

คู่มือการกำกับข้อมูล
Thai Dependency Tree Bank
ตามแนว Universal Dependencies v.2

จัดทำโดย

วิโรจน์ อรุณมานะกุล
ธารทอง แจ่มไพบูลย์

ศูนย์วิจัยการประมวลผลภาษาและวัจนะ
ภาควิชาภาษาศาสตร์
คณะอักษรศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย
๒๕๖๖

คำนำ

คู่มือเล่มนี้เป็นข้อสรุปที่ได้จากการทำงานร่วมกันกับผู้ร่วมงานในการกำกับข้อมูลภาษาไทยจำนวนหนึ่งจากคลังข้อมูลภาษาไทยแห่งชาติ (ประมาณ 148,000 คำ) โดยตั้งความมุ่งหวังเพื่อสร้างคลังข้อมูลที่เป็น Tree bank เผยแพร่แก่สาธารณะให้มีข้อมูลร่วมกันในการวิเคราะห์งานที่ต้องอาศัยข้อมูลทางวากยสัมพันธ์ภาษาไทย การกำกับข้อมูลอาศัยกรอบการวิเคราะห์ของ Universal Dependencies¹ ซึ่งเป็นกรอบการวิเคราะห์มาตรฐานสากลที่ต้องการให้ข้อมูลที่วิเคราะห์จากหลากหลายภาษามีข้อมูลวากยสัมพันธ์ที่ใช้และเข้าใจร่วมกัน คือ หมวดคำ หรือ Part of Speech และความสัมพันธ์เชิงพึ่งพา หรือ dependency relation ที่เชื่อมโยงระหว่างคำที่เป็นคำหลักและคำพึ่งพา

การกำกับข้อมูลทางภาษาศาสตร์นี้เป็นงานที่ต้องอาศัยความละเอียดและการวิเคราะห์ที่สม่ำเสมอ ต้องอาศัยการประชุมหารือและนำข้อสงสัยในการวิเคราะห์ข้อมูล หลาย ๆ ครั้งที่มีจะต้องกลับมาแก้ไขการวิเคราะห์และการกำกับข้อมูลที่ได้ทำไปแล้ว เพราะความยากของการกำกับข้อมูลทางวากยสัมพันธ์ในภาษาไทยอยู่ที่การไม่มีรูปแบบที่ประเภของคำและความสัมพันธ์ทางวากยสัมพันธ์ จึงทำให้สามารถคิดวิเคราะห์ได้หลายแนวทาง ข้อมูลที่จะนำไปใช้ประโยชน์ได้จริงจึงต้องเป็นข้อมูลที่วิเคราะห์ได้สม่ำเสมอ ไม่เปลี่ยนแปลงในแต่ละตัวอย่าง หรือแปรไปตามผู้วิเคราะห์และกำกับข้อมูล ทางโครงการคลังข้อมูลภาษาไทยแห่งชาติ จึงได้จัดทำคู่มือนี้เพื่อเป็นแนวทางสำหรับผู้สนใจงานกำกับและวิเคราะห์ข้อมูลภาษาไทยได้ใช้เป็นแนวทางการทำงานได้ทำงานกำกับข้อมูลออกมาได้สม่ำเสมอและคงที่มากที่สุด

งานนี้สำเร็จเป็นรูปร่างเบื้องต้นได้โดยความร่วมมือร่วมใจของนิสิตบัณฑิตศึกษาคณะศึกษาศาสตร์ที่ได้ช่วยทำข้อมูลและนำปัญหาที่พบมาพูดคุยเพื่อแลกเปลี่ยนความเห็นและตัดสินใจเลือกการวิเคราะห์ที่เหมาะสมกับภาษาไทย จึงต้องขอขอบคุณนิสิตต่อไปนี้ ซึ่งต่างก็ได้สำเร็จการศึกษาจากภาควิชาไปแล้ว

ธารทอง แจ่มไพบูลย์

นัชชา ธีระสารโรช

ศศิวิมล กาลันสีมา

อิสราภ ล้อรัตน์ไชยวงศ์

ศุภวิชญ์ แต่รุ่งเรือง

ภัทธิดา โสตาบัน

และก็ต้องขอขอบคุณนางสาวชลธร ขวัญขจรเกียรติ นิสิตบัณฑิตศึกษาคณะศึกษาศาสตร์คอมพิวเตอร์ที่ติดตั้งระบบ Brat ที่ใช้ในการกำกับข้อมูลหมวดคำและความสัมพันธ์พึ่งพาที่ใช้ในโครงการนี้ ณ ขณะนี้ ข้อมูลที่มีการกำกับหมวดคำและความสัมพันธ์ที่บอกคำหลักคำพึ่งพาเพื่อสร้าง dependency tree ได้ทำเสร็จรอบแรกแล้ว กำลังอยู่ระหว่างการตรวจสอบความสม่ำเสมอของการวิเคราะห์ และยังอยู่ในขั้นตอนการวิเคราะห์ชนิดของความสัมพันธ์ในต้นไม้นี้แต่ละต้น และกำกับชนิดของความสัมพันธระหว่างคำต่าง ๆ โดยมีนิสิตภาควิชาภาษาไทย คือ นางสาวนภัสสร เกษศรี เป็นผู้กำกับความสัมพันธ์และตรวจแก้ไขการกำกับหมวดคำที่ตกหล่นหรือไม่สม่ำเสมอ การกำกับข้อมูลทั้งหมด

¹ <http://universaldependencies.org/>

ขณะนี้ ยังไม่แล้วสิ้น โดยโครงการนี้ได้เริ่มทำงานกำกับข้อมูลนี้มาตั้งแต่ปี พ.ศ.๒๕๖๐ หากเสร็จสิ้นจะเผยแพร่เป็นข้อมูลสาธารณะสำหรับนำไปใช้งานได้

วิโรจน์ อรุณมานะกุล

๑๔ ตุลาคม ๒๕๖๓

หมวดคำภาษาไทย

หมวดคำเป็นข้อมูลพื้นฐานทางวากยสัมพันธ์ ในตำราหลักภาษาไทยของพระยาอุปกิตศิลปสารได้กำหนดหมวดคำไว้ 7 ประเภท ได้แก่ คำนาม คำสรรพนาม คำกริยา คำวิเศษณ์ คำบุพบท คำสันธาน และคำอุทาน ซึ่งเป็นการจำแนกด้วยเกณฑ์หลายเกณฑ์ เช่น จำแนกตามเกณฑ์ทางความหมาย ให้นามเป็นคำที่ใช้อ้างถึงคน สัตว์ สิ่งของ คำกริยาเป็นคำแสดงอาการ เป็นต้น จำแนกตามเกณฑ์หน้าที่ ให้คำวิเศษณ์เป็นคำใช้ประกอบคำอื่น จำแนกตามเกณฑ์ตำแหน่ง ให้บุพบทเป็นคำนำหน้านาม เป็นต้น แต่การจัดหมวดคำแบบนี้มีที่มาจากไวยากรณ์ดั้งเดิมที่มีต้นแบบมาจากไวยากรณ์กรีกและไวยากรณ์ลาติน หมวดคำลักษณะนี้จึงอาจจะไม่ได้สะท้อนธรรมชาติที่แท้จริงของภาษาไทย นักภาษาศาสตร์จำนวนหนึ่งจึงพยายามจำแนกหมวดคำตามแนวคิดทฤษฎีไวยากรณ์สมัยใหม่ เช่น การจำแนกด้วยเกณฑ์หน้าที่ (นววรรณ พันธุเมธา 2527) การจำแนกตามหลักไวยากรณ์โครงสร้าง (วิจิตร ภาณุพงศ์ 2532) การจำแนกตามเกณฑ์วากยสัมพันธ์ตามหลักทฤษฎีไวยากรณ์ศัพทการก (อมรา ประสิทธิ์รัฐสินธุ์ 2543) บ้างก็ใช้เกณฑ์หลากหลายผสมกันโดยแจกแจงรายละเอียดมากขึ้นเพื่องานประมวลผลภาษาไทย เช่น หมวดคำในคลังข้อมูลอรรถาภิธาน (Virach Sornlertlamvanich et al. 1997) การจำแนกหมวดคำที่ผ่านมาจึงมีความแตกต่างกันไปตามลักษณะเกณฑ์และแนวคิดที่ใช้

เนื่องจากหมวดคำเป็นหน่วยพื้นฐานทางวากยสัมพันธ์ แม้การนิยามหมวดคำบางครั้งจะดูเหมือนว่าเป็นเรื่องของความหมาย แต่โดยพฤติกรรมแล้ว คำในหมวดคำเดียวกันก็มีลักษณะการเกิดคล้าย ๆ กัน เช่น คำที่เป็นกริยาก็มักปรากฏหลังคำนามที่ทำหน้าที่เป็นประธาน บางกลุ่มก็อาจมีนามตามหลังที่ทำหน้าที่เป็นกรรมได้ เป็นต้น เกณฑ์ทางวากยสัมพันธ์หรือการปรากฏร่วมกับคำอื่น ๆ จึงควรเป็นหลักสำคัญในการจำแนกหมวดคำ หมวดคำจึงเป็นลักษณะเฉพาะของแต่ละภาษา หมวดคำภาษาไทยจึงควรจำแนกโดยอาศัยข้อมูลภาษาไทยเป็นสำคัญ ไม่สามารถนำชุดหมวดคำที่ใช้ในภาษาอื่นมาใช้ได้โดยตรงทั้งหมด

อย่างไรก็ดี ในการทำงานประมวลผลภาษาต่าง ๆ หากแต่ละภาษามีการกำหนดใช้ชุดหมวดคำในลักษณะต่าง ๆ ที่แตกต่างกันโดยสิ้นเชิง จะทำให้การทำงานร่วมกันหลาย ๆ ภาษาเป็นไปได้ด้วยความลำบากมากขึ้น จึงเกิดแนวคิดที่จะกำหนดชุดหมวดคำที่สามารถใช้ได้กับทุก ๆ ภาษา จะมีเฉพาะบางส่วนที่ยอมให้เป็นเรื่องเฉพาะภาษาได้ แต่กรอบการวิเคราะห์พื้นฐานแล้วจะใช้เหมือนกัน กลุ่มคนที่มีแนวคิดลักษณะนี้ จึงได้เสนอโครงการ universal dependencies (UD) ซึ่งเป็นโครงการสากลสำหรับกำกับหมวดคำและกำกับความสัมพันธ์ระหว่างคำผ่านการวิเคราะห์แบบฟิงพา โดยได้กำหนด POS และกำหนดความสัมพันธ์แบบฟิงพา (dependency relation) ที่เกิดขึ้นระหว่างคำในประโยคที่เป็นสากลขึ้นมา POS สำหรับใช้แท็กข้อมูลภาษาไทยในที่นี้ จึงยึดตามรายการที่นำเสนอใน <http://universaldependencies.org/> ข้อควรสังเกตคือ จำนวน POS ที่เสนอให้ใช้นั้นมีไม่มาก และไม่แนะนำให้ตั้งแท็กที่เป็นเรื่องเฉพาะแต่ละภาษา จำนวนแท็กจึงมีน้อยกว่าที่ใช้กันในงาน POS Tagging ทั่วไป อย่างไรก็ตาม จำนวน POS ที่น้อยนั้นถูกชดเชยด้วยการกำหนดความสัมพันธ์แบบฟิงพาที่ค่อนข้างมากแทน ทำให้ยังคงสามารถแยกความต่างทางวากยสัมพันธ์ออกได้ เช่น คำนามที่เคยแยกเป็นประเภทย่อยต่าง ๆ ในกรอบ UD มีเพียง NOUN, PROPN คำกลุ่มที่พบในภาษาไทย จีน อย่างคำ

ลักษณะนามถูกจัดเป็น NOUN ซึ่งต้องไปแยกด้วยวิธีการอื่นต่อไป ในภาษาจีนมีผู้ใช้ nmod:clf เป็นตัวกำหนดความสัมพันธ์นี้แทน (Leung et al. 2016) ซึ่งต่อมาใน UD version 2.0 กำหนดให้ใช้ relation “clf” เป็นการบอกความสัมพันธ์ระหว่าง head noun กับ classifier ในการกำกับข้อมูลภาษาไทยในที่นี้จึงจะใช้ NounType:class ที่มีเสนอในบางภาษามาใช้ (ให้ class เป็น value ของ attribute “NounType” ภายใต้แท็ก POS: NOUN) และใช้ relation “clf” ประกอบด้วยเมื่อเชื่อมโยงกับคำนามหลัก เป็นต้น

จำนวนแท็กและรายการ POS ต่าง ๆ ใน UD พัฒนาต่อมาจากหลายโครงการ ได้แก่ Stanford dependencies (de Marneffe et al., 2006, 2008, 2014), Google universal part-of-speech tags (Petrov et al., 2012), และ Intersect interlingua for morphosyntactic tagsets (Zeman, 2008) เพื่อให้ได้ชุดรายการที่เป็นมาตรฐานเดียวกันสำหรับใช้ได้กับภาษาต่าง ๆ หลักการนอกจากจะให้ได้กับหลากหลายภาษา ก็ยังต้องมีความถูกต้องทางภาษาศาสตร์ และไม่ยากเกินไปที่จะให้คนทั่วไปสามารถใช้งานได้ด้วย เวอร์ชันแรกของ UD ออกมาในเดือนตุลาคม ปี 2014 UD ปัจจุบัน เป็น UD เวอร์ชัน 2

สำหรับภาษาไทย จำนวน POS tag และ attribute เบื้องต้นที่จะใช้กับภาษาไทย ทางโครงการคลังข้อมูลภาษาไทยแห่งชาติ ได้กำหนดวิธีการกำกับข้อมูล ดังนี้

การกำหนดหมวดคำภาษาไทยใน UD

การกำหนดหมวดคำ นอกจาก POS แล้ว เรายังสามารถระบุ attribute : value เพื่อใช้จำแนกประเภทย่อยของ POS นั้นอีกได้ UD POS tags และ feature ต่าง ๆ ที่ใช้ในที่นี่ใช้โดยเทียบเคียงจากคำอธิบายที่เขียนไว้ในเว็บ <http://universaldependencies.org/> แต่ก็มีคำบางกลุ่มที่อาจต้องปรับเปลี่ยนอีกเพื่อความเหมาะสม

กลุ่มคำเปิด		
NOUN		คำนามทั่วไปรวมถึงคำนำหน้าชื่อทั้งหลาย นาย นาง พล.ต. เหล่านี้ก็ให้กำหนดเป็น NOUN ด้วย แล้วโยงความสัมพันธ์ compound ระหว่าง คำนำหน้าชื่อ กับ ชื่อ (นาย)-(สมชาย) คำในกลุ่มชื่อสัตว์ (นก/พาราบ นก/กระยาง ปลาทุ ปลาสรวย) หรือชื่อพืช (ผัก/บุ้ง ต้น/ตำลึง ต้น/มะม่วง) เหล่านี้ให้กำกับเป็น NOUN
	NounType: class	Feature สำหรับใส่ให้คำที่เป็นลักษณนาม กำหนดไว้ก่อน แต่อาจจะลบทิ้งได้ เพราะสามารถใช้ relation clf บอกความเป็น classifier ได้
PROPN		คำนามเฉพาะ ได้แก่ ชื่อคน สถานที่ องค์กร ผลิตภัณฑ์ ประเทศ ส่วนชื่ออาหาร ชื่อปลา สัตว์ ต้นไม้ จะให้เป็น NOUN
	NameType: Prs NameType: Giv NameType: Sur NameType: Geo NameType: Com NameType: Nat NameType: Pro NameType: Oth	Prs: person name (ใช้กรณีไม่รู้ว่าเป็น given หรือ surname) Giv: given name ชื่อต้น เช่น ประยุทธ์ อัครวิน เป็นต้น Sur: surname ชื่อสกุล เช่น จันทร์โอชา ขวัญเมือง เป็นต้น Geo: geographical name เช่น ชื่อสถานที่ ชื่อประเทศ ชื่อจังหวัด ชื่ออำเภอ ชื่อมหาวิทยาลัย (ในกรณีที่ในบริบทนั้นต้องการบ่งบอกสถานที่ตั้ง เช่น ฉันขับรถผ่านจุฬาฯ เป็นต้น) Com: company organization name ชื่อบริษัท ชื่อประเทศ ชื่อมหาวิทยาลัย (ในกรณีที่บริบทนั้นบ่งถึงลักษณะขององค์กร เช่น จุฬาฯ ส่งนิสิตเข้าแข่งขันในรายการต่าง ๆ เป็นต้น) Nat: nationality คำแสดงเชื้อชาติ สัญชาติ Pro: product เช่น ไอโฟน Oth: others ชื่อเฉพาะอื่น ๆ ที่ไม่สามารถจัดเข้าประเภทต่าง ๆ ข้างต้นได้
VERB		คำกริยาทั้งกรรมและกรรม
	VerbType: Cop	Featureที่กำหนดให้กับ กริยา copular ได้แก่คำว่า 'คือ' (ไม่ใช่ทุกกรณีที่เป็น copular) จะกำกับให้ VERB นั้นเป็น Cop ก็ต่อเมื่อฝั่งซ้ายและฝั่งขวาสื่อถึงสิ่งเดียวกันเท่านั้น เช่น เขาคือเพื่อนแท้ → เขา กับ เพื่อนแท้ เป็นสิ่งเดียวกัน เช่นนี้จะกำกับ คือ/VERB:{VerbType: Cop} ข้อสังเกตคือฝั่งซ้ายและขวามักจะมีสถานะเป็น NOUN PRON หรือ PROPN คือ ที่ทำหน้าที่เชื่อมความในประโยค จะวิเคราะห์เป็น CONJ เช่น การสื่อสารที่มีประสิทธิภาพที่สุดคือ การที่ผู้ส่งสารและผู้รับสารมีลักษณะเหมือนกัน
INTJ		คำอุทานทั้งหลาย
ADJ		คำคุณศัพท์ เป็นคำ ขยายหลังคำนาม รวมถึงคำที่บ้างจัดให้เป็นคำกริยาคุณศัพท์ จะต้องเป็นคำที่แสดงคุณลักษณะบางอย่างของคำนามที่อยู่

		ข้างหน้า เช่น ดี สวย อ้วน ผอม ใหญ่ เล็ก เป็นต้น ในกรณีที่คำซึ่งปรากฏหลังคำนามนั้นไม่ได้แสดงลักษณะของคำนาม แต่บ่งความชี้เฉพาะหรือไม่ชี้เฉพาะ จะกำกับเป็น DET
ADV		คำวิเศษณ์ ขยายกริยา VERB หรือขยายคำ ADJ หรือ ADV ปกติปรากฏหลังคำกริยา แต่มีบ้างที่ปรากฏหน้ากริยา เช่น คำว่า 'สุด' ใน สุดขยัน มีคำ ADV จำพวกหนึ่งที่ไม่ได้ขยายกริยา แต่ขยาย numeral, determiner, noun, pronoun ปรากฏหน้านาม เช่น เพียง อีก ทั้ง สัก ทุก แต่ละ หลากหลาย (หลากหลายรูปแบบ) เป็นต้น ให้กำกับเป็น ADV ลักษณะนี้พบในตัวอย่างการวิเคราะห์ภาษาอื่นด้วย เช่น ภาษาจีน
กลุ่มคำปิด		
AUX		คำกริยาช่วยทั้งที่ปรากฏหน้าและหลังกริยาในภาษาไทย เช่น ควร จะ นำ ได้ ไป มา เป็นคำที่บ่งบอก tense, aspect, modality ในกรณีที่ปรากฏหน้ากริยา จะกำกับเป็น AUX แต่หากปรากฏหลังกริยาจะกำกับเป็น ADV ได้
ADP	AdpType: Prep	Adposition ในภาษาไทยเป็น preposition อย่างเดียว ไม่มี postposition ปรากฏหน้าคำนาม เช่น บน ใน หน้า ทาง เพื่อ feature นี้จึงเป็น default สำหรับ ADP ในภาษาไทย นำหน้า นามวลี นาม สรรพนาม หรืออนุพากย์ ทำหน้าที่นาม
PRON	PronType: Prs	คำสรรพนามบุคคล เช่น ฉัน เขา เธอ อะไร (เขาทำอะไร เธอก็ชอบทั้งนั้น) นี่ (นี่คือสิ่งที่เขาต้องการ) ใครต่อใคร ล้วนแต่ ต่าง (ผู้คนต่างยกย่องเขา)
	PronType: Rcp	คำสรรพนาม reciprocal → กัน ซึ่งกันและกัน
NUM		คำบอกจำนวน ปริมาณ ทำหน้าที่เหมือน DET, ADJ, PRON บอกความหมายเกี่ยวกับจำนวน เป็น ตัวเลข คำจำนวนนับ เช่น 123 สิบ/เอ็ด ตัว เลขที่ประกอบคำ เช่น 7 ใน Windows 7 ก็ให้กำกับ NUM
DET		คำบ่งชี้เฉพาะหรือไม่ชี้เฉพาะ ใช้ขยายหลังคำนาม เช่น นี่ นั้น ไหน อื่น ทั้งหมด ต่างๆ ทั่วไป (กรณีที่มีการละคำนามและเหลือเพียงคำประเภที่คำเดียว เช่น 'ทั้งหมด' ก็คงให้เป็น DET)
CCONJ		Coordinate conjunction คำเชื่อมสองส่วนที่มีสถานะเท่ากัน เช่น และ (หมา และแมว) แต่ (ฉันอยากกินแต่เธออยากนอน) หรือ (ข้าว กว๊ายเดี่ยว หรือสแต็ก) เป็นต้น รวมถึงคำที่ทำหน้าที่เป็น discourse marker ที่เชื่อมความ เช่น อย่างไรก็ตาม แต่ถึงกระนั้น
SCONJ		Subordinate conjunction คำเชื่อมส่วนที่ทำหน้าที่เป็นส่วนขยายหรือ subordinate clause เช่น ที่ว่า ให้ จึง เพื่อ รวมถึงคำว่า เช่น คือ ที่ทำหน้าที่เชื่อมความในประโยค
PART		คำอนุภาคทั่วไป กรณีคำอนุภาคเฉพาะให้เติม feature ข้างล่างนี้
	PartType: Emp	emphasis ใช้กำกับหลังวลีหรืออนุประโยค เพื่อเป็นการเน้นเหตุการณ์ ดังกล่าวได้อีกด้วย เช่น 'นั่น' ใน ขณะที่กำลังผสมส่วนผสมต่างๆเข้าด้วยกันนั้น 'เสียอีก' 'ซะอีก' เป็นต้น
	PartType: Res	Response ครับ ค่ะ นะ จ๊ะ
	PartType: Int	Interrogative หรือไม่ หรือ ไหม ทำไม
	PartType: Neg	Negation ไม่ มี บ่ ใช่ว่า ไม่ใช่ว่า

	PartType: Adj PartType: Adv	Adjective นำ เป็นคำที่เป็น Adjective marker เมื่อปรากฏหน้าคำใด ทำให้คำนั้นทำหน้าที่เป็น adjective ได้ เช่น น่ารัก น่าเอ็นดู Adverb อย่าง โดย เป็นคำที่เป็น adverb marker เมื่อปรากฏร่วมกับคำที่ตามมา จะทำให้คำนั้นทำหน้าที่เป็น Adv ได้ เช่น อย่างเร็ว โดยด่วน ด้วยดี (Particle สองประเภทหลังกำหนดใช้ในภาษาไทยเป็นการเฉพาะ)
PUNCT		เครื่องหมายวรรคตอนต่าง ๆ ที่เป็นส่วนหนึ่งของข้อความมีหน้าที่ในทางภาษา เช่น ๆ ฯลฯ , ?
SYM		เครื่องหมายสัญลักษณ์ต่าง ๆ ที่ไม่ใช่ PUNCT เช่น สัญลักษณ์ทางคณิตศาสตร์ + - * # @ รวมถึง email, url dummy@gmail.com www.dummy.com เป็นต้น
X		คำที่ยังไม่สามารถจัดประเภทได้ตามข้างบนนี้ ให้กำกับเป็น X ไปพลางก่อน
Feature		เป็น feature ที่ใช้กำกับคำใด ๆ เป็นการกำกับลักษณะเพิ่มเติมของคำนั้น ๆ ซึ่งถูกกำกับด้วย POS บางอย่างไว้แล้ว จากนั้นจึงพิจารณาต่อไปว่าคำดังกล่าวมีลักษณะใด ๆ เพิ่มเติมหรือไม่ ลักษณะดังกล่าวนี้ ได้แก่ คำนำหน้า (Prefix) คำย่อ (Abbr) และคำต่างประเทศ (Foreign)
	Prefix: Yes Abbr: Yes Foreign: Yes	Prefix ประกอบคำ เช่น ชาว นัก ผู้ การ ความ นำ อย่าง คำย่อ (POS ของคำย่อเป็นไปตามรูปเต็มคำนั้น เช่น กรุงเทพฯ/PROPN {NameType:geo, Abbr:Yes}/ฯ/PUNCT รร./NOUN {Abbr:Yes}) คำทับศัพท์ต่างประเทศ

ข้อสังเกตเพิ่มเติมในการกำกับ POS:

1. การกำกับคำร่วมกับเครื่องหมายไม้มก (๑)

คำบางคำปรากฏร่วมกับเครื่องหมายไม้มก (๑) โดยปกติจะกำกับแยกเครื่องหมาย ๆ ออกเป็น PUNCT อย่างไรก็ดี ในบางกรณีคำดังกล่าวจำเป็นต้องปรากฏซ้ำ หรือต้องปรากฏร่วมกับ ๆ เท่านั้น จึงจะสามารถสื่อความหมายที่ถูกต้อง เช่น

- บนโต๊ะเต็มไปด้วยเอกสารต่าง ๆ
- หลัง ๆ เขามักจะมาตามนัด

จะเห็นว่า ต่าง ๆ หรือ หลัง ๆ จำเป็นต้องปรากฏด้วยกัน หากปรากฏเพียงคำเดียวโดด ๆ จะสื่อความหมายที่ต่างออกไป ดังนั้น การกำกับคำว่า ต่าง ๆ กับ หลัง ๆ จำเป็นต้องกำกับรวมเป็น 1 คำ ต่าง ๆ/DET หลัง ๆ/CCONJ เป็นต้น

2. การกำกับนามวลีในกลุ่ม การ- ความ-

นามวลีในกลุ่ม การ- ความ- โดยปกติจะกำกับแยก การ- และ ความ- ออกมาเป็นหน่วย NOUN:{Prefix: Yes} แต่บางกรณีจำเป็นต้องรวม การ- หรือ ความ- ไว้เป็นส่วนหนึ่งของคำนั้นด้วย เช่น ความทรงจำ การบ้าน การเมือง การงาน เป็นต้น ขอให้พิจารณาจากความหมายของคำนั้น ๆ เป็นหลัก การวิเคราะห์ความสัมพันธ์ระหว่าง การ และ หน่วยกริยาวลีที่ตามมา สามารถวิเคราะห์ได้สองแบบ

แบบแรก ให้ การ เป็นคำหลักโดยมี POS เป็น NOUN ที่มี feature {prefix} และมีคำกริยาเป็นคำพึ่งพา เช่น การสูญเสียบุคคลใกล้ชิด การ >สูญเสีย เพราะมองในเชิงวากยสัมพันธ์ว่า ทั้งหมดทำหน้าที่เป็น นามวลี ไปเชื่อมโยงกับคำอื่นต่อ การ จึงควรเป็นคำหลักและมี POS: NOUN และหากเติม บุพบทลีแสดงความเป็นเจ้าของ เช่น การสูญเสียบุคคลใกล้ชิดของเขา ของเขา จะสัมพันธ์กับ การสูญเสีย เหมือนอย่าง หนังสือของเขา หากวิเคราะห์แบบนี้ความสัมพันธ์ระหว่าง การ และ กริยาลีข้างหลังควรเป็น acl คือ clausal modifier of noun

แบบที่สอง วิเคราะห์ให้กริยาเป็นหลักเพราะมองว่าเป็นคำเนื้อหา การ มีความเป็นคำไวยากรณ์จึงควรเป็นหน่วยพึ่งพาตามหลักที่ให้คำไวยากรณ์ไม่สามารถเป็นหน่วยหลัก การ <สูญเสีย โดยมีความสัมพันธ์เป็น mark และการสูญเสียบุคคลใกล้ชิดของเขา ก็จะโยงความสัมพันธ์ระหว่างคำหลัก สูญเสีย กับ >(ของ<เขา) ในเชิงความหมาย การวิเคราะห์แบบนี้ดูเหมือนจะทำให้คำเนื้อหาเชื่อมโยงกันได้ดีกว่า [สูญเสีย – บุคคล, สูญเสีย - เขา] แต่วิธีนี้ทำให้มีปัญหายุ่งยากในการแยกว่า เขา มีบทบาททางความหมายอะไรผ่าน relation อะไร เพราะ (ของเขา) ปกติเป็น nmod ของคำนามข้างหน้า กรณีนี้จะยังคงให้เป็น nmod ของ สูญเสีย ได้ไหม ถ้ายอมได้ ก็จะกลายเป็นมีกรณี nmod ขยาย VERB ได้ด้วย และจะเป็นปัญหาต่อไปในกระบวนการ parsing หรือถ้าแก้โดยให้ สูญเสีย เป็น NOUN เพราะมีการ อยู่ข้างหน้า ก็จะทำให้ต้องมีอีกกระบวนการหนึ่งเปลี่ยน POS ของคำได้

ในที่นี้จึงเลือกการวิเคราะห์แบบแรกให้ การ เป็น NOUN{prefix} ที่มีกริยาลีตามหลังด้วยความสัมพันธ์ acl

3. POS ของคำอาจขึ้นกับตำแหน่งในการปรากฏ

คำหนึ่ง ๆ อาจสามารถกำกับได้หลาย POS ขึ้นกับว่าคำนั้น ๆ ปรากฏในตำแหน่งใดของข้อความ ยกตัวอย่างคำว่า หลากหลาย จากข้อความต่อไปนี้

- อาชีพหลากหลาย หลากหลาย กำกับเป็น ADJ เป็น dependent ของ อาชีพ (ปรากฏหลังนาม)



- อาชีพที่หลากหลาย หลากหลาย กำกับเป็น VERB เป็น head ของ ที่



- หลากหลายอาชีพ หลากหลาย กำกับเป็น ADV เป็น dependent ของ อาชีพ (ปรากฏหน้านาม)



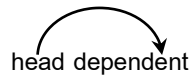
UD เปลี่ยนวิธีการวิเคราะห์ ellipsis จากเดิมที่ใช้ relation remnant มาเป็นการโปรโมทหน่วยรองมาเป็น head แทน เช่น ถ้าละ head nominal จะโปรโมท dependents ตามลำดับความสัมพันธ์นี้มาแทน amod > nummod > det > nmod > case ถ้าละกริยา ให้โปรโมท aux หรือ cop มาแทน ถ้าไม่มีก็มีสองแนวทาง ทางแรกคือให้โยงตัว orphan กับ grandparent และใช้ composite relation คือมี relation ของ headrel ต่อด้วย orphanrel เช่น I like tea and you coffee จะได้ like-(conj>nsubj)>you like-(conj>obj)>coffee ทางที่สองคือใช้ relation “orphan” ในตัวอย่างนี้จะเป็น you-(orphan)>coffee ในที่นี้จะเลือกใช้วิธีที่สอง relation orphan

4. การวิเคราะห์คำกริยาเรียง

- ประโยคที่มีกริยามากกว่าหนึ่งคำมาประกอบกัน ในแต่ละกรณีจะวิเคราะห์ต่างกัน กรณีที่เป็น กริยาเรียง หรือ serial verb construction จะวิเคราะห์ทุกคำเป็น VERB และมีความสัมพันธ์แบบ compound เช่น *เขาหยิบหนังสือมาอ่าน* หยิบ-compound>มา, มา-compound>อ่าน
- กรณีกริยาการเคลื่อนที่ที่วิเคราะห์ได้ว่าไม่ได้ใช้ในความหมายการเคลื่อนที่แต่กลายเป็นคำไวยากรณ์ ให้วิเคราะห์คำนั้นเป็น ADV และใช้ความสัมพันธ์ advmod ส่วนใหญ่เป็นประโยคที่มีคำเช่น *ไป มา ขึ้น ลง เข้า ออก* อยู่ท้ายประโยค เช่น *เขากลับมาแล้ว* กลับ-advmod>มา, กลับ-aux>แล้ว
แต่ถ้าเป็นกรณีที่เป็นคำกริยาที่ยังมีความหมายการเคลื่อนที่ เช่น *เขาเดินไปโรงเรียน* ก็จะวิเคราะห์เป็น VERB และความสัมพันธ์แบบ compound เหมือนกริยาเรียงปกติ
- กรณีที่ดูคล้ายเป็นคำกริยาเรียง แต่สิ่งที่กริยา subcategorize เป็นอนุพากย์ไม่ใช่คำกริยา ให้วิเคราะห์เป็น VEBR แต่ความสัมพันธ์เป็น xcomp, ccomp, หรือ advcl แล้วแต่กรณี เช่น *เขาชอบกินอาหารนอกบ้าน* สิ่งที่ชอบคือ การกินอาหารนอกบ้าน จึงวิเคราะห์เป็น *ชอบ-xcomp>กิน* ที่เป็น xcomp เพราะอนุพากย์นี้ไม่สามารถใส่ประธานได้ ประธานของ *กิน* ผูกกับประธานของ *ชอบ* เป็นคนเดียวกัน *เขาให้(แดง)ดูรูป* คำว่า *ให้* ในที่นี้ subcategorize อนุพากย์โดยที่อนุพากย์นี้อาจมีหรือไม่มีประธานและไม่ผูกกับประธานประโยคหลัก จึงวิเคราะห์เป็น *ให้-ccomp>ดู*
พายุพัดต้นไม้โค่น ไม่วิเคราะห์ *พัดโค่น* เป็น compound จะวิเคราะห์เหมือนประโยค *พายุพัดจนต้นไม้โค่น* จึงเป็น *พัด-advcl>โค่น* หรือถ้าเป็นประโยค *พายุพัดต้นไม้จนโค่น* ก็ยังวิเคราะห์ได้เหมือนเดิม คือ *พัด-advcl>โค่น* ส่วนประธานของ *โค่น* ไม่ได้ผูกกับกรรมของประโยคหลัก เป็น zero pronoun ที่ต้องหาจากบริบท เพราะสามารถเขียนประโยคเต็มเป็น *พายุพัดต้นไม้จนต้นไม้โค่น*

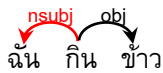
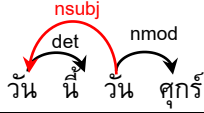
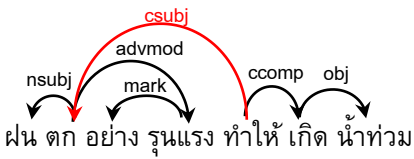
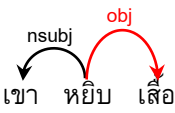
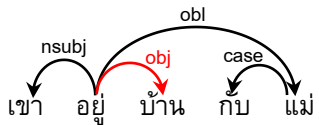
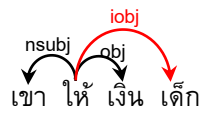
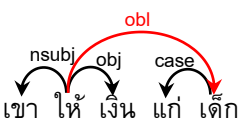
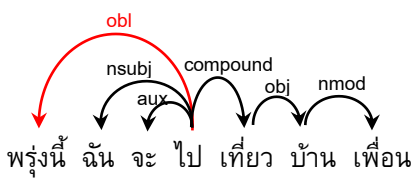
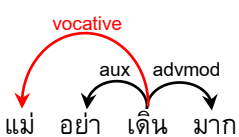
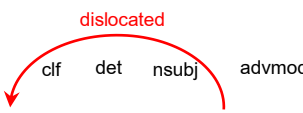
การกำกับความสัมพันธ์ระหว่างคำ

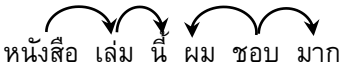
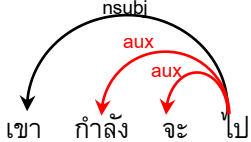
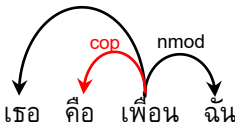
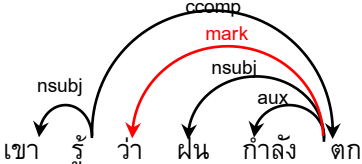
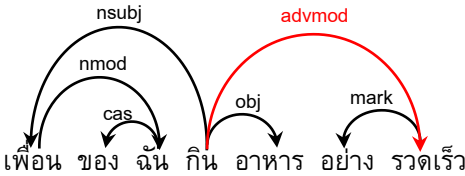
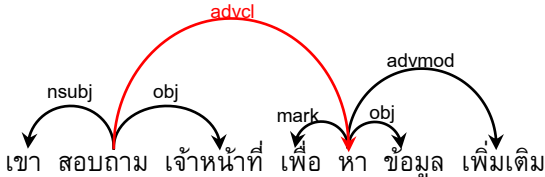
- ความสัมพันธ์แบบพึ่งพา dependency relation จะกำกับระหว่าง head กับ dependent ลูกศรตั้งต้นมาจาก head ไปที่ dependent



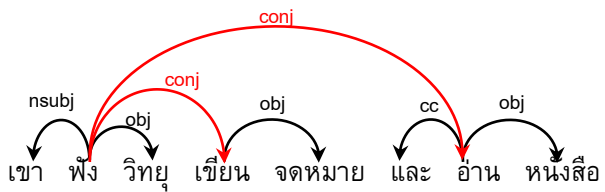
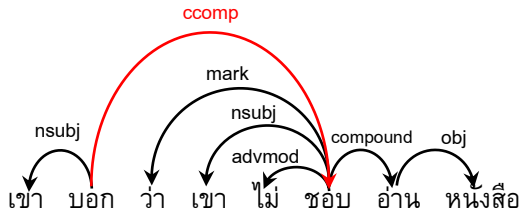
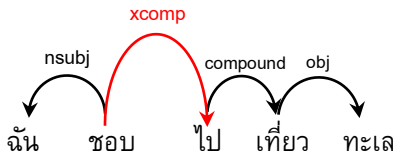
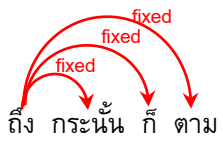
- โครงสร้างที่ได้จากความสัมพันธ์นี้จะออกมาเป็น dependency tree dependency จะโยงระหว่างคำเนื้อหาเป็นหลัก ในโครงสร้าง coordinate จึงโยงระหว่าง noun กับ noun หรือ verb กับ verb dependency tree ใน UD จึงต่างจาก dependency tree ใน approach อื่น เข้าใจว่า การโยงแบบ UD ทำให้เห็น relation ระหว่าง content word โดยตรง ซึ่งจะเอื้อต่อการวิเคราะห์ความหมายต่อไปได้ง่ายกว่า
- ส่วนคำไวยากรณ์จะเป็น dependent เป็นหลัก ไม่มีคำที่มา depend ด้วย กลุ่มคำไวยากรณ์จึงไม่ต่อกันเป็นสาย ถ้ามีหลาย ๆ ตัว ก็จะไปขยาย head เดียวกัน
- คำที่เป็น multiword เช่น in spite of จะโยงความสัมพันธ์ fixed ระหว่าง in >spite, in >of กรณีภาษาไทย ถ้าเป็นคำที่กำหนดให้เป็นคำแล้วก็ไม่ต้องทำแบบนี้ เช่น อย่างไรก็ตาม แต่หากโปรแกรมตัดคำไม่ได้ตัดเป็นคำเดียว เช่น ถึง|กระนั้น|ก็|ตาม ก็จะต้องกำกับความสัมพันธ์แบบ fixed นี้ ถึง->กระนั้น ถึง->ก็ ถึง->ตาม และใช้ relation "fixed"
- การละ จะไม่มีการเติมคำที่หายไป จะวิเคราะห์ไปตามรูปปรากฏ ถ้าคำที่หายเป็น dependent ก็ไม่ต้องทำอะไร ถ้าเป็นคำที่ควรเป็น head เช่นคำนาม ก็ให้ promote คำที่ควรเป็น dependent ของคำนามที่หายไปมาเป็น head แทน ตามลำดับดังนี้ amod > nummod > det > nmod > case เช่น He buys himself a green car and she (buys) herself a >obj>(red) สำหรับภาษาไทย ลำดับการ promote อาจไม่เป็นตามนี้ แต่หลักการคงเดิม คือไม่มีการเติมสิ่งที่ละไป
- Dependency structure ที่ใช้ในที่นี้เป็น basic tree สำหรับการใช้งานลักษณะ shallow ไม่ใช่ enhanced dependencies ที่ให้มี null node มีการบอกถึง controlled/raised subject ที่ใช้ร่วมกัน แบบ enhanced จึงมีรายละเอียดมากขึ้น มี cross dependency เกิดขึ้นได้ด้วย

รายการความสัมพันธ์

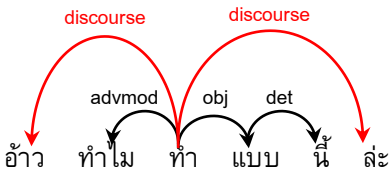
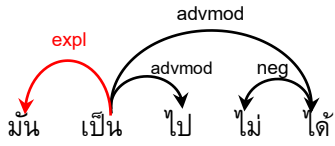
nsubj	Nominal subject	ความสัมพันธ์ระหว่างกริยากับประธาน ในกรณีที่ไม่มี predicate เช่น วันนี้วันศุกร์ ก็ให้ใช้ nsubj (วัน นี้) < (วัน ศุกร์)
		
		
csubj	Clausal subject	โยง subject ที่เป็น clause เช่น What she (said) < is (interesting)
		
obj	Object	ความสัมพันธ์ระหว่างกริยากับกรรม เช่น เขา (หยิบ) > (เสื่อ)
		
		
iobj	Indirect object	ความสัมพันธ์ระหว่างกริยากับกรรมรอง (ให้) เงิน > (เด็ก)
		
		
obl	oblique nominal	ใช้กับ non-core argument หรือ adjunct ดูจากรูปประโยคจะมีคำ ไวยากรณ์กำกับด้วย เช่น (ให้) เงิน แก่ > (เด็ก)
		
vocative	vocative	โยงกริยากับนามที่เป็นคำเรียกขาน (แม่) < อย่า (เดิน) มาก
		
dislocated	dislocated	โยงคำกริยากับนามที่ถูกย้ายที่ (หนังสือ) < เล่ม นี้ ผม (ชอบ) มาก
		

		
aux	auxiliary	โยงกริยากับคำไวยากรณ์ที่บอก tense, aspect, mood, voice, evidentiality เช่น เขา (กำลัง)<(จะ)<(ไป)
		
cop	copula	โยงระหว่าง function word ที่เชื่อม subject กับ nonverbal predicate เช่น he (is)<(honest), เธอ (คือ)<(เพื่อน) ฉัน, เขา (เป็น)<(คน) ดี แต่ ถ้าเป็นประโยคซับซ้อน เช่น การพินิจเป็นสิ่งที่มีอรรถาธิบายให้ วิเคราะห์ เป็น เป็นกริยาหลักที่มี nsubj และ obj relation กับประธาน และการกรรม nsubj
		
mark	marker	โยงระหว่างกริยากับคำไวยากรณ์ที่บอกถึงการเป็น subordinate clause เช่น เขา รู้(ว่า)<ฝน กำลัง (ตก) , นักเรียน (ที่)<(มา) วันนี้ หรือคำว่า อย่าง ที่ใช้นำหน้าคำขยายกริยา เช่น กิน (อย่าง)<(รวดเร็ว) ปกติเป็นคำในกลุ่ม complementizer, subordinate conjunction รวมถึงคำ เช่น คือ ที่เป็น discourse marker เชื่อมความต่อไป
		
advmod	Adverbial modifier	โยงระหว่างกริยากับคำขยายกริยา (กิน) อย่าง >(รวดเร็ว) หรือ ADV ที่ขยายคำอื่น (เกือบ)<(2000) (ไม่)< ชอบ
		
advcl	Adverbial clause modifier	โยงกับกริยาของ clause ที่ทำหน้าที่ขยายกริยาหลัก เช่น เขา (สอบถาม) เจ้าหน้าที่ เพื่อ >(หา) ข้อมูล เพิ่มเติม
		

acl	Clausal modifier of noun (adjectival clause)	คล้ายกับ advcl ที่อันนี้ใช้ขยายคำนาม เช่น ผม เจอ (หนังสือ) ที่ คุณ >(ชอบ) ความสัมพันธ์ระหว่าง การ >กริยาวิเศษณ์ เช่น การ >(สูญเสีย) บุคคลใกล้ชิดของเขา ก็วิเคราะห์เป็น acl
det	determiner	โยงคำนึงชี้เฉพาะหรือไม่เฉพาะที่ขยายคำนาม เช่น (หนังสือ) >(นี้), หนังสือ (กอง) >(นี้)
case	Case marking	โยงระหว่างคำนามกับบุพบท ให้ เงิน (แก่)< (เด็ก)
clf	Classifier	โยงคำนามที่เกี่ยวข้องกับคำลักษณนาม เข้ากับคำลักษณนาม เช่น (หนังสือ) สาม >(เล่ม) (ในภาษาอื่น classifier เป็นคำไวยากรณ์ จึงเป็น dependent เสมอ แต่ในภาษาไทย อาจใช้โดยไม่มีคำนามและทำหน้าที่เป็นคำนามแทน เช่น เล่ม นี้ จึงถูกจัดเป็น NOUN)
nmod	Nominal phrase modifier	คำนามหรือนามวลีที่มาขยายอีกคำนามหนึ่ง เช่น (ห้องทำงาน) ของ >(หัวหน้า)
amod	Adjectival modifier	โยงระหว่างคำคุณศัพท์ที่มาขยายคำนาม เช่น (เนื้อ) >(ดิบ)
appos	Appositional modifier	โยงระหว่าง nominal phrase ที่เท่ากันหรือ coreference เช่น (สมชาย) >(ผู้จัดการ) โรงแรม แห่ง นี้
nummod	Numeric modifier	โยงระหว่าง numeral กับคำนามที่เป็น head เช่น เงิน (60)< (บาท)

conj	conjunction	โยงระหว่างสองคำที่ coordinate กัน เช่น เขา (เขียน) จดหมาย และ > (อ่าน) หนังสือ
		
cc	Coordinating conjunction	โยงระหว่าง คำ ที่ coordinate กับคำสันธาน เช่น เขา เขียน จดหมาย (และ)< (อ่าน) หนังสือ
ccomp	Clausal complement	โยงระหว่างกริยาแรกกับกริยาสองที่เป็น dependent clause เมื่อทั้งกริยาตัวแรกและกริยาตัวที่สองต่างก็มีประธานของตนเอง เช่น เขา (บอก) ว่า เขา >(ชอบ) อ่าน หนังสือ (กรณีที่มีการละประธานตามปกติของภาษาไทย เช่น เขากล่าวว่าไม่ชอบอ่านหนังสือ ก็วิเคราะห์เป็น ccomp เหมือนเดิม)
		
xcomp	Open clausal complement	คล้ายกับ ccomp แต่กริยาของ dependent clause จะไม่มี subject คือใช้ subject ร่วมกับกริยาแรก เช่น He (likes) to >(swim) ภาษาไทย เช่น ฉัน (ชอบ) >(ไป) เทียว ทะเล (ถ้าเติมประธานไปประโยคจะใช้ไม่ได้ *ฉันชอบฉันไปเที่ยวทะเล) ส่วนกรณีถ้าเป็นความสัมพันธ์ของ serial verb ให้วิเคราะห์เป็น compound
		
fixed	Fixed multiword expression	เป็น 1 ใน 3 relation ของ MWE (multi-word expression) ใช้กรณีคำที่เป็น multiword ที่จะต้องรวบเป็นคำเดียว เช่น ถึงกระนั้นก็ตาม = (ถึง) > (กระนั้น), (ถึง) >(ก็), (ถึง) >(ตาม)
		
compound	compound	เป็น 1 ใน 3 relation ของ MWE โยงความสัมพันธ์ของคำประสม เช่น (สมุด) >(โทรศัพท์) Relation นี้ใน UD v2 ให้ใช้สำหรับ serial verb ด้วยโดยใช้ compound: svc ส่วน Compound: prt ใช้กับ particle verb (put)>(up)

flat	Flat multiword expression	เป็น 1 ใน 3 relation ของ MWE คล้ายกับ compound แต่ flat ใช้กับ exocentric MWE ในขณะที่ compound เป็นลักษณะแบบ endocentric เช่น (1) >(มกราคม) >(2017), (นายก) >(ชวน), (นาย)< (สมชาย), (สมชาย) >(กล้าหาญ) ภาษาอังกฤษ (Mr) >(Somchai) ภาษาไทย ชื่อ น่าจะเป็นคำหลักมากกว่า ยกเว้นกรณี นายก >ชื่อ ประเทศ >ชื่อ รวมถึงกรณีคำที่เป็นชื่อต่าง ๆ เช่น ชื่อผลิตภัณฑ์ บริการ เช่น “สินเชื่о เร็วทันใจ” “บัตรสไมล์ทันใจ” ก็ให้ใช้ flat เชื่อมระหว่างคำในชื่อนั้น
parataxis	parataxis	ความสัมพันธ์แบบต่อกันโดยไม่มีการเชื่อม มักใช้กับการให้วงเล็บขยายความ เช่น (Washington) (>(CNN))
list	list	ความสัมพันธ์แบบเป็นรายการต่อๆ กัน เช่น (แอปเปิ้ล) >(มะละกอ) >(กล้วย) >(ส้ม)
orphan	orphan	ความสัมพันธ์ของ dependent ที่ไม่เข้าตามมาตรฐานที่กล่าวมา ใช้กรณีที่มีการละคำ ทำให้เกิดความสัมพันธ์ของคู่คำที่ปกติไม่เกิด orphan จึงเป็นตัวบอกว่ามีบางอย่างหายไป ณ ตรงนั้น
punct	Punctuation	โยงระหว่าง predicate หรือคำหลักอื่นกับเครื่องหมายวรรคตอน เช่น ทำไม(กล้า) เรียน punct>? , อุบัติเหตุมักเกิด(ซ้ำ) punct>ฯ

reparandum	Overridden disfluency	ใช้กับถ้อยคำที่พูดผิดแล้วแก้ส่วนใหม่ที่ถูกมาแทน เช่น ไป ทาง (ซ้าย) < – ทาง (ขวา) แก่คำพูดเปลี่ยนจาก ซ้าย เป็น ขวา
dep	Unspecified dependency	ใช้เมื่อไม่สามารถระบุว่าเป็นความสัมพันธ์แบบใดเลย จึงให้เป็น dep ไว้ก่อน
root	root	โยง ROOT เข้ากับคำที่เป็น head หลักของประโยคหรือส่วนนั้น
discourse	discourse element	ความสัมพันธ์นี้ใช้โยงคำในกลุ่ม INTERJECTIONS เช่น โอ้ อืม โอ้โฮ, กลุ่ม discourse marker (actually, well, like) โดยโยงกริยาหลักไปยังคำอุทานเหล่านั้น
		
expl	expletive	ใช้โยงความสัมพันธ์ระหว่างกริยากับสิ่งที่ไม่ได้เป็นคำนามชี้เฉพาะชัดเจน เช่น มัน ใน (มัน) < (เป็น) ไป ไม่ได้ จะเห็นได้ว่าคำว่า มัน อาจละทิ้งได้โดยไม่ส่งผลต่อความหมาย กล่าวคือ มัน ในตัวอย่างนี้เป็นคำนาม แต่ไม่มี semantic role กับ predicate
		

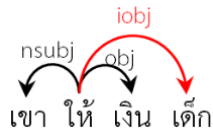
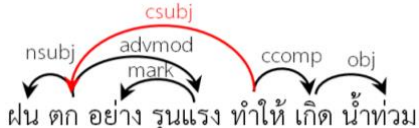
ตารางแสดงความสัมพันธ์รูปแบบต่าง ๆ ระหว่างคำ

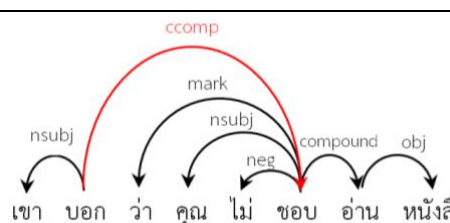
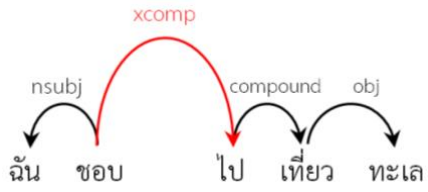
	Nominals	Clauses	Modifier Words	Function Words
Core arguments	nsubj obj iobj	csubj ccomp xcomp		
Non-core dependents	obl vocative expl dislocated	advcl	advmod discourse	aux cop mark
Nominal dependents	nmod appos nummod	acl	amod	det clf case
Coordination	MWE ²	Loose	Special	Other
conj cc	fixed flat compound	list parataxis	orphan goeswith reparandum	punct root dep

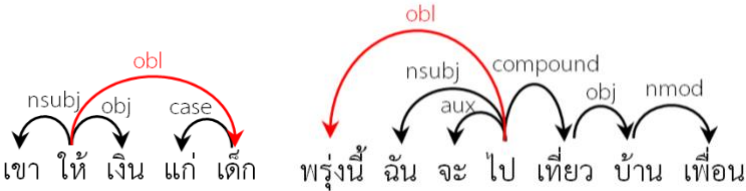
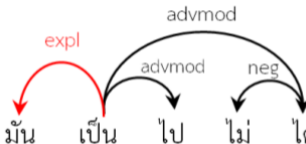
² MWE: multi-word expression

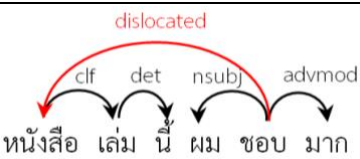
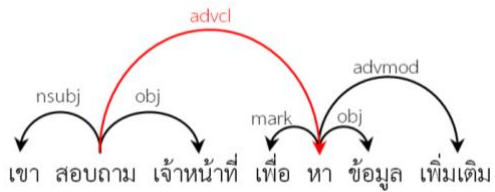
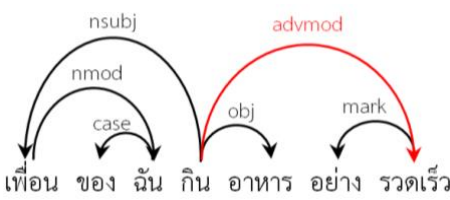
คำอธิบายความสัมพันธ์โดยละเอียด

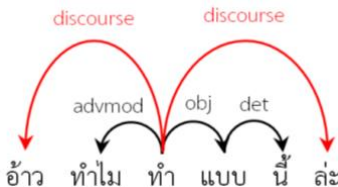
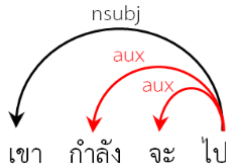
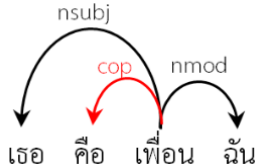
Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
Core arguments - Nominals	1	nsubj	nominal subject	ความสัมพันธ์ระหว่างกริยากับประธาน ประธานของประโยค อาจเป็น คำนาม (NOUN)สรรพนาม (PRON) หรืออาจเป็นคำคุณศัพท์ ADJ ได้ในกรณีที่เกิดการละ โดยที่ประธานของประโยคอาจไม่จำเป็นต้องเป็นผู้กระทำกริยานั้น ๆ (agent) โดยตรงก็ได้ หากประโยคมีความสัมพันธ์แบบกรรมวาจก (passive) จะระบุความสัมพันธ์ของกริยาโยงไปยังประธาน เป็น nsubj:pass ความสัมพันธ์แบบ nsubj อาจไม่จำเป็นต้องตั้งต้นมาจาก VERB เสมอไป ในกรณีที่ VERB ที่ปรากฏในประโยคนั้นเป็น COPULAR VERB ส่วนตั้งต้นหรือ head จะเป็นสิ่งที่มาเติมเต็ม COPULAR VERB นั้น ซึ่งอาจเป็น ADJ หรือ NOUN ก็ได้ ในกรณีที่ไม่มี predicate เช่น <i>วันนี้วันศุกร์</i> ก็ให้ใช้ nsubj (<i>วัน นี้</i>) < (<i>วัน ศุกร์</i>) เมื่อไหร่ก็ตามที่ประธานมีลักษณะเป็น empty argument (ประโยคที่ขึ้นต้นด้วย there) ความสัมพันธ์ระหว่าง VERB กับประธานจะเป็น <u>expl</u> แทน ในภาษาไทย อาจพบในประโยค เช่น <i>มันเป็นไปได้ว่า... มัน</i> <-expl-เป็น มันเป็นเหมือน dummy subject	
Core arguments - Nominals	2	obj	object	ความสัมพันธ์ระหว่างกริยากับกรรมตรงของกริยานั้น โดยปกติจะเป็นคำนาม และหากกริยานั้นมีกรรมเพียงแค่ตัวเดียวจะระบุความสัมพันธ์เป็น obj แต่หากกริยาเป็นกริยาทวิกรรมคือมีกรรมสองตัว หรือมีกรรมมากกว่า 2 ตัว จะโยงความสัมพันธ์ให้กรรมตัวหนึ่งเป็น obj (เป็นกรรมที่รับผลจากการกระทำหรือกริยานั้นโดยตรง) ส่วนที่เหลือเป็น <u>iobj</u>	

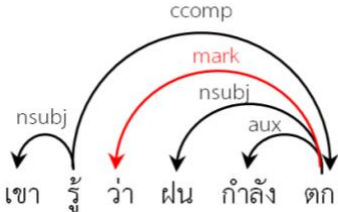
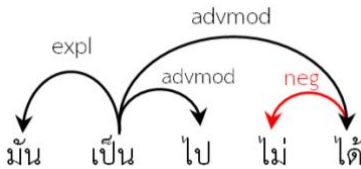
Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
Core arguments - arguments - Nominals	3	iobj	indirect object	iobj เป็นความสัมพันธ์หลักอีกความสัมพันธ์หนึ่ง โยงระหว่างกริยากับกรรมรอง ไม่มีคำบุพบทมาคั่นกลาง ในกรณีที่กรรมรองนำหน้าด้วยคำบุพบท จะถือว่ากริยากับกรรมรองนั้นมีความสัมพันธ์แบบ obl	
					
Core arguments - Clauses	4	csbj	clausal subject	เป็นรูปแบบความสัมพันธ์ของกริยากับประธานของประโยค โดยประธานไม่ใช่คำนามเดิยวโดด ๆ แต่มีลักษณะเป็นอนุประโยค (clause) การโยงความสัมพันธ์จะโยงจากกริยาหลัก (หรือ head ของส่วนเติมเต็มของกริยา COPULAR) ไปยัง head ของอนุประโยคที่เป็นประธาน ซึ่งอาจจะเป็นกริยาหรือนามก็ได้ ในกรณีที่เป็นประโยคกรรมวาจก (passive) การโยงจะระบุเพิ่มเติมเป็น csbj:pass	
					
Core arguments - Clauses	5	ccomp	clausal complement	เป็นประเภทการโยงกริยาหลัก (VERB หรือ ADJ) เข้ากับ head ของส่วนที่มาเติมเต็ม อันปรากฏในตำแหน่งกรรมของประโยค ข้อสังเกตการโยงความสัมพันธ์นี้คือ ประธานของประโยคหลักกับประธานของอนุประโยคที่มาเติมเต็มสามารถเป็นคนละคนหรือคนละสิ่งกันก็ได้ เช่น เขา (บอก) ว่า คุณ >(ชอบ) อ่าน หนังสือ แม้กรณีที่ไม่ปรากฏประธาน เช่น เขา (บอก) ว่า >(ชอบ) อ่าน หนังสือ ก็ยังคงเป็นความสัมพันธ์ ccomp	

Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
					
Core arguments - Clauses	6	xcomp	open clausal complement	เป็นประเภทการโยงกริยาหลัก (VERB หรือ ADJ) เข้ากับ head ของส่วนที่มาเติมเต็ม ความแตกต่างระหว่าง ccomp กับ xcomp คือ ความสัมพันธ์แบบ xcomp นั้น ประธานในอนุประโยคเติมเต็ม ไม่สามารถมีได้จะเป็น subject เดียวกับของประโยคหลัก เช่น <i>He (likes) to >(swim)</i> ภาษาไทย เช่น ฉัน (ชอบ) >(ไป) เทียว ทะเล อาจมีการใช้คำเชื่อมหน้ากริยาหลังเป็น <i>ที่จะ...</i> <i>ที่ว่า....</i> ก็ได้ เช่น เขา (ชอบ) <i>ที่จะ >(ไป) ทำงาน สาย</i> คำเชื่อมเหล่านี้บอกความเป็น xcomp คือไม่สามารถมีประธานหน้ากริยาได้ ถ้าเทียบกับภาษาอังกฤษก็จะเป็น non-finite (กรณีประโยคอย่าง <i>ฉันชอบหนังสืออ่านแล้วตื่นเต้น</i> ไม่วิเคราะห์ <i>อ่านแล้วตื่นเต้น</i> เป็น xcomp แต่วิเคราะห์เป็น (หนังสือ) -acl->(อ่าน) แล้ว ตื่นเต้น คล้ายกับ <i>ฉันชอบหนังสือที่อ่านแล้วตื่นเต้น</i>)	
					
Non-core dependent - Nominals	7	obl	oblique nominal	ใช้โยงความสัมพันธ์ระหว่างคำกริยาหลัก คำคุณศัพท์หรือวิเศษณ์ กับคำนามในตำแหน่งกรรมของกริยานั้น แต่ต้องไม่ใช่กรรมตรง ใช้กับ non-core argument หรือ adjunct ข้อสังเกตของความสัมพันธ์แบบ obl คือคำนามนั้น ๆ มักนำหน้าด้วย ADPOSITION และความสัมพันธ์ระหว่างคำนามกับ ADP ข้างหน้าคือ case นอกจากนี้ obl ยังใช้โยงความสัมพันธ์ระหว่างคำขยายที่เป็นคำนามอยู่ข้างหน้าประโยคอีกด้วย เช่น <i>(Last night) <, I (swam) in the pool</i> หรือภาษาไทยเช่น <i>(พรุ่งนี้) < ฉัน จะ (ไป) เทียว บ้าน เพื่อน</i>	

Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
					
Non-core dependent - Nominals	8	vocative	vocative	แสดงความสัมพันธ์ระหว่างคำกริยากับคำนามที่เป็นคำเรียกขาน เช่น (แม่)<อย่า (เดิน) มาก	
					
Non-core dependent - Nominals	9	expl	expletive	ใช้โยงความสัมพันธ์ระหว่างกริยากับสิ่งที่ไม่ได้เป็นคำนามชี้เฉพาะชัดเจน เช่น มัน ใน (มัน)<(เป็น) ไป ไม่ได้ จะเห็นได้ว่าคำว่า มัน อาจจะทิ้งได้โดยไม่ส่งผลต่อความหมาย กล่าวคือ มัน ในตัวอย่างนี้เป็นคำนาม แต่ไม่มี semantic role กับ predicate	
					
Non-core dependent - Nominals	10	dislocated	dislocated element	โยงคำกริยากับคำนามที่ถูกย้ายที่ เช่น (หนังสือ)<เล่ม นี้ ผม (ชอบ) มาก	

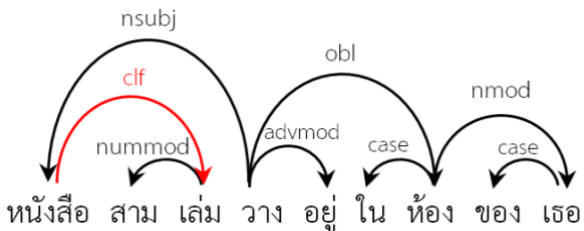
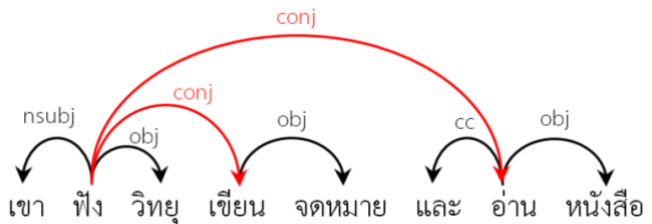
Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
 หนังสือ เล่ม นี่ ผม ชอบ มาก					
Non-core dependent - Clauses	11	advcl	adverbial clause modifier	โยงความสัมพันธ์ระหว่างภาคแสดงกับส่วนขยายภาคแสดงซึ่งมีลักษณะเป็นอนุประโยค เช่น เขา (สอบถาม) เจ้าหน้าที่ เพื่อ >(หา) ข้อมูล เพิ่มเติม	
 เขา สอบถาม เจ้าหน้าที่ เพื่อ หา ข้อมูล เพิ่มเติม					
Non-core dependent - Modifier words	12	advmod	adverbial modifier	โยงความสัมพันธ์ระหว่างกริยากับคำขยายกริยา เช่น (กิน) อย่าง >(รวดเร็ว) รวมถึงคำขยายอื่น ๆ ด้วย เช่น (เกือบ)< (2000) (หลาย)< (เล่ม)	
 เพื่อน ของ ฉัน กิน อาหาร อย่าง รวดเร็ว					
Non-core dependent - Modifier words	13	discourse	discourse element	ความสัมพันธ์นี้ใช้โยงคำในกลุ่ม INTERJECTIONS เช่น โอ้ อืม โอ้โฮ, กลุ่ม discourse marker (actually, well, like) โดยโยงกริยาหลักไปยังคำอุทานเหล่านั้น นอกจากนี้ยังสามารถใช้โยงคำกริยาไปยังคำลงท้าย (PART:RES) ได้อีกด้วย	

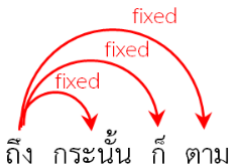
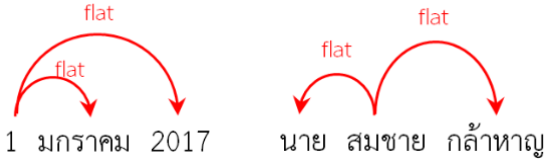
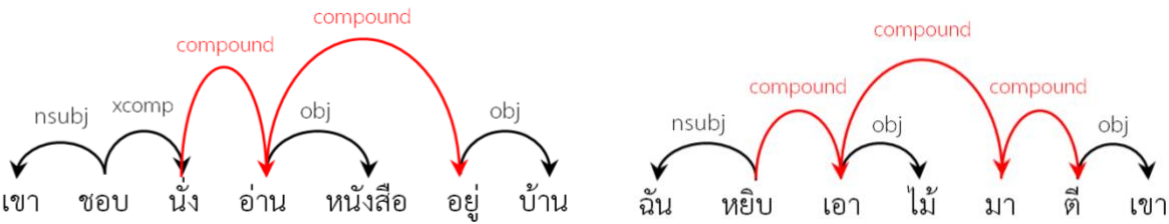
Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
					
Non-core dependent - Function Words	14	aux	auxiliary	เป็นความสัมพันธ์ระหว่างคำกริยากับคำช่วยกริยาต่าง ๆ (คำไวยกรณ์ที่บอก tense, aspect, mood, voice, evidentiality) โดยคำกริยาจะเป็น head เสมอ เป็นตัวตั้งต้นโยงไปยังคำช่วยกริยาแต่ละคำ เช่น เขา (กำลัง)< (จะ)< (ไป)	
					
Non-core dependent - Function Words	15	cop	copula	โยงระหว่าง function word ที่เชื่อม subject กับ nonverbal predicate คำนามทางด้านซ้ายและขวาของ VERB:COP ต้องอ้างถึงสิ่งเดียวกันหรือบอกคุณลักษณะสิ่งนั้น เช่น he (is)< (honest) เธอ (คือ)< (เพื่อนฉัน) ในภาษาไทยจะพบความสัมพันธ์นี้ในประโยคที่มีคำว่า คือ แต่กรณีที่เป็นการโยงความด้วยอนุพากย์ ให้วิเคราะห์คำนั้นเป็นกริยาทั่วไป เช่น สิ่งที่สำคัญคือต้องพยายามทำงานให้หนัก (สิ่ง)<nsubj- (คือ), (คือ)-ccomp->(ทำงาน) หรือ การพนันเป็นสิ่งที่มีมอมเมาประชาชน ก็ให้เป็น การพนัน<nsubj (เป็น), (เป็น) obj>สิ่ง	
					

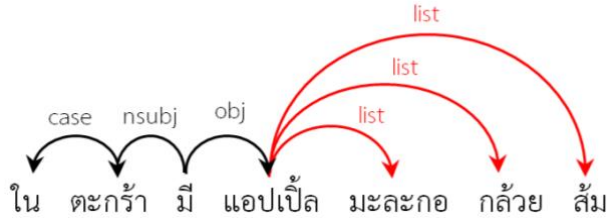
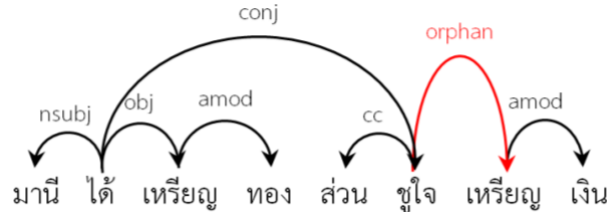
Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
Non-core dependent - Function Words	16	mark	marker	โยงระหว่างกริยากับคำไวยากรณ์ที่บอกถึงการเป็น subordinate clause นั่นคือปกติจะโยงคำกริยาในอนุประโยคเติมเต็มเข้ากับคำในกลุ่ม complementizer, subordinate conjunction เช่น เขา <u>รู้</u> (ว่า) < ผ่นกำลัง (ตก) หรือคำว่า <u>อย่าง</u> ที่ใช้นำหน้าคำขยายกริยา เช่น กิน (อย่าง) < (รวดเร็ว) กรณีของ discourse particle ที่ไม่ใช่ final particle ก็ให้วิเคราะห์ความสัมพันธ์เป็น mark ได้ เช่น (เพลง) <u>ที่</u> ได้เข้าชิงรางวัล (ได้แก่) < mark < (X) Y Z.	
					
Non-core dependent - Function Words	17	neg	negation	ใช้โยงความสัมพันธ์ระหว่างกริยาหลักกับคำปฏิเสธ ในภาษาไทย บางกรณีคำปฏิเสธไม่ได้ปรากฏกับคำกริยาเสมอไป อาจปรากฏร่วมกับคำขยาย เช่น ทำ <u>ไม่ได้</u> ได้ เป็น ADV ขยาย ทำ เราก็จะโยงความสัมพันธ์ neg ระหว่าง <u>ได้</u>/ADV กับ <u>ไม่</u>/PART:NEG ที่อยู่ข้างหน้า เนื่องจากใน UD ภาษาส่วนใหญ่ให้ ความสัมพันธ์ระหว่างคำปฏิเสธกับกริยาเป็น advmod ปัจจุบันจึงเปลี่ยนมาใช้ advmod	
					
Nominal dependent - Nominals	18	nmod	nominal phrase modifier	คำนามหรือนามวลีที่มาขยายอีกคำนามหนึ่ง เช่น (ห้องทำงาน) ของ > (หัวหน้า)	

Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
				ห้องทำงาน ของ หัวหน้า ของ ฝ่าย บริหาร ของ โรงเรียน	
Nominal dependent - Nominals	19	appos	appositional modifier	โยงระหว่าง nominal phrase ที่เท่ากันหรือ coreference สิ่งที่อยู่ในวงเล็บ หรือคำอธิบายอักษรย่อ (สองคำอาจวางสลับตำแหน่งกันได้) โดย appos จะโยงจากซ้ายไปขวาเสมอ สมชาย ผู้จัดการ โรงแรม แห่ง นี้	
Nominal dependent - Nominals	20	nummod	numeric modifier	เป็นความสัมพันธ์ที่โยงระหว่างคำนามกับตัวเลข เช่น เงิน (60) < (บาท) กรณีคำขยายบอกปริมาณ เช่น หลาย บาง ทุก แต่ละ ให้วิเคราะห์เป็น advmod ขยายนามที่ปรากฏรวม	
				ฉัน มี เงิน 60 บาท	
Nominal dependent - Clauses	21	acl	clausal modifier of noun (adjectival clause)	คล้ายกับ advcl แต่ในกรณีนี้ใช้เพื่อขยายคำนาม เช่น ผม เจอ (หนังสือ) ที่ คุณ > (ชอบ)	

Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
Nominal dependent - Modifier words	22	amod	adjectival modifier	โยงระหว่างคำคุณศัพท์ที่มาขยายคำนาม เช่น (เนื้อ) >(ดิบ)	
Nominal dependent - Function Words	23	det	determiner	โยงคำบ่งชี้เฉพาะหรือไม่เฉพาะที่ขยายคำนาม เช่น (หนังสือ) >(นี้), หนังสือ (กอง) >(นี้) โดยทั่วไปจะโยงคำนามเข้ากับคำที่มี POS เป็น DET	
Nominal dependent - Function Words	24	clf	classifier	ใน UD v2 ใช้โยงตัวเลขบอกจำนวนไปยังคำลักษณะนามโดยยกตัวอย่างภาษาจีน three CLF student และวิเคราะห์เป็น three-clf>CLF, three<-nummod-student ส่วน this CLF bus วิเคราะห์เป็น this-clf->CLF this<-det-bus แต่ในกรณีที่มีการละคำนามหลักก็ให้ CLF เป็นคำหลัก three CLF วิเคราะห์เป็น three<-nummod-CLF ซึ่งดูไม่คงที่ และตำราไวยากรณ์ไทยมักวิเคราะห์คำลักษณะนามเป็นคำนามประเภทหนึ่ง จึงควรวิเคราะห์ให้เลขจำนวนนับสัมพันธ์กับคำลักษณะนามมากกว่าคำนามหลัก เช่น นักเรียนสามคน จะโยงความสัมพันธ์เป็น (นักเรียน)-clf->(คน), (สาม)<-nummod-(คน) , รถคันนี้ จะวิเคราะห์เป็น (รถ)-clf->(คัน), (คัน)-det->(นี้) และเมื่อมีการละคำนามหลัก สามเล่มนี้ ก็ยังวิเคราะห์แบบเดิมได้ (สาม)<-nummod-(เล่ม), (เล่ม)-det->(นี้)	

Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
 หนังสือ สาม เล่ม วาง อยู่ ใน ห้อง ของ เธอ					
Nominal dependent - Function Words	25	case	case marking	โยงระหว่างคำนามกับบุพบท (เมื่อไหร่ที่ข้างหน้าเป็น SCONJ ความสัมพันธ์จะเป็น mark แต่ถ้าข้างหน้าเป็น ADP ความสัมพันธ์จะเป็น case)	
Coordination	26	conj	conjunct	โยงระหว่างสองคำที่ coordinate กัน เช่น เขา (เขียน) จดหมาย และ >(อ่าน) หนังสือ	
 เขา ฟัง วิทย์ เขียน จดหมาย และ อ่าน หนังสือ					
Coordination	27	cc	coordinating conjunction	โยงระหว่างคำที่ coordinate เป็นคำสุดท้ายกับคำสันธาน	
MWE	28	fixed	fixed multiword expression	ใช้กรณีคำที่เป็น multiword ที่จะต้องรวมเป็นคำเดียว เช่น ถึงกระนั้นก็ตาม = ถึง >กระนั้น, ถึง >ก็, ถึง >ตาม	

Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
					
MWE	29	flat	flat multiword expression	flat ใช้กับ exocentric MWE คล้ายกับ compound แต่ flat ใช้กับ exocentric MWE (ในขณะที่ compound เป็นลักษณะแบบ endocentric) ใช้กับชื่อคน (โยงชื่อต้น ไปหา ชื่อกลาง, ชื่อต้น ไปหา นามสกุล) โยงวัน-เดือน-ปี หรือคำภาษาต่างประเทศ โดยโยงจากซ้ายไปขวา เช่น (1) >(มกราคม) >(2017), (นาย) >(ชวน), (นาย) <(สมชาย), (สมชาย) >(กล้าหาญ) (ภาษาอังกฤษ (Mr.) >(Somchai) แต่ภาษาไทย ชื่อ น่าจะเป็นคำหลักมากกว่า ยกเว้นกรณี นาย >ชื่อ ประเทศ >ชื่อ)	
					
MWE	30	compound	compound	ใช้โยงคำในกลุ่มคำประสม เช่น สมุด>โทรศัพท์ รวมถึงคำกริยาเรียง ใน UD v2 ก็จะทำให้ความสัมพันธ์นี้ด้วย วิธีโยงความสัมพันธ์คือโยงต่อกันไป เช่น หม้อหุงข้าว หม้อ>หุง หุง>ข้าว	
					

Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
Loose	31	list	list	ใช้โยงคำในรายการ โดยต้องมีจำนวนมากกว่า 2 รายการ ตัวตั้งต้นเป็นรายการแรกเสมอ	
					
Loose	32	parataxis	parataxis	ความสัมพันธ์แบบต่อกันโดยไม่มีการเชื่อม เช่น (Washington) (>(CNN)) โดยปกติจะโยงจากซ้ายไปขวา ใช้โยงเข้ากับคำหรือข้อความที่อยู่ในวงเล็บ	
Special	33	orphan	orphan	ความสัมพันธ์ของ dependent ที่ไม่เข้าตามมาตรฐานที่กล่าวมา ใช้กรณีที่มีการละ ทำให้เกิดความสัมพันธ์ของคู่คำที่ปกติไม่เกิด เช่น มานี้ ได้ เหรียญทอง ส่วน (ซูโจ) >(เหรียญเงิน) ในตัวอย่างนี้มีการละกริยาในประโยคส่วนที่สอง ปกติความโยง ได้/VERB เข้ากับกริยาหลักของประโยคส่วนที่ 2 แต่เมื่อเกิดการละเช่นนี้ จะ promote ให้ ซูโจ เป็น head ของประโยคส่วนที่สองโดยที่ยังคงมี POS เป็น PROPN แล้วโยงความสัมพันธ์ระหว่าง ซูโจ กับ เหรียญเงิน เป็น orphan	
					

Type	ที่	Tagging UD	ชื่อเต็ม	คำอธิบาย	หมายเหตุ
Special	34	goeswith	goes with	ใช้โยงความสัมพันธ์ของคำที่อื่นที่จริงต้องเขียนติดกัน แต่อาจจะเขียนผิดหรือแก้ไขผิดพลาดจนทำให้ตัวหนังสือแยกจากกัน เช่น <i>with out, never the less</i> ในภาษาไทยอาจไม่พบกรณีเช่นนี้มากนัก อย่างไรก็ตาม หากเกิดกรณีตัดคำมาผิด เช่น เกื่อ-กุล และไม่สามารถแก้ไขการตัดคำได้ ให้โยง (เกื่อ) > (กุล) เข้าหากันโดยใช้ความสัมพันธ์นี้ โยงจากซ้ายไปขวา	
Special	35	reparandum	overridden disfluency	ใช้กับถ้อยคำที่พูดผิดแล้วแก้ส่วนใหม่ที่ถูกมาแทน เช่น ไป ทาง (ซ้าย) < - ทาง (ขวา) แก้คำพูดเปลี่ยนจาก ซ้าย เป็น ขวา	
Other	36	punct	punctuation	โยงระหว่าง predicate กับเครื่องหมายวรรคตอน	
Other	37	root	root	โยง root เข้ากับคำที่เป็น head หลักของประโยคหรือส่วนนั้น	
Other	38	dep	unspecified dependency	ใช้เมื่อไม่สามารถระบุว่าเป็นความสัมพันธ์แบบใดเลย หากเป็นไปได้ ขอให้หลีกเลี่ยงการโยงรูปแบบความสัมพันธ์นี้	