

ชานน อ่อนมัน 2555: การจับคู่การแปลระดับประโยคภาษาไทย – อังกฤษ  
โดยใช้ฟังก์ชันค่าใช้จ่ายแบบถ่วงน้ำหนัก ปริญาวิศวกรรมศาสตรมหาบัณฑิต  
(วิศวกรรมคอมพิวเตอร์) สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์  
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: รองศาสตราจารย์อัครนิษฐ์ ก่อตระกูล, D.Eng. 81 หน้า

การแปลภาษาด้วยเครื่องเป็นการแปลข้อความจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่งโดยอัตโนมัติ ปัจจุบันการแปลภาษาด้วยคอมพิวเตอร์มีบทบาทมากขึ้นเนื่องจากมีความต้องการแปลเอกสารจำนวนมากขึ้นภายใต้เวลาและบุคลากรที่จำกัด เทคนิคที่ใช้ในการแปลภาษาด้วยเครื่องมี 2 เทคนิค ได้แก่ การแปลด้วยกฎ และการแปลด้วยคลังประโยค การแปลด้วยกฎจะรับคำศัพท์และกฎการแปลจากนักภาษาศาสตร์คำนวณโดยตรง ซึ่งจะแตกต่างจากการแปลด้วยคลังประโยคที่จะสกัดความรู้สำหรับการแปลจากตัวอย่างการแปลในคลังประโยคของตนเองโดยอัตโนมัติ การแปลด้วยคลังประโยคจึงสามารถขยายผลไปสู่คู่ภาษาหรือโดเมนอื่นได้ง่ายกว่าการแปลด้วยกฎ

การสร้างตัวอย่างการแปลที่ละประโยคด้วยมือต้องใช้ค่าใช้จ่ายสูง ดังนั้นจึงมีงานวิจัยที่พัฒนาเทคนิคการจับคู่การแปลระดับประโยคจากเอกสารขนานโดยอัตโนมัติ เทคนิคการจับคู่การแปลระดับประโยคในงานที่มีมาแล้วมักคำนวณค่าใช้จ่ายการจับคู่การแปลจากคู่คำแปลหรือความน่าจะเป็นของคู่คำแปลที่สกัดจากคลังประโยคคู่ขนานตั้งต้น จากทรัพยากรทางภาษาที่มีอยู่แล้ว เช่น พจนานุกรม เวิร์ดเน็ต เป็นต้น หรือจากลักษณะเด่นร่วมกันของคู่ภาษา เช่น เครื่องหมายวรรคตอน เป็นต้น จากงานวิจัยที่มีมาแล้วพบว่า การจับคู่การแปลระดับประโยคมักได้ความถูกต้องต่ำสำหรับเอกสารที่แปลโดยมุ่งเน้นการสื่อความหมายและไม่มุ่งเน้นความครบถ้วนของเอกสารภาษาต้นทางหรือแปลโดยอรรถ ซึ่งทำให้เกิดการจับคู่การแปลแบบแทรกและแบบลบปริมาณมาก นอกจากนี้คู่เอกสารภาษาไทย – อังกฤษมีลักษณะเด่นร่วมกันของคู่ภาษาและทรัพยากรทางภาษาน้อยเมื่อเทียบกับคู่ภาษาอื่นจึงไม่สามารถประยุกต์ใช้วิธีการที่มีมาแล้วได้โดยตรง ดังนั้นวิทยานิพนธ์นี้จึงมีวัตถุประสงค์เพื่อศึกษาและพัฒนาเทคนิคสำหรับการจับคู่การแปลระดับประโยคภาษาไทย – อังกฤษ วิธีที่เสนอเป็นการใช้ฟังก์ชันค่าใช้จ่ายแบบถ่วงน้ำหนักซึ่งคำนวณจากข้อสนเทศ 3 กลุ่ม ได้แก่ รูปแบบการจับคู่ ความยาวประโยค และความน่าจะเป็นการแปล โดยตัวแปรและค่าถ่วงน้ำหนักของแต่ละฟังก์ชันค่าใช้จ่ายจะคำนวณจากคลังประโยคคู่ขนานตั้งต้น

งานวิจัยนี้ทำการทดลองไขว้แบบ 5 ทบเพื่อวัดประสิทธิภาพของโปรแกรมจับคู่คำแปลระดับประโยคที่พัฒนาขึ้น กับเอกสารที่แปลโดยอรรถเกี่ยวกับการท่องเที่ยวจำนวน 115 ย่อหน้าซึ่งประกอบด้วย 941 คู่ประโยค ผลการทดลองพบว่ามีความแม่นยำ ค่าความครอบคลุม และค่าคะแนนเอฟ ที่ดีที่สุดเป็น 0.728, 0.752 และ 0.740 ตามลำดับ นอกจากนี้ยังทดลองกับเอกสารที่แปลแบบทุกข้อความหรือที่เรียกว่าแปลโดยพยัญชนะในโดเมนกฎหมายจำนวน 1,632 คู่ประโยค พบว่าผลความแม่นยำ ความครอบคลุม และค่าคะแนนเอฟมีค่าเป็น 0.998