใช้งานจริง ทั้งการแจงส่วนแบบย้อนรอยกองซ้อนโครงสร้างกราฟ และการแจงส่วนแบบจีแอลอาร์-ไอจีแอลอาร์ สามารถที่จะใช้จัดการกับข้อผิดพลาดในประโยคได้มากกว่า 1 จุด

State	Action Field				Goto Field			
	n	v	p	\$	NP	PP	VP	S
0	sh3				2			1
1			sh5	acc		4		
2		sh8	sh5			7	6	
3		re3	re3	re3				
4			re2	re2				
5	sh3				9			
6			re1	re1				
7		re4	re4	re4				
8	sh3		sh5		10	11		
9		re5	re5/sh5	re5		7		
10			re6/sh5	re6		7		
11			re7	re7				

ตารางที่ 7.1 ตารางแอลอาร์ที่ใช้ในการแจงส่วนประโยค

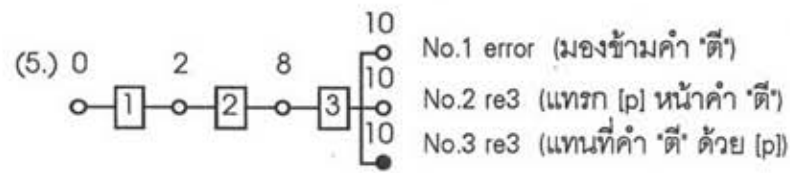
ประโยคชนิดรูปแบบที่ทำการแจงส่วน : เขา กัด เธอ ดี ฉันทัน

ชนิดของคำที่ใช้ในการแจงส่วน : n v n v n



รูปที่ 7.2 แสดงขั้นตอนการทำงานและการตรวจพบข้อผิดพลาดที่ 'ดี'

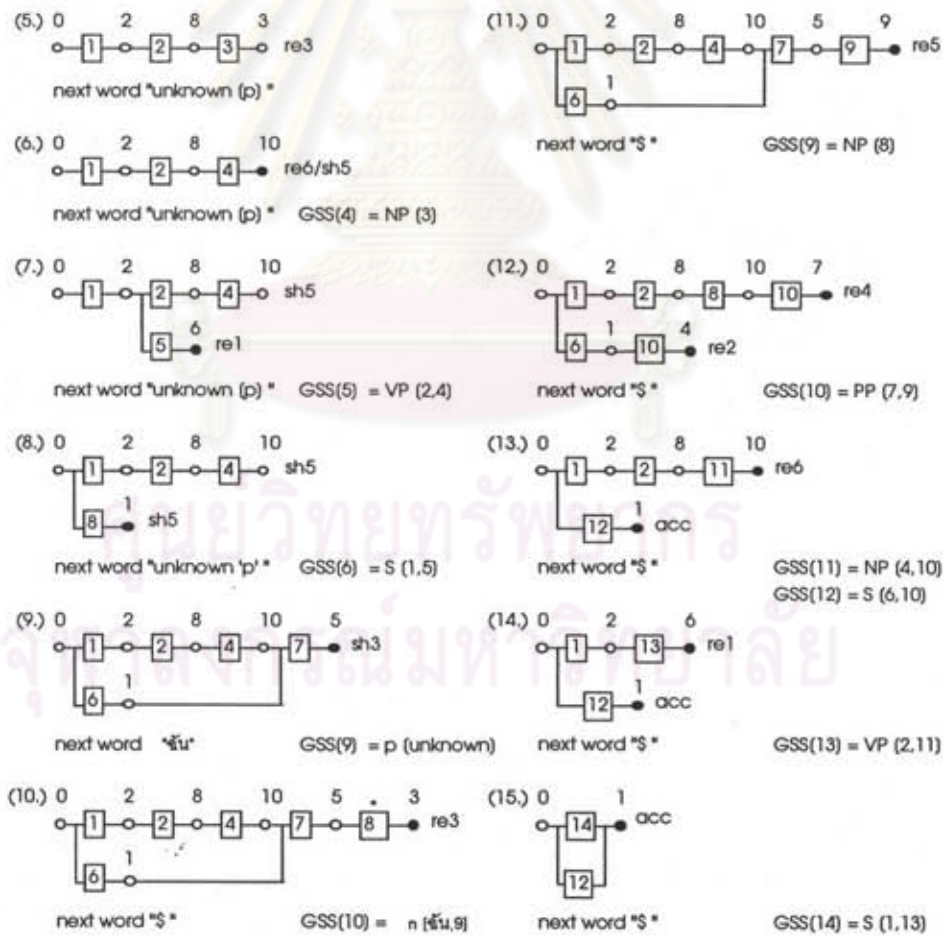
เมื่อมีการตรวจพบข้อผิดพลาดที่ขั้นตอนที่ (6) ตัวแจงส่วนจะทำการถอยหลังไปขั้นตอนที่ (5) เพื่อตรวจสอบตำแหน่งของคำที่มีโอกาสเป็นจุดเริ่มต้นที่จะแก้ไขประโยค ซึ่งในกรณีนี้ก็คือ คำที่เป็นองค์ประกอบอิสระทั้ง 3 คือ 'กัด', 'เธอ' และ 'ดี' และภายใต้เงื่อนไขที่กำหนดให้มีข้อผิดพลาดได้เพียง 1 จุด เท่านั้นในประโยคนี้ จะมีเพียงการแก้ไขประโยคที่คำ 'กัด' และคำ 'ดี' เท่านั้นที่สามารถทำการแจงส่วนประโยคได้จนเสร็จสมบูรณ์



path No.1 next word = 'จัน'
 path No.2 next word = 'unknown [p]'
 path No.3 next word = 'unknown [p]'

รูปที่ 7.3 แสดงการแยกโครงสร้างข้อมูลในขั้นตอนที่ (5) เพื่อแก้ไขประโยคที่คำ 'ตี'

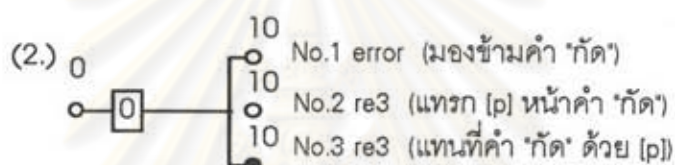
การแก้ไขประโยคที่ตำแหน่งของคำกริยา 'ตี' โครงสร้างข้อมูลในขั้นตอนที่ (5) ที่เป็นขั้นตอนที่เริ่มแจງ ส่วนคำ 'ตี' จะถูกแยกออกเป็น 3 สายเพื่อจัดการกับข้อผิดพลาดที่กำหนด และมีแต่สายที่ 3 เท่านั้นที่สามารถทำ การแจງส่วนต่อไปได้จนจบประโยค



รูปที่ 7.4 แสดงการแก้ไขประโยคในสายที่ 3 ของการแก้ไขประโยคที่ 'ตี'

จากรูปที่ 7.4 สายที่ 3 ของการแก้ไขประโยคที่ "ดี" ถูกแยกออกมาเพื่อทำการแก้ไขข้อผิดพลาดที่เกิดจากการแทนที่องค์ประกอบใดๆ ในประโยค ในที่นี้ตัวแฉงส่วนได้พิจารณาจากตารางแอลอาร์ สถานะพร้อมที่ 3 ชนิดของคำที่มีโอกาสเป็นไปได้ในตำแหน่งนี้คือ คำกริยา และคำบุพบท แต่คำกริยาเป็นชนิดคำแบบเดียวกับ "ดี" ที่ตรวจพบข้อผิดพลาด ดังนั้นชนิดของคำที่เหมาะสมในการนำมาแทนที่คำ "ดี" ในที่นี้จึงมีคำเดียวคือ คำบุพบท เพราะฉะนั้นตัวแฉงส่วนจึงแทนที่คำ "ดี" ด้วยชนิดคำบุพบท [p] ในขั้นตอนที่ (5) และทำการแฉงส่วนประโยคต่อไปจนจบในสายโครงสร้างนี้ และได้ประโยคที่ผ่านการแก้ไขของตัวแฉงส่วนออกมา คือ ประโยค "เรา กัด เธอ [p] ฉััน"

ตำแหน่งต่อมาที่สามารถทำการแก้ไขประโยคได้จนสมบูรณ์ตามเงื่อนไขที่กำหนดคือ การแก้ไขประโยคที่ตำแหน่งของคำ "กัด" ดังนั้นโครงสร้างข้อมูลในขั้นตอนที่ (2) ซึ่งเป็นขั้นตอนที่คำกริยา "กัด" ถูกแฉงส่วน จึงถูกแยกออกเป็น 3 สาย ดังแสดงในรูปที่ 7.5



path No.1 next word = "เธอ"

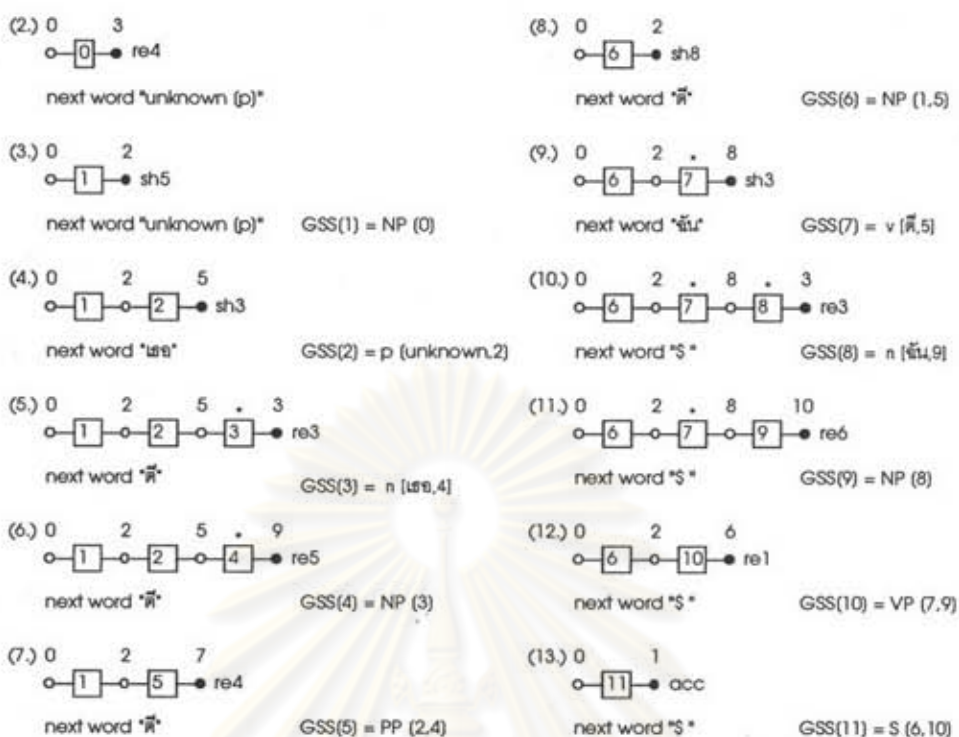
path No.2 next word = "unknown [p]"

path No.3 next word = "unknown [p]"

รูปที่ 7.5 แสดงการแยกโครงสร้างข้อมูลในขั้นตอนที่ (2) เพื่อแก้ไขประโยคที่ "กัด"

และด้วยเหตุผลเดียวกันกับการแก้ไขประโยคที่ตำแหน่งของคำ "ดี" จะมีเพียงโครงสร้างสายที่ 3 เท่านั้นที่สามารถทำงานต่อไปจนจบประโยคดังแสดงในรูปที่ 7.6 และได้ประโยคที่ผ่านการแก้ไขของตัวแฉงส่วนออกมา คือ ประโยค "เรา [p] เธอ ดี ฉััน"

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 7.6 แสดงการแก้ไขประโยคในโครงสร้างสายที่ 3 ของตำแหน่งคำ "กัด"

หลังจากการแก้ไขประโยคที่ตำแหน่งคำทั้งสอง ประโยคที่แก้ไขโดยตัวแฉงส่วนที่ได้มาคือ

การแก้ไขประโยคที่คำ "ดี" : เขา กัด เธอ [p] ฉัน

การแก้ไขประโยคที่คำ "กัด" : เขา [p] เธอ ดี ฉัน

จะเห็นได้ว่าการแฉงส่วนแบบย้อนรอยกองซ้อนโครงสร้างกราฟ สามารถที่จะแก้ไขปัญหาที่ถูกมองข้ามไปจากการแก้ไขข้อผิดพลาด ณ ตำแหน่งที่ตรวจพบเพียงอย่างเดียวได้ นอกจากนี้โดยทั่วไปแล้วถึงแม้การแฉงส่วนแบบนี้จะยากในการเขียนโปรแกรมเพื่อควบคุมการทำงาน และใช้เวลาในการทำงานมากกว่าการแฉงส่วนแบบจีแอลอาร์-ไอจีแอลอาร์ก็ตาม จากการทดลองแฉงส่วนประโยคผิดไวยากรณ์ที่ซับซ้อนขึ้น จะพบว่าการแฉงส่วนแบบย้อนรอยกองซ้อนโครงสร้างกราฟ สามารถครอบคลุมข้อผิดพลาดได้มากกว่าการแฉงส่วนแบบจีแอลอาร์-ไอจีแอลอาร์ ลองพิจารณาจากตัวอย่างสุดท้ายที่ได้กล่าวถึงในบทก่อน

ประโยคที่ถูกไวยากรณ์ : เขา กับ เธอ ดี ฉัน กับ น้อง

ประโยคที่ผิดไวยากรณ์ : เขา กัด เธอ ดี ฉัน กัด น้อง

ประโยคที่แก้ไขแล้ว จากการแฉงส่วนแบบจีแอลอาร์-ไอจีแอลอาร์

ขั้นตอนแรก : เขา กัด เธอ [p] ฉัน [p] น้อง

ขั้นตอนที่ 2 : เขา [p] เธอ [p] ฉัน กัด น้อง

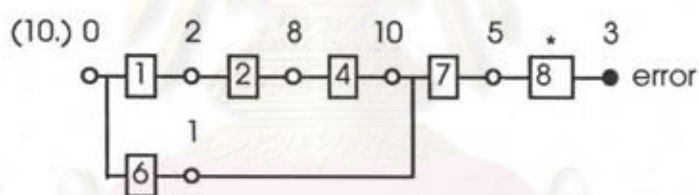
อย่างที่ได้ออกมาแล้วว่าประโยคที่ได้ออกมาทั้งสองประโยค ยังคงมีข้อผิดพลาดที่ถูกมองข้ามไปปรากฏอยู่ในประโยค ถ้านำประโยคนี้ออกมาบางส่วนโดยใช้การแจقس่วแบบย้อนรอยกองซ้อนโครงสร้างกราฟ จะสามารถจัดการกับปัญหาในลักษณะนี้ได้ ดังประโยคที่ขีดเส้นใต้สามารถจัดการกับข้อผิดพลาดทั้งสองจุดได้

ประโยคที่แก้ไขแล้ว จากการแจقس่วแบบย้อนรอยกองซ้อนโครงสร้างกราฟ

- เขา กัด เธอ [p] ฉัน [p] น่อง
- เขา [p] เธอ ตี ฉัน [p] น่อง
- เขา [p] เธอ [p] ฉัน กัด น่อง

7.2 ข้อควรระวังในการพัฒนาโปรแกรม

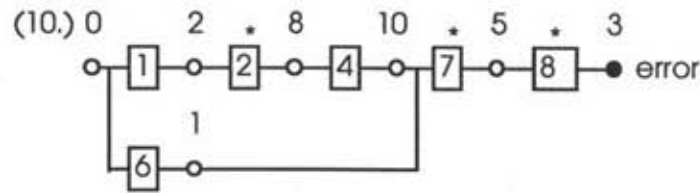
1. การเกิดการวนซ้ำไม่รู้จบ ในกรณีที่มีการเรียกใช้ส่วนการแจقس่วแบบย้อนรอยโครงสร้างกราฟนั้น หลังจากทำการแก้ไขประโยคที่ตำแหน่งต่างๆ แล้ว จะต้องถือเสมือนว่านับจากจุดที่เราเริ่มแก้ไขประโยค ถอยหลังกลับไปถึงต้นประโยค จะไม่มีองค์ประกอบอิสระอีกแล้วในสายโครงสร้างนั้น เพื่อว่าในการแก้ไขข้อผิดพลาดที่อาจจะเกิดต่อไปข้างหน้าจะได้เริ่มแก้ไขประโยคที่จุดถัดไปข้างหน้า โดยไม่ย้อนกลับมาทำให้เกิดการวนซ้ำไม่รู้จบ



รูปที่ 7.7 แสดงตัวอย่างการตรวจพบข้อผิดพลาดจุดที่ 2 ในขั้นตอนที่ (10) จากรูปที่ 7.4

จากตัวอย่างในรูปที่ 7.4 ขั้นตอนที่ (10) สมมติว่ามีการตรวจพบข้อผิดพลาดในขั้นตอนนี้อีก ถ้าจะพิจารณาจริงๆ แล้วในขั้นตอนนี้ องค์ประกอบที่เป็นองค์ประกอบอิสระในโครงสร้างสายบนก็คือ GSS(2) 'กัด' , GSS(7) 'unknown [p]' และ GSS(8) 'ฉัน' แต่ขั้นตอนเหล่านี้มาจากการแก้ไขประโยคโดยการแทนที่ 'ตี' ด้วย 'unknown [p]' ในขั้นตอนที่ (5) ดังนั้นในที่นี้เราจะถือว่า GSS(2) และ GSS(7) เป็นองค์ประกอบที่ผ่านการแก้ไขมาแล้วจะไม่แก้ไขซ้ำอีก จุดเริ่มต้นที่จะทำการแก้ไขประโยคสำหรับการตรวจพบข้อผิดพลาดครั้งที่ 2 ในกรณีนี้จึงเหลือเพียงจุดเดียวที่ คำนาม 'ฉัน'

2. การเลือกสายโครงสร้างข้อมูล ที่จะใช้ในการพิจารณาหาจุดเริ่มต้นในการแก้ไขประโยค ในกรณีนี้โครงสร้างข้อมูลของขั้นตอนที่ใช้พิจารณาหาจุดเริ่มต้นของการแก้ไขประโยค มีการแยกเป็นหลายสายโครงสร้างในที่นี้ให้เลือกพิจารณาองค์ประกอบอิสระจากสายที่มีองค์ประกอบอิสระมากที่สุด เพื่อว่าจะได้มีโอกาสครอบคลุมข้อผิดพลาดได้มากขึ้น



รูปที่ 7.8 แสดงตัวอย่างองค์ประกอบอิสระในแต่ละสายโครงสร้างข้อมูล

จากตัวอย่างในรูปที่ 7.8 สมมติว่าโครงสร้างข้อมูลในสายนี้ ยังไม่เคยผ่านการแก้ไขประโยคมาก่อน และมีองค์ประกอบของข้อมูลในกองซ้อนดังกล่าว คือโครงสร้างสายบนมีองค์ประกอบอิสระ 3 จุดคือ GSS(2), GSS(7) และ GSS(8) ส่วนโครงสร้างสายล่างมีองค์ประกอบอิสระ 2 จุด คือ GSS(7) และ GSS(8) ในกรณีนี้จะยึดองค์ประกอบสายบนเป็นหลัก และทำการแก้ไขประโยคที่ 3 จุดคือ GSS(2), "กั๊ด", GSS(7), "unknown [p]" และ GSS(8), "ฉัน" อย่างไรก็ตามในระบบที่ต้องการความเร็วในการประมวลผล ก็อาจจะเลือกสายโครงสร้างที่มีองค์ประกอบอิสระน้อยที่สุดก็ได้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย