

บทที่ 5

การประยุกต์ผลการวิเคราะห์กับเครื่องคอมพิวเตอร์

หลังจากได้ผลการวิเคราะห์ทางภาษาศาสตร์ในบทที่ 3 และ 4 ในบทนี้ จะเป็นการประยุกต์ผลการวิเคราะห์กับเครื่องคอมพิวเตอร์ แต่ผู้วิจัยไม่สามารถใช้เทคนิคการเรียนรู้ของเครื่อง (machine learning) ได้ตามที่คาดการณ์ไว้เนื่องจากปริมาณข้อมูลที่เป็นตัวอย่างของโครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์นั้นน้อยเกินไป กล่าวคือ

(1) จำนวนโครงสร้างที่พบมีเพียง 42 โครงสร้าง

(2) ปริมาณของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่ปรากฏในแต่ละโครงสร้างมีความเหลื่อมล้ำกัน ดังนี้

(2.1) มีเพียงโครงสร้างเดียวที่ปรากฏคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์มากถึง 536 คำ

(2.2) มีหลายโครงสร้างที่ปรากฏคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์เพียงคำเดียวหรือ 2 คำเท่านั้น ทำให้ไม่เพียงพอที่จะเป็นตัวอย่างให้เครื่องเรียนรู้ได้ เพราะคำอื่นที่อยู่นอกกลุ่มข้อมูลอาจมีโครงสร้างอย่างเดียวกันนี้ได้ แต่มีคุณสมบัติภายใน เช่น ชนิดของคำ หมวดหมู่ทางความหมาย เป็นต้น แตกต่างไปจากที่พบในคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ ด้วยลักษณะเช่นนี้ เครื่องคอมพิวเตอร์จะไม่สามารถพบคำดังกล่าวได้แม้จะมีโครงสร้างเดียวกัน เพราะไม่ตรงกับตัวอย่างที่เคยสอนเครื่องไว้ ดังนั้น หากมีปริมาณตัวอย่างไม่มากพอ จะทำให้ไม่สามารถสอนเครื่องให้เรียนรู้โครงสร้างหรือรูปแบบได้

ด้วยปริมาณของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่ปรากฏแตกต่างกันในแต่ละโครงสร้าง คือ มีเพียงโครงสร้างเดียวที่คำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ปรากฏสูงมาก หลายโครงสร้างมีคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ปรากฏประมาณ 10 กว่าคำ และยังมีอีกหลายโครงสร้างที่คำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ปรากฏเพียง 1 – 2 คำเท่านั้นซึ่งน้อยเกินไปที่จะ

เป็นตัวอย่างของคำที่มีโครงสร้างลักษณะดังกล่าวได้ จากความเหลื่อมล้ำของปริมาณคำนามแสดง
อุปกรณ์ด้านวิทยาศาสตร์ที่ปรากฏในแต่ละโครงสร้าง ทำให้ไม่เหมาะสมที่จะใช้เทคนิคการเรียนรู้
ของเครื่องเพื่อสอนให้เครื่องรู้จักคำจากตัวอย่างคำในแต่ละโครงสร้างที่มีปริมาณคำปรากฏมากน้อย
แตกต่างกัน

ดังนั้น ผู้วิจัยจึงเลือกใช้วิธีอื่นในการประยุกต์ผลการวิเคราะห์ คือ การใช้โปรแกรม Unixtex
ซึ่งเป็นโปรแกรมที่ใช้สำหรับประมวลผลคลังข้อมูล ดังจะกล่าวโดยละเอียดในหัวข้อ 5.1 ต่อไป

ผู้วิจัยได้นำโครงสร้างความสัมพันธ์ระหว่างส่วนประกอบในบทที่ 4 มาเขียนเป็นกราฟ
แล้วนำกราฟไปทดสอบกับคลังข้อมูลตัวอย่างเพื่อทดสอบการตรวจจับคำจากโครงสร้างที่กำหนด
ด้วยโปรแกรมประยุกต์ Unixtex โดยแบ่งเป็นหัวข้อต่าง ๆ ดังต่อไปนี้

5.1 โปรแกรม Unixtex

5.2 ขั้นตอนการใช้งาน โปรแกรม Unixtex

5.3 การกำกับชนิดของคำภาษาไทยที่ใช้ใน โปรแกรม Unixtex

5.4 การกำกับหมวดหมู่ตามความหมาย

5.5 กฎโครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์

5.6 การประเมินค่าที่ได้จากการทดสอบ

5.7 ผลการทดสอบ

5.8 สรุป

5.1 โปรแกรม Unitex

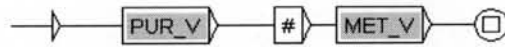
การทดสอบผลการวิเคราะห์กับเครื่องคอมพิวเตอร์เพื่อหาขอบเขตของคำนามแสดง
อุปกรณ์ด้านวิทยาศาสตร์เป็นการประยุกต์กับโปรแกรมที่มีชื่อว่า “Unitex”

โปรแกรม Unitex (Online, 2007) เป็นโปรแกรมประมวลผลคลังข้อมูลที่ใช้เทคโนโลยี
automata-oriented technology เป็นพื้นฐาน โปรแกรมนี้พัฒนาขึ้นเพื่อใช้วิเคราะห์ภาษาธรรมชาติ
องค์ประกอบสำคัญของการวิเคราะห์ด้วยโปรแกรม Unitex ประกอบด้วยพจนานุกรม
อเล็กทรอนิกส์ กฎไวยากรณ์ และตารางแสดงไวยากรณ์ของคำ ในเบื้องต้น Sébastien Paumier แห่ง
สถาบันวิจัย Gaspard – Monge (IGM) เป็นผู้พัฒนาโปรแกรมนี้เพื่อใช้วิเคราะห์ภาษาฝรั่งเศส ต่อมา
โปรแกรม Unitex ได้พัฒนามากขึ้นเพื่อใช้วิเคราะห์ภาษาอื่นๆ ที่เกี่ยวข้องกับเครือข่ายห้องทดลอง
RELEX

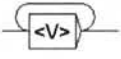
คุณสมบัติเด่นของพจนานุกรมอเล็กทรอนิกส์ที่ใช้ในโปรแกรมนี้ คือ สามารถระบุคำศัพท์
ที่เป็นคำศัพท์ทั่วไปและคำศัพท์ที่เป็นคำประสมพร้อมทั้งรากศัพท์และคำไวยากรณ์ (ทั้งทางด้าน
ความหมายและวิภคิตปัจจัย) คุณลักษณะเช่นนี้จะประ โยชน์ในการช่วยค้นหาและจับคู่
โครงสร้างทางภาษาที่ต้องการจากข้อมูล พจนานุกรมอเล็กทรอนิกส์นี้ยังสามารถอธิบายกลุ่ม
คำศัพท์ขนาดใหญ่ด้วยโครงสร้างง่ายๆ โดยจะแสดงรายการคำศัพท์ในเมนู DELA ต่อมา
นักภาษาศาสตร์หลายท่านได้ช่วยกันพัฒนาพจนานุกรมอเล็กทรอนิกส์นี้เพื่อให้วิเคราะห์ได้หลาย
ภาษา เช่น ภาษาฝรั่งเศส ภาษาอังกฤษ ภาษากรีก ภาษาอิตาเลียน ภาษาสเปน ภาษาเยอรมัน
ภาษาไทย ภาษาเกาหลี ภาษาโปแลนด์ ภาษานอร์เวย์ และภาษาโปรตุเกส เป็นต้น

กฎไวยากรณ์ในโปรแกรมนี้จะอยู่ในรูปกราฟ¹⁸ ที่ผู้ใช้สามารถสร้างและปรับปรุงแก้ไขได้
โดยง่าย กฎไวยากรณ์ดังกล่าวสร้างขึ้นจากกฎทางภาษาศาสตร์ที่ผู้ใช้วิเคราะห์จากข้อมูล
ตัวอย่างเช่น

¹⁸ ในทางคอมพิวเตอร์ กราฟ หมายถึง รูปภาพที่ใช้แทนโครงสร้างหรือความสัมพันธ์ต่าง ๆ ไม่จำเป็นต้องเป็น
แผนภูมิหรือกราฟในทางคณิตศาสตร์เท่านั้น

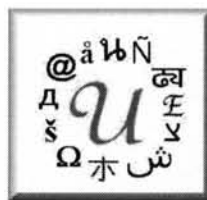


ภาพที่ 23 การสร้างกฎทางภาษาศาสตร์ที่ว่า “วัตถุประสงค์+วิธีการ (PUR+MET)” ให้เป็นกฎทางไวยากรณ์ในรูปของกราฟแสดงความสัมพันธ์ในโปรแกรม Unix

โปรแกรม Unix จะใช้กฎการเวียนเกิดซ้ำ (Recursive Transition Networks: RTN) เพื่อกำหนดให้เครื่องค้นหาส่วนใดส่วนหนึ่งในโครงสร้างซ้ำ ๆ จนพบสิ่งที่ผู้ใช้ต้องการ เช่น ผู้ใช้ต้องการให้เครื่องคอมพิวเตอร์ค้นหาคำกริยาแสดงอาการที่เรียงต่อกันในรูปของกริยาเรียงหลายตัวให้ครบก่อน ไปยังคำนามคำต่อไป เครื่องก็จะค้นหาคำกริยาแสดงอาการที่เรียงต่อกันจนครบและพบคำนามที่เกิดเรียงลำดับต่อไป เครื่องจึงจะหยุดกระบวนการเวียนเกิดซ้ำนี้ การเวียนเกิดซ้ำจะแทนด้วยเครื่องหมาย 

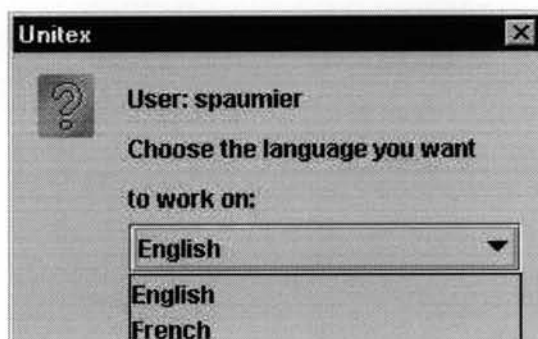
กฎไวยากรณ์ในโปรแกรมจะสิ้นสุดเมื่อพบเครื่องหมายวงกลมที่มีรูปสี่เหลี่ยมจัตุรัสซ้อนด้านใน  หรือเรียกว่า Final State อันเป็นเครื่องหมายที่แสดงการสิ้นสุดของโครงสร้าง

โปรแกรม Unix เป็นเครื่องมือที่ตรวจจับโครงสร้างของภาษาด้วยการสร้างเป็นกราฟ โปรแกรมนี้มีลักษณะโดดเด่นหลายประการ เช่น ใช้งานง่าย ใช้งานได้กับหลายภาษารวมถึงภาษาที่มีตัวอักษรลักษณะเฉพาะแบบ เช่น ภาษาอังกฤษ ภาษาฝรั่งเศส ภาษาไทย ภาษาเกาหลี ภาษาอาหรับ เป็นต้น นอกจากนี้ยังสามารถประยุกต์ใช้กับสำนวนและคำประสมได้ด้วย



ภาพที่ 24 สัญลักษณ์ของโปรแกรม Unix

เมื่อติดตั้งโปรแกรมในเครื่องคอมพิวเตอร์เรียบร้อยแล้ว หากจะเริ่มใช้งาน ให้คลิกที่สัญลักษณ์ของโปรแกรม แล้วโปรแกรมจะถามว่าต้องการดำเนินการโปรแกรมด้วยภาษาใด ดังนี้



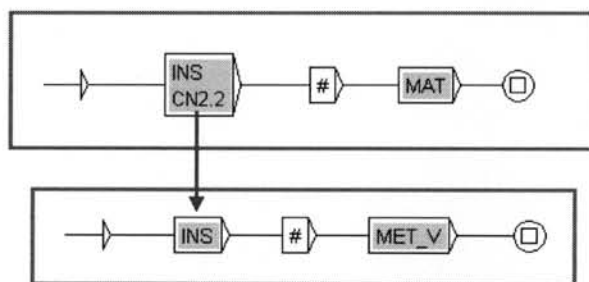
ภาพที่ 25 การเลือกภาษาสำหรับใช้งาน

เนื่องจากวิทยานิพนธ์เล่มนี้มีจุดมุ่งหมายทดสอบการบอกขอบเขตของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ในภาษาไทย ผู้วิจัยจึงเลือกภาษาที่ใช้กับโปรแกรมเป็นภาษาไทย จากนั้นจะพบกับหน้าต่างของโปรแกรม Unitex ซึ่งประกอบด้วยเมนูการใช้งาน 8 เมนู ได้แก่

เมนู Text เป็นเมนูที่สั่งให้โปรแกรมดำเนินการทดสอบและแสดงผลการทดสอบออกมาในรูปแบบของ concordance

เมนู DELA เป็นเมนูที่สร้างพจนานุกรมคำศัพท์ต่าง ๆ สำหรับโปรแกรม Unitex โดยมีพจนานุกรมประจำโปรแกรมในชื่อว่า DELA

เมนู FSGraph เป็นเมนูที่สร้างกราฟที่ได้จากกฎโครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่ได้วิเคราะห์ไว้ โดยกราฟนี้จะแสดงจุดเริ่มต้น ความสัมพันธ์ของส่วนต่าง ๆ ยกตัวอย่างเช่น



ภาพที่ 26 ตัวอย่างกราฟแสดงความสัมพันธ์แบบ “อุปกรณ์+วัสดุ”

ภาพที่ 26 ประกอบด้วยกราฟใหญ่และกราฟย่อย กราฟใหญ่แสดงความสัมพันธ์แบบ “อุปกรณ์+วัสดุ” ปรากฏในกรอบบนซึ่งอธิบายเส้นทางได้ว่า กราฟเริ่มต้นด้วยเครื่องหมายลูกศร ตามด้วยอุปกรณ์ซึ่งประกอบด้วย “อุปกรณ์” ที่แทนด้วย INS และคำประสมแบบ 2.2 (แทนด้วย CN2.2) ตามด้วยเงื่อนไขไม่ให้มีการเว้นวรรค (แทนด้วย #) ตามด้วย “วัสดุ” ที่แทนด้วย MAT และสิ้นสุดที่เครื่องหมายวงกลมที่มีรูปสี่เหลี่ยมจัตุรัสซ้อนด้านใน (final state)

สำหรับกราฟย่อยในกรอบล่าง กราฟดังกล่าวเป็นกราฟแสดงความสัมพันธ์แบบ “อุปกรณ์+วิธีการ” ของคำประสมแบบ 2.2 โดยเริ่มต้นจากเครื่องหมายลูกศร ตามด้วยอุปกรณ์ซึ่งแทนด้วย INS ตามด้วยเงื่อนไขไม่ให้มีการเว้นวรรค (แทนด้วย #) จากนั้นเป็นวิธีการซึ่งเป็นคำกริยาแทนที่ด้วย MET_V และสิ้นสุดที่เครื่องหมายวงกลมที่มีรูปสี่เหลี่ยมจัตุรัสซ้อนด้านใน (final state)

เมนู Lexicon-Grammar เป็นเมนูที่สามารถเรียกคุณสมบัติที่ผู้ใช้กำหนดให้แก่ส่วนต่าง ๆ นอกจากนี้ ยังใช้เมนูนี้ในการเรียกใช้คุณสมบัติที่กำหนดให้แก่ส่วนต่างๆ ข้างต้นกับกราฟที่สร้างขึ้นได้

เมนู Edit เป็นเมนูที่เปลี่ยนแปลงแก้ไขข้อมูลระหว่างที่ผู้ใช้จัดการกับข้อมูลบนหน้าต่างของโปรแกรมนี้

เมนู File Edition เป็นเมนูที่จัดการกับไฟล์ข้อมูล เช่น สร้างไฟล์ใหม่ เปิดไฟล์เดิม เป็นต้น

เมนู Windows เป็นเมนูที่จัดการกับหน้าต่างที่เปิดขึ้นมาภายใน โปรแกรมนี้ เช่น จัดเรียงหน้าต่างหลายๆ หน้าต่างให้เรียงต่อกันหรือซ้อนทับกัน เป็นต้น

เมนู Info เป็นเมนูที่ให้รายละเอียดเกี่ยวกับ โปรแกรม Unitex และคำอธิบายเมื่อผู้ใช้ต้องการความช่วยเหลือเกี่ยวกับ โปรแกรมนี้

หลังจากทราบรายละเอียดเกี่ยวกับเมนูการใช้งานของ โปรแกรม Unitex แล้ว จะเป็นขั้นตอนการใช้งานโปรแกรม Unitex เพื่อสร้างกราฟแสดงโครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์

5.2 ขั้นตอนการใช้งานโปรแกรม Unitex

เมื่อได้โครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์แล้ว ต้องนำโครงสร้างหลายระดับของคำนามดังกล่าวมาเขียนเป็นโครงสร้างที่แสดงคุณสมบัติของคำแสดงและองค์ประกอบบังคับเพื่อให้สามารถสร้างกราฟได้ ในที่นี้ ผู้วิจัยเลือกโครงสร้างที่แสดงความสัมพันธ์ทางความหมายเป็นโครงสร้างหลักของการสร้างกราฟเนื่องจาก

1. ความสัมพันธ์ทางความหมายเป็นส่วนสำคัญของการรวมส่วนประกอบต่าง ๆ ของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์เพื่อร่วมกันบ่งชี้ถึงอุปกรณ์
2. การใช้โครงสร้างที่แสดงความสัมพันธ์ทางความหมายระหว่างส่วนประกอบสามารถแสดงความสัมพันธ์ระหว่างส่วนประกอบได้ดีขึ้น ทำให้เข้าใจลักษณะการรวมความหมายได้
3. ความสัมพันธ์ทางความหมายแสดงลักษณะเด่นหรือลักษณะเฉพาะของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ได้





ด้วยเหตุผลดังกล่าว ผู้วิจัยจึงเลือกโครงสร้างที่แสดงความสัมพันธ์ทางความหมายเป็นตัวแทนในการสร้างและตั้งชื่อกราฟ โดยประกอบด้วยกราฟใหญ่และกราฟย่อย กราฟใหญ่เป็นการแสดงความสัมพันธ์ทางความหมาย ส่วนกราฟย่อยเป็นคุณสมบัติด้านอื่น ๆ ของคำแสดงและองค์ประกอบบังคับ ได้แก่ ชนิดของคำ และอาจมีโครงสร้างความสัมพันธ์ทางความหมายเกิดขึ้นภายในเป็นกราฟย่อยอีกเป็นลำดับชั้นได้ ขั้นตอนการใช้งานโปรแกรม Unitex เพื่อสร้างกราฟแสดงโครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์มีดังนี้

5.2.1 ขั้นตอนการสร้างกราฟ

หลังจากได้โครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์จากบทที่ 4 มาแล้ว โดยมีโครงสร้างหลักเป็นความสัมพันธ์ทางความหมายที่ประกอบด้วยคุณสมบัติต่าง ๆ ของส่วนประกอบ¹⁹ จึงนำโครงสร้างดังกล่าวมาเขียนเป็นกราฟ ในที่นี้ ผู้วิจัยตั้งชื่อกราฟตามโครงสร้างจากบทที่ 4 กราฟที่สร้างขึ้นใช้สัญลักษณ์ต่าง ๆ ดังนี้

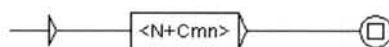
¹⁹ ดูรายละเอียดเพิ่มเติมได้ที่ภาคผนวก ข.

ตารางที่ 13 สัญลักษณ์ที่ใช้ในการเขียนกราฟ

สัญลักษณ์	ความหมาย
	จุดเริ่มต้นของกราฟ
—	การเชื่อมโยงความสัมพันธ์
	ส่วนประกอบของโครงสร้าง
	จุดสิ้นสุดของกราฟ
<>	การใส่ชนิดของคำ
+	การเพิ่มคุณสมบัติของคำ เช่น ประเภทย่อยของชนิดของคำ คำ AKO ฯลฯ
	การเวียนเกิดซ้ำ
“ ”	การเว้นวรรค
#	เงื่อนไขไม่ให้มีเว้นวรรค
[]	เงื่อนไขบริบทให้ตรวจจับสิ่งที่ปรากฏหน้าหรือหลังสิ่งที่อยู่ในเงื่อนไข

กราฟที่สร้างขึ้นแบ่งเป็น 2 ชนิด คือ

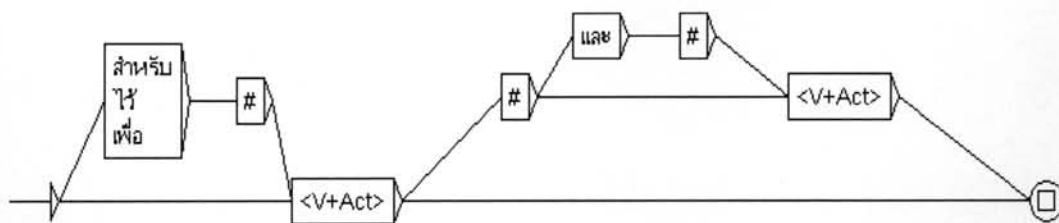
(1) กราฟย่อยเป็นกราฟที่แสดงคุณสมบัติของส่วนประกอบต่าง ๆ ในด้านชนิดของคำ ทั้งนี้ ชนิดของคำที่ใช้ในการทดสอบด้วยโปรแกรม Unitex เป็นชนิดของคำที่ได้รับการปรับปรุงมาจากชนิดของคำ 2 หมวดที่ใช้สำหรับการประมวลผลภาษาธรรมชาติ ได้แก่ ORCHID และ MT²⁰ หากมีโครงสร้างความสัมพันธ์ทางความหมายอื่นเกิดขึ้นจะเพิ่มเติมต่อไป ตัวอย่างเช่น



ภาพที่ 27 กราฟ [อุปกรณ์]

ภาพที่ 27 เป็นกราฟแสดงคุณสมบัติของส่วนประกอบ [อุปกรณ์] ซึ่งแทนชื่อกราฟว่า INS ประกอบด้วยคำนามทั่วไป (N+Cmn)

²⁰ ชนิดของคำที่ใช้ในการทดสอบครั้งนี้จะกล่าวถึงโดยละเอียดในหัวข้อ 5.3 เรื่องการกำกับชนิดของคำภาษาไทยที่ใช้ในโปรแกรม Unitex



ภาพที่ 28 กราฟ [วัตถุประสงค์]

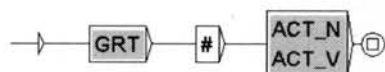
ภาพที่ 28 เป็นกราฟแสดงคุณสมบัติของส่วนประกอบ [วัตถุประสงค์] ซึ่งแทนชื่อกราฟว่า PUR_V ประกอบด้วยคำกริยาแสดงอาการ (V+Act) ที่อาจมีคำเชื่อม “สำหรับ” “ไว้” “เพื่อ” ปรากฏหน้าคำกริยาแสดงอาการ และอาจมีคำกริยาแสดงอาการคำที่ 2 ปรากฏหลังคำกริยาแสดงอาการคำแรกโดยมีคำเชื่อม “และ” ปรากฏระหว่างคำกริยาแสดงอาการทั้งสองหรือไม่ก็ได้

(2) กราฟหลักเป็นกราฟที่แสดงโครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ กราฟหลักประกอบด้วยกราฟย่อยต่าง ๆ ตามโครงสร้างที่ได้กำหนดไว้ ตัวอย่างเช่น



ภาพที่ 29 กราฟแสดงความสัมพันธ์แบบ [อุปกรณ์-วิธีการ]

ภาพที่ 29 เป็นกราฟหลักที่แสดงความสัมพันธ์แบบ [อุปกรณ์-วิธีการ] โดยเกิดจากกราฟย่อย 2 กราฟ คือ กราฟอุปกรณ์ (แทนด้วย INS) และกราฟวิธีการ (แทนด้วย MET_V)



ภาพที่ 30 กราฟแสดงความสัมพันธ์แบบ [ความยิ่งใหญ่ ความสำคัญ-การกระทำ]

ภาพที่ 30 เป็นกราฟหลักที่แสดงความสัมพันธ์แบบ [ความยิ่งใหญ่ ความสำคัญ-การกระทำ] โดยเกิดจากกราฟย่อย 3 กราฟ คือ กราฟความยิ่งใหญ่ ความสำคัญ (แทนด้วย GRT) และกราฟการกระทำซึ่งประกอบด้วยกราฟย่อย 2 กราฟ คือ กราฟการกระทำที่เป็นค่านาม (แทนด้วย ACT_N) และกราฟการกระทำที่เป็นคำกริยา (แทนด้วย ACT_V)

เมื่อสร้างกราฟแสดงโครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์เสร็จแล้ว ขั้นตอนต่อไปจะเป็นการสร้างพจนานุกรมให้กับโปรแกรมเพื่อให้โปรแกรมรู้ว่า คำศัพท์แต่ละคำเป็นคำชนิดใดซึ่งช่วยให้โปรแกรมสามารถอ่านกราฟแล้วตรวจจับคำที่มีโครงสร้างตรงตามที่กำหนดไว้

5.2.2 ขั้นตอนการสร้างพจนานุกรม

โดยปกติ โปรแกรม Unitex มีพจนานุกรมคำศัพท์ประจำโปรแกรมอยู่แล้วในชื่อว่า DELA ซึ่ง ดร.กฤษณ์ โกสวัสด์ (Krit Kosawat, 2003) กล่าวว่า พจนานุกรมภาษาไทยในโปรแกรม Unitex ได้แบ่งชนิดของคำตามแนวคิดของพจนานุกรมราชบัณฑิตยสถาน พุทธศักราช 2525 เป็นหลัก ซึ่งแบ่งคำเป็น 11 ชนิด ดังแสดงในตารางที่ 9 ดังนี้

ตารางที่ 14 ชนิดของคำในพจนานุกรมภาษาไทยประจำโปรแกรม Unitex

สัญลักษณ์	ชนิดของคำ	ตัวอย่าง
N	คำนาม	กตिका ธรรมบท
PRO	คำสรรพนาม	แก คุณ
V	คำกริยา	กระดิก เทิน
ADJV	คำวิเศษณ์ (คุณศัพท์หรือกริยาวิเศษณ์)	กตัญญู ธรรมดา
PREP	คำบุพบท	กระทั่ง ของ
CONJ	คำเชื่อมอนุพจน์	จึง ฉวย โดยที่
NIBAT	คำนิบาต	ก็ ก็ดี ก็ตาม
INTJ	คำอุทาน	ชะ ถึงว่า
EXP	สำนวน	ปลาช้องเดียวกัน น้ำสั่งฟ้า ปลาสั่งฝน
PFX	วิภัติ	กรรม กรรตุ โทร ²¹
SFX	ปัจจัย	กระเทียบ กระเฟียด

²¹ ปกติจะมีเครื่องหมายขั้วกำกับ (-) ปรากฏหน้าปัจจัยหรือหลังวิภัติเพื่อแสดงว่าเป็นคำเติมหน้าหรือคำเติมหลัง เช่น กรรม- กรรตุ- โทร- กระเทียบ- กระเฟียด- เป็นต้น แต่การใส่เครื่องหมายดังกล่าวจะทำให้โปรแกรม Unitex ไม่สามารถตัดคำได้ตรงตามชนิดของคำที่กำหนด เพราะเครื่องหมายดังกล่าวไม่ปรากฏในคำที่เดิมวิภัติหรือปัจจัยแล้ว เช่น กรรตุจาก โทรศัพท์ กระเทียบ เป็นต้น

อย่างไรก็ดี แม้โปรแกรม Unitex จะมีพจนานุกรมภาษาไทยประจำโปรแกรมอยู่แล้ว แต่การกำหนดชนิดของคำเป็น 11 ชนิดตามแนวคิดของพจนานุกรมราชบัณฑิตยสถาน พุทธศักราช 2525 ดังกล่าวไม่เอื้อต่อชนิดของคำที่ปรากฏในโครงสร้างความสัมพันธ์ของคำนาม แสดงอุปกรณ์ด้านวิทยาศาสตร์ เพราะไม่มีการกำหนดชนิดย่อยของคำ แต่ชนิดของคำที่ใช้กับคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์เป็นชนิดของคำที่ใช้สำหรับการประมวลผลภาษาธรรมชาติ ซึ่งมีการกำหนดรายละเอียดมากขึ้นเป็นชนิดย่อยของชนิดของคำเพิ่มเติมในเชิงหน้าที่และความหมายด้วย ตัวอย่างเช่น

กรณีคำนาม มีการกำหนดชนิดย่อยของคำนามออกเป็นหลายชนิด เช่น คำนามทั่วไป (แทนด้วย N+Cmn) คำนามชื่อเฉพาะ (แทนด้วย N+Prp) คำนามที่เป็นคำนำหน้าชื่อ (แทนด้วย N+Tle) คำนามที่แสดงจำนวนนับ (แทนด้วย N+Crd) ฯลฯ แต่ในพจนานุกรม DELA กำหนดเป็นเพียงคำนาม (แทนด้วย N)

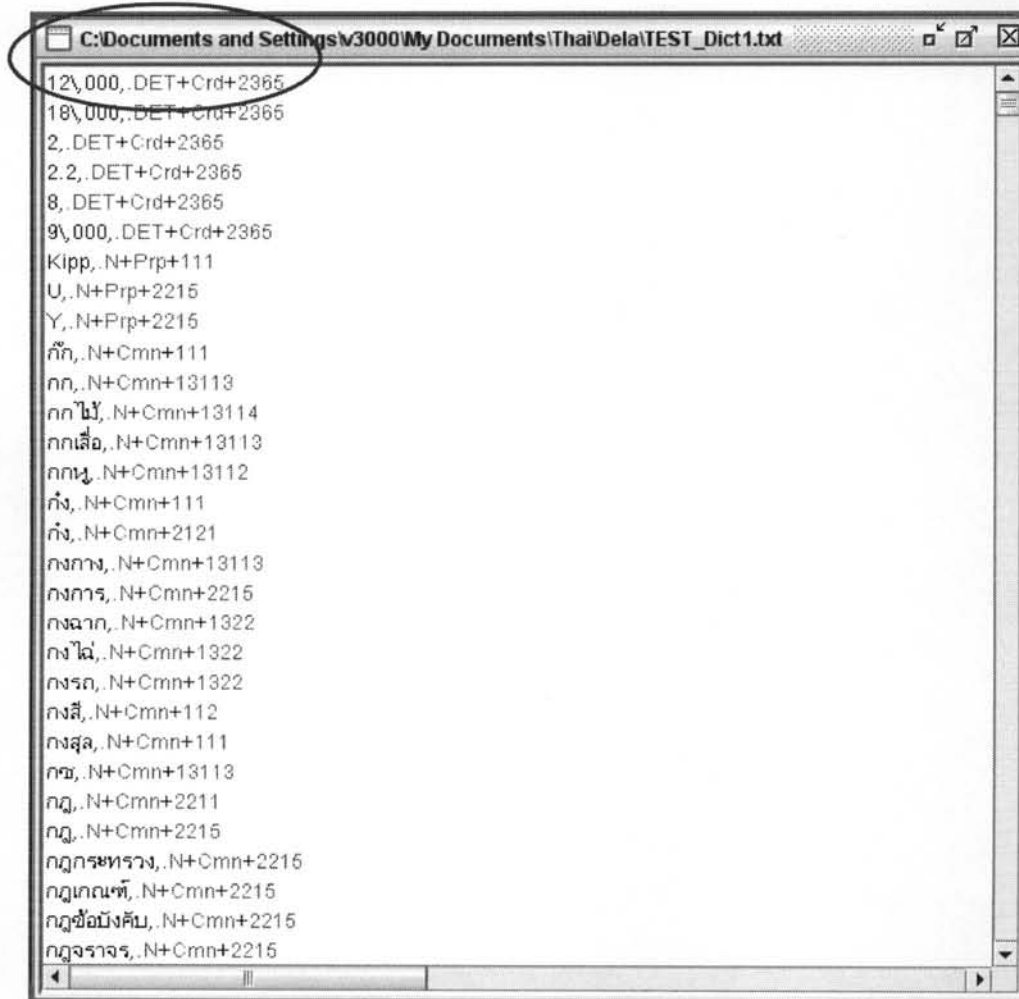
ด้วยสาเหตุดังกล่าว ผู้วิจัยจึงพัฒนาพจนานุกรมอิเล็กทรอนิกส์ขึ้นมาใหม่ เรียกว่า “TEST_DICT” พจนานุกรมอิเล็กทรอนิกส์ดังกล่าวสามารถกำกับชนิดของคำของคำศัพท์ในพจนานุกรมอิเล็กทรอนิกส์ตามแบบชนิดของคำที่ใช้สำหรับการประมวลผลภาษาธรรมชาติ ทั้งนี้พจนานุกรมที่นำมาปรับปรุงเป็นพจนานุกรมอิเล็กทรอนิกส์ภาษาไทยซึ่งได้รับการพัฒนาโดยคณะนักวิจัยของศูนย์วิจัยและพัฒนาปัญญาประดิษฐ์ สถาบันเทคโนโลยีพระจอมเกล้าธนบุรีในโครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์ระหว่าง NECTEC – CICC (นวลทิพย์ ดันดี เสวตรรัตน์ และคณะ, 2536: 77-84) ในที่นี้ใช้คำย่อว่า MT เพราะชนิดของคำที่ใช้ในพจนานุกรมนี้เป็นชนิดของคำที่ใช้สำหรับแปลภาษาด้วยคอมพิวเตอร์

ขั้นตอนการสร้างพจนานุกรมมีดังต่อไปนี้

(1) ปรับปรุงชนิดของคำที่ปรากฏในพจนานุกรมให้ตรงกับชนิดของคำที่ได้ปรับปรุงแล้วจากหมวดคำ 2 ชุดที่ใช้ในงานประมวลผลภาษาธรรมชาติ คือ ORCHID และ โครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์ระหว่าง NECTEC – CICC ดังปรากฏในหัวข้อ 5.3

(2) ตัดคำศัพท์ที่เป็นคำประสมและบ่งชี้อุปกรณ์ด้านวิทยาศาสตร์ออกจากรายการคำศัพท์ของพจนานุกรมเพื่อผลการทดสอบที่ถูกต้อง

(3) เพิ่มคำศัพท์ในกรณีที่ปรากฏคำนั้นในคลังข้อมูลแต่ไม่ปรากฏในพจนานุกรม หากไม่เพิ่มคำศัพท์ลงไปแล้ว จะทำให้เครื่องคอมพิวเตอร์ไม่สามารถตรวจจับคำที่มีโครงสร้างตามที่กำหนดได้ ทั้งนี้ การเพิ่มคำศัพท์ลงไปต้องทำให้ถูกต้องตามรูปแบบของพจนานุกรมซึ่งก็คือ คำศัพท์ ตามด้วยเครื่องหมายจุดภาคและมหัพภาค (,) และชนิดของคำ โดยห้ามเว้นวรรค ในกรณีที่ มีเครื่องหมายอื่นปรากฏในคำ เช่น เครื่องหมายจุดภาค (,) ใน 12,000 18,000 และ 9,000 เป็นต้น ให้ใส่เครื่องหมาย \ (back slash) หน้าเครื่องหมายดังกล่าว ดังภาพที่ 24



ภาพที่ 31 รูปแบบรายการคำที่ปรากฏในพจนานุกรมที่ปรับปรุงใหม่

5.2.3 ขั้นตอนการสร้างคลังข้อมูล

คลังข้อมูลที่ใช้ทดสอบครั้งนี้แบ่งเป็น 2 ลักษณะ คือ

(1) คลังข้อมูลที่เป็นรายการคำศัพท์ เป็นคลังข้อมูลรายการคำนามแสดงอุปกรณณ์ด้านวิทยาศาสตร์ที่รวบรวมจากพจนานุกรมที่นำมาใช้เป็นข้อมูล 4 เล่ม คลังข้อมูลนี้จะใช้สำหรับทดสอบความสามารถของโปรแกรมในการตรวจจับโครงสร้างทั้งหมดที่วิเคราะห์ได้จากคำนามแสดงอุปกรณณ์ด้านวิทยาศาสตร์ว่าสามารถตรวจจับได้ถูกต้องและแม่นยำเพียงใด

(2) คลังข้อมูลที่เป็นบทความ เป็นคลังข้อมูลที่รวบรวมจากบทความนิตยสารคลังข้อมูลนี้จะใช้สำหรับทดสอบความสามารถของโปรแกรมในการตรวจจับโครงสร้างทั้งหมดที่วิเคราะห์ได้ว่า โปรแกรมสามารถตรวจจับคำที่มีโครงสร้างดังกล่าวได้ถูกต้องและแม่นยำเพียงใด แม้ว่าคำเหล่านั้นจะปรากฏในประโยค แต่ไม่ได้ปรากฏเป็นรายการคำศัพท์ก็ตาม ทั้งนี้ ผู้วิจัยได้จัดทำรายการคำเหล่านั้นเพื่อตรวจสอบความถูกต้องของการตรวจจับคำด้วยโปรแกรม Unitex คำดังกล่าวแยกได้เป็น 2 กลุ่ม คือ คำนามแสดงอุปกรณณ์ด้านวิทยาศาสตร์และคำนามแสดงอุปกรณณ์ที่อยู่ในอกสาขาวิทยาศาสตร์

บทความที่นำมาเป็นคลังข้อมูลมาจากสิ่งตีพิมพ์และทางอินเทอร์เน็ต และแบ่งได้เป็น 4 สาขา ได้แก่

(2.1) บทความด้านวิทยาศาสตร์และเทคโนโลยีมาจากคอลัมน์ “สารคดีวิทยาศาสตร์” และ “มองไทย” นิตยสารอัปเดต (Update) เดือนมกราคม พฤษภาคม และกันยายน พ.ศ. 2547 และเดือนมกราคมถึงเดือนเมษายน พ.ศ. 2548

(2.2) บทความด้านวิทยาการทางการแพทย์และสุขภาพมาจากคอลัมน์ “โลกกว้างทางการแพทย์” “ตามประสาหมอ” และ “Drug & Supplement” นิตยสารใกล้หมอ เดือนกุมภาพันธ์ มิถุนายน และตุลาคม พ.ศ. 2547 และเดือนพฤษภาคม พ.ศ. 2548

(2.3) บทความด้านบ้านและการตกแต่งมาจาก 2 แหล่ง คือ

(2.3.1) จากสิ่งตีพิมพ์ คือ คอลัมน์ “ผลิตภัณฑ์” นิตยสารบ้านและสวน เดือนมีนาคม กรกฎาคม และพฤศจิกายน พ.ศ. 2547

(2.3.2) จากอินเทอร์เน็ต คือ คอลัมน์ “ผลิตภัณฑ์” นิตยสารบ้านและสวน ทางเว็บไซต์ <http://www.baanlaesuan.com> ซึ่งเคยตีพิมพ์ลงนิตยสารบ้านและสวนในเดือนมิถุนายน พ.ศ. 2550

(2.4) บทความด้านขนานยนต์ นิตยสารฟอร์มูลาซึ่งเผยแพร่ผ่านทางเว็บไซต์ <http://autoinfo.co.th/formula> ซึ่งเคยตีพิมพ์ลงนิตยสารฟอร์มูลาในเดือนมิถุนายน พ.ศ. 2547 เดือนธันวาคม พ.ศ. 2548 เดือนเมษายน พฤษภาคม และพฤศจิกายน พ.ศ. 2550

5.3 การกำกับชนิดของคำภาษาไทยที่ใช้ในโปรแกรม Unitex

เมื่อนำผลการวิเคราะห์ทางภาษาศาสตร์มาประยุกต์กับเครื่องคอมพิวเตอร์แล้ว ชนิดของคำแนวไวยากรณ์โครงสร้างที่ได้วิเคราะห์ไว้ไม่สามารถนำมาใช้กับโปรแกรม Unitex ได้ทันที เนื่องจากพจนานุกรมอิเล็กทรอนิกส์ที่เลือกใช้ใน โปรแกรมนี้เป็นพจนานุกรมที่กำหนดชนิดของคำสำหรับการประมวลผลภาษาธรรมชาติ จึงต้องมีการปรับปรุงชนิดของคำให้เข้ากับโปรแกรม

ทั้งนี้ ชนิดของคำภาษาไทยเป็นสิ่งที่ได้มีผู้เสนอแนวทางเอาไว้มากมาย ขึ้นอยู่กับมุมมองของนักไวยากรณ์แต่ละท่าน มีทั้งแนวภาษาศาสตร์และแนวการประมวลผลภาษาธรรมชาติ เช่น พระยาอุปกิตศิลปสาร (2546) บรรจบ พันธุเมธา (2522) กำชัย ทองหล่อ (2540) วิจิตร ภาณุพงศ์ (2530) นววรรณ พันธุเมธา (2549) อุดม วโรตมสถิตยัตต์ (2535) เรืองเดช ปิ่นเขื่อนขันธ์ (2541) วิรัช ศรีเลิศล้ำวนิช และคณะ (Virach Sornlertlamvanich et al., 1997) อมรา ประสิทธิ์รัฐสินธุ์ (2543) อ้างถึงใน นัฐวุฒิ ไชยเจริญ (2544: 16) นัฐวุฒิ ไชยเจริญ (2544: 60 – 77) เป็นต้น

การพิจารณาชนิดของคำของนักไวยากรณ์ข้างต้น ผู้วิจัยเห็นว่า การปรับปรุงชนิดของคำภาษาไทยให้เข้าโปรแกรม Unitex ควรเป็นชนิดของคำที่ใช้ในงานประมวลผลภาษาธรรมชาติ โดยปรับปรุงจากชนิดของคำ 2 ชุดที่ใช้ในงานประมวลผลภาษาธรรมชาติ คือ ORCHID และ โครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์ระหว่าง NECTEC – CICC เนื่องจากเป็นชุดชนิดของคำที่สร้างขึ้นเพื่อการประมวลผลภาษาธรรมชาติและคลังข้อมูลภาษา จึงน่าจะเหมาะสมกับงานวิจัยนี้

5.3.1 ORCHID

ORCHID (Virach Sornlertlamvanich et al., 1997) เป็นชนิดของคำภาษาไทยที่ใช้ในการกำกับชนิดของคำในคลังข้อมูลซึ่งเป็นส่วนหนึ่งของงานประมวลผลภาษาธรรมชาติ ชนิดของคำ ORCHID ได้รับการสร้างขึ้นในโครงการความร่วมมือระหว่างศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ประเทศไทย และหน่วยปฏิบัติการวิจัยการสื่อสาร (Communications Research Laboratory: CRL) ประเทศญี่ปุ่น

ชนิดของคำชุดนี้ได้แบ่งชนิดของคำโดยใช้หลายเกณฑ์ร่วมกัน คือ เกณฑ์ความหมาย ตำแหน่ง และหน้าที่ แบ่งเป็นชนิดของคำหลัก 14 ประเภท คำหลักบางประเภทมีชนิดของคำย่อยด้วย โดยสรุป ชนิดของคำชุดนี้ซึ่งรวมทั้งชนิดของคำหลักและชนิดของคำย่อยแล้วจะมีทั้งสิ้น 47 ประเภท ได้แก่

5.3.1.1 คำนาม (NOUN) แบ่งชนิดย่อยได้เป็น 6 ประเภท คือ Proper noun (NPRP), Cardinal number noun (NCNM), Ordinal number noun (NONM), Label noun (NLBL), Common noun (NCMN) และ Title noun (NTTL)

5.3.1.2 คำสรรพนาม (PRONOUN) แบ่งชนิดย่อยได้เป็น 4 ประเภท คือ Personal pronoun (PPRS), Demonstrative pronoun (PDMN), Interrogative pronoun (PNTR) และ Relative pronoun (PREL)

5.3.1.3 คำกริยา (VERB) แบ่งชนิดย่อยได้เป็น 3 ประเภท คือ Active verb (VACT), Stative verb (VSTA) และ Attributive verb (VATT)

5.3.1.4 คำช่วยกริยา (AUXILIARY) แบ่งชนิดย่อยได้เป็น 5 ประเภท คือ Pre-verb auxiliary, before negator (XVBM), Pre-verb auxiliary, after negator (XVAM), Pre-verb auxiliary, before or after negator (XVMM), Pre-verb auxiliary in imperative mood (XVBB) และ Post-verb auxiliary (XVAE)

5.3.1.5 คำบ่งชี้ (DETERMINER) แบ่งชนิดย่อยได้เป็น 9 ประเภท คือ Definite determiner: after a noun without a classifier in between (DDAN), Definite determiner: allowing a classifier in between (DDAC), Definite determiner: between a noun and a classifier, or preceding a quantitative expression (DDBQ), Definite determiner: following a quantitative expression (DDAQ), Indefinite determiner: following a noun and allowing a classifier in between (DIAC), Indefinite determiner: between a noun and a classifier, or preceding a quantitative expression

(DIBQ), Indefinite determiner: following a quantitative expression (DIAQ), Determiner: cardinal number expression (DCNM) และ Determiner: ordinal number expression (DONM)

5.3.1.6 คำกริยาวิเศษณ์ (ADVERB) แบ่งชนิดย่อยได้เป็น 4 ประเภท คือ Adverb with a normal form (ADV N), Adverb with an iterative form (ADV I), Adverb with a prefixed form (ADV P) และ Sentential adverb (ADV S)

5.3.1.7 คำลักษณนาม (CLASSIFIER) แบ่งชนิดย่อยได้เป็น 5 ประเภท คือ Unit classifier (CNIT), Collective classifier (CLTV), Measurement classifier (CMTR), Frequency classifier (CFQC) และ Verbal classifier (CVBL)

5.3.1.8 คำเชื่อม (CONJUNCTION) แบ่งชนิดย่อยได้เป็น 3 ประเภท คือ Coordinating conjunction (JCRG), Comparative conjunction (JCMP) และ Subordinate conjunction (JSBR)

5.3.1.9 คำบุพบท (PREPOSITION: RPRE)

5.3.1.10 คำอุทาน (INTERJECTION: INT)

5.3.1.11 คำเติมหน้า (PREFIX) แบ่งชนิดย่อยได้เป็น 2 ประเภท คือ Nominal prefix (FIXN) และ Adverbial prefix (FIXV)

5.3.1.12 อนุภาค (SENTENCE PARTICLE) แบ่งชนิดย่อยได้เป็น 2 ประเภท คือ Ending for affirmative sentence (EAFF) และ Ending for interrogative sentence particle (EITT)

5.3.1.13 คำปฏิเสธ (NEGATOR: NEG)

5.3.1.14 เครื่องหมาย (PUNCTUATION: PUNC) ทั้งนี้ เครื่องหมายที่ปรากฏในข้อความจะถูกจัดให้เป็นชนิดหนึ่งของคำสำหรับงานประมวลผลภาษาธรรมชาติด้วย เพราะการประมวลผลภาษาธรรมชาติเป็นการประมวลผลภาษามนุษย์โดยอัตโนมัติด้วยเครื่องคอมพิวเตอร์ ขั้นตอนพื้นฐานของการประมวลผลภาษาธรรมชาติเป็นการตัดคำอัตโนมัติซึ่งต้องกำหนดชนิดของคำให้กับคำต่าง ๆ ที่ปรากฏในข้อความรวมทั้งเครื่องหมายด้วย อีกทั้งเครื่องคอมพิวเตอร์ไม่มีระบบการรับรู้ (cognition) เช่นมนุษย์ทำให้ไม่สามารถรับรู้และไม่สามารถแยกแยะได้ว่า เครื่องหมายที่พบคืออะไรและใช้ประโยชน์อย่างไร ดังนั้น หากไม่กำหนดชนิดของคำให้กับเครื่องหมายที่ปรากฏในข้อความ จะทำให้เครื่องคอมพิวเตอร์เกิดความผิดพลาดในขั้นตอนการประมวลผลได้เพราะมีส่วนที่ไม่ได้รับการระบุชนิดของคำปรากฏอยู่

5.3.2 MT: โครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์ระหว่าง NECTEC – CICC

ชนิดของคำที่ใช้ในโครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์ระหว่าง NECTEC – CICC (นวลทิพย์ ดันติเสวตรรัตน์ และคณะ, 2536: 77-84) ในที่นี้ใช้คำย่อว่า MT เป็นหมวดคำภาษาไทยที่ใช้ในการกำกับชนิดของคำให้กับคำศัพท์ในพจนานุกรมอิเล็กทรอนิกส์ภาษาไทย ซึ่งได้รับการพัฒนาโดยคณะนักวิจัยของศูนย์วิจัยและพัฒนาปัญญาประดิษฐ์ สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี

ชนิดของคำชุดนี้ได้แบ่งชนิดของคำโดยใช้เกณฑ์หน้าที่ของคำในประโยค เพื่อมุ่งให้คอมพิวเตอร์เข้าใจภาษาเพื่อการประมวลผล โดยแบ่งเป็นชนิดของคำหลัก 13 ประเภท คำหลักบางประเภทมีชนิดของคำย่อยด้วย โดยสรุป ชนิดของคำชุดนี้ซึ่งรวมทั้งชนิดของคำหลักและชนิดของคำย่อยแล้วจะมีทั้งสิ้น 45 ประเภท ได้แก่

5.3.2.1 คำนาม (NOUN: N) แบ่งชนิดย่อยได้เป็น 7 ประเภท คือ Proper Noun: NPRP, Cardinal Number: NCNM, Ordinal Number: NONM, Adverb Like Noun: NALN, a Noun which is Conceptually an Adjective: NAJN, Label: NLBL และ Common Noun: NCMN

5.3.2.2 สรรพนาม (PRONOUN: P) แบ่งชนิดย่อยได้เป็น 6 ประเภท คือ Personal Pronoun: PPRS, Demonstrative Pronoun: PDMN, Definite Pronoun: PDEF, Indefinite Pronoun: PIND, Interrogative Pronoun: PNTR และ Relative Pronoun: PREL

5.3.2.3 กริยา (VERB: V) แบ่งชนิดย่อยได้เป็น 3 ประเภท คือ Active Verb: VACT, Stative Verb: VSTA และ Attributive Verb: VATT

5.3.2.4 คำช่วยกริยา (AUXILIARY VERB: AUX) แบ่งชนิดย่อยได้เป็น 5 ประเภท คือ Before Verb and Before “ไม่”: XVBM, Before Verb and After “ไม่”: XVAM, Before Verb; Before or After “ไม่”: XVMM, Before Verb; At the Beginning of Sentence: XVBB และ After Verb; At the End of Verb Phrase: XVAE

5.3.2.5 คำบ่งชี้ (DETERMINER: DET) แบ่งชนิดย่อยได้เป็น 9 ประเภท คือ After Noun: DDAN, After Noun and Classifier: DDAC, Between Noun and Classifier: DDBC, After Quantitative Noun: DDAQ, After Classifier: DIAC, Before Quantitative Noun: DIBQ, After Quantitative Noun: DIAQ, Cardinal Number: DCNM และ Ordinal Number: DONM

5.3.2.6 กริยาวิเศษณ์ (ADVERB) แบ่งชนิดย่อยได้เป็น 4 ประเภท คือ Adverb with Normal Form: ADVN, Adverb with Iterative Form: ADVI, Adverb with Prefixed Form: ADVP และ Sentential Adverb: ADVS

5.3.2.7 ลักษณนาม (CLASSIFIER)

5.3.2.8 คำสันธาน (CONJUNCTION) แบ่งชนิดย่อยได้เป็น 3 ประเภท คือ Coordinating Conjunction: CCRG, Comparative Conjunction: CCMP และ Subordinating Conjunction: CSBR

5.3.2.9 คำบุรพบท (PREPOSITION: RPRE) แบ่งชนิดย่อยได้เป็น 2 ประเภท คือ Verbal Preposition: RVBP และ Preposition except "RVBP": RPRE

5.3.2.10 คำอุทาน (INTERJECTION: INT)

5.3.2.11 คำอุปสรรค (PREFIX)

5.3.2.12 คำลงท้าย (ENDING: END) แบ่งชนิดย่อยได้เป็น 2 ประเภท คือ Ending for affirmative sentence: EAFF และ Ending for interrogative sentence: EITT

5.3.2.13 คำปฏิเสธ (NEGATOR: NEG)

5.3.3 ชนิดของคำที่ผ่านการปรับปรุงเพื่อใช้ในโปรแกรม Unitex

ชนิดของคำที่ผ่านการปรับปรุงโดยผนวกชนิดของคำทั้งสองชุดข้างต้นเข้าด้วยกัน และได้ปรับชื่อชนิดของคำเสียใหม่เพื่อให้สื่อความได้ชัดเจนมากขึ้น ในที่นี้ใช้ชื่อว่า CW ซึ่งมาจากคำว่า "Compound Words" เพราะคำนามแสดงอุปกรณณ์ด้านวิทยาศาสตร์ส่วนใหญ่เป็นคำประสม โดยแบ่งเป็นชนิดของคำหลัก 14 ประเภท คำหลักบางประเภทมีชนิดของคำย่อยด้วย โดยสรุป ชนิดของคำชุดนี้ซึ่งรวมทั้งชนิดของคำหลักและชนิดของคำย่อยแล้วจะมีทั้งสิ้น 46 ประเภท ได้แก่

5.3.3.1 คำนาม (NOUN: N) คือ คำที่ใช้เรียกคน สัตว์ สิ่งของ สถานที่ และสิ่งที่เป็นนามธรรม แบ่งชนิดย่อยได้เป็น 6 ประเภท คือ

5.3.3.1.1 Proper noun: N+Prp คือ คำนามชื่อเฉพาะ ซึ่งเป็นคำนามชื่อเฉพาะ เป็นชื่อเฉพาะของคน สถานที่ หรือสิ่งใดสิ่งหนึ่ง เช่น วินโดวส์ 95 เนคเทค เป็นต้น

5.3.3.1.2 Label noun: N+Lbl คือ คำนามที่เป็นเครื่องหมายหรือสัญลักษณ์ เช่น 1 2 3

5.3.3.1.3 Common noun: N+Cmn คือ คำนามทั่วไปที่ใช้เรียกคน สัตว์ สิ่งของ สถานที่ เช่น หนังสือ อาหาร ตัวเงิน

5.3.3.1.4 Title noun: N+Ttl คือ คำนำหน้าชื่อเพื่อบอกอาชีพ สถานภาพยศ ตำแหน่ง ฯลฯ เช่น อาจารย์ ดร.

5.3.3.1.5 Adjective like noun: N+Ajn คือ คำนามคล้ายคำวิเศษณ์ (Adjective) ที่แสดงการขยายคำอื่นได้เช่นเดียวกับคำวิเศษณ์ (Adjective) ได้แก่ ทางสุขภาพ ส่วนรวม

5.3.3.1.6 Adverb like noun: N+Aln คือ คำนามคล้ายคำกริยาวิเศษณ์ (adverb) ที่ทำหน้าที่ได้เช่นเดียวกับคำกริยาวิเศษณ์ (adverb) มักจะเป็นคำนามที่ใช้แสดงเวลา เช่น เมื่อเช้า รัชสมัย

5.3.3.2 คำบอกจำนวน (NUMBER: NB) คือ คำที่ใช้บอกจำนวน แบ่งชนิดย่อยได้เป็น 2 ประเภท คือ

5.3.3.2.1 Cardinal number: NB+Crđ คือ คำจำนวนนับซึ่งไม่ปรากฏในวลีแสดงจำนวน เช่น หนึ่ง สอง

5.3.3.2.2 Ordinal number: NB+Ord คือ คำลำดับที่ เช่น ที่หนึ่ง ที่สุดท้าย

5.3.3.3 คำสรรพนาม (PRONOUN: PRO) คือ คำที่ใช้แทนคำนามหรือนามวลี แบ่งชนิดย่อยได้เป็น 6 ประเภท คือ

5.3.3.3.1 Personal pronoun: PRO+Prs คือ บุรุษสรรพนามที่ใช้แทนตัวผู้พูด ผู้ที่พูดด้วย หรือคนหรือสิ่งที่กล่าวถึง เช่น ฉัน คุณ เขา

5.3.3.3.2 Interrogative pronoun: PRO+Int คือ สรรพนามแสดงการถาม เช่น อะไร ที่ไหน

5.3.3.3.3 Relative pronoun: PRO+Rel คือ ประพันธสรรพนามที่ใช้แทนนามที่อยู่ข้างหน้า และทำหน้าที่เชื่อมประโยคเข้าด้วยกัน เช่น ที่ ว่า ผู้

5.3.3.3.4 Demonstrative pronoun: PRO+Dmn คือ สรรพนามชี้เฉพาะแสดงตำแหน่ง ใช้แทนสิ่งที่ผู้พูดชี้เฉพาะเจาะจงถึงในตำแหน่งใกล้หรือไกลจากตัวผู้พูด เช่น นี้ โน่น

5.3.3.3.5 Definite pronoun: PRO+Def คือสรรพนามชี้เฉพาะแสดงขอบเขต โดยใช้แทนเพื่อบอกขอบเขตของสิ่งที่พูดถึง ได้แก่ เหล่านั้น สิ่งหนึ่ง

5.3.3.3.6 Indefinite pronoun: PRO+Ind คือสรรพนามไม่ชี้เฉพาะที่ใช้แทนคนหรือสิ่งต่างๆ ไป ได้แก่ บ้างทั้งหลาย

5.3.3.4 คำกริยา (VERB: V) คือคำที่แสดงกริยาอาการ สภาพ หรือคุณลักษณะของประธานในประโยค แบ่งชนิดย่อยได้เป็น 3 ประเภท คือ

5.3.3.4.1 Attributive verb: V+Att คือคำกริยาแสดงลักษณะ เช่น ไปงแวนว๊ว

5.3.3.4.2 Active verb: V+Act คือคำกริยาแสดงอาการ เช่น ทำงาน กระชิบ

5.3.3.4.3 Stative verb: V+Sta คือคำกริยาแสดงสภาพ ความรู้สึก เช่น กังวลใจ ฟุ้ง

5.3.3.5 คำช่วยกริยา (AUXILIARY: V+Aux) คือคำที่เกิดร่วมกับคำกริยาเพื่อทำหน้าที่ทางไวยากรณ์ เช่น กาล มาลา วาจก การณ์ลักษณะ ฯลฯ และค่อนข้างเกิดในตำแหน่งที่สัมพันธ์กับคำกริยาของประโยค แบ่งชนิดย่อยได้เป็น 5 ประเภท คือ

5.3.3.5.1 After verb; at the end of the verb phrase: V+Aux+av+e คือคำช่วยหลังกริยาที่วางหลังกริยาหรือกริยาวลี เช่น ขึ้น ลง เข้า

5.3.3.5.2 Before verb; after “ไม่” in negation: V+Aux+bv+am คือคำช่วยหน้ากริยาที่วางอยู่หน้ากริยาและหลังคำปฏิเสธ “ไม่” ในประโยคปฏิเสธ เช่น ค่อย นำ ได้

5.3.3.5.3 Before verb; before “ไม่” in negation: V+Aux+bv+bm คือคำช่วยหน้ากริยาที่วางอยู่หน้ากริยาและหน้าคำปฏิเสธ “ไม่” ในประโยคปฏิเสธ เช่น กำลัง อาจจะ

5.3.3.5.4 Before verb; before or after “ไม่” in negation: V+Aux+bv+mm คือคำช่วยหน้ากริยาที่วางหน้ากริยาและหน้าหรือหลังคำปฏิเสธ “ไม่” ในประโยคปฏิเสธ เช่น ต้อง อยาก ชอบ

5.3.3.5.5 Before verb; at the top of sentence: V+Aux+bv+bb คือคำช่วยหน้ากริยาที่วางอยู่หน้าคำกริยา โดยปกติจะอยู่ต้นประโยค เช่น กรุณา จง

5.3.3.6 คำบ่งชี้ (DETERMINER: DET) คือคำที่ใช้แสดงความเฉพาะเจาะจงของคำนาม แบ่งชนิดย่อยได้เป็น 9 ประเภท คือ

5.3.3.6.1 Definite determiner: After a noun; without classifier in between: DET+Def+an คือ คำบ่งชี้เฉพาะซึ่งวางอยู่หลังคำนามโดยไม่มีคำลักษณนาม (Classifier) มากั้น เช่น อย่างไม่อย่างหนึ่ง ส่วนใหญ่

5.3.3.6.2 Definite determiner: After a noun; allow classifier in between: DET+Def+ac คือ คำบ่งชี้เฉพาะซึ่งวางอยู่หลังคำนามหรือคำลักษณนาม เช่น นั้น โน้น

5.3.3.6.3 Definite determiner: After quantitative: DET+Def+aq คือ คำบ่งชี้หลังจำนวน เช่น เท่านั้น พอดี ถ้วน

5.3.3.6.4 Indefinite determiner: Between noun and classifier, or before quantitative expression: DET+Ind+bq คือ คำไม่ชี้เฉพาะซึ่งวางอยู่ระหว่างคำนามและคำลักษณนาม หรือหน้าจำนวน เช่น บาง ประมาณ

5.3.3.6.5 Definite determiner: Between noun and classifier, or before quantitative expression: DET+Def+bq คือ คำบ่งชี้เฉพาะซึ่งวางอยู่ระหว่างคำนามและคำลักษณนาม หรืออยู่หน้าจำนวน เช่น อีก สัก

5.3.3.6.6 Indefinite determiner: Between noun; may allow classifier in between: DET+Ind+ac คือ คำไม่ชี้เฉพาะซึ่งวางอยู่หลังคำนามและอาจมีคำลักษณนามอยู่ระหว่างคำนามและคำบ่งชี้ที่ไม่ชี้เฉพาะชนิดนี้ได้ เช่น ต่าง ๆ ถัดไป

5.3.3.6.7 Indefinite determiner: After quantitative expression: DET+Ind+aq คือ คำไม่ชี้เฉพาะหลังจำนวน เช่น พอดี ถ้วน

5.3.3.6.8 Determiner: Cardinal number expression: DET+Crd คือ คำแสดงจำนวนนับและใช้ในวลีแสดงจำนวน เช่น สาม เดียว

5.3.3.6.9 Determiner: Ordinal number expression: DET+Ord คือ คำแสดงลำดับที่ เช่น ที่หนึ่ง ที่สุดท้าย

5.3.3.7 คำกริยาวิเศษณ์ (ADVERB: ADV) คือ คำขยายกริยา คำกริยาวิเศษณ์อื่นหรือประโยค แบ่งชนิดย่อยได้เป็น 4 ประเภท คือ

5.3.3.7.1 Adverb with normal form: ADV+Nrm คือ คำกริยาวิเศษณ์ทั่วไปที่ไม่ได้เป็นคำซ้ำหรือเกิดมาจากคำกริยาที่เติมคำเติมหน้า “โดย” “อย่าง” เช่น จำเจ เร็ว

5.3.3.7.2 Adverb with iterative form: ADV+Itr คือ คำกริยาวิเศษณ์ที่เป็นคำซ้ำ เช่น เร็ว ๆ เสมอ ๆ

5.3.3.7.3 Adverb with prefixed form: ADV+Pfx คือ คำกริยาวิเศษณ์ที่เกิดจากการเติมคำเติมหน้า “โดย” “อย่าง” เช่น อย่างเร็ว โดยเร็ว

5.3.3.7.4 Sentential adverb: ADV+Snt คือ คำกริยาวิเศษณ์ที่ขยายประโยค มักปรากฏต้นประโยค เช่น ปกติธรรมดา

5.3.3.8 คำลักษณนาม (NOUN CLASSIFIER: CLAS) คือ คำที่แสดงลักษณะของคำนาม เช่น ชูค คู่ คณะ กระจ่าง

5.3.3.9 คำเชื่อม (CONJUNCTION: CONJ) คือ คำที่ใช้เชื่อมคำ วลี หรือประโยคเข้าด้วยกัน แบ่งชนิดย่อยได้เป็น 3 ประเภท คือ

5.3.3.9.1 Coordinating conjunction: CONJ+Crg คือ คำเชื่อมที่ใช้เชื่อมประโยคความรวม (Compound Sentence) หรือเชื่อมประโยคความเดียว (Simple Sentence) 2 ประโยคเข้าด้วยกันเพื่อแสดงความสัมพันธ์ในลักษณะใดลักษณะหนึ่ง เช่น ทั้ง และ

5.3.3.9.2 Comparative conjunction: CONJ+Cmp คือ คำเชื่อมแสดงความเปรียบเทียบ เช่น กว่า ที่สุด

5.3.3.9.3 Subordinate conjunction: CONJ+Sbr คือ คำเชื่อมอนุพากย์เพื่อแสดงเวลา เงื่อนไข วัตถุประสงค์ เหตุ หรือผล เช่น ว่า

5.3.3.10 คำบุพบท (PREPOSITION: PREP) คำที่นำหน้าคำนามหรือคำสรรพนามเพื่อบอกตำแหน่ง สถานที่ และความสัมพันธ์ระหว่างคำกริยาและคำนามหรือสรรพนามที่ตามมา แบ่งชนิดย่อยได้เป็น 2 ประเภท คือ

5.3.3.10.1 Verbal Preposition: PREP+Vrb คือ คำบุพบทที่มาจากคำกริยา เช่น จาก ข้าม

5.3.3.10.2 Preposition except “PREP+Vrb”: PREP+Nrm คือ คำบุพบททั่วไปที่ไม่ได้มาจากคำกริยา เช่น ก่อน บน ได้

5.3.3.11 คำปฏิเสธ (NEGATOR: NEG) คือ คำที่แสดงการปฏิเสธ เช่น ไม่ ไม่ได้ มิได้ มิ

5.3.3.12 คำเติมหน้า (PREFIX: PFX) คือ คำที่ใช้ประกอบหน้าคำเพื่อเปลี่ยนความหมายหรือหน้าที่ของคำ เช่น การ ความ โดย อย่าง

5.3.3.13 คำลงท้าย (ENDING: END) คือ คำที่ปรากฏท้ายประโยคเพื่อแสดงการถามหรือแสดงความสุภาพ แบ่งชนิดย่อยได้เป็น 2 ประเภท คือ

5.3.3.13.1 Ending for interrogative sentence: END+Int คือ คำลงท้ายประโยคเพื่อแสดงการถาม เช่น หรือ เหนอ ไหม มั้ย

5.3.3.13.2 Ending for affirmative sentence: END+Aff คือ คำลงท้ายในประโยคบอกเล่า เช่น นะ หน่อย เอะ ชี

5.3.3.14 คำอุทาน (INTERJECTION: INTJ) คือ คำที่ใช้ในการอุทาน เช่น โอ๊ย
ไอ้ เออ เอ้

ตารางที่ 15 ชนิดของคำที่ผ่านการปรับปรุงเพื่อใช้ในโปรแกรม Unitex

Part-of-Speech							
NO	ORCHID	Description	Example	MT	CW	Description	Example
1	NPRP	Proper noun	กรุงเทพฯ, ไค้ก	NPRP	N+Ppp	Proper noun	พระอาทิตย์, วินโดวส์
2	NTTL	Title noun	ผู้พัน, ด.ช.	NTTL	N+Ttl	Title noun	อาจารย์, มหาลัย
3	NCMN	Common noun	ประมวลกฎหมาย, หนังสือ	NCMN	N+Cmn	Common noun	ตัวเงิน, หนังสือ
4	NONM	Ordinal number	ปฐม, อันดับหนึ่ง	NONM	NB+Ord	Ordinal number	ที่หนึ่ง, ที่สุดท้าย
5	NCNM	Cardinal number	สิ่งทีเท่ากัน, หนึ่ง	NCNM	NB+Crd	Cardinal number	หนึ่ง, สอง, ครั้ง
6	NLBL	Label noun	1, 2, 3, 4, 5	NLBL	N+Lbl	Label noun	1, 2, 3, 4, 5
7	-	-	-	NAJN	N+Ajn	Adjective like noun	ทางสุขภาพ, ส่วนรวม
8	-	-	-	NALN	N+AIn	Adverb like noun	เมื่อเช้า, รัชสมัย
9	PPRS	Personal pronoun	หล่อน, พวกคุณ	PPRS	PRO+Prs	Personal pronoun	เขา, คุณ
10	PNTR	Interrogative pronoun	อันใด, ไหน	PNTR	PRO+Int	Interrogative pronoun	อะไร, ที่ไหน
11	PREL	Relative pronoun	ซึ่ง (ที่, ซึ่ง, อัน)	PREL	PRO+Rel	Relative pronoun	ผู้, ที่ว่า
12	PDMN	Demonstrative pronoun	นั่น, นี้, อยู่นั้น	PDMN	PRO+Dmn	Demonstrative pronoun	นี่, โน้น

ตารางที่ 15 ชนิดของคำที่ผ่านการปรับปรุงเพื่อใช้ในโปรแกรม Unitex (ต่อ)

Part-of-Speech							
NO	ORCHID	Description	Example	MT	CW	Description	Example
13	-	-	-	PDEF	PRO+Def	Definite pronoun	สิ่งหนึ่ง, เหล่านี้
14	-	-	-	PIND	PRO+Ind	Indefinite pronoun	บ้าง, ทั้งหลาย
15	VATT	Attributive verb	เดือน, บาน ขิมเยาะ, แยกกัน	VATT	V+Att	Attributive verb	ไปง, แหว้วม
16	VACT	Active verb	แยกกัน	VACT	V+Act	Active verb	กระซิบ, ซ่าน
17	VSTA	Stative verb	กังวลใจ, ควบ	VSTA	V+Sta	Stative verb	เวียน, ฟูง, เป็น
18	XVAE	Post-verb auxiliary	ขึ้น, ไป, มา	XVAE	V+Aux+av+e	After verb at the end of verb phrase	ขึ้น, ลง, เข้า
19	XVAM	Pre-verb auxiliary after negator “ไม่”	ค่อย, น่า, ได้	XVAM	V+Aux+bv+am	Before verb after “ไม่” in negation	ค่อย, น่า, ได้
20	XVBM	Pre-verb auxiliary before negator “ไม่”	เกิด, เกือบ, กำลัง	XVBM	V+Aux+bv+bm	Before verb before “ไม่” in negation	กำลัง, อาจจะ

ตารางที่ 15 ชนิดของคำที่ผ่านการปรับปรุงเพื่อใช้ในโปรแกรม Unitex (ต่อ)

Part-of-Speech							
NO	ORCHID	Description	Example	MT	CW	Description	Example
21	XVMM	Pre-verb before or after negator “ไม่”	ควร, เคย, ต้อง	XVMM	V+Aux+bv+mm	Before verb before or after “ไม่” in negation	ต้อง, อยาก, ชอบ
22	XVBB	Pre-verb auxiliary, in imperative mood	กรุณา, โปรด	XVBB	V+Aux+bv+bb	Before verb at the top of sentence	กรุณา, จง
23	DDAN	Definite determiner, after noun without classifier in between	ทั้งหมด, เพียงอย่างเดียว	DDAN	DET+Def+an	After noun, without classifier in between	อย่างใดอย่างหนึ่ง, ส่วนใหญ่
24	DDAC	Definite determiner, allowing classifier in between	นั่น, นั้น, อยู่นั้น	DDAC	DET+Def+ac	After noun, allow classifier in between	โน้น, นั่น
25	DDAQ	Definite determiner, following quantitative expression	เท่านั้น, เท่านั้น	DDAQ	DET+Def+aq	After quantitative	เท่านั้น, พอดี, ถ้วน

ตารางที่ 15 ชนิดของคำที่ผ่านการปรับปรุงเพื่อใช้ในโปรแกรม Unitex (ต่อ)

Part-of-Speech							
NO	ORCHID	Description	Example	MT	CW	Description	Example
26	DIBQ	Indefinite determiner, between noun and classifier or preceding quantitative expression	หลายๆ, บาง	DIBQ	DET+Ind+bq	Between noun and classifier or or before quantitative express	บาง, ประมาณ
27	DDBQ	Definite determiner, between noun and classifier or before quantitative expression	เพียง, อีก	DDBC	DET+Def+bq	Between noun and classifier or before quantitative expression	อีก, ถัก
28	DCNM	Determiner, cardinal number expression	หนึ่ง, สอง	DCNM	DET+Crđ	Cardinal number expression	สาม, เดียว
29	DIAC	Indefinite determiner, following noun and allowing classifier in between	อื่นๆ, ต่างๆ	DIAC	DET+Ind+ac	Between noun may allow classifier in between	ต่างๆ, ถัดไป
30	DIAQ	Indefinite determiner, following quantitative expression	เท่านั้น, พอดี	DIAQ	DET+Ind+aq	After quantitative expression	พอดี, ล้วน

ตารางที่ 15 ชนิดของคำที่ผ่านการปรับปรุงเพื่อใช้ในโปรแกรม Unix (ต่อ)

Part-of-Speech							
NO	ORCHID	Description	Example	MT	CW	Description	Example
31	DONM	Determiner, ordinal number expression	ที่หนึ่ง, ที่สอง	DONM	DET+Ord	Ordinal number expression	ที่หนึ่ง, ที่สุดท้าย
32	ADVI	Adverb with iterative form	บ่อยๆ, เร็วๆ	ADVI	ADV+Itr	Adverb with iterative form	เร็วๆ เสมอๆ
33	ADVN	Adverb with normal form	เร็ว, ช้า	ADVN	ADV+Nrm	Adverb with normal form	ช้าๆ, เร็ว
34	ADVP	Adverb with prefixed form	โดยเร็ว, ตามลำดับ	ADVP	ADV+Pfx	Adverb with prefixed form	อย่างรวดเร็ว, โดยเร็ว
35	ADVS	Sentential adverb	โดยปกติ, ธรรมดา	ADVS	ADV+Snt	Sentential adverb	ปกติ, ธรรมดา
36	CMTR	Measurement classifier	บาท, ปี				
37	CNIT	Unit classifier	ใบ, ขวด, ลูก				
38	CLTV	Collective classifier	คู่, กลุ่ม, ทาง	CLAS	CLAS	Noun classifier	ชุด, คู่, คณะ, กระจ่าง
39	CVBL	Verbal classifier	ม้วน, มัด				
40	CFQC	Frequency classifier	ครั้ง, เทียว, รอบ				
41	JSBR	Subordinating conjunction	ที่, แม้ว่า	CSBR	CONJ+Sbr	Subordinating conjunction	ว่า
42	JCMP	Comparative conjunction	เหมือนกัน, เท่ากัน	CCMP	CONJ+Cmp	Comparative conjunction	กว่า, ที่สุด

ตารางที่ 15 ชนิดของคำที่ผ่านการปรับปรุงเพื่อใช้ในโปรแกรม Unitex (ต่อ)

Part-of-Speech							
NO	ORCHID	Description	Example	MT	CW	Description	Example
43	JCRG	Coordinating conjunction	และ, หรือ, แต่	CCRG	CONJ+Crg	Coordinating conjunction	ทั้ง, และ
44	RPRE	Preposition	จาก, บน	RPRE	PREP+Nrm	Preposition except "RVBP"	ก่อน, ได้, บน
45	-		-	RVBP	PREP+Vrb	Verbal preposition	จาก, ช้าม
46	NEG	Negator	ไม่ได้, ไม่ได้	NEG	NEG	Negation	มิได้, ไม่ได้
47	PUNC	Punctuation	, ; : . ?	-	-	-	-
48	FIXV	Adverbial prefix	อย่างรวดเร็ว, โดยเร็ว	FIXP	PFX	Prefix	อย่าง, การ, ความ, โดย
49	FIXN	Nominal prefix	การ..., ความ...				
50	EITT	Ending for interrogative sentence	หรือ, หรือ, ไหม, มั้ย	EITT	END+Int	Ending for interrogative sentence	หรือ, หรือ, ไหม, มั้ย
51	EAFF	Ending for affirmative sentence	ใช่, ใช่, ค่ะ, ครับ	EAFF	END+Aff	Ending for affirmative sentence	นะ, หน่อย, เกอะ, ซิ
52	INT	Interjection	โธ่, โธ่, เออ, เอ	INT	INTJ	Interjection	โธ่, โธ่, เออ, เอ

5.3.4 การเปลี่ยนชนิดของคำจากแนวไวยากรณ์โครงสร้างเป็นชุดที่ใช้ในโปรแกรม Unix

เมื่อปรับปรุงชนิดของคำสำหรับใช้ในโปรแกรม Unix เรียบร้อยแล้ว ชนิดของคำที่ใช้ตามแนวไวยากรณ์โครงสร้างในบทที่ 3 จึงต้องปรับเปลี่ยนด้วย เพื่อให้สามารถนำผลการวิเคราะห์มาเขียนกราฟด้วยโปรแกรม Unix ได้ ผู้วิจัยจึงนำชนิดของคำแนวไวยากรณ์โครงสร้างมาเปรียบเทียบกับชนิดของคำที่ปรับปรุงแล้ว และปรับเปลี่ยนจากชนิดของคำแนวไวยากรณ์โครงสร้างให้เป็นไปตามชนิดของคำชุดใหม่ที่มีการปรับปรุงซึ่งให้ชื่อว่า CW ชื่อนี้มาจากคำว่า “Compound Words” เพราะค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ส่วนใหญ่เป็นคำประสม ทั้งนี้ การเปลี่ยนแปลงชนิดของคำนั้นไม่มีผลกระทบเท่าใดนัก เพราะชนิดของคำชุดใหม่เป็นชุดที่อิงแนวทางของไวยากรณ์โครงสร้างอยู่แล้ว การเปรียบเทียบชนิดของคำปรากฏดังตารางข้างล่างนี้

ตารางที่ 16 การเปรียบเทียบชนิดของคำแนวไวยากรณ์โครงสร้างและชุดที่ปรับปรุงใหม่

ที่	ไวยากรณ์โครงสร้าง	ตัวอย่าง	CW	ตัวอย่าง
1	ค่านาม	ฝน, นาฬิกา	N+Prp	พระอาทิตย์, วินโดวส์
2			N+Titl	อาจารย์, มเหสี
3			N+Cmn	ตัวเงิน, หนังสือ
4	คำคุณศัพท์	ส่วนตัว, เปล่า	N+Ajn	ทางสุขภาพ, ส่วนรวม
5			DET+Def+an	อย่างใดอย่างหนึ่ง, ส่วนใหญ่
6	คำบอกเวลา	กลางคืน, เช้า, ชั่วโมง	N+Aln	เมื่อเช้า, รัชสมัย
7	คำสรรพนาม	เขา, เธอ	PRO+Prs	เขา, คุณ
8			PRO+Int	อะไร, ที่ไหน
9			PRO+Dmn	นี่, โน่น
10			PRO+Def	สิ่งหนึ่ง, เหล่านั้น
11			PRO+Ind	บ้าง, ทั้งหมด
12	คำกริยากรรมย่อ	เย็น, เก่า	V+Att	โป่ง, แหว็บ

ตารางที่ 16 การเปรียบเทียบชนิดของคำแนวไวยากรณ์โครงสร้างและชุดที่ปรับปรุงใหม่ (ต่อ)

ที่	ไวยากรณ์ โครงสร้าง	ตัวอย่าง	CW	ตัวอย่าง
13	คำกริยากรรม	ตี, ตก	V+Act	กระชับ, ช่าน
14	คำกริยาทวิกรรม	ให้		
15	คำกริยากรรม	ทำ		
16			V+Sta	เวียน, ฟุ้ง, เป็น
17	คำช่วยหลังกริยา	แล้ว, อยู่, อยู่แล้ว	V+Aux+av+e	ขึ้น, ลง, เข้า
18	คำหลังกริยา	ขึ้น, ลง, เข้า		
19	คำหน้ากริยา	ไป, มา	V+Aux+bv+am	ค่อย, นำ, ได้
20	คำช่วยหน้ากริยา	เพิ่ง, จะ, เคย	V+Aux+bv+bm	กำลัง, อาจจะ
21			V+Aux+bv+mm	ต้อง, ยาก, ชอบ
22			V+Aux+bv+bb	กรุณา, จง
23	คำบอกกำหนดเสียงตรี	นี้, นั้น, โน้น, อื่น,	DET+Ind+ac	ต่างๆ, ถัดไป
24	และเสียงจัตวา	ต่าง ๆ	DET+Def+ac	โน้น, นั้น
25	คำบอกกำหนดเสียงโท	นี้, โน้น, นู่น		
26	คำหน้าจำนวน	อีก, สัก, ประมาณ	DET+Ind+bq	บาง, ประมาณ
27			DET+Def+bq	อีก, สัก
28	คำจำนวนนับ	หนึ่ง, บาง, ครั้ง	DET+Crd	สาม, เดียว
29			NB+Crd	หนึ่ง, สอง, ครั้ง
30			N+Lbl	1, 2, 3, 4, 5
31	คำหลังจำนวน	พอดี, ถ้วน, เศษ	DET+Def+aq	เท่านั้น, พอดี, ถ้วน
32			DET+Ind+aq	พอดี, ถ้วน
33	คำลำดับที่	แรก, สุดท้าย	DET+Ord	ที่หนึ่ง, ที่สุดท้าย
34			NB+Ord	ที่หนึ่ง, ที่สุดท้าย
35	คำกริยาวิเศษณ์	จ้ง, เหลือเกิน, อ้อก	ADV+Itr	เร็วๆ เสมอๆ
36			ADV+Nrm	จำใจ, เร็ว
37			ADV+Pfx	อย่างรวดเร็ว, โดยเร็ว

ตารางที่ 16 การเปรียบเทียบชนิดของคำแนวไวยากรณ์โครงสร้างและชุดที่ปรับปรุงใหม่ (ต่อ)

ที่	ไวยากรณ์โครงสร้าง	ตัวอย่าง	CW	ตัวอย่าง
38	คำพิเศษ	ปกติ, ธรรมดา	ADV+Snt	ปกติ, ธรรมดา
39	คำลักษณนาม	เต็ม, หลัง	CLAS	ชุด, คู่, คณะ, กระจ่าง
40	คำเชื่อมอนุพจน์	ว่า, ที่, นอกเสียจาก, หรือ, เลย, ตอนที่	CONJ+Sbr	ว่า
41			CONJ+Cmp	กว่า, ที่สุด
42			CONJ+Crg	ทั้ง, และ
43			PRO+Rel	ผู้, ที่ว่า
44	คำเชื่อมนาม	กะ, กับ, ของ	-	-
45	คำบุพบท	ใต้, จาก	PREP+Nrm	ก่อน, ใต้, บน
46			PREP+Vrb	จาก, ข้าม
47	คำปฏิเสธ	ไม่	NEG	มิได้, ไม่ได้
48	-	-	PFX	อย่าง, การ, ความ, โดย
49	คำลงท้าย	คะ, จ๊ะ, น้า	END+Int	หรือ, เหนือ, ไหม, มั้ย
50			END+Aff	นะ, หน่อย, เอะ, ซิ
51	-	-	INTJ	โอย, โอ้, เออ, เอ

5.4 การกำกับหมวดหมู่ตามความหมาย (AKO)

จากการทดสอบโครงสร้างค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์กับโปรแกรม Unitek ในขั้นต้นแล้ว พบว่า การกำหนดคุณสมบัติของส่วนประกอบด้วยชนิดของคำนั้นยังไม่เพียงพอ ทำให้เกิดข้อผิดพลาดในการตรวจจับ โครงสร้างคำ กล่าวคือ มีโครงสร้างที่ไม่ถูกต้องถูกตรวจจับรวมมาด้วยจำนวนมาก ผู้วิจัยจึงแก้ปัญหาด้วยการกำกับหมวดหมู่ทางความหมายให้แก่คำแต่ละคำเพื่อลดความผิดพลาดจากการตรวจจับในขั้นแรก

การกำกับหมวดหมู่ทางความหมายของคำได้ใช้การแบ่งกลุ่มทางความหมายที่เรียกว่า AKO (A Kind Of) ในโครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์ระหว่าง NECTEC-

CICC (นวลทิพย์ ตันติเสวตรรัตน์ และคณะ, 2536: 86-89) โดยคณะนักวิจัยของศูนย์วิจัยและพัฒนา
ปัญญาประดิษฐ์ สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี

AKO (A Kind Of) เป็นข้อมูลที่ใช้จัดหมวดหมู่ทางความหมายของคำในภาษาเพื่อ
ประโยชน์ในการวิเคราะห์ความหมายของประโยคและช่วยลดความกำกวมที่อาจเกิดขึ้นในประโยค
ได้ ตัวอย่างเช่น

ประโยค “ตากลับบ้าน” อาจเกิดความกำกวมที่คำนาม “ตา” เพราะเครื่องคอมพิวเตอร์ไม่
สามารถบอกได้ว่า คำนี้หมายถึง คุณตาที่เป็นคน (grandfather) หรือนัยน์ตา (eye) ที่เป็นอวัยวะได้
จึงต้องอาศัยคำที่อยู่ใกล้เคียงมาช่วยในการตัดสินใจ คือ คำกริยา “กลับ” จะพบว่า คำกริยาตัวนี้ต้องการ
ประธานที่เป็นสิ่งมีชีวิต ในที่นี้คือ “ตา” (grandfather)

จากตัวอย่างข้างต้น จึงทำให้ต้องมีการแบ่งกลุ่มทางความหมายให้กับคำแต่ละคำ โดย
จัดเป็นหมวดหมู่ทางความหมายเป็นลำดับขั้น แบ่งเป็น 2 ประเภทใหญ่ ดังนี้

1 Concrete: รูปธรรม

11 Subject: บุคคลและกลุ่มที่ประกอบด้วยคน

111 Person: บุคคล เช่น ผู้ปกครอง เขา ผู้ดูแล คู่รัก เพื่อน

112 Organization: กลุ่มหรือองค์กรต่าง ๆ เช่น บริษัท กรรมการ กลุ่ม คณะ

12 Concrete place: สถานที่

121 Region: สถานที่ที่เป็นดินแดนหรือมีขอบเขตที่แน่นอน เช่น จังหวัด

อำเภอ หมู่บ้าน

122 Nature: สถานที่ที่เป็นธรรมชาติ

1221 Topography: สถานที่ที่เป็นภูมิประเทศ เช่น ทุ่ง เทือกเขา

ชายทะเล

1222 Astronomy: สถานที่ที่เกี่ยวข้องกับดาราศาสตร์ เช่น โลก

พระอาทิตย์

123 Man-made place: สถานที่ที่มนุษย์สร้างขึ้น เช่น ทางด่วน ดึก บ้าน

13 Concrete thing: สิ่งของ

131 Nature thing: สิ่งของที่เกิดตามธรรมชาติ

1311 Living thing: สิ่งมีชีวิต

13111 Animal: สัตว์ เช่น หนอน หมา แมว

13112 A part of animal or its derivative: ส่วนของ
สิ่งมีชีวิตหรือสิ่งที่ได้จากสิ่งมีชีวิต เช่น มือ เท้า แขน ไข่ น้ำตา

13113 Plant: พืช เช่น ดอกไม้ มะนาว กุหลาบ

13114 A part of plant or its derivative: ส่วนของพืช
หรือสิ่งที่ได้จากพืช เช่น กิ่งไม้ ราก เปลือก

1312 Substance: สสาร

13121 Solid: ของแข็ง เช่น ทอง เงิน เพชร

13122 Liquid: ของเหลว เช่น น้ำ ปรอท น้ำมัน

13123 Gas: ก๊าซ เช่น อากาศ ไออน้ำ ก๊าซธรรมชาติ

132 Man-made thing: สิ่งของที่มนุษย์สร้างขึ้น

1321 Material: วัตถุดิบ เช่น พลาสติก ผ้า

1322 Finished product: ผลิตภัณฑ์ เช่น โต๊ะ ดวงไฟ รถยนต์

1323 A part of product or its derivative: ส่วนของผลิตภัณฑ์หรือ
ผลที่ได้จากผลิตภัณฑ์ เช่น ขาโต๊ะ ขี่ยางลบ ไบมีด

2 Abstract: นามธรรม

21 Abstract matter: นามธรรมที่ครอบคลุมเหตุการณ์และกริยาอาการต่าง ๆ

211 Activity: กริยาและอาการต่าง ๆ

2111 Action: การกระทำ เช่น ปฏิบัติการเล่น (การ) เดินทาง ใช้

2112 Mental action: การกระทำที่เกี่ยวกับความรู้สึกนึกคิด

21121 Rational action: การกระทำที่ต้องใช้เหตุผลและ
ความคิด เช่น คิด ตัดสินใจ ควบคุม

21122 Emotional action: การกระทำที่เกี่ยวข้องกับ
ความรู้สึก เช่น รัก หวัง โกรธ เสียใจ

212 Phenomenon: ปรากฏการณ์

2121 Event: เหตุการณ์ เช่น ดวง โชค ความสำเร็จ

2122 Natural phenomenon: ปรากฏการณ์ธรรมชาติ เช่น พายุฝน
สุริยุปราคา

2123 Social phenomenon: ปรากฏการณ์ทางสังคม เช่น สงคราม
การประท้วง การจลาจล

2124 Physiological phenomenon: ปรากฏการณ์ทางสรีระหรือทางร่างกาย เช่น อาการป่วย การเกิด การตาย การย่อยอาหาร

22 Abstract thing: นามธรรมที่มีลักษณะคล้ายสิ่งที่สามารถจับต้องได้

221 Product of activity: ผลของกริยาและอาการต่าง ๆ

2211 Result of action: ผลของการกระทำ เช่น คำอธิบายแผนการณื ข้อเสนอ

2212 Result of mental action: ผลของการกระทำที่เกี่ยวข้องกับความรูสึกนึกคิด เช่น ความจำ ความเป็นห่วง

2213 Abstract thing fixed as socially: สิ่งที่เป็นนามธรรมที่เกี่ยวข้องกับสังคม เช่น งานปีใหม่ งดลอยกระทง

2214 Abstract thing fixed as culture: สิ่งที่เป็นนามธรรมที่เกี่ยวข้องกับวัฒนธรรม เช่น ภาษาไทย รำวง มวยไทย

2215 Immaterial man-made thing: สิ่งที่ไม่ใช่วัตถุที่มนุษย์สร้างขึ้น เช่น วรรณคดี เพลง บทความ โปรแกรมคอมพิวเตอร์

23 Abstract relation: ความสัมพันธ์เชิงนามธรรม

231 Place: สถานที่ เช่น ขวา ซ้าย บน แลว ๆ

232 Time: เวลา เช่น เมื่อวานนี้ สัปดาห์นี้ สมัยธนบุรี

233 Existence: การดำรงอยู่ เช่น มี อยู่

234 Set: กลุ่ม เช่น เป็นแถว เรียงราย เป็นกลุ่ม

235 Attribute: คุณสมบัติของวัตถุหรือสิ่งของ

2351 Quality: คุณลักษณะ เช่น ลักษณะ อุปนิสัย

2352 Condition: สภาวะ เช่น สภาพร่างกาย สถานการณ์

2353 Apperance: ลักษณะที่ปรากฏให้เห็น เช่น รูปร่าง สี ตาย

2354 Sort/Type: ประเภทหรือชนิด เช่น แบบ ชนิด ประเภท เผ่า

2355 Number/Quantity: จำนวน เช่น ปริมาตร ความสูง

2356 Relation: ความสัมพันธ์ เช่น ความเกี่ยวข้อง

2357 Function: หน้าที่และประสิทธิภาพ เช่น ความสามารถ

ประสิทธิภาพ คุณภาพ

236 Value: ค่าของวัตถุหรือสิ่งของ

- 2361 Quality: ค่าที่แสดงลักษณะ เช่น ใจดี สูง ง่าย
- 2362 Condition: ค่าที่แสดงสภาวะ เช่น เป็นอันตราย ปลอดภัย
- 2363 Apperance: ค่าที่แสดงลักษณะที่ปรากฏ เช่น กลุ่ม สี เหลี่ยม
สว่าง แดง
- 2364 Sort/Type: ค่าที่แสดงประเภทหรือชนิด เช่น รุ่น 2503
- 2365 Number/Quantity: ค่าที่แสดงจำนวนหรือปริมาตร เช่น
แบบดิจิทัล
10 วัน มาก นาน
- 2366 Relation: ค่าที่แสดงความสัมพันธ์ เช่น ใกล้เคียง คล้าย
ติดกัน
- 2367 Function: ค่าที่แสดงหน้าที่และประสิทธิภาพ เช่น
(ประสิทธิภาพ) สูง (คุณภาพ) เยี่ยม

5.5 กฎโครงสร้างของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์

โครงสร้างของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่นำมาสร้างเป็นกฎสำหรับเขียนกราฟในโปรแกรม Unitek นั้นนำมาจากผลการวิเคราะห์โครงสร้างหลายระดับของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ในบทที่ 3 และ 4 โดยแบ่งได้เป็น 2 โครงสร้าง คือ โครงสร้างของส่วนประกอบของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ซึ่งแสดงด้วยกราฟย่อย และ โครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ซึ่งแสดงด้วยกราฟหลัก ดังนี้

(1) โครงสร้างของส่วนประกอบของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ได้มาจาก 3 โครงสร้าง คือ

(1.1) โครงสร้างระดับชนิดของค่า เป็น โครงสร้างที่แสดงคุณสมบัติของส่วนประกอบทางด้านชนิดของค่าที่เป็นส่วนประกอบ

(1.2) โครงสร้างระดับหน้าที่ เป็น โครงสร้างที่แสดงคุณสมบัติเชิงหน้าที่ของส่วนประกอบและช่วยบอกความแตกต่างของส่วนประกอบว่าเป็นส่วนประกอบประเภทใด เช่น นามวลี กริยาวลี หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา อนุภาค เป็นต้น

(1.3) โครงสร้างระดับตรรกะเป็นโครงสร้างที่ช่วยในการคำนวณการปรากฏของ ส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ว่า คำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ ประกอบด้วยส่วนประกอบที่เป็นคำแสดงและองค์ประกอบบังคับจำนวนเท่าใด ดังจะเห็นได้จาก โครงสร้างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ว่ามีความซับซ้อนมากขึ้น เพียงใด เช่น บางโครงสร้างมีส่วนประกอบ 2 ส่วน บางโครงสร้างมีส่วนประกอบ 3 ส่วน เป็นต้น

ผู้วิจัยได้นำโครงสร้างเหล่านี้มาเขียนเป็น กราฟย่อยของคำนามแสดงอุปกรณ์ด้าน วิทยาศาสตร์

(2) โครงสร้างหลักของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ซึ่งนำมาเขียนเป็น กราฟหลัก นั้น ได้มาจาก โครงสร้างระดับความหมายซึ่งแสดงความสัมพันธ์ทางความหมายระหว่าง ส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ เนื่องจาก

(2.1) ความสัมพันธ์ทางความหมายเป็นส่วนสำคัญของการรวมส่วนประกอบ ต่าง ๆ ของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์เพื่อร่วมกันบ่งชี้อุปกรณ์

(2.2) โครงสร้างแสดงความสัมพันธ์ทางความหมายระหว่างส่วนประกอบสามารถ แสดงความสัมพันธ์ระหว่างส่วนประกอบได้ดีขึ้น ทำให้เข้าใจลักษณะการรวมความหมายได้

(2.3) ความสัมพันธ์ทางความหมายแสดงลักษณะเด่นหรือลักษณะเฉพาะของ คำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ได้

สำหรับการตั้งชื่อกราฟย่อยและกราฟหลัก ผู้วิจัยตั้งชื่อกราฟย่อยตามชื่อความสัมพันธ์ทาง ความหมายเพราะสามารถแสดงความสัมพันธ์ระหว่างส่วนประกอบของคำนามแสดงอุปกรณ์ วิทยาศาสตร์ได้ดังเหตุผลที่กล่าวมาข้างต้น หากส่วนประกอบของคำนามแสดงอุปกรณ์วิทยาศาสตร์ มีความสัมพันธ์เป็นลำดับชั้น กล่าวคือ ส่วนประกอบเกิดการรวมเป็นลำดับชั้นก่อน เช่น ส่วนประกอบบางส่วนเป็นคำประสม เป็นอนุพากย์ เป็นต้น ชื่อกราฟย่อยจะใช้คำว่า CN ซึ่งย่อมาจาก Compound Nouns (แปลว่า คำนามประสม) เพื่อแสดงว่าเป็นคำประสม และ CL ซึ่งย่อมาจาก Clause (แปลว่า อนุพากย์) เพื่อแสดงว่าเป็นอนุพากย์ ส่วนชื่อกราฟหลัก ผู้วิจัยใช้ชื่อว่า ST ซึ่งย่อมาจาก Structure (แปลว่า โครงสร้าง) เพราะ โครงสร้างหลักของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ มีความซับซ้อนเป็นลำดับชั้น และเกิดจากการรวม โครงสร้างย่อยหลายรูปแบบเข้าด้วยกัน ทำให้มี

ลักษณะคล้ายไขเมงมูม อย่างไรก็ตาม หากโครงสร้างหลักของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ เป็นหน่วยสร้างคำประสมที่ไม่ค่อยซับซ้อน และมีลักษณะเช่นเดียวกับโครงสร้างส่วนประกอบที่เป็นกราฟย่อยชนิดคำประสม (มีชื่อว่า CN) ผู้วิจัยจะใช้ชื่อ CN เป็นชื่อกราฟหลักของโครงสร้างหลักที่มีลักษณะดังกล่าว ดังแสดงต่อไปนี้

5.5.1 กฎโครงสร้างของส่วนประกอบ

ส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์มีโครงสร้างที่ใช้เขียนเป็นกราฟย่อย โดยแบ่งตามชื่อกราฟย่อยได้ 35 กราฟ ดังต่อไปนี้

5.5.1.1 การกระทำ อากา แยกเป็น 5 กราฟ คือ การกระทำที่เป็นคำนาม (แทนด้วย ACT_N) การกระทำที่เป็นคำกริยา (แทนด้วย ACT_V) การกระทำที่เป็นนามวลี (แทนด้วย ACT_NP) การกระทำที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย ACT_VP1, ACT_VP2 และ ACT_VP3) และการกระทำที่เป็นหน่วยสร้างไร้ศูนย์ที่ขึ้นต้นด้วยคำบุพบท (แทนด้วย ACT_PP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 2111, 2211 และ 2124 หรือคำนามทั่วไปที่เกิดจากกราฟนามวลี “การ”
 - คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 หรือ 2124 และอาจมีคำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 ปรากฏตามมาก็ได้
 - คำบุพบททั่วไปที่มีค่า AKO เป็น 231 ตามด้วยกราฟ NP_K1
 - นามวลีที่เกิดจากกราฟ ACT_N ตามด้วย
 - (1) กราฟ SUBS, NAME หรือ LOC_PP
 - (2) “ใน” ซึ่งตามด้วยกราฟ PART
 - (3) “ของ” ซึ่งตามด้วยกราฟ PART, PHE, SYS_N, SYS_NP

หรือ SUBS

- หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาที่เกิดจากกราฟ ACT_V ตามด้วย

(1) กราฟ OBJ_NP, OBJ_CN, OBJ, MAN, MET_PP, CHA_V, LOC_N, LOC_P, LOC_PP, RES_N, RES_VP หรือ STA_VI

(2) กราฟ OBJ ตามด้วยกราฟ MET_PP

5.5.1.2 ความยิ่งใหญ่ ความสำคัญ (แทนด้วย GRT) เกิดจากคำนามทั่วไปที่มีค่า AKO เป็น 2361

5.5.1.3 จำนวน (แทนด้วย NUM) เกิดจาก

(1) คำบ่งชี้ที่แสดงจำนวนนับที่มีค่า AKO เป็น 2365

(2) คำบ่งชี้ที่แสดงลำดับที่มีค่า AKO เป็น 2365

(3) คำบ่งชี้เฉพาะที่มีค่า AKO เป็น 2355

(4) คำบ่งชี้ประเภทไม่ชี้เฉพาะหน้าจำนวนที่มีค่า AKO เป็น 2355

5.5.1.4 ชื่อเฉพาะ (แทนด้วย NAME) เกิดจากคำนามชื่อเฉพาะที่มีค่า AKO เป็น 111, 1322, 1323, 2111 หรือ 2215

5.5.1.5 ทั้งส่วน (แทนด้วย WHL) เกิดจาก

(1) คำนามทั่วไปที่มีค่า AKO เป็น 13111, 1312, 1322, 1323 หรือ 232

(2) กราฟคำประสม CN26

5.5.1.6 ทิศทาง แยกเป็น 4 กราฟ คือ ทิศทางที่เป็นคำนาม (แทนด้วย DIR_N) ทิศทางที่เป็นนามวลี (แทนด้วย DIR_NP1 และ DIR_NP2) ทิศทางที่เป็นคำกริยา (แทนด้วย DIR_V) และทิศทางที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย DIR_VP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏใน โครงสร้างมีดังต่อไปนี้

- คำนามสามัญที่มีค่า AKO เป็น 231

- คำบุพบททั่วไปและคำบุพบทที่มาจากคำกริยาที่มีค่า AKO เป็น 231 ตามด้วยกราฟ PART_NP1

- คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111

- คำช่วยกริยาที่วางหลังคำกริยาที่มีค่า AKO เป็น 2111

- นามวลีที่เกิดจาก

(1) กราฟ PART_NP1

(2) กราฟ DIR_N ตามด้วยกราฟ LOC_N, LOC_P และ PHE

(3) กราฟ DIR_N ที่มีการเวียนซ้ำ คือ มีคำนามแสดงทิศทาง

เกิดขึ้น 2 ตัวเรียงกัน

- กริยาวลีที่เกิดจากกราฟ DIR_V ที่มีการเวียนซ้ำ คือ มีคำกริยาหรือคำช่วยกริยาแสดงทิศทางเกิดขึ้น 2 ตัวเรียงกัน แล้วตามด้วยกราฟ MAN

5.5.1.7 **ที่มา** (แทนด้วย SRC) เกิดจากคำนามทั่วไปที่มีค่า AKO เป็น 13112 ซึ่งอาจมีคำบุพบท “จาก” นำหน้า

5.5.1.8 **นามวลี “การ” “ความ”** แยกเป็น 4 กราฟ คือ นามวลี “การ” แบบที่ 1 (แทนด้วย NP_K1) นามวลี “การ” แบบที่ 2 (แทนด้วย NP_K2) นามวลี “การ” แบบที่ 3 (แทนด้วย NP_K3) และนามวลี “ความ” (แทนด้วย NP_K4)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำเติมหน้า “การ” ตามด้วย

◆ คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 ซึ่งตามด้วยคำกริยาแสดงอาการที่มีค่า AKO เป็น 2111

◆ คำกริยาแสดงอาการที่มีค่า AKO เป็น 2124 ซึ่งอาจตามด้วย

(1) คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111

(2) กราฟ OBJ ซึ่งอาจมีกราฟ PUR_V ตามกราฟ OBJ

◆ กราฟ MET_V ตามด้วยกราฟ OBJ_NP และตามด้วยกราฟ ACT_V

◆ กราฟ ACT_V ตามด้วย กราฟ MET_N หรือ MET_PP

- คำเติมหน้า “ความ” ตามด้วย

◆ คำกริยาแสดงลักษณะที่มีค่า AKO เป็น 2361 และ 2363

◆ คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 ซึ่งอาจตามด้วย

(1) คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111

(2) กราฟ OBJ

◆ คำกริยาแสดงสภาพที่มีค่า AKO เป็น 2362 ซึ่งอาจตามด้วยกราฟ COMP

5.5.1.9 ประเภทย่อย แยกเป็น 10 กราฟ ได้แก่ ประเภทย่อยที่เป็นค่านาม (แทนด้วย TYPE) ประเภทย่อยแสดงรูปร่างลักษณะ (แทนด้วย TYPE_CHA1, TYPE_CHA2 และ TYPE_CHA3) ประเภทย่อยแสดงสถานที่ (แทนด้วย TYPE_LOC1 และ TYPE_LOC2) ประเภทย่อยแสดงลักษณะการกระทำ (แทนด้วย TYPE_MAN) ประเภทย่อยแสดงวิธีการ (แทนด้วย TYPE_MET1 และ TYPE_MET2) ประเภทย่อยแสดงส่วนประกอบ (แทนด้วย TYPE_PART) ประเภทย่อยแสดงสภาพ (แทนด้วย TYPE_STA1 และ TYPE_STA2) ประเภทย่อยแสดงระบบ (แทนด้วย TYPE_SYS1 และ TYPE_SYS2) ประเภทย่อยแสดงหัวข้อ (แทนด้วย TYPE_TOPIC) และประเภทย่อยหลัก (แทนด้วย TYPE_NP1 และ TYPE_NP2)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- ค่านามทั่วไปที่มีค่า AKO เป็น 2354
- คำบุพบททั่วไปที่มีค่า AKO เป็น 2366 ตามด้วยกราฟ TYPE แล้วตามด้วยกราฟ TOPIC

- กราฟ CHA_N2, CHA_N3, CHA_V, CHA_NP5, CHA_NP7, PHE_NP, MET_N, MET_CN, MET_V, MET_VP1, MET_VP2, MET_VP3, LOC_NP, PART_NP5, PART_NP6, QNT1, QNT3, CL1 หรือ CL2

- กราฟ TYPE ตามด้วยกราฟ CHA_V, CHA_N2, CHA_NP7, DIR_V, LOC_PP, LOC_VP, MAN, MAT, MET_CN, MET_VP1, MET_VP2, MET_VP3, NAME, STA_N, STA_V1, STA_VP2, SYS_N, SYS_VP, SYS_A&V, TYPE_CHA1, TYPE_MET1, TYPE_SYS1, TYPE_PART หรือ TYPE_STA2

5.5.1.10 ปรากฏการณ์ธรรมชาติ แยกเป็น 2 กราฟ ได้แก่ ปรากฏการณ์ธรรมชาติที่เป็นค่านาม (แทนด้วย PHE) และปรากฏการณ์ธรรมชาติที่เป็นนามวลี (แทนด้วย PHE_NP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- ค่านามชื่อเฉพาะที่มีค่า AKO เป็น 2122
- ค่านามทั่วไปที่มีค่า AKO เป็น 2122 หรือ 2353
- กราฟ CL1
- นามวลีที่เกิดจากกราฟ PHE ตามด้วยกราฟ CHA_N3, LOC_P, SUBS, ACT_V, STA_V1 หรือ STA_V2

5.5.1.11 ปริมาณ แยกเป็น 3 กราฟ ได้แก่ ปริมาณที่เป็นค่านาม (แทนด้วย QNT_N) ปริมาณที่เป็นนามวลี (แทนด้วย QNT_NP) และปริมาณที่เป็นหน่วยจำนวน (ซึ่งแทนด้วย QNT1, QNT2 และ QNT3)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- ค่านามทั่วไปที่มีค่า AKO เป็น 2122 หรือ 2355
- นามวลีที่เกิดจากกราฟ QNT_N ตามด้วย
 - (1) กราฟ ANI, INS, STA_N, STA_V2 หรือ ACT_VP1
 - (2) กราฟ ACT_N, PHE หรือ SUBS ซึ่งอาจมีคำบุพบท “ของ” นำหน้า
- หน่วยจำนวนที่เกิดจาก
 - (1) กราฟ UNIT ตามด้วยกราฟ NUM
 - (2) กราฟ NUM ตามด้วยกราฟ UNIT ทั้งนี้ จะปรากฏเป็น

ส่วนประกอบของกราฟ QNT1 และ QNT3 แต่ต่างกันตรงที่กราฟ QNT1 จะใส่เครื่องหมาย # ระหว่างกราฟ NUM และกราฟ UNIT เพื่อแสดงเงื่อนไขไม่ให้มีการเว้นวรรคระหว่างคำ ในขณะที่กราฟ QNT3 จะไม่ใส่เครื่องหมาย # ผู้วิจัยได้สร้างกราฟสองกราฟขึ้นเพื่อแก้ปัญหาความผิดพลาดของการตรวจจับโครงสร้างดังจะกล่าวถึงโดยละเอียดต่อไปในหัวข้อ 5.7 ผลการทดลอง

5.5.1.12 ผล แยกเป็น 4 กราฟ ได้แก่ ผลที่เป็นค่านาม (แทนด้วย RES_N) ผลที่เป็นคำกริยา (แทนด้วย RES_V) ผลที่เป็นนามวลี (แทนด้วย RES_NP) และผลที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย RES_VP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- ค่านามชื่อเฉพาะที่มีค่า AKO เป็น 2122
- ค่านามทั่วไปที่มีค่า AKO เป็น 13111, 13112, 1312, 13121, 13123, 1322, 1323, 2122, 2124, 2211, 234, 2353 หรือ 2362
- คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111
- คำกริยาแสดงลักษณะที่มีค่า AKO เป็น 2363
- คำกริยาแสดงสภาพที่มีค่า AKO เป็น 2362
- กราฟ STA_VP3
- นามวลีที่เกิดจากกราฟ PHE ตามด้วยกราฟ STA_V2

- หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาที่เกิดจากคำกริยา “เป็น” ตาม
ด้วยกราฟ RES_N

5.5.1.13 ผู้ถูก แยกเป็น 4 กราฟ ได้แก่ ผู้ถูกที่เป็นคำนาม (แทนด้วย OBJ) ผู้ถูกที่เป็น
เป็นคำนามประสม (แทนด้วย OBJ_CN) ผู้ถูกที่เป็นนามวลี (แทนด้วย OBJ_NP) และผู้ถูกที่เป็น
หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย OBJ_VP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏใน โครงสร้างมีดังต่อไปนี้

- คำนามชื่อเฉพาะที่มีค่า AKO เป็น 2122
- คำนามทั่วไปที่มีค่า AKO เป็น 1222, 13111, 13112, 1312, 13121,
13122, 13123, 132, 1321, 1322, 1323, 2111, 2122, 2124, 2211, 2215, 231, 232, 2353, 2355, 2361,
2362 หรือ 2363
- กราฟคำประสม CN1.1, CN1.4, CN1.5, CN8, CN15 หรือ CN24
- กราฟ ACT_NP, ACT_VP2, ANI_NP, CHA_NP2, CHA_NP4,
DIR_NP1, INS_NP, LOC_NP, PART_NP1, PART_NP3, PHE_NP, QNT_NP, STA_NP,
SUBS_NP, THN_NP หรือ TYPE_TOPIC

5.5.1.14 ผู้ทรงสภาพ (แทนด้วย PAT) เกิดจาก

- (1) คำนามทั่วไปที่มีค่า AKO เป็น 1323, 2355, 2362 หรือ 2363
- (2) กราฟคำประสม CN13

5.5.1.15 ผู้ทำ (แทนด้วย AGT) เกิดจาก

- (1) คำนามสามัญที่มีค่า AKO เป็น 13112, 13122, 1323 และ 2122
- (2) นามวลีที่เกิดจากกราฟ QNT1 ตามด้วย “หรือ” แล้วตามด้วยกราฟ QNT1

5.5.1.16 ผู้เสริม (แทนด้วย COMP) เกิดจาก

- (1) คำนามสามัญที่มีค่า AKO เป็น 13111, 1312, 1322, 1323, 2111, 2361
หรือ 2363 แล้วอาจตามด้วยเครื่องหมายขั้วติงค์ (-) แล้วตามด้วยคำนามสามัญที่มีค่า AKO เป็น 1312
- (2) กราฟคำประสม CN1.4 หรือ CN3.1
- (3) นามวลี “การ” ที่เกิดจากกราฟ NP_K1

5.5.1.17 ระบบ แยกเป็น 4 กราฟ ได้แก่ ระบบที่เป็นคำกริยาหรือคำวิเศษณ์ (แทนด้วย SYS_A&V) ระบบที่เป็นคำนาม (แทนด้วย SYS_N) ระบบที่เป็นนามวลี (แทนด้วย SYS_NP) และระบบที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย SYS_VP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำกริยาแสดงลักษณะที่มีค่า AKO เป็น 2361
- คำวิเศษณ์ที่เกิดจากการเติมคำเติมหน้า และมีค่า AKO เป็น 2361
- คำนามชื่อเฉพาะที่มีค่า AKO เป็น 111 หรือ 2215
- คำนามทั่วไปที่มีค่า AKO เป็น 2211 หรือ 2215
- นามวลีที่เกิดจากกราฟ SYS_N ตามด้วยกราฟ ACT_N
- หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาที่เกิดจากกราฟ ACT_VP1 หรือ

ACT_VP2

5.5.1.18 รูปร่างลักษณะ แยกเป็น 4 กราฟ ได้แก่ รูปร่างลักษณะที่เป็นคำนาม (แทนด้วย CHA_N1, CHA_N2 และ CHA_N3) รูปร่างลักษณะที่เป็นคำกริยา (แทนด้วย CHA_V) รูปร่างลักษณะที่เป็นนามวลี (แทนด้วย CHA_NP1, CHA_NP2, CHA_NP3, CHA_NP4, CHA_NP5, CHA_NP6, CHA_NP7 และ CHA_NP8) และรูปร่างลักษณะที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย CHA_VP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 13112, 13121, 1322, 2215, 234, 2353, 2362 หรือ 2363

- คำนามชื่อเฉพาะที่มีค่า AKO เป็น 2215
- คำกริยาแสดงลักษณะที่มีค่า AKO เป็น 2361 หรือ 2363
- กราฟคำประสม CN25.2
- นามวลีที่เกิดจาก

(1) กราฟ PART_NP1 หรือ PART_NP2

(2) กราฟ CHA_N1 ตามด้วย

(2.1) กราฟ NAME, PART, QNT3, CHA_NP1

(2.2) “ของ” และตามด้วยกราฟคำประสม CN14

มีการเขียนซ้ำ (3) กราฟ CHA_N1 หรือ PART แล้วตามด้วยกราฟ CHA_V ซึ่ง

(4) กราฟ CHA_N1 หรือกราฟ CHA_N2 ตามด้วยกราฟ CHA_N2

(5) กราฟ CHA_N3 ตามด้วยกราฟ ACT_V
- หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาที่เกิดจากกราฟ STA_V1 ซึ่งอาจมีกราฟ ACT_V อยู่ข้างหน้าหรือไม่ก็ได้ แล้วตามด้วยกราฟ COMP

5.5.1.19 ลักษณะการกระทำ (แทนด้วย MAN) เกิดจาก

- (1) คำกริยาวิเศษณ์ทั่วไปที่มีค่า AKO เป็น 2361 หรือ 2362
- (2) คำกริยาวิเศษณ์ที่เกิดจากการเติมคำเติมหน้าที่มีค่า AKO เป็น 2361
- (3) คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111
- (4) คำกริยาแสดงลักษณะที่มีค่า AKO เป็น 2361 ซึ่งอาจตามด้วยคำกริยาวิเศษณ์ทั่วไปที่มีค่า AKO เป็น 2367

5.5.1.20 วัตถุ แยกเป็น 2 กราฟ ได้แก่ วัตถุที่เป็นคำนาม (แทนด้วย THN) และวัตถุที่เป็นนามวลี (แทนด้วย THN_NP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 1222, 1321, 1322 หรือ 2215
- นามวลีที่เกิดจากกราฟ THN ตามด้วยกราฟ TOPIC

5.5.1.21 วัตถุประสงค์ แยกเป็น 6 กราฟ ได้แก่ วัตถุประสงค์ที่เป็นคำนาม (แทนด้วย PUR_N) วัตถุประสงค์ที่เป็นคำนามชนิดคำประสม (แทนด้วย PUR_CN) วัตถุประสงค์ที่เป็นคำกริยา (แทนด้วย PUR_V) วัตถุประสงค์ที่เป็นนามวลี (แทนด้วย PUR_NP) วัตถุประสงค์ที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย PUR_VP1, PUR_VP2, PUR_VP3, PUR_VP4, PUR_VP5, PUR_VP6, PUR_VP7 และ PUR_VP8) และวัตถุประสงค์ที่เป็นอนุภาค (แทนด้วย PUR_CL)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 1221, 1312, 13121, 13122, 13123, 1322, 1323, 2111, 2122, 2211, 2215, 231, 232, 2353, 2355 หรือ 2362 ทั้งนี้ อาจมีคำเชื่อม “สำหรับ” ปรากฏข้างหน้าก็ได้

- คำนามคล้ายคำวิเศษณ์ (Adjective) ที่มีค่า AKO เป็น 234

- คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 ซึ่งอาจมีคำเชื่อม “สำหรับไว้เพื่อ” อยู่ข้างหน้า หรืออาจมีคำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 คำที่สองตามมา ซึ่งอาจมีคำเชื่อม “และ” ปรากฏระหว่างคำกริยาแสดงอาการคำที่หนึ่งและคำที่สองหรือไม่ก็ได้

- นามวลี “การ ความ” ที่เกิดจากกราฟ NP_K1, NP_K3 หรือ NP_K4 ซึ่งอาจมีคำเชื่อม “สำหรับ” ปรากฏหน้ากราฟทั้งสามก็ได้

- กราฟคำประสม CN1.6, CN3.2 หรือ CN2

- นามวลีที่เกิดจากกราฟ CHA_NP2, CHA_NP8, LOC_NP, PART_NP1, PHE_NP, QNT_NP, STA_NP, STA_VP1, STA_VP3 หรือ SUBS_NP

- กริยาวลีที่เกิดจากกราฟ PUR_V ตามด้วย

(1) กราฟ OBJ, OBJ_CN, OBJ_NP, RES_N, ACT_PP, LOC_N, LOC_P, LOC_NP, LOC_PP, MAN, MET_N, TIME2, CL1 หรือ CL2

(2) กราฟ NP_K1 หรือ NP_K4 ซึ่งอาจมีคำบุพบท “ใน” อยู่ข้างหน้าได้

(3) กราฟ ACT_V ซึ่งอาจตามด้วยกราฟ LOC_PP

(4) กราฟ ACT_VP2 ซึ่งอาจมีกราฟ OBJ นำหน้า

(5) กราฟ MET_PP ซึ่งอาจมีกราฟ OBJ_NP นำหน้า

(6) กราฟ SRC ซึ่งอาจมีกราฟ RES_N นำหน้า

(7) กราฟ PUR_VP1 หรือ PUR_VP3 ซึ่งอาจมีคำเชื่อม “และ” นำหน้ากราฟทั้งสอง

(8) กราฟ OBJ แล้วตามด้วยกราฟ SYS_A&V, STA_V1, MET_CL, MET_PP, RES_VP หรือ PUR_VP1

(9) กราฟ OBJ_CN ตามด้วยกราฟ SYS_A&V

- อนุพากย์ที่เกิดจาก

- (1) คำเชื่อม “ที่” ตามด้วยกราฟ PUR_VP1
- (2) คำเชื่อม “เพื่อให้” ตามด้วยกราฟ CL2

5.5.1.22 วัสดุ แยกเป็น 4 กราฟ ได้แก่ วัสดุที่เป็นคำนาม (แทนด้วย MAT) วัสดุที่เป็นคำนามชนิดคำประสม (แทนด้วย MAT_CN) วัสดุที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย MAT_VP) และวัสดุที่เป็นอนุพากย์ (แทนด้วย MAT_CL)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 1321
- กราฟคำประสม CN1.2, CN2.1, CN2.2, CN3.1, CN3.3, CN4 หรือ CN6
- หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาซึ่งเกิดจากกราฟ ACT_V ตามด้วย “ด้วย” และกราฟ MAT

- อนุพากย์ที่เกิดจาก

- (1) “ให้” ตามด้วยกราฟ MET_V และตามด้วยกราฟ DIR_NP1
- (2) “ที่” ตามด้วยกราฟ MET_V แล้วตามด้วย
 - (2.1) กราฟ MET_N ซึ่งอาจมีคำบุพบท “ด้วย” อยู่ข้างหน้า
 - (2.2) กราฟ MAN ซึ่งตามด้วยกราฟ SYS_A&V
 - (2.3) กราฟ OBJ ซึ่งตามด้วยกราฟ SYS_A&V STA_VP1 หรือ DIR_V ซึ่งตามด้วย “ได้”
- (3) “โดย” ตามด้วยคำปฏิเสธ และ “ต้อง” แล้วตามด้วยกราฟ MET_V ซึ่งตามด้วยกราฟ OBJ หรือ OBJ_NP
- (4) “ซึ่ง” ตามด้วยกราฟ MET_V ซึ่งตามด้วยกราฟ OBJ หรือ OBJ_NP

5.5.1.23 วิธีกร แยกเป็น 5 กราฟ ได้แก่ วิธีกรที่เป็นคำนาม (แทนด้วย MET_N) วิธีกรที่เป็นคำกริยา (แทนด้วย MET_V) วิธีกรที่เป็นนามวลี (แทนด้วย MET_NP) วิธีกรที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย MET_VP1, MET_VP2 และ MET_VP3) และวิธีกรที่เป็นหน่วยสร้างไร้ศูนย์ที่ขึ้นต้นด้วยคำบุพบท (แทนด้วย MET_PP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 1221, 1222, 13112, 1312, 13121, 13122, 13123, 1322, 1323, 2111, 2122, 2215, 2353, 2362 หรือ 2363 ซึ่งอาจตามด้วยเครื่องหมายอัฒจันทร์ (-) แล้วตามด้วยคำนามทั่วไปที่มีค่า AKO เป็น 13123

- คำกริยาแสดงอาการที่มีค่า AKO เป็น 2124

- คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 หรือ 2124 ซึ่งอาจตามด้วยคำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 โดยมีคำเชื่อม “หรือ” ปรากฏอยู่ระหว่างคำกริยาแสดงอาการทั้งสองหรือไม่ก็ได้

- หน่วยสร้างไร้ศูนย์ที่ขึ้นต้นด้วยคำบุพบทที่เกิดจากคำบุพบททั่วไปที่มีค่า AKO เป็น 231 หรือ 2366 แล้วตามด้วยกราฟ MET_N, NP_K1 หรือ NP_K2

- นามวลีที่เกิดจาก

(1) กราฟ DIR_NP2, LOC_NP หรือ PHE_NP

(2) คำว่า “เชิง” ตามด้วยกราฟ MET_N

- หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาซึ่งเกิดจากกราฟ MET_V ซึ่งอาจมี “เชิง” นำหน้า จากนั้นตามด้วย

(1) กราฟ OBJ_NP, QNT1, QNT2, STA_V1, DIR_NP1, DIR_NP2 หรือ LOC_N

(2) กราฟ MET_N ซึ่งอาจมีคำบุพบท “ด้วย” นำหน้า

(3) กราฟ OBJ หรือ DIR_V ซึ่งอาจมี “ได้” ตามหลัง

(4) กราฟ OBJ ตามด้วยกราฟ SYS_A&V, PUR_VP6 หรือ DIR_V

(5) กราฟ MAN ตามด้วยกราฟ SYS_A&V

5.5.1.24 เวลา แยกเป็น 2 กราฟ ได้แก่ เวลาที่เป็นอนุพากย์ (แทนด้วย TIME1) และ เวลาที่เป็นคำกริยา (แทนด้วย TIME2)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- อนุพากย์ที่เกิดจากคำว่า “ก่อน ขณะ” ตามด้วยกราฟ ACT_VP1

- คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111

5.5.1.25 สถานที่ แยกเป็น 6 กราฟ ได้แก่ สถานที่ที่เป็นคำบ่งชี้แสดงจำนวนหรือลำดับที่ (แทนด้วย LOC_D) สถานที่ที่เป็นคำนาม (แทนด้วย LOC_N) สถานที่ที่เป็นคำบุพบททั่วไป (แทนด้วย LOC_P) สถานที่ที่เป็นนามวลี (แทนด้วย LOC_NP) สถานที่ที่เป็นหน่วยสร้างไวยากรณ์ที่ขึ้นต้นด้วยคำบุพบท (แทนด้วย LOC_PP) และสถานที่ที่เป็นหน่วยสร้างไวยากรณ์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย LOC_VP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำบ่งชี้แสดงจำนวนหรือลำดับที่มีค่า AKO เป็น 2365
- คำนามทั่วไปที่มีค่า AKO เป็น 1221, 1222, 123, 13112, 1322, 1323, 2211, 231 หรือ 2353
- คำนามทั่วไปที่มีค่า AKO เป็น 2354 แล้วตามด้วยกราฟ TOPIC
- คำบุพบททั่วไปที่มีค่า AKO เป็น 231
- นามวลีที่เกิดจากกราฟ LOC_N ตามด้วย
 - (1) กราฟ STA_N, PART, PHE, SUBS หรือ MET_V
 - (2) “ของ” แล้วตามด้วยกราฟ STA_N
- หน่วยสร้างไวยากรณ์ที่ขึ้นต้นด้วยคำบุพบทที่เกิดจากกราฟ LOC_P ตามด้วยกราฟ LOC_N, CL1 หรือกราฟคำประสม CN1.4
- หน่วยสร้างไวยากรณ์ที่ขึ้นต้นด้วยคำกริยาที่เกิดจากกราฟ ACT_V ตามด้วยกราฟ LOC_N

5.5.1.26 สภาพ แยกเป็น 4 กราฟ ได้แก่ สภาพที่เป็นคำนาม (แทนด้วย STA_N) สภาพที่เป็นคำกริยา (แทนด้วย STA_V1 และ STA_V2) สภาพที่เป็นนามวลี (แทนด้วย STA_NP) และสภาพที่เป็นหน่วยสร้างไวยากรณ์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย STA_VP1, STA_VP2 และ STA_VP3)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 13122, 1322, 2211, 2355, 2361, 2362 หรือ 2363
- กราฟนามวลี “ความ” (แทนด้วย NP_K4)
- คำกริยาแสดงสภาพที่มีค่า AKO เป็น 233, 2362

- คำกริยาแสดงลักษณะที่มีค่า AKO เป็น 2361 หรือ 2363 แล้วตามด้วย คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 หรือคำกริยาแสดงลักษณะที่มีค่า AKO เป็น 2361 ทั้งนี้ อาจมีคำเชื่อม “และ” ปรากฏนำหน้าหรือมี “กัน” ปรากฏตามหลังก็ได้

- นามวลีที่เกิดจากกราฟ STA_N ตามด้วย

(1) กราฟ CHA_N2, INS, LOC_N, SUBS หรือ STA_V2

(2) กราฟ PHE ซึ่งอาจมีคำบุพบท “ของ” นำหน้าก็ได้

- หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาที่เกิดจากกราฟ STA_V1 ตามด้วยกราฟ COMP, LOC_N, MAN หรือ QNT1 หรือคำว่า “ได้”

5.5.1.27 ส่วนประกอบ แยกเป็น 3 กราฟ ได้แก่ ส่วนประกอบที่เป็นคำนาม (แทนด้วย PART) ส่วนประกอบที่เป็นนามวลี (แทนด้วย PART_NP1, PART_NP2, PART_NP3, PART_NP4, PART_NP5 และ PART_NP6) และส่วนประกอบที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาหรือเป็นอนุภาค (แทนด้วย PART_VP&CL)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 13112 หรือ 1323

- นามวลีที่เกิดจาก

(1) กราฟ PART ตามด้วย

(1.1) กราฟ WHL และอาจมีคำเชื่อม “ของ” ปรากฏระหว่าง ส่วนประกอบทั้งสองได้

(1.2) กราฟ ACT_V, CHA_V, CHA_N2, CHA_NP5, DIR_VP, LOC_VP, UNIT หรือ QNT1

(1.3) กราฟ STA_V1 ซึ่งอาจตามด้วยกราฟ LOC_N

(2) กราฟ PART_NP1 ตามด้วยคำเชื่อม “ของ” แล้วตามด้วยกราฟ WHL

(3) กราฟ PART หรือ PART_NP6 ตามด้วยกราฟ LOC_PP หรือ LOC_VP

(4) กราฟ PART หรือ PART_NP2 ตามด้วยกราฟ LOC_P กราฟ CL1 หรือ CL2

- ส่วนประกอบที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาหรือเป็นอนุภาคขึ้นเกิดจากกราฟ STA_VP1 ซึ่งอาจมีคำเชื่อม “ที่” ปรากฏข้างหน้า

5.5.1.28 สสาร แยกเป็น 2 กราฟ ได้แก่ สสารที่เป็นคำนาม (แทนด้วย SUBS) และสสารที่เป็นนามวลี (แทนด้วย SUBS_NP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 1312, 13121, 13122 หรือ 13123

- นามวลีที่เกิดจากกราฟ SUBS ตามด้วยกราฟ QNT1, QNT2 หรือ

TYPE_STA1

5.5.1.29 สิ่งมีชีวิต แยกเป็น 2 กราฟ คือ สิ่งมีชีวิตที่เป็นคำนาม (แทนด้วย ANI) และสิ่งมีชีวิตที่เป็นนามวลี (แทนด้วย ANI_NP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามสามัญที่มีค่า AKO เป็น 111 หรือ 13111

- นามวลีที่เกิดจากกราฟ ANI ตามด้วยกราฟ CHA_NP4

5.5.1.30 หน่วยวัด (แทนด้วย UNIT) เกิดจาก

- คำนามสามัญที่มีค่า AKO เป็น 2355

- คำลักษณนามที่มีค่า AKO เป็น 1323, 2111, 231, 232, 2353, 2354, 2355

หรือ 2357

5.5.1.31 หน้าที่ แยกเป็น 3 กราฟ ได้แก่ หน้าที่ที่เป็นคำนาม (แทนด้วย FUNC_N) หน้าที่ที่เป็นคำกริยา (แทนด้วย FUNC_V) และหน้าที่ที่เป็นหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยา (แทนด้วย FUNC_VP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 123, 1322 หรือ 2367

- คำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 ซึ่งอาจตามด้วยคำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 อีก 1 คำก็ได้

- หน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาที่เกิดจากกราฟ FUNC_V ตามด้วยกราฟ OBJ, OBJ_NP, RES_N, LOC_P, ACT_V, STA_V2 หรือ CL2

5.5.1.32 หัวข้อ (แทนด้วย TOPIC) เกิดจากคำนามทั่วไปที่มีค่า AKO เป็น 1222, 123 หรือ 13121

5.5.1.33 อุปกรณ์ แยกเป็น 2 กราฟ ได้แก่ อุปกรณ์ที่เป็นคำนาม (แทนด้วย INS) และอุปกรณ์ที่เป็นนามวลี (แทนด้วย INS_NP)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- คำนามทั่วไปที่มีค่า AKO เป็น 1321, 1322, 1323, 2353 หรือ 2363
- นามวลีที่เกิดจากกราฟคำประสม CN27 ตามด้วยกราฟ TYPE_CHA2

จากส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33 ข้างต้น จะเห็นได้ว่า ส่วนประกอบเหล่านั้นเป็น โครงสร้างที่ประกอบด้วยคำและหน่วยสร้างที่แสดงความสัมพันธ์ทางความหมาย และเป็น โครงสร้างที่นำมาใช้เขียนกราฟย่อย ทั้งยังใช้ตั้งเป็นชื่อของกราฟย่อยด้วย คำและหน่วยสร้างเหล่านั้นมีความซับซ้อนและหลากหลายรูปแบบ ดังนั้น เพื่อให้สามารถเข้าใจคำอธิบายข้างต้น ได้ ผู้วิจัยจะนำเสนอตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ดังกล่าวตามชื่อความสัมพันธ์ทางความหมายของคำที่ปรากฏ ดังตารางข้างล่างนี้

อย่างไรก็ดี ตารางข้างล่างนี้จะไม่ปรากฏตัวอย่างของข้อ 5.5.1.8 นามวลี “การ” “ความ” ด้วย เพราะเป็นข้อที่แสดงกระบวนการกลายคำด้วยวิธีการแปลงให้เป็นนามวลี (Nominalization) กระบวนการนี้เป็นกระบวนการที่กริยาลูกถูกเปลี่ยนให้เป็นนามวลี และเป็นกระบวนการกลายคำซึ่งไม่เปลี่ยนความหมาย แต่เปลี่ยนเพียงโครงสร้างเท่านั้น (อมรา ประสิทธิ์รัฐสินธุ์, 2537) ข้อ 5.5.1.8 นี้ไม่ได้แสดงความสัมพันธ์ทางความหมาย แต่เหตุที่ผู้วิจัยได้วิเคราะห์โครงสร้างนี้ด้วยเนื่องจากต้องการให้เครื่องคอมพิวเตอร์สามารถตรวจจับคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่มีส่วนประกอบส่วนหนึ่งเป็นนามวลี “การ” “ความ” ได้ด้วย แม้ว่าคำกริยาที่ตามหลัง “การ” “ความ” จะไม่ใช่คำเดียวกับคำกริยาที่ปรากฏในคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ก็ตาม นอกจากนี้ การวิเคราะห์นามวลี “การ” “ความ” ยังช่วยให้ไม่ต้องเพิ่มนามวลี “การ” “ความ” ทุกรูปแบบที่ปรากฏในคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ลงไปในพจนานุกรมของโปรแกรม Unitex ด้วย ตัวอย่างเช่น

คำว่า “มาตรการไต้ยีน” (วรบ.) เป็นอุปกรณสำหรับวัดการไต้ยีน ปรากฏเป็น คำนามแสดงอุปกรณด้านวิทยาศาสตร์ของการวิจัยครั้งนี้ คำนี้มีโครงสร้างเป็น [อุปกรณ-วัดอุปสรรค] เกิดจากส่วนประกอบที่เป็นคำนาม คือ “มาตร” และส่วนประกอบที่เป็นนามวลี คือ “การไต้ยีน” ส่วนประกอบที่เป็นนามวลีเกิดจาก การ + คำกริยาแสดงอาการ “ไต้ยีน”

เมื่อนำโครงสร้างหลักของคำนามแสดงอุปกรณด้านวิทยาศาสตร์ไปทดสอบกับ แหล่งข้อมูลอื่น เช่น เอกสารทางการแพทย์ บทความทางวิศวกรรมศาสตร์ เป็นต้น อาจปรากฏว่า แหล่งข้อมูลนั้นมีคำนามแสดงอุปกรณด้านวิทยาศาสตร์ที่มีโครงสร้างคล้ายกับตัวอย่างได้ เช่น คำว่า “มาตรการหายใจ” ซึ่งเป็นอุปกรณทางการแพทย์สำหรับวัดการหายใจ คำนี้มีโครงสร้างเป็น [อุปกรณ-วัดอุปสรรค] เกิดจากส่วนประกอบที่เป็นคำนาม คือ “มาตร” และส่วนประกอบที่เป็น นามวลี คือ “การหายใจ” ส่วนประกอบที่เป็นนามวลีเกิดจาก การ + คำกริยาแสดงอาการ “หายใจ” ส่วนประกอบที่เป็นนามวลีนี้มีคุณสมบัติเหมือนกัน คือ “การ + คำกริยาแสดงอาการ” แต่คำกริยา แสดงอาการนั้นเป็นคนละคำกัน หากไม่ได้วิเคราะห์โครงสร้างของนามวลี “การ” “ความ” เครื่อง คอมพิวเตอร์จะไม่สามารถตรวจจับคำว่า “มาตรการหายใจ” เพราะ “การหายใจ” ไม่ปรากฏเป็น คำศัพท์ในพจนานุกรมของโปรแกรม Unitex ลักษณะดังกล่าวอาจส่งผลให้ความสามารถในการ ตรวจจับคำของ โปรแกรม Unitex คลาดเคลื่อนได้

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณด้านวิทยาศาสตร์ตามชื่อ ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33)

ที่	ความหมาย	POS	AKO	คำ
1	[ACT]	V+Act+	2111	กระตุ้น ควบคุม จุดระเบิด แยก
	[ACTION]		2124	หายใจ
	[การกระทำ อาการ]	N+Cmn+	2111	การทำงาน ออสโมซิส
			2211	แรงเคลื่อน แรงดัน
			2124	การหายใจ

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ
2	[AGT]	N+Cmn+	13122	น้ำ
	[AGENT]		1323	ก้ามปู เกือก แทน ลูกสูบ
	[ผู้ทำ]		2122	แสง ไฟ
3	[ANI]	N+Cmn+	111	เด็กอ่อน
	[ANIMATED THING] [สิ่งมีชีวิต]		13111	เชื้อ
4	[CHA]	V+Att+	2361	ต่ำ เข็ม มัน ยาว
			2363	กลม กลวง โค้ง โป่ง
	[CHARACTERISTIC] [รูปร่างลักษณะ]	N+Cmn+	13112	พื้น
			13121	ทราย
			1322	จาน ถ้วย แหวน
			234	ชุด
			2353	ขนาด ตัว รูป
			2361	ความสูง
			2363	กากบาท ทรงกระบอก ป้อม พวง
			2365	อรงค์ค์ เอกรงค์ค์
	N+Prp+	2363	ที่	
DET+Crd+	2365	ครึ่ง		

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ
5	[COMP] [COMPLEMENT] [ผู้เสริม]	N+Cmn+	13111	เชื้อ
			1312	กรด เบส
			1323	ไดอะแฟรม ยางใน
			2111	การบำรุงรักษา
			2361	ความคลาด
6	[DIR] [DIRECTION] [ทิศทาง]	N+Cmn+	231	ทาง ทิศทาง แนว รัศมี
		PREP+Nrm+	231	หน้า
		PREP+Vrb+	231	ตาม
		V+Act+	2111	กลับ ไชวี้ ตั้ง นอน
		V+Aux+av+e+	2111	ขึ้น เข้า ลง
		V+Att+	2361	คิง
7	[FUNC] [FUNCTION] [หน้าที่]	V+Act+	2111	กรอง กำเนิด ด้านทาน รับรู้
		N+Cmn+	123	สะพาน
			1322	ตะเกียง เทียน สว่าน
			2215	ตรา
			231	ทาง
8	[GRT] [GREATNESS] [ความยิ่งใหญ่ ความสำคัญ]	N+Cmn+	2361	แม่

ตารางที่ 17 ตัวอย่างของส่วนประกอบของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ
9	[INS] [INSTRUMENT] [อุปกรณ์]	N+Cmn+	13122	น้ำ น้ำมัน
			1321	กระดาษ ผ้า ไม้ เหล็ก
			1322	กล้อง เครื่อง ตะเกียง บารอมิเตอร์
			1323	เฟือง
			231	ที่
			234	ชุด
			2353	ตัว
			2363	ก้าน เต้า สาย หลอด
10	[LOC] [LOCATION] [สถานที่]	PREP+Nrm+	231	กลาง ใต้ ใน ภายนอก
		DET+Crd+	2365	ศูนย์
		DET+Ord+	2365	ที่ห้า
		N+Cmn+	1221	ดิน บรรยากาศ อากาศ
			1222	เส้นศูนย์เที่ยง
			123	พื้น
			13112	แขน ช่องปาก เส้นเลือดแดง
			1322	รถ
			1323	กระบอกสูบ ผิว ฝาสูบ ลิ้นไอเสีย
			2211	แผล
			231	โซน ตำแหน่ง มุม สนาม
			2366	หลัก

ตารางที่ 17 ตัวอย่างของส่วนประกอบของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ
11	[MAN]	ADV+Nrm+	2361	กัน
	[MANNER]		2362	อย่างละเอียด
	[ลักษณะการกระทำ]	ADV+Nrm+	2365	ยวดยิ่ง ยิ่งยวด
		V+Att+	2361	เร็ว
12	[MAT] [MATERIAL] [วัสดุ]	N+Cmn+	1321	กระเบื้องขาว แก้ว เซรามิก ดินขาว ผ้า แพลทินัม ยาง ลวด วัสดุ เหล็ก
13	[MET] [METHOD] [วิธีการ]	V+Act+	2111	กด กวาง ดูก ฟัน
		V+Att+	2363	ฟู
		N+Cmn+	1222	เส้นศูนย์เที่ยง
			13112	เท้า มือ เส้นเลือด
			1312	อะตอม อิเล็กตรอน
			13121	ควอดซ์ ทราบ
			13122	แกโซลีน ของเหลว ดีเซล น้ำ
			13123	แก๊ส ลม อากาศ ไอ้
			1322	จักร เทอร์มอสแตต
			1323	ไดอะแฟรม ใบพัด เฟือง ลูกสูบ
			2111	โทรทรรศน์ ออกซิเจน แอนิรอยด์
			2122	กระแสไฟฟ้า คลื่นสั้น ไฟฟ้า เลเซอร์
			2215	โฟโตแกรมเมตรี เรเดียม
			231	อากาศ
2351	เชิง			
2362	โซลิดสเตต สูญญากาศ อุณหภูมิ			

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ
14	[NAME]	N+Prp+	111	ไกเกอร์มูลเลอร์ เทศกู วังเกล
	[NAME]		1322	กอล เคเบิต ซีโอดิสไต์
	[ชื่อเฉพาะ]		1323	คีโอเอชซี โอเอชซี โอเอชวี
			2111	ซีไอ ออกซิเดชั่น เอสไอ ไอซี
			2215	วี เอฟ แอล Y
15	[NUM]	DET+Crd+	2365	2 8 สาม สี่
	[NUMBER]	DET+Ord+	2365	เดียว
	[จำนวน]	DET+Def+bq+	2355	เฉพาะ
		DET+Ind+bq+	2355	บาง
16	[OBJ]	N+Cmn+	1222	ดวงอาทิตย์ ดาว
	[OBJECTIVE]		13111	เชื้อโรค
	[ผู้ถูก]		13112	ท่อไต กรวยไต กระดูก
			1312	เชื้อเพลิง ยา ยาสลบ สาร
			13121	แก้ว จาระบี ฝุ่น ลวด
			13122	ของเหลว โซดา น้ำ น้ำเกลือ
			13123	แก๊ส ออกซิเจน อากาศ ไขมัน
			132	วัตถุ
			1322	กระบอก บิวเรท รถ หลอด
			1323	กระบอกสูบ เผลา ล้อ
			2111	ปฏิกิริยา
			2122	กัมมันตภาพรังสี คลื่นวิทยุ ไฟฟ้า
			2124	คลื่นไฟฟ้าหัวใจ ซีพจร
	2211	กำลัง งาน แรง แรงดัน		

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ				
16	[OBJ]	N+Cmn+	2215	ข้อมูล	ปริพันธ์	มาตรฐาน		
	[OBJECTIVE]		231	ตำแหน่ง	ทิศทาง	พิสัย	มุม	
	[ผู้ถูก]		232	เวลา				
			2353	ขนาด	ภาพ	สี	แสง	
			2355	ค่า	ครรชนีหักเห	น้ำหนัก	ส่วนสูง	
			2361	ความเร็ว				
			2362	ความชื้น	ความร้อน			
			2363	ทัศนวิสัย	แผ่น			
17	[PART]	N+Cmn+	13112	ตา	ไต	ปอด	ปาก	
	[PART]		1323	แกน	ข้อเหวี่ยง	คอ	เฟือง	
18	[ส่วนประกอบ]							
	[PAT]	N+Cmn+	1323	ก้ามปู	คอคอด	นมหนู	ผิว	
	[PATIENT]		2122	ไฟ	แสง	อุณหภูมิต		
	[ผู้ทรงสภาพ]		2211	แรง				
			231	ทาง				
			2353	ขนาด	วิว			
			2355	อัตราเร็ว				
			2362	สูญญากาศ				
	2363		แผ่น					

ตารางที่ 17 ตัวอย่างของส่วนประกอบของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ
19	[PHE]	N+Cmn+	2122	กระแส กัมมันตรังสี กลิ่น เอกซเรย์
	[PHENOMENA]		2353	ลำแสง
	[ปรากฏการณ์ธรรมชาติ]	N+Prp+	2122	ดีซี เอซี
20	[PUR]	V+Act+	2111	กด ควบแน่น จำย ชั่ง
	[PURPOSE]	N+Cmn+	1221	ดิน
	[วัตถุประสงค์]		1312	เชื้อเพลิง สารหล่อเย็น
			13121	น้ำตาล แคลเซียมคลอไรด์
			13122	น้ำมันเครื่อง น้ำหล่อเย็น
			13123	ลม อากาศ ไอดี ไอเสีย
			1322	ฟิวส์ สำลี
			1323	ไดอะแฟรม ล้อ
			2111	จุลทรรศน์ ปฏิกรณ์ ปรีทรศน์
			2122	กระแส ไฟฟ้า รังสี
			2211	กำลังไฟฟ้า แรงดึง
			2215	ทันตกรรม ฟลักซ์ มุมดเวลล์ อักษร
			231	ทิส พื้นที่ ระดับ สนาม
			232	เวลา
			234	ส่วนบุคคล
			2353	ภาพ
			2355	อัตรารอบ อัตราเร็ว อัตรากาไรไหล
			2362	ความสว่าง ความหนืด นิรภัย

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ	
21	[QNT]	N+Cmn+	2122	อุณหภูมิ	
	[QUANTITY] [ปริมาณ]		2355	ช่วงชัก รอบ ระยะทาง สักย	
22	[RES]	N+Cmn+	13111	เชื้อจุลินทรีย์ เชื้อแบคทีรี	
	[RESULT]		13112	ฟัน	
	[ผล]		1312	เชื้อเพลิง อนุภาค	
			13121	ละออง	
			13123	แก๊ส	
			1322	ปากกา แผนที่	
			1323	โครง	
			2122	ไฟฟ้าสลับ สัญญาณกวาด แสง	
			2124	ซีพจร	
			2211	สำเนา	
			234	แถว เอกเทศ	
			2353	ภาพ	
			2362	ความเย็น ความร้อน	
			N+Prp+	2122	เอกซเรย์
			V+Att+	2361	บาง
	V+Sta+	2362	ขนาน เย็น		
	V+Act+	2111	ระเหย แล่น สลับที่		
23	[SRC] [SOURCE] [ที่มา]	N+Cmn+	13112	ปาก	

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคํานามแสดงอุปกรณัด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ	
24	[STA]	V+Sta+	233	มี	
	[STATE]		2362	ขาด แข็ง คงตัว เหนียว	
	[สภาพ]		2366	เท่า	
			V+Att+	2361	ไกล สูง
			V+Att+	2363	กว้าง เล็ก
			N+Cmn+	13122	น้ำมัน
				1322	กาว
				2211	ความกดดัน ความดัน
				2355	ความถี่
				2361	ความตึง ความเร็ว ความสูง
				2362	ความชื้น ความร้อน จุด สุญญากาศ
				2363	ความเข้ม
	25	[SUBS]	N+Cmn+	1312	ของไหล เชื้อเพลิง
[SUBSTANCE]		13121		แม่เหล็กไฟฟ้า	
[สาร]		13122		ของเหลว น้ำฝน เลือด	
		13123		ไนโตรสออกไซด์ ออกซิเจน ไอ	
26	[SYS]	N+Cmn+	2211	วัฏจักรออกโต โรทารี	
	[SYSTEM]		2215	กล เทอร์โบ แอเดียแบติก ไฮดรอลิก	
	[ระบบ]	N+Prp+	111	Kipp	
			2215	แลมบ์ดา	
		ADV+Pfx+	2361	โดยอัตโนมัติ	
		ADV+Nrm+	2361	เอง	

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อ
ความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ
27	[THN]	N+Cmn+	1222	ดาว
	[THING]		1321	กระจก กระดาษ ไม้
	[วัตถุ]		1322	พาหนะ ภาพ маты รถ
			2215	อักษร
28	[TIME]	V+Act+	2111	เร่ง
	[TIME] [เวลา]	CONJ+Sbr+	232	ก่อน ขณะ
29	[TOPIC]	N+Cmn+	1222	ดาว
	[TOPIC]		123	ห้องฟ้าจำลอง
	[หัวข้อ]		13121	โลหะ
30	[TYPE] [TYPE] [ประเภท]	N+Cmn+	2354	ชนิด ด้าน แบบ
31	[UNIT]	N+Cmn+	2355	เมตร
	[UNIT]	CLAS+	1323	คอ ตา ท่อ รุ
	[หน่วยวัด]		2111	จังหวัด
			231	ทาง
			2353	ช่อง ตอน ตัว ส่วน
			2354	ชนิด
			2355	ครั้ง เมตร
			2357	ทาง

ตารางที่ 17 ตัวอย่างของส่วนประกอบของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ตามชื่อความสัมพันธ์ทางความหมายของคำที่ปรากฏ (ข้อ 5.5.1.1 ถึงข้อ 5.5.1.33) (ต่อ)

ที่	ความหมาย	POS	AKO	คำ
32	[WHL]	N+Cmn+	13111	กา ช้าง ปู แร้ง
	[WHOLE]		1312	สาร
	[ทั้งส่วน]		1322	กรรไกร กุญแจ นาฬิกา เบรก
			1323	กระบอกสูบ
			232	ชั่วโมง

5.5.1.34 อนุพากย์ แยกเป็น 3 กราฟ ได้แก่ อนุพากย์แบบที่ 1 (แทนด้วย CL1) อนุพากย์แบบที่ 2 (แทนด้วย CL2) และอนุพากย์แบบที่ 3 (แทนด้วย CL3)

คุณสมบัติของคำหรือหน่วยสร้างที่ปรากฏในโครงสร้างมีดังต่อไปนี้

- กราฟ PAT ตามด้วยกราฟ STA_V1 หรือ STA_VP3
- กราฟ AGT ตามด้วยกราฟ ACT_V หรือ ACT_VP2
- กราฟ OBJ ตามด้วยกราฟ ACT_V

5.5.1.35 คำประสม ปรากฏเป็นทั้งโครงสร้างหลักและโครงสร้างส่วนประกอบของโครงสร้างคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ ในที่นี้ จะขึ้นต้นชื่อกราฟคำประสมทั้งหมดด้วยคำว่า CN เพราะ CN มาจากคำว่า “Compound Nouns” ซึ่งหมายถึงคำนามประสม และคำประสมที่ปรากฏในคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์เป็นคำนามทั้งหมดด้วย

กราฟคำประสมเกิดจากการรวมกราฟแสดงความสัมพันธ์ทางความหมายเข้าด้วยกันเป็นโครงสร้างต่าง ๆ กราฟคำประสมแยกได้เป็น 30 กราฟ คือ คำประสมแบบที่ 1 ถึงคำประสมแบบที่ 30 ดังแสดงในตารางข้างล่างนี้

ตารางที่ 18 โครงสร้างของส่วนประกอบที่เป็นคำประสม

แบบ ที่	ชื่อ กราฟ	โครงสร้าง ความสัมพันธ์	โครงสร้าง	ประเภทย่อย	
				ชื่อกราฟ	โครงสร้าง
1	CN1	[อุปกรณ์ - วัตถุประสงค์]		CN1.1	INS + PUR_V
				CN1.2	INS + PUR_N
				CN1.3	INS + PUR_NP CN1.2
				CN1.4	INS + PUR_VP1 / PUR_VP2 / PUR_VP3 / PUR_VP5
				CN1.5	CN16 + PUR_N
				CN1.6	CN27 + PUR_N
				CN1.7	CN2.1 + PUR_V CN2.1/ CN15/ + PUR_VP1 CN27
				CN1.8	CN1.1+PUR_N
2	CN2	[อุปกรณ์ - วิธีการ]		CN2.1	INS + MET_N
				CN2.2	INS + MET_V
				CN2.3	INS + QNT1 QNT2
				CN2.4	INS + MET_VP3 CN1.1 CN1.4 + MET_VP1
				CN2.5	CN1.4 + MET_CL
3	CN3	[อุปกรณ์ - หน้าที่]		CN3.1	INS + FUNC_V
				CN3.2	INS + FUNC_VP
				CN3.3	CN27 + FUNC_V
4	CN4	[อุปกรณ์ - สภาพ]	INS + STA_N CN1.1		

ตารางที่ 18 โครงสร้างของส่วนประกอบที่เป็นคำประสม (ต่อ)

แบบ ที่	ชื่อ กราฟ	โครงสร้าง ความสัมพันธ์	โครงสร้าง	ประเภทย่อย	
				ชื่อกราฟ	โครงสร้าง
5	CN5	[อุปกรณ์ - ชื่อเฉพาะ]	INS / CN2.1 / CN7.3 / CN23	+ NAME	
6	CN6	[อุปกรณ์ - วัสดุ]	INS CN2.2	+ MAT	
7	CN7	[อุปกรณ์ - รูปร่างลักษณะ]			CN7.1 INS + CHA_N2 CHA_V
				CN7.2	INS + CHA_N3
				CN7.3	CN27 + CHA_V
				CN7.4	INS + CN26.1
				CN7.5	INS + PART_NP2
8	CN8	[อุปกรณ์ - สถานที่]	INS + LOC_N LOC_P		
9	CN9	[อุปกรณ์ - ผล]	CN1.2 CN10.1	+ RES_NP	
10	CN10	[อุปกรณ์ - ระบบ]			CN10.1 INS CN1.2 + SYS_N
				CN10.2	INS / CN1.4 / CN2.1 + QNT1
				CN10.3	INS + SYS_A&V
11	CN11	[อุปกรณ์ - ทิศทาง]	INS + DIR_V QNT2		

ตารางที่ 18 โครงสร้างของส่วนประกอบที่เป็นคำประสม (ต่อ)

แบบ ที่	ชื่อ กราฟ	โครงสร้าง ความสัมพันธ์	โครงสร้าง	ประเภทย่อย	
				ชื่อกราฟ	โครงสร้าง
12	CN12	[จำนวน - ส่วนประกอบ]	QNT1		
13	CN13	[ระบบ-อุปกรณ์]	SYS_N + INS		
14	CN14	[หน้าที่ - วัตถุประสงค์]	FUNC_N + PUR_N		
15	CN15	[วัสดุ - ชื่อเฉพาะ]	MAT + NAME		
16	CN16	[ความยิ่งใหญ่ ความสำคัญ - การกระทำ]	GRT + ACT_N ACT_V		
17	CN17	[วัตถุ-สภาพ]	THN + STA_V2		
18	CN18	[วัตถุประสงค์ - ผู้ถูก]	PUR_V + OBJ OBJ_CN		
19	CN19	[วัตถุประสงค์ - การกระทำ]	PUR_V + ACT_V		
20	CN20	[วัตถุประสงค์ - วิธีการ]	PUR_V + MET_V		
21	CN21	[ทั้งส่วน - ส่วนประกอบ]	WHL + CL1 CL2		

ตารางที่ 18 โครงสร้างของส่วนประกอบที่เป็นคำประสม (ต่อ)

แบบ ที่	ชื่อ กราฟ	โครงสร้าง ความสัมพันธ์	โครงสร้าง	ประเภทย่อย	
				ชื่อกราฟ	โครงสร้าง
22	CN22	[ส่วนประกอบ - วัตถุประสงค์]	PART + PUR_V PUR_VP1		
23	CN23	[ส่วนประกอบ - หน้าที่]	PART + FUNC_N		
24	CN24	[ส่วนประกอบ - ทั้งส่วน]	PART + WHL		
25	CN25	[รูปร่างลักษณะ - รูปร่างลักษณะ]		CN25.1	CHA_N3 + CHA_V
				CN25.2	CHA_N3 + CHA_N2
26	CN26	[รูปร่างลักษณะ - วัตถุ]	CHA_N3 + THN		
27	CN27	[รูปร่างลักษณะ - วัสดุ]	CHA_N2 + MAT CHA_N3		
28	CN28	[รูปร่างลักษณะ - อุปกรณ์]	CHA_N3 + INS / CN1.2 / CN1.3 / CN1.4 / CN10.2 / CN10.3 / CN24 / CN26		
29	CN29	[รูปร่างลักษณะ - สภาพ]	CHA_N3 + STA_N		
30	CN30	[อุปกรณ์ - ส่วนประกอบ]	INS + QNT1 QNT2 CN1.1 + QNT1 CN1.2 PART_N		

5.5.2 กฎโครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์

กฎโครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์เป็น โครงสร้างที่เกิดจากการรวมส่วนประกอบของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์เข้าด้วยกัน ส่วนประกอบของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์แสดงด้วยกราฟย่อย กฎโครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์เป็นกฎที่ใช้เขียนเป็นกราฟหลัก ดังนั้น กราฟหลักจึงเกิดจากการรวมกราฟย่อยหลาย ๆ กราฟเข้าด้วยกันตามแต่ละกฎโครงสร้าง

สำหรับชื่อกราฟหลัก ผู้วิจัยใช้ชื่อว่า ST ซึ่งย่อมาจาก Structure (แปลว่า โครงสร้าง) เพราะโครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์มีความซับซ้อนเป็นลำดับชั้น และเกิดจากการรวมโครงสร้างย่อยหลายรูปแบบเข้าด้วยกัน ทำให้มีลักษณะคล้ายใยแมงมุม อย่างไรก็ตาม หากโครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์เป็นหน่วยสร้างคำประสมที่ไม่ค่อยซับซ้อน และมีลักษณะเช่นเดียวกับโครงสร้างส่วนประกอบที่เป็นกราฟย่อย ชนิดคำประสม (มีชื่อว่า CN) อยู่แล้ว ผู้วิจัยจะใช้ชื่อ CN ประเภทนั้น ๆ มาเป็นชื่อกราฟหลักของโครงสร้างหลักที่มีลักษณะดังกล่าว เช่น กฎโครงสร้างหลักที่มีความสัมพันธ์แบบ [รูปร่างลักษณะ – วัตถุ] ตรงกับโครงสร้างคำประสมแบบ CN26 ผู้วิจัยจะใช้ชื่อ CN26 มาเป็นชื่อกราฟหลัก ทั้งนี้ เพื่อลดความซ้ำซ้อนของโครงสร้างและช่วยให้เครื่องคอมพิวเตอร์สามารถประมวลผลได้สะดวกขึ้น

กฎโครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์สามารถนำมาสร้างกราฟได้ 37 กราฟ ดังต่อไปนี้

5.5.2.1 [อุปกรณ์-วัตถุประสงค์] (แทนด้วย ST1) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟคำประสมต่าง ๆ ตามด้วยกราฟวัตถุประสงค์ทุกแบบ (แทนด้วย PUR_N, PUR_V, PUR_NP, PUR_VP1-8, PUR_CL) ดังนี้

- กราฟคำประสม CN1.1, CN1.2, CN1.4, CN1.7 หรือ CN1.8
- กราฟคำประสม CN1.1, CN1.2, CN1.4 หรือ CN6 ตามด้วยกราฟ PUR_N
- กราฟคำประสม CN1.1, CN1.2, CN6 หรือ CN7.5 ตามด้วยกราฟ PUR_V
- กราฟ INS หรือกราฟคำประสม CN1.1, CN1.2, CN1.4, CN1.8, CN2.4, CN4, CN6, CN7.1, CN7.5, CN13, CN22 หรือ CN29 ตามด้วยกราฟ PUR_VP1

- กราฟ INS หรือกราฟคำประสม CN1.2, CN2.1, หรือ CN6 ตามด้วยกราฟ PUR_VP2
- กราฟ INS หรือกราฟคำประสม CN2.1 ตามด้วยกราฟ PUR_VP3, PUR_VP5, PUR_VP6 หรือ PUR_VP8
- กราฟ INS หรือกราฟคำประสม CN1.7 หรือ CN2.1 ตามด้วยกราฟ PUR_VP4
- กราฟ INS หรือกราฟคำประสม CN1.2, CN2.1, CN6 หรือ CN9 ตามด้วยกราฟ PUR_VP7
- กราฟคำประสม CN17 ตามด้วยกราฟ PUR_CL
- กราฟ INS ตามด้วยกราฟ PUR_CN หรือ PUR_NP
- กราฟ INS หรือกราฟคำประสม CN6 หรือ CN29 ตามด้วยกราฟ PUR_V แล้วตามด้วยกราฟ PUR_VP1

ทั้งหมดแสดงเป็นกฎได้ว่า

ST1 → CN1.1 / CN1.2 / CN1.4 / CN1.7 / CN1.8
 CN1.1 / CN1.2 / CN1.4 / CN6 + PUR_N
 CN1.1 / CN1.2 / CN6 / CN7.5 + PUR_V

INS / CN1.1 / CN1.2 / CN1.4 / CN1.8 / CN2.4 / CN4 / CN6 / CN7.1 / CN7.5 / CN13 / CN22 / CN29	+ PUR_VP1
---	-----------

 INS / CN1.2 / CN2.1 / CN6 + PUR_VP2
 INS / CN2.1 + PUR_VP3 / PUR_VP5 / PUR_VP6 / PUR_VP8
 INS / CN1.7 / CN2.1 + PUR_VP4
 INS / CN1.2 / CN2.1 / CN6 / CN9 + PUR_VP7
 CN17 + PUR_CL
 INS + PUR_CN / PUR_NP
 INS / CN6 / CN29 + PUR_V + PUR_VP1

5.5.2.2 [อุปกรณ์-วิธีการ] (แทนด้วย ST2) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟคำประสมต่าง ๆ ตามด้วยกราฟวิธีการทุกแบบ (แทนด้วย MET_N, MET_V, MET_NP, MET_VP, MET_CL) รวมถึงกราฟอื่น ๆ ที่เกี่ยวข้อง ดังนี้

- กราฟคำประสม CN2.1, CN2.2, CN2.3, CN2.4 หรือ CN2.5
- กราฟคำประสม CN1.1, CN1.2, CN1.4, CN2.1, CN8 หรือ CN25.1 ตาม

ด้วยกราฟ MET_N

- กราฟ INS หรือกราฟคำประสม CN1.1 หรือ CN2.1 ตามด้วยกราฟ

MET_CN

- กราฟคำประสม CN5 ตามด้วยกราฟ MET_V
- กราฟ INS ตามด้วยกราฟ MET_NP หรือ MET_PP
- กราฟคำประสม CN2.2 ตามด้วยกราฟ MET_VP1
- กราฟคำประสม CN1.1 หรือ CN1.2 ตามด้วยกราฟ MET_VP2
- กราฟคำประสม CN1.2 หรือ CN1.4 ตามด้วยกราฟ MET_CL
- กราฟคำประสม CN1.2 ตามด้วยกราฟ DIR_NP1 หรือ PHE_NP
- กราฟคำประสม CN1.1 ตามด้วยกราฟ STA_NP
- กราฟคำประสม CN2.5 ตามด้วยกราฟ QNT1
- กราฟคำประสม CN18 ตามด้วยกราฟ CL1
- กราฟ INS ตามด้วยกราฟ CL3

ทั้งหมดแสดงเป็นกฎได้ว่า

ST2 → CN2.1 / CN2.2 / CN2.3 / CN2.4 / CN2.5
 CN1.1 / CN1.2 / CN1.4 / CN2.1 / CN8 / CN25.1 + MET_N
 INS / CN1.1 / CN2.1 + MET_CN
 CN5 + MET_V
 INS + MET_NP / MET_PP
 CN2.2 + MET_VP1
 CN1.1 / CN1.2 + MET_VP2
 CN1.2 / CN1.4 + MET_CL
 CN1.2 + DIR_NP1 / PHE_NP
 CN1.1 + STA_NP
 CN2.5 + QNT1
 CN18 + CL1
 INS + CL3

5.5.2.3 [อุปกรณ์-ประเภทย่อย] (แทนด้วย ST3) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟค่าประสม CN1.1, CN1.2, CN1.3, CN1.4, CN1.5, CN2.1, CN3.1, CN3.2, CN7.1, CN11, CN16, CN18 หรือ CN22 ตามด้วยกราฟประเภทย่อยแบบที่ 1 (แทนด้วย TYPE_NP1) แสดงเป็นกฎได้ว่า

$$ST3 \rightarrow \boxed{\begin{array}{l} INS / CN1.1 / CN1.2 / CN1.3 / CN1.4 / \\ CN1.5 / CN2.1 / CN3.1 / CN3.2 / CN7.1 / \\ CN11 / CN16 / CN18 / CN22 \end{array}} + TYPE_NP1$$

5.5.2.4 [อุปกรณ์-หน้าที่] (แทนด้วย ST4) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟค่าประสมต่าง ๆ ตามด้วยกราฟหน้าที่ ดังนี้

- กราฟค่าประสม CN3.1, CN3.2 หรือ CN3.3
- กราฟ INS หรือกราฟค่าประสม CN1.2 ตามด้วยกราฟ FUNC_N ทั้งหมดแสดงเป็นกฎได้ว่า

$$ST4 \rightarrow \begin{array}{l} CN3.1 / CN3.2 / CN3.3 \\ INS / CN1.2 + FUNC_N \end{array}$$

5.5.2.5 [อุปกรณ์-รูปร่างลักษณะ] (แทนด้วย ST5) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟค่าประสมต่าง ๆ ตามด้วยกราฟรูปร่างลักษณะทุกแบบ (แทนด้วย CHA_N, CHA_V, CHA_NP, CHA_VP) ดังนี้

- กราฟค่าประสม CN7.1
- กราฟค่าประสม CN1.1, CN1.2, CN1.4 หรือ CN7.1 ตามด้วยกราฟ CHA_N3

- กราฟค่าประสม CN19 ตามด้วยกราฟ CHA_V
 - กราฟ INS ตามด้วยกราฟ CHA_NP5 หรือ CHA_NP8
 - กราฟค่าประสม CN1.1 ตามด้วยกราฟ CHA_NP6
 - กราฟ INS หรือกราฟค่าประสม CN1.1 ตามด้วยกราฟ CHA_NP7
 - กราฟค่าประสม CN1.1 หรือ CN6 ตามด้วยกราฟ CHA_VP
- ทั้งหมดแสดงเป็นกฎได้ว่า

ST5 → CN7.1
 CN1.1 / CN1.2 / CN1.4 / CN7.1 + CHA_N3
 CN19 + CHA_V
 INS + CHA_NP5 / CHA_NP8
 CN1.1 + CHA_NP6
 INS / CN1.1 + CHA_NP7
 CN1.1 / CN6 + CHA_VP

5.5.2.6 [อุปกรณ์-ระบบ] (แทนด้วย ST6) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟค่าประสมต่าง ๆ ตามด้วยกราฟระบบ (แทนด้วย SYS_A&V, SYS_N, SYS_VP) ดังนี้

- กราฟค่าประสม CN10.1 หรือ CN10.2
- กราฟ INS หรือกราฟค่าประสม CN1.4, CN1.7, CN2.1 หรือ CN14 ตามด้วยกราฟ SYS_A&V
- กราฟค่าประสม CN2.1, CN3.1, CN3.2 หรือ CN10.1 ตามด้วยกราฟ SYS_N
- กราฟ INS หรือกราฟค่าประสม CN1.4 หรือ CN2.1 ตามด้วยกราฟ SYS_VP

ทั้งหมดแสดงเป็นกฎได้ว่า

ST6 → CN10.1 / CN10.2
 INS / CN1.4 / CN1.7 / CN2.1 / CN14 + SYS_A&V
 CN2.1 / CN3.1 / CN3.2 / CN10.1 + SYS_N
 INS / CN1.4 / CN2.1 + SYS_VP

5.5.2.7 [อุปกรณ์-สถานที่] (แทนด้วย ST7) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟค่าประสมต่าง ๆ ตามด้วยกราฟสถานที่ (แทนด้วย LOC_D, LOC_VP, LOC_PP) ดังนี้

- กราฟค่าประสม CN8
- กราฟ INS ตามด้วยกราฟ LOC_D
- กราฟค่าประสม CN1.2 ตามด้วยกราฟ LOC_VP

กราฟ LOC_PP

- กราฟ INS หรือกราฟค่าประสม CN1.4, CN1.7 หรือ CN16 ตามด้วย

ทั้งหมดแสดงเป็นกฎได้ว่า

ST7 → CN8
 INS + LOC_D
 CN1.2 + LOC_VP
 INS / CN1.4 / CN1.7 / CN16 + LOC_PP

5.5.2.8 [อุปกรณ์-สภาพ] (แทนด้วย ST8) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟค่าประสมต่าง ๆ ตามด้วยกราฟสภาพ (แทนด้วย STA_V, STA_VP) รวมถึงกราฟอื่น ๆ ที่เกี่ยวข้อง ดังนี้

- กราฟค่าประสม CN4
- กราฟ INS หรือกราฟค่าประสม CN1.1, CN1.7, CN2.2, CN15 หรือ CN23 ตามด้วยกราฟ SAT_V1 หรือ STA_VP1

- กราฟค่าประสม CN1.2 ตามด้วยกราฟ PHE_NP

- กราฟค่าประสม CN2.1 ตามด้วยกราฟ QNT_NP

ทั้งหมดแสดงเป็นกฎได้ว่า

ST8 → CN4
 INS / CN1.1 / CN1.7 / CN2.2 / CN15 / CN23 + SAT_V1 / STA_VP1
 CN1.2 + PHE_NP
 CN2.1 + QNT_NP

5.5.2.9 [อุปกรณ์-ชื่อเฉพาะ] ตรงกับกราฟค่าประสมแบบที่ 5 ซึ่งแทนด้วย CN5 เกิดจากกราฟ INS หรือกราฟค่าประสม CN2.1 CN7.3 หรือ CN23 ตามด้วยกราฟ NAME แสดงเป็นกฎได้ว่า

CN5 → INS / CN2.1 / CN7.3 / CN23 + NAME

5.5.2.10 [อุปกรณ์-วัสดุ] ตรงกับกราฟค่าประสมแบบที่ 6 ซึ่งแทนด้วย CN6 เกิดจากกราฟ INS หรือกราฟค่าประสม CN2.2 ตามด้วยกราฟ MAT แสดงเป็นกฎได้ว่า

CN6 → INS / CN2.2 + MAT

5.5.2.11 [อุปกรณ์-ผล] ตรงกับกราฟคำประสมแบบที่ 9 ซึ่งแทนด้วย CN9 เกิดจากกราฟคำประสม CN1.2 หรือ CN10.1 ตามด้วยกราฟ RES_NP แสดงเป็นกฎได้ว่า

$$\text{CN9} \rightarrow \text{CN1.2 / CN10.1 + RES_NP}$$

5.5.2.12 [อุปกรณ์-ทิศทาง] ตรงกับกราฟคำประสมแบบที่ 11 ซึ่งแทนด้วย CN11 เกิดจากกราฟ INS ตามด้วยกราฟ DIR_V หรือ QNT2 แสดงเป็นกฎได้ว่า

$$\text{CN11} \rightarrow \text{INS + DIR_V / QNT2}$$

5.5.2.13 [อุปกรณ์-เวลา] (แทนด้วย ST9) เกิดจากกราฟคำประสม CN1.1 ตามด้วยกราฟเวลาแบบที่ 1 (แทนด้วย TIME1) แสดงเป็นกฎได้ว่า

$$\text{ST9} \rightarrow \text{CN1.1 + TIME1}$$

5.5.2.14 [รูปร่างลักษณะ-อุปกรณ์] ตรงกับกราฟคำประสมแบบที่ 28 ซึ่งแทนด้วย CN28 เกิดจากกราฟ CHA_N3 ตามด้วยกราฟ INS หรือกราฟคำประสม CN1.2 CN1.3 CN1.4 CN10.2 CN10.3 CN24 หรือ CN26 แสดงเป็นกฎได้ว่า

$$\text{CN28} \rightarrow \text{CHA_N3 + } \left\{ \begin{array}{l} \text{INS / CN1.2 / CN1.3 / CN1.4 / CN10.2 /} \\ \text{CN10.3 / CN24 / CN26} \end{array} \right.$$

5.5.2.15 [อุปกรณ์-ส่วนประกอบ] ตรงกับกราฟคำประสมแบบที่ 30 ซึ่งแทนด้วย CN30 เกิดจากกราฟ INS ตามด้วยกราฟ QNT1 หรือ QNT2 หรือกราฟคำประสม CN1.1 หรือ CN1.2 ตามด้วยกราฟ QNT1 หรือ PART_N แสดงเป็นกฎได้ว่า

$$\text{CN30} \rightarrow \text{INS + QNT1 / QNT2} \\ \text{CN1.1 / CN1.2 + QNT1 / PART_N}$$

5.5.2.16 [รูปร่างลักษณะ-วัสดุ] ตรงกับกราฟคำประสมแบบที่ 27 ซึ่งแทนด้วย CN27 เกิดจากกราฟ CHA_N2 หรือ CHA_N3 ตามด้วยกราฟ MAT แสดงเป็นกฎได้ว่า

$$\text{CN27} \rightarrow \text{CHA_N2 / CHA_N3 + MAT}$$

5.5.2.17 [รูปร่างลักษณะ-รูปร่างลักษณะ] ตรงกับกราฟค่าประสมแบบที่ 25.2 ซึ่งแทนด้วย CN25.2 เกิดจากกราฟ CHA_N3 ตามด้วยกราฟ CHA_N2 แสดงเป็นกฎได้ว่า

$$\text{CN25.2} \rightarrow \text{CHA_N3} + \text{CHA_N2}$$

5.5.2.18 [รูปร่างลักษณะ-วัตถุ] ตรงกับกราฟค่าประสมแบบที่ 26 ซึ่งแทนด้วย CN26 เกิดจากกราฟ CHA_N3 ตามด้วยกราฟ THN แสดงเป็นกฎได้ว่า

$$\text{CN26} \rightarrow \text{CHA_N3} + \text{THN}$$

5.5.2.19 [ทั้งส่วน-ส่วนประกอบ] (แทนด้วย ST10) เกิดจากกราฟแสดง [ทั้งส่วน] ตามด้วยกราฟแสดง [ส่วนประกอบ] ดังนี้

- กราฟค่าประสม CN21
- กราฟค่าประสม CN2.1 ตามด้วยกราฟ PART_NP2, PART_NP5, CL1 หรือ CL2
- กราฟค่าประสม CN6, CN10.1 หรือ CN23 ตามด้วยกราฟ PART_VP&CL
- กราฟ INS หรือกราฟค่าประสม CN1.1 หรือ CN2.1 ตามด้วยกราฟ QNT1 หรือ QNT2

ทั้งหมดแสดงเป็นกฎได้ว่า

$$\begin{aligned} \text{ST10} \rightarrow & \text{CN21} \\ & \text{CN2.1} + \text{PART_NP2} / \text{PART_NP5} / \text{CL1} / \text{CL2} \\ & \text{CN6} / \text{CN10.1} / \text{CN23} + \text{PART_VP\&CL} \\ & \text{INS} / \text{CN1.1} / \text{CN2.1} + \text{QNT1} / \text{QNT2} \end{aligned}$$

5.5.2.20 [ส่วนประกอบ-วัตถุประสงค์] ตรงกับกราฟค่าประสมแบบที่ 22 ซึ่งแทนด้วย CN22 เกิดจากกราฟ PART ตามด้วยกราฟ PUR_V หรือ PUR_VP1 แสดงเป็นกฎได้ว่า

$$\text{CN22} \rightarrow \text{PART} + \text{PUR_V} / \text{PUR_VP1}$$

5.5.2.21 [ส่วนประกอบ-ทั้งส่วน] (แทนด้วย ST11) เกิดจากกราฟแสดง [ส่วนประกอบ] ตามด้วยกราฟแสดง [ทั้งส่วน] ดังนี้

- กราฟคำประสม CN24
 - กราฟคำประสม CN18 ตามด้วยกราฟคำประสม CN2.1
- ทั้งหมดแสดงเป็นกฎได้ว่า

ST11 → CN24
CN18 + CN2.1

5.5.2.22 [ส่วนประกอบ-หน้าที่] ตรงกับกราฟคำประสมแบบที่ 23 ซึ่งแทนด้วย CN23 เกิดจากกราฟ PART ตามด้วยกราฟ FUNC_N แสดงเป็นกฎได้ว่า

CN23 → PART + FUNC_N

5.5.2.23 [วัตถุประสงค้-ผู้ถูก] ตรงกับกราฟคำประสมแบบที่ 18 ซึ่งแทนด้วย CN18 เกิดจากกราฟ PUR_V ตามด้วยกราฟ OBJ หรือ OBJ_CN แสดงเป็นกฎได้ว่า

CN18 → PUR_V + OBJ / OBJ_CN

5.5.2.24 [วัตถุประสงค้-การกระทำ] ตรงกับกราฟคำประสมแบบที่ 19 ซึ่งแทนด้วย CN19 เกิดจากกราฟ PUR_V ตามด้วยกราฟ ACT_V แสดงเป็นกฎได้ว่า

CN19 → PUR_V + ACT_V

5.5.2.25 [วัตถุประสงค้-วิธีการ] ตรงกับกราฟคำประสมแบบที่ 20 ซึ่งแทนด้วย CN20 เกิดจากกราฟ PUR_V ตามด้วยกราฟ MET_V แสดงเป็นกฎได้ว่า

CN20 → PUR_V + MET_V

5.5.2.26 [ความยิ่งใหญ่-ความสำคัญ-การกระทำ] ตรงกับกราฟคำประสมแบบที่ 16 ซึ่งแทนด้วย CN16 เกิดจากกราฟ GRT ตามด้วยกราฟ ACT_N หรือ ACT_V แสดงเป็นกฎได้ว่า

CN16 → GRT + ACT_N / ACT_V

5.5.2.27 [วัตถุ-สภาพ] ตรงกับกราฟคำประสมแบบที่ 17 ซึ่งแทนด้วย CN17 เกิดจากกราฟ THN ตามด้วยกราฟ STA_V2 แสดงเป็นกฎได้ว่า

CN17 → THN + STA_V2

5.5.2.28 [จำนวน-ส่วนประกอบ] ตรงกับกราฟคำประสมแบบที่ 12 ซึ่งแทนด้วย CN12 เกิดจากกราฟ QNT1 แสดงเป็นกฎได้ว่า

CN12 → QNT1

5.5.2.29 [ระบบ-อุปกรณ์] ตรงกับกราฟคำประสมแบบที่ 13 ซึ่งแทนด้วย CN13 เกิดจากกราฟ SYS_N ตามด้วยกราฟ INS แสดงเป็นกฎได้ว่า

CN13 → SYS_N+INS

5.5.2.30 [หน้าที่-วัตถุประสงค์] ตรงกับกราฟคำประสมแบบที่ 14 ซึ่งแทนด้วย CN14 เกิดจากกราฟ FUNC_N ตามด้วยกราฟ PUR_N แสดงเป็นกฎได้ว่า

CN14 → FUNC_N+PUR_N

5.5.2.31 [อุปกรณ์-วัตถุประสงค์-วัตถุประสงค์/วิธีการ/ระบบ/เวลา/สถานที่/สภาพ/รูปร่างลักษณะ] (แทนด้วย ST12) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟคำประสมต่าง ๆ ตามด้วยกราฟวัตถุประสงค์ (แทนด้วย PUR_V, PUR_VP) แล้วตามด้วยกราฟวัตถุประสงค์ กราฟวิธีการ กราฟระบบ กราฟเวลา กราฟสถานที่ กราฟสภาพ หรือกราฟรูปร่างลักษณะ ดังนี้

- กราฟ INS ตามด้วยกราฟ PUR_V ที่มีการเวียนซ้ำ และอาจตามด้วยกราฟ TIME1, TIME2, MET_V หรือ MET_VP2

- กราฟ INS ตามด้วยกราฟ PUR_VP2 แล้วตามด้วยกราฟ PUR_N

- กราฟ INS ตามด้วยกราฟ PUR_VP3 แล้วตามด้วยกราฟ LOC_PP

- กราฟ INS ตามด้วยกราฟ PUR_VP7 ที่มีการเวียนซ้ำ

- กราฟ INS หรือกราฟคำประสม CN2.1, CN5 หรือ CN27 ตามด้วยกราฟ PUR_VP1 แล้วตามด้วยกราฟ MET_VP1, MET_VP3, PUR_VP2, CHA_N3, SYS_A&V, STA_VP1, LOC_PP หรือ TYPE_NP2

ทั้งหมดแสดงเป็นกฎได้ว่า

ST12 → INS + PUR_V + PUR_V + TIME1 / TIME2 / MET_V / MET_VP2

INS + PUR_VP2 + PUR_N

INS + PUR_VP3 + LOC_PP

INS + PUR_VP7 + PUR_VP7

INS / CN2.1 / CN5 / CN27

+ PUR_VP1 +

MET_VP1 / MET_VP3 / PUR_VP2 / CHA_N3 / SYS_A&V / STA_VP1 / LOC_PP / TYPE_NP2
--

5.5.2.32 [อุปกรณ์-วิธีการ-วัตถุประสงค์/วิธีการ/สภาพ/ประเภทย่อย] (แทนด้วย ST13) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟค่าประสมต่าง ๆ ตามด้วยกราฟวิธีการ (แทนด้วย MET_V, MET_VP) แล้วตามด้วยกราฟวัตถุประสงค์ กราฟวิธีการ กราฟสภาพ หรือกราฟประเภทย่อย ดังนี้

- กราฟ INS หรือกราฟค่าประสม CN29 ตามด้วยกราฟ MET_V แล้วตามด้วยกราฟ MET_V, MET_VP3, PUR_V, PUR_VP1, PUR_VP7 หรือ STA_VP1

- กราฟ INS หรือกราฟค่าประสม CN1.1 ตามด้วยกราฟ MET_VP3 แล้วตามด้วยกราฟ PUR_VP2 หรือ TYPE_LOC2

ทั้งหมดแสดงเป็นกฎได้ว่า

ST13 → INS / CN29 + MET_V +

MET_V / MET_VP3 / PUR_V / PUR_VP1 / PUR_VP7 / STA_VP1
--

INS / CN1.1 + MET_VP3 + PUR_VP2 / TYPE_LOC2

5.5.2.33 [อุปกรณ์-รูปร่างลักษณะ-วัตถุประสงค์/รูปร่างลักษณะ/ชื่อ] (แทนด้วย ST14) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟค่าประสม CN27 ตามด้วยกราฟรูปร่างลักษณะ แล้วตามด้วยกราฟวัตถุประสงค์ กราฟรูปร่างลักษณะ หรือกราฟชื่อ ดังนี้

- กราฟ INS ตามด้วยกราฟ CHA_NP4 แล้วตามด้วยกราฟ CHA_NP4, PUR_V หรือ PUR_VP1

- กราฟ INS ตามด้วยกราฟค่าประสม CN25.2 แล้วตามด้วยกราฟ PUR_VP7

- กราฟค่าประสม CN27 ตามด้วยกราฟ CHA_V แล้วตามด้วยกราฟ
NAME หรือ CHA_V
ทั้งหมดแสดงเป็นกฎได้ว่า

ST14 →
$$\text{INS} + \begin{cases} \text{CHA_NP4} + \text{CHA_NP4} / \text{PUR_V} / \text{PUR_VP1} \\ \text{CN25.2} + \text{PUR_VP7} \end{cases}$$

CN27 + CHA_V + NAME / CHA_V

5.5.2.34 [อุปกรณ์-ประเภทย่อย-วิธีการ/รูปร่างลักษณะ/วัตถุประสงค์] (แทนด้วย ST15) เกิดจากกราฟค่าประสมแสดง [อุปกรณ์] ตามด้วยกราฟประเภทย่อย แล้วตามด้วยกราฟวิธีการ กราฟรูปร่างลักษณะ หรือกราฟวัตถุประสงค์ ดังนี้

- กราฟค่าประสม CN1.1 ตามด้วยกราฟ TYPE_MAN แล้วตามด้วยกราฟ
PUR_N

- กราฟค่าประสม CN1.2 ตามด้วยกราฟ TYPE_MET2 แล้วตามด้วยกราฟ
CHA_NP3

- กราฟค่าประสม CN1.4 ตามด้วยกราฟ TYPE_SYS2 แล้วตามด้วยกราฟ
MET_CL

- กราฟค่าประสม CN1.4 ตามด้วยกราฟ TYPE_LOC1 แล้วตามด้วยกราฟ
MET_VP2

ทั้งหมดแสดงเป็นกฎได้ว่า

ST15 →
$$\text{CN1.1} + \text{TYPE_MAN} + \text{PUR_N}$$

$$\text{CN1.2} + \text{TYPE_MET2} + \text{CHA_NP3}$$

$$\text{CN1.4} + \begin{cases} \text{TYPE_SYS2} + \text{MET_CL} \\ \text{TYPE_LOC1} + \text{MET_VP2} \end{cases}$$

5.5.2.35 [อุปกรณ์-วัสดุ- วัตถุประสงค์] (แทนด้วย ST16) เกิดจากกราฟค่าประสม CN1.1 ตามด้วยกราฟวัสดุ (แทนด้วย MAT_VP) แล้วตามด้วยกราฟวัตถุประสงค์ (แทนด้วย PUR_VP1) แสดงเป็นกฎได้ว่า

ST16 →
$$\text{CN1.1} + \text{MAT_VP} + \text{PUR_VP1}$$

5.5.2.36 [อุปกรณ์-ระบบ-วิธีการ/ผล] (แทนด้วย ST17) เกิดจากกราฟอุปกรณ์ (แทนด้วย INS) หรือกราฟคำประสม CN1.2 ตามด้วยกราฟระบบที่เป็นคำนาม (แทนด้วย SYS_N) แล้วตามด้วยกราฟวิธีการที่เป็นคำนาม (แทนด้วย MET_N) หรือกราฟผลที่เป็นนามวลี (แทนด้วย RES_NP) แสดงเป็นกฎได้ว่า

$$\text{ST17} \rightarrow \text{INS} / \text{CN1.2} + \text{SYS_N} + \text{MET_N} / \text{RES_NP}$$

5.5.2.37 [อุปกรณ์-ส่วนประกอบ-ประเภทย่อย] (แทนด้วย ST18) เกิดจากกราฟคำประสม CN1.1 ตามด้วยกราฟส่วนประกอบที่เป็นนามวลีแบบที่ 2 (แทนด้วย PART_NP2) แล้วตามด้วยกราฟประเภทย่อยที่แสดงวิธีการแบบที่ 2 (แทนด้วย TYPE_MET2) แสดงเป็นกฎได้ว่า

$$\text{ST18} \rightarrow \text{CN1.1} + \text{PART_NP2} + \text{TYPE_MET2}$$

5.6 การประเมินค่าที่ได้จากการทดสอบ

หลังจากทดสอบกราฟแสดงโครงสร้างของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ จะเป็นการประเมินค่าที่ได้จากการทดสอบว่าเครื่องคอมพิวเตอร์สามารถตรวจจับโครงสร้างได้ถูกต้องและแม่นยำเพียงใดนั้น การประเมินค่าดังกล่าวเป็นการคำนวณหาค่าความครบถ้วน (recall) และค่าความแม่นยำ (precision) ของเครื่องคอมพิวเตอร์ซึ่ง Baeza-Yates and Ribeiro-Neto (1999: 75) ได้กล่าวถึงค่าทั้งสองไว้ว่า

Recall is the fraction of the relevant documents (the set R) which has been retrieved.

$$\text{Recall} = \frac{|Ra|}{|R|}$$

Precision is the fraction of the retrieved documents (the set A) which is relevant.

$$\text{Precision} = \frac{|Ra|}{|A|}$$

$|Ra|$ is the number of documents in the intersection of the sets R and A .

5.6.1 ค่าความครบถ้วน (Recall)

ค่าความครบถ้วน (Recall) คือ ค่าที่ได้จากการนำเอาค่าที่เครื่องคอมพิวเตอร์บอกขอบเขตของคำได้และตรวจสอบแล้วว่าเป็นคำที่ถูกต้อง ซึ่งในที่นี้ คือ จำนวนคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่เครื่องคอมพิวเตอร์บอกขอบเขตของคำได้ถูกต้อง หารด้วยจำนวนคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่ใช้ในการทดสอบทั้งหมด แล้วคูณด้วย 100 เพื่อให้ได้ค่าความครบถ้วนเป็นเปอร์เซ็นต์

สูตรในการหาค่าความครบถ้วน มีดังนี้

$$\frac{\text{คำที่ถูกต้อง}}{\text{คำที่ใช้ในการทดสอบทั้งหมด}} \times 100 = \text{Recall (\%)}$$

ตัวอย่างเช่น

จำนวนคำที่ใช้ในการทดสอบทั้งหมดมี 1,000 คำ เครื่องบอกขอบเขตของคำได้ถูกต้องจำนวน 700 คำจากจำนวนคำที่เครื่องคอมพิวเตอร์บอกขอบเขตของคำได้ทั้งหมด 800 ประโยค จำนวนค่าความครบถ้วนได้ดังนี้

$$\frac{700}{1,000} \times 100 = 70 \%$$

ดังนั้น ค่าความครบถ้วนของตัวอย่างข้างต้น คือ 70 %

5.6.2 ค่าความแม่นยำ (Precision)

ค่าความแม่นยำ (Precision) คือ ค่าที่ได้จากการนำเอาค่าที่เครื่องคอมพิวเตอร์บอกขอบเขตคำได้และตรวจสอบแล้วว่าเป็นคำที่ถูกต้อง ซึ่งในที่นี้ คือ คำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่เครื่องคอมพิวเตอร์บอกขอบเขตของคำได้ถูกต้อง หารด้วยจำนวนคำที่เครื่องคอมพิวเตอร์บอกขอบเขตของคำได้ทั้งหมด แล้วคูณด้วย 100 เพื่อให้ได้ค่าความแม่นยำเป็นเปอร์เซ็นต์

สูตรในการหาค่าความแม่นยำ มีดังนี้

$$\frac{\text{คำที่ถูกต้อง}}{\text{คำที่พบ}} \times 100 = \text{Precision (\%)}$$

ตัวอย่างเช่น

จำนวนคำที่ใช้ในการทดสอบทั้งหมดมี 1,000 คำ เครื่องบอกรอบเขตของคำได้ถูกต้องจำนวน 700 คำจากจำนวนคำที่เครื่องคอมพิวเตอร์บอกรอบเขตประโยคได้ทั้งหมด 800 ประโยค คำนวณค่าความแม่นยำได้ดังนี้

$$\frac{700}{800} \times 100 = 87.50 \%$$

ดังนั้น ค่าความแม่นยำของตัวอย่างข้างต้น คือ 87.50 %

5.7 ผลการทดสอบ

เมื่อสร้างกราฟแสดงโครงสร้างของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ทั้งกราฟย่อยและกราฟใหญ่²² จึงได้สร้างกราฟหลักสำหรับการทดสอบโครงสร้างรวมเพื่อทดสอบการตรวจจับค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ โดยแบ่งเป็น 2 ส่วน คือ กราฟหลักสำหรับทดสอบการตรวจจับโครงสร้างกับคลังข้อมูลที่เป็นรายการคำศัพท์จากพจนานุกรม และกราฟหลักสำหรับทดสอบการตรวจจับโครงสร้างกับคลังข้อมูลที่ได้จากการรวบรวมบทความในนิตยสารและที่เผยแพร่ทางอินเทอร์เน็ต

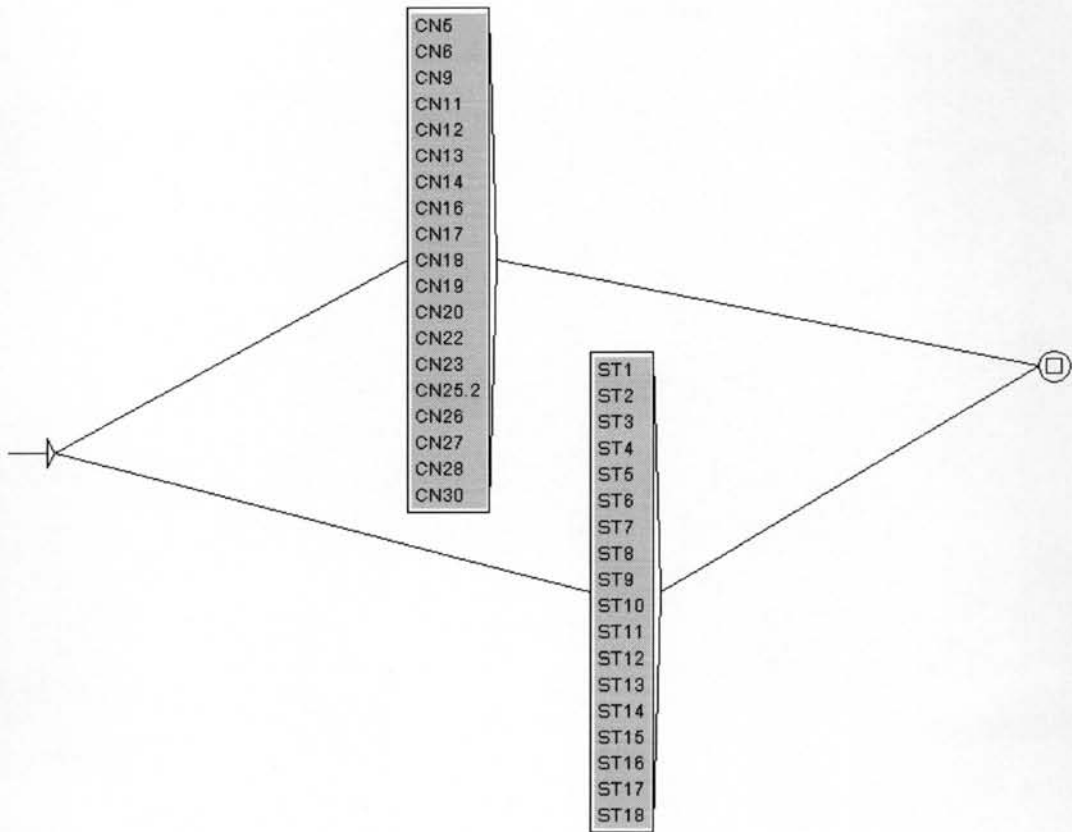
5.7.1 ผลการทดสอบกับคลังข้อมูลที่เป็นรายการคำศัพท์

คลังข้อมูลที่เป็นรายการคำศัพท์นั้นเป็นรายการค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่รวบรวมจากพจนานุกรมที่นำมาใช้เป็นข้อมูล 4 เล่มเท่านั้น โดยไม่มีค่านอกสาขาหรือค่านามทั่วไปอื่น ๆ คลังข้อมูลนี้จะใช้สำหรับทดสอบความสามารถของโปรแกรมในการตรวจจับโครงสร้างทั้งหมดที่วิเคราะห์ได้จากค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ว่า โปรแกรมสามารถตรวจจับคำได้ถูกต้องและแม่นยำเพียงใด

สำหรับการทดสอบโครงสร้างกับคลังข้อมูลที่เป็นรายการคำศัพท์ ผู้วิจัยได้สร้างกราฟหลักขึ้นมา 2 กราฟ โดยให้ชื่อเดียวกันว่า MAIN1 ทั้งหมด กราฟเหล่านี้เป็นกราฟรวม

²² ดูภาพกราฟและรายละเอียดเพิ่มเติมได้ที่ภาคผนวก ค.

โครงสร้างหลักทั้งหมดของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ กราฟดังกล่าวแตกต่างกันตรงรายละเอียดที่ปรากฏเป็นคุณสมบัติของส่วนประกอบและเงื่อนไข 2 แบบ คือ (1) ระบุเพียงชนิดของคำ และ (2) ระบุชนิดของคำกับค่า AKO และมีการกำหนดเงื่อนไขขอบเขต ดังภาพ



ภาพที่ 32 กราฟ MAIN1 แสดงโครงสร้างหลักของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ โดย (1) ระบุเพียงชนิดของคำ และ (2) ระบุชนิดของคำกับค่า AKO และมีการกำหนดเงื่อนไขขอบเขต

ค่านามทั้งหมดที่เป็นกลุ่มตัวอย่างของการศึกษาโครงสร้างมีทั้งสิ้น 973 คำ เมื่อนำกราฟ MAIN1 มาทดสอบแล้ว พบว่า

เมื่อไม่ระบุค่า AKO เครื่องสามารถตรวจจับได้ 974 คำ เป็นคำที่ถูกต้อง 973 คำ และตรวจจับผิด 1 คำ ซึ่งเป็นการตรวจจับเพียงส่วนหนึ่งของคำ ค่าความครบถ้วนเป็นร้อยละ 100 และค่าความแม่นยำเป็นร้อยละ 99.90 โดยคำนวณได้ดังนี้

$$\begin{aligned} \text{ค่าความครบถ้วน (Recall)} &= \frac{973}{973} \times 100 = 100\% \\ \text{ค่าความแม่นยำ (Precision)} &= \frac{973}{974} \times 100 = 99.90\% \end{aligned}$$

เมื่อระบุค่า AKO และมีการกำหนดเงื่อนไขบริบท เครื่องสามารถตรวจจับได้ 976 คำ เป็นคำที่ถูกต้อง 972 คำ และตรวจจับผิด 4 คำ แบ่งเป็น 2 ลักษณะ คือ (1) เป็นคำที่อยู่ในโครงสร้างแต่ไม่สามารถตรวจจับได้ครบทั้งคำ มี 1 คำ และ (2) เป็นการตรวจจับเพียงส่วนหนึ่งของคำ มี 3 คำ ค่าความครบถ้วนเป็นร้อยละ 99.90 และค่าความแม่นยำเป็นร้อยละ 99.59 โดยคำนวณได้ดังนี้

$$\begin{aligned} \text{ค่าความครบถ้วน (Recall)} &= \frac{972}{973} \times 100 = 99.90\% \\ \text{ค่าความแม่นยำ (Precision)} &= \frac{972}{976} \times 100 = 99.59\% \end{aligned}$$

เครื่องคอมพิวเตอร์สามารถตรวจจับคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่เป็นรายการคำศัพท์ได้ครบถ้วนและถูกต้องมาก อีกทั้ง ค่าความครบถ้วนเป็นร้อยละ 100 ในกรณีที่ใช้เพียงชนิดของคำ แสดงให้เห็นว่า บริบทและเงื่อนไขค่า AKO ไม่ใช่สิ่งจำเป็นสำหรับรายการคำศัพท์ เนื่องจากรายการคำศัพท์จะปรากฏ 1 คำ ต่อ 1 บรรทัด หมายความว่า มีการแบ่งคำให้อยู่แล้ว เครื่องคอมพิวเตอร์จึงสามารถตรวจจับคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่มีโครงสร้างจากกราฟ MAIN1 ทั้งสองแบบได้ถูกต้องมาก

5.7.2 ผลการทดสอบกับคลังข้อมูลที่เป็นบทความ

คลังข้อมูลที่เป็นบทความเป็นคลังข้อมูลที่รวบรวมบทความนิตยสารต่าง ๆ ทั้งที่เป็นสิ่งตีพิมพ์และที่เผยแพร่ทางอินเทอร์เน็ต และแบ่งได้เป็น 4 สาขา ได้แก่

- (1) บทความด้านวิทยาศาสตร์และเทคโนโลยีจากนิตยสารอัปเดต
- (2) บทความด้านวิทยาการทางการแพทย์และสุขภาพจากนิตยสารใกล้หมอ

(3) บทความด้านบ้านและการตกแต่งจากนิตยสารบ้านและสวนทั้งที่เป็นสิ่งตีพิมพ์และที่เผยแพร่ทางอินเทอร์เน็ต ที่เว็บไซต์ <http://www.baanlaesuan.com>

(4) บทความด้านยานยนต์จากนิตยสารฟอร์มูลาซึ่งเผยแพร่ผ่านทางเว็บไซต์ <http://autoinfo.co.th>

คลังข้อมูลนี้จะใช้สำหรับทดสอบความสามารถของโปรแกรมในการตรวจจับโครงสร้างทั้งหมดที่วิเคราะห์ได้ว่า โปรแกรมสามารถตรวจจับคำที่มีโครงสร้างดังกล่าวได้ถูกต้องและแม่นยำเพียงใด แม้ว่าคำเหล่านั้นจะปรากฏในประโยค แต่ไม่ได้ปรากฏเป็นรายการคำศัพท์ซึ่งจะปรากฏ 1 คำ ต่อ 1 บรรทัดก็ตาม

ทั้งนี้ ผู้วิจัยได้จัดทำรายการคำเหล่านั้นเพื่อตรวจสอบความถูกต้องของการตรวจจับคำด้วยโปรแกรม Unitex คำดังกล่าวแยกได้เป็น 2 กลุ่ม คือ คำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์จำนวน 815 คำ และคำนามแสดงอุปกรณ์ที่อยู่นอกสาขาวิทยาศาสตร์จำนวน 1,209 คำ

สำหรับการทดสอบโครงสร้างกับคลังข้อมูลดังกล่าว ผู้วิจัยได้นำกราฟ MAIN1 ที่ระบุเพียงชนิดของคำมาทำการทดสอบ พบว่า โปรแกรมตรวจจับคำที่มีโครงสร้างดังกล่าวได้มากถึง 17,373 คำ

คำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์

ค่าความครบถ้วน (Recall)	=	$\frac{150}{815}$	X 100 =	18.40%
ค่าความแม่นยำ (Precision)	=	$\frac{150}{17,373}$	X 100 =	0.86%

คำนามที่อยู่นอกสาขา

ค่าความครบถ้วน (Recall)	=	$\frac{439}{1,209}$	X 100 =	36.31%
ค่าความแม่นยำ (Precision)	=	$\frac{439}{17,373}$	X 100 =	2.53%

จะเห็นได้ว่า การทดสอบกราฟดังกล่าวเพื่อตรวจจับค่านามทั้งสองกลุ่มได้ผลออกมาไม่แตกต่างกันมากนัก กล่าวคือ ค่าความครบถ้วนไม่สูง และค่าความแม่นยำต่ำลงมาก เพราะค่านามทั้งสองกลุ่มไม่ได้ปรากฏเป็นรายการคำศัพท์ แต่ปรากฏในประโยค การปรากฏของคำในประโยคจะมีบริบทแวดล้อม ไม่ได้ปรากฏเพียงคำเดียวโดด ๆ ดังเช่นที่ปรากฏในรายการคำศัพท์ ทำให้เครื่องคอมพิวเตอร์ตรวจจับคำได้ยาวเกินความต้องการ อีกทั้งยังเกิดความผิดพลาดอื่นขึ้นด้วย คือ เครื่องคอมพิวเตอร์ตรวจจับคำ วลี หรือหน่วยสร้างอื่นที่มีการเว้นวรรคระหว่างส่วนประกอบเข้ามาด้วย จึงจำเป็นต้องอาศัยเงื่อนไขอื่นเพิ่มเติมเพื่อช่วยคัดกรองคำที่ถูกตรวจจับผิดพลาดออกไป

ดังนั้น ผู้วิจัยจึงกำกับหมวดหมู่ทางความหมาย (ค่า AKO) ให้กับส่วนประกอบ คือ กราฟย่อยต่าง ๆ เพื่อช่วยคัดกรองคำที่ไม่มีความหมายเกี่ยวข้องและแยกค่านามคำอื่นหรือนามวลีอื่นที่ไม่เกี่ยวข้องออกไปด้วย และยังระบุเงื่อนไขบริบทเพิ่มเติม เงื่อนไขดังกล่าว คือ เงื่อนไขไม่ให้ตรวจจับคำที่มีการเว้นวรรคระหว่างส่วนประกอบ โดยใส่เครื่องหมาย # ลงไประหว่างส่วนประกอบต่าง ๆ²³ กราฟ MAIN1 อันใหม่ที่ใช้ทดสอบจึงได้รับการกำกับหมวดหมู่ทางความหมาย (ค่า AKO) ให้กับกราฟย่อยทั้งหมดและใส่เครื่องหมาย # ระหว่างส่วนประกอบด้วย จากนั้น จึงทดสอบใหม่อีกครั้ง ผลการทดลองเป็นดังนี้

ค่านามแสดงอุปกรณณ์ด้านวิทยาศาสตร์

$$\text{ค่าความครบถ้วน (Recall)} = \frac{236}{815} \times 100 = 28.96\%$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{236}{16,495} \times 100 = 1.43\%$$

ค่านามที่อยู่นอกสาขา

$$\text{ค่าความครบถ้วน (Recall)} = \frac{642}{1,209} \times 100 = 53.10\%$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{642}{16,495} \times 100 = 3.89\%$$

จากการทดลองครั้งที่ 2 พบว่า การกำกับหมวดหมู่ทางความหมาย (ค่า AKO) และการใส่เครื่องหมาย # ให้กับส่วนประกอบ (กราฟย่อย) ทำให้ปริมาณความผิดพลาดลดลง 878 คำ ทั้ง

²³ ดูรายละเอียดกราฟได้ในภาคผนวก ก. หัวข้อ “กราฟแสดงโครงสร้างของค่านามแสดงอุปกรณณ์ด้านวิทยาศาสตร์”

ยังสามารถตัดคำ วลี หรือหน่วยสร้างที่มีการเว้นวรรคออกไปได้ การแก้ปัญหาด้วยวิธีการดังกล่าว ทำให้สามารถคัดกรองคำนามที่ไม่เกี่ยวข้องได้พอสมควร ทำให้ค่าความครบถ้วน (recall) เพิ่มขึ้น ประมาณร้อยละ 10.56 สำหรับคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ และประมาณร้อยละ 16.79 สำหรับคำนามที่อยู่นอกสาขา และค่าความแม่นยำ (Precision) เพิ่มขึ้นประมาณร้อยละ 0.57 สำหรับคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ และประมาณร้อยละ 1.36 สำหรับคำนามที่อยู่นอกสาขา

อย่างไรก็ดี ค่าความครบถ้วน (recall) และค่าความแม่นยำ (Precision) ยังคงต่ำมากเกินไป เพราะมีความผิดพลาดอื่นอีกหลายลักษณะ ได้แก่

(1) การตรวจจับหน่วยสร้างเข้าสู่ศูนย์ที่ขึ้นต้นด้วยคำกริยาแล้วตามด้วยคำกริยาหรือคำนาม หน่วยสร้างดังกล่าวส่วนใหญ่ไม่ใช่คำนาม แต่เป็นส่วนหนึ่งของหน่วยประโยค คือ ภาคแสดงที่ขึ้นต้นด้วยคำกริยา และตามด้วยคำนามซึ่งทำหน้าที่เป็นกรรม ความผิดพลาดดังกล่าวเกิดขึ้นเพราะโครงสร้าง “คำกริยา + คำกริยา / คำนาม” เช่น พัดลม ไขควง ห้ามล้อ ฯลฯ ปรากฏเป็นโครงสร้างหนึ่งในโครงสร้างคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์

(2) ความผิดพลาดที่มาจากคำหลายหน้าที่ กรณีนี้พบบ่อยในคำว่า “ที่” คำว่า “ที่” อาจเป็นคำนาม คำลักษณนาม หรือประพันธสรรพนาม ในกรณีที่เป็นส่วนหลักของคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์ คำว่า “ที่” จะเป็นคำนาม แต่เครื่องคอมพิวเตอร์ตรวจจับทุกคำที่มีคำว่า “ที่” และมีโครงสร้างตรงกับกราฟที่สร้างไว้ คือ “ที่ + คำกริยาแสดงอาการ + คำนาม” คำว่า “ที่” ที่เป็นประพันธสรรพนามจึงถูกตรวจจับมาเป็นจำนวนมาก

(3) ความผิดพลาดที่เกิดจากการไม่กำหนดเงื่อนไขบริบท โปรแกรมนี้ตรวจจับคำที่ยาวที่สุดก่อน เมื่อไม่กำหนดเงื่อนไขบริบท คำที่ตรวจจับได้จึงมีความยาวมาก และไม่ใช่ว่าคำที่ต้องการ แต่โปรแกรมตรวจจับทุกสิ่งที่มีโครงสร้างตามกราฟ

จากความผิดพลาดข้างต้น ผู้วิจัยพบว่า คำที่ถูกต้องสำหรับข้อที่ (1) มีเพียง 14 คำ จาก 5,122 คำ เป็นคำในสาขา 11 คำ คำนอกสาขา 3 คำ และคำที่ถูกต้องสำหรับข้อที่ (2) มีเพียง 16 คำ จาก 620 คำ เป็นคำในสาขา 3 คำ คำนอกสาขา 13 คำ ผู้วิจัยจึงแก้ปัญหาดังกล่าวโดยสร้างกราฟจากการระบุคำในส่วนหลัก แล้วตามด้วยโครงสร้างที่เหลือ และเอาคำว่า “กัน ไข ที่ พัด ห้าม” ออกจากรายการคำศัพท์ในพจนานุกรมด้วยเพื่อลดความผิดพลาดในการตรวจจับ ทั้งนี้ ผู้วิจัยสร้างกราฟโดยอิงจากโครงสร้างคำนามแสดงอุปกรณ์ด้านวิทยาศาสตร์เป็นหลัก กราฟดังกล่าวได้แก่

กราฟ “กัน” แทนด้วย kan เกิดจากคำว่า “กัน” ตามด้วยคำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 แสดงเป็นกฎได้ว่า

kan \rightarrow กัน + V+Act+2111



ภาพที่ 33 กราฟแสดงส่วนหลัก “กัน”

กราฟ “ไข” แทนด้วย khai เกิดจากคำว่า “ไข” ตามด้วยคำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 แสดงเป็นกฎได้ว่า

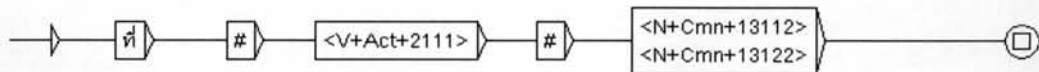
khai \rightarrow ไข + V+Act+2111



ภาพที่ 34 กราฟแสดงส่วนหลัก “ไข”

กราฟ “ที่” แทนด้วย thii เกิดจากคำว่า “ที่” ตามด้วยคำกริยาแสดงอาการที่มีค่า AKO เป็น 2111 ตามด้วยคำนามทั่วไปที่มีค่า AKO เป็น 13112 หรือ 13122 แสดงเป็นกฎได้ว่า

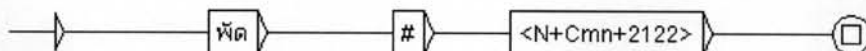
thii \rightarrow ที่+V+Act+2111+N+Cmn+13112,13122



ภาพที่ 35 กราฟแสดงส่วนหลัก “ที่”

กราฟ “พัด” แทนด้วย phad เกิดจากคำว่า “พัด” ตามด้วยคำนามทั่วไปที่มีค่า AKO เป็น 2122 แสดงเป็นกฎได้ว่า

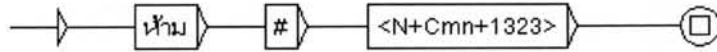
phad \rightarrow พัด + N+Cmn+2122



ภาพที่ 36 กราฟแสดงส่วนหลัก “พัด”

กราฟ “ห้าม” แทนด้วย ham เกิดจากคำว่า “ห้าม” ตามด้วยคำนามทั่วไปที่มีค่า AKO เป็น 1323 แสดงเป็นกฎได้ว่า

ham \rightarrow ห้าม + N+Cmn+1323



ภาพที่ 37 กราฟแสดงส่วนหลัก “ห้าม”

ผู้วิจัยได้สร้างกราฟหลักอันใหม่ซึ่งมีชื่อว่า MAIN2 กราฟนี้ประกอบด้วยกราฟโครงสร้างหลักเดิมที่ถูกดึงโครงสร้างที่มีปัญหาข้างต้นออกมา กราฟที่ถูกดึงออกมา คือ กราฟคำประสม CN18, CN19 และ CN20 กราฟเหล่านี้ได้รับการแทนที่ด้วยกราฟทั้งสี่ข้างต้น นอกจากนี้ยังได้ระบุเงื่อนไขทางบริบทเพิ่มเติมลงไปด้วย คือ คำที่มีโครงสร้างหลักตามที่กำหนดไว้จะปรากฏหน้าเงื่อนไขต่าง ๆ ต่อไปนี้

(1) การเว้นวรรค

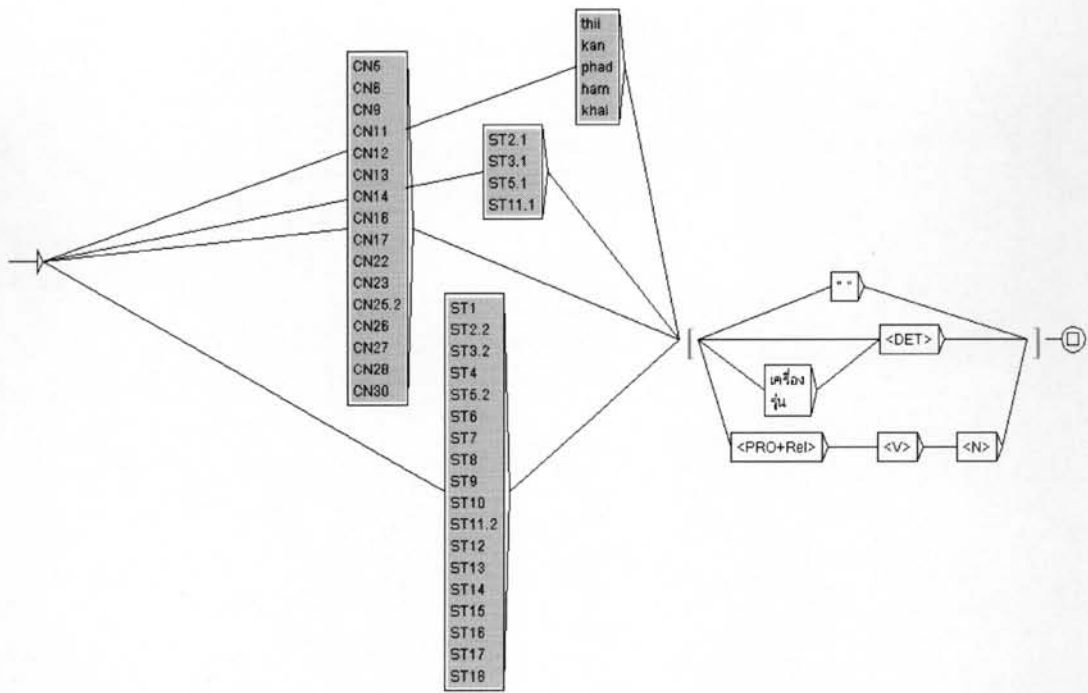
(2) คำบ่งชี้เฉพาะ เช่น นี้ นั้น ฯลฯ ซึ่งอาจมีคำลักษณะนาม “เครื่อง รุ่น ชนิด”

นำหน้าได้

(3) อนุภาคย่ที่ขึ้นต้นด้วยประพันธสรรพนาม “ที่ ซึ่ง” ตามด้วยคำกริยา และคำนาม

เงื่อนไขบริบทข้างต้นปรากฏอยู่ในเครื่องหมาย [] เพื่อกำหนดให้

โปรแกรมไม่แสดงเงื่อนไขเป็นสิ่งที่ตรวจจับได้ด้วย กราฟหลัก MAIN2 ที่ได้รับการกำหนดเงื่อนไขบริบทและเพิ่มกราฟเฉพาะโครงสร้างที่เป็นปัญหาลงไปมีลักษณะ ดังภาพ



ภาพที่ 38 กราฟ MAIN2 แสดงโครงสร้างหลักทั้งหมดของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์
โดยมีการกำหนดเงื่อนไขบริบท

จากนั้น จึงทดสอบกราฟ MAIN2 ผลการทดลองเป็นดังนี้

กราฟที่ไม่ได้กำกับค่า AKO มีผลเป็น

ค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์

ค่าความครบถ้วน (Recall) = $\frac{297}{815} \times 100 = 36.44\%$

ค่าความแม่นยำ (Precision) = $\frac{297}{5,250} \times 100 = 5.66\%$

ค่านามที่อยู่นอกสาขา

ค่าความครบถ้วน (Recall) = $\frac{615}{1,209} \times 100 = 50.86\%$

ค่าความแม่นยำ (Precision) = $\frac{615}{5,250} \times 100 = 11.71\%$

กราฟที่กำกับค่า AKO มีผลเป็น

ค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์

$$\text{ค่าความครบถ้วน (Recall)} = \frac{386}{815} \times 100 = 47.36\%$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{386}{2,397} \times 100 = 16.10\%$$

ค่านามที่อยู่นอกสาขา

$$\text{ค่าความครบถ้วน (Recall)} = \frac{790}{1,209} \times 100 = 65.34\%$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{790}{2,397} \times 100 = 32.96\%$$

จากการทดลองครั้งที่ 3 พบว่า กราฟหลัก MAIN2 สามารถลดความผิดพลาดได้มากขึ้น แต่ยังคงเกิดความผิดพลาดสำคัญ คือ การตัดคำของโปรแกรม Unitek ผิดพลาด ความผิดพลาดลักษณะนี้เกิดขึ้นเพื่อให้โปรแกรมสามารถตรวจจับคำในโครงสร้างได้ อันอาจทำให้เกิดการแยกคำ ตัวอย่างเช่น

ตัดคำว่า “พลังงาน” เป็น *ลังงาน* มาจากโครงสร้าง “ค่านาม + ค่านาม” คือ ค่านาม “ลัง” และค่านาม “งาน”

ตัดคำว่า “หลากสีสัน” เป็น *ลากสีสัน* มาจากโครงสร้าง “คำกริยา + ค่านาม” คือ คำกริยา “ลาก” และค่านาม “สีสัน”

ตัดคำว่า “ตุลาคม” เป็น *ลาคม* มาจากโครงสร้าง “ค่านาม + คำกริยา” คือ ค่านาม “ลา” และคำกริยา “คม”

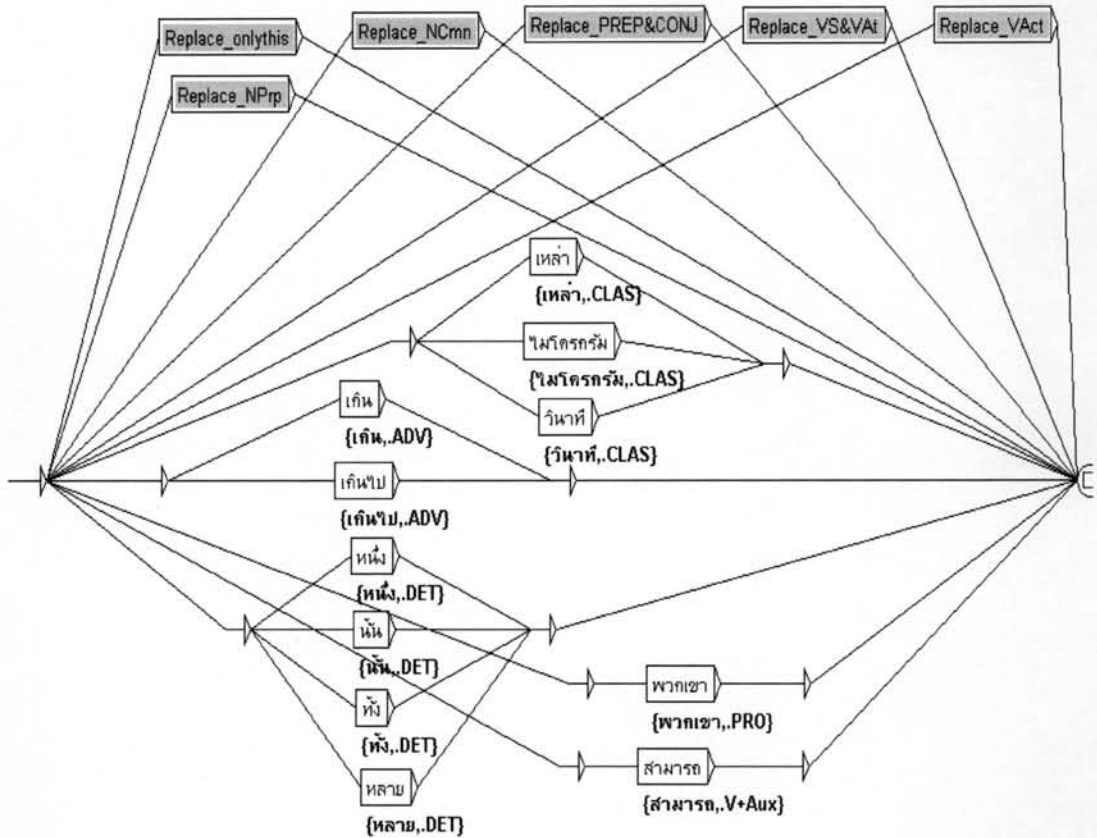
ตัวอย่างการตัดคำผิดของโปรแกรม Unitek ปรากฏดังภาพด้านล่าง

ยอาหารจึงเกิดความสนใจ เครื่องสองชั้นนี้ ก็สามารถเก็บยุงได้ เครื่องเขม่นเขม่นหมากก็วางรพ
 ากใช้ของเหม็นๆ เป็นเหยื่อล่อแมลงวัน อีโคบอล 2 คงจะมีกลิ่นเหม็นตุลึกล้นน้อย (S)ชื่อเรื่
 องเปิดประตูออกมาจนน้ำทะเลลึกเข้ารถ เดินลุยน้ำท่วม ทั้งรถไว้ รอให้น้ำลดแล้วให้ช่างมาค
 ของนักเรียน เพื่อป้องกันการลักพาตัว (S)อย่างไรก็ตามเมื่อไม่นานมานี้ สหพันธ์เสรีภาพ
 นอายุแบบตะวันออกที่เป็นเอกลักษณ์เฉพาะตัว ให้ทั้งความงามและสะดวกสบายในการใช้
 ฝังอยู่ที่ตัวผู้จะกระเด็นเพื่อผลักดันน้ำบานออกมา สามารถใช้ได้กับบานพับทุกชนิด โดยไม่
 การปลีกตัว หาเวลาออกกำลังกายให้ได้ 3. สำงาเป็นก็โซ้ยาชวย 4. การปรับลดความฮ่ว
 และแสงที่มนุษย์สร้างขึ้น) พลังงานจากสนามแม่เหล็กรอบ ๆ สายไฟฟ้า แม้กระทั่งการลั่นล
 โฟบนต้นคริสต์มาสหนึ่งหลอด พลังงานที่ได้นี้อาจจ่ายให้กับเครื่องซาวจเมตเตอร์ หรือไม
 ภาให้เมื่ออาหารกับผลผลิตพลังงานเพิ่มขึ้น 2. กลุ่มที่กระตุ้นการหลั่งสารซีโรโตนิน (ser
 อย่าง และใช้มนุษย์เป็นแหล่งพลังงานไฟฟ้า เครื่องจักรนี้ "เพาะเลี้ยง" มนุษย์ไว้ในฟาร์มอด
 ทุกๆ ชาของสัตว์โลกนั้นมีกำลังเท่ากัน พูลพบว่ามีชีวิตต่างๆ สามารถสร้างพลังงานเชิง
 กของอังกฤษ ในบริสทอล กำลังพัฒนาหุ่นยนต์ที่จับแมลงวันเป็นอาหาร โดยตั้งชื่อหุ่นย
 อบอยู่เหนือเตา ใช้พัดลมกำลังแรงเป็นตัวดูด บางรุ่นมีการติดไฟ ฟิลเตอร์กรองอากาศและค
 กันในหลายๆจุดก็ใช้มีมที่มีกำลังวัดด้สงขึ้น *ฝักบัวอาบน้ำแบบ Rain Shower (S)ฝักบัวมี
 โกลส์เดียวกัน ต้องใช้มีมที่มีกำลังวัดด้สงขึ้น คิดงายๆสำเป็นบ้านเดี่ยวสองชั้น มีห้องน้ำ 2-3
 ไฟฉายคาดศีรษะ (S)ให้กำลังสองสร้างสูง เพียงนำมาคาดบนศีรษะเพื่อส่องสว่างไปยังวัด
 ทั่วทั้งป่าใบพีวีซี และอะคริลิกหลากสีสั้น *ไฟลวดระบบม้วนเก็บ (S)สามารถเก็บซ่อนได้มีดขีด
 ภาใช้อย่างสำงหน้าบอยที่ตลอดเวลาเข้าห้องน้ำ สวนอ่างอาบน้ำที่ใช้เฉพาะเวลาอาบน้ำก็ควรอ
 โกลส์โลกที่พบเมื่อวันที่ 10 ตุลาคม พ.ศ. 2547 ได้สร้างความประหลาดใจให้กับนักดาราศา
 วมิสภาสหรัฐ เมื่อวันที่ 2 ตุลาคม 2545 ซึ่งมีเนื้อความโดยสรุปว่า "นักวิทยาศาสตร์ศึกษา
 ปีที่ 28 ฉบับที่ 10 เดือนตุลาคม 2547 (S)ชื่อหนังสือ: โกลส์หมอ (S)คอลัมน์: Drug & Sup
 ปีที่ 28 ฉบับที่ 10 เดือนตุลาคม 2547 (S)ชื่อหนังสือ: โกลส์หมอ (S)คอลัมน์: โลกกว้างทา
 ปีที่ 28 ฉบับที่ 10 เดือนตุลาคม 2547 (S)ชื่อหนังสือ: โกลส์หมอ (S)คอลัมน์: โลกกว้างทา
 ในเชิงพาณิชย์ ตั้งแต่เดือนตุลาคม พ.ศ. 2544 มีขนาดใหญ่กว่ากล้องไม้ขีดเล็กน้อย ไม่ก
 ลว่าลูกสูบเบรคติด (S)ต้องสร้างกระบอกสูบเบรค เปลี่ยนชุดซ่อมใหม่ ถายน้ำมันเบรค ไสลม
 ้นามากมาย ไม่ว่าจะเป็นการสำงหน้า อ่างน้ำ หรือทำกับข้าว และในวันหยุดด้วยแล้วจะมีก
 สละวันค่อนข้างเร่รีบและมีเวลาดำกัด สามารถบดสับ ผสมอาหาร และปั่นน้ำผลไม้ได้ในเด
 โกล้ม และต้องประสบปัญหาเวลาใช้ เพราะมันบิดไม่สะดวกเอาเสียเลย สาเหตุมาจากใบมีด

ภาพที่ 39 ตัวอย่างการตัดคำผิดของโปรแกรม Unitex

ความผิดพลาดดังกล่าวเกิดจากการวิเคราะห์คำภาษาไทยของโปรแกรม Unitex
 โดยวิเคราะห์ตัวอักษรเรียงตัว (Analysing by character) และถือว่า ตัวอักษรแต่ละตัวเท่ากับหนึ่งคำ
 ตามแบบภาษาอังกฤษ ทั้งนี้ เพื่อแก้ปัญหาที่ภาษาไทย ไม่มีช่องว่างระหว่างคำ ผู้วิจัยจึงคิดสร้างกราฟ
 ไวยากรณ์เพื่อแก้ไขความกำกวมด้วยการกำหนดชนิดของคำตามตำแหน่งที่ปรากฏ แต่วิธีดังกล่าว
 ไม่สามารถทำได้เพราะเป็นข้อจำกัดของโปรแกรมที่ยังไม่สามารถเชื่อมโยงกราฟไวยากรณ์ให้มีผล
 ต่อข้อมูลที่ใช้ในการทดสอบได้ตั้งแต่กระบวนการก่อนการประมวลผล (preprocessing)
 หากสามารถกำหนดชนิดของคำตามตำแหน่งที่ปรากฏในขั้นตอนนี้ได้ จะทำให้ตรวจจับคำได้
 ถูกต้องขึ้นและลดความกำกวมได้

เมื่อไม่สามารถสร้างกราฟไวยากรณ์ได้ ผู้วิจัยจึงได้ทดลองสร้างกราฟสำหรับแก้ไขการตัดคำผิดในขั้นตอนก่อนการประมวลผล (Preprocessing) กราฟดังกล่าวมีชื่อว่า Replace.grf กราฟนี้จะกำหนดคำที่ถูกต้องซึ่งโปรแกรมเคยตัดผิด ทำให้โปรแกรมระบุลงในข้อมูลอย่างชัดเจนว่าคำคำนั้นเป็นคำใด เป็นคำชนิดใด และมีหมวดหมู่ทางความหมาย (คำ AKO) ไດ โปรแกรมจะกำหนดลักษณะที่ถูกต้องทันทีโดยไม่ตรวจสอบในพจนานุกรมอีก วิธีการดังกล่าวทำได้โดยการใส่คำที่ตรวจจับผิด ตามด้วยเครื่องหมายทับ (/) ตามด้วยคำที่ถูกต้อง แล้วตามด้วยการระบุชนิดของคำ และอาจมีหมวดหมู่ทางความหมาย (คำ AKO) ด้วยภายในเครื่องหมายปีกกา ({ }) เช่น ไม่ต้องการให้ตรวจจับคำว่า “สามารถ” เป็นลักษณะอื่น จะสร้างเป็น สามารถ/{สามารถ,.V+Aux} ดังภาพ



ภาพที่ 40 กราฟ Replace.grf

เมื่อเรียกใช้กราฟ Replace.grf ทำให้คำที่กำหนดนั้นได้รับการตัดคำอย่างถูกต้องตั้งแต่ขั้นตอนนี้ จากนั้น จึงทดสอบคลังข้อมูลที่ผ่านมาการแก้ไขการตัดคำแล้วกับกราฟหลัก MAIN2 ผลการทดลองเป็นดังนี้

กราฟที่ไม่ได้กำกับค่า AKO มีผลเป็น

ค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์

$$\text{ค่าความครบถ้วน (Recall)} = \frac{390}{815} \times 100 = 47.85\%$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{390}{4,794} \times 100 = 8.14\%$$

ค่านามที่อยู่นอกสาขา

$$\text{ค่าความครบถ้วน (Recall)} = \frac{663}{1,209} \times 100 = 54.84\%$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{663}{4,794} \times 100 = 13.82\%$$

กราฟที่กำกับค่า AKO มีผลเป็น

ค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์

$$\text{ค่าความครบถ้วน (Recall)} = \frac{582}{815} \times 100 = 71.41\%$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{582}{1,901} \times 100 = 30.62\%$$

ค่านามที่อยู่นอกสาขา

$$\text{ค่าความครบถ้วน (Recall)} = \frac{985}{1,209} \times 100 = 81.47\%$$

$$\text{ค่าความแม่นยำ (Precision)} = \frac{985}{1,901} \times 100 = 51.81\%$$

จะเห็นได้ว่า ค่าความครบถ้วน (Recall) และค่าความแม่นยำ (Precision) เพิ่มขึ้นร้อยละ 71.41 และร้อยละ 30.62 ตามลำดับ สำหรับค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ และค่าความครบถ้วน (Recall) และค่าความแม่นยำ (Precision) เพิ่มขึ้นเป็นร้อยละ 81.47 และ

ร้อยละ 51.81 ตามลำดับ สำหรับค่านามที่อยู่นอกสาขา ผลที่ได้ถือว่าน่าพอใจ เพราะสามารถปรับปรุงค่าความครบถ้วนและความแม่นยำขึ้นมาได้ร้อยละ 53.01 และร้อยละ 29.76 ตามลำดับ สำหรับค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ และเพิ่มค่าความครบถ้วนและความแม่นยำขึ้นมาได้ร้อยละ 45.16 และร้อยละ 49.28 ตามลำดับ สำหรับค่านามที่อยู่นอกสาขา เมื่อเทียบกับผลการทดลองที่ 1

ทั้งนี้ การที่ค่าความแม่นยำสำหรับค่านามที่อยู่นอกสาขามากกว่าค่าความแม่นยำสำหรับค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ เนื่องด้วยเหตุผล ดังนี้

(1) ค่านามแสดงอุปกรณ์ทั่วไปที่อยู่นอกสาขาส่วนใหญ่มีโครงสร้างไม่ซับซ้อน โครงสร้างนั้นคล้ายกับโครงสร้างของค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ที่มีค่าปรากฏในโครงสร้างจำนวนมาก ตัวอย่างเช่น

โครงสร้าง [อุปกรณ์ - วัตถุประสงค์] มีค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ในรายการคำศัพท์จำนวน 536 คำ

โครงสร้าง [อุปกรณ์ - วิธีการ] มีค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ในรายการคำศัพท์จำนวน 95 คำ

(2) ค่านามแสดงอุปกรณ์ทั่วไปที่อยู่นอกสาขามีจำนวนคำในคลังข้อมูลมากกว่าค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ คือ ค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์มีจำนวน 815 คำ ในขณะที่ค่านามแสดงอุปกรณ์ที่อยู่นอกสาขาวิทยาศาสตร์มีจำนวน 1,209 คำซึ่งมีจำนวนคำมากกว่า 394 คำ

(3) ผู้วิจัยได้คัดเลือกค่านามแสดงอุปกรณ์ทั้งด้านวิทยาศาสตร์และนอกสาขาวิทยาศาสตร์จากการอ่านบทความทั้งหมด โดยไม่ได้วิเคราะห์โครงสร้างภายในของค่านามแสดงอุปกรณ์เหล่านั้น จึงมีความเป็นไปได้ที่อาจมีค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์บางคำไม่ถูกตรวจจับได้

ด้วยสาเหตุดังกล่าว ค่าความแม่นยำสำหรับค่านามที่อยู่นอกสาขาจึงมากกว่าค่าความแม่นยำสำหรับค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์

5.8 สรุป

ในบทนี้เป็นการประยุกต์ผลการวิเคราะห์ทางภาษาศาสตร์ในบทที่ 3 และ 4 กับเครื่องคอมพิวเตอร์ ด้วยโปรแกรม Unitex การประยุกต์ผลการวิเคราะห์ดังกล่าวทำได้โดยการสร้างกราฟแสดงความสัมพันธ์ ได้แก่ กราฟย่อยและกราฟหลัก

จากการทดสอบ พบว่า เมื่อทดสอบกับรายการคำศัพท์ ทั้งค่าความครบถ้วน (Recall) และค่าความแม่นยำ (Precision) สูงมาก คือ ประมาณร้อยละ 99 แต่เมื่อทดสอบกับคลังข้อมูลค่าครบถ้วน (Recall) ต่ำ คือ ประมาณร้อยละ 18 สำหรับค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ และประมาณร้อยละ 36 สำหรับค่านามที่อยู่นอกสาขา และค่าความแม่นยำ (Precision) ต่ำมากเช่นกัน คือ โดยเฉลี่ยเพียงร้อยละ 1 ผู้วิจัยจึงได้แก้ปัญหาด้วยวิธีต่าง ๆ ทำให้สามารถเพิ่มค่าความครบถ้วน (Recall) เป็นร้อยละ 71.41 สำหรับค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ และเป็นร้อยละ 81.47 สำหรับค่านามที่อยู่นอกสาขา และเพิ่มค่าความแม่นยำ (Precision) เป็นร้อยละ 30.62 สำหรับค่านามแสดงอุปกรณ์ด้านวิทยาศาสตร์ และเป็นร้อยละ 51.81 สำหรับค่านามที่อยู่นอกสาขา

นอกเหนือจากการแก้ไขปัญหาที่เกิดขึ้นด้วยการกำหนดเงื่อนไขทางบริบทให้แก่กราฟการกำกับหมวดหมู่ทางความหมาย (ค่า AKO) ให้แก่คำซึ่งเป็นส่วนประกอบเป็นส่วนสำคัญอย่างมาก เพราะการกำกับหมวดหมู่ทางความหมาย (ค่า AKO) ให้แก่คำช่วยให้สามารถคัดกรองคำที่ไม่เกี่ยวข้องได้สูงกว่าการระบุเพียงชนิดของคำ