

1 *Introduction*

1.1. Geographical and historical contexts

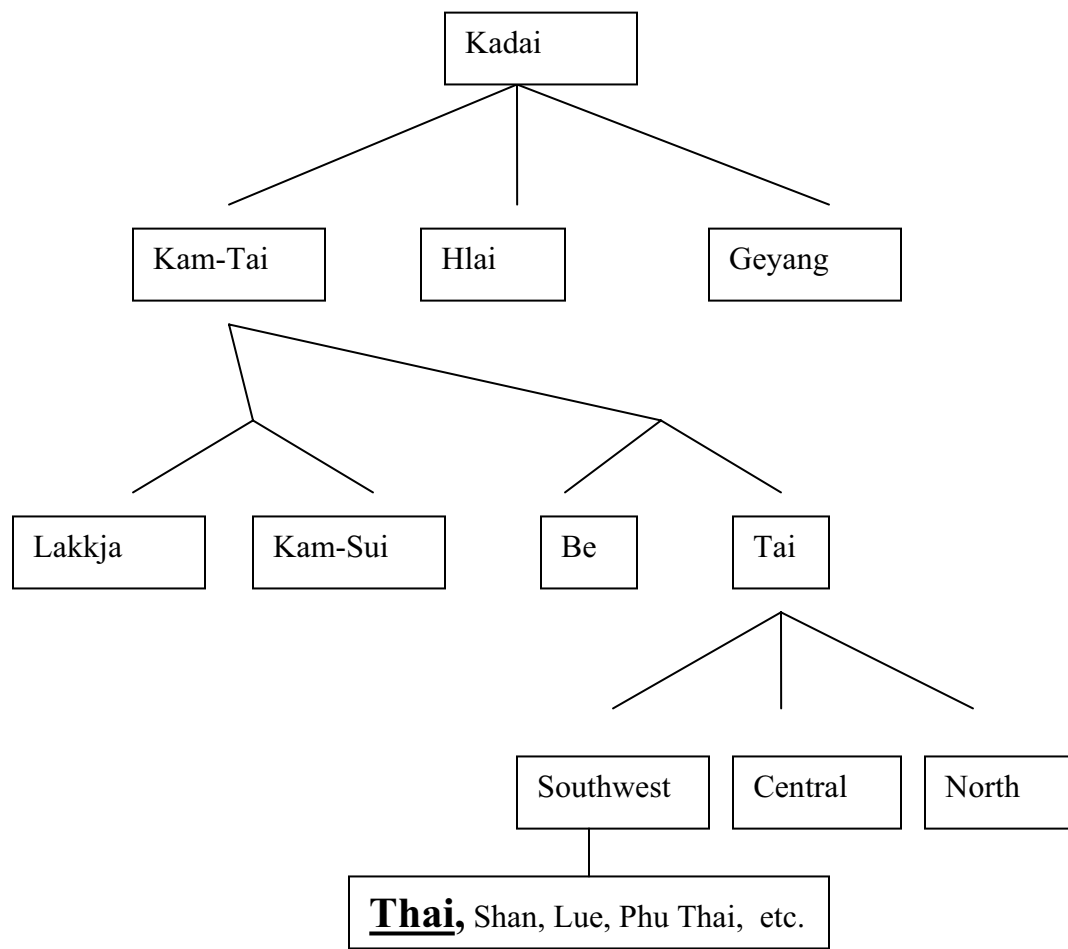
Thailand, a major Southeast Asian country, is situated in the middle of mainland Southeast Asia with a population of 60.9 million as of the year 2000 (National Statistical Office). It borders Laos to the east and northeast, Cambodia to the southeast, Myanmar to the west and northwest, and Malaysia to the south. The total area is 513,000 square kilometers (198,000 square miles). Bangkok, the country's capital, is the largest city with a population of 6.3 million.

'Thai' used in this book refers to the 'Standard Thai,' or the language used by the educated, middle class population in the central region of Thailand, particularly in Bangkok. Although Standard Thai is the native language of only about 19.5% of the population (Smalley 1994:367), it is used and understood widely throughout the country due to its use at school and in the mass media. The Central dialect, to which Standard Thai belongs, is one of the four major dialects and has 20 to 25 million speakers. The other major dialects are the Northeastern dialect (Isarn or Lao) with about 23 million speakers, the Northern dialect (Kam Muang, Lan Na or Yuan) with 6 million speakers, and the Southern dialect with 5 million speakers (Summer Institute of Linguistics Website).

Besides these major dialects, a number of related languages are spoken in the country. Almost all of these languages belong to the Southwestern branch of Tai (Li 1977). Speakers of these Tai languages comprise about 90% of the whole population. Non-Tai languages include Khmer (Austro-Asiatic), Malay (Austronesian), Karen (Tibeto-Burman), and Hmong (Hmong-Mien). The 2000 census reports that 2.3% of the population speaks Khmer and 2.3 % Malay. In addition, in cities and towns, southern varieties of Chinese such as Teochiu (or Swatow) and Hakka are commonly found.

Languages of the Southwestern branch of Tai are found not only in Thailand but also in Laos, northern Vietnam, Myanmar, India, and also in southern China. The other two branches are the Central and Northern branches. The Central branch includes languages spoken in northern Vietnam and southern China (e.g. Nùng, Tày). The Northern branch includes other languages of southern China (e.g. N. Zhuang, Bouyei, Seak). The Tai language family with these three sub-groups is related to other sister and parent branches to make up a larger language stock called Kadai or Tai-Kadai. However, the exact relationship within these branches is still being worked out. The following illustration is one proposed by Edmondson and Solnit (1997).

2 Introduction



Kadai language family

The current geographical spread of Tai languages is a result of the southwestward movement of the original Tai people from what is believed to be the Guizhou and Guangxi areas in southern China, just northeast of present-day Vietnam. The migration began during the first few centuries of the Christian era. Between the 5th to 8th centuries, the ancestors of the speakers of the Southwestern group of the Tai language family settled near the Black River, which runs from southern China into northern Vietnam, and northeastern Laos. From this location, this group of people spread out in a fan-shaped fashion. One group went northwestward as far as Assam in the northeastern corner of India. Some moved along the Mekong River, and settled in Laos and northern Thailand. Yet others went further south and settled in northeastern Thailand. By the 13th century, Tais had already settled in many parts of present-day Thailand (Wyatt 1982:6).

As the once-powerful Khmer empire of Ankor and Burman empire of Pagan started to decline in the 13th century, Tais began to gain power in the region. Lan Na and Payao in the north, and Sukhothai in the central region are important Tai kingdoms in this period. Among them, Sukhothai is the most significant not only because

it marked the beginning of the official history of Thailand, but also because it marked the beginning of the written record using the Thai script. King Ramkhamhaeng (1275-1317), the third king of Sukhothai, is said to have created the Thai writing system, and he is believed to have inscribed the first historical record on a stone monument dated in 1292. It is generally held that the Sukhothai dialect is the ancestor of present-day Thai (cf. Hudak 1987:798). After the Ayutthaya Period (1351-1767) and a brief Thonburi period, the current Bangkok period started in 1782 with King Rama I.

The writing system of King Ramkhamhaeng was modeled after the Khmer writing system based on the system originated in India. Unlike the modern version, it places all the vowel symbols on the same line as the consonant symbols, but by the middle of the 14th century, some vowels were already placed either above or under a consonant letter, as in the modern version. By the late 17th century, the system added more consonant and vowel letters. There were only two tone marks (\square and ') in the Ramkhamhaeng system, but in the late 18th century the remaining two tone marks were added.

Thai has several important typological characteristics. First, it is a tone language with five distinct tones (Section 1.2.4 below). Second, it is an isolating language. This type of language, unlike familiar European languages, lacks inflectional morphology to code grammatical information such as number and gender for nouns and verbs. Third, it is a classifier language, which uses different classifiers for counting different types of objects (Chapter 5). Other features are further discussed in Sections 1.6 and 1.7 below.

1.2. Phonological structure

1.2.1. Syllable structure

The basic syllable structure is as follows. [C = consonant including glides, V = vowel, T = tone; elements in parentheses are optional.]

$$(C) (C) \mathbf{V}^T (V) (C)$$

The only obligatory elements in a syllable are a vowel and a tone. A vowel may be short, V, or long, V_iV_i, or a diphthong, V_iV_j. V_j in a diphthong must be /a/. (The subscript indicates the identity of the vowel; V_iV_i means that the two vowels are identical, while V_iV_j means that they are not. This notation is also used for the discussion of consonants immediately below.) There may be an initial single consonant, C, or consonant cluster, C_iC_j. C_i is restricted to some stop consonants (see below) and C_j is /l/, /r/, or /w/. A syllable may have a final consonant (a stop, nasal or a glide).

The traditional morphological description divides syllable types between ‘live’ syllables (/kham pen/) and ‘dead’ syllables (/kham taay/). Dead syllables are those syllables ending with a stop consonant, /p, t, k, ʔ/. Live syllables have all other types of syllable endings (i.e. long vowels, glides, and nasal). This distinction becomes crucial for tone specification (below).

4 Introduction

1.2.2. Consonants and glides

There are 21 simple consonants including 2 glides. (See the first table below). All these consonants and glides can appear as an initial consonant. The two glides and seven consonants can also appear as a final consonant. (See the second table below).

	Labial	Alveolar	Palatal	Velar	Other
Stop					
voiced	/ b /	/ d /			
vl. unaspirated	/ p /	/ t /	/ c /	/ k /	/ ʔ /
vl. aspirated	/ ph /	/ th /	/ ch /	/ kh /	
Fricative					
voiceless	/ f /	/ s /			/ h /
Nasal	/ m /	/ n /		/ ŋ /	
Liquid		/ l /			/ r /
Glide	/ w /		/ y /		

Inventory of Consonants and Glides

Stop	/ -p /	/ -t /	/ -k /	/ ʔ /
Nasal	/ -m /	/ -n /	/ -ŋ /	
Glide	/ -w /		/ -y /	

Inventory of Final Consonants and Glides

Glottal stop occurrence is largely predictable; (i) it appears when no initial consonant is present ([ʔaacaan] ‘teacher’), and (ii) when no final consonant appears after a short vowel ([tóʔ] ‘table’). In these cases, a glottal stop is not indicated in the transcript. The presence or absence of a glottal stop is not predictable after a diphthong, e.g. /iaʔ/ vs. /ia/, and such unpredictable glottal stops are marked in the transcript. (See more comments on the glottal stop in Section 1.2.5 below.) The lateral liquid /l/ and the trill or tap /r/ are distinguished in the orthography and pronounced differently in very carefully articulated speech. However, in most speech /r/ is pronounced /l/. Some voiceless stops may form initial consonant clusters with a liquid, /l/ or /r/, or the bilabial glide, /w/. The possible 11 clusters are shown in the table on the next page. (/thr-/ is possible, but very rare. /fr/ and /fl/ also appear in some recent loan words. These are, however, not included here.) The initial clusters with /l/ and /r/ tend

to be pronounced as a single consonant without them (cluster simplification). Thus, /plaa/ ‘fish’ is usually pronounced /paa/, and /khray/ ‘who’ as /khay/.

/ pr- /	/ tr- /	/ kr- /	/ kw- /
/ pl- /		/ kl- /	
/ phr- /		/ khr- /	/ khw- /
/ phl- /		/ khl- /	

Inventory of Initial Consonant Clusters

1.2.3. Vowels

There are nine short vowels, as shown in the table below.

	Front	Central	Back
High	/ i /	/ ʊ /	/ u /
Mid	/ e /	/ ɤ /	/ o /
Low	/ ɛ /	/ a /	/ ɔ /

Inventory of Vowels

Each short vowel has a contrasting long vowel, such as /ii/ and /aa/. In addition, there are three diphthongs, consisting of a high vowel followed by /a/, as shown below. There are both short and long versions, and the short version is specified with the glottal stop; /iaʔ/, /ʉaʔ/ and /uaʔ/.

/iaʔ/	/ʉaʔ/	/uaʔ/
/ia/	/ʉa/	/ua/

Inventory of Diphthongs

1.2.4. Tones

There are five phonologically contrasting tones.

1 st tone: Mid	(no marking)	a	(ก้า / khaa / ‘to be unsettled’)
2 nd tone: Low	˘	à	(ก้า / khàa / ‘galangal root’)
3 rd tone: Falling	ˆ	â	(ก้า / khâa / ‘to kill’)
4 th tone: High	ˊ	á	(ก้า / kháa / ‘to trade’)
5 th tone: Rising	ˋ	ǎ	(ก้า / khǎa / ‘leg’)

Inventory of Tones

6 Introduction

In addition, an emphatic high tone (or sixth tone) may appear on emphasized words. In the transcription adopted in this book, tone marks are placed on the (first) vowel. No indication of emphatic high tone will be provided, except when such an indication is necessary for a discussion (e.g. Chapter 2.2.2.2).

1.2.5. Glottal stop deletion and tone neutralization

The glottal stop is deleted when a short vowel with no final consonant appears as the first element of a compound word. This is often accompanied by a neutralization of tone especially when the original tone is low (/bà[mǐi/ > [bamǐi] ‘egg noodle’). The word /sà[àat/ ‘clean’ is normally pronounced [sa[àat], after the glottal stop associated with the initial /sà[is deleted and the tone neutralized. The glottal stop associated with /[àat/, however, is still present. This type of word is transcribed in this book as, for example, [sa[àat] not [saàat]. Similarly, /prà[yòot/ ‘value’ is pronounced and transcribed as /pra[yòot/ or /pa[yòot/ after consonant-cluster simplification.

Tone neutralization sometimes does not occur on a high-tone syllable after glottal stop deletion (e.g. /wí[thii/ > [wíthii] ‘method’) (Li 1977:8). Also, glottal stop deletion and tone neutralization do not occur on reduplicated words: [y[yé[] ‘a lot’ and [kè[kà[] ‘to be in the way’ (Peyasantiwong 1986:222). In this book, glottal stops are in general not indicated except where non-indication would cause confusion.

1.2.6. Stress and syllable duration

A monosyllabic word with a short vowel may be pronounced with a longer duration when it is emphasized and when it appears at the end of a sentence; e.g. /dây/ ‘can’ is often pronounced [dây] at the end of an utterance. When a word consists of two syllables, the first syllable tends to be pronounced with shorter-than-normal duration and the second with longer-than-normal duration. For example, [pàak] ‘mouth’ is pronounced with long vowel duration in isolation, but it is pronounced with short duration in the compound [pàk kaa] ‘pen < mouth + crow.’ Also when /nám/ ‘water,’ which has a short vowel, appears as the first element in a compound, it is pronounced with shorter duration than normal, [nám khěŋ] ‘ice.’ When /nám/ appears as the second element, however, it is pronounced with long duration, e.g. [hōŋ nám] ‘toilet.’ This pattern applies to many compounds and reduplicated words, but there are also some exceptions (Peyasantiwong 1986). There is another type of vowel duration alternation, which is not predictable. This is idiosyncratic pronunciation of an orthographic long vowel as a short vowel. The word for ‘to play’ is written with a long vowel เล่น /lêen/, but pronounced with a short vowel [lēn] (Dhananjayananda 1997).

1.3. Writing system

The traditional chart of Thai letters is known as /kɔɔ kày khǎw khày/ (กไก่ไข่). This name comes from the two initial letters in the chart: the first letter ก is pronounced /kɔɔ/ in isolation and is found in the initial letter of the word ไก่ /kày/ ‘chicken’; and the second letter ข is pronounced /khǎw/ in isolation and is found in the initial letter of ไข่ /khày/ ‘eggs.’ As the above suggests, consonant letters in isolation are pronounced with the vowel /ɔɔ/.

The /kɔɔ kày khǎw khày/ chart currently lists 44 consonant letters. However, two are completely obsolete, and Modern Thai employs only 42 letters. The reason why there are more letters than the actual number of consonants (21 consonants including 2 glides) is because many consonants are represented by more than one letter; there are six different letters for the consonant /th/ (ก ข ฃ ค ฅ ฆ), and four different letters for the consonant /s/ (ซ ณ ด ถ), and so forth. The 42 consonant letters are arranged according to the traditional three-part classification, Mid, High, and Low consonants in the table shown at the end of this chapter (p. 22). This classification is crucial for tone specification (below).

Mid-consonant Letters (9 letters): (i) All voiced consonant letters (บ ป ฝ ฎ, (ii) all unaspirated consonant letters (จ ฎ ฌ ฎ), and (iii) the letter อ, which is the vowel /ɔɔ/, but classified as a Mid-consonant letter.

High-consonant Letters (10 letters): The first set of aspirated consonant letters (ค ฆ ฃ ฅ จ ช ด ถ น บ).

Low-consonant Letters (23 letters): (i) The second set of aspirated consonant letters (ท ฐ ฑ ฒ ณ ด ถ น บ), and (ii) sonorant consonant letters for the sounds / m ɲ n ɳ r ʎ (ม น ญ ณ ด ถ น บ), and (iii) glides (ว ย ฌ).

The /kɔɔ kày khǎw khày/ chart lists 32 vowel letters, including a few rarely used ones (see the table at the end of this chapter on p. 23). Some letters are simplex letters consisting of one symbol, but others are complex letters consisting of more than one symbol. All long vowels except /a:/ are represented by simplex letters. In addition, there are symbols (mostly complex) for diphthongs and /ay, aw, am/. The writing system also includes four tone markers and special symbols for abbreviation among others. Vowel symbols and these other symbols are presented at the end of this chapter (p. 23).

The Thai writing system is extremely efficient in terms of specifying the tone associated with a syllable. There are four determining factors for tones; (a) initial consonant’s class (‘High’, ‘Mid’ or ‘Low’); (b) syllable type (‘Live’ or ‘Dead’); (c) vowel length (Long or Short); and (d) tone mark. The following table taken from Tomita (1990:25) shows all the possible tone patterns. Only live syllables with an initial Mid-consonant letter display all five tone patterns. Live syllables with an initial High- or Low-consonant letter display three tone patterns each. The association between a tone mark and an actual tone pattern is different depending on the class of the initial consonant. The tone of a dead syllable with a Mid or High consonant is always

8 Introduction

low (the 2nd tone), but that with a Low consonant is either high (the 4th tone) when the vowel is short or falling (the 3rd tone) when the vowel is long.

Live Syllable	Tone mark	Mid Consonant	High Consonant	Low Consonant
	(none)	กา kaa 1	ขา khǎa 5	คา khaa 1
	'	ก่า kàa 2	ข่า khàa 2	ค่า khâa 3
	ˊ	ก้า kâa 3	ข้า khâa 3	ค้ำ kháa 4
	ˋ	ก้า káa 4		
	◻	ก้า kǎa 5		

Dead Syllable	Vowel type	Mid Consonant	High Consonant	Low Consonant
	Short vowel	กัต kàt 2	ขัต khàt 2	คัต khát 4
	Long vowel	กาด kàat 2	ขาด khàat 2	กาด khâat 3

Consonant Class and Tone Patterns

1.4. Word classes

Categorizing Thai words into the familiar word classes, or parts-of-speech, is not always easy. First, the word-class classification established in the West may not be completely relevant for non-Western languages. Second, distinguishing words on morphological grounds is impossible in an isolating language like Thai. Finally, Thai words are often very fluid, and many words are still undergoing the process of ‘grammaticalization,’ a process in which content words such as nouns and verbs become function words like prepositions and auxiliary verbs. This means some words have to be listed in two or more different categories. No complete agreement on how to classify words has been reached among linguists working on this language (cf. Noss 1964, Panupong 1970, Yoshikawa 1987, Pankhuenkhat 1998).

Notwithstanding the difficulty, it is still possible and important to establish several distinct word classes according to structural, semantic, and functional criteria. In this book 14 different word classes are recognized. These classes may be arranged in the following four groups.

- I. Noun-related words:

(1) Noun(2) Pronoun(3) Demonstrative(4) Preposition(5) Classifier(6) Numeral
- II. Verb-related words:

(7) Verb(8) Auxiliary verb(9) Negator
- III. Modifying words:

(10) Adjective(11) Adverb

IV. Miscellaneous words:

(12) Linker	(13) Particle
(14) Exclamative	

Some of these classes are described in detail in separate chapters. In such cases, chapter numbers are given for further information, and their description here is kept minimal.

1.4.1. Noun-related words

(1) Nouns: There are both simple nouns and compound nouns (Chapter 2). They can be a subject, an object of a verb or preposition, or a predicate nominal after the copula /pen/ and /khəu/. They can be followed by various modifiers such as adjectives, demonstratives, numerals, genitive (possessive) phrases, and relative clauses (Chapter 4).

(2) Pronouns (Chapters 3 and 30): There are a great number of pronouns in Thai. Speakers will choose the most appropriate one according to sociolinguistic criteria. Structurally, pronouns are similar to nouns in some respects; both can be a subject or an object of a verb or preposition. Also, both can be modified by a demonstrative (Sornhiran 1978:67): /aacaan khon nán/ (teacher – CLS – that) ‘that teacher’ and /kháw khon nán/ (he – CLS – that) ‘that he (person).’ However, while nouns can be modified by adjectives, numeral phrases, genitive phrases, and relative clauses, pronouns cannot. Also pronouns cannot appear after the copulas, /pen/ and /khǔu/.

(3) Demonstratives (Chapter 6): This class includes words like the English ‘this’ and ‘that.’ Thai demonstratives are divided between the proximate /nǐi/ (close to the speaker), medial /nán/ (away from the speaker), and distal /nóon/ (farther away from the speaker). They are further distinguished between the modifier and pronominal forms.

(4) **Prepositions:** Prepositions appear directly before a noun. Some prepositions, e.g. /hây/ ‘for’ (< ‘to give’) and /aw/ ‘with’ (< ‘to take’), are grammaticalized prepositions from verbs (see Chapter 8 for prepositions used in adverbial phrases).

(5) Classifier (Chapter 5): Classifiers are a special category used when nouns are counted and modified.

(6) Numerals: Thai has borrowed the Chinese number system. The numbers from one to ten are as follows. ‘Zero’ is /sǔun/, which is an Indic loan.

{ 1 นึ่ง /nùŋ/ 2 สอง /sǎwŋ/ } [3 สาม /sǎam/ 4 สี่ /sǐi/ 5 ห้า /hâa/
6 หก /hòk/ 7 เจ็ด /cèt/ 8 แปด /pèet/ 9 เก้า /kâw/ 10 สิบ /sǐp/]

10 Introduction

Numbers 3 through 10 come from the same set of numerals in Chinese (Set A), while numbers 1 (/nəŋ/) and 2 (/sǎŋ/) come from a different set (Set B). Number 1 in Set A is /èt/, and it appears in the one's place in numbers greater than 10; e.g. 11 /sǐp èt/, 51 /hâa sǐp èt/. Number 2 in Set A is /yǐi/, and it appears in the ten's place of numbers in the 20's; e.g. 20 /yǐi sǐp/, 21 /yǐi sǐp èt/, 22 /yǐi sǐp sǎŋ/. Among large numbers, only 10,000 /mùun/ is a Chinese loan. The etymology for 100 /rǎŋ/, 1000 /phan/, 100,000 /sǎen/, and 1,000,000 /láan/ is not clear (Tomita 1990:1489).

1.4.2. Verb-related words

(7) Verbs (Chapter 9): There are intransitive, transitive, and ditransitive verbs. Adjectives when used as a predicate are considered stative (intransitive) verbs.

(8) Auxiliary verbs (Chapters 11, 12, 27, 28, 29): Auxiliary verbs supply additional information for the main verb, such as aspect, modality, direction, and potentiality.

(9) Negators: Negators include /mây/, /mây-chây/, /yàa/, and /plàaw/. The negator /mây/ is a general negator which appears before a verb, adjective, or certain auxiliary verbs, e.g. /mây pay/ 'not go' and /mây tŋ/ 'not have to.' On the other hand, /mây-chây/ negates a nominal predicate, e.g. /mây-chây khon thai/ 'not a Thai person.' /yàa/ is placed before a verb to form a negative imperative, e.g. /yàa pay/ 'Don't go.' /plàaw/, originally an adjective 'blank,' is used as 'No' in a negative answer to negate a whole proposition for certain questions (Chapter 23).

1.4.3. Modifying words

(10) Adjectives (Chapter 7): Adjectives are used either as a modifier or a predicate. Predicate adjectives are normally treated as stative verbs for they do not require a copula as in English. However, adjectives differ from other stative verbs in that they can appear in comparative and superlative constructions. Some adjectives can function as manner adverbs.

(11) Adverbs (Chapter 8): Besides manner adverbs derived from adjectives, there are also genuine adverbs. These adverbs tend to modify a whole sentence rather than a verb alone, e.g. /sam-sam/ 'always' and /baan-thii/ 'sometimes.'

1.4.4. Miscellaneous words

(12) Linkers: The linker (or conjunction) consists of several sub-types. The lexical linkers connect individual words: /kâp/ 'and,' /kà/ 'and,' /lé/ 'and,' /rǎu/ 'or.' The clause linkers combine clauses and sentences: /léew (kô)/ 'and,' /rǎu/ 'or,' /tèe/ 'but.' The discourse linkers (or discourse markers) combine (sets of) propositions in discourse: /tèe-wâa/ 'but,' /tèe-kô/ 'but.' Some linkers appear as a set: /phró cŋ cŋ 'because....,' /thâa kô kŋf ...then...' (Chapter 22). There are also more formal discourse linkers such as /phróchanán/ 'therefore.'