

### บทคัดย่อ

ภาษาไทยเป็นภาษาที่มีลักษณะของคำที่ซับซ้อนภาษาหนึ่ง และมีลักษณะการเขียนที่ติดเรียงกันทั้งประโยค ไม่มีการเว้นวรรคระหว่างคำอย่างในภาษาอังกฤษ ดังนั้นการประมวลผลคำภาษาไทยจึงมีความซับซ้อนกว่าภาษาอื่นๆ เนื่องจากลักษณะเฉพาะของภาษา ตัวอย่างเช่น ไม่มีขอบเขตที่ชัดเจนของคำ กลุ่มคำ วลี หรือแม้แต่ประโยค เหตุเพราะไม่ใช้การเว้นวรรคระหว่างคำทุกคำ ไม่ใช้เครื่องหมายวรรคตอนที่ชัดเจนเพื่อบอกจุดสิ้นสุดของวลีและประโยค ไม่ใช้อักษรพิเศษเพื่อแสดงการขึ้นประโยคใหม่หรือแสดงชื่อเฉพาะ ไม่มีการเปลี่ยนรูปคำเพื่อแสดงพจน์ เพศ กาล มาลา การก วาก อีกทั้งยังมีการใช้กริยามากกว่าหนึ่งตัวเรียงกันในประโยค และมีการละคำในประโยคที่สามารถยอมรับได้เกิดขึ้นบ่อยครั้ง ลักษณะที่ซับซ้อนดังกล่าวเป็นอุปสรรคอย่างยิ่งต่อการพัฒนาระบบประมวลผลภาษาไทย ดังนั้นจึงจำเป็นต้องมีการพัฒนาซอฟต์แวร์ที่สามารถแบ่งแยกคำในประโยคให้สามารถเห็นได้อย่างชัดเจนถึงคำแต่ละคำในประโยคว่ามีคำใดบ้าง เพื่อให้คอมพิวเตอร์รู้จักได้ว่าคำแต่ละคำในประโยคว่ามีความหมายหรือลักษณะอย่างไร โดยทฤษฎีการตัดคำไทยโดยเทียบคำที่ยาวที่สุดเป็นทฤษฎีหนึ่งที่สามารถใช้ในการตัดคำไทยได้ประสิทธิภาพดีและมีเปอร์เซ็นต์ความถูกต้องดีที่สุด และเพื่อเป็นการสร้างจุดเริ่มต้นของการต่อยอดในการพัฒนาการวิเคราะห์คำจากเสียงพูดหรืออื่นๆ ที่อีกมากมายจากการตัดคำ