

บทที่ 3

การวิเคราะห์โครงสร้างของประโยคโดยลิงค์แกรมมาฟาสเซอร์

เราสามารถอธิบายลิงค์แกรมมาได้โดยให้คำแต่ละคำในประโยคมีการเชื่อมต่อกันโดยลิงค์ที่มีหลากหลายชนิดตามหน้าที่ของคำในประโยคนั้น ลิงค์เกิดจากคอนเนคเตอร์ (connectors) ชนิดเดียวกันของคำสองคำทำให้สามารถเชื่อมคำสองคำด้วยกันได้ ดังนั้นคอนเนคเตอร์ของคำแต่ละคำจึงทำให้เกิดลิงค์ชนิดต่าง ๆ ที่ทำหน้าที่แตกต่างกันออกไปในแต่ละประโยค ซึ่งชนิดของลิงค์ที่แตกต่างกันนั้นจะมีผลต่อการวิเคราะห์หัวลีในประโยค เนื่องจากลิงค์แต่ละชนิดที่เชื่อมต่อกันนั้นจะทำให้เกิดวลีชนิดต่าง ๆ ขึ้น ซึ่งวลีที่ได้จากลิงค์นั้นจะเป็นประโยชน์ในการคัดแยกข้อความต่อไปเพื่อให้ได้ข้อความที่มีความหมายมากขึ้น สำหรับคอนเนคเตอร์แต่ละตัวนั้นจำเป็นจะต้องมีการระบุทิศทางซึ่งอาจชี้ไปทางซ้ายหรือชี้ไปทางขวาของคำ คอนเนคเตอร์ที่ชี้ไปทางซ้ายจะเชื่อมต่อกับคอนเนคเตอร์ชนิดเดียวกันของคำอีกคำหนึ่งที่ชี้ไปทางขวาทำให้เกิดลิงค์ขึ้น โดยลิงค์นั้นจะทำหน้าที่ในการแสดงความสัมพันธ์ระหว่างคำแต่ละคู่ในประโยคตามความสัมพันธ์ที่มีต่อกัน คอนเนคเตอร์ที่ชี้ไปทางขวาจะแสดงด้วยเครื่องหมาย “+” และคอนเนคเตอร์ที่ชี้ไปทางซ้ายจะแสดงด้วยเครื่องหมาย “-” คำแต่ละคำจะมีกฎเกี่ยวกับคอนเนคเตอร์ของคำนั้นโดยเฉพาะ

รายละเอียดเกี่ยวกับกฎของคำที่ทำให้เกิดลิงค์และขั้นตอนในการวิเคราะห์โครงสร้างของประโยคโดยลิงค์แกรมมาฟาสเซอร์นั้นจะกล่าวดังต่อไปนี้

3.1 กฎของคำ ยกตัวอย่างคำที่มีในคิกชันนารี

blah: A+;

หมายความว่า ถ้ามีคำว่า “blah” ปรากฏอยู่ในประโยค จะต้องมีการเชื่อมลิงค์ “A” ไปยังอีกคำในประโยค ซึ่งคำนั้นจะต้องมีคอนเนคเตอร์เป็น “A-” เครื่องหมายที่ตามหลังเครื่องหมายทวิภาค (Colon) “:” จะเรียกว่า “link requirement” ของคำนั้น คำหนึ่งคำอาจมีคอนเนคเตอร์ได้มากกว่าหนึ่งตัว เช่น

blah: A+ & B-;

คำนี้จะสามารถเชื่อมต่อกับคำอื่นได้โดยคอนเนคเตอร์ทั้งสองตัวตามกฎ โดยคอนเนคเตอร์ “A” จะเชื่อมไปทางขวาและคอนเนคเตอร์ “B” เชื่อมไปยังทางซ้าย

blah: A+ or B-;

หมายความว่า คำนี้จะต้องการเชื่อมคอนเนคเตอร์ “A” ไปทางขวา หรือคอนเนคเตอร์ “B” ไปทางซ้าย อย่างใดอย่างหนึ่ง

กฎเหล่านี้สามารถรวมเข้าด้วยกันได้ ยกตัวอย่างเช่น

blah: A+ or (B- & C+);

หมายความว่า คำนี้สามารถจะเชื่อมต่อกับคอนเนคเตอร์ “A” ที่เชื่อมไปทางขวา หรือคอนเนคเตอร์ “B” ที่เชื่อมไปทางซ้ายและคอนเนคเตอร์ “C” ที่เชื่อมไปทางขวา กฎเหล่านี้สามารถเขียนได้อย่างเป็นระเบียบและไม่มีขีดจำกัด เช่น

blah: (A+ or B-) & ((C+ & A+ & (D- or E-)) or F+);

สำหรับบางคอนเนคเตอร์ซึ่งเป็นทางเลือกจะถูกกำกับไว้ด้วยเครื่องหมายปีกกา เช่น

blah: A+ & {B+};

หมายความว่า คำนี้จะต้องมีคอนเนคเตอร์ “A” เชื่อมไปทางขวา และสามารถมีคอนเนคเตอร์ “B” เชื่อมไปทางขวาหรือไม่ก็ได้ สำหรับเครื่องหมายปีกกานั้นจะใช้เมื่อกฎของคำนั้นมีความซับซ้อน เช่น

blah: (A+ or B-) & {C+ & (D+ or E-)} & {@F+};

การเขียนด้วยเครื่องหมายปีกกา เช่น “{X-}” จะมีความหมายเหมือนกับ “(X- or ()”

คำที่มีลักษณะเป็น “multi-connector” จะแสดงด้วยสัญลักษณ์ “@” ซึ่งหมายความว่าคำนั้น จะไม่มีการจำกัดจำนวนของคอนเนคเตอร์ที่เป็นชนิดเดียวกัน ยกตัวอย่างเช่น

blah: (A+ or B-) & {C+ & (D+ or E-)} & {@F+};

หมายความว่า คำนี้มีคอนเนคเตอร์ชนิด “F” เชื่อมไปทางขวามากกว่าหนึ่งตัว

สำหรับคอนเนคเตอร์ที่มีทิศทางเดียวกันและเชื่อมด้วยสัญลักษณ์ “&” นั้น ลำดับที่มีใน “connector expression” นั้นมีความสำคัญ เนื่องจากลำดับที่มีจะเป็นตัวกำหนดคำที่มีความสัมพันธ์กันที่อยู่ใกล้ที่สุดที่เชื่อมกันได้ ตัวอย่างเช่น

blah: A+ & B+;

หมายความว่า “blah” จะมีคอนเนคเตอร์ “A” เชื่อมไปทางขวา และคอนเนคเตอร์ “B” เชื่อมไปทางขวา โดยที่คำที่เชื่อมด้วยคอนเนคเตอร์ “A” จะต้องอยู่ใกล้ “blah” มากกว่าคำที่เชื่อมด้วยคอนเนคเตอร์ “B” แต่สำหรับคอนเนคเตอร์ที่มีทิศทางตรงข้ามกันนั้นลำดับจะไม่มีผลสำคัญ ดังนั้น

blah: A+ & B-; จะมีความหมายเหมือนกับ blah: B- & A+;

ดิคชันนารีประกอบไปด้วยคำศัพท์ตามด้วยเครื่องหมายทวิภาค “:” ตามด้วยคอนเนคเตอร์ และสุดท้ายตามด้วยเครื่องหมายอัฒภาค (semicolon) “;” ในดิคชันนารีจะประกอบไปด้วยลำดับของคำศัพท์จำนวนหลายคำในรายการ โดยถูกแยกจากกันด้วยการเว้นวรรค ยกตัวอย่างเช่น

blah blee blay: A+;

3.2 กฎทั่วไป

เป็นกฎที่ใช้ในดิคชันนารีเพื่อใช้ในการควบคุมการเชื่อมกันของคำในประโยค ประกอบไปด้วยกฎดังนี้

1. ลิงค์ทั้งหมดไม่สามารถไขว้กันได้ (crossing-links or planarity rule)
2. คำทั้งหมดในประโยคจะต้องมีการเชื่อมกับคำอื่น (connectivity rule)
3. ลิงค์จะต้องเพียงพอที่จะเชื่อมโยงคำทั้งหมดตามลำดับอย่างสัมพันธ์กัน

3.3 ระบบลิงค์แกรมมา

โครงสร้างของประโยคที่ได้จากลิงค์แกรมมานั้นจะแสดงความสัมพันธ์ของคำแต่ละคู่ ตัวอย่างเช่น ประโยค “the black dog has gone” ได้ผลลัพธ์จากลิงค์แกรมมา แสดงดังรูป 3.1



รูป 3.1 แสดงผลลัพธ์ที่ได้จากลิงค์แกรมมา

อธิบายได้ว่า “S” (subject) แสดงความสัมพันธ์ระหว่าง “dog” และ “has” คำที่อยู่ซ้ายสุดของลิงค์ “S” จะเป็นประธานของอนุประโยคนั้น และทางขวาสุดของลิงค์จะเป็นคำกริยา

“PP” (past-participle) แสดงความสัมพันธ์ระหว่าง “has” และ “gone”

“D” (determiner) แสดงความสัมพันธ์ระหว่าง “the” และ “dog” คำที่อยู่ซ้ายสุดของลิงค์ “D” จะเป็นคำบ่งชี้ซึ่งใช้นำหน้าคำนาม

คำนาม คำกริยาและคำคุณศัพท์ทั้งหมดที่มีในดิคชันนารีจะมีตัวอักษรพิมพ์เล็กแสดงหน้าที่ของคำปรากฏอยู่ด้วย เช่น “.n” “.v” หรือ “.a”

3.3.1 คำอธิบายหน้าที่ของคำศัพท์ (Word Subscripts) คำศัพท์หนึ่งคำอาจมีได้หลาย

ความหมาย ดังนั้นในดิคชันนารีจึงแสดงหน้าที่ของคำตามหลังคำศัพท์นั้นด้วยตัวอักษรพิมพ์เล็ก ยกตัวอย่างเช่น

run.n: A+ or B+...

run.v: C+ or D+...

หน้าที่หลักของคำที่แสดงด้วยตัวอักษรนั้นมีดังนี้

“.n” หมายถึง คำนาม (noun)

“v” หมายถึง คำกริยา (verb)

“a” หมายถึง คำคุณศัพท์ (adjective)

“e” หมายถึง คำวิเศษณ์ (adverb)

“p” หมายถึง คำบุพบท (preposition)

“s” หมายถึง คำเอกพจน์ (single)

“p” หมายถึง คำพหูพจน์ (plural)

“t” หมายถึง คำนำหน้าชื่อ (title)

3.3.2 ค่าของระบบ (The cost system) พาสเซอร์จะมีการประเมินค่าของลิงค์ที่พบในแต่ละประโยค โดยจะแสดงค่านั้นอยู่ในเครื่องหมาย “[]” พาสเซอร์จะใช้ค่านี้เป็นเกณฑ์ในการเลือกลิงค์ที่จะแสดงเป็นผลลัพธ์ที่ได้ โดยเลือกจากลิงค์ที่มีค่าต่ำที่สุดเป็นอันดับแรก ซึ่งปกติจะมีค่าเป็น 0 1 หรือ 2 สำหรับค่าที่ประเมินได้จะแสดงดังนี้

Unique linkage, cost vector = (UNUSED=1 DIS=0 AND=0 LEN=1)

“DIS” แสดงค่าของคอนเนกเตอร์ หรือค่าที่แสดงความไม่ต่อเนื่องกันของลิงค์ (disjunct cost)

“AND” แสดงค่าความต่างระหว่าง and-list element

“LEN” แสดงค่าความยาวของลิงค์ทั้งหมดในประโยค

“UNUSED” แสดงจำนวนของ null-links

3.4 ลักษณะพิเศษในดิกชันนารี (Special Features of the Dictionary)

3.4.1 Capitalization เป็นหมวดหนึ่งในดิกชันนารี เรียกว่า “CAPITALIZED_WORDS” ทำหน้าที่ในการเก็บคำศัพท์ที่ขึ้นต้นด้วยตัวอักษรพิมพ์ใหญ่ ถ้าในประโยคนั้นขึ้นต้นด้วยตัวอักษรพิมพ์ใหญ่ พาสเซอร์จะทำการค้นหาเข้าไปในหมวดนี้ก่อน ถ้าไม่มีจะทำการเพิ่มคำดังกล่าวเข้าไปในหมวด “CAPITALIZED_WORDS” ถ้าคำนั้นมีปรากฏอยู่ทั้งสองหมวด คือหมวดอักษรพิมพ์ใหญ่ และหมวดทั่วไป พาสเซอร์จะใช้ทั้งสองหมวดเพราะหมายถึงว่าคำนั้นเป็นคำศัพท์ธรรมดาทั่วไป

3.4.2 Hyphenated Expression เป็นหมวดในดิกชันนารี เรียกว่า

“HYPHENTED_WORDS” ถ้าในสายอักขระ (string) นั้นมีเครื่องหมายยัติภังค์ (Hyphenated)

ปรากฏอยู่และ ไม่มีค่านั้นอยู่ในดิกชันนารี พาสเซอร์จะทำการเพิ่มค่านั้นเข้าไปในหมวด

“HYPHENTED_WORDS”

3.4.3 Unknown word เป็นหมวดที่เกี่ยวกับคำศัพท์ที่พาสเซอร์ไม่รู้จัก ซึ่งคำดังกล่าวอาจแสดงหน้าที่ของคำแตกต่างกันออกไปได้ อาจเป็นคำนาม คำกริยา คำคุณศัพท์หรือคำวิเศษณ์ ซึ่งโดยทั่วไปแล้วพาสเซอร์จะเดาว่าเป็นคำนาม แล้วแสดงหน้าที่ด้วยอักษร “n” ในการเดานั้นจะมี

พื้นฐานอยู่บนการสะกดของคำนั้น เช่น คำที่ลงท้ายด้วย“-s” จะเดาว่าเป็นคำพหูพจน์หรือคำกริยาของคำนามเอกพจน์ ซึ่งจะถูกเก็บอยู่ในหมวดของ “S-WORDS” ในดิกชันนารีต่อไป

3.4.4 Punctuation พาสเซอร์สามารถจัดการกับเครื่องหมายวรรคตอนต่าง ๆ ได้ โดยเครื่องหมายเหล่านั้นจะถูกเก็บไว้ในดิกชันนารีเช่นเดียวกันกับคำศัพท์แต่จะต้องอยู่ในเครื่องหมายอัญประกาศ ยกตัวอย่างเช่น

“;”: A+ or B-

“+”: C+ or D-

3.4.5 The sorting of unknown strings พาสเซอร์จำเป็นต้องมีการตัดสินใจในกรณีที่กำลังคำไม่มีปรากฏในดิกชันนารีนั้น ดังนั้นจึงต้องมีกระบวนการในการเลือกคำที่เหมาะสมที่สุดสำหรับประโยคนั้น ซึ่งพาสเซอร์จะดำเนินการดังนี้

1. ถ้าคำนั้นเป็นคำแรกในประโยคและตัวอักษรตัวแรกนั้นเป็นตัวอักษรพิมพ์ใหญ่แล้วให้สร้างตัวอักษรตัวแรกเป็นตัวพิมพ์เล็ก และปฏิบัติตามขั้นตอนต่อไปทั้งในกรณีที่เป็นอักษรพิมพ์ใหญ่และพิมพ์เล็ก
2. ทำการตัดตัวอักษรที่สามารถตัดออกได้จากทางซ้ายสุดของสายอักขระและทำตามกระบวนการต่อไปทั้งกับตัวอักษรที่ตัดออกและสายอักขระที่เหลือ
3. ทำการค้นหาคำนั้นในดิกชันนารี ถ้าตรงกับคำที่มีก็ให้เลือกคำนั้น
4. ทำการตัดเครื่องหมายวรรคตอนต่าง ๆ ทางด้านขวาของสายอักขระออก และทำตามข้อที่ 3 ทั้งในกรณีของตัวอักษรที่เป็นเครื่องหมายวรรคตอนและตัวอักษรที่เหลือ ถ้ายังไม่พบให้ทำตามข้อที่ 3 และข้อที่ 4 ซ้ำอีก
5. ถ้าสายอักขระนั้นขึ้นต้นด้วยตัวพิมพ์ใหญ่ให้ระบุลงในหมวด
“CAPITALIZED_WORDS”
6. ถ้าสายอักขระนั้นมีเครื่องหมายยัติภังค์ปรากฏอยู่ด้วยให้ระบุลงในหมวด
“HYPHENTED_WORDS”
7. ถ้าสายอักขระนั้นมีจำนวนตัวเลขใด ๆ ปรากฏอยู่ให้ระบุลงในหมวด “NUMBERS”
8. ถ้าสายอักขระนั้นยังไม่สามารถระบุลงในหมวดใด ๆ ได้ ให้ค้นหาในหมวด
“GUESSED-WORD” และระบุลงในหมวดนี้
9. ถ้าสายอักขระนั้นยังไม่สามารถระบุลงในหมวดใด ๆ ให้ระบุลงในหมวด
“UNKNOWN-WORD”
10. ถ้าทำตามกระบวนการดังกล่าวแล้วก็ยังไม่ได้ข้อสรุป พาสเซอร์จะสรุปว่าคำดังกล่าวไม่มีอยู่ในดิกชันนารี และไม่แสดงผลลัพธ์อื่น ๆ อีกสำหรับประโยคนั้น

เมื่อกระทำตามกระบวนการดังกล่าวแล้วพาสเซอร์พบว่ามีความเป็นไปได้ในหลายทางที่จะพิจารณาทุกกรณีที่จะเป็นไปได้ แต่ต้องอยู่ภายใต้เงื่อนไขดังนี้

- 1) ถ้ารูปแบบของคำที่เกิดขึ้นนั้นทุกกรณีมีปรากฏอยู่ในดิคชันนารีด้วยหน้าที่ของคำที่แตกต่างกันออกไป
- 2) ถ้าอักษรตัวแรกของสายอักขระนั้นเป็นอักษรตัวพิมพ์ใหญ่และมีอยู่ในดิคชันนารีทั้งตัวอักษรที่เป็นพิมพ์ใหญ่และอักษรพิมพ์เล็ก
- 3) ถ้ามีหมวด UNKNOWN-WORD มากกว่าหนึ่ง

3.4.6 The wall(s) ในกรณีที่คำศัพท์เริ่มต้นของทุกประโยคเป็นคำศัพท์ที่ไม่รู้จัก จึงจำเป็นต้องมี “the wall” ขึ้นโดยมีลิงค์กำหนดเช่นเดียวกับคำศัพท์ทั่วไปที่เก็บอยู่ในดิคชันนารี แบ่งเป็น “LEFT-WALL” จะมีการแทรกที่จุดเริ่มต้นของประโยค โดยเชื่อมต่อกับประโยคด้วยคอนเนคเตอร์ “Wd” และ “RIGHT-WALL” เชื่อมต่อกับ “LEFT-WALL” ด้วยคอนเนคเตอร์ “RW” เพื่อให้เป็นไปตามกฎของการเชื่อมต่อของคำในประโยค

3.4.7 ส่วนวน สำหรับคำที่มีการรวมกันเพื่อเกิดเป็นส่วนวนต่าง ๆ นั้น สามารถเก็บอยู่ในดิคชันนารีเหมือนเป็นคำศัพท์คำหนึ่งได้ โดยเชื่อมคำเหล่านั้นเข้าด้วยกันโดยใช้เครื่องหมาย “_” คำที่มีลักษณะเป็นส่วนวนนั้นจะไม่มีกระบวนการหน้าทีของคำในดิคชันนารี ยกตัวอย่างเช่น

a_la_mode: A+ or B-;

3.5 คำเชื่อม

สำหรับคำเชื่อมที่จะกล่าวถึงในส่วนนี้จะเจาะจงไปที่คำเหล่านี้ “and” “or” “but” “either-or” “neither-nor” “both-and” และ “not only-but”

3.5.1 การจัดการกับคำเชื่อม กำหนดให้ S คือ คำที่เรียงต่อเนื่องกัน (sequence of words) และมีคำว่า “and” ปรากฏอยู่

“well-formed ‘and’ list” L คือ ส่วนย่อยของคำที่เรียงต่อกัน (subsequence of words) ที่มีเงื่อนไขดังนี้

1. สามารถแบ่ง L ออกเป็นส่วนประกอบย่อย (element) ได้ องค์ประกอบย่อยแต่ละส่วนจะถูกแยกออกจากองค์ประกอบย่อยที่อยู่ติดกันโดยเครื่องหมายจุลภาค หรือ “and” (หรือทั้งสอง) ซึ่งทั้งเครื่องหมายจุลภาคและ “and” นั้น ไม่ได้เป็นส่วนหนึ่งขององค์ประกอบนั้น องค์ประกอบย่อยคู่สุดท้ายจะถูกแยกออกจากกันโดย “and” หรือเครื่องหมายจุลภาค (หรือเครื่องหมายจุลภาคและตามด้วย “and”)

ยกตัวอย่างเช่น ประโยค “The dog, cat, and mouse ran” องค์ประกอบย่อยที่แบ่งได้จากประโยคดังกล่าวคือ “dog”, “cat”, and “mouse” ซึ่งเป็นองค์ประกอบย่อยของ “dog, cat, and mouse”

2. คำที่เรียงต่อเนื่องกันที่ได้จากการแทนที่ L ด้วยองค์ประกอบย่อยหนึ่งของ L คือประโยคของลิงก์แกรมมา
3. มีการเลือกลิงก์สำหรับแต่ละประโยคนั้นยกเว้น “and” ถ้าตัดลิงก์ที่เชื่อมต่่องค์ประกอบย่อยกับส่วนที่เหลือของประโยค ขัวยของค์ประกอบย่อยนั้นออกจากประโยคแล้วแทนที่ด้วยองค์ประกอบย่อยอื่น ถ้าลิงก์นั้นยังสามารถเชื่อมต่อกับองค์ประกอบย่อยนั้นได้อย่างถูกต้อง จึงสรุปได้ว่าประโยคนั้นมีการสร้างลิงก์ที่ถูกต้อง

ยกตัวอย่างเช่น พิจารณาประโยค “We ate popcorn and watched movies on TV for three days.” พบวลีที่เป็น “well-formed ‘and’ list” คือ “ate popcorn and watched movies on TV” สามารถแบ่งเป็นองค์ประกอบย่อยได้ 2 ส่วนคือ “ate popcorn” และ “watched movies on TV” องค์ประกอบย่อยทั้งสองส่วนนี้จะเชื่อมต่อกับประโยคที่เหลือด้วยคอนเนคเตอร์ “S” ที่เชื่อมไปทางซ้าย และ “MV” เชื่อมไปทางขวา ซึ่งปัญหาที่พบคือ ไม่มีลิงก์ที่เชื่อมจาก “and” ไปยังคำอื่น ๆ ของประโยค ดังนั้นการพาสเซอร์จึงเกิดความกำกวมต่อประโยคนั้น จึงได้มีการแก้ไขปัญหาดังกล่าวด้วยวิธีแรกคือ ทำการจำกัดชนิดของคอนเนคเตอร์ที่สามารถเชื่อมต่่องค์ประกอบของ “and” ไปยังส่วนที่เหลือของประโยค ซึ่งชนิดของคอนเนคเตอร์ดังกล่าวนี้จะถูกเก็บอยู่ใน “ANDABLE-CONNECTORS” ในดิกชันนารี และวิธีที่สองคือ ทำการจำกัดนิยามของ “well-formed ‘and’ list” ให้เป็นไปตามเงื่อนไขที่ว่า แต่ละองค์ประกอบย่อยจะต้องเชื่อมกับส่วนที่เหลือของประโยคด้วยคอนเนคเตอร์หนึ่งของประโยคอย่างถูกต้อง

3.5.2 ตัวห้อย (Subscripts) ในกรณีที่มีการเชื่อมคำด้วยการใช้คำสันธานนั้นทำให้เกิดการเชื่อมคำตั้งแต่สองคำขึ้นไปที่มีคอนเนคเตอร์ที่มีตัวห้อยแตกต่างกัน ดังนั้นจึงจะต้องเลือกคอนเนคเตอร์ที่สามารถเชื่อมได้กับทุกคำ ยกตัวอย่าง พิจารณาคำในดิกชันนารีดังนี้

a: Ds+;

the: D+;

those: Dm+;

cats dogs: Dm-;

cat dog: Ds-;

เมื่อในประโยคปรากฏวลี “cats and dog” ซึ่งมีคอนเนคเตอร์เป็น “Dm-” และ “Ds-” ตามลำดับ

ดังนั้นคำนำหน้าคำนามที่จะสามารถใช้ได้กับทั้งสองคำคือ “the” เพราะมีคอนเนคเตอร์เป็น “D+” ซึ่งสามารถเชื่อมกับคอนเนคเตอร์ “Dm-” และ “Ds-” ได้

3.6 ปัญหาบางประการที่พบ พิจารณาประโยคดังนี้

“I gave Bob a doll and Mary a gun”

“This is a problem Moscow created and failed to solve”

พิจารณาประโยคแรก พบว่าวลี “I gave Bob a doll” นั้น “gave” จะถูกลิงค์ไปยัง “Bob” และ “doll” ซึ่ง “Bob a doll” นั้นไม่ได้เป็นองค์ประกอบของ “and” list รูปแบบของประโยคแรกจึงถูกละทิ้งไป สำหรับในประโยคที่สองนั้น “Moscow” จะต้องเชื่อมต่อกับ “failed” และ “problem” จะต้องเชื่อมต่อกับ “solve” ดังนั้น “problems to solve” จึงไม่ได้เป็นองค์ประกอบของ “and” list ซึ่งยังพบปัญหานี้อยู่ในระบบ

สำหรับปัญหาอื่น ๆ ในระบบนั้นเกิดขึ้นจากอนุประโยคที่อยู่ในประโยค พิจารณาลิงค์ที่ได้จากประโยค ยกตัวอย่างเช่นประโยค

“I think John and Dave ran” จะมีลิงค์ที่เกิดจากการรวมกันของสองประโยคด้วยกันคือ “I think John ran” และ “Dave ran” ซึ่งเกิดปัญหากับลิงค์ที่ได้เนื่องจากลิงค์ “S” ที่ได้ระหว่าง “and” และ “ran” นั้นจะเหมือนกับที่เกิดระหว่าง “I” และ “think” และลิงค์ “S” เชื่อม “and” และ “ran” น่าจะต้องเป็นลิงค์ที่เกิดระหว่าง “Dave” และ “ran” จึงจะถูกต้อง ต่อมาจึงมีการแก้ปัญหาดังกล่าวด้วยกระบวนการที่เรียกว่า “post-processing” ซึ่งจะอธิบายในส่วนต่อไป จากกระบวนการที่ได้กล่าวไปแล้วข้างต้นนั้นเป็นการจัดการกับคำสันธานโดยการแยกประโยคเดิมออกเป็นประโยคย่อย แล้วประมวลผลหาลิงค์ประกอบของกลุ่มคำที่ได้จากลิงค์ของแต่ละประโยค สุดท้ายลิงค์ที่ได้จากต้นฉบับที่เชื่อว่าไม่ถูกต้องนั้น ถ้าองค์ประกอบของลิงค์คู่่นั้นมีการซ้อนทับกันจากลิงค์ที่เป็นชนิดเดียวกัน โดยที่ไม่ใช่กลุ่มคำกลุ่มเดียวกัน เช่น ลิงค์ “S” ที่พบในสองประโยคข้างต้น

3.7 กระบวนการหลังการประมวลผล (Post-Processing)

3.7.1 ตรรกะของกระบวนการหลังการประมวลผล (The logic of post-processing) ในการพัฒนาระบบกระบวนการหลังการประมวลผลนั้นก็เพื่อแก้ปัญหาในข้อ 3.6 ซึ่งมีพื้นฐานอยู่ภายใต้แนวความคิดที่เรียกว่า “domains” โดยที่โดเมนนั้นจะเป็นสับเซตของลิงค์ที่ทำให้เกิดประโยคขึ้น หลังจากพบความสัมพันธ์ของลิงค์แล้ว กระบวนการหลังการประมวลผลจะทำการท่องไปในลิงค์ที่พบแล้วทำการแบ่งประโยคออกเป็นโดเมนต่าง ๆ ซึ่งก็ใช้ชนิดของลิงค์ที่พบในประโยคเป็นพื้นฐาน

ในการจัดการดังกล่าว จากนั้นจึงทำการแบ่งลิงค์ออกเป็นกลุ่ม ซึ่งกลุ่มนี้จะเป็นเซตของลิงค์ที่เป็นสมาชิกโดเมนร่วมกัน

3.7.2 โครงสร้างของโดเมน โดเมนจะเริ่มต้นด้วยชนิดของลิงค์ที่เรียกว่า “root link” ของโดเมน คำที่อยู่ทางซ้ายสุดของลิงค์จะเรียกว่า “root word” เมื่อลิงค์เริ่มต้นมีชนิดที่แตกต่างกันชนิดของโดเมนที่ได้ก็จะแตกต่างกันเช่นกัน โดเมนของลิงค์จะประกอบไปด้วยลิงค์ทั้งหมดที่สามารถท่องไปถึงได้จากลิงค์นั้นโดยไม่ผ่าน “root word” หรืออาจกล่าวได้ว่าโดเมนประกอบไปด้วยลิงค์ที่สามารถไปถึงได้ทั้งหมดจากขวาสุดของ “root link” โดยไม่ผ่าน “root link” เอง ยกตัวอย่าง สมมติให้ลิงค์ “C” เป็นจุดเริ่มของโดเมนในประโยค ดังนั้นโดเมนนี้จะประกอบด้วยลิงค์ “Sp” ลิงค์ “I” แต่ไม่รวมลิงค์ “Ss” และ “O”

-----C-----			
---Ss---	---O---	---Sp---	---I---
He	told	me	they would go

3.7.3 กลุ่ม (Groups) กลุ่มของลิงค์ คือเซตของลิงค์ที่มีสมาชิกในโดเมนเหมือนกัน กลุ่มที่ได้นี้จะมัลักษณะคล้ายกับเป็นการแสดงประธานและกริยาของประโยค กลุ่มของลิงค์จึงเป็นส่วนหนึ่งของอนุประโยค แต่ไม่ใช่ส่วนหนึ่งของอนุประโยคที่มีความอิสระต่อกันในอนุประโยคนั้น

3.7.4 กฎของกระบวนการหลังการประมวลผล จากโครงสร้างของโดเมนที่ได้นั้น ทำให้สามารถแบ่งประโยคออกเป็นกลุ่มของลิงค์ได้ ดังนั้นจึงสามารถทำให้กลุ่มของลิงค์ชนิดต่าง ๆ เป็นการกำหนดหน้าที่ในการแสดงประธานและกริยาของประโยค ซึ่งในกรณีนี้จะเป็นประโยชน์เมื่อมีข้อกำหนดในการรวมกันของลิงค์ที่สามารถแสดงได้ในอนุประโยคนั้น แต่ว่าลิงค์อาจจะถูกแยกได้อีกจึงเป็นการยากที่จะใช้ข้อกำหนดเดิมดังกล่าวกับกรณีเช่นนี้ จึงสร้างกฎขึ้นมาใช้เรียกว่า “contains_one” ซึ่งกำหนดว่า ถ้ากลุ่มนั้นมีลิงค์ที่เป็นชนิดที่แน่นอนแล้ว จำเป็นจะต้องมีลิงค์ชนิดอื่นอยู่ในกลุ่มนั้นด้วย และอีกหนึ่งกฎเรียกว่า “contains_none” ซึ่งกำหนดว่าถ้ากลุ่มนั้นมีลิงค์ชนิดใดชนิดหนึ่งอยู่แล้ว อาจจะไม่ต้องมีลิงค์ที่มีชนิดของลิงค์ที่แน่นอนอีกก็ได้

ในแต่ละกรณีนั้น จะต้องมีการ “triggering link-type” ซึ่งเป็นลิงค์ที่นำไปสู่การใช้กฎใดกฎหนึ่ง และจะต้องมี “criterion link-type” เป็นการกำหนดนิยามของลิงค์ ซึ่งเขียนอยู่ในรูปแบบดังนี้

X, YZ, “Warning!”;

เมื่อ “X” เป็น “triggering link-type”, “Y” และ “Z” เป็น “criterion link-type” และ “Warning!” เป็นข้อความที่จะแสดงเมื่อถูกถูกฝ่าฝืน

ถ้าข้อกำหนดดังกล่าวอยู่นอกภายใต้กฎ “contains_one” จะหมายความว่า กลุ่มที่มีคอนเนคเตอร์ X จะต้องมียอนเนคเตอร์ Y หรือไม่ก็คอนเนคเตอร์ Z ด้วย แต่ถ้าข้อกำหนดดังกล่าวอยู่ภายใต้

กฎ “contains_none” จะหมายความว่า กลุ่มที่มีคอนเนคเตอร์ X อาจจะไม่ต้องมีคอนเนคเตอร์ Y หรือคอนเนคเตอร์ Z ก็ได้

นอกจากนี้ยังสามารถนิยามเครื่องหมายเหมือนเป็นเซตของลิงค์ได้เช่นกัน ยกตัวอย่างเช่น

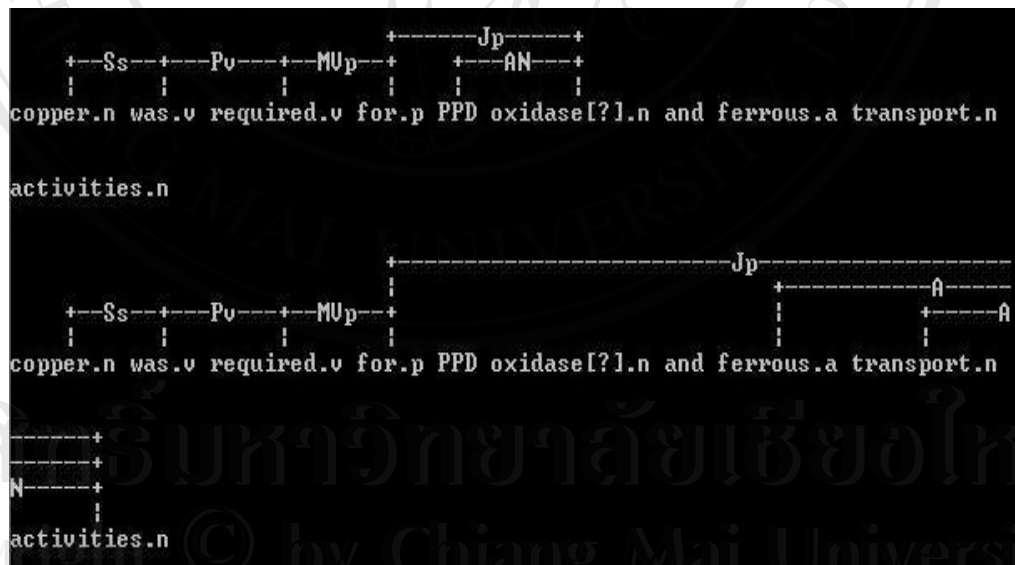
@Blah: Y Z

ซึ่ง “@Blah” หมายความว่าถึงเซตของลิงค์ {Y, Z} ดังนั้นเมื่อเขียนว่า X, Blah, “Warning!”, ก็จะมี ความหมายเช่นเดียวกันกับ X, YZ, “Warning!”

3.8 การวิเคราะห์ห่วงโซ่ในประโยค

พาสเชอร์นั้นเป็นส่วนประกอบหนึ่งของลิงค์แกรมมาพาสเชอร์ โดยนำเอาลิงค์ที่มีการเชื่อมกันที่ได้จากกระบวนการของลิงค์แกรมมาทำการสร้างโครงสร้างของประโยคซึ่งอาจประกอบไปด้วยวลีต่าง ๆ เช่น นามวลี กริยาวลี และบุพบทวลี เป็นต้น

ยกตัวอย่างเช่น ประโยค “Copper was required for the PPD oxidase and ferrous transport activities” ผลลัพธ์ที่ได้จากลิงค์แกรมมาเป็นดังรูปที่ 3.2



รูป 3.2 แสดงผลลัพธ์ที่ได้จากลิงค์แกรมมา

Ss หมายถึง ลิงค์ที่เชื่อมระหว่างคำนามที่ทำหน้าที่เป็นประธานและคำกริยา

Pv หมายถึง ลิงค์ที่เชื่อมระหว่างคำกริยาช่วยและคำกริยาหลัก

MVp หมายถึง ลิงค์ที่เชื่อมระหว่างคำกริยาและคำบุพบท

Jp หมายถึง ลิงค์ที่เชื่อมระหว่างคำบุพบทและคำนามที่ทำหน้าที่เป็นกรรมของคำบุพบทนั้น

A หมายถึง ลิงค์ที่เชื่อมระหว่างคำคุณศัพท์ที่ทำหน้าที่ขยายคำนามและคำนาม

AN หมายถึง ลิงค์ที่เชื่อมระหว่างคำนามที่ทำหน้าที่ขยายคำนามและคำนาม

จากนั้นพิจารณาลิงค์ที่มีการเชื่อมกัน ซึ่งพาสเซอร์จะทำการวิเคราะห์ลิงค์แต่ละส่วนตามกฎโครงสร้างของโดเมนดังที่ได้กล่าวไว้แล้วในข้อ 3.7.2 และจะอธิบายหลักการในการวิเคราะห์หัวลิในประโยคในหัวข้อถัดไป

3.9 หลักการในการวิเคราะห์หัวลิในประโยค

กระบวนการในการสร้างโครงสร้างของประโยคจากผลลัพธ์ที่ได้จากลิงค์แกรมมานั้น ประกอบไปด้วยหลายขั้นตอนด้วยกัน ขั้นตอนแรกนั้น โปรแกรมจะต้องสร้างเซตของวลีต่างๆ สำหรับแต่ละสับลิงค์ (sublinkage) ของประโยค แล้วทำการรวมเซตเหล่านั้นเข้าด้วยกัน ก่อนที่จะมีการวิเคราะห์โครงสร้างของประโยคจะต้องผ่านกระบวนการที่เรียกว่า “post-processes” ก่อน ซึ่งได้อธิบายไปแล้วในหัวข้อข้างต้น ในขั้นตอนที่สองจะทำการสร้างส่วนที่เรียกว่า “domain structure” ขึ้น โดยโครงสร้างดังกล่าวนี้จะได้จากการแบ่งประโยคออกเป็นกลุ่มของคำที่มีความสัมพันธ์กับวลีต่างๆ

เมื่อโครงสร้างของโดเมนดังกล่าวได้ถูกสร้างขึ้นแล้ว พาสเซอร์จะทำการทอ้งผ่านไปยังรายการของโดเมนที่ได้จึงทำให้เกิดวลีสำหรับแต่ละโดเมน ซึ่งในบางกรณีก็สามารถสร้างได้สองวลีต่อหนึ่งโดเมนก็ได้เช่นกัน บางกรณีลิงค์เริ่มต้นของโดเมนก็สามารถทำให้เกิดวลีได้ โดยทั่วไปขอบเขตของวลีจะเริ่มที่คำทางซ้ายสุดและสิ้นสุดที่คำทางขวาสุดในโดเมน ผลลัพธ์จากพาสเซอร์จะแสดงองค์ประกอบต่างๆ ของประโยคดังรูปที่ 3.3

```
<S <NP Copper>
  <UP was
    <UP required
      <PP for
        <NP <NP the PPD oxidase>
          and
            <NP ferrous transport activities>>>>>>>
```

รูป 3.3 แสดงผลลัพธ์ที่ได้จากลิงค์แกรมมาพาสเซอร์

ตารางต่อไปนี้จะแสดงรายการของวลีที่ถูกสร้างขึ้นหลังจากมีโครงสร้างของโดเมนเกิดขึ้นแล้ว โดยคอลัมน์แรกแสดงชนิดของโดเมน (Domain) คอลัมน์ที่ 2 แสดงชนิดของลิงค์เริ่มต้น (Starting links)

คอลัมน์ที่ 3 แสดงวลี (Constituent) ที่เกิดขึ้นและคอลัมน์สุดท้ายแสดงคำอธิบาย (Comment)

ตารางที่ 3.1 แสดงรายการของวลีที่ถูกสร้างขึ้น

Domain	Starting links	Constituent	Comment
P	MVp, Mp, MVt, MX#, MG, OF, Pp	PP	บุพบทวลี เช่น “The dog ran [PP in the park]”
V	S (ยกเว้น S##d), Pg, Pv, I, PP, PF, SF, SX, Mv, Mg	VP	กริยาวลี เช่น “The dog [VP will [VP ran in the park]]”
S	Wd, Wc, Ce, Cs, Rn, Cc, Cr, Cs, Cpi, CP*, Zc, Bc, Coq, CX, B#w, B#d	S	อนุประโยค เช่น “[S He said [S he left when [S he saw the dog [S I bought]]]]”
-	Mj	WHPP, WHNP	สำหรับความสัมพันธ์ที่เป็นบุพบท คำสองคำที่เป็นส่วนประกอบของ WHPP จะประกอบด้วย คำบุพบท และสรรพนาม สำหรับคำที่เป็นส่วนประกอบของ WHNP จะมีแค่ คำสรรพนามหนึ่งคำเท่านั้น เช่น “The park [WHPP in [WHNP which]] we played was big”
-	Ss#d, B#d	WHNP, SBAR	ประโยคที่ปรากฏคำ “what” ซึ่งเป็น “WHNP” และอนุประโยคทั้งหมดเป็น “SBAR” เช่น “[SBAR[WHNP what] goes up] must come down”
a	Pa, Ma, MX#a, L	ADJP	คุณศัพท์วลี เช่น “He is [ADJP happy]”
u	A	ADJP	คุณศัพท์วลีที่แสดงคุณลักษณะ เช่น “The [ADJP big] dog ran”

ตารางที่ 3.1(ต่อ) แสดงรายการของวลีที่ถูกสร้างขึ้น

Domain	Starting links	Constituent	Comment
Z	S##d, Wi, S##w, RS, MVg, Mgp, MX#p, PF	VP, S	ในกรณีนี้ ทั้ง S และ VP จะถูกสร้างขึ้น ซึ่งเป็นไปตามกฎเกณฑ์ของรูปแบบของคำกริยาที่เติม -ing หรือ -ed ที่ตามหลังคำบุพบท จะต้องเป็นอนุประโยค เช่น “He escaped by [S[VP digging a tunnel]]”
C	Cop	VP, S	ใช้สำหรับวลีที่เป็นรูปแบบของคำกริยาที่เติม -ing หรือ -ed แบบเปิด เช่น “[S[VP Carrying the dog], he left”
B	TH, R*, MVs, QI, MX*r, MVh, MX#d, Mr, Mj	SBAR	SBAR ถูกสร้างขึ้นได้ในหลายกรณี เช่น อนุประโยคที่อยู่ในประโยคด้วยคำว่า “that” ประโยคคำถาม หรือ อนุประโยคที่มีการให้คำเชื่อม เช่น “He said [SBAR that he was coming]”
F	CO#s	SBAR	SBAR ที่อยู่ทางซ้าย ซึ่งมักจะพบในประโยคเปิด เช่น “[SBAR When he arrived], I left”
-	Ce, Rn	SBAR	Rn ใช้ในประโยคที่มีอนุประโยคที่ทำหน้าที่ขยายนามโดยไม่ใช่คำบุพบท และ Ce* ใช้ในประโยคที่มีอนุประโยคฝังอยู่โดยไม่มี “that” เช่น “The dog [SBAR[S we saw]] was black”

ตารางที่ 3.1(ต่อ) แสดงรายการของวลีที่ถูกสร้างขึ้น

Domain	Starting links	Constituent	Comment
N	O, J (ยกเว้น Jw), SI, OX, MX#*, TI, BIt, IN, ON, JG, U, JT, OD	NP	นามวลี ที่มีหน้าที่เป็นกรรมของประโยค กรรมของบุพบท และ วลีที่ประธานถูกสลับตำแหน่ง เช่น “I saw [NP the dog]”
Y	YS, YP, Yt, Yd, GN, OD	NP	นามวลีที่อยู่ทางซ้ายสุดของวลีที่มีความสัมพันธ์ที่แสดงความเป็นเจ้าของ เช่น “[NP The dog]’s nose was black”
-	R*, MX*r	WHNP	คำเดี่ยวที่เป็นคำสรรพนาม เช่น “The dog [WHNP who] chased me was big”
E	E, EA	ADVP	กริยาวิเศษณ์วลี คำขยายหน้าคำกริยา เช่น “He [quickly] ran”
I	Mva, MVb	ADVP	กริยาวิเศษณ์วลี คำขยายหลังคำกริยา เช่น “He ran [ADVP quickly]”
Q	CPx	SINV	อนุประโยคที่มีการสลับตำแหน่งของประธานและกริยา ใช้ในการถ่ายทอดข้อความ เช่น “He ran, [SINV said Joe]”
G	CO***, CO#n, COd#	PP	คำบุพบทที่อยู่ทางด้านซ้าย เช่น “[PP On Tuesday], we saw him”
K	K	PRT	คำศัพท์เล็กน้อย เช่น “He came [PRT in]”
D	D##n, ND	QP	ตัวเลขที่อยู่ทางกึ่งด้านซ้าย เช่น “I have [QP 500] dogs”

ตารางที่ 3.1 (ต่อ) แสดงรายการของวลีที่ถูกสร้างขึ้น

Domain	Starting links	Constituent	Comment
H	NIax	QP	ตัวเลขที่อยู่ทางกึ่งด้านขวา (ส่วนใหญ่ใช้กับ “\$”) เช่น “I have \$ [QP 5]”
-	QI, Mr, MX#d	WHADVP, WHNP	ในกรณีที่มีการใช้คำแสดงคำถาม (question word) กับนามวลี เช่น “I wonder [WHNP who] he saw” หรือในกรณีที่คำถามนั้นทำหน้าที่เป็นคำนำหน้านาม เช่น “I wonder [WHNP which movies] he saw” หรือในกรณีที่มิอนุประโยคต่อท้าย อนุประโยคนั้นจะถูกจัดกลุ่มด้วย S เช่น “I wonder [WHNP which movies] [S he saw]”
-	CP	S	สำหรับประโยคที่มีการถ่ายทอดข้อความ “S” จะต้องครอบคลุมประโยคทั้งหมด เช่น “[S John ran, he said]”
-	MVs	WHADVP	ในกรณีที่มีการใช้ “when” เป็นคำเหมือนคำสันธาน เช่น “I ran [WHADVP when] I saw him”
T	TO, MVi	S, VP	รูปกริยาที่เป็นนามที่ขึ้นต้นด้วย “to” เช่น “I intend [S[VP to go]]”

โครงสร้างของประโยคที่สามารถวิเคราะห์ได้จากลิงก์แกรมมาฟาสเซอร์นั้นจะถูกจัดให้เป็นโครงสร้างแบบต้นไม้ ซึ่งแต่ละโหนดจะประกอบไปด้วยตัวชี้ที่ชี้ไปโหนดถัดไปและตัวชี้ที่ชี้ไปที่โหนดลูก ซึ่งจะกล่าวถึงในบทต่อไป

3.10 องค์ประกอบของส่วนที่ทำให้สมบูรณ์ (Complement Constituents)

เมื่อองค์ประกอบถูกสร้างขึ้น ทำให้เกิดองค์ประกอบหลักและองค์ประกอบที่เป็นส่วนขยายต่าง ๆ หลายชนิดแตกต่างกันออกไป ในบางกรณีที่องค์ประกอบหลักของประโยคนั้นประกอบด้วยทุกส่วนในองค์ประกอบ X ที่มีขนาดใหญ่กว่าองค์ประกอบอื่น ซึ่งถือเป็นส่วนหลักของประโยค เช่น เป็นประธานของประโยค แต่ไม่อยู่ในส่วนองค์ประกอบ Y ซึ่งเป็นส่วนขยาย ยกตัวอย่างเช่น อนุประโยคซึ่งมี NP ที่ประกอบด้วยประธาน และทุกส่วนอยู่ในองค์ประกอบ S แต่ไม่อยู่ในองค์ประกอบ VP เช่น [S The dog [VP ran in the park]] ---> [S[NP The dog] [VP ran in the park]]

ดังนั้นจึงมีหลายกรณีที่จำเป็นจะต้องมีส่วนที่เรียกว่า “complement constituents” ซึ่งมีรายการดังนี้

1. สำหรับทุก ๆ อนุประโยคเปิด (SBAR) ในองค์ประกอบ S จะต้องเพิ่ม complement S เช่น “[S[SBAR When I saw him], [S he ran]]”
2. สำหรับทุก ๆ PP ใน S จะต้องเพิ่ม complement S เช่น “[S[PP On Tuesday], [S we saw him]]”
3. สำหรับทุก ๆ participle (S) ใน S จะต้องเพิ่ม complement S เช่น “[S[S Carrying the dog], [S he left]]”
4. สำหรับทุก ๆ VP (เริ่มต้นด้วยลิงค์ S) ใน S จะต้องเพิ่ม NP สำหรับวลีที่เป็นประธาน เช่น “[S[NP The dog][VP ran]]”
5. สำหรับทุก ๆ relative clause (SBAR) ใน NP จะต้องเพิ่ม NP เช่น “[NP[NP The dog] [SBAR we saw]] ran”
6. สำหรับ participle modifiers (VP) ใน NP จะต้องเพิ่ม NP เช่น “[NP[NP The dog][VP chasing the cat]] was black”
7. สำหรับทุก ๆ PP ใน NP จะต้องเพิ่ม NP เช่น “[NP[NP The dog][PP in the park]] was big”
8. สำหรับทุก ๆ NP ที่ทำหน้าที่ขยายความใน NP จะต้องเพิ่ม NP เช่น “[NP[NP The dog], [NP a terrier]], was black”
9. สำหรับทุก ๆ NP ในอนุประโยคที่มีการสลับตำแหน่งของประธานและกิริยา(SINV) จะต้องเพิ่ม VP เช่น “He ran, [S[VP said][NP John]]”

3.11 เครื่องหมายจุลภาค (Commas)

เครื่องหมายจุลภาคที่ใช้ยังเป็นปัญหอยู่บ้าง เครื่องหมายจุลภาคที่คั่นอยู่ระหว่างสององค์ประกอบที่ไม่ได้ถูกรวมอยู่ในองค์ประกอบใดองค์ประกอบหนึ่ง เช่น “[SBAR When we saw

him], [S we left]” จะไม่เกิดขึ้นในกรณีปกติ ซึ่งจากประโยคตัวอย่างดังกล่าวเครื่องหมายจุลภาคนั้นจะเป็นส่วนหนึ่งของโดเมน f ดังนั้นเครื่องหมาย “;” จึงเป็นส่วนหนึ่งในอนุประโยคเปิด ดังได้แสดงต่อไปนี “[SBAR When we saw him,] [S we left]” จึงมีการย้ายขอบเขตของ SBAR ใหม่โดยการแยกเครื่องหมายจุลภาคออกทำให้ได้ผลลัพธ์เป็นดังตัวอย่างแรกข้างต้น

อย่างไรก็ตาม เมื่อมีองค์ประกอบที่ทำให้สมบูรณ์ (complement) S เกิดขึ้น จะต้องรวมเครื่องหมายจุลภาคเข้าไปด้วย (เมื่อองค์ประกอบที่ทำให้สมบูรณ์ นั้นประกอบไปด้วยทุกส่วนที่ไม่ได้อยู่ใน SBAR) ดังนั้นจึงไม่มีการพิจารณาขอบเขตทางขวาของ SBAR จนกระทั่งหลังจากได้มีการสร้าง S ขึ้นแล้วจึงจะทำการปรับเอาเครื่องหมายจุลภาคออกจากขอบเขตดังกล่าว

3.12 การรวมองค์ประกอบต่าง ๆ (Merging the constituents)

เมื่อมีองค์ประกอบต่าง ๆ ของประโยคเกิดขึ้นแล้วจะต้องมีการรวมองค์ประกอบต่าง ๆ เข้าด้วยกัน ยกตัวอย่างเช่น ประโยค “The dog and the cat ran” ซึ่งสามารถแบ่งออกได้เป็นสองส่วนด้วยกันคือ “The dog ran” และ “The cat ran” ซึ่งมีโครงสร้างดังนี้

[S [NP The dog] [VP ran] S]

[S [NP The cat] [VP ran] S]

เราสามารถเริ่มต้นด้วยการรวมองค์ประกอบดังกล่าวเข้าด้วยกันอย่างง่าย ๆ โดยไม่มีหลักการจะได้ผลลัพธ์ดังนี้

[S [NP The dog NP] and [S [NP The cat NP] [VP [VP ran VP] VP] S] S]

การรวมกันด้วยวิธีข้างต้นจะทำให้เกิดปัญหามากมาย อันดับแรกจะทำให้เกิดองค์ประกอบต่าง ๆ มากเกินจำเป็น ซึ่งจากตัวอย่างข้างต้นจะต้องมี S และ VP เพียงอย่างละหนึ่งเท่านั้น สำหรับกรณีที่เป็น VP นั้นก็แก้ปัญหาง่าย ๆ ด้วยการลบหนึ่งในจำนวนนั้นออก โดยทั่วไปจะทำการลบองค์ประกอบที่ซ้ำและมีขอบเขตเหมือนกันออก แต่สำหรับกรณีที่เป็น S นั้นจะไม่สามารถลบออกได้ เนื่องจากขอบเขตไม่เหมือนกัน จะต้องทำการสืบค้นหา “andlist” ซึ่งจะเป็นรายการขององค์ประกอบที่เป็นชนิดเดียวกันและแต่ละองค์ประกอบนั้นจะต้องไม่อยู่สืบถึงชนิดเดียวกัน แล้วจะทำการรวมแต่ละองค์ประกอบเข้าด้วยกัน แล้วสร้างองค์ประกอบใหม่ขึ้นเป็นขอบเขตรอบ “andlist” ทั้งหมดที่พบด้วยองค์ประกอบชนิดเดียวกันกับองค์ประกอบที่อยู่ใน “andlist” นั้น ซึ่งจากประโยคตัวอย่างข้างต้น จะต้องมีการรวม NP ที่อยู่ใน “andlist” นั้นเข้าด้วยกันโดยการเพิ่ม NP เป็นขอบเขตของ “andlist” นั้นไว้ ได้ผลดังนี้

[S [NP [NP The dog NP] and [NP The cat NP] NP] [VP ran VP] S]

3.13 ประโยคที่พาสเซอร์ไม่สามารถวิเคราะห์โครงสร้างได้

การที่พาสเซอร์ไม่สามารถวิเคราะห์โครงสร้างของประโยคบางประโยคได้ อาจเกิดจากสาเหตุดังนี้

1. ประโยคบางประโยคที่มีลักษณะเป็นคำพูด เช่น “Hi, my name is John” หรือ “John laughed himself silly” พาสเซอร์จะไม่สามารถวิเคราะห์ได้ เนื่องจากระบบถูกสร้างมาเพื่อใช้กับประโยคที่มีลักษณะแบบบรรยาย
2. ตัวอักษรพิมพ์ใหญ่พิมพ์เล็กก็มีผลต่อการวิเคราะห์โดยพาสเซอร์เช่นกัน ยกตัวอย่างเช่น “I saw bob” จะไม่ถูกวิเคราะห์ เพราะว่า “bob” ที่ปรากฏในดิคชันนารีนั้นถูกกำหนดว่าเป็นคำกริยาดังนั้นประโยค “I saw Bob” เท่านั้นที่พาสเซอร์จะทำการวิเคราะห์ได้
3. เมื่อมีการใช้สำนวนใด ๆ ก็ตามที่ไม่มีระบุอยู่ในดิคชันนารีประโยคนั้นจะไม่ถูกวิเคราะห์ เช่น “Now is the time to come” ซึ่งไม่มีระบุในดิคชันนารีจึงไม่ถูกวิเคราะห์
4. สำหรับในบางประโยคที่มีการใช้คำเชื่อมพาสเซอร์จะไม่สามารถวิเคราะห์ได้ ซึ่งเป็นข้อจำกัดอย่างหนึ่งของพาสเซอร์ดังที่ได้กล่าวมาแล้วในหัวข้อ 3.5 แต่กรณีนี้ก็มีเกิดขึ้นไม่มากนัก
5. อาจเกิดจากข้อผิดพลาดบางประการที่เกี่ยวกับหลักไวยากรณ์ ซึ่งถือว่าเกิดขึ้นได้น้อยมาก

สำหรับในบทต่อไปจะกล่าวถึงวิธีการดำเนินการในการคัดแยกข้อความโดยนำเอาหลักการของลิงแกรมมาพาสเซอร์มาใช้ในการวิเคราะห์โครงสร้างของประโยคแล้วทำการคัดแยกข้อความตามโครงสร้างของประโยคนั้นอย่างไร ผลการทดลองที่ได้จากการนำหลักการนี้ไปใช้