



## 1.1 ความเป็นมาของปัญหา

### 1.1.1 ลักษณะของภาษาไทยและปัญหา

ในการเขียนข้อความภาษาไทยนั้น แต่ละคำในประโยคจะเขียนติดต่อกันไป โดยไม่มีช่องว่างแยกแต่ละคำเหมือนในภาษาอังกฤษ ลักษณะเช่นนี้ทำให้การประมวลผลภาษาไทยด้วยคอมพิวเตอร์ ประสบปัญหาพอสมควร เช่น การจัดหลังให้ด้านขวาของบรรทัดตรงกัน ในโปรแกรมประมวลผลคำภาษาไทย (Thai word processor) การทำดัชนี (indexing) รวมทั้งงานเกี่ยวกับการวิเคราะห์ข้อความทางภาษา (language parsing) เช่น การแปลภาษาด้วยเครื่อง (machine translation) (Keretho, 1983) หรือการประมวลผลภาษารธรรมชาติ (natural language processing) เป็นต้น

### 1.1.2 การตัดคำกับโปรแกรมประมวลผลคำ

การพัฒนาขั้นตอนวิธี (algorithm) การตัดคำภาษาไทยนั้น ได้มีการพัฒนามาเป็นเวลานานแล้ว โดยจะพัฒนาเป็นส่วนหนึ่งของโปรแกรมประมวลผลคำ ดังนั้น โปรแกรมประมวลผลคำภาษาไทยแต่ละตัว ก็จะมีขั้นตอนวิธีการตัดคำภาษาไทยเป็นของตัวเอง และฝังตัวอยู่ในโปรแกรมประมวลผลคำ

ลักษณะของการที่ขั้นตอนวิธีการตัดคำภาษาไทย ถูกมองเป็นส่วนประกอบของโปรแกรมประมวลผลคำ ทำให้เกิดปัญหา คือ

1. การพัฒนาขั้นตอนวิธีการตัดคำภาษาไทย จะเกิดขึ้นเมื่อมีการพัฒนาตัวโปรแกรมประมวลผลคำเท่านั้น ทำให้พัฒนาได้ช้า และไม่ใช่อิสระ
2. การนำขั้นตอนวิธีการตัดคำภาษาไทยไปใช้ในโปรแกรมประยุกต์ (application Program) อื่น นอกเหนือจากโปรแกรมประมวลผลคำทำได้ยาก เนื่องจากลักษณะการติดต่อกับขั้นตอนวิธีการตัดคำภาษาไทย มักจะต้องใช้โครงสร้างข้อมูล (data

structure) ที่เป็นลักษณะเฉพาะของ โปรแกรมประมวลผลคำ เช่น ต้องสงตำแหน่งของกั้นหน้าขวา (right margin) ไปให้ขั้นตอนวิธีการตัดคำภาษาไทยด้วย

## 1.2 วัตถุประสงค์ของวิทยานิพนธ์

พัฒนาส่วนเชื่อมโยง (interface) และโปรแกรมอรรถประโยชน์ (utilities) เพื่อเอื้ออำนวยต่อการตัดคำในภาษาไทย โดยมีองค์ประกอบที่จะพิจารณาคือ

1. พัฒนาส่วนเชื่อมโยงระหว่างมอดูลของขั้นตอนวิธีการตัดคำภาษาไทยกับโปรแกรมที่นำผลของการตัดคำภาษาไทยไปใช้ ในลักษณะที่มีความยืดหยุ่นสูงเพื่อที่จะสามารถนำมอดูลการตัดคำนั้นไปใช้ได้โดยสะดวก กับทุกๆ โปรแกรมประยุกต์
2. พัฒนาขั้นตอนวิธีในการแทรกอักขระแบ่งคำ (separator) หลังคำที่ถูกตัด โดยไม่ให้สับสนกับข้อมูลจริง รวมทั้งพัฒนาขั้นตอนวิธีที่จะแยกอักขระแบ่งคำออก เพื่อให้เหลือแต่ข้อมูลจริง
3. พัฒนาส่วนเชื่อมโยงระดับสูง เพื่อเสริมส่วนเชื่อมโยงที่ได้จากวัตถุประสงค์ข้อที่ 1 และข้อที่ 2 เพื่อให้สะดวกต่อการเขียน โปรแกรมประยุกต์เพื่อเรียกใช้ระบบตัดคำ

## 1.3 ขอบเขตของวิทยานิพนธ์

1. ขั้นตอนวิธีการตัดคำที่ใช้ทดลอง จะเป็นขั้นตอนวิธีการตัดคำที่ใช้ในโปรแกรม CU-writer โดยสามารถเปลี่ยนไปใช้ขั้นตอนวิธีการตัดคำอื่นได้
2. การพัฒนาส่วนเชื่อมโยงของขั้นตอนวิธีการตัดคำ จะไม่เกี่ยวข้องกับเปอร์เซ็นต์ของความถูกต้องในการตัดคำของขั้นตอนวิธีการตัดคำภาษาไทยที่นำมาใช้
3. รหัสภาษาไทยที่ใช้ จะใช้รหัสของสำนักงานมาตรฐานผลิตภัณฑ์อุตสาหกรรม (ส.ม.อ.)
4. การกำหนดส่วนเชื่อมโยงและการพัฒนาโปรแกรมอรรถประโยชน์ จะอยู่ภายใต้ระบบภาษาซี (C language) เนื่องจากมีความยืดหยุ่นสูงและเป็นภาษาที่ใช้ในการพัฒนา

โปรแกรมบรรณประโยชน์บนเครื่องไมโครคอมพิวเตอร์ และสามารถโยกย้ายไปใช้ได้กับเครื่องคอมพิวเตอร์หลาย ๆ แบบ

#### 1.4 ประโยชน์ที่คาดว่าจะได้รับจากการทำวิทยานิพนธ์

1. เป็นแนวทางให้เกิดแบบแผนซึ่งเป็นมาตรฐาน เพื่อให้ผู้ที่พัฒนาขั้นตอนวิธีการตัดคำและโปรแกรมประยุกต์ สามารถนำไปใช้ได้เป็นอย่างดีเป็นมาตรฐานเดียวกัน
2. ส่วนเชื่อมโยงและโปรแกรมสนับสนุนที่พัฒนาขึ้นสามารถนำไปทดลองใช้ในงานจริง
3. สามารถได้ผลตอบสนองจากการใช้จริง ทำให้เข้าใจกลไกและธรรมชาติของการใช้ส่วนเชื่อมโยงขั้นตอนวิธีการตัดคำ

#### 1.5 เนื้อหาของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้จะกล่าวถึงขั้นตอนในการออกแบบและพัฒนาตัวเชื่อมโยงขั้นตอนวิธีการตัดคำและการแทรกอักขระแบ่งคำ โดยแบ่งเนื้อหาออกเป็นบทดังนี้

บทที่ 1 กล่าวถึง ความเป็นมาของปัญหา วัตถุประสงค์ ขอบเขต และเนื้อหาของวิทยานิพนธ์

บทที่ 2 กล่าวถึง รูปแบบของขั้นตอนวิธีการตัดคำ และการนำไปใช้งาน

บทที่ 3 กล่าวถึง การออกแบบตัวเชื่อมโยงขั้นตอนวิธีการตัดคำ

บทที่ 4 กล่าวถึง ลักษณะของอักขระแบ่งคำ และนโยบายในการเลือกอักขระแบ่งคำ

บทที่ 5 กล่าวถึง การออกแบบขั้นตอนวิธี การแทรกและตรวจรู้อักขระแบ่งคำ

บทที่ 6 กล่าวถึง การออกแบบฟังก์ชันบรรณประโยชน์

บทที่ 7 สรุปผลการทำวิทยานิพนธ์ และข้อเสนอแนะ