

## บทที่ 4

### การรู้จำชื่อเฉพาะ

ในบทนี้จะกล่าวถึงการรู้จำชื่อเฉพาะ โดยเป็นการใช้กฎที่เขียนขึ้นเพื่อพิจารณารายการชื่อเฉพาะที่เลือกมาด้วยวิธีการทางสถิติว่าเป็นชื่อเฉพาะจริงหรือไม่และเป็นชื่อเฉพาะประเภทใด ซึ่งบทนี้จะแบ่งเนื้อหาออกเป็น 4 หัวข้อ โดยหัวข้อแรกจะกล่าวถึงการสร้างกฎจากหลักฐานภายใน (internal evidence) รวมถึงหลักฐานจากบริบทข้างเคียง (external evidence) และในหัวข้อส่วนที่สองจะกล่าวถึงกฎที่ถูกสร้างขึ้น ส่วนที่สามและสี่จะกล่าวถึงผลของการใช้กฎที่สร้างขึ้นเพื่อตัดสินใจจำแนกประเภทของชื่อเฉพาะว่ามีความถูกต้องมากน้อยเพียงใดและการอภิปรายผลการทดสอบที่ได้

#### 4.1 การจำแนกประเภทชื่อเฉพาะ

หลังจากที่ได้เลือกใช้วิธีการทางสถิติแบบ Mutual Information ร่วมกับการใช้ Localmax algorithm เพื่อหาขอบเขตและเลือกชื่อเฉพาะออกมาแล้ว ขั้นตอนต่อไป คือ การตัดสินใจว่าแต่ละหน่วยที่ดึงมานั้นเป็นชื่อเฉพาะจริงหรือไม่และเป็นชื่อเฉพาะประเภทใด ขั้นตอนนี้จะอาศัยกฎที่สร้างขึ้นเพื่อใช้รู้จำและจำแนกประเภทของชื่อเฉพาะ โดยใช้หลักฐานจากบริบทภายใน (internal evidence) และบริบทข้างเคียง (external evidence) ซึ่งจะมีส่วนช่วยในการรู้จำตำแหน่งและจำแนกประเภทของชื่อเฉพาะ

##### 4.1.1 หลักฐานภายใน

หลักฐานภายใน (internal evidence) หมายถึงคำที่มักจะพบว่าปรากฏร่วมกับชื่อเฉพาะบ่อยๆ และกลุ่มคำที่เป็นหลักฐานภายในนี้ก็มีลักษณะเป็นกลุ่มคำแบบปิด หรือ closed class กล่าวคือสามารถคาดเดาจำนวนของคำทั้งหมดที่เป็นหลักฐานภายในได้ หากมีคลังข้อมูลที่ใหญ่เพียงพอ ยกตัวอย่างเช่น หลักฐานภายในที่แสดงอรรถลักษณะเป็น <ปัจเจกบุคคล> และ <บอกอายุ เพศ และสถานภาพการสมรส> ซึ่งจัดอยู่ในประเภทชื่อคนนั้น ก็จะมีกลุ่มคำนำหน้าชื่อ ได้แก่ "นาย", "นาง", "นางสาว" เป็นต้น

หลักฐานภายใน ซึ่งได้แก่คำนำหน้าและคำที่ตามหลังชื่อเฉพาะนั้นจะสามารถหาได้จากการพิจารณาจากคลังข้อมูลใช้ฝึก ซึ่งหลักฐานภายในเหล่านี้จะช่วยในการเขียนกฎเพื่อกำหนดขอบเขตของชื่อเฉพาะโดยอาจจะใช้บ่งบอกถึงจุดเริ่มต้นของชื่อเฉพาะ รวมทั้งจุดสิ้นสุดของชื่อ

เฉพาะได้ โดยหลักฐานภายในจะถูกแบ่งออกเป็น 3 ประเภทตามลักษณะการปรากฏของมันกับชื่อเฉพาะ ดังนี้

1. หลักฐานภายในที่ปรากฏอยู่หน้าชื่อเฉพาะ ได้แก่คำนำหน้าชื่อเฉพาะต่างๆ ซึ่งรายการคำนำหน้าชื่อที่เก็บได้มาจากคลังข้อมูลจะแสดงอยู่ด้านล่าง
 

"อ.", "นาย", "นาง", "ดร.", "พล.ต.", "พล.ท.", "พ.ญ.", "พ.อ.", "ส.ก.", "ร.ต.", "ร.อ.", "ภ.ญ.", "ร.ท.", "แม่จ้อย", "จอมพล", "น.ส.", "นางสาว", "น.ต.", "อาจารย์", "หัวหน้า", "พล.อ.", "น.พ.", "ม.ล.", "คุณหญิง", "หลวงพ่อ", "ร.อ.หญิง", "พ.ต.ท.", "พ.ต.หญิง", "ส.ต.อ.", "พล.ร.ท.", "ร.ต.อ.", "พล.ต.ท.", "ม.ร.ว.", "พล.อ.อ.", "พล.ต.ต.", "พล.ต.อ.", "ส.ต.ต.", "พ.ต.ต.", "พ.ต.อ.", "พล.ต.ท.", "ร.ต.ต.", "จ.ส.ต.", "ร.ต.ท.", "พ.ญ.คุณหญิง", "ว่าที่ร.ต.", "พ.ต.อ.น.พ.", "กรม", "พรรค", "องค์การ", "กระทรวง", "สำนักงาน", "สำนัก", "กองทัพบก", "บริษัท", "ธนาคาร", "องค์กร", "กอง", "โรงเรียน", "โรงพยาบาล", "วิทยาลัย", "บรรษัท", "สหภาพ", "สภา", "มูลนิธิ", "ศาล", "สถาบัน", "ศูนย์", "ทวีป", "ร้านกาแฟ", "ร้านอาหาร", "สวนสัตว์", "ค่าย", "สถานทูต", "สนามกอล์ฟ", "ท่าอากาศยาน", "สุเหร่า", "บึง", "คุก", "มัสยิด", "สวนสาธารณะ", "ร.ร.", "พระธาตุ", "เกาะ", "สโมสร", "สมาคม", "สาธารณรัฐ", "ราชอาณาจักร", "จังหวัด", "แม่น้ำ", "ลุ่มน้ำ", "เขา", "อ่างเก็บน้ำ", "ทำน้ำ", "ท่าเรือ", "ท่าเรือน้ำลึก", "ปราสาท", "ปราสาทหิน", "คลอง", "วัด", "วัง", "รพ.", "อนุสาวรีย์", "พระบรมราชานุสาวรีย์", "เขื่อน", "อนุสรณ์", "มณฑล", "นคร", "ต.", "อาคาร", "ตึก", "โรงแรม", "สนามบิน", "ประเทศ", "แขวง", "หมู่บ้าน", "สนามม้า", "แฟลต", "ซอย", "ร้าน", "เคหะชุมชน", "โรงงาน", "สวน", "บ้าน", "ธนาคาร", "ศาลเจ้า", "สน.", "อุทยาน", "ถ.", "เขตเทศบาลตำบล", "ศูนย์ประชุม", "หอประชุมอำเภอ", "ทางด่วนสาย", "ถนนสาย", "สำนักงานสาขา", "หอประชุม", "ที่ว่าการ", "มูลนิธิ", "ภูมิภาค", "สถานี", "ถนน", "ที่ทำการสาขา", "ห้อง", "สถาบัน", "อำเภอ", "หอ", "ที่ทำการ", "ภาค", "ตลาดสดเทศบาล", "ตลาดสด", "ตลาด", "สนามกีฬาเทศบาล", "สนามกีฬากลาง", "สนามกีฬา", "สนาม", "ศาลา", "จ.", "เมือง", "กรุง", "เขต"
2. หลักฐานภายในที่ปรากฏอยู่ด้านหน้าและหลังชื่อเฉพาะ ได้แก่ บริษัท...จำกัด
3. หลักฐานภายในที่อาจปรากฏอยู่ด้านหน้าหรือด้านหลังชื่อเฉพาะก็ได้ ได้แก่ มหาวิทยาลัย เช่น มหาวิทยาลัยธรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

จากการพิจารณาหลักฐานภายในเหล่านี้ พบว่าชื่อเฉพาะประเภทชื่อคนจะมีหลักฐานชนิดนี้มากที่สุด โดยจะมีคำนำหน้าชื่อถึง 5408 จากจำนวนชื่อคน 5431 ชื่อโดยคิดเป็น 99.577% ชื่อองค์กรมีคำนำหน้าชื่อ 3140 จากจำนวนชื่อองค์กร 3222 ชื่อโดยคิดเป็น 97.455% และชื่อสถานที่ มีคำนำหน้าชื่ออยู่ 2158 จากจำนวนชื่อสถานที่ 3030 ชื่อโดยคิดเป็น 71.221% ดังนั้น จากการพบหลักฐานภายในเหล่านี้ การสร้างกฎจึงใช้คำนำหน้าชื่อนี้เป็นหลักในการระบุขอบเขตของชื่อเฉพาะ โดยจะเป็นการกำหนดจุดเริ่มต้นของชื่อนั้นเอง นอกจากนั้นแล้ว คำนำหน้าชื่อยังสามารถบ่งบอกประเภทของชื่อเฉพาะได้อีกด้วย ซึ่งการแบ่งประเภทของชื่อเฉพาะตามคำนำหน้าชื่อนั้น อาจทำได้โดยการศึกษาความหมายของคำนำหน้าชื่อแต่ละคำ ด้วยการวิเคราะห์องค์ประกอบของอรรถลักษณะ (componential analysis) ซึ่งเป็นวิธีที่นำความหมายของคำนำหน้าชื่อแต่ละคำมา แจกแจงเป็นส่วนความหมายย่อยที่เรียกว่าอรรถลักษณะ (semantic feature) (อมรทิพย์, 2546) ซึ่งพบว่าจากอรรถลักษณะร่วมของกลุ่มคำ จะทำให้แบ่งประเภทของคำนำหน้าชื่อเฉพาะได้ดังตารางที่ 4 ด้านล่าง โดยจะแบ่งออกเป็น 3 ประเภทคือ ชื่อคน ชื่อองค์กร และชื่อสถานที่

ตารางที่ 4 การแบ่งประเภทกลุ่มคำนำหน้าชื่อ

ประเภทของชื่อเฉพาะ	อรรถลักษณะร่วมของกลุ่มคำนำหน้าชื่อ	คำนำหน้าชื่อ
ชื่อคน	<ปัจเจกบุคคล> <แสดงฐานะหรือลำดับชั้น>	"พล.ต." , "พล.ท." , "พ.อ." , "พล.อ." , "ส.ก." , "ร.ต." , "ร.อ." , "ร.ท." , "พล.อ.อ." , "พล.ต.ต." , "พล.ต.อ." , "ส.ต.ต." , "พ.ต.ต." , "พ.ต.อ." , "พล.ต.ท." , "ร.ต.ต." , "จ.ส.ต." , "ร.ต.ท." , "ร.อ.หญิง" , "พ.ต.ท." , "พ.ต.หญิง" , "ส.ต.อ." , "พล.ร.ท." , "ร.ต.อ." , "พล.ต.ท." , "จอมพล" , "น.ต." , "ว่าที่ร.ต." , "พ.ต.อ.น.พ." , "คุณหญิง"
	<ปัจเจกบุคคล> <แสดงระดับการศึกษา>	"ดร."

	<ปัจเจกบุคคล> <แสดงฐานันดร>	"ม.ร.ว.", "ม.ล."
	<ปัจเจกบุคคล> <แสดงสถานะทางศาสนา>	"หลวงพ่อ"
	<ปัจเจกบุคคล> <คำเรียกทางวิชาชีพ>	"พ.ญ.", "ภ.ญ.", "น.พ.", "อาจารย์", "พ.ญ.คุณหญิง"
	<ปัจเจกบุคคล> <บอกอายุ เพศ หรือ/และ สถานภาพการสมรส>	"นาย", "นาง", "น.ส.", "นางสาว", "แม่อยู่"
ชื่อองค์กร	<กลุ่มบุคคล> <มีการรวมตัวกัน> <ประกอบขึ้นเป็นหน่วยเดียว> <มีสภาพเป็นนิติบุคคล> <เพื่องานราชการ>	"กรม"* , "องค์การ" "กระทรวง"* , "กองทัพ" "กอง", "ศาล"
	<กลุ่มบุคคล> <มีการรวมตัวกัน> <ประกอบขึ้นเป็นหน่วยเดียว> <มีสภาพเป็นนิติบุคคล> <เพื่อทำธุรกิจ>	"บริษัท"* , "ธนาคาร"* "บรรษัท", "บริษัท...จำกัด"
	<กลุ่มบุคคล> <มีการรวมตัวกัน> <ประกอบขึ้นเป็นหน่วยเดียว> <มีสภาพเป็นนิติบุคคล> <เพื่อทำกิจการตาม วัตถุประสงค์ใดวัตถุประสงค์ หนึ่ง>	"พรรค" , "สำนักงาน"* "สำนัก", "องค์กร", "สหภาพ" "สภา", "มูลนิธิ"* , "สถาบัน" "ศูนย์"
ชื่อสถานที่	<พื้นที่หรือบริเวณ>	"ร้านกาแฟ" , "ร้านอาหาร"

\* คำนำหน้าชื่อที่สามารถจัดให้ชื่อเฉพาะอยู่ในประเภทที่ต่างกัน 2 ประเภท

	<p>&lt;สร้างหรือกำหนดขึ้น&gt;          &lt;สำหรับใช้ประโยชน์หรือทำ          กิจกรรม&gt;</p>	<p>"สวนสัตว์" , "ค่าย"          "สถานทูต" , "สนามกอล์ฟ"          "ท่าอากาศยาน" , "สุเหร่า"          "คุก" , "มัธยม" , "โรงเรียน"          "สวนสาธารณะ" , "ร.ร." , "พระ          ธาตุ" , "สโมสร" , "สมาคม"          "อ่างเก็บน้ำ" , "ทำน้ำ"          "ท่าเรือ" , "ท่าเรือน้ำลึก"          "ปราสาท" , "ปราสาทหิน"          "วัด" , "วัง" , "รพ."          "อนุสาวรีย์" , "พระบรมราชานุ          สาวรีย์" , "เขื่อน" , "อนุสรณ์"          "อาคาร" , "ตึก" , "โรงแรม"          "สนามบิน" , "สนามม้า"          "แฟลต" , "ซอย" , "ร้าน"          "คณะชุมชน" , "โรงงาน"          "สวน" , "ศาลเจ้า" , "สน."          "ถ." , "ศูนย์ประชุม"          "หอประชุมอำเภอ" , "ทางด่วน          สาย" , "ถนนสาย" , "สถานี"          "สำนักงานสาขา"          "หอประชุม" , "ที่ว่าการ"          "ถนน" , "ที่ทำการสาขา"          "ห้อง" , "หอ" , "ศาลา" , "ที่          ทำ          การ" , "ตลาดสดเทศบาล"          "ตลาดสด" , "ตลาด" , "สนาม          กีฬาเทศบาล" , "สนามกีฬา          กลาง"          "สนามกีฬา"          "สนาม" , "มหาวิทยาลัย"          "โรงพยาบาล"</p>
--	---	--

	<พื้นที่หรือบริเวณ> <สร้างหรือกำหนดขึ้น> <เป็นเขตปกครอง>	"ทวีป" , "สาธารณรัฐ" "ราชอาณาจักร" , "จังหวัด" "มณฑล" , "นคร" , "ต." , "ประเทศ" , "แขวง" , "หมู่บ้าน" "บ้าน" , "เขตเทศบาลตำบล" "อำเภอ" , "ภาค" , "จ." , "เมือง" "กรุง" , "เขต" , "อ."
	<พื้นที่หรือบริเวณ> <สร้างหรือกำหนดขึ้น> <ระบุตำแหน่งหรือบริเวณ>	"ภูมิภาค"
	<พื้นที่หรือบริเวณ> <เกิดขึ้นตามธรรมชาติ> <ส่วนพื้นน้ำ>	"บึง" , "แม่น้ำ" , "ลุ่มน้ำ" "คลอง"
	<พื้นที่หรือบริเวณ> <เกิดขึ้นตามธรรมชาติ> <ส่วนพื้นดิน>	"เกาะ" , "เขา" , "อุทยาน"

จากตารางที่ 4 จะพบว่ามีคำนำหน้าชื่อเฉพาะที่สามารถจัดประเภทชื่อเฉพาะให้อยู่ในสองประเภทได้ โดยอาจจะกำกวมอยู่ระหว่างชื่อเฉพาะประเภทชื่อองค์กรและชื่อสถานที่ ซึ่งคำนำหน้าชื่อดังกล่าวได้แก่ "กรม" , "กระทรวง" , "บริษัท" , "ธนาคาร" , "มูลนิธิ" , "สำนักงาน"

จากคำนำหน้าชื่อทั้ง 6 คำนี้ หากพบว่ามีหลักฐานภายนอก (external evidence) ที่ได้มาจากบริบทข้างเคียง เช่น "พื้นที่" , "บริเวณ" , "ที่" , "รอบ" เป็นต้น ปรากฏอยู่ด้านหน้าจะทำให้สามารถนำมาใช้ในความหมายอ้างอิงถึงสถานที่ ดังตัวอย่าง

"เวลา 09.00น. วันเดียวกัน ที่ กรม<location>ประชาสัมพันธ์</location>"

"ไปไหว้สิ่งศักดิ์สิทธิ์ ที่ กระทรวง<location>มหาดไทย</location>ต่อไป"

"พื้นที่ บริษัท<location>ไทยเมล่อน</location>"

"บริเวณ รอบ บริษัท<location>ไทยเมล่อน</location>เป็นพื้นที่ของนักการเมือง"

ดังนั้นหลักฐานภายนอกที่ได้มาจากบริบทข้างเคียงจึงเป็นส่วนสำคัญอีกส่วนหนึ่งที่จะนำมาใช้ประกอบการเขียนกฎเพื่อให้สามารถระบุประเภทของชื่อเฉพาะได้อย่างแม่นยำ

#### 4.1.2 หลักฐานภายนอก

หลักฐานภายนอก (external evidence) เป็นหลักฐานที่สามารถหาได้จากบริบทข้างเคียงที่มีชื่อเฉพาะนั้นๆ ปรากฏอยู่ด้วย ซึ่งจะสามารถหาได้จากการพิจารณาลักษณะของรูปภาพในคลังข้อมูลและดึงหลักฐานที่ได้มาใช้เขียนกฎ ซึ่งจากคลังข้อมูลพบว่าหลักฐานภายนอกนี้มีความหลากหลาย อีกทั้ง หลักฐานแบบนี้จะไม่สามารถคาดเดาจำนวนของกลุ่มคำได้ (opened class)

หลักฐานภายนอกนี้ยังสามารถจำแนกและลดความกำกวมในการจำแนกประเภทของชื่อเฉพาะได้ ดังตัวอย่าง

“ไปไหว้สิ่งศักดิ์สิทธิ์ ที่ กระทรวง<location>มหาดไทย</location>ต่อไป”

ซึ่ง “กระทรวง” เป็นคำนำหน้าชื่อที่บ่งบอกประเภทของชื่อเฉพาะว่าเป็นชื่อองค์กร แต่ในบริบทดังตัวอย่าง คำว่า “ที่” จะช่วยบ่งชี้ได้ว่า “กระทรวงมหาดไทย” ในบริบทนี้ เป็นชื่อสถานที่มิใช่ชื่อองค์กร คำว่า “ที่” นี้จึงจัดเป็นหลักฐานภายนอก

นอกจากนั้น หลักฐานภายนอกยังเป็นส่วนสำคัญในการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของขอบเขตชื่อเฉพาะได้อีกด้วย โดยสำหรับกรณีที่ชื่อเฉพาะนั้นๆ ไม่มีคำนำหน้าชื่อเป็นหลักฐานภายในเพื่อบ่งบอกจุดเริ่มต้นของขอบเขตชื่อเฉพาะ หลักฐานภายนอกก็จะมีส่วนช่วยในการแก้ปัญหานี้ได้ เช่นตัวอย่าง

“ที่ดินย่าน<location>เกียกกาย</location>”

จะเห็นได้ว่า “เกียกกาย” ซึ่งเป็นชื่อเฉพาะประเภทสถานที่ ไม่มีคำนำหน้าชื่อเป็นหลักฐานภายในเพื่อบ่งบอกจุดเริ่มต้นของขอบเขตชื่อเฉพาะ ดังนั้น การใช้หลักฐานภายนอกที่ได้จากบริบทข้างเคียงอย่างคำว่า “ย่าน” จะมีส่วนช่วยในการระบุตำแหน่งและจำแนกประเภทของชื่อเฉพาะ

#### 4.2 การสร้างกฎการจำแนกประเภทชื่อเฉพาะ

ในส่วนนี้จะกล่าวถึงการสร้างกฎเพื่อใช้ในการตัดสินใจว่ารายการชื่อที่ได้จากวิธีการสกัดนั้น รายการไหนเป็นชื่อเฉพาะและเป็นชื่อเฉพาะประเภทใด โดยอาศัยข้อมูลจากหลักฐานภายในและหลักฐานภายนอกดังที่กล่าวมาแล้ว ซึ่งจากการพิจารณาลักษณะของรูปภาพในคลังข้อมูลพบว่าการระบุตำแหน่งเพื่อหาขอบเขตของชื่อเฉพาะภาษาไทยเป็นไปได้ยาก เพราะภาษาไทยไม่มีสัญลักษณ์หรือลักษณะพิเศษ เช่น ตัวพิมพ์ใหญ่ (capitalization) ที่สามารถใช้ระบุจุดเริ่มต้นของชื่อเฉพาะได้เหมือนในภาษาอังกฤษ อีกทั้งยังไม่มีกรเว้นวรรคระหว่างคำเพื่อให้การระบุจุดสิ้นสุดของขอบเขตชื่อเฉพาะเป็นไปได้ง่ายด้วย อย่างไรก็ตาม จากคลังข้อมูล พบว่าชื่อเฉพาะจะมีหลักฐานที่ได้มาจากหลักฐานภายในหรือคำนำหน้าชื่อนำอยู่ด้านหน้า โดยพบว่าชื่อเฉพาะประเภทชื่อคนจะมีหลักฐานชนิดนี้มากที่สุด โดยจะมีคำนำหน้าชื่อถึง 5408 จากจำนวนชื่อคน 5431 ชื่อ

โดยคิดเป็น 99.577% ชื่อองค์กรมีคำนำหน้าชื่อ 3140 จากจำนวนชื่อองค์กร 3222 ชื่อโดยคิดเป็น 97.455% และชื่อสถานที่ที่มีคำนำหน้าชื่ออยู่ 2158 จากจำนวนชื่อสถานที่ 3030 ชื่อโดยคิดเป็น 71.221% จากหลักฐานภายในที่เห็นได้จากคำนำหน้าชื่อนี้เองทำให้รูปแบบของกฎจะเริ่มด้วยการใช้คำนำหน้าชื่อเป็นหลักในการกำหนดจุดเริ่มต้นของขอบเขตของชื่อเฉพาะ โดยกฎจะมีลักษณะเป็นไปดังนี้

“หากพบ [คำนำหน้าชื่อ] คำที่ตามหลังก็จะเป็น [ชื่อเฉพาะ]”

อย่างไรก็ตาม จากการพิจารณาคำหลังข้อมูลใช้ฝึก พบว่า รายการของคำนำหน้าชื่อ บางส่วนก็มิได้นำหน้าชื่อเฉพาะเพียงอย่างเดียว รายการของคำดังกล่าวยังนำหน้าคำทั่วไปด้วย เช่น “พรรค” อาจจะนำหน้าคำว่า “ขนาด” ก็ได้ โดย “ขนาด” มิได้เป็นชื่อเฉพาะแต่อย่างใด ดังตัวอย่าง

“จนจะกลายเป็นพรรคขนาดเล็ก”

ดังนั้น กฎที่ใช้อิงกับคำนำหน้าชื่อนี้จึงเพิ่มกฎการตรวจสอบคำที่อาจตามหลังคำนำหน้าชื่อแต่ไม่ได้เป็นชื่อเฉพาะ เพื่อคัดเอารายการของกลุ่มพยางค์ที่ไม่น่าจะใช้ชื่อเฉพาะออกไป

ซึ่ง [รายการของคำยกเว้นหลัง] อาจประกอบด้วยคำหลายประเภท ตามการกำกับหมวดคำในแนวพระยาอุปกิตศิลปสาร (อุปกิตศิลปสาร, 2541) ได้แก่

คำนำหน้ากริยา เช่น “คง”, “จะ”, “อย่า”

คำกริยา เช่น “สั่ง”, “ประกาศ”, “กล่าว”, “เห็น”

คำวิเศษณ์ เช่น “ใหญ่”, “เล็ก”, “เหม็น”

คำประพันธ์สรรพนาม เช่น “ที่”, “ซึ่ง”, “อัน”

คำบุพบท เช่น “ทั้ง”

คำสันธาน เช่น “กับ”, “และ”, “เพราะ”, “ก็”

เหตุที่เป็นเช่นนี้ เพราะคำนำหน้าชื่อ เช่น “พรรค” นอกจากจะใช้นำหน้าชื่อเฉพาะแล้ว ยังเป็นคำนามด้วย ซึ่งลักษณะแบบนี้ทำให้คำที่จะตามหลังคำนำหน้าชื่อเฉพาะนี้ มีโอกาสที่จะเป็นส่วนขยายของคำนาม เช่น “พรรคเล็กๆ” หรือเป็นคำในหมวดอื่นๆ ก็ได้ เช่น คำกริยา คำสันธาน คำบุพบท เป็นต้น (ดังที่แสดงไว้ด้านบน) เพราะถือเป็นการเริ่มต้นวลีใหม่ที่ไม่ใช่คำนามวลีเดิม

ดังนั้น กฎจึงจำเป็นต้องมีส่วนที่จะคัดกรอง [รายการของคำยกเว้นหลัง] เหล่านี้ออก เพื่อให้ความแม่นยำและความครบถ้วนในการรู้จำและจำแนกประเภทชื่อเฉพาะเป็นไปได้ดีขึ้น ผู้วิจัยจึงได้เลือกรายการคำยกเว้นเหล่านี้จากคำที่เป็นคำที่ใช้ทั่วไปและไม่พบเป็นส่วนเริ่มต้นของชื่อเฉพาะ (แต่ปัญหาของวิธีนี้ก็ทำให้ชื่อใหม่ๆ ที่อาจเกิดขึ้นและนำหน้าด้วยคำเหล่านี้ เช่น “พรรคใหญ่เหนือฟ้า” ถูกกันออกไปด้วย) ทำให้รูปแบบของกฎจะเป็นดังนี้



“พบ [คำนำหน้าชื่อ] และคำที่ตามหลังมิได้อยู่ใน [รายการของคำยกเว้นหลัง] คำที่ตามหลังก็จะเป็น [ชื่อเฉพาะ]”

ซึ่ง [รายการของคำยกเว้นหลัง] เช่นคำว่า “ขนาด” จะแสดงไว้ในภาคผนวก

นอกจากกรณีดังกล่าวไปแล้วนั้น ในคลังข้อมูลยังพบว่ามีการที่พบคำนำหน้าชื่อซึ่งเป็นหลักฐานภายในแล้ว และคำที่ตามหลังก็อาจมองว่าเป็นชื่อเฉพาะได้ แต่ก็ไม่สามารถจัดให้คำตามหลังเหล่านั้นเป็นชื่อเฉพาะได้ เช่น

หัวหน้า “พรรค” ประชาธิปัตย์

ในที่นี้จะไม่ถือว่า “ประชาธิปไตย” ในที่นี้เป็นชื่อองค์กร เนื่องจากคำที่นำหน้า “พรรค” นั้นคือ คำว่า “หัวหน้า” ซึ่งเป็นหลักฐานภายนอกที่บ่งบอกให้ทราบว่า “หัวหน้าพรรคประชาธิปัตย์” นี้เป็นตำแหน่งทางการเมืองที่อาจอ้างถึงบุคคล มิได้ใช้ชื่อองค์กร ซึ่งจากรูปภาษาดังกล่าวทำให้การพิจารณา “หัวหน้าพรรคประชาธิปัตย์” นั้นจะไม่ได้มองว่า “พรรค” เป็นส่วนหลักของวลี แต่จะมองที่ “หัวหน้า” ซึ่งเป็นส่วนหลัก (head) ของวลีแทน ซึ่งจากลักษณะดังกล่าวนี้ ยังมีกลุ่มคำที่จัดอยู่ในลักษณะเดียวกันนี้ ได้แก่ กลุ่มคำในรายการของคำยกเว้นหน้าของชื่อเฉพาะประเภทองค์กร ดังแสดงในภาคผนวก

และจากรูปแบบภาษาดังกล่าว จะทำให้กฎมีการเปลี่ยนแปลงรูปแบบ โดยจะเป็นดังที่แสดงด้านล่าง

“พบ [คำนำหน้าชื่อ] และคำที่นำหน้ามิได้อยู่ใน [รายการของคำยกเว้นหน้า] และคำที่ตามหลังมิได้อยู่ใน [รายการของคำยกเว้นหลัง] คำที่ตามหลังก็จะเป็น [ชื่อเฉพาะ]”

จากกฎข้างต้นจะทำให้สามารถระบุได้เพียงตำแหน่งเริ่มต้นของชื่อเฉพาะเท่านั้น แต่ในการหาขอบเขตสิ้นสุดของชื่อเฉพาะนั้น ในภาษาไทยซึ่งไม่มีสัญลักษณ์หรือลักษณะพิเศษเช่น การเว้นวรรค เช่นในภาษาอังกฤษ จึงทำให้การระบุหาขอบเขตสิ้นสุดของชื่อเฉพาะเป็นไปได้ยาก แต่อย่างไรก็ตาม จากการพิจารณาคคลังข้อมูล พบว่าสามารถใช้หลักฐานภายนอกที่ได้มาจากบริบทข้างเคียงเป็นการบอกจุดสิ้นสุดของขอบเขตของชื่อเฉพาะได้ แต่รายการของหลักฐานภายนอกดังกล่าว ก็จะมีขนาดใหญ่ เพราะไม่อาจจะคาดเดาได้ว่าจะมีคำจากกลุ่มคำประเภทใดบ้างที่จะเป็นจุดสิ้นสุดของชื่อเฉพาะ ทำให้รายการของหลักฐานภายนอกที่สามารถระบุจุดสิ้นสุดของขอบเขตชื่อเฉพาะซึ่งได้มาจากการพิจารณาคคลังข้อมูลใช้ฝึกนี้ จะสามารถใช้กับงานวิจัยนี้เท่านั้น หากมีการเพิ่มเติมขนาดของคลังข้อมูล หรือเปลี่ยนแปลงรูปแบบของข้อมูลในคลังข้อมูล อาจทำให้รายการของคำที่สามารถระบุจุดสิ้นสุดของขอบเขตชื่อเฉพาะนี้ใช้ไม่ได้ผลเต็มที่

ซึ่งในงานวิจัยนี้ พบว่ารายการของคำที่ได้มาจากหลักฐานภายนอกจากบริบทข้างเคียงเป็นตัวกำหนด ดังเช่นรายการของกลุ่มคำทั่วไปที่หาได้จากคลังข้อมูลใช้ฝึก เช่น คำว่า “และ”

“อย่า” เป็นต้น ซึ่งคำเหล่านี้แสดงว่าได้มีการขึ้นต้นวลีใหม่จึงใช้บอกจุดสิ้นสุดได้ หรืออาจเป็นกลุ่มของคำนำหน้าชื่อที่ตามมาด้านหลังเอง ดังตัวอย่าง “พ.ต.ท. ทักษิณ ชินวัตร นาง สิริกร” ซึ่งจุดเริ่มของชื่อจะเริ่มจาก “ทักษิณ” และจบที่ “ชินวัตร” และเมื่อพบคำนำหน้าชื่อว่า “นาง” อีกคำก็จะรู้ว่าชื่อแรกได้จบแล้ว จึงสามารถนำมาใช้เพื่อเป็นจุดสิ้นสุดของขอบเขตของชื่อเฉพาะได้

ซึ่งกลุ่มคำที่ถูกรู้ใช้เป็นจุดสิ้นสุดของขอบเขตชื่อเฉพาะนี้ นอกจากกลุ่มคำนำหน้าชื่อแล้วยังอาจแบ่งกลุ่มคำออกเป็น 7 ประเภท ตามการกำกับหมวดคำในแนวพระยาอุปกิตศิลปสาร (อุปกิตศิลปสาร, 2541) ได้แก่

1. คำนาม เช่น “คน”, “บ้าน”, “บริษัท”, “ตัว”
2. คำนำหน้ากริยา เช่น “คง”, “จะ”, “อย่า”
3. คำกริยา เช่น “กล่าว”, “เปิด”, “เห็น”, “อยาก”, “สู้”, “เป็น”, “เหมือน”, “คล้าย”, “ทราบ”, “ตอบ”, “ยื่น”, “ประกาศ”
4. คำวิเศษณ์ เช่น “ใกล้”, “ไกล”, “มาก”, “เล็ก”
5. คำประพันธ์สรรพนาม เช่น “ที่”, “ซึ่ง”, “อัน”
6. คำบุพบท เช่น “ที่”, “ทั้ง”, “ประมาณ”, “ตลอด”
7. คำสันธาน เช่น “กับ”, “และ”, “เพราะ”, “ก็”, “ส่วน”

ซึ่งสำหรับกรณีคำที่ใช้ในการบอกขอบเขตของชื่อเฉพาะจำนวนหนึ่งเป็นคำนาม ยกตัวอย่างเช่น

“บ้านเมืองไม่ใช่ของพ.ต.ท.<person>ทักษิณชินวัตร</person>คนเดียว”

ซึ่งในกรณีนี้ “คน” เป็นลักษณะนามของ “พ.ต.ท.ทักษิณ ชินวัตร”

สำหรับกรณีคำที่ใช้ในการบอกขอบเขตสิ้นสุดของชื่อเฉพาะจำนวนหนึ่งเป็นคำกริยา ทั้งนี้เป็นเพราะชื่อเฉพาะเมื่ออยู่ในประโยคอาจจะทำหน้าที่เป็นประธานของประโยคทำให้เมื่อพบคำกริยาจึงทำให้ทราบว่าพบจุดสิ้นสุดของขอบเขตชื่อเฉพาะแล้ว เช่น

“พ.ต.ท.<person>ทักษิณชินวัตร</person>ประกาศนโยบายของพรรคชุดใหม่”

สำหรับกรณีคำที่ใช้ในการบอกขอบเขตของชื่อเฉพาะจำนวนหนึ่งเป็นคำวิเศษณ์นั้น ทั้งนี้เป็นเพราะชื่อเฉพาะที่มีลักษณะเหมือนคำนามที่ปรากฏหลังคำกริยาจึงอาจตามด้วยคำวิเศษณ์ที่ขยายคำกริยานั้นได้ เช่น

“ทำให้พรรค<organization>ประชาธิปไตย</organization>เล็กลง”

สำหรับกรณีคำที่ใช้ในการบอกขอบเขตของชื่อเฉพาะจำนวนหนึ่งเป็นคำบุพบทนั้น ทั้งนี้เป็นเพราะชื่อเฉพาะอาจทำหน้าที่เป็นกรรมของประโยคได้ ดังนั้นคำบุพบททำหน้าที่เป็นคำเชื่อม

ระหว่างความที่อาจจะมาขยายกริยาของประโยค จึงทำให้สามารถมองเห็นจุดสิ้นสุดของขอบเขตชื่อเฉพาะได้ เช่น

“ปิดเรื่องของตนกับเรื่องของพรรค<organization>ไทยรักไทย</organization>

ตลอดเวลา”

ซึ่งในที่นี้ “ตลอด” เป็นบุพบทบอกเวลา

สำหรับกรณีที่ว่าที่ใช้ในการบอกขอบเขตของชื่อเฉพาะจำนวนหนึ่งเป็นคำสันธานนั้น ทั้งนี้เป็นเพราะคำสันธานจะทำหน้าที่เชื่อมความให้ต่อกัน ทำให้นอกจากจะทราบว่าประโยคสิ้นสุดลงแล้ว ชื่อเฉพาะที่มีตำแหน่งอยู่ท้ายประโยคก็จะพบจุดสิ้นสุดของขอบเขตชื่อเฉพาะด้วย เช่น

“พรรค<organization>ประชาธิปไตย</organization>ส่วนกรณีอื่น”

จากการพิจารณาคำข้อมูลและพบว่ากลุ่มคำซึ่งถูกใช้เป็นจุดสิ้นสุดของขอบเขตชื่อเฉพาะนี้มีความหลากหลายมาก โดยอาจจัดอยู่ในประเภทต่างๆ ตามที่ได้กล่าวไปแล้ว อีกทั้งยังมีจำนวนมากซึ่งไม่อาจคาดเดาจำนวนทั้งหมดได้ ทั้งนี้เป็นเพราะกลุ่มคำซึ่งถูกใช้เป็นจุดสิ้นสุดของขอบเขตชื่อเฉพาะนี้ อาจเป็นคำขยายคำนาม ซึ่งชื่อเฉพาะก็จัดว่าเป็นคำนามประเภทหนึ่ง หรืออาจเป็นการสิ้นสุดนามวลี แล้วเริ่มต้นวลีอื่นๆ ดังนั้นกลุ่มคำซึ่งถูกใช้เป็นจุดสิ้นสุดของขอบเขตชื่อเฉพาะนี้จึงมีได้มากมาย

เมื่อพิจารณาคำข้อมูลใช้ฝึก จนได้รายการของกลุ่มคำบอกจุดสิ้นสุดมาแล้ว ดังนั้นรูปแบบของกฎจะเพิ่มส่วนของการหาจุดสิ้นสุดของชื่อเฉพาะเป็นดังนี้

“พบ [คำนำหน้าชื่อ] และคำที่นำหน้ามีได้อยู่ใน [รายการของคำยกเว้นหน้า] และคำที่ตามหลังมีได้อยู่ใน [รายการของคำยกเว้นหลัง] คำที่ตามหลังก็จะเป็น [ชื่อเฉพาะ] และมีจุดสิ้นสุดของขอบเขตชื่อเฉพาะเมื่อพบ [รายการของคำบอกจุดสิ้นสุดหรือรายการของคำนำหน้าชื่อ]”

จากกฎที่กล่าวไปข้างต้น เมื่อนำมาใช้กับชื่อเฉพาะประเภทชื่อบุคคลนั้น กฎจะถูกลดให้เหลือเพียง

กฎสำหรับชื่อบุคคล : “พบ [คำนำหน้าชื่อ] คำที่ตามหลังก็จะเป็น [ชื่อเฉพาะ] และมีจุดสิ้นสุดของขอบเขตชื่อเฉพาะเมื่อพบ [รายการของคำบอกจุดสิ้นสุดหรือรายการของคำนำหน้าชื่อ]”

ทั้งนี้เป็นเพราะ [คำนำหน้าชื่อ] ของชื่อเฉพาะประเภทคนนั้นไม่พบว่าปรากฏนำหน้าคำทั่วไปอื่นๆ เช่น “นาย” จะไม่นำหน้าคำทั่วไปอย่าง “ฝ่าย” เพราะฉะนั้นจึงทำให้กฎเป็นไปตามที่กล่าวมาแล้ว

ส่วนชื่อองค์กรและชื่อสถานที่จะเป็นไปตามกฎนี้

กฎสำหรับชื่อองค์กร / ชื่อสถานที่ : "พบ [คำนำหน้าชื่อ] และคำที่นำหน้ามิได้อยู่ใน [รายการของคำยกเว้นหน้า] และคำที่ตามหลังมิได้อยู่ใน [รายการของคำยกเว้นหลัง] คำที่ตามหลังก็จะเป็น [ชื่อเฉพาะ] และมีจุดสิ้นสุดของขอบเขตชื่อเฉพาะเมื่อพบ [รายการของคำบอกจุดสิ้นสุดหรือรายการของคำนำหน้าชื่อ]"

อย่างไรก็ตาม กฎสำหรับชื่อองค์กรและชื่อสถานที่อาจปรับไปตามสิ่งที่พบจากคลังข้อมูลใช้ฝึก โดยอาจมีหรือไม่มีส่วนของ [รายการของคำยกเว้นหน้า] ซึ่งนำหน้าคำนำหน้าชื่อ หรืออาจมีหรือไม่มีส่วนของ [รายการของคำยกเว้นหลัง] ที่ตามหลังคำนำหน้าชื่อ โดย [รายการของคำยกเว้นหน้า] และ [รายการของคำยกเว้นหลัง] ก็จะแตกต่างกันไปตามคำนำหน้าแต่ละคำที่ใช้ด้วย

ซึ่งรายการของคำยกเว้นหน้า คำยกเว้นหลัง และคำบอกจุดสิ้นสุดดังกล่าวจะแสดงอยู่ในภาคผนวก

จากกฎที่กล่าวไปข้างต้น จะช่วยให้ระบุตำแหน่งของชื่อเฉพาะได้ และจากคำนำหน้าชื่อทั้ง 3 ประเภทก็จะทำให้จำแนกประเภทของชื่อเฉพาะดังกล่าวได้ อย่างไรก็ตาม จากคลังข้อมูลใช้ฝึกยังพบว่ามีกรณีชื่อเฉพาะนั้นๆ ไม่มีคำนำหน้าชื่อ ซึ่งจากการพิจารณาค้นคลังข้อมูลใช้ฝึก พบว่า ชื่อเฉพาะประเภทชื่อสถานที่จะมีกรณีที่ชื่อเฉพาะไม่มีคำนำหน้าชื่อมากที่สุดจากชื่อเฉพาะทั้ง 3 ประเภท กล่าวคือกรณีที่ชื่อสถานที่ไม่มีคำนำหน้าชื่อจะมีถึง 28.779% โดยที่ชื่อคนจะไม่มีคำนำหน้าชื่อเพียง 0.423% และชื่อองค์กรจะไม่มีคำนำหน้าชื่อเพียง 2.545% ดังนั้นเพื่อเป็นการเพิ่มประสิทธิภาพของระบบที่ใช้กฎซึ่งอิงคำนำหน้าชื่อเป็นหลัก ผู้วิจัยจึงพิจารณานานุกรณภายนอกซึ่งได้มาจากบริบทข้างเคียงเพื่อช่วยระบุจุดเริ่มต้นของขอบเขตของชื่อเฉพาะในกรณีที่ชื่อเฉพาะนั้นๆ ไม่มีคำนำหน้าชื่อเป็นหลักฐานภายในเพื่อบ่งบอกขอบเขต

ดังตัวอย่าง "ที่ดินย่าน<location>เกียกกาย</location>"

จะเห็นได้ว่า "เกียกกาย" ซึ่งเป็นชื่อเฉพาะประเภทสถานที่ไม่มีคำนำหน้าชื่อเป็นหลักฐานภายในเพื่อบ่งบอกจุดเริ่มต้นของชื่อเฉพาะ ดังนั้น การใช้หลักฐานภายนอกที่ได้จากบริบทข้างเคียงอย่างคำว่า "ย่าน" จะมีส่วนช่วยในการระบุตำแหน่งและจำแนกประเภทของชื่อเฉพาะ ซึ่งบริบทข้างเคียงดังกล่าวสามารถหาได้จากการใช้โปรแกรม Collocation Extract 3.06 ซึ่งสามารถดาวน์โหลดจากเว็บไซต์ <http://pioneer.chula.ac.th/~awirote/colloc/index.html> (Aroonmanakun, 2004) โดยวิธีทางสถิติที่เลือกใช้โปรแกรมคือวิธีการหาค่า Mutual Information (MI) เช่นเดียวกับวิธีทางสถิติที่เลือกใช้ในงานวิจัยนี้ เพื่อหาคำที่ปรากฏร่วมกับชื่อสถานที่ที่มากเป็นพิเศษและนำมาประกอบการใช้กฎเพื่อทำให้ประสิทธิภาพของกฎเป็นไปได้ดีขึ้น และรายการคำที่หาได้จะมาจากบริบทข้างเคียงหรือ external evidence เพื่อนำไปใช้เป็นตัวกำหนดขอบเขตเริ่มต้นของชื่อเฉพาะประเภทสถานที่ ได้แก่คำว่า "ย่าน" , "ฝั่ง" , "เขื่อน" , "บริเวณ" , "ห้อง" , "ลาน" ซึ่ง

หลักฐานภายนอกที่ได้มาจากบริบทข้างเคียงเหล่านี้นอกจากจะช่วยลดจำนวนรายการของกลุ่มพยางค์ที่อาจเป็นชื่อเฉพาะ (candidate) ลงแล้ว ยังช่วยจำแนกประเภทของชื่อเฉพาะให้มีความแม่นยำมากขึ้นด้วย

ดังนั้น รูปแบบของกฎจะเป็นดังนี้

กฎสำหรับชื่อสถานที่ : "พบ [คำนำหน้าชื่อหรือคำ 6 คำที่ได้จากหลักฐานภายนอก] และคำที่นำหน้ามีได้อยู่ใน [รายการของคำยกเว้นหน้า] และคำที่ตามหลังมีได้อยู่ใน [รายการของคำยกเว้นหลัง] คำที่ตามหลังก็จะเป็น [ชื่อเฉพาะ] และมีจุดสิ้นสุดของขอบเขตชื่อเฉพาะเมื่อพบ [รายการของคำบอกจุดสิ้นสุดหรือรายการของคำนำหน้าชื่อ]"

#### 4.3 ผลการตัดสินใจและจำแนกประเภทชื่อเฉพาะ

ชื่อเฉพาะที่เลือกมาด้วยวิธีการทางสถิติที่ได้กล่าวไว้แล้วในบทที่ 3 จะถูกนำมาตัดสินใจด้วยโปรแกรมกฎที่เขียนขึ้นเพื่อระบุและจำแนกประเภทของชื่อเฉพาะตามที่ได้อธิบายในตอนที่แล้ว ผลทดสอบที่ได้จะแสดงในตารางที่ 5 โดยคอลัมน์ที่ 3 จะแสดงจำนวนชื่อที่โปรแกรมกฎตัดสินใจเป็นชื่อเฉพาะ และในจำนวนนี้มีจำนวนหนึ่งที่โปรแกรมตัดสินใจถูกต้องซึ่งแสดงไว้ในคอลัมน์ที่ 2 ส่วนคอลัมน์ที่ 4 เป็นจำนวนชื่อเฉพาะจริงที่พบในข้อมูล และในตารางจะแสดงผลแยกเป็นชื่อเฉพาะแต่ละประเภท เพราะกฎที่เขียนขึ้นนั้นเป็นกฎที่แยกตามประเภทของชื่อเฉพาะ

ตารางที่ 5 ผลการทดสอบค่าความแม่นยำ ค่าความครบถ้วน และค่า F เมื่อใช้โปรแกรมกฎ

ประเภทของชื่อเฉพาะ	จำนวนชื่อเฉพาะที่ตัดสินใจถูกต้อง	จำนวนชื่อเฉพาะทั้งหมดที่เลือกมา	จำนวนชื่อเฉพาะทั้งหมด	ค่าความแม่นยำ (%)	ค่าความครบถ้วน (%)	ค่า F
ชื่อคน	3786	7009	3940	54.000	96.117	69.150
ชื่อบุคคล	2854	5996	3071	47.598	92.934	62.953
ชื่อสถานที่	873	2757	1735	31.665	50.317	38.869

จากตารางที่ 5 พบว่าชื่อเฉพาะประเภทชื่อคนและชื่อบุคคลนั้น แม้ว่าจะมีอัตราการทำจำหรือค่า F เพียง 69.150% และ 62.953% ตามลำดับ แต่ก็พบว่าการระบุตำแหน่งและการจำแนก

ประเภทของชื่อเฉพาะทั้ง 2 ประเภท ก็มีค่าความครบถ้วนมากกว่า 90% ซึ่งแสดงให้เห็นว่ากฎที่ใช้สามารถจำแนกประเภทของชื่อเฉพาะได้ดี แต่สำหรับชื่อเฉพาะประเภทชื่อสถานที่ที่มีจำนวนชื่อเฉพาะที่ไม่มีคำนำหน้าชื่อเป็นจำนวนมากที่สุดในชื่อเฉพาะ 3 ประเภทนั้น จะให้ค่าความครบถ้วนน้อยที่สุดด้วย แม้ว่าในกฎจะมีการใช้หลักฐานที่ได้จากบริบทข้างเคียงทั้ง 6 คำได้แก่คำว่า “ย่าน” , “ฝั่ง” , “เขื่อน” , “บริเวณ” , “ท้อง” , “ลาน” มาเสริมกับการใช้คำนำหน้าชื่อก็ทำให้ผลค่าอัตราการรู้จำมีค่าเพียง 38.869%

ตารางที่ 6 ผลการจำแนกประเภทของโปรแกรมกฎที่เขียนขึ้น

ประเภทของชื่อเฉพาะ	ชื่อคน	ชื่อองค์กร	ชื่อสถานที่
ชื่อคน	3786	0	0
ชื่อองค์กร	0	2854	0
ชื่อสถานที่	0	3	873

ในการใช้กฎเพื่อจำและจำแนกประเภทชื่อเฉพาะนั้น อาจเป็นไปได้ว่าโปรแกรมกฎอาจตัดสินใจจำแนกประเภทชื่อเฉพาะผิดไปก็ได้ ดังนั้นผู้วิจัยจึงนับและแสดงผลจำนวนการตัดสินใจผิดพลาดไว้ในตารางที่ 6

จากข้อมูลในตาราง พบว่ากฎที่ใช้สามารถจำแนกประเภทของชื่อเฉพาะประเภทชื่อคน และชื่อองค์กรได้ถูกต้องโดยไม่พบการรู้จำและจำแนกชื่อเฉพาะผิดประเภท แต่สำหรับชื่อสถานที่นั้น พบว่ามีกรณีที่กฎจำแนกประเภทให้ชื่อสถานที่ผิดประเภทไปเป็นชื่อองค์กร โดยชื่อเฉพาะที่ถูกจำแนกผิดประเภทดังกล่าวพบว่าเป็นกรณีที่มีการใช้คำนำหน้าชื่อเฉพาะเป็นตัวบ่งบอก และคำนำหน้าชื่อนั้นมีความกำกวม คือสามารถนำหน้าชื่อองค์กรหรือชื่อสถานที่ก็ได้ (คำว่า “กระทรวง” และ “กรม”) ทั้งสามกรณีที่ตัดสินใจผิด ได้แก่

“ผู้สื่อข่าวรายงานจาก กระทรวงเกษตรและสหกรณ์ ว่าเมื่อวันที่ 13 มิ.ย.”

“เวลา 09.00 น. วันเดียวกันที่ กรมประชาสัมพันธ์”

“สำหรับสถานศึกษาที่ทดลองปลูกพืชจีเอ็มโอทั้ง 3 แห่ง ได้แก่ กรมวิชาการเกษตร”

ซึ่งสาเหตุที่โปรแกรมกฎรู้จำและจำแนกประเภทได้ไม่ผิดประเภทนั้น เป็นเพราะกฎที่เขียนขึ้นใช้ในงานนั้นๆ ได้ปรับให้เข้ากับข้อมูลเป็นอย่างดี จึงมีกรณีที่ผิดพลาดเพียงสามตัวอย่างดังที่แสดงไว้

#### 4.4 อภิปรายผล

ผลที่ได้จากตารางที่ 5 แสดงให้เห็นว่าชื่อเฉพาะทั้งหมดที่เลือกมาด้วยวิธีการทางสถิติ เพื่อให้โปรแกรมกฎตัดสินว่าเป็นชื่อเฉพาะนั้นมีจำนวนค่อนข้างมาก จึงทำให้ค่าความแม่นยำของการตัดสินในส่วนของการใช้กฎมีค่าน้อย ซึ่งทำให้เมื่อคำนวณอัตราการเรียนรู้หรือค่า F แล้ว ทำให้มีค่าต่ำด้วย ทั้งนี้อาจเป็นเพราะกฎที่เขียนขึ้นนั้นมีข้อจำกัดเรื่องการหาขอบเขตของชื่อเฉพาะ ถึงแม้ว่าการหาจุดเริ่มต้นของชื่อเฉพาะจะอาศัยรายการคำนำหน้าชื่อซึ่งจัดเป็นหลักฐานภายในที่น่าจะให้ผลดี แต่การระบุจุดสิ้นสุดของชื่อเฉพาะนั้นทำได้ไม่ดีนัก ทั้งนี้เพราะชื่อเฉพาะอาจทำหน้าที่เป็นประธานของประโยคหรือเป็นกรรมของประโยคก็ได้ ทำให้รายการของคำบอกจุดสิ้นสุดของขอบเขตชื่อเฉพาะจึงมีความหลากหลายมากตามไปด้วย ซึ่งนอกจากคำบอกจุดสิ้นสุดอาจเป็นคำนำหน้าชื่อแล้ว คำบอกจุดสิ้นสุดยังอาจจัดอยู่ในกลุ่มคำประเภทใดก็ได้ ได้แก่ คำนาม คำนำหน้ากริยา คำกริยา คำประพันธ์สรรพนาม คำวิเศษณ์ คำบุพบท และคำสันธาน โดยหากชื่อเฉพาะทำหน้าที่เป็นประธานของประโยคก็อาจจะทำให้คำบอกจุดสิ้นสุดของขอบเขตชื่อเฉพาะจัดอยู่ในหมวดคำกริยา หรือกรณีที่ชื่อเฉพาะทำหน้าที่เป็นกรรมของประโยค เมื่อจะกล่าวถึงประโยคถัดไป ก็อาจใช้คำสันธานมาเชื่อมต่อกความได้ อีกทั้งชื่อเฉพาะยังอาจมีคำวิเศษณ์มาขยายความได้ด้วย

อีกทั้งการระบุจุดสิ้นสุดของชื่อเฉพาะ ยังอาจมีข้อจำกัด โดยคำบอกจุดสิ้นสุดอาจครอบคลุมกรณีบางกรณี เช่น เราอาจจะกำหนดให้คำว่า “ว่า” เป็นหนึ่งในรายการของคำที่เป็นจุดสิ้นสุดของขอบเขตของชื่อเฉพาะได้ แต่คำว่า “ว่า” จะครอบคลุมเพียงไม่กี่กรณี ซึ่งหากเกิดกรณีนอกเหนือจากนั้น ก็จะทำให้กฎนี้ใช้ไม่ได้ผลและทำให้มีกลุ่มพยางค์ที่อาจเป็นชื่อเฉพาะเพิ่มจำนวนขึ้นมาอีก เช่น

กระทรวง <candidate>การคลังก็รับ</candidate>ว่า

จะสังเกตได้ว่า จากคำนำหน้าชื่อซึ่งเป็นจุดเริ่มต้นคือคำว่า “กระทรวง” ผ่านเงื่อนไขของโปรแกรมเรียบร้อยแล้ว แต่มาพบปัญหาอยู่ที่จุดสิ้นสุดของชื่อเฉพาะทำให้ได้ชื่อเฉพาะที่ไม่ถูกต้องปนเข้ามาอีกตัวหนึ่ง ยังผลให้ค่าความแม่นยำลดลงไปอีก

สำหรับชื่อเฉพาะประเภทสถานที่นั้น นอกจากกฎที่เขียนขึ้นเพื่อจัดการกับชื่อเฉพาะที่มีหลักฐานที่ได้จากบริบทภายในหรือที่มีรายการของคำนำหน้าชื่อแล้ว ก็ยังมีการเพิ่มกฎเพื่อใช้กับกรณีที่ชื่อเฉพาะปรากฏแบบไม่มีคำนำหน้าชื่อนำอยู่ด้านหน้าด้วย ดังนั้น ในกรณีแบบนี้ ผู้วิจัยใช้โปรแกรม Collocation Extract ในการหารายการของคำที่มีความสัมพันธ์กับชื่อเฉพาะประเภท

สถานที่ซึ่งรายการของคำเหล่านี้จัดเป็นหลักฐานที่หาได้จากบริบทข้างเคียง ซึ่งได้ออกมา 6 คำ ดังนี้คือ ย่าน , ฝั่ง , เยือน , บริเวณ , ท้อง , ลาน ซึ่งผลการทดสอบเมื่อรับเอาคำเหล่านี้เข้าไปใช้ในโปรแกรมกฎ พบว่าค่าความแม่นยำ ความครบถ้วน และอัตราการรู้จำหรือค่า F ก็ยังไม่สูงนัก

ส่วนผลของค่าความครบถ้วนนั้น พบว่ากฎที่เขียนขึ้นทำให้สามารถจำแนกประเภทชื่อเฉพาะประเภทชื่อคนกับชื่อองค์กรได้เกิน 90% แต่ให้ผลไม่ดีนักกับชื่อสถานที่ ทั้งนี้เป็นเพราะ ชื่อสถานที่ที่เข้ามาสู่ระบบกฎนั้น ส่วนใหญ่มักไม่มีคำนำหน้าชื่อ เช่น

"ในเขตที่ไม่มีส.ส.ของพรรคเช่น <location>ยานนาวา</location> <location>ดินแดง</location> <location>ห้วยขวาง</location> <location>บางคอแหลม</location> และ <location>สัมพันธวงศ์</location>"

ทำให้การดึงชื่อสถานที่ที่ถูกต้องออกมาโดยโปรแกรมกฎนั้นเป็นไปได้ยากและทำให้ได้ค่าความครบถ้วนและค่าความแม่นยำน้อย ยังผลให้อัตราในการรู้จำหรือค่า F ต่ำด้วย

สำหรับกรณีที่คำนำหน้าชื่อซึ่งถูกใช้เป็นตัวเริ่มต้นของขอบเขตชื่อเฉพาะมีความกำกวมในการจำแนกประเภทชื่อเฉพาะระหว่างชื่อองค์กรและชื่อสถานที่นั้น จากผลในตารางที่ 6 ทำให้ทราบว่าโปรแกรมกฎจำแนกประเภทชื่อเฉพาะที่มีคำนำหน้าชื่อซึ่งกำกวมระหว่างชื่อองค์กรและชื่อสถานที่ ได้แก่คำว่า "กรม" , "กระทรวง" ผิดประเภท ซึ่งชื่อเฉพาะดังกล่าวควรถูกจัดอยู่ในชื่อเฉพาะประเภทชื่อสถานที่ แต่โปรแกรมกฎจัดให้อยู่ในประเภทชื่อองค์กร โดยกรณีการจำแนกประเภทชื่อสถานที่ผิดเป็นชื่อองค์กรจะแสดงไว้อีกครั้งที่ด้านล่างนี้

"ผู้สื่อข่าวรายงานจาก กระทรวง<organization>เกษตรและสหกรณ์</organization> ว่าเมื่อวันที่13มิ.ย."

"เวลา09.00น.วันเดียวกันที่ กรม<organization>ประชาสัมพันธ์</organization>"

"สำหรับสถานศึกษาที่ทดลองปลูกพืชจีเอ็มโอทั้ง 3 แห่งได้แก่ กรม<organization>วิชาการเกษตร</organization>"

ซึ่งจากบริบทข้างเคียงที่ปรากฏรอบชื่อเฉพาะนี้ พบว่าไม่มีคำที่สามารถดึงมาเพื่อใช้เป็นหลักฐานภายนอกเพื่อช่วยระบุประเภทของชื่อเฉพาะประเภทชื่อสถานที่ ดังเช่นคำว่า "บริเวณ" , "รอบ" , "พื้นที่" จึงทำให้กฎไม่สามารถรู้จำและจำแนกประเภทชื่อเฉพาะที่มีความกำกวมได้ แต่อย่างไรก็ตาม พบว่ากรณีการจำแนกประเภทผิดพลาดมีเพียง 3 ชื่อเฉพาะเท่านั้น ซึ่งคิดเป็น 0.17% จากชื่อเฉพาะทั้งหมด

และนอกจากนั้น กรณีดังกล่าวก็ยังพบปัญหาเรื่องการหาขอบเขตสิ้นสุดของชื่อเฉพาะเหมือนชื่อเฉพาะอื่นๆ ที่คำนำหน้าชื่อไม่ได้มีความกำกวม ซึ่งยังผลให้เมื่อคำนวณค่าความแม่นยำ



และค่าความครบถ้วนแล้ว ทำให้อัตราในการรู้จำหรือค่า F มีค่าต่ำกว่า 90% ซึ่งทำให้ระบบมีประสิทธิภาพไม่ดีเท่าที่ควร