

เกณฑ์เบื้องต้นในการตัดคำ โดย

สาโรจน์ บัวพันธุ์งาม; พรวิภา ไชยสมคุณ; พงษ์ศักดิ์ รัตนวงศ์ และ ชญานิน บุญสงค์ศักดิ์

1.1 คำเดี่ยว คือคำที่ประกอบด้วยหน่วยคำอิสระ 1 หน่วย อาจมี 1 พยางค์ หรือมากกว่า เช่น ทาง คน สอง กระทะ ตะกร้า กะทัดรัด นาฬิกา

1.2 คำประสม คือคำที่ประกอบด้วยหน่วยคำอิสระอย่างน้อย 2 หน่วยขึ้น กลายเป็นคำประสม 1 คำ ที่มีความหมายเฉพาะ หรือมีความหมายเปรียบเทียบ เช่น พัดลม หมูสับ หวานใจ ดอกกุหลาบ ปลาตุ๊ก ปลาช่อน ตำกระท้อน ต่างจังหวัด ยาสระสม คนขับรถ รถดับเพลิง กระจกสองหน้า ไข่มั่นอึดตัว ลอดช่องแดงไทย ไก่ย่างส้มตำ เส้นทางศึกษาธรรมชาติ

- เกณฑ์ที่ใช้ระบุคำประสม

1) คำประสมต้องมีความหมายเฉพาะซึ่งอาจคงเค้าความหมายเดิมของแต่ละหน่วยคำที่นำมาประสมกัน เช่น ผ้าเช็ดตัว น้ำแข็ง หวานเย็น รถยนต์

2) คำประสมจะแทรกคำใดระหว่างหน่วยคำที่นำมาประสมกันไม่ได้ เช่น ลูกเสือ คอห่าน

(1) วันนี้ ลูกเสือหลายคนต้องร่วมกิจกรรมเดินทางไกล (คำประสม)

(2) วันนี้ ที่สวนสัตว์เชียงใหม่มีลูกเสือ/ลูกของเสือเกิดใหม่ (ไม่ใช่คำประสม)

(3) วันนี้ พ่อเรียกช่างมาซ่อมคอห่าน (คำประสม)

(4) พ่อเลี้ยงห่านไว้หลายตัว ลูกสาวชอบจับคอห่าน/คอของห่าน (ไม่ใช่คำประสม)

(5) สิ่งที่ดีที่สุดคือคุณ “สิ่งที่ดีที่สุด” เป็นนามวลี เพราะแทรกคำว่า “ที่” ระหว่าง “สิ่ง” กับ “ที่สุด” และ “ที่ดีที่สุด” เป็นหน่วยขยายซึ่งทำหน้าที่ขยาย “สิ่ง”

(6) เส้นทางศึกษาธรรมชาติสายนี้ได้รับการพัฒนาจากหน่วยงานส่วนท้องถิ่น “เส้นทางศึกษาธรรมชาติ” เป็นคำประสม เพราะไม่สามารถแทรกคำว่า “ที่” ระหว่าง “เส้นทาง” กับ “ศึกษาธรรมชาติ”

1.3 คำประสม คือคำที่ประกอบด้วยหน่วยคำอย่างน้อย 2 หน่วย โดยต้องเป็นหน่วยคำไม่อิสระ อย่างน้อย 1 หน่วย เช่น นักเรียน การเรียน ความดี ทั้งนี้ คำประสมให้ตัดคำเป็น 1 คำ เช่น

ความกังวล = /ความกังวล/, ความกลัว = /ความกลัว/

1.4 คำซ้อน คือคำที่ประกอบด้วยหน่วยคำอย่างน้อย 2 หน่วย โดยหน่วยคำที่นำมารวมกันนั้น ต้องมีความสัมพันธ์กันทางความหมาย คือเหมือนกัน คล้ายกัน หรือตรงกันข้าม เช่น หนทาง ลู่ทาง เสาหา เต็ดขาด ป่วยไข้ ไกลเคียง ดิษฐ์ ถูกผิด พี่น้อง ลูกหลาน

นอกจากนี้ หน่วยคำที่นำมาซ้อนกันอาจจะมี ความสัมพันธ์กันทางด้านเสียง โดยอาจมีเสียงพยัญชนะเหมือนกัน หรือเสียงสระเหมือนกัน เช่น จุกจิก จัวเจียว จัดจ้าน ตกต่ำ งุนงง รูปโฉมโฉมพรรณ

1.5 คำซ้ำ คือคำที่ประกอบด้วยหน่วยคำเดียวกันซ้ำกัน ส่วนใหญ่ใช้เครื่องหมายไม้ยมก (๗) แทน การเขียนซ้ำหน่วยคำเดิม เช่น

(1) เพื่อน ๆ “เพื่อน ๆ มากันแล้ว”

(2) ไกล่ ๆ “บ้านของเขายู่ไกล ๆ ห่างสรรพสินค้า”

1.6 คำสมาส คือคำที่ประกอบด้วยหน่วยคำที่ยืมมาจากคำภาษาบาลีและสันสกฤตอย่างน้อย 2 หน่วยมารวมกัน เช่น วิศวกรรม ราชการ เศรษฐศาสตร์

1.7 คำยืม คือคำที่ยืมมาจากภาษาต่างประเทศอื่น ๆ เช่น บุฟเฟต์ คาเฟ่ (คำยืมจากภาษาฝรั่งเศส)

1.8 คำเลียนเสียงธรรมชาติ (Onomatopoeic words) เช่น พิว, ชู

1.9 คำอุทาน คือคำที่เปล่งออกมาเพื่อแสดงอารมณ์และความรู้สึกต่าง ๆ ของผู้พูด เช่น แง้ว, ฮ่า, อ้อ

1.10 ตัวอักษรพิเศษหรือสัญลักษณ์ ถ้าสามารถสื่อความหมายให้เข้าใจไปในทางเดียวกันให้ตัดคำ เช่น ฿ (THB = บาท), & (ampersand = และ), # (hashtag)

1.11 hashtag ให้ตัดทั้งชุดคำ ยกเว้น hashtag ที่มีชื่อบุคคล หรือชื่อสถานที่ ให้ติดแท็ก (tag) ชื่อเฉพาะด้วย แต่ถ้าเป็นภาษาอังกฤษให้ตัดรวมกันไม่ต้องแท็ก

- เกณฑ์ในการติดแท็กชื่อเฉพาะ (tag) และ แท็กคำประสม (compound)

ชื่อเฉพาะ หมายถึง ชื่อบุคคล สถานที่ สิ่งประดิษฐ์ และวิสามานยนามอื่น ๆ การเขียนชื่อเฉพาะอาจเขียนไม่ตรงกับที่ปรากฏในพจนานุกรม ฉบับราชบัณฑิตยสถาน ต้องรู้ได้โดยทั่วไปว่าเป็นใคร อันไหน หรือที่ไหน แบ่งการติดแท็กตามกรณีต่าง ๆ ดังนี้

กรณีที่ 1 ชื่อเฉพาะที่ไม่สามารถแยกคำออกจากกันได้ หรือเป็นที่รู้จักกันโดยทั่วไปในชุดคำเดียวกัน ให้ติดแท็กเพียงอย่างเดียว เช่น **รอยัล พารากอน ฮอลล์** (แท็กชื่อเฉพาะ) **พารากอน ซีพีเพล็กซ์** (แท็กชื่อเฉพาะ)

กรณีที่ 2 ชื่อเฉพาะที่เป็นส่วนหนึ่งของคำประสม ให้ ติดแท็กคำประสม + ติดแท็กชื่อเฉพาะ ตามลำดับ เช่น **วัดวรจรรยาวาส**

1) วัดวรจรรยาวาส ติดแท็กคำประสม [วัดวรจรรยาวาส]

2) [วัดวรจรรยาวาส] ติดแท็กชื่อเฉพาะ [วัดวรจรรยาวาส]
จังหวัดเชียงใหม่

1) จังหวัดเชียงใหม่ ติดแท็กคำประสม [จังหวัดเชียงใหม่]

2) [จังหวัดเชียงใหม่] ติดแท็กชื่อเฉพาะ [จังหวัดเชียงใหม่]

เพจแมวเต๋ยชนกิน

1) เพจแมวเต๋ยชนกิน ติดแท็กคำประสม [เพจแมวเต๋ยชนกิน]

2) [เพจแมวเต๋ยชนกิน] ติดแท็กชื่อเฉพาะ [เพจแมวเต๋ยชนกิน]

ผศ. ดร.พรวิภา ไชยสมคุณ

1) ผศ. ดร.พรวิภา ไชยสมคุณ ติดแท็กคำประสม [ผศ. ดร.พรวิภา ไชยสมคุณ]

คุณ]

2) [ผศ. ดร.พรวิภา ไชยสมคุณ] ดิดแท้กชื่อเฉพาะ

[ผศ. ดร.พรวิภา ไชยสม

กรณีที่ 3 ชื่อเฉพาะและนามสกุลของบุคคลให้ดิดแท้กชื่อเฉพาะทั้งชื่อและนามสกุล

สาโรจน์ บัวพันธุ์งาม

ดิดแท้กชื่อเฉพาะ

สาโรจน์ บัวพันธุ์

งาม

- เกณฑ์ในการแก้ไขคำ

1. คำที่เป็นภาษาอินเทอร์เน็ต เช่น 5555 แก้ไขเป็น ฮา, จิง แก้ไขเป็น จริง

2. คำย่อ เช่น มีย แก้ไขเป็น มิ.ย.

3. คำทับศัพท์ เช่น มอไซค์ แก้ไขเป็น มอเตอร์ไซค์, โปรโมชั่น แก้ไขเป็น โปรโมชัน

การแก้ไขคำทับศัพท์ให้ตรวจสอบกับคำที่ปรากฏในพจนานุกรม ฉบับราชบัณฑิตยสถาน พ.ศ. 2554 ก่อน หากคำใดไม่ปรากฏในพจนานุกรม ฉบับราชบัณฑิตยสถาน พ.ศ. 2554 ให้ยึดหลักเกณฑ์ การเขียนคำทับศัพท์ของราชบัณฑิตยสถาน

4. คำตัด เช่น คอน แก้ไขเป็น คอนเสิร์ต

5. เครื่องหมาย เช่น ตกใจสุด !!!! แก้ไขเป็น ตกใจสุด !

6. ไม่ได้คู่ใช้เพื่อให้แยกเสียงได้สะดวก เช่น ซีนิเพลกซ์ แก้ไขเป็น ซีนิเพล็กซ์

Basic Guidelines for Word Segmentation

by

*Saraj Buaphanngam; Pornwipa Chaisomkhun; Pongsak Rattanawong and
Chayanin Boonsongsak*

1.1 Single word is composed of one free morpheme that contains one or more syllables, e.g. ทาง “way”, คน “people, สอง “two” กระทะ “pan”, ตะกร้า “basket, กะทัดรัด “compact” and นาฬิกา “clock”.

1.2 Compound word requires at least two free morphemes to make it specifically or comparatively meaningful, e.g. พัดลม “fan”, หมูสับ “minced pork”, หวานใจ “sweetheart”, ดอกกุหลาบ “rose”, ปลาตุก “catfish”, ปลาช่อน “snakehead fish, ตำกระท้อน “santol salad”, ต่างจังหวัด “upcountry”, ยาสระผม “shampoo”, คนขับรถ “driver”, รถดับเพลิง “fire engine”, กระจกส่องหน้า “mirror”, ไขมันอิ่มตัว “saturated fat”, ลอดช่องแตงไทย “Lod-chong Taeng-Thai” (coconut milk-soaked rice noodle and Thai melon), ไก่ย่างส้มตำ “papaya salad and grilled chicken”, and เส้นทางศึกษาธรรมชาติ “nature trail”.

Compound Word Identification

1. A compound word must contain a specific meaning. It can represent the original meaning of each compounding morpheme, e.g. ผ้าเช็ดตัว “towel” (body-drying cloth), น้ำแข็ง “ice” (solid water), หวานเย็น “sweet and cold” (Wan-yen-cold and sweet desert served in a bowl with syrup), and รถยนต์ “automobile” (engine-car).

2. A compound word must not contain any word between two morphemes, e.g. ลูกเสือ “boy scout”, and คอห่าน “toilet”. The following sentences are examples:

(1) วันนี้ **ลูกเสือ**หลายคนต้องร่วมกิจกรรมเดินทางไกล (คำประสม)

Many *boy scouts* must attend hiking activity today. (Compound)

(2) วันนี้ ที่สวนสัตว์เชียงใหม่มี**ลูกเสือ/ลูกของเสือ**เกิดใหม่ (ไม่ใช่คำประสม)

Many *tiger cubs* were born at Chiang Mai zoo today. (Not compound)

(3) วันนี้ พ่อเรียกช่างมาซ่อม**คอห่าน** (คำประสม)

My dad called the plumber to fix the *toilet*. (Compound)

(4) พ่อเลี้ยงห่านไว้หลายตัว ลูกสาวชอบจับ**คอห่าน/คอของห่าน** (ไม่ใช่คำประสม)

The father raised many geese. The daughter likes touching a *goose's neck*.

(Not compound)

(5) *สิ่งที่ดีที่สุดคือคุณ* The best thing is you. “The best thing” is a noun phrase with a noun modifier identifying the degree of goodness. Thus, it’s not compound word.

(6) *เส้นทางศึกษาธรรมชาติ* สายนี้ได้รับการพัฒนาจากหน่วยงานท้องถิ่น “This nature trail was developed by the local authority”. (Nature trail is a compound word because it does not need a relative pronoun to describe it as “the trail which is natural”.)

1.3 Complex word is composed of at least two morphemes. The one morpheme thereof must be bound morpheme. Here are some examples: นักเรียน “student”, การเรียน “learning”, and ความดี “goodness”. The complex words should be segmented as one word, e.g. ความกังวล “worry” = /ความกังวล/ and ความกลัว “fear” = /ความกลัว/.

1.4 Duplicated compound word consists of at least two free morphemes that are either concordant, similar, or opposite to each other, e.g. หนทาง “way”, ลู่วาง “way”, เสาะหา “to seek”, เด็ดขาด “to be decisive”, ป่วยไข้ “to be sick”, ใกล้เคียง “nearby”, ดีชั่ว “good and bad”, ถูกผิด “right and wrong”, พี่น้อง “older brother and younger brother, sibling”, ลูกหลาน “child and grand-child”

Furthermore, the two morphemes might have similar consonant or vowel sound, e.g. จุกจิก “picky”, ง่วงเงีย “drowsy” จัดจ้าน “spicy”, ตกต่ำ “declining”, งุนงง “confusing”, รูปโฉมโนมพรรณ “overall appearance”.

1.5 Reduplication is any word composed of the repeating morphemes. Mostly, it is written with the repetition symbol (ๆ) instead of reduplicating the same word. The following are some examples.

(1) เพื่อน ๆ (Read as friend-friend, but it means friends)

- เพื่อน ๆ มากันแล้ว “Friends have arrived”

(2) ใกล้ ๆ (Read as nearby-nearby, but it means very close to something.)

- บ้านของเขายู่ใกล้ ๆ ห้างสรรพสินค้า “His house is close to the department store”.

1.6 Samasa words are Pali-Sanskrit compound words, e.g. วิศวกรรม (engineering), ราชการ (governmental/official), เศรษฐศาสตร์ (economy).

1.7 Loanwords are any words borrowed from other languages, e.g. บุฟเฟต์ “buffet” and คาเฟ่ “café. (borrowed from French.)

1.8 Onomatopoeic words are those created by imitating sounds. This word formation is productive in Thai, e.g. shuooow (ชู่) or fewwww (ฟิ้ว).

1.9 Interjection includes any words that the speaker utters out to express the emotion or feeling, e.g. อ้อ “oh”, and ฮ่า “huh”.

1.10 Special alphabets / symbols are used to replace words purposely to let the readers concordantly understand, e.g. ฿ (Thai Baht), & (and), and # (hashtag).

1.11 Hashtag must be counted as a whole set of words. In case of a specific name of person or place, the tag must cover a specific name. Additionally, English words must be inseparable as well. The samples are

#อร่อยไปลอง = /#อร่อยไปลอง/ #แบบสวยสุด = /#แบบสวยสุด/ #BNK48 = /#BNK48/.

The guideline for specific name tagging and compound word tagging

“Specific name” includes names of persons, locations, and other proper nouns. They could be written differently from the Royal Institute Dictionary. They, nevertheless, can be commonly understood what they specify. Specific names can be tagged under the following rules.

Rule 1: Inseparable specific names

This type of name is known as a set of words, e.g. รอยัล พารากอน ฮอลล์ Royal Paragon Hall (Specific name tag), and Paragon Cineplex (Specific name tag).

Rule 2: Separable specific names

This can be divided into specific name and another compound word. “วัดวรจรยยาวาส” (Warachanyawat Temple) is the first example.

1. Compound word tagging = วัดวรจรยยาวาส “*Warachanyawat temple*”

2. Specific name tagging = วรจรยยาวาส “*Warachanyawat*”

Secondly, “จังหวัดเชียงใหม่” (Chiang Mai province) can be tagged as below.

1. Compound word tagging = จังหวัดเชียงใหม่ *Chiang Mai province*

2. Specific name tagging = เชียงใหม่ *Chiang Mai*

Thirdly, “เพจแมวตั๋ยชวนกิน” (Toey Cat, the Foodie page)

1. Compound word tagging = เพจแมวตั๋ยชวนกิน *Toey Cat, the Foodie page*

2. Specific name tagging = แมวตั๋ยชวนกิน *Toey Cat, the Foodie*

Fourthly, “ขนมจีนเส้นป๋าย่อย” (San-Pa-Koy Rice noodle)

1. Compound word tagging = ขนมจีนเส้นป๋าย่อย *San-Pa-Koy Rice noodle*

2. Specific name tagging = เส้นป๋าย่อย *San-Pa-Koy*

Lastly, ผศ. ดร.พรวิภา ไชยสมคุณ (Assistant Professor Pornwipa Chaisomkun, Ph.D.)

1. Compound word tagging = ผศ. ดร.พรวิภา ไชยสมคุณ *Assistant Professor Pornwipa Chaisomkun, Ph.D.*

2. Specific name tagging = พรวิภา ไชยสมคุณ *Pornwipa Chaisomkun*

Rule 3: Person name

Full names must be specifically tagged together, e.g.

สาโรจน์ บัวพันธุ้งาม (Saroj Buaphanngam).

Word editing criteria

1. Words used on the internet (Netspeak), e.g. 5555 → ฮา (Haha), จี๋ง → จี๋ริง (Real!)

2. Abbreviations, e.g. มิย → มิ.ย. (June)

3. Transliterated loanwords, e.g. มอไซด์ → มอเตอร์ไซด์ (motorbike), โปรโมชั่น → โปรโมชัน (Promotion).

4. Shortened words, e.g. คอน → คอนเสิร์ต (concert)

*Royal Institute Dictionary B.E. 2554 must be used in word editing. To write any word absent from the dictionary, transliteration guideline of the Royal Society must be applied.

5. Symbols e.g. ตกใจสุด !!!! → ! (Be frightened)

6. The short vowel symbol (ไม้ไต่คู้) should be added to precisely identify the tone and vowel, e.g. ซีนีเพลกซ์ → ซีนีเพล็กซ์ (Cineplex)