

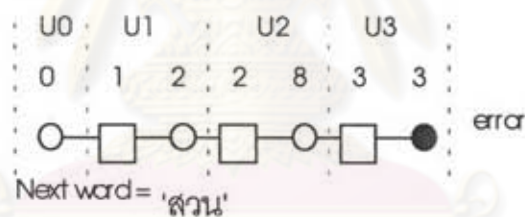
บทที่ 4

การตรวจพบการผิดพลาดไวยากรณ์

4.1 การตรวจพบการผิดพลาดไวยากรณ์

การตรวจพบการผิดพลาดไวยากรณ์ของตัวแบ่งส่วนแบบแอลอาร์ ก็คือการที่ตัวแบ่งส่วนตรวจหาค่าการกระทำในขั้นต่อไปจากตารางแอลอาร์ แต่ไม่มีการกระทำในตารางที่สถานะปัจจุบันให้กระทำ เช่น จากตัวอย่างที่ผ่านมา ประโยคที่นำมาแบ่งส่วนคือ 'ฉัน เห็น เขา ใน สวน' ถ้าประโยคที่นำมาแบ่งส่วนเกิดข้อผิดพลาดโดยการขาดคำบุพบท 'ใน' ในประโยคไป ประโยคที่นำมาแบ่งส่วนกลายเป็นประโยค 'ฉัน เห็น เขา สวน' ดังนั้นในขั้นตอนที่ 5 จากรูปที่ 3.11 ในบทที่ 3 คำต่อไปที่จะนำมาแบ่งส่วนต่อไปจะกลายเป็น คำนาม 'สวน'

5)



รูปที่ 4.1 การตรวจพบข้อผิดพลาดที่คำนาม 'สวน' ขณะที่สถานะพร้อมคือ สถานะที่ 3

จากตารางแอลอาร์ (ตารางที่ 3.1) สถานะที่ 3 คอลัมน์ (n) ไม่มีการกระทำในสถานะนี้ แสดงว่ามีการผิดพลาดไวยากรณ์เกิดขึ้นในประโยคที่กำลังทำการแบ่งส่วน

4.2 การจำแนกประเภทของข้อผิดพลาดที่ตรวจพบ

การผิดพลาดทางไวยากรณ์ที่ตรวจพบ อาจมีสาเหตุมาจากหลายกรณี ในการทำวิทยานิพนธ์เพื่อแก้ไขการผิดพลาดไวยากรณ์ที่เกิดขึ้นนี้ ได้จำแนกข้อผิดพลาดที่เกิดขึ้นออกเป็น 3 กรณี (Malone and Felshin, 1991; Saito and Tomita, 1991) คือ

4.2.1 ข้อผิดพลาดที่เกิดจากการแทรกองค์ประกอบใดๆ ลงในประโยค (Inserted element error) เช่น
ประโยคที่ถูกไวยากรณ์: ฉัน เห็น เขา ใน สวน

ประโยคที่ผิดไวยากรณ์: ฉัน เห็น เขา ใน ใน สวน

จากตัวอย่างประโยคที่ให้ จะมีคำบุพบท 'ใน' แทรกเข้ามาอีกหนึ่งคำ ทำให้เกิดการผิดไวยากรณ์ขึ้น

4.2.2 ข้อผิดพลาดที่เกิดจากการลบหรือการขาดองค์ประกอบใดๆ ในประโยค (Deleted element error) เช่น

ประโยคที่ถูกไวยากรณ์: ฉัน เห็น เขา ใน สวน

ประโยคที่ผิดไวยากรณ์: ฉัน เห็น เขา สวน

ประโยคนี้การผิดไวยากรณ์เกิดจากการขาดคำบุพบท 'ใน' ในประโยค

4.2.3 ข้อผิดพลาดที่เกิดจากการแทนที่องค์ประกอบใดๆ ในประโยค (Substituted element error) เช่น

ประโยคที่ถูกไวยากรณ์: ฉัน เห็น เขา กับ เธอ

ประโยคที่ผิดไวยากรณ์: ฉัน เห็น เขา กัด เธอ

การผิดพลาดในข้อที่ 4.2.3 นี้มักเกิดจากการพิมพ์คำผิด หรือการสะกดคำผิด จากตัวอย่างเป็นการพิมพ์คำผิด เป็นคำกริยา 'กัด' แทนที่คำบุพบท 'กับ' ทำให้ในประโยคตัวอย่างมีคำกริยามากกว่า 1 คำในประโยค ดังนั้นคำกริยาคำที่สอง 'กัด' จะถูกตรวจพบเป็นข้อผิดพลาด (อ้างอิงตามไวยากรณ์ที่กำหนดไว้ในรูปที่ 3.5 แต่ละประโยคจะมีคำกริยาได้ 1 ตัวเท่านั้น)

4.3 การแก้ไขการผิดไวยากรณ์ที่เกิดขึ้น

เมื่อมีการตรวจพบการผิดไวยากรณ์ในประโยค ในขณะนี้ตัวเจงส่วนยังไม่สามารถแยกได้อย่างแน่นอนว่าข้อผิดพลาดที่เกิดขึ้นนั้นเกิดจากสาเหตุใดกันแน่ ดังนั้นตัวเจงส่วนจะทำการแยก (split) โครงสร้างข้อมูลการเจงส่วน (parsed tree) ณ ขณะที่มีการตรวจพบข้อผิดพลาด ออกเป็น 3 สาย (path) เพื่อแก้ไขข้อผิดพลาดประเภทต่างๆ ทั้ง 3 แบบดังที่กล่าวมาแล้ว คือ

4.3.1 การแก้ไขข้อผิดพลาดที่เกิดจากการแทรกองค์ประกอบใดๆ ลงในประโยค (Handling inserted element error) เช่น

ประโยคที่ถูกไวยากรณ์: ฉัน เห็น เขา ใน สวน

ประโยคที่ผิดไวยากรณ์: ฉัน เห็น เขา ใน ใน สวน

ประโยคที่แก้ไขแล้ว : ฉัน เห็น เขา ใน สวน

การแก้ไขการผิดพลาดไวยากรณ์แบบนี้ตัวแจะส่วนจะทำการข้าม หรือผ่าน คำที่ตรวจพบข้อผิดพลาดไปแล้ว ดั่งคำที่จะทำการแจะส่วนต่อไปเข้ามาประมวลผล จากตัวอย่างประโยคที่ให้ ข้อผิดพลาดจะถูกตรวจพบที่คำบุพ บท "ใน" คำที่สอง ดั่งนั้นตัวแจะส่วนจะข้ามคำนี้ไปและดั่งค่านาม "สวน" เข้ามาประมวลผลต่อจนจบ

4.3.2 การแก้ไขข้อผิดพลาดที่เกิดจากการลบหรือการขาดองค์ประกอบใดๆ ในประโยค (Handling deleted element error) เช่น

ประโยคที่ถูกไวยากรณ์: ฉัน เห็น เขา ใน สวน

ประโยคที่ผิดพลาดไวยากรณ์: ฉัน เห็น เขา สวน

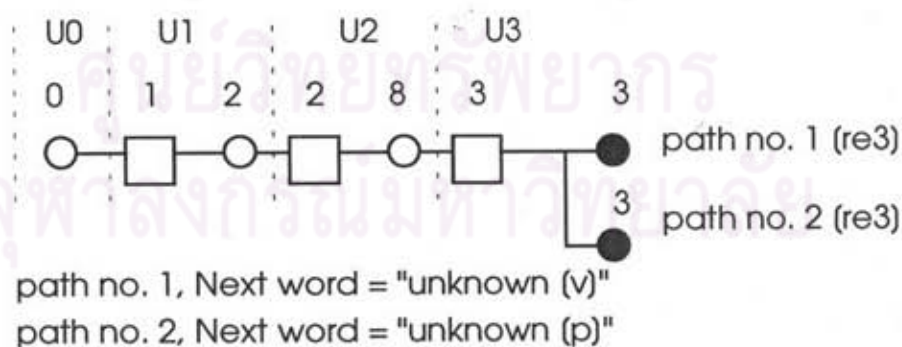
ประโยคที่แก้ไขแล้ว : ฉัน เห็น เขา [p] สวน

เมื่อ [p] คือคำบุพบทที่แทรกลงในประโยค โดยตัวแจะส่วน

ประโยคนี้การผิดพลาดไวยากรณ์จะถูกตรวจพบเมื่อค่านาม "สวน" ถูกประมวลผล ตัวแจะส่วนจะแก้ไขข้อผิดพลาดประเภทนี้ ด้วยการแทรกชนิดของคำที่เหมาะสมลงไปหน้าคำที่ตรวจพบข้อผิดพลาด ชนิดของคำที่จะนำมาแทรกในประโยคนั้นจะพิจารณาจากตารางแอลอาร์ โดยอาจจะมีชนิดคำที่เติมได้มากกว่า 1 ชนิดคำก็ได้ จากตัวอย่างที่กำหนดให้ข้อผิดพลาดจะถูกพบในขั้นตอนที่ 5 ดังรูปที่ 4.1

จากตัวอย่าง ชนิดของคำที่สามารถนำมาแทรกลงในประโยคได้ ณ ตำแหน่งนี้ คือคำบุพบท และคำกริยา แต่ถ้าเติมคำกริยาลงไปในกรณีนี้จะทำให้เกิดข้อผิดพลาดขึ้นมาอีก (ทำให้ในประโยคมีคำกริยา 2 ตัว ผิดจากไวยากรณ์ที่กำหนด) ดั่งนั้นการทำงานในขั้นตอนที่ 5 คือ

5) จากตารางแอลอาร์ สถานะพร้อมที่ 3 ชนิดของคำที่เป็นไปได้ในกรณีนี้คือ คำบุพบท กับคำกริยา ตัวแจะส่วนจะแยกโครงสร้างข้อมูลในขั้นนี้ออกเป็น 2 สายเพื่อแทรกชนิดคำที่เป็นไปได้จากตารางแอลอาร์ลงไป



รูปที่ 4.2 แสดงการแยกสายโครงสร้างในขั้นตอนที่ 5 เพื่อแทรกชนิดคำที่เหมาะสมลงในประโยค

หลังจากทำการแก้ไขประโยคแล้ว ประโยคที่สามารถแจງส่วนจนจบสมบูรณ์คือ ประโยคที่เติมคำบุพบท "to" ลงไปในประโยค ดังนั้นประโยคที่ผ่านการแก้ไขของตัวแจງส่วนแล้วคือ ประโยค "ฉัน เห็น เขา [to] ส่วน"

4.3.3 การแก้ไขข้อผิดพลาดที่เกิดจากการแทนที่องค์ประกอบใดๆ ในประโยค (Handling substituted element error) เช่น

ประโยคที่ถูกไวยากรณ์ : ฉัน เห็น เขา กับ เธอ

ประโยคที่ผิดไวยากรณ์ : ฉัน เห็น เขา กัด เธอ

ประโยคที่แก้ไขแล้ว : ฉัน เห็น เขา [to] เธอ

ในกรณีนี้ข้อผิดพลาดจะถูกตรวจพบที่คำกริยาตัวที่สอง "กัด" หลังจากตรวจพบข้อผิดพลาดแล้ว ตัวแจງส่วนจะทำงานคล้ายกับการแก้ไขข้อผิดพลาดในหัวข้อ 4.3.2 ต่างกันที่ในการแก้ไขข้อผิดพลาดประเภทนี้ตัวแจງส่วนจะเติมชนิดของคำที่เป็นไปได้ลงไปแทนที่คำ ณ ตำแหน่งที่ตรวจพบข้อผิดพลาดนั้น จากตัวอย่างจึงแทนที่ "กัด" ด้วยคำบุพบท "to" (เช่นเดียวกับข้อ 2 การแทนที่ด้วยคำกริยาจะทำให้เกิดข้อผิดพลาดอีก กรณีนี้จึงแทนที่ด้วยคำบุพบทอย่างเดียว)

การแก้ไขข้อผิดพลาดที่กล่าวมาทั้ง 3 วิธี นั้นเป็นพื้นฐานการทำงานเพื่อจัดการกับข้อผิดพลาดที่ได้จำแนกไว้ทั้ง 3 แบบ ปัญหาต่อมาที่เราพบคือเราจะทำการแก้ไขการผิดไวยากรณ์นี้ที่ตำแหน่งใดในประโยค โดยทั่วไปแล้วตัวแจງส่วนแบบจีแอลอาร์จะตรวจพบข้อผิดพลาดได้ในทันทีที่เป็นไปได้ แต่ในบางครั้งก็มีความเป็นไปได้เช่นกัน ที่ตำแหน่งที่มีการตรวจพบข้อผิดพลาดนั้น อาจจะไม่ใช่ตำแหน่งที่เป็นต้นเหตุของการผิดไวยากรณ์ที่แท้จริง ดังนั้นในการทำงานขั้นต่อไปเราจะจำแนกวิธีการแก้ไขประโยคที่ผิดรูปแบบไวยากรณ์ ออกตามประเภทของการกำหนดตำแหน่งเพื่อทำการแก้ไขประโยค ออกเป็น 3 หัวข้อด้วยกัน คือ 1. การแก้ไขข้อผิดพลาด ณ ตำแหน่งที่มีการตรวจพบข้อผิดพลาดเท่านั้น 2. การแก้ไขข้อผิดพลาดโดยใช้การแจງส่วนแบบจีแอลอาร์-ไอจีแอลอาร์ และ 3. การแก้ไขข้อผิดพลาดโดยใช้การแจງส่วนแบบย้อนรอยโครงสร้างกราฟ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย