การแปลภาษาล้านนาเป็นภาษาไทย
Lanna to Thai Language Translation

พุธษดี ศิริแสงตระกูล (Pusadee Seresangtakul)[1*]
นิกร ยาพรม (Nikorn Yaprom)[2]

บทคัดย่อ

บทความนี้ ผู้วิจัยได้นำเสนอการพัฒนาระบบการแปลภาษาล้านนาเป็นภาษาไทย โดยในการพัฒนาระบบงานนี้ ผู้วิจัยได้สร้างพจนานุกรมล้านนา-ไทย ซึ่งประกอบด้วยคำล้านนาจำนวน 7,497 คำ ข้อความที่ป้อนเข้าสู่ระบบประกอบด้วยคำ ประโยค และวลีล้านนา โดยข้อความที่ป้อนเข้าสู่ระบบจะถูกตัดคำโดยใช้หลักการเทียบคำแบบยาวที่สุด (Longest Matching) การวิเคราะห์และแปลประโยคและวลีล้านนาเป็นภาษาไทยใช้เทคนิคที่ผสมผสานกันระหว่างพจนานุกรมและข่ายงานการเปลี่ยนเพิ่มขยาย (Augmented Transition Networks: ATNs) และได้ประยุกต์ออโตมาตาจำกัดเชิงไม่กำหนด (Nondeterministic Finite Automata: NFA) ในการปริวรรตคำล้านนาที่ไม่พบในพจนานุกรมเป็นภาษาไทย

ในการวัดประสิทธิภาพของระบบผู้วิจัยได้ทำการทดสอบแปลประโยคซึ่งพิมพ์จากอักษรธรรมล้านนาจำนวน 558 ประโยค ผลการทดลองพบว่าระบบมีประสิทธิภาพความถูกต้องในการแปลเท่ากับ 83.69%

ABSTRACT

The purpose of this research was to develop a Lanna to Thai language translation system. In order to develop the system, we created a Lanna to Thai dictionary. The dictionary comprises 7,497 Lanna words. The input of this system is a sequence of Lanna words, sentences and phrases. The input texts are segmented into sequences of Lanna words by using the longest matching algorithm. The Lanna sentences and phrases are analyzed and translated into the Thai language by using a hybrid of dictionary and Augmented Transition Networks (ATNs). We applied the Nondeterministic Finite Automata (NFA) to translate unknown words. In order to evaluate the efficiency of the system, 558 sentences of the Lanna Dharma script were input to test the system. The results of experiments showed that the correctness of the system is 83.69%.

คำสำคัญ: ภาษาล้านนา, ข่ายงานการเปลี่ยนเพิ่มขยาย, การแปลภาษา
Keywords: Lanna, Augmented Transition Networks, ATNs, Translation, NFA

[1]*ผู้ช่วยศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น*
[2]*อาจารย์สถาบันการพลศึกษา วิทยาเขตสุโขทัย*
[*]*Corresponding author, e-mail: pusadee@kku.ac.th*

บทนำ

ภาษาล้านนา (อุดม, 2524) เป็นภาษาประจำราชอาณาจักรล้านนา ปัจจุบันใช้สื่อสารในเขตจังหวัดภาคเหนือตอนบนของไทย มีภาษาพูดและภาษาเขียนเป็นของตนเอง มีอักษรธรรมล้านนาสำหรับจารึกวรรณกรรม ตำรา และภูมิปัญญาท้องถิ่นอันเป็นประโยชน์ อักษรธรรมล้านนามีสัณฐานเป็นเส้นโค้งอ่อนไหว มีส่วนกว้างมากกว่าส่วนสูง ประกอบด้วยพยัญชนะ 42 ตัว รูปสระประกอบด้วยสระลอย 8 รูป สระจม 29 รูป และสระพิเศษ 6 รูป มีวรรณยุกต์ 2 รูป แต่สามารถผันวรรณยุกต์ได้ถึง 8 เสียง

ภาษาล้านนาต่างจากภาษาอื่นๆ ที่ลำดับการเขียน โดยสามารถเขียนวรรณยุกต์ก่อนหรือหลังตัวสะกดได้ เช่น ᨠᩮᩢᨶ/kɛːn^3/ “แก่น”(เขียนวรรณยุกต์ก่อนตัวสะกด) และ ᨠᩮᨶᩢ (เขียนวรรณยุกต์หลังตัวสะกด) ซึ่งผลที่ได้ไม่ต่างกัน ทำให้การแปลแบบคำต่อคำไม่สามารถกระทำได้

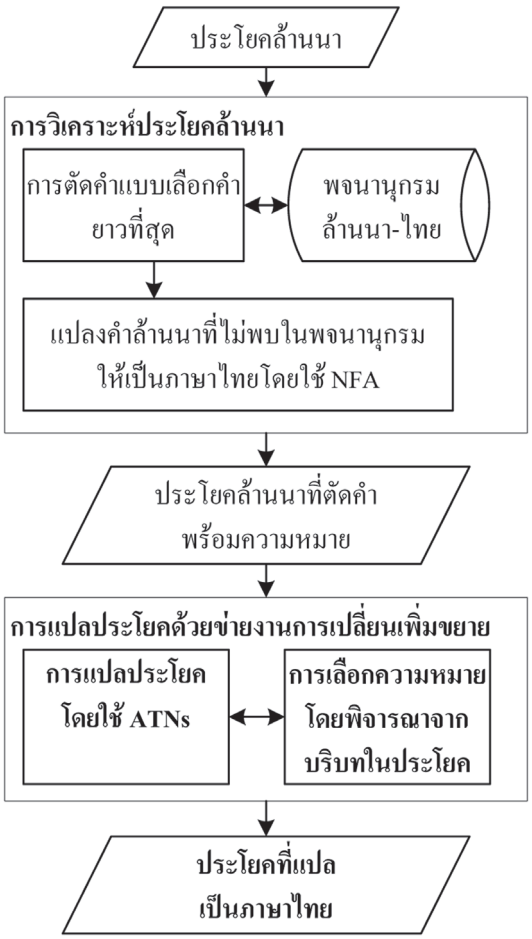
ในงานวิจัยนี้จะใช้เสียงวรรณยุกต์ล้านนาแบบ 8 เสียง (กรรณิการ์, 2549) ในปัจจุบันมีผู้อ่านและแปลภาษาล้านนาได้น้อยมาก ทำให้สิ่งที่จารึกไว้ซึ่งมีอยู่มากมาย ไม่สามารถนำมาเผยแพร่ให้เกิดประโยชน์ต่อสังคมได้

การแปลภาษาล้านนานิยมใช้วิธีการปริวรรตอักษร แบบคำต่อคำ ส่วนระบบการแปลภาษาล้านนาเป็นภาษาไทยด้วยคอมพิวเตอร์นั้น นิกรและพุธษดี (2550) ได้เสนอวิธีการแปลโดยใช้ ข่ายงานการเปลี่ยนเพิ่มขยาย (Augmented Transition Networks: ATNs) สร้างกฎโครงสร้างประโยคสามัญ 3 กฎ และกฎโครงสร้างวลี 4 กฎ และสร้างพจนานุกรมล้านนา-ไทย เป็นเครื่องมือในการแปล การตัดคำใช้วิธีการเทียบคำที่ยาวที่สุด และสร้างกฎปริวรรตอักษรล้านนาเป็นอักษรไทย 226 กฎ มีประสิทธิภาพความถูกต้องในการแปล 78.48% การแปลผิดเนื่องจากโครงสร้างประโยคสามัญจะประกอบด้วยภาคประธาน ภาคแสดง และภาคขยายเท่านั้น ซึ่งไม่ครอบคลุมรูปแบบประโยคทั้งหมดในภาษาล้านนา

ในงานวิจัยนี้ ผู้วิจัยได้นำเสนอการแปลภาษาล้านนาเป็นภาษาไทยโดยใช้เทคนิคที่ผสมผสานกันระหว่างพจนานุกรมและ ATNs โครงสร้างประโยคที่สร้างขึ้นประกอบด้วยส่วนมูลฐานและส่วนเสริมทั้งที่อยู่ลำพัง และนำมารวมกัน ซึ่งทั้งสองส่วนนี้ประกอบด้วยกฎโครงสร้างวลี 6 กฎ และได้ประยุกต์ออโตมาตาจำกัดเชิงไม่กำหนด (Nondeterministic Finite Automata: NFA) ในการปริวรรตคำล้านนาที่ไม่พบในพจนานุกรมเป็นภาษาไทย

วิธีการวิจัย

วิธีดำเนินงานวิจัยแบ่งออกเป็น 3 ขั้นตอนคือ 1) สร้างพจนานุกรมภาษาล้านนา-ไทย 2) วิเคราะห์ประโยคล้านนา ซึ่งประกอบด้วยการตัดคำและการปริวรรตคำล้านนาที่ไม่พบในพจนานุกรมให้เป็นภาษาไทย 3) แปลประโยค ATNs ซึ่งประกอบด้วย การนำ ATNs สร้างกฎโครงสร้างประโยค และโครงสร้างวลีภาษาไทยเพื่อการแปลประโยค และการเลือกความหมายของคำโดยพิจารณาจากบริบทในประโยค ในกรณีที่ศัพท์ที่ค้นพบมีหลายความหมาย และเป็นคำประเภทเดียวกัน โดยแสดงภาพรวมของระบบการแปลภาษาล้านนาเป็นภาษาไทยในรูปที่ 1

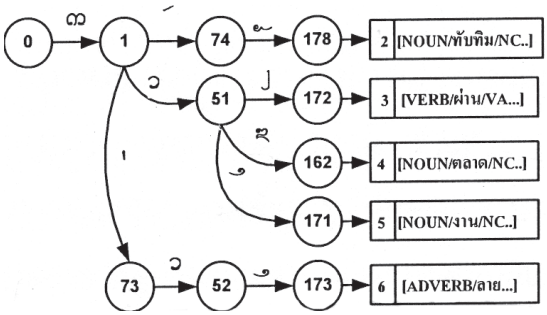


รูปที่ 1. ภาพรวมของระบบการแปลภาษาล้านนาเป็นภาษาไทย

1. การสร้างพจนานุกรมเพื่อการแปลภาษาล้านนา-ไทย

ในงานวิจัยนี้ได้ใช้พจนานุกรมเพื่อการแปลภาษาล้านนาเป็นภาษาไทยซึ่งเป็นส่วนหนึ่งของงานวิจัยการแปลภาษาล้านนาเป็นภาษาไทย (นิกร และพุธษดี, 2550) มีจำนวนคำทั้งหมด 7,497 คำ มีองค์ประกอบ 3 ส่วนคือรูปคำล้านนา เครื่องหมายเท่ากับ และส่วนวิเคราะห์ความหมายของคำ ซึ่งประกอบด้วย ประเภทคำ คำที่ถอดจากภาษาล้านนาเป็นไทย ส่วนกำกับหมวดคำ ลักษณะประจำคำ ความหมายภาษาอังกฤษ และสัทอักษร

พจนานุกรมที่สร้างขึ้น จัดเก็บในโครงสร้างข้อมูลแบบทรัย (ฐาปนี และพุธษดี, 2548, Morita et al., 2004) ตัวอย่างการจัดเก็บคำในพจนานุกรมแสดงได้ดังรูปที่ 2



รูปที่ 2. ตัวอย่างการจัดเก็บพจนานุกรมในโครงสร้างข้อมูลแบบทรัย

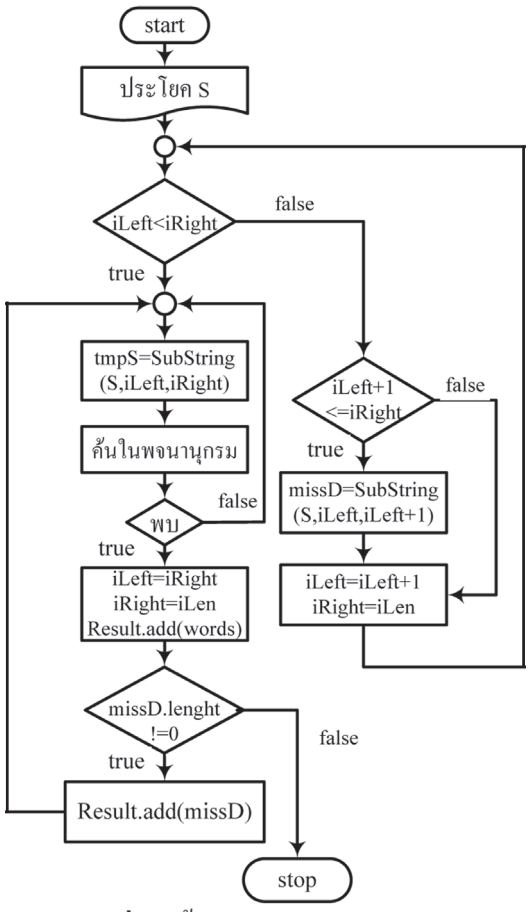
2. การวิเคราะห์ประโยคล้านนา

2.1 การตัดคำล้านนาแบบเลือกคำยาวที่สุด

เนื่องจากคำล้านนามีการประสมคำทำให้เกิดความหมายใหม่ เช่นคำ ᩁᩮᩣᨸᩣᨠ(/khon^{1}pa:k^{3}/) “ขนปาก” (หนวด) การเลือกคำยาวที่สุดจะทำให้ได้ความหมายถูกต้อง ดังนั้นงานวิจัยนี้จึงใช้วิธีการตัดคำล้านนาแบบเลือกคำยาวที่สุด (ยืน และวิวรรธน์, 2529) โดยกำหนดให้

S	เป็นประโยคที่จะทำการตัดคำ
iLeft	เป็นตำแหน่งด้านซ้ายประโยค
iLen	เป็นความยาวของประโยค
iRight	เป็นตำแหน่งด้านขวาประโยค
tmpS	เป็นอักษรที่จะค้นในพจนานุกรม
missD	เป็นตัวอักษรที่ค้นไม่พบ
result	เป็นลิสต์เก็บผลการตัดคำ

ขั้นตอนการตัดคำแสดงดังรูปที่ 3



รูปที่ 3. ขั้นตอนการตัดคำล้านนา

2.2 การใช้ NFA สร้างกฎการปริวรรตอักษรล้านนาเป็นอักษรไทย

จารึกล้านนามักจะมีคำบาลี และคำเฉพาะ เช่นชื่อคน ซึ่งไม่พบในพจนานุกรม จึงต้องมีการปริวรรตคำเหล่านี้เป็นภาษาไทย ซึ่งนิกร และพุธษดี (2550) ได้สร้างกฎการปริวรรตอักษรล้านนาเป็นไทย 226 กฎ และใช้วิธีเทียบคำที่ยาวที่สุดเพื่อตัดคำให้กับกฎที่สร้างขึ้น ปัญหาที่พบคือความกำกวมของโครงสร้างกฎ เช่น การปริวรรตคำ ᨾᩮᩯᩅ (/mɛ:ŋ^{2}waw^2/)“แมงเวา” ซึ่งมีโครงสร้างกฎเป็น VCC VCVV เมื่อใช้วิธีเทียบคำที่ยาวที่สุดจะได้โครงสร้างของคำเป็น VCCV เสมอ ทำให้โครงสร้างที่เหลือเป็น CVV ซึ่งไม่ตรงกับกฎใดๆ ทำให้การปริวรรตไม่ถูกต้อง

ในงานวิจัยนี้แก้ปัญหาโดยการประยุกต์ออโตมาตาจำกัดเชิงไม่กำหนด NFA (Mark, 2002) แทนโครงสร้างพยางค์ล้านนา เพื่อปริวรรตอักษรล้านนาเป็นอักษรไทย ซึ่ง NFA ที่สร้างขึ้นมีทั้งหมด 3 แบบ ตามลักษณะการเขียนคำในภาษาล้านนาโดยพิจารณาจากการประสมคำและการวางตำแหน่งของสระเป็นหลัก ประกอบด้วยคำที่ประสมด้วยสระในบรรทัด คำที่ประสมด้วยสระที่เขียนบนและล่างพยัญชนะ และคำที่ประสมด้วยสระประสม

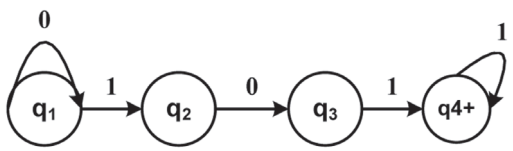
NFA สามารถแก้ปัญหาความกำกวมของโครงสร้างกฎได้ จากการปริวรรตคำล้านนาดังกล่าว NFA จะให้โครงสร้างพยางค์เป็น VCC (สระในบรรทัด) VCSV (สระประสม) ทำให้ผลการปริวรรตถูกต้อง

ออโตมาตาจำกัดเชิงไม่กำหนด (NFA) เป็นแบบจำลองทางคณิตศาสตร์ของเครื่องจักรสถานะจำกัด ประกอบด้วยสถานะ (states) และฟังก์ชั่นการเปลี่ยนสถานะ (transition function) ในรูปที่ 4 เป็นตัวอย่าง NFA โดยที่

สถานะแทนด้วยวงกลม มี 4 สถานะคือ q1, q2, q3 และ q4 โดยที่สถานะเริ่มต้น (initial states) จะไม่มีเส้นโค้งจากสถานะอื่นชี้เข้ามาในที่นี้คือ q1 และสถานะยอมรับ (accepting states) จะถูกกำกับด้วยเครื่องหมาย “+” ในที่นี้คือ q4

ฟังก์ชั่นการเปลี่ยนสถานะ แทนด้วยเส้นโค้งซึ่งชี้ไปยังสถานะต่างๆ

การเปลี่ยนสถานะหนึ่งไปยังสถานะใดๆ จะเกิดขึ้นเมื่อมีการรับค่าที่ตรงกับอักขระที่กำกับในแต่ละเส้นโค้ง ดังนั้นสายอักขระที่สามารถไปถึงสถานะยอมรับจะต้องเริ่มต้นด้วย 0 กี่ตัวก็ได้ ตามด้วย “101” และจบด้วย 1 กี่ตัวก็ได้ ตัวอย่างเช่น 01011, 0010111, 000101111 ฯลฯ แต่จะไม่ยอมรับอักขระใดๆ ที่เริ่มต้นด้วย 1 และจบด้วย 0



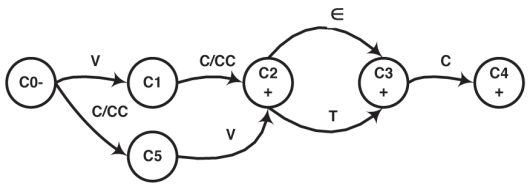
รูปที่ 4. ตัวอย่างของ NFA

การใช้ NFA สร้างกฎการปริวรรตอักษรล้านนาเป็นอักษรไทยที่ใช้ในงานวิจัยนี้ ผู้วิจัยได้ปรับปรุงมาจากกฎโครงสร้างพยางค์ล้านนา (กรรณิการ์, 2549) โดย NFA ที่สร้างขึ้นประกอบด้วยสถานะ C0 ถึง C7 สถานะเริ่มต้นคือ C0 สถานะยอมรับ มีหลายสถานะ ส่วนของอักขระ ได้ปรับปรุงโครงสร้างพยางค์ล้านนา (กรรณิการ์, 2549) ซึ่งประกอบด้วย

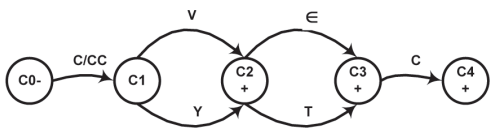
C แทนพยัญชนะต้น และชนะท้าย
CC แทนพยัญชนะควบกล้ำ
T แทนวรรณยุกต์
V แทนสระ
O แทนสระ ออ มีพยัญชนะท้าย (◌ั)
Y แทนสระ เอีย ที่เขียนแทนด้วย ◌ีย
สัญลักษณ์ ∈ คือไม่มีการรับค่าอักขระใดๆ

ฟังก์ชั่นเปลี่ยนสถานะแทนด้วยทิศทางของเส้นโค้งที่เชื่อมไปยังสถานะต่างๆ

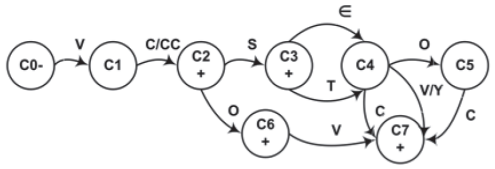
รูปที่ 5, 6 และ 7 แสดง NFA ของคำที่ประสมด้วยสระในบรรทัด คำที่ประสมด้วยสระที่เขียนบนและล่างพยัญชนะ และคำที่ประสมด้วยสระประสม ตามลำดับ



รูปที่ 5. NFA ของคำที่ประสมด้วยสระในบรรทัด



รูปที่ 6. NFA ของคำที่ประสมด้วยสระที่เขียนบนและล่างพยัญชนะ



รูปที่ 7. NFA ของคำที่ประสมด้วยสระประสม

การปริวรรตอักษร จะแปลงอักษรล้านนาให้เป็นอักขระตามกฎ และส่งอักขระเข้าไปยังเครื่องจักรทีละตัว เมื่อถึงสถานะยอมรับ จะทำการปริวรรตให้เป็นอักษรไทย ตารางที่ 1 เป็นตัวอย่างการปริวรรตคำว่า ᨠᩮᩬᨾ /lɯ:a[1]/(เหลือ)

ตารางที่ 1. การปริวรรตคำว่า ᨠᩮᩬᨾ /lɯ:a[1]/(เหลือ)

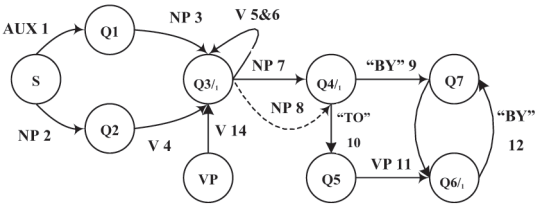
โครงสร้างกฎ / รายการ	V	C	C	S	O	C
รูปคำล้านนา	ᨠ	ᨻ	ᩬ	ᩮ	◌ั	ᨾ
กฎการปริวรรต	V	C	C	◌ั	C	
ปริวรรตเป็น	เ	ห	ล	◌ั	อ	
ผลการปริวรรต	เหลือ					

3. การแปลประโยคด้วย ข่ายงานการเปลี่ยนเพิ่มขยาย

ข่ายงานการเปลี่ยนเพิ่มขยาย (Augmented Transition Networks: ATNs) (Wood, 1970) เป็นกราฟชนิดหนึ่ง ประกอบด้วยสถานะ (state) ซึ่งแทนด้วยรูปวงกลม และการส่งผ่านค่า (transition) ซึ่งแทนด้วยเส้นโค้ง

สถานะเริ่มต้นจะไม่มีเส้นโค้งชี้เข้ามา ส่วนสถานะสุดท้ายจะกำกับด้วย “$/_1$” ในแต่ละเส้นโค้งจะมีตัวเก็บค่า (register) และหมายเลขกำกับอยู่ เมื่อส่งผ่านคำในประโยคเข้าสู่ข่ายงาน ตัวเก็บค่าจะทำหน้าที่วิเคราะห์และบันทึกบทบาทของคำตามเงื่อนไขที่สร้างขึ้น ในรูปที่ 8 เป็นตัวอย่างข่ายงานโครงสร้างประโยคภาษาอังกฤษ สัญลักษณ์ S NP และ

VP แทนประโยค นามวลี และกริยาวลี ตามลำดับ สัญลักษณ์ AUX แทนกริยาช่วย และ V แทนคำกริยา ส่วน “TO” และ “BY” นั้นจะระบุว่าคำที่รับเข้ามาจะต้องเป็น “TO” หรือ “BY” เท่านั้น สถานะเริ่มต้นคือ S และ VP สถานะสุดท้ายประกอบด้วย Q3 Q4 และ Q6 เส้นโค้งที่เป็นเส้นประหมายถึงเส้นเสมือนซึ่งจะถือว่านามวลีที่ได้นั้นเป็นประธานของประโยค และจะผ่านนามวลีดังกล่าวเข้าไปยังสถานะต่อไปได้โดยไม่ต้องมีการรับคำใดๆ เข้าสู่ข่ายงาน



รูปที่ 8. ตัวอย่างข่ายงานการเปลี่ยนเพิ่มขยาย

จากรูปที่ 8 สามารถวิเคราะห์ประโยค “John was believed to have been shot” ได้ดังนี้

(1) ส่งผ่านคำว่า “John” ซึ่งเป็นนามเฉพาะจากสถานะ S ไปยังสถานะ Q2 ตามเส้นโค้งที่ 2 จากนั้นทำการบันทึก “John” เป็นนามวลี และกำหนดให้เป็นประธานของประโยค

(2) คำที่สองคือ “was” เป็นกริยา “BE” จึงส่งผ่านเข้าสู่สถานะ Q3 ตามเส้นโค้งที่ 4 และบันทึกใน V โดยข่ายงานจะทราบว่ากริยาที่จะรับเข้ามานั้นจะต้องเติม ed ต่อท้าย

(3) คำที่สามคือ “believed” เป็นคำกริยา จึงส่งผ่านเข้าสู่สถานะ Q3 อีกครั้งตามเส้นโค้งที่ 5&6 จากนั้นทำการบันทึกเพิ่ม “believed” ใน V ซึ่งขณะนี้ V จะเก็บค่า “was believed” และข่ายงานจะกำหนดให้ V เป็นกริยาวลีหลักของประโยค

(4) คำที่สี่คือ “to”ไม่สามารถผ่านไปยังสถานะ Q4 ได้ จึงใช้นามวลีจาก NP ในเส้นโค้งที่ 2 และส่งผ่านไปยังสถานะ Q4 ตามเส้นโค้งที่ 8

(5) ส่งผ่าน “to” ไปยังสถานะ Q5 ตามเส้นโค้งที่ 11

(6) ส่งผ่านคำ “have been shot” ซึ่งเป็นกริยาวลี ไปสถานะ Q6 ตามเส้นโค้งที่ 11 แล้วจบประโยค

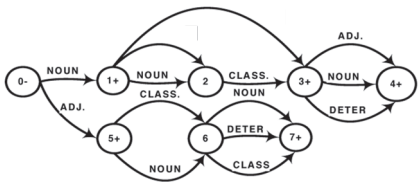
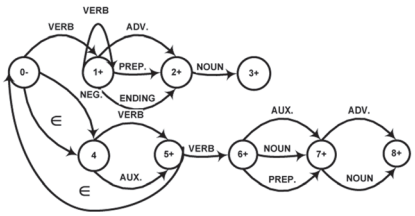
จากการศึกษาระบบโครงสร้างประโยคล้านนา พบว่าจะเหมือนกับภาษาไทยมาตรฐาน (กรรณิการ์, 2549) งานวิจัยนี้จึงได้ประยุกต์ ATNs สร้างกฎโครงสร้างวลี (เรืองเดช, 2540) และโครงสร้างประโยคภาษาไทย (วิจินตน์, 2547) ซึ่งมีรายละเอียดดังนี้

3.1 การสร้างกฎโครงสร้างวลีภาษาไทยด้วย ATNs

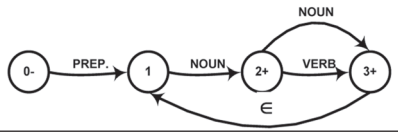
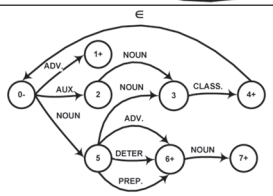
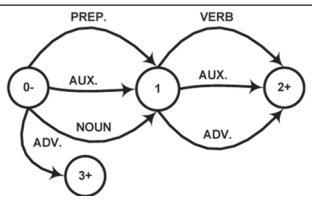
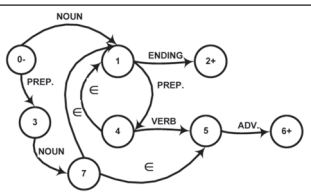
ATNs ที่สร้างขึ้นประกอบด้วยสถานะ 0 ถึง 7 การส่งผ่านค่าจะมีตัวเก็บค่ากำกับอยู่ ซึ่งตัวเก็บค่าจะแทนด้วยประเภทของคำ ได้แก่ คำนาม (NOUN) คำลักษณะนาม (CLASS) คำกำหนด (DETER) คำคุณศัพท์ (ADJ.) คำกริยา (VERB) กริยาช่วย (AUX.) คำเชื่อม (PREP.) คำวิเศษณ์ (ADV.) และคำลงท้าย (ENDING)

โครงสร้างวลีภาษาไทยประกอบด้วยสองส่วนคือ ส่วนมูลฐาน ทำหน้าที่เป็นภาคประธาน ภาคแสดง และส่วนเสริม ทำหน้าที่เป็นส่วนขยาย ในตารางที่ 2 และ 3 แสดงรายละเอียดโครงสร้างวลีภาษาไทยที่เป็นส่วนมูลฐานและส่วนเสริมที่ใช้ในงานวิจัยตามลำดับ

ตารางที่ 2. โครงสร้างวลีภาษาไทยที่เป็นส่วนมูลฐาน

วลี	ส่วนประกอบ	โครงสร้าง
(1) นามวลี (noun phrase: NP)	คำนามและหน่วยขยายซึ่งอาจเป็นคำลักษณะนาม คำกำหนด และคำคุณศัพท์	
(2) กริยาวลี (verb phrase: VP)	คำกริยาคำเดียว ซึ่งเป็นคำหลัก และคำกริยากับคำขยาย เช่น กริยาช่วย คำเชื่อม คำวิเศษณ์ และคำลงท้าย อีกกี่คำก็ได้	

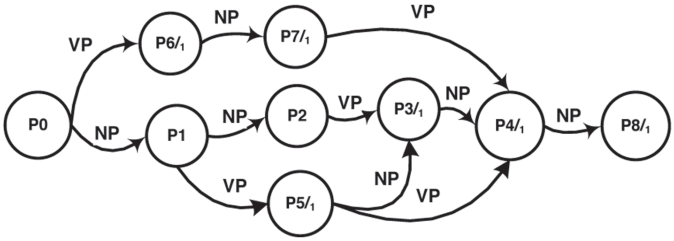
ตารางที่ 3. โครงสร้างวลีภาษาไทยที่เป็นส่วนเสริม

วลี	ส่วนประกอบ	โครงสร้าง
(1) สถานวิเศษณ์วลี	ใช้บ่งบอกสถานที่ ประกอบด้วยคำเชื่อม (ที่ บน ใน ฯลฯ) หน้าคำนามและคำกริยา	
(2) กาลวิเศษณ์วลี	ใช้บ่งบอกถึงกาลเวลา ประกอบด้วยคำนาม คำนับ คำกริยาช่วย คำเชื่อม คำวิเศษณ์ และคำลงท้าย	
(3) พิเศษวิเศษณ์วลี	เป็นหน่วยเสริมประโยค เป็นการรวมกันของคำเชื่อม คำนาม คำกริยา กริยาช่วย และคำคุณศัพท์	
(4) อาลปนะวลี	ทำหน้าที่เป็นคำเรียกขาน ประกอบด้วยคำนาม คำขยาย เช่นคำเชื่อม คำกริยา คำคุณศัพท์ และคำลงท้าย	

3.2 การสร้างกฎโครงสร้างประโยคภาษาไทยด้วย ATNs

โครงสร้างประโยคภาษาไทยประกอบด้วยส่วนมูล และส่วนเสริม ทั้งที่อยู่ลำพัง และนำมารวมกัน โดยจำแนกประโยคดังนี้

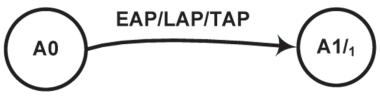
3.2.1 ส่วนมูลฐาน ทำหน้าที่เป็นหน่วยประธาน หน่วยกรรม หน่วยนามเดี่ยว (NP) และทำหน้าที่เป็นหน่วยกริยา (VP) ข่ายงานความสัมพันธ์ของส่วนมูลฐานในประโยค แสดงในรูปที่ 9



รูปที่ 9. ข่ายงานความสัมพันธ์ของส่วนมูลฐานในประโยค

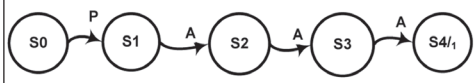
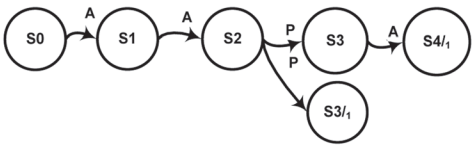
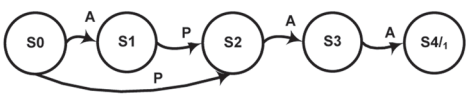
3.2.2 ส่วนเสริม สามารถเรียงลำดับโดยเริ่มต้นจากวลีใดๆ และสลับตำแหน่งได้โดยไม่ทำให้ความหมายของประโยคเปลี่ยนแปลง รูปที่ 10 เป็นข่ายงานการเรียงลำดับของส่วนเสริม โดย EAP คือ พิเศษวิเศษณ์วลี LAP คือ สถานวิเศษณ์วลี และ TAP คือกาลวิเศษณ์วลี

จากความสัมพันธ์ของส่วนมูลฐาน และการเรียงลำดับของส่วนเสริม สามารถสร้างโครงสร้างประโยคภาษาไทยด้วย ATNs ได้ทั้งหมด 4 โครงสร้าง โดยให้ P เป็นส่วนมูลฐาน และ A เป็นส่วนเสริม ดังแสดงในตารางที่ 4



รูปที่ 10. ข่ายงานการเรียงลำดับของส่วนเสริม

ตารางที่ 4. แสดงการสร้างโครงสร้างประโยคด้วย ATNs

แบบที่	ส่วนประกอบของประโยค	โครงสร้างประโยค
1	ส่วนเสริมทั้งสามชนิดปรากฏอยู่ต้นประโยค	
2	ส่วนเสริมทั้งสามชนิดอาจปรากฏอยู่ท้ายประโยค	
3	มีส่วนเสริมสองชนิดอยู่ต้นประโยค อีกชนิดอยู่ท้ายประโยคหรือมีอยู่เฉพาะต้นประโยค สองชนิด	
4	ส่วนเสริมชนิดเดียวอยู่ต้นประโยค ส่วนอีกสองชนิดอยู่ท้ายประโยค หรือมีเฉพาะท้ายประโยคสองชนิด	

4. การเลือกรูปแบบประโยคที่เหมาะสม

เป็นการค้นหารูปแบบประโยคที่เหมาะสมกับประโยคที่ทำการแปล (Mark, 2002) ใช้วิธีการผ่านคำในประโยคเข้าสู่รูปแบบประโยคซึ่งสร้างด้วย ATNs ทีละคำตามลำดับ จนครบทุกรูปแบบประโยค แล้วเลือกรูปแบบประโยคที่ยอมให้คำในประโยคผ่านเข้าไปได้มากที่สุด

5. การแปลและแจงประโยค

มีวิธีการคือส่งคำแรกในประโยคเข้าสู่ ATNs ถ้าคำนั้นตรงกับโครงสร้างประโยคในข่ายงาน ตัวเก็บค่าจะบันทึกคำแปลพร้อมกับส่งคำถัดไปเข้าสู่ข่ายงานทีละคำ เมื่อหมดประโยค ตัวเก็บค่าทุกตัวจะคืนค่าที่บันทึกไว้ตามลำดับพร้อมทั้งกำกับชนิดวลีเป็นชื่อข่ายงานนั้นๆ ในตารางที่ 5 เป็นตัวอย่างการแปลประโยคล้านนา ᩉᩣᨲᩢᩅᩣᩉᨾᩮᩴᨾᩮᨦᨾᩮᩢᨾᩥᨦᨾᩮᩢᩈᩮᩢᨶ /ta²wa:²nɔ:ŋ⁴pay²ʔɛ:w³ti:⁴ci:aŋ²sɛ:n¹/ (ตะวาน้องไปแอ่วที่เชียงแสน) โดยมีรายการคำศัพท์ดังนี้

ᩉᩣᨲᩢ = ADVERB/เมื่อวาน
ᨶᩮᩢᨦ = NOUN/น้อง
ᩅᩣᩉ = VERB/ไป
ᨾᩮᩴᨾᩮᨦ = VERB/เที่ยว, NOUN/คุตีข้าว
ᨾᩥᨦ = PREPOSITION/เที่ยว
ᨾᩮᩢᨾᩥᨦᩈᩮᩢᨶ = NOUN/เชียงแสน

โครงสร้างประโยคคือ

S ⟶ AdvP+NP+VP+AdvP

ตารางที่ 5. การแปลประโยค ᩉᩣᨲᩢᩅᩣᩉᨾᩮᩴᨾᩮᨦᨾᩮᩢᨾᩥᨦᨾᩮᩢᩈᩮᩢᨶ /ta²wa:²nɔ:ŋ⁴pay²ʔɛ:w³ti:⁴ci:aŋ²sɛ:n¹/ ลำดับประโยค ₁ᩉᩣᨲᩢ /ta²wa:²/ ₂ᨶᩮᩢᨦ /nɔ:ŋ⁴/ ₃ᩅᩣᩉ /pay²/ ₄ᨾᩮᩴᨾᩮᨦ /ʔɛ:w³/ ₅ᨾᩥᨦ /ti:⁴/ ₆ᨾᩮᩢᨾᩥᨦᩈᩮᩢᨶ /ci:aŋ²sɛ:n¹/,

ขั้นที่	ตำแหน่ง	คำศัพท์	ข่ายงาน	การกระทำ	คำอธิบาย
1	1			นำคำศัพท์แรกเข้าสู่ข่ายงาน AdvP	
2	1	ᩉᩣᨲᩢ	AdvP	เก็บคำแปลไว้ใน *AdvP* (กาล) ← "เมื่อวาน"	
3	2	ᨶᩮᩢᨦ	NP	เก็บคำแปลไว้ใน *NP* ← "พี่"	
4	3	ᩅᩣᩉ	VP	เก็บคำแปลไว้ใน *VP* ← "ไป"	ใช้โครงสร้างกริยาวลีแบบ กริยา-กริยา
5	4	ᨾᩮᩴᨾᩮᨦ	VP	เพิ่มคำแปลต่อท้าย *VP* ← "เที่ยว"	
6	5	ᨾᩥᨦ	AdvP	เก็บคำแปลไว้ใน *AdvP* (สถาน)← "ที่"	ใช้โครงสร้างสถานวิเศษณ์วลีแบบ วิเศษณ์-นาม
7	6	ᨾᩮᩢᨾᩥᨦᩈᩮᩢᨶ	AdvP	เพิ่มคำแปลต่อท้าย *AdvP* (สถาน) ← "เชียงแสน"	
8	7	-	-	คืนค่า *AdvP* (กาล) ← "เมื่อวาน", *NP* ← "น้อง", *VP* ← "ไปเที่ยว", *AdvP* (สถาน) ←"ที่เชียงแสน"	จบการทำงานเนื่องจากหมดข้อมูล

6. การเลือกความหมายที่เหมาะสม โดยพิจารณาจากบริบท และการเกิดร่วมของคำ

ศัพท์หนึ่งคำอาจมีหลายความหมายที่เป็นคำประเภทเดียวกัน เช่น คำว่า ᨾᩪᩬᩣᨶ᩠ / mu:an^{4}/ (ม่วน) เป็นคำกริยา 4 ความหมายคือ สนุก เพราะ สบาย และดีในพจนานุกรมมีข้อมูลคำเกิดร่วมให้กับศัพท์ประเภทนี้ เช่น “เพราะ” เกิดร่วมกับบริบท “ฟัง เพลง พูด” และ “สบาย” เกิดร่วมกับบริบท “นุ่ง สวม ใส่ เสื้อ ผ้า” การเลือกความหมายที่เหมาะสมจะเลือกความหมายที่มีความถี่มากที่สุดในการเกิดร่วมกับทุกบริบทในประโยค

7. การทดลองและผลการทดลอง

ข้อมูลที่นำมาทดสอบเป็นวรรณกรรมและตำราล้านนา มีหลากหลายสำนวนทั้งบทร้อยแก้ว บทร้อยกรอง และบทสวดบาลี เป็นคำล้านนาดั้งเดิม มีประโยชน์ต่อการเรียนรู้ พิมพ์ด้วยอักษรธรรมล้านนา จากการแปลพบว่า ATNs สามารถแปลประโยคได้ถูกต้อง ตัวอย่างเช่น

ᨣᩮᨿ /k^{h}ɔ:ŋ^{5}/ (ข้อง) มีสองความหมายคือ สะดุด (คำกริยา) และกระเชอ (คำนาม) ATNs สามารถแปลได้ถูกต้อง ดังนี้

(1) ᨠᩮᩢᨶ᩠ ᨣᩮᨿ ᨣᩮᨿ ᩉᩳᩬ
/k^{h}on^{2}t^{h}aw^{5}k^{h}ɔ:ŋ^{5}k^{h}ɔ:n^{1}may^{6}/
NP/ คนแก่ VP/ สะดุด NP/ ขอน NP/ ไม้

(2) ᨻᩣᨣᩮᨿ ᩉᩳᨿ ᨸᩪᩣ
/ʔaw^{2}k^{h}ɔ:ŋ^{5}pay^{2}say^{3}pa:^{1}/
VP/ เอา NP/ กระเชอใส่ Noun/ ปลา

ประเมินผลการวิจัย

ประเมินประสิทธิภาพของการแปล โดยใช้ตัวชี้วัดค่าในการทดสอบ คือจำนวนประโยคที่แปลถูกต้องต่อจำนวนประโยคทั้งหมด คิดเป็นเปอร์เซ็นต์ตามสมการที่ 1

$$\text{ประสิทธิภาพ } (\%) = \frac{X}{N} \times 100 \quad (1)$$

โดยที่ X คือจำนวนประโยคที่แปลถูก
N คือประโยคทั้งหมด

ตารางที่ 6. แสดงประสิทธิภาพของการแปลภาษาล้านนาเป็นภาษาไทย จากการแปลเอกสารล้านนา 11 เอกสารที่แบ่งประโยคแล้ว

ตารางที่ 6. ประสิทธิภาพของการแปลภาษาล้านนาเป็นภาษาไทย

ที่	ชื่อเอกสาร	จำนวนประโยค	แปลถูก	คิดเป็น%
1	นิทาน “ปูกับนกยาง”	40	31	77.50%
2	นิทาน “คนไม่รู้บุญคุณ”	29	23	79.31%
3	นิทาน “กินข้าวลำคู่คาบ”	58	51	87.93%
4	เพลงน้อยใชยยา-แว่นแก้ว	83	69	83.13%
5	คร่าวร่ำ “มาเรียนหนังสือเมืองกันเต๊อะ”	81	66	81.48%
6	ตำรายาสักกันพิษงู	41	37	90.24%
7	ตำรายาวิเศษ	11	9	81.81%
8	ยาผีเบื่อ	9	6	66.67%
9	คาถาเมตตามหานิยม	10	8	80%
10	คติสอนใจล้านนา	189	160	84.65%
11	ประโยคทดสอบ	7	7	100%
	ค่าเฉลี่ย	**558**	**467**	**83.69%**

อภิปรายและสรุปผลการวิจัย

จากผลการแปลภาษาล้านนาเป็นภาษาไทย โดยใช้เทคนิคพจนานุกรมร่วมกับข่ายงานการเปลี่ยนเพิ่มขยาย โดยพจนานุกรมล้านนา-ไทย ที่สร้างขึ้นมีคำทั้งหมด 7,497 คำ และประยุกต์ข่ายงานการเปลี่ยนเพิ่มขยาย สร้างกฎโครงสร้างประโยค 4 กฎ สร้างกฎโครงสร้างวลี 6 กฎ กำหนดความถูกต้องของการแปลอยู่ในระดับประโยค จากการทดลองแปลประโยคจำนวน 558 ประโยคพบว่ามีประสิทธิภาพในการแปล 83.69%

สาเหตุการแปลผิดเกิดจากโครงสร้างวลีและประโยคในข่ายงานไม่สอดคล้องกับโครงสร้างประโยคที่แปล เนื่องจากเอกสารทดสอบส่วนใหญ่เป็นวรรณกรรม โครงสร้างประโยคที่ปรากฏจึงไม่เป็นไปตามโครงสร้างประโยคสามัญ

ข้อเสนอแนะ

1. ควรมีการวิเคราะห์โครงสร้างวลี และประโยคล้านนาที่ปรากฏในจารึกโดยผู้เชี่ยวชาญภาษาศาสตร์ล้านนา เพื่อให้ได้โครงสร้างประโยคล้านนาที่ถูกต้อง

2. ผู้เชี่ยวชาญภาษาล้านนาควรสร้างเกณฑ์มาตรฐานในลำดับการพิมพ์คำล้านนาให้เป็นมาตรฐานเดียวกัน

เอกสารอ้างอิง

กรรณิการ์ วิมลเกษม. 2549. **ภาษาไทยถิ่นเหนือ.** พิมพ์ครั้งที่ 2. กรุงเทพฯ: ธรรมสาร.

นิกร ยาพรม และ พุธษดี ศิริแสงตระกูล. 2550. **การแปลภาษาล้านนาเป็นภาษาไทย.** The 11th National Computer Science and Engineering Conference (NCSEC). 19-21 พฤศจิกายน 2550, หน้า 180-189.

ฐาปนี เฮงสนั่นกูล และ พุธษดี ศิริแสงตระกูล. **การตัดคำโดยใช้เทคนิค Fast and Compact Updating Algorithm.** The 2nd Join Conference on Computer Science and Software Engineering 2005; 17-18 พฤศจิกายน 2548. หน้า 144-150.

ยืน ภู่วรวรรณ และ วิวรรธน์ อิ่มอารมณ์. 2529. **การแบ่งแยกพยางค์ไทยด้วยดิกชันนารี.** การประชุมวิชาการวิศวกรรมไฟฟ้า ครั้งที่ 9.

เรืองเดช ปันเขื่อนขัติย์. 2540. ภาษาศาสตร์.**ภาษาไทย.** พิมพ์ครั้งที่ 1. กรุงเทพฯ: โรงพิมพ์มหาจุฬาลงกรณ์ราชวิทยาลัย.

วิจินตน์ ภาณุพงศ์. 2547. ไวยากรณ์โครงสร้าง: ประโยคและวลี. ใน สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช.**เอกสารการสอนชุดวิชาภาษาไทย 3.** กรุงเทพฯ: มหาวิทยาลัยสุโขทัยธรรมธิราช: 127-172.

อุดม รุ่งเรืองศรี. 2524. **ระบบการเขียนอักษรลานนา.** เชียงใหม่: คณะมนุษยศาสตร์ มหาวิทยาลัยเชียงใหม่.

Watson M, 2002. **Practical Artificial Intelligence Programming in Java.** [online]. [cite 2007 July 6]. Available from: http://www.markwatson.com/ opencontent/ javaai_lic.htm.

Morita K, Tanaka, A., and Fuketa, M. 2004. Fast and compact updating algorithms of a double-array structure. **Information Sciences 2004.** 159(1-2): 53-67.

Woods, W. 1970. Transition network grammar for natural language analysis. comn. **ACM**, 10: 591-66.