

วิทยานิพนธ์ฉบับนี้ ได้นำเสนอวิธีการสร้างเครือข่ายคำไทยของมโนทัศน์พื้นฐานร่วมของ เอนทิตีลำดับที่หนึ่งด้วยมือ โดยใช้พจนานุกรมอังกฤษ-ไทย 5 ชุด ได้คำแปลภาษาไทยทั้งสิ้น 14,804 คำ แต่คำแปลที่ได้ต้องผ่านขั้นตอนการคัดเลือก 4 ขั้นตอน คือ (1) เก็บคำแปลที่เหมาะสมกับมโนทัศน์เอาไว้ (2) คัดคำแปลภาษาไทยที่ไม่ตรงกับมโนทัศน์ออก (3) ค้นคว้าคำจากแหล่งวิชาการอื่น หรือผู้รู้ หรือวลีเพื่อแทนมโนทัศน์ที่ไม่มีคำในภาษาไทย และ (4) ตรวจสอบคำภาษาไทยกับผู้เชี่ยวชาญ จนได้เป็นเครือข่ายคำไทยของมโนทัศน์พื้นฐานร่วมของเอนทิตีลำดับที่หนึ่งที่เป็นตัวมาตรฐานเทียบ ซึ่งจากกลุ่มคำไวยากรณ์ของเอนทิตีลำดับที่หนึ่ง จำนวน 493 กลุ่ม ผู้วิจัยได้คำภาษาไทยที่เหมาะสมทั้งสิ้น 1,310 คำ จากนั้นจึงพัฒนาระบบอัตโนมัติที่ใช้ในการสร้างเครือข่ายไทยด้วยวิธีการแปลสองทาง โดยใช้ข้อมูลพจนานุกรมที่มีวิธีการสร้างแตกต่างกัน 2 ชุด คือ พจนานุกรมอิเล็กทรอนิกส์ อังกฤษ->ไทย เล็กชิตรอน และพจนานุกรม อังกฤษ->ไทย ของ วิทย์ เทียงบูรณะธรรม ซึ่งเป็นพจนานุกรมที่สร้างด้วยวิธีที่ต่างกัน พจนานุกรมเล็กชิตรอนสร้างโดยอาศัยข้อมูลจริงที่รวบรวมมา ส่วนพจนานุกรมของวิทย์ เทียงบูรณะธรรม สร้างโดยผู้เชี่ยวชาญเป็นผู้ทำพจนานุกรมเอง ผลการทดลองพบว่า ระบบอัตโนมัติที่ใช้ในการสร้างเครือข่ายคำไทย ด้วยวิธีการแปลสองทาง ที่ใช้พจนานุกรมอิเล็กทรอนิกส์เล็กชิตรอน มีค่าเอฟ-เมเชอร์ในมิติกลุ่มคำไวยากรณ์เท่ากับ 50.36 และมิติคำศัพท์เท่ากับ 25.01 ส่วนระบบที่ใช้พจนานุกรม ของ วิทย์ เทียงบูรณะธรรม มีค่าเอฟ-เมเชอร์ในมิติกลุ่มคำไวยากรณ์เท่ากับ 64.51 และมิติคำศัพท์เท่ากับ 34.54 และเมื่อรวมพจนานุกรมสองเล่มเข้าด้วยกัน ค่าเอฟ-เมเชอร์ในมิติกลุ่มคำไวยากรณ์เพิ่มเป็น 67.16 และมิติคำศัพท์เพิ่มเป็น 36.27

This research presents a method of the construction of Thai WordNet of 1st Order Entity Common Base Concepts, containing 493 synsets. By using five English-Thai dictionaries, 14,804 Thai words are obtained; however, irrelevant Thai words have to be filtered out of each synset. The steps are as follow: (1) retain words that their meanings fit to the definition of the synset, (2) delete words that their meanings are irrelevant to the definition, (3) find relevant Thai words from other academic resources or create a new Thai word or phrase if the concept is not yet lexicalised in Thai, and (4) verify new terms by experts in that field. As a result, 1,310 Thai words are created as a Thai WordNet. This Thai WordNet can be used as a gold standard test set for evaluating any automatic or semi-automatic approaches of Thai WordNet construction. After that, an automatic bi-directional translation system has been developed and tested with two English<->Thai dictionaries, LEXITRON and Dr. Wit Thiengburanathum Dictionary. The former was compiled using a corpus-based approach, while the latter was compiled on the basis of the author's expertise. The results show that F-measure of the system using LEXITRON rates at 50.36 in synset aspect, 25.01 in word aspect; furthermore, F-measure of the system using Dr. Wit Thiengburanathum Dictionary rates at 64.51 in synset aspect, 34.54 in word aspect. When combining two dictionaries, F-measure increases to 67.16 in synset aspect, and to 36.27 in word aspect.