

หัวข้อวิทยานิพนธ์	ระบบการวิเคราะห์ประโยคภาษาไทยที่มีการละประธานที่ซ้ำกันในประโยค
นักศึกษา	นายสิงห์ ตรงงาม
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร.รัตติกร วรากุลศิริพันธุ์
ระดับการศึกษา	วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า
ภาควิชา	วิศวกรรมอิเล็กทรอนิกส์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.	2540

บทคัดย่อ

การประมวลผลภาษาธรรมชาติ (Natural Language Processing - NLP) เป็นส่วนหนึ่งของศาสตร์ที่เรียกว่า “ปัญญาประดิษฐ์” (Artificial Intelligence - AI) ซึ่งเป็นงานวิจัยที่ได้รับความสนใจอย่างมาก เพราะเป็นที่คาดหวังว่า จะสามารถนำไปประยุกต์เพื่องานต่างๆ ได้มากมาย เช่น แปลภาษาด้วยเครื่องคอมพิวเตอร์ การให้เครื่องจักรหรือคอมพิวเตอร์เข้าใจภาษามนุษย์ เป็นต้น แต่เมื่อมีการศึกษามาระยะหนึ่ง เริ่มประสบปัญหาทางด้านภาษาศาสตร์มากขึ้น ดังนั้นในปัจจุบันนี้ จึงมีการศึกษาร่วมกันระหว่างการประมวลผลภาษาธรรมชาติกับภาษาศาสตร์ โดยต้องมีการประยุกต์หลักการและโครงสร้างทางภาษาศาสตร์เพื่อนำไปใช้ในการประมวลผลภาษาธรรมชาติหลายระดับ เช่น ระดับหน่วยคำ (morphology) ระดับโครงสร้าง (syntax) ระดับความหมาย (semantics) และระดับการใช้ภาษา (pragmatics) โดยเฉพาะเรื่องของการศึกษาวิเคราะห์โครงสร้างไวยากรณ์ (syntactic analysis) และการวิเคราะห์เชิงความหมาย (semantics)

ปัญหาหนึ่งในการวิเคราะห์โครงสร้างไวยากรณ์และการวิเคราะห์เชิงความหมายของประโยคภาษาไทย คือข้อความภาษาไทยที่ประกอบขึ้นเป็นประโยค ซึ่งมีลักษณะเป็นการนำหน่วยคำมาเขียนติดกัน โดยไม่มีช่องว่างหรือเครื่องหมายระบุการจบของหน่วยคำ บางข้อความจะมีการนำประโยคภาษาไทยหลายประโยคมารวมกันเป็นย่อหน้า และมีการปรับแต่งข้อความให้สวยงามกระชับรัด เช่น การลดคำที่ซ้ำกันในประโยคออก ทำให้การแยกหน่วยคำและการแยกประโยคออกจากย่อหน้าทำได้ยากขึ้น จากปัญหานี้ มีผลทำให้การวิเคราะห์โครงสร้างไวยากรณ์ของประโยคภาษาไทย มีความยุ่งยากซับซ้อนมากขึ้นเช่นเดียวกัน เพราะต้องมีการแยกแยะประโยคภาษาไทยออกจากย่อหน้าให้ถูกต้องเสียก่อน การวิเคราะห์โครงสร้างไวยากรณ์จึงจะมีความถูกต้องสมบูรณ์น่าเชื่อถือมากที่สุด

วิทยานิพนธ์นี้จะเสนอ ระบบการวิเคราะห์ประโยคภาษาไทยที่มีการละประธานที่ซ้ำกันในประโยคออกจากย่อหน้า โดยลักษณะของประโยคในย่อหน้านี้จะประกอบด้วยประโยคที่มีเนื้อความเพียงอย่างเดียว ตั้งแต่ 2 ประโยคขึ้นไปและมีคำสันธานเป็นตัวเชื่อมอยู่ระหว่างประโยคนั้น จากนั้นมีการปรับปรุงทำให้รูปประโยคสั้นลงโดยการละประธานที่ซ้ำซ้อนกันออกไป ซึ่งวิทยานิพนธ์นี้ใช้วิธีการ

วิเคราะห์ประโยคตามวิธีการของ “ไวยากรณ์การก” เป็นหลักในการเลือกความสัมพันธ์ทางด้านโครงสร้างที่ถูกต้องของหน่วยคำ และได้มีการเพิ่มฐานความรู้ในการจัดเรียงตำแหน่งของหน่วยคำและการผสมหน่วยคำในประโยคภาษาไทย โดยที่ฐานความรู้นี้ได้พัฒนาและประยุกต์มาจากโครงสร้างประโยคทางด้านภาษาศาสตร์ เพื่อสามารถแยกประโยคภาษาไทยที่มีการละประธานที่ซ้ำกันในประโยคย่อหน้า ออกมาเป็นประโยคได้อย่างถูกต้อง