Malayic varieties of Kelantan and Terengganu: description and linguistic history

Wu, J.

CHAPTER 1

Introduction

1.1 Overview

This dissertation examines the Malayic varieties spoken in Kelantan and Terengganu, two Malaysian states located on the northeast coast of the Malay Peninsula. It focuses on three varieties, namely Kelantan Malay, Coastal Terengganu Malay and Inland Terengganu Malay, all belonging to the Malayic subgroup within the Austronesian language family. The area where these varieties are spoken is indicated in the map in Figure 1.1. The primary objectives of this study are twofold: first, to provide a synchronic description of these languages, and second, to offer a historical account of their development, which could shed light on the migration history of the speakers.

The following abbreviations will be used throughout the dissertation: KM for Kelantan Malay, CTM for Coastal Terengganu Malay and ITM for Inland Terengganu Malay. For ease of reference, the three varieties are also collectively referred to as Northeastern Peninsular Malayic varieties (henceforth NEPMs). The term "variety" is chosen to avoid the fuzzy distinction between "language" and "dialect". As will be discussed in more detail in §1.2, there is no clear differentiation between "non-Malay Malayic

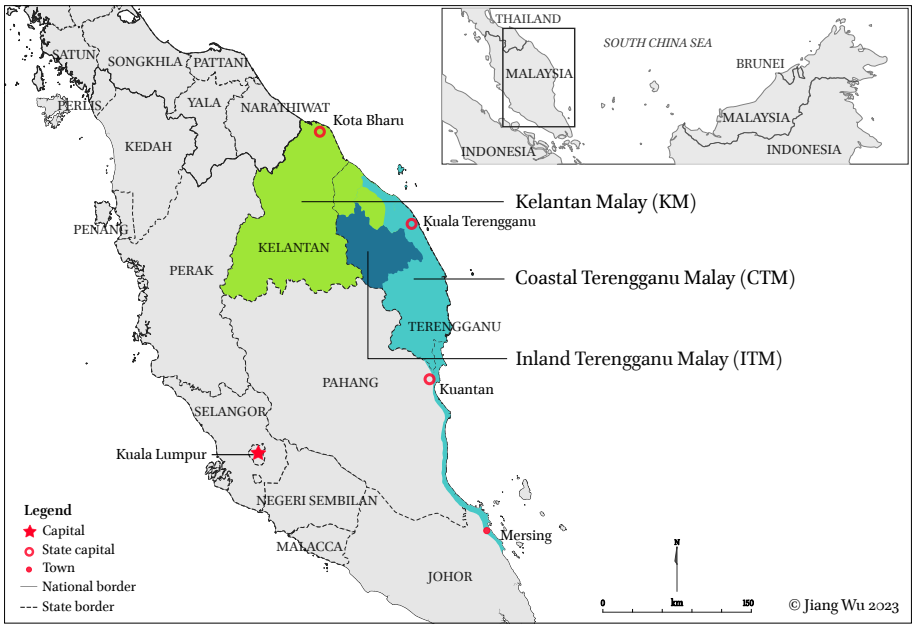

Figure 1.1: Malayic varieties in Kelantan and Terengganu

languages" and "Malay dialects" within the Malayic group. While the labels "non-Malay" and "Malay" can refer to ethnic groups, often based on religious and cultural practices, and sometimes political considerations, the distinction does not necessarily apply to the languages spoken by these ethnic groups. In Malaysia, the speech forms of ethnic Malays are considered dialects of Malay. However, as has been noted repeatedly, some of these "Malay dialects" are distinct enough to be unintelligible to speakers of Standard Malay (henceforth SM) (e.g., Blust 1988; Adelaar 2004b, 2018); those spoken in the northern states such as Kelantan and Terengganu are prime examples. It is likely that NEPMs should be considered separate languages in their own right, and for this reason, they are referred to as "Malayic varieties" rather than "Malay varieties".

In the field of Malayic and Austronesian linguistics, NEPMs, especially KM, are widely recognised for their unique structural features. They have attracted an extensive amount of scholarly interests since the late 19th century, and most Malaysian linguists from Kelantan and Terengganu have written about their own speech varieties. Nonetheless, despite the abundance

of literature that has provided valuable insights, many issues have not been fully addressed, and there remain a number of reasons why NEPMs deserve further investigation in this dissertation (see §1.5 for a more comprehensive appreciation of previous studies).

First of all, previous studies were typically engaged in the comparison between NEPMs and SM, rather than treating NEPMs as linguistic entities on their own. More importantly, focus was often given to the sound system and lexicon alone, with little attention paid to the morphosyntactic aspects. One objective of the present study is to provide a modern linguistic description of NEPMs by adopting a structural approach, covering both their phonology and basic morphosyntax.

Second, the study of NEPMs holds a significant place in Malayic historical linguistics. The Malay Peninsula is generally viewed as a late settlement of the Malayic-speaking people following their migration from the homeland in West Borneo via Sumatra (Blust 1985; Adelaar 2004b). This suggests that Peninsular Malayic varieties have a relatively short history, and they are often considered offshoots of court Malay as documented from the fifteenth century (from which SM is a direct descendant). Contrary to expectations, however, NEPMs exhibit some noteworthy retentions that are not present in other Peninsular Malayic varieties, as previously noted by Collins (1983a) for ITM. The second aim of this study is therefore to establish the diachronic development of NEPMs from Proto Malayic (henceforth PM), which can contribute to a more fine-grained internal classification of the Malayic subgroup and a deeper understanding of the Malayic migration history.

Third, the history of NEPMs is interesting from the perspective of contact linguistics. Along socio-historical lines, NEPMs are categorised as vernacular or "inherited" Malayic varieties (Adelaar & Prentice 1996). Nevertheless, NEPMs share certain similarities with contact varieties or so-called "Pidgin Derived Malay", as will be demonstrated in the current description. The region where present-day NEPMs are spoken is indeed a contact zone, with Aslian languages spoken in the inland of the Malay Peninsula and historical presence of Mon-Khmer languages, both groups belonging to the Austro-Asiatic (henceforth AA) family. Some earlier studies have posited that the peculiarities of NEPMs might be attributed to an AA substratum (e.g., Winstedt 1923: 96; Benjamin 1987, 1997). This hypothesis will be tested in this dissertation.

Finally, the documentation of NEPMs, ITM in particular, is of utmost importance in view of their language vitality and endangerment. ITM is spoken by only approximately 50,000 to 70,000 people in the inland area of Terengganu, and it is not being passed down to younger generations who tend to switch to the more prestigious CTM, which is the *de facto* standard variety in Terengganu (see more in §1.4.2). KM and CTM each have over a million speakers, and they are vigorously spoken across generations as an essential part of the local people's cultural identities. Even so, they are increasingly being influenced by SM.

The remainder of this introductory chapter provides further information on NEPMs and the area where they are spoken. §1.2 presents an overview of the Malayic languages with a focus on their historical background and classifications. §1.3 introduces the vernacular Malayic varieties spoken on the Malay Peninsula. §1.4 takes a closer look at NEPMs, examining the geo-historical settings of Kelantan and Terengganu and providing basic linguistic facts about NEPMs. §1.5 reviews previous studies on NEPMs. §1.6 explains the methodology, data collection and data processing in this study, and offers a summary of the transcription conventions. §1.7 outlines the structure of this dissertation.

1.2 The Malayic languages

The Malayic languages are a group of languages belonging to the Malayo-Polynesian branch of the Austronesian family, spoken primarily in island Southeast Asia. The Malayic subgroup includes Malay proper, the standardised forms of which are the national languages of Malaysia, Indonesia, Brunei and Singapore, a large variety of Malay dialects, and various languages that are sufficiently close to Malay. The total number of Malayic-speaking population is difficult to estimate, but Malay proper alone has almost 280 million speakers (including those who speak Indonesian as a second language, Adelaar 2018: 571). The dispersal and distribution of the Malayic languages are depicted in Figure 1.2. The figure also shows that the core Malayic-speaking areas are West Borneo, Sumatra and the Malay Peninsula.

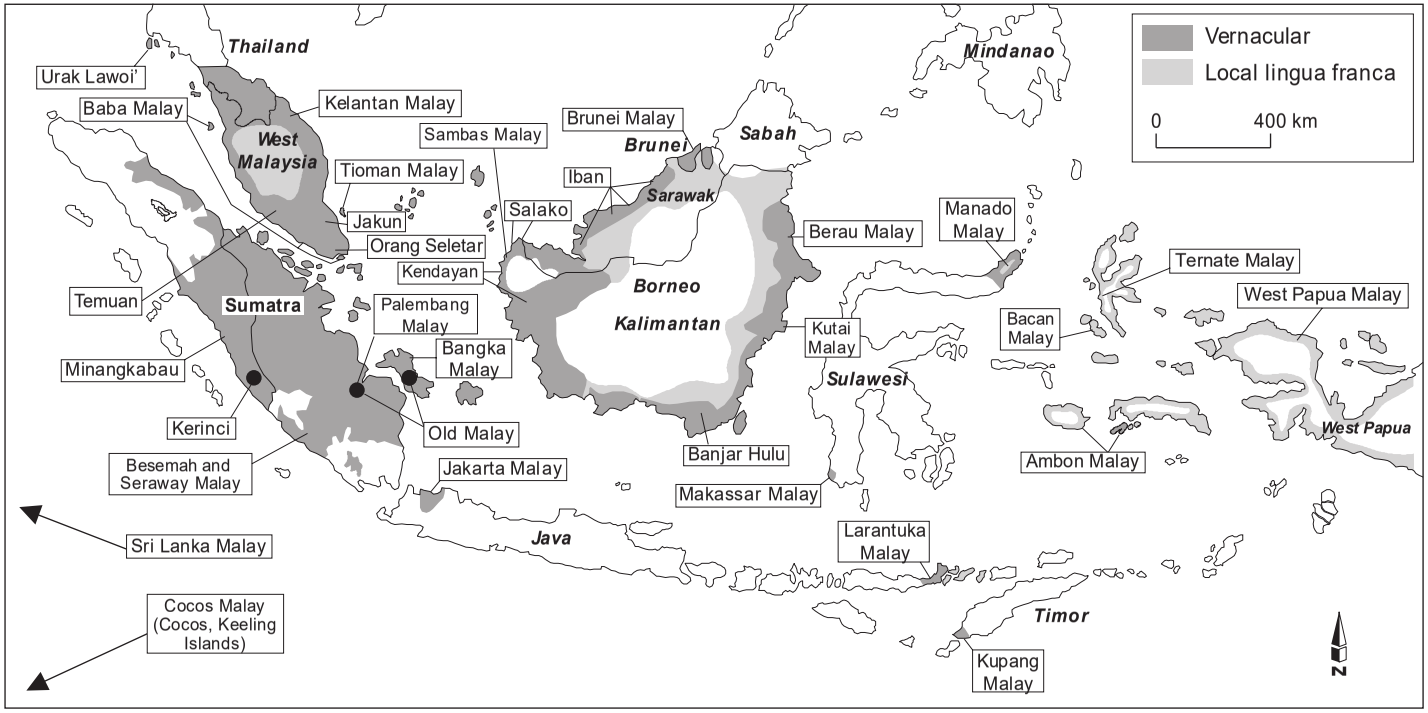

Figure 1.2: The spread of Malayic varieties (modified from Adelaar 2005c: 203)

The establishment of Malayic as a subgroup within the Malayo-Polynesian languages is on the basis of a set of shared innovations that all Malayic languages have undergone since Proto Malayo-Polynesian (henceforth PMP), which can be reconstructed to a common ancestral language, namely Proto Malayic (PM). The reconstruction of PM in Adelaar (1992) was primarily based on six Malayic varieties: SM, Minangkabau (central-west Sumatra), Banjar Hulu (southeast Borneo), Seraway (southwest Sumatra), Iban (northwest Borneo) and Jakarta Malay (Java).

The internal subgrouping of Malayic, on the other hand, has been much disputed. Even before Malayic was well recognised as subgroup, a distinction was often made between Malay dialects and non-Malay languages, usually based on non-linguistic criteria. In Borneo, for instance, Malay dialects refer to the speech forms of ethnic Malays who are Muslims, whereas if the speakers do not consider themselves as Malay on ethnic, cultural or religious grounds, their speech forms are regarded as separate languages (Hudson 1970). This differentiation has its usefulness, but an undesired linguistic implication is that the so-called Malay dialects are perceived as genetically closer to one another, and that they constitute a lower-level group, i.e., a Malay group within Malayic. However, the demarcation between the hypothetical Malay group and the Malayic group, whose members supposedly descend from two distinct proto languages, has never been made explicit, and the scope of languages descending from "Proto Malay" remains unclear. As Blust (1988: 1–5) shows, Banjar Malay is commonly taken as a Malay dialect whereas Iban is not, but on the basis of lexical evidence, Banjar Malay and Iban are equally distinct from SM. Blust (1988: 6–7) further suggests that northern Peninsular Malayic varieties such as Kedah Malay and Terengganu Malay, which are traditionally taken as Malay dialects, might not be much closer to SM than non-Malay languages such as Minangkabau and Kerinci. Asmah (1995) intended to reconstruct Proto Malay (*bahasa Melayu induk*), where all Peninsular Malayic varieties were conveniently included, but the boundary of languages belonging to her Malay group was not well defined, and some varieties showing important retentions such as ITM were overlooked. The labels "Malay dialects" and "non-Malay languages" can still be found in more recent literature, but it should be borne in mind that the distinction is often arbitrary and not indicative of genetic distance.

Various subgrouping proposals based on more convincing linguistic evidence have been put forth by Adelaar (1992, 1993, 2008), Collins (1994),

Nothofer (1996, 1997), Ross (2004), Anderbeck (2012) and Smith (2017), but a detailed classification has not yet been reached (see appraisals in Adelaar 2005d: 17–19 and Anderbeck in print). A general consensus is that West Bornean languages such as Iban and Kendayan (also known as Kanayatn) are distantly related to SM, representing primary branch(es) in the Malayic family tree. The majority of other Malayic varieties, including all Peninsular varieties, cannot be satisfactorily classified into finer-grained groups due to the lack of clear exclusively shared innovations. They are often grouped together as belonging to one single branch, which has been referred to variously as "Nuclear Malayic" (Ross 2004, which serves as the basis for the classification on *Glottolog 4.7*, Hammarström et al. 2023), "other Malayic" (Smith 2017), or simply "Malay" (Anderbeck 2012). Based on the highest linguistic diversity and retentions attested in languages in West Borneo, scholars generally agree that this region is the prehistorical homeland of the Malayic languages (Blust 1985, 1988, 1994; Adelaar 1988, 1992, 1995, 2004b; Nothofer 1996, 1997; Collins 2001, 2006). Additionally, the spread of languages towards the interior in Borneo also suggests a longer period of diffusion. Southeast Sumatra is traditionally taken as the cradle of Malay civilisation and culture, where Malayic speakers founded the maritime empire Srivijaya and developed a separate Malay identity, leaving behind the earliest inscriptions written in Old Malay dating back to the seventh century (Andaya 2001: 317; Andaya & Andaya 2017: 31–32; Adelaar 2004b: 4–5). The Malay Peninsula, on the other hand, is commonly considered as a late settlement of Malayic-speaking population, as evidenced by the demographic pattern where Malays dominate the coasts and push Orang Asli (Malay for 'aboriginal people') further inland (Skeat & Blagden 1906: 434; Bellwood 1993; Adelaar 1988: 74, 2004b: 4).

A supplementary classification has been made along socio-historical lines. Three broad categories were recognised by Adelaar & Prentice (1996): 1) literary Malay, 2) lingua franca Malay and 3) "inherited" Malay. A number of other terms with similar meanings were used in later publications: literary Malay has been referred to as Court Malay, Classical Malay or standard varieties; lingua franca Malay as vehicular Malay, trade Malay or Pidgin Derived Malay; and "inherited" Malay as vernacular varieties (Adelaar 2005c, 2018; Paauw 2008; Anderbeck in print).[1] Originally intended for

[1] The scope of these terms is not always the same, and they are not necessarily mutually

categorising different forms of Malay, this classification was later expanded to include other Malayic languages (see, for instance, Adelaar 2005c). Literary Malay is the predecessor of present-day SM, which developed in Malay courts across the region from the fifteenth century. Vehicular Malay refers to the contact varieties spoken mainly in Eastern Indonesia and other ports throughout island Southeast Asia, which likely arose against a certain socio-historical background with a pidginised form of Malay as a common source. "Inherited" or vernacular Malay(ic) are varieties spoken in traditional Malayic speech communities in Borneo, Sumatra and the Malay Peninsula, practically comprising all other Malayic varieties that appear to have directly inherited from PM, including non-Malay languages such as Iban and Kendayan.

1.3 Peninsular vernacular Malayic varieties

With few exceptions (e.g., SM as the literary variety and Baba Malay, which is a vehicular variety), Peninsular Malayic varieties are vernaculars along socio-historical lines. They are primarily spoken by ethnic Malay groups, hence typically known as Malay dialects. There are also some Malayic-speaking Orang Asli groups such as Temuan and Jakun, and a few groups of Malayic-speaking Orang Laut (Malay for 'sea people') including Orang Seletar and Urak Lawoi' (see Figure 1.2).

This diversity already captured the attention of British lexicographers and grammarians during colonial times. While there had not been dedicated studies on any particular non-standard Malay(ic) variety, notes on regional variation were included in some early Malay dictionaries and grammars in the nineteenth century. The grammar by Crawfurd (1852: 75–76) briefly mentions that Malay dialects often differ in pronunciation and the usage of personal pronouns. The Malay–English dictionary by Clifford & Swettenham (1894: vi) contains a section more specifically on local peculiarities of the Peninsular dialects, where the authors outline the pronunciation of various dialects and note that "the states of Patani and Kelantan are more rich in local words than any other places in the Peninsula and there the low-

exclusive. For example, some vernacular varieties such as Minangkabau and Jakarta Malay are also used as lingua francas. It is therefore best to avoid the term "lingua franca Malay" and restrict the second category to "vehicular Malay".

est form of Malay is spoken." Another English–Malay dictionary (Shellabear 1916: iv) comments that the Malay language is spoken in considerably diverse dialects across the islands in the archipelago, and the varieties spoken on the east coast of the peninsula differ particularly from those on the west coast.

Contemporary Peninsular Malaysia consists of eleven states, and it is often claimed that each state has its own dialect (e.g., Asmah 1977, 1985 and their revised versions published in 1991 and 2008; Nik Safiah et al. 1986: 30–32; Abdul Hamid 1994: 1–2). However, as Collins (1989) points out, this "canon of Malay dialects" corresponding to the state boundary grid is far from the reality. As an example, consider "Terengganu Malay". This name is commonly used to refer to the Malay dialect spoken in the state of Terengganu as if it is a homogeneous variety, but the label is ill-defined for several reasons. First, not all populations in Terengganu speak "Terengganu Malay"; people from the northern district Besut and part of Setiu are predominantly KM-speaking. Second, "Terengganu Malay" is not only spoken in Terengganu, but also in fisherman's villages thinly stretching southwards along the coast to Johor (Ismail 1973; Collins 1989). The term Coastal Terengganu Malay (CTM) is therefore more appropriate. Lastly, the population in the inland area of Terengganu speaks a highly distinct variety which has been referred to as Ulu Terengganu or Inland Terengganu Malay (ITM). While ITM is often considered a subdialect of "Terengganu Malay", it is in fact not necessarily closer to CTM than to KM.

The exact number of distinct Malayic varieties and their boundaries remain to be studied further, but suffice it to say, among all Peninsular varieties, those spoken in the northern states stand out with marked features. This observation was already noted at the turn of the twentieth century (Clifford & Swettenham 1894; Winstedt 1923), and it has been confirmed by later studies on some of the varieties spoken in northern states including Kelantan and Terengganu (see §1.5). There have also been a number of overviews of the diversity of Malay dialects on the Malay Peninsula (Ismail 1973; Farid 1976: 112–132; Teoh 1994: 104–107), or in Malaysia as a whole (Asmah 1977, 1995). From these, it is evident that the varieties spoken in Kelantan and Terengganu are among the most divergent ones, most notably for the remarkable sound changes they have undergone and the specific usage of some local words.

1.4 Malayic varieties in Kelantan and Terengganu

Before outlining the basic linguistic facts about the Malayic varieties spoken in Kelantan and Terengganu, it is crucial to first examine the geo-historical settings of the region, as presented in §1.4.1. This examination is essential as it illuminates how the development and distribution of languages are intrinsically connected to the geographic environment and the historical background within which they have evolved. It also lays the foundation for a deeper understanding of the linguistic characteristics and historical evolution of NEPMs.

1.4.1 Geo-historical settings

1.4.1.1 Geography, demographics and livelihoods

In the Malay Peninsula, as in many other places in Southeast Asia, the most important natural features with which local communities interact are the rivers and the sea (Dobby 1942; Miksic 1978: 170). Prior to the nineteenth century, traditional Malay communities typically depended on the rivers and the sea for their livelihood; the Malays in Kelantan and Terengganu were no exceptions. Villages were established by riverbanks or coastlines before roads were built, where water routes served as the primary means of movements and communication. Another geographical trait characterising Kelantan and Terengganu is the surrounding mountain ranges, which largely isolate these states from the rest of the peninsula and have posed great impediments to trans-peninsular movements until recent times (see, for example, Swettenham 1885; Clifford 1897). The geographical details of Kelantan and Terengganu are provided in greater depth in this section, which also encompasses information about the demographics and livelihoods of the populations in these states.

Kelantan is the largest Malaysian state on the Malay Peninsula, spanning a total area of 15,040 km^2. It is bordered by the Narathiwat Province of Thailand to the north, Perak to the west, Pahang to the south and Terengganu to the southeast (see Figure 1.1). Its geographic boundaries are relatively well defined, with the Golok River marking the Malaysian-Thai border, the jungle-clad Titiwangsa Range extending over the Kelantan-Perak boundary, the Tahan Range delimiting Kelantan from Pahang, and

the Pantai Timur Range covering a large part of the Terengganu frontier. Bounded by these mountain ranges lies a low-lying and flat alluvial plain, with the Kelantan Delta situated at the estuary of the Kelantan River.

The Kelantan River, which is named after the name of the state (or might have given its name to the state), is fed by several major tributaries that originate in the south and southwest of the state, as illustrated in Figure 1.3. The Nenggiri River (also known as the Betis River in its upper reaches) has its headwaters in the Titiwangsa Mountains and flows eastward, merging with the Galas River at Kuala Sungai. From there the Galas River flows northeast wards and merges with the Lebir River. The Galas River and the Lebir River both originate in southern Kelantan near the border with Pahang, and they converge at Kuala Krai to form the Kelantan River.

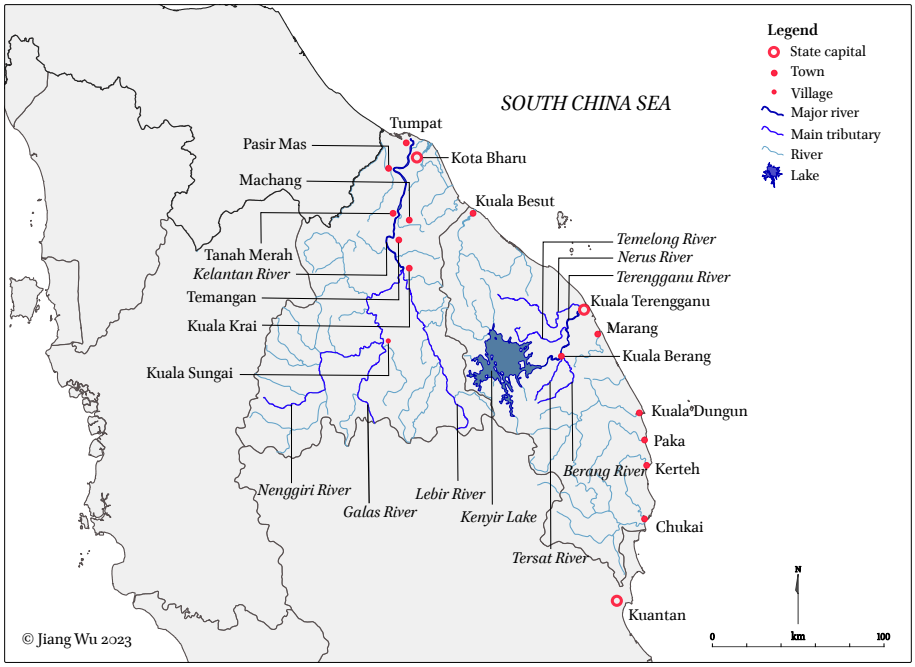

Figure 1.3: River systems of Kelantan and Terengganu

The Kelantan River and its tributaries play a vital role for the local communities, providing water for agriculture and supporting fishing, which are crucial sources of livelihood. The importance of the river system can be seen from the placement of main settlements in the state: from Kuala Krai, ma-

jor towns along the banks of the Kelantan River include Tanah Merah, Pasir Mas, Kota Bharu, the capital city of Kelantan, and Tumpat. The rivers must have been of great importance even in ancient times, as evidenced by the discovery of Gua Cha, one of the most significant archaeological sites on the peninsula, located on the bank of the Nenggiri River (Sieveking 1954; Adi 1985).

According to the 2020 Malaysian Population and Housing Census (available at https://www.dosm.gov.my/v1/), Kelantan has a population of approximately 1.79 million, among which the Bumiputera (Malay for 'sons/daughters of the soil', comprising Malays and indigenous groups including Orang Asli of the peninsula) make up 96.6%. Chinese make up another 2.5% of the population, and Indians constitute 0.3%. The Kelantanese Malays practise Islam. Politically, the Islamic Party of Malaysia has been ruling Kelantan uninterruptedly for over three decades. The party has been pushing for the gradual enforcement of Shari'a laws in the state, leading to the state's reputation as one of the most conservative Malay heartlands alongside Terengganu.[2]

The Kelantanese Malays have a long tradition of practising intensive wet-rice agriculture, and the fertile soil of the Kelantan Delta has made it a major centre of rice production in Malaysia (Dobby 1951; Hill 1951; Cheng 1969). Rice cultivation remains a significant part of the state's economy, along with rubber-tapping, which is another traditional economic activity in the village sphere (Downs 1960; Nash 1974). Other crops grown in Kelantan include oil palms, coconuts, cassava and various vegetables and fruits such as durians, papayas and rambutans. Fishing has also been an important source of livelihood for fisherman's villages along the coastline (Graham 1908: 65; Firth 1943, 1966; Norfatiha & Nor Hayati 2022). In more recent years, Kelantan's economy has become more diversified, with increasing investments in manufacturing activities and tourism. Traditional agriculture is becoming less attractive to young generations, and as the population grows, some parts of the traditional rice paddies have been cleared to make room for housing developments.

[2] The Islamic customs, such as the way of dressing, appear to have been rather different a century ago, see Graham (1908: 24–26, 31–33).

Situated to the southeast of Kelantan, Terengganu (formerly also spelled Trengganu or Tringganu) has an elongated shape covering an area of 13,035 km^2. Its geography is characterised by a long coastline along the South China Sea, stretching over 200 kilometres. The coast strip consists largely of open sandy beaches, which extend further south into Pahang and north into Kelantan. From the east to the west, the flat coastal plain gradually rises to hilly terrain in the interior, where the mountains form Terengganu's borders with Kelantan and Pahang.

Terengganu has several rivers that stem from the mountainous interior and flow towards the sea, each supporting an important town at their estuaries along the coast (see Figure 1.3). This pattern is quite different from the one in Kelantan, where traditional settlements are primarily located along the Kelantan River and its tributaries. The Terengganu River, which flows through the state capital Kuala Terengganu, is the largest and most prominent river in Terengganu. It originates in the highlands in the central part of the state, fed by several main tributaries including the Berang, Tersat, Telemong and the Nerus River. Another major town along the Terengganu River is Kuala Berang, which is near the confluence of the Berang, Tersat and the Terengganu River. In addition to the Terengganu River, other notable rivers in the state include the Besut River in the north, the Marang River, the Dungun River and the Kemaman River in the south, which support the towns of Kuala Besut, Marang, Kuala Dungun and Chukai respectively. Just like in Kelantan, these rivers serve as important waterways for transportation and commerce, as well as providing water for agriculture and other uses.

Figure 1.3 also shows that a large portion of the upstream Terengganu River is now submerged by the Kenyir Lake, which is the largest man-made lake in Southeast Asia. The lake was formed by damming several tributaries of the Terengganu River for the purpose of generating hydroelectric power. The construction of the Kenyir Dam and the creation of the Kenyir Lake between 1978 to 1985 have considerably altered the landscape of the interior of Terengganu, as evident from a comparison between the present-day map of river systems and the depiction in Firth (1943: 194), as shown in Figure 1.4. The project also led to the relocation of several villages, both Malay and Orang Asli ones.

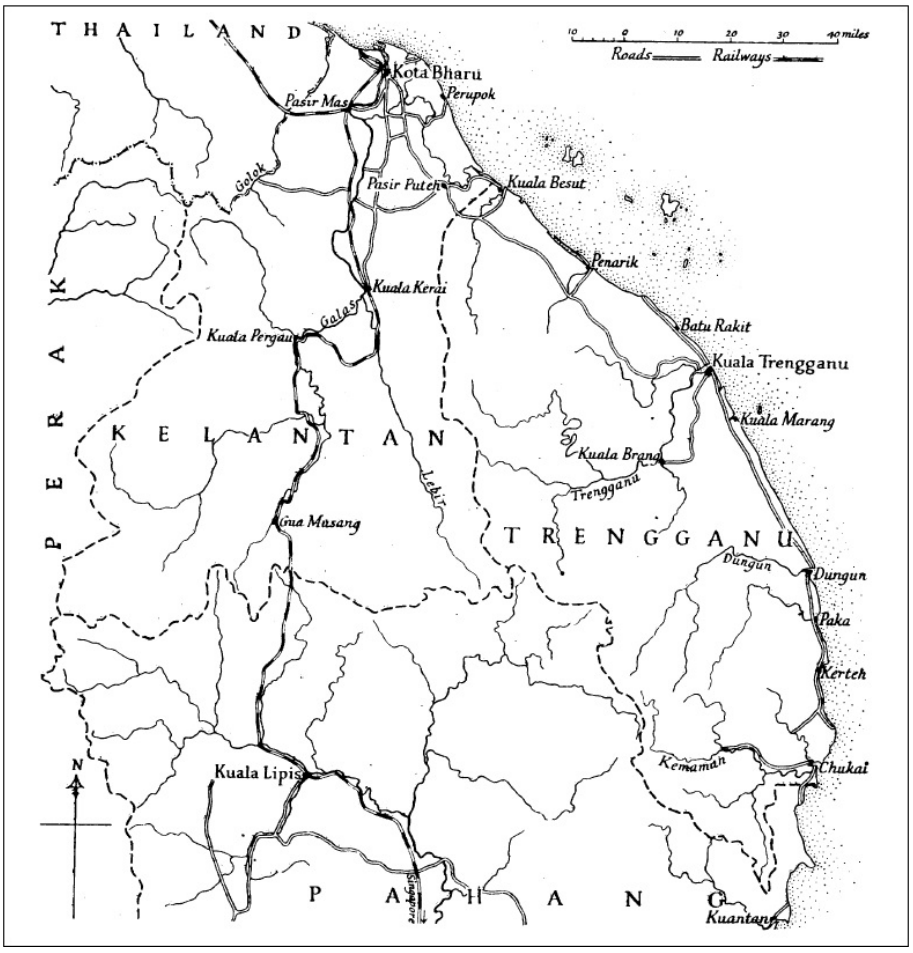

Figure 1.4: A map of Kelantan and Terengganu before the 1980s (Firth 1943: 194)

As of 2020, Terengganu had a population of around 1.15 million, with the percentage of Bumiputera standing at 97.6%. Chinese and Indians make up a small percentage of the population at 2.1% and 0.2% respectively. Along with Kelantan, Terengganu is one of the Malaysian states with the highest concentration of ethnic Malays who practise Islam. Currently, Terengganu is also ruled by the Islamic Party of Malaysia. The primary economic activities in Terengganu used to be agriculture and fishing. Apart from rice, rubber, oil palms and coconuts, other important crops grown in the state in-

clude corns, peanuts, peppers, cucumbers and watermelons (p.c. with consultants). Thanks to the state's extensive coastline, fishing and other related industry such as boatbuilding continue to play an important role in the economy and cultural heritage (Norfatiha & Nor Hayati 2022, for the historical significance of the fishing industry in Terengganu, see Firth 1943, 1966 and Gosling 1978). However, the discovery of oil wells off the coast in the 1970s, especially in the southern areas of Kerteh and Paka, has significantly transformed Terengganu's economic structure to become heavily reliant on the oil and gas industry. In recent years, Terengganu has also seen growth in the tourism sector, particularly with the popularity of its coastal islands.

1.4.1.2 History

While ethnic Malays dominate contemporary Kelantan and Terengganu, this was likely not the case in the past. The northern part of the Malay Peninsula and the Kra Isthmus further north have been a contact zone for centuries, if not millennia. This region has long been where the speakers of Austronesian languages from the south and east met with the Austro-Asiatic and Thai people passing down to the peninsula from the north. Such interactions often led to conflicts, but during peaceful times, this area benefited from its strategic location along early trade routes connected to the South China Sea, attracting foreign travellers and traders from China, India, the Middle East and Europe. This section provides a concise history of Kelantan and Terengganu, highlighting the interactions and power transitions between various groups of people.

Ancient kingdoms established in the region can be dated back to as early as the first centuries of the Christian era. Ptolemy's map, which was drawn based on the Roman geographer's book *Geography* composed in the second century, shows two ports on the east coast on the Malay Peninsula, marked as Perimula and Coli polis (or Koli polis, Kole polis). The locality of these two ports has been variously identified by historians as corresponding to present-day Nakhon Si Thammarat (Ligor) and Kelantan (Gerini 1909: 105–111), or at the mouths of the Terengganu River and the Kemaman River (Braddell 1936: 37), or somewhere near the Kuantan River in present-day Pahang (Linehan 1951: 94; Wheatley 1955: 16). Chinese historical records also attest to the existence of political entities in this region in the early years of the Christian era. *Han-shu* [The book of the Han Dynasty], which includes a

Treatise on Geography, notes a country named 都元 (now read Du-yuan) in the first century, which is sometimes believed to be related to present-day Dungun in southern Terengganu (Hsu 1961: 97). The seventh-century *Liang-shu* [The book of the Liang Dynasty] recorded a mission to Funan (southern Indochina, present-day Cambodia and southern Vietnam) during the Wu Dynasty in the third century. The record also mentioned that Funan attacked several countries including 屈都昆 (Qu-du-kun), 九稚 (Jiu-zhi) and 典孙 (Dian-sun). As these countries were documented as being situated across the gulf from Funan, they must have been on the east coast of the Malay Peninsula.[3] 九稚 (Jiu-zhi) was also known elsewhere as 拘利 (Ju-li), which, according to Wheatley (1955: 15–16), likely corresponds to Kole in Ptolemy's map. While we know little more than the names and approximate locations of these ancient kingdoms (see a summary of different interpretations in Wheatley 1973: 14–25, 152–155), these early records demonstrate that the northeast coast of the Malay Peninsula was already home to important settlements that attracted travellers from both the west and the east.

The northern Malay Peninsula attained more prominence between the fifth and the seventh century, owing to the decline of Funan's power to the north and the emergence of extensive trading networks. A number of Hindu-Buddhist kingdoms existed in this region around the sixth century, as attested by Chinese dynasty annals such as *Liang-shu* and *Sui-shu* [The book of the Sui Dynasty]. Some of the most important kingdoms include 狼牙修 (Lang-ya-xiu or Langkasuka), 赤土 (commonly transliterated as Chi'tu) and 丹丹 (Dan-dan or Tan-tan), and from various sources it is clear that they lay next to each other from the north to the south along the east coast of the peninsula. Hsu (1947, 1961: 161–166) identifies Chi'tu as in present-day Songkhla and Tan-tan in Kelantan, whereas Wheatley (1973: 36, 55) locates Chi'tu in the upper reaches of the Kelantan River, and Tan-tan in Terengganu. As recorded in *Sui-shu*, Chi'tu was an advanced kingdom, to which an embassy was sent in the year 607. It was described as a Mon-Khmer kingdom founded by the descendants of Funan, where Hinduism was practised. Langkasuka was located in the northern part of the peninsula, generally suggested as the predecessor of the later Patani

[3] 屈都昆 (Qu-du-kun) is probably the same country that was referred to as 屈都乾 (Qu-du-qian), 屈都 (Qu-du) or 都昆 (Du-kun) in several other early works (Wheatley 1973: 21–22). *Jin-shu* [The book of the Jin Dynasty] recorded an event of 屈都乾 (Qu-du-qian) being invaded by Champa in the fourth century.

Kingdom (Wheatley 1956, 1973; Teeuw & Wyatt 1970: 1–3; Bougas 1990; Andaya & Andaya 2017: 74). The archaeological sites discovered in the city of Yarang seem to support this proposal (Wales 1974; Jacq-Hergoualc'h 2002: 166–191). *Liang-shu* mentions that Langkasuka was probably established around the second century, and its king sent an envoy to China in 515. Various names related to Langkasuka recurred in Chinese historical records until the fifteenth century, making it one of the most long-lasting kingdoms in the region. It is likely that the territory of Langkasuka extended to Kelantan and Terengganu in its heydays after Chi'tu and Tan-tan went into decline (Sheppard 1949).

From the seventh century onwards, Srivijaya rose to power as a maritime empire centred in Sumatra, and eventually gained control of much of the Malay Peninsula and surrounding islands. Langkasuka, together with other kingdoms located on the Malay Peninsula at that time, was recorded as paying tributes to Srivijaya in the thirteenth-century *Zhu-fan-zhi* [A description of barbarian nations]. It is also in this record that the names Kelantan (吉蘭丹 Ji-lan-dan, which is currently the state's official Chinese name) and Terengganu (登牙侬 Deng-ya-nong) first appeared, indicating that these states assumed independent identity before the thirteenth century, although still vassals of Srivijaya. The same source also referred to a neighbouring country called 佛罗安 (Fo-luo-an or Fo-lo-an), which is thought to be located in present-day Kuala Berang in the interior of Terengganu, where the Terengganu inscription stone was later discovered (Wheatley 1973: 70, also see below). These countries (or city-states) practised Buddhist culture (just like their suzerain Srivijaya), and produced local goods such as gharuwood, lakewood, sandalwood and ivory.

By the early fourteenth century, Srivijaya had fallen. The Siamese Ayutthaya Kingdom in the north and the Javanese Majapahit Kingdom in the south began to rise and exert influence on the Malay Peninsula. The states on the peninsula likely maintained semi-independence as small principalities. The names 吉蘭丹 (Ji-lan-dan) and 丁家盧 (Ding-jia-lu) occurred in *Dao-yi-zhi-lüe* [A brief account of island barbarians] written around 1339 (Rockhill 1915), and 古蘭丹 (most likely a misprint of 吉蘭丹, i.e., Ji-lan-dan) and 丁架路 (Ding-jia-lu) are shown in the Mao Kun Map which documents the voyages of Admiral Zheng He (or Cheng-ho) between 1403 and 1433. Both states are mentioned as dependencies of Majapahit in the fourteenth century Javanese poem *Nagarakṛtāgama* (Winstedt 1935: 30; Pigeaud

1960: 17), and Terengganu (written as 丁機宜 Ding-ji-yi) is also recorded in *Ming-shi* [The history of the Ming Dynasty] as a vassal of Majapahit.

The fourteenth century also saw the transition from Hindu-Buddhism to Islam in island Southeast Asia. The Terengganu inscription stone, written in Classical Malay in the Jawi script (a writing system based on the Arabic script), symbolises the presence of Malay and Islamic influence on the peninsula. The inscription has a date that can be read variously between 1303 and 1387, and it describes a proclamation by a Terengganu ruler who claimed that Terengganu was the first state to receive Islam and provided basic Shari'a laws for the guidance of his subjects (Paterson 1924; Andaya & Ishii 1992: 514). The introduction of Islam to the east coast of the peninsula therefore predated Malacca's conversion to Islam, which probably took place during the reign of Sultan Megat Iskandar Shah around 1414 (Wake 1964; Cœdès 1968: 246; also see Mills 1930: 49; Teeuw & Wyatt 1970: 4).[4] The Islamisation of Kelantan presumably happened around the same time or somewhat later in the mid-fifteenth century. According to *Ming-shi*, the Maharaja of Kelantan 苦马儿 (Ku-ma-er) sent an embassy to China in 1411 (Rentse 1934: 47), and *Sejarah Melayu* [The Malay annals] describes an event of Malacca invading Kelantan around 1500 for not paying homage and mentions the name Sultan Mansur Shah of Kelantan (Winstedt 1938: 12). The names and the titles of the rulers suggest that Kelantan was still an Indianised state in 1411, but it had already embraced Islam by the end of the fifteenth century.

As Malacca quickly grew in power in the fifteenth century, Terengganu became integrated into the Malacca Sultanate by the time of Sultan Muhammad Shah (ca. 1424–1444). Kelantan also became a vassal of Malacca following the attack around 1500 (Winstedt 1938: 5, 12). According to *Sejarah Melayu*, the ruling family of Terengganu was allegedly murdered by the sultan of Pahang (who was an elder brother of the sultan of Malacca) in 1478, after which a former Pahang governor's family ruled Terengganu for over a century (Linehan 1936: 14–15; Sheppard 1949: 5–6). Following the Portuguese conquest of Malacca in 1511, the last sultan of Malacca retreated to Johor, where his heir established a new ruling dynasty, with which Terengganu maintained close ties. The history of Kelantan in this

[4] Islam mostly likely spread to the peninsula directly from Sumatra, where it had already been introduced by 1281 (Cœdès 1968: 202, 231).

period, on the other hand, was intertwined with that of Patani to the north. Built on the ground of former Langkasuka, Patani emerged around the middle of the fifteenth century and soon adopted Islam (Teeuw & Wyatt 1970: 3–4; Bougas 1990: 115). After the fall of Malacca, Patani became a prosperous and important trade centre that was favoured by Chinese and Muslim merchants. It was also frequently visited by Portuguese, Dutch and English traders. Despite some internal disruption and strong Siamese influences, Patani reached the peak of its prosperity in the early seventeenth century (Teeuw & Wyatt 1970: 5–20; Ibrahim Syukri 1985: 13–38; Andaya & Andaya 2017: 73–75). It appears that the significance of Kelantan diminished dramatically following the rise of Patani and the southward expansion of Siamese pressure, and it is likely that Kelantan was divided into small chiefdoms that subordinated either to Patani or Terengganu (Graham 1908: 38–39; Andaya & Andaya 2017: 73). According to some sources, much of Kelantan had already been incorporated into Patani by the time of the reign of Raja Biru (ca. 1616–1624) (Abdullah Mohamed 1981: 21–22; Andaya & Andaya 2017: 75). This is also testified by the Chinese record *Dong-xi-yang-kao* [Notes on Eastern and Western Oceans], compiled in 1617, in which Kelantan was described as a port of Patani. There was nevertheless a period of Kelantanese rule in Patani in the second half of the seventeenth century or the early eighteenth century, suggesting a close relationship between the two states (Teeuw & Wyatt 1970: 20–22; Andaya & Andaya 2017: 76). Oral traditions, as summarised in Rentse (1934), also tell that the ancestors of the royal family of Kelantan came from overseas and first arrived in Patani. The prosperity of Patani came to an end when the Patani-Siam relationship deteriorated in the late seventeenth century. The city was eventually invaded and destroyed by the Thais in 1786, leading to Patani's complete subjugation to Thai rule (Teeuw & Wyatt 1970: 23; Ibrahim Syukri 1985: 41–44).

The current Terengganu Sultanate was established in 1725 by Sultan Zainal Abidin, a younger brother of a former Johor sultan who took refuge in Terengganu after being expelled (Sheppard 1949: 8–11). Meanwhile, Kelantan was ruled by many local chieftains after the decline of Patani. Following a period of disorder, a local chief named Long Pandak from Kubang Labu came into power. Eventually, Long Yunus, the son of an admiral to Raja Long Pandak, managed to unify Kelantan with the assistance of Sultan Mansur Shah of Terengganu, and was enthroned as the Sultan of Kelantan

in 1793 (Marriott 1916: 17; Rentse 1934: 51–53). Both states were nevertheless struggling to maintain their independence after attaining sultanate status, as the Siamese Kingdom of Rattanakosin once again began to expanded its influence southwards with greater demands. Both Kelantan and Terengganu were sending *bunga mas* (Malay for 'golden tree', a form of tribute) to the Thai king by the late eighteenth century and early nineteenth century, and it was clear that Siam viewed Kelantan and Terengganu as its tributary states (Newbold 1839: 61–65; Rentse 1934: 59; Sheppard 1949: 19; Wyatt 1974; Andaya 1986). In fear of absorbed by the Thais, Kelantan and Terengganu sought assistance from the British, who had already gained a strong hold in the southern part of the peninsula. In 1822, Kelantan petitioned the British to be accepted as a vassal state, but their plea was in vain (Andaya & Andaya 2017: 128–129).

The Thai claim to suzerainty over Kelantan and Terengganu lasted until the beginning of the twentieth century. With the Anglo-Siamese Treaty 1909, Kelantan and Terengganu (together with the Kedah and Perlis on the west coast) were transferred to the British control and became British protectorates known as the "Unfederated Malay States". Each state received a British advisor while keeping their own local ruler. After the Second World War, Kelantan and Terengganu joined the Malayan Union in 1946 and subsequently became part of the Federation of Malaya in 1948, which ultimately gained independence in 1957.

The history of Kelantan and Terengganu summarised above reveals that these states have undergone several distinct phases of development, and they have been populated by different groups of people over time. There is strong evidence indicating that Malays, who adopted Islam in the fourteenth to fifteenth centuries, have dominated both states since then. However, little is known with certainty about the populations or languages spoken in this region before the fourteenth century. Chinese historical records suggest that sixth-century kingdoms such as Chi'tu were possibly Mon-Khmer in their culture and population composition. By the thirteenth century, the northern peninsula came under the influence of Srivijaya, which was a Malay state (Cœdès 1968: 82–83). Nevertheless, it is unclear whether Srivijiya's vassal states on the peninsula were still Mon-Khmer states or if they had already been dominated by Malays. According to the founding legend of Patani, the kingdom developed from a coastal village established by Malays from the southern peninsula and Sumatra, while its

direct predecessor was an inland Mon-Khmer kingdom (Ibrahim Syukri 1985: 13–14, who used the term "Siam-Asli" to refer to Mon-Khmer). The legend may not be far from the truth. As mentioned earlier, the predecessor of Patani was presumably Langkasuka, which lay to the north of Chi'tu in the sixth century. Like its neighbour, Langkasuka was probably a Mon-Khmer kingdom as well, and it might have remained so until it was replaced by the Malay Kingdom of Patani in the fifteenth century. Citing Benjamin (1997), Andaya (2001: 324–328) also concludes that the Malay Peninsula was not considered part of the "Melayu lands" before the time of the Malacca Sultanate, and the northern part of the peninsula received a particularly greater influence from Mon-Khmer culture. More concrete evidence of this influence can be found in present-day Sathing Phra to the north of Patani, where artefacts and the citadel discovered on the archaeological sites are believed to be of Mon character dating back to the sixth to the thirteenth century (Stargardt 1983: 32). However, no such evidence can be found in Patani, let alone further south in Kelantan and Terengganu. In short, historical evidence suggests that the northern part of the peninsula likely underwent a transition from being dominated by Mon-Khmer culture and population to its present-day Malay dominance, yet the exact period of this transition cannot be precisely determined.

1.4.2 Basic linguistic facts

There are at least three sufficiently distinct Malayic varieties spoken in Kelantan and Terengganu. KM is believed to exhibit relative homogeneity across the state (Ismail 1973), and this observation is largely confirmed by my personal experiences visiting various districts in Kelantan including Kota Bharu, Tumpat, Pasir Puteh, Pasir Mas, Tanah Merah and Machang. More recently, it has been pointed out that there are notable variations between the coastal variety and the inland variety spoken along the Nenggiri River (Mohd Tarmizi 2018a, b, c), which calls for further study on the regional variation of KM.[5] The KM-speaking area extends to the northern

[5] A reappraisal of Tarmizi's data gives the impression that both varieties share most of the typical KM features, with minor differences that can be seen in the reflection of two sound changes. First, penultimate high vowels have been sporadically lowered in the coastal variety, but they are retained in some inland varieties, e.g., PM *ikur > coastal [ɛkɔː], inland [ikuː] 'tail'; *uraŋ > coastal [ɔɣɛ̃], inland [uɣaŋ] 'person' (following the author's tran-

districts in Terengganu including Besut and Setiu, and potentially spans across the Kelantan-Pahang and Kelantan-Perak borders. KM also shares a close relationship with Patani Malay spoken in the southernmost provinces of Thailand across the border.[6] To what extent KM and Patani Malay resemble or differ from each other still needs to be demonstrated through systematic comparisons, but the available material suggests that they may be considered the same variety, as they are closely related on the one hand, and sufficiently different from other Malayic varieties on the other hand (Tadmor 1995: 13). The ISO 639-3 code *mfa* is assigned to Kelantan-Patani Malay (Eberhard et al. 2023). It is estimated that there are around two million speakers of KM in Malaysia, while Patani Malay has over one million speakers in southern Thailand.

In Terengganu, as previously noted, the varieties spoken along the coast and in the inland area are highly divergent in many aspects. In fact, throughout this dissertation, it will become clear that CTM is closer to KM than to ITM. Collins (1989: 251) reports that CTM is spoken in a narrow strip of sometimes discontiguous villages along the the east coast, from the north of Kuala Terengganu to at least Mersing in Johor (see Figure 1.1). It is also spoken by dwellers in the middle course of various rivers in Terengganu, likely as a result of the movement of inhabitants from the coastal area towards the inland. In the upper valleys of the Terengganu River and its tributaries, which form the district of Hulu Terengganu, villagers residing along the rivers speak various forms of ITM. However, Kuala Berang, the main town of the Hulu Terengganu district, is primarily CTM-speaking. Another vernacular variety spoken in the upper valleys of the Dungun River, known as the Pasir Raja dialect, appears to be closer to the Ulu Tembeling dialect of Pahang Malay (Mohd Tarmizi 2020). Unfortunately, the scanty data on this variety does not allow further discussion. Neither CTM nor ITM has been recognised as a distinct Malay(ic) variety by *Ethnologue*, and neither has been assigned an ISO code. It is estimated that there are around one million speakers of CTM, while the number of ITM speakers is significantly

scription). Second, *-an, *-am and *-aŋ merged to [-ɛ̃] in the coastal variety, whereas in the inland variety *-an and *-am merged to [-ɛ] but *-aŋ is retained as [-aŋ], e.g., PM *ik**an** > coastal [ikɛ̃], inland [ikɛ] 'fish'; *mal**am** > coastal [malɛ̃], inland [malɛ] 'night'; but *ur**aŋ** > coastal [ɔɣɛ̃], inland [uɣ**aŋ**] 'person'.

[6] There is a variant spelling of Pattani Malay (with two "t"s), especially in English and Thai contexts. In the present study, I use the Malay spelling with one "t" for consistency.

lower, likely ranging from 50,000 to 70,000. Additionally, there is concern for the endangerment of ITM as it is not being passed down to younger generations. People under the age of twenty generally do not speak ITM, or only have passive knowledge of it.

Like other vernacular varieties, NEPMs are considered low varieties or basilects, and they are only used in informal settings. Formal education and administration are conducted in SM, while religious matters are typically handled in Arabic. For example, during a funeral in a village in Kelantan, the official welcome would be given in (colloquial) SM by the master of ceremony, while the eulogy would be delivered in Arabic. Guests would likely converse with each other in KM. A diglossic situation like this has probably persisted for centuries in the region, with commoners using the vernacular variety and the royal courts using some form of literary Malay. Language use in Hulu Terengganu is particularly interesting, as villagers who grow up speaking ITM often acquire both CTM and SM. When a villager goes to the market in Kuala Berang, conversations with vendors typically occur in CTM. CTM is also used when conversing with local police officers or in restaurants outside the village. In comparison, during an elementary school sporting event that I attended in Kampung Dusun, all official announcements were made in SM. Hulu Terengganu therefore represents a triglossic situation where ITM is the basilect or the lowest variety, CTM is a mesolect, and SM represents an acrolect or a high form of Malay. Nowadays, there is a radio programme called *GEGAR* with the slogan *nombor satu di Pantai Timur* 'number one on the East Coast', which is broadcast in vernacular varieties targeting East Coast Peninsular listeners, but it is typically mixed with colloquial SM. There are also some famous *syair* and *sajak* (forms of Malay poetry) in ITM in Hulu Terengganu, but it is probable that they were originally written in SM and later translated into ITM. Traces of formal usage of SM can be seen in the morphosyntactic structure of these poems, which deviates from the everyday usage of ITM and may seem unnatural to many speakers.

Given their exclusive use in informal settings, NEPMs lack a standardised orthography and are typically not written. For literate younger generations, SM is the preferred written language, although informal communication such as texting and social media may contain unsystematic forms of spelling that reflect local pronunciation. For older generations who did not receive formal education in SM but had religious education, literacy is

largely restricted to reading the Qur'an. Their texting conventions tend to reflect the pronunciation of the vernacular variety, but they are not always used systematically.

It is worth noting that speakers of NEPMs have different ways of referring to their local vernacular, and they are well aware of the distinction between the three varieties. In Kelantan, KM is commonly known as /lɔɣaʔ klatɛ/ 'the Kelantan dialect' (commonly written as ⟨loghat klate⟩), and there is a popular phrase ⟨kecek klate⟩ /kɛcɛʔ klatɛ/ 'to speak Kelantanese', which is also well known outside Kelantan as it reflects the peculiarities of both word usage and pronunciation of KM (cf. SM *cakap kəlantan* 'to speak Kelantanese'). In coastal Terengganu, CTM is referred to as ⟨loghat tganong⟩ /lɔɣaʔ tganoŋ/ 'the Terengganu dialect'. In village settings, speakers also use the phrase /cakaʔ kapoŋ/ (cf. SM *cakap kampuŋ*, 'to speak the village variety') to refer to speaking CTM. SM, on the other hand, is referred to as /bahasə suɣaʔ/ (cf. SM *bahasa surat*, 'letter language'). In Hulu Terengganu, speakers use the word /uləʊ/ 'inland, upstream' (or its cognates with variable pronunciation, cf. SM *hulu*) to refer to ITM. They also refer to SM as /bahasɛ suɣaʔ/ 'letter language', and CTM is considered the "city variety", as in the phrase /cakaʔ bandɔ/ 'to speak the city variety' (cf. SM *cakap bandar*). The reported mutual intelligibility is that KM and CTM speakers may be able to understand each other, and ITM speakers can understand both CTM and KM, but neither CTM nor KM speakers understand ITM without sufficient exposure.

Linguistically, the distinctions among NEPMs are primarily marked in the phonological systems (see Chapters 2 and 4), but the three varieties also share many common sound patterns, as shown in Table 1.1. In all three varieties, only three consonants /ʔ, ŋ, h/ are allowed in word-final position, which reflect the merger of earlier final stops to /ʔ/, final nasals to /ŋ/ (with further nasal deletion following *a in KM), and PM *s and *h to /h/. The morphosyntactic features of NEPMs also exhibit more similarities than differences in various aspects (see Chapters 5 and 6).

Table 1.1: A comparison of some words in NEPMs

KM	CTM	ITM	SM	PM	Gloss
Merger of final stops to /ʔ/					
*saya**ʔ***	*saya**ʔ***	*saya**ʔ***	*saya**p***	*sayap	‘wing’
*laŋi**ʔ***	*laŋi**ʔ***	*laŋi**ʔ***	*laŋi**t***	*laŋit	‘sky’
*tase**ʔ***	*tase**ʔ***	*tasɛi**ʔ***	*tasi**k***	*tasik	‘lake’
Merger of final nasals to /ŋ/					
tanɛ	*tana**ŋ***	*tana**ŋ***	*tana**m***	*tanam	‘to plant’
*kɨɣɨ**ŋ***	*kɨɣɨ**ŋ***	*kɨɣɨ**ŋ***	*kiri**m***	*kirim	‘to send’
*buɣo**ŋ***	*buɣo**ŋ***	*buɣəʊ**ŋ***	*buru**ŋ***	*buruŋ	‘bird’
Merger of *s and *h to /h/					
*ata**h***	*ata**h***	*ata**h***	*atas*	*atas	‘top’
*pəcɔ**h***	*pəcɔ**h***	*pəcɔ**h***	*pəca**h***	*pəcah	‘to break’

Based on these shared characteristics, a “Northeastern Peninsular Malay dialect subgroup” has been proposed, which, according to Tadmor (1995: 13–14), includes the varieties spoken in Kelantan, Terengganu, Pahang and southeastern Thailand. Collins (1989: 253–254) previously presented a similar version of this subgrouping, suggesting that KM, CTM, ITM and Pahang Malay must have formed a single dialect group at some point in the past. In *Glottolog 4.7* (Hammarström et al. 2023), “Northeastern Peninsular Malay” is considered a single branch consisting of three subbranches: Kedah-Perak Malay, Kelantan-Patani Malay and Urak Lawoi’. “Terengganu Malay” is classified as a member of the Kelantan-Patani branch, alongside Kelantan, Pahang and Patani-Nonthaburi Malay. The classification of this “Northeastern Peninsular Malay” group is illustrated in Figure 1.5.

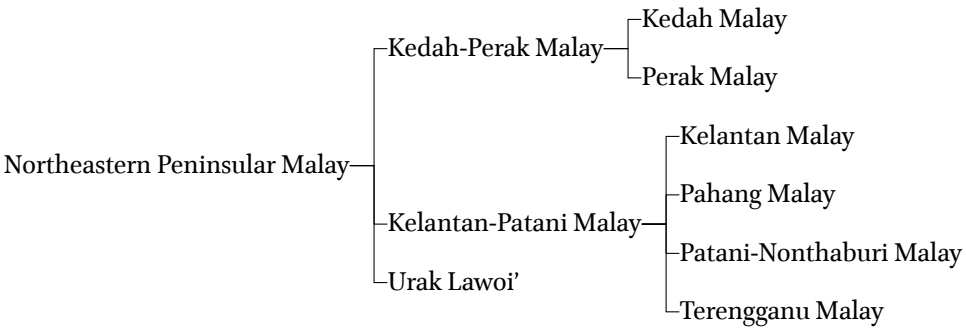

Figure 1.5: Subgrouping of Northeastern Peninsular Malay on *Glottolog 4.7*

However, the basis of these subgrouping proposals is unclear, and the proposed subgroups are not defined by exclusively shared innovations.[7] With a more detailed examination of the historical phonology of NEPMs, I will show that the three varieties do not, in fact, form a lower-level subgroup within Malayic (see Chapter 7). The similarities in their synchronic sound patterns likely have resulted from later diffusion rather than being inherited from an immediate common ancestor.

In the present study, the term "Northeastern Peninsular Malayic" (NEPM) will be used to collectively refer to the varieties spoken in Kelantan and Terengganu for the sake of convenience, but it should be noted that the term is not intended to define a genealogical relationship, but only serves to characterise the geographical area where these varieties are spoken.

1.5 Previous research

Studies on NEPMs so far have covered a wide range of topics with varying degrees of quality. This section presents an overview of the essential literature,

[7] Ajid (2008) also proposed a Patani-Kelantan-Terengganu subgroup, suggesting that Patani Malay and KM have a closer relationship against Terengganu Malay. Unfortunately the methodology used in this study was flawed, and no concrete evidence was presented to support the proposed subgroup. The result of KM and Patani Malay sharing a closer relationship is also hardly surprising, as these two varieties could well be considered the same variety, as mentioned earlier. Moreover, even if all the varieties shown in Figure 1.5 form a tightly-knit subgroup, a more accurate term would be "Northern Peninsular Malay(ic)", as neither Kedah-Perak nor Urak Lawoi' is spoken in the northeast region of the peninsula.

which can be divided into two phases, with the 1960s serving as a dividing line. Early studies primarily consisted of unsystematic observations on the structural peculiarities of NEPMs, while more systematic linguistic research began in the 1960s.

1.5.1 Early studies

The earliest mention of Malayic varieties spoken in Kelantan and Terengganu probably occurs in Munshi Abdullah's *Kisah pelayaran Abdullah ke Kelantan* [The story of Abdullah's voyage to Kelantan], which recounts the writer's experiences on a voyage from Singapore to Kelantan via Pahang and Terengganu in 1838. On the Malay spoken in Kelantan, Abdullah (1949: 44) (translated by A. E. Coope) wrote: "they speak Malay, but their pronunciation is very ugly; they lisp as Tamils do when speaking Malay. Often they leave out a final 't' and add final 'g' and change 'a' to 'o'. But they do not make these changes when writing." In Terengganu,

> "though the people speak Malay, their Malay differs from that of other Malays and sounds strange to the ear; their accent is like that of Kedah Malays. They have a trick of adding a 'g' at the end of words; thus where we say "tuan", they say "tuang"; for "jangan', they say "jangang" and for "bulan", "bulang"; and they say "Alloh" instead of "Allah". This trick however extends only to their speech; they write as we do." (Abdullah 1949: 20–21)

In another early handbook on Kelantan, Graham (1908: 34) also noted that "the Kelantan dialect is a fearsome-sounding jargon in the ear of the Malay of other parts, full of strange clippings and contortions, and sprinkled with words of local manufacture of a Siamese origin, unknown in any other parts of Malaya". These observations were mostly anecdotal, but even nowadays, Malay speakers in Malaysia generally believe that the dialects spoken in Kelantan and Terengganu are unintelligible to speakers from other parts of the country, except perhaps Pahang.

Scholarly studies on Peninsular Malayic varieties began to appear in the early twentieth century. Among the initial contributions to the studies on NEPMs were Sturrock's *Some notes on the Kelantan dialect, and some comparisons with the dialects of Perak and Central Pahang* (1912), Pepys' *A Kelantan glossary* (1916) and McKerron's *A Trengganu vocabulary* (1931), all

authored by officers of the Malayan Civil Service during the colonial period. These studies aimed to document words and expressions specific to these states, along with observations on obvious differences in pronunciation. Lloyd (1921) is another notable contribution, containing transcriptions of native songs and chants recorded in the states of Patani, Kelantan and Kedah ("Lower Siam" in the author's words) sung in the local Malay dialect.[8] The paper offers the first phonetic transcriptions of Kelantan-Patani Malay using International Phonetic Alphabets with great precision. Additionally, the observations made from the comparison between Kelantan-Patani Malay and SM, such as Kelantan-Patani Malay corresponding to SM with changes "occurring chiefly at the end of words, or, at times, of syllables" (Lloyd 1921: 37), are particularly insightful by the standards of their time.

The most important sources from this period are Brown (1927) on KM and Brown (1935) on CTM (referred to as "Terengganu Malay"), later recompiled together with Brown (1921) on Perak Malay and reprinted as Brown (1956). These books comprise of dialogues preceded by brief introductions and general remarks on the usage of personal pronouns, local words and expressions, as well as the pronunciation of these dialects. Brown's works contain informative and mostly accurate observations, and they can further be appreciated in their value in comparative dialectal study. For instance, it was already noted that KM and CTM share many similarities, while ITM reveals striking dissimilarity (Brown 1935: 1, 1956: 124). Unfortunately, the data are transcribed in a confusing and inconsistent orthography that does not correspond to actual pronunciations; instead, the orthography represents how words would have been spelled in SM, which reflects the "true" forms or "a representation of the Malay words in a familiar guise" (Brown 1927: 14) – a doctrine that was also practised in most previous studies on Malay dialects. For examples, KM [tɔʔsɛ] 'not want to' is written as ⟨tak sir⟩ and [ɣɔyaʔ] 'to tell' is written as ⟨ruwiyat⟩ (Brown 1927: 6); the latter is also written as ⟨royat⟩ and ⟨riwayat⟩ elsewhere. While the "standardised" spelling may inform us of the origins of local words, e.g., KM [ɣɔyaʔ] 'to tell' corres-

[8] Though not stated explicitly, it was implied that a single dialect was spoken in these northern states, for which some phonological notes were provided. However, upon examination of the transcriptions and phonological notes, it becomes evident that the recorded dialect was some form of Kelantan-Patani Malay. It also appears that the song recorded in Kedah was in the same dialect, which led to the misconception that northern Malay dialects were homogeneous.

ponds to SM *riwayat* and ultimately comes from Arabic *riwāya*, it disguises the peculiarities of these Malay dialects and makes it difficult for readers with no prior knowledge of these dialects to understand, as they have to remember all the rules while reading the words and dialogues.

On the whole, earlier studies tended to take SM as the point of departure, with which dialects were compared. This approach was based on the unspoken assumption that dialects are secondary to the written language, i.e., they are Malay pronounced in a deviant or an improper way. As a result, the goal was often to identify the "true" forms of local words and expressions, which could be disguised by the use of SM-based orthography (also see comments in Teeuw 1961: 43). As Teeuw (1957: 295) points out in his review of Brown (1956), "it would be important to consider also the Malay dialects in their own rights, and not apriori to make them secondary to and dependent upon the written standard". In spite of these criticisms, earlier works such as Brown's compilations of dialogues remain important sources for understanding the history of NEPMs.

1.5.2 Linguistic studies from 1960s

Modern linguistic research on NEPMs (and Peninsular Malay varieties in general) began in the 1960s. With the exception of works by James T. Collins, almost all studies were carried out by local Malaysian scholars, most of whom are natives of Kelantan or Terengganu.

Nik Safiah's MA thesis (1965) on KM phonology, along with two subsequent articles (1966, 1967) in *Dewan Bahasa*, is among the first studies that treat KM as an independent linguistic entity and offer a systematic analysis of its sound system. While the overall quality is high, the analysis is not always consistent. For instance, in Nik Safiah (1965), she lists nineteen phonemic consonants (/p, b, t, d, c, ɟ, k, g, ʔ, m, n, ɲ, ŋ, s, ɣ, h, l, w, j/) and provides illustrations for contrastive nasal vowels (although it is not specified whether they are phonemic). In contrast, in her later works (1966, 1967), she charts twenty phonemic consonants with the addition of /z/, and no mention is made of nasal vowels. Abdul Hamid (1971) is a BA thesis on the phoneme inventory of KM, and the book titled *Sintaksis dialek Kelantan* [Syntax of the Kelantan dialect] (1994) by the same author is to date the most comprehensive grammar sketch of KM. A summary of the KM phonological history based on Abdul Hamid's data can be found in Adelaar (2005c: 210–212).

However, Abdul Hamid's transcription is quite inconsistent throughout his book, and his analysis of the KM morphological system is particularly debatable. Additionally, Hashim's MA thesis (1974) on KM morphemes and Ajid's (1985) work on the phonology and lexicon of KM (represented by the variety spoken in Pasir Mas) are also noteworthy.

Compared to KM, CTM has received less scholarly attention. While not dedicated to studying CTM, Collins (1980) provides a comparison of Ambon Malay with CTM based on the dialogues compiled by Brown (1935), highlighting the isolating structure of CTM. Abdul Hamid (1990) offers an overview of CTM phonology and some aspects of its morphosyntax. Other BA theses that have been cited in the literature include Othman Omar (1983) on CTM phonology and Kamsiah Salleh (1990) on CTM morphology, but unfortunately I do not have access to them.

ITM has been studied even less extensively but with considerable depth, thanks to Collins' book (1983a), which remains one of the most influential works on a Peninsular Malay dialect. Collins identifies some distinct features of ITM, including the retention of historical high vowels in penultimate syllables and diphthongisation of high vowels in final syllables. It is also noted that ITM has a small inventory of affixes, and passive constructions in ITM are formed with a pre-verbal anaphoric marker *ŋə* or *ɲi*. Also importantly, Collins argues that ITM should not be taken as a subdialect of the ill-defined "Terengganu Malay", but rather a distinct variety on its own. Other published works on related topics include Collins & Naseh Hassan (1981) and Collins (1983b), as well as two unpublished BA theses by Abdullah Junus (1977) and Naseh Hasan (1981). Many of Collins' earlier observations on ITM will be further elaborated and advanced in this dissertation.

In the past two decades, a significant amount of research has been conducted on NEPMs, with a primary focus on phonology and the application of theoretical considerations. Dialectology studies are typically conducted in the context of geolinguistics (see Mackey 1988), aiming to map the differentiation and boundaries of dialects (e.g., Nor Hashimah, Wan Athirah & Khairul 2021; Nor Hashimah, Wan Athirah & Harishon 2021). Mohd Tarmizi (2018a) is an important study focusing on the Malay(ic) varieties spoken in the inland/upstream area on the east coast of the Malay Peninsula and their history, which is particularly relevant to the present dissertation. He hypothesises that the inland area preserves older forms of Malay, as indicated by certain archaic features, and proposes that the spread of Malay varieties

originated from inland regions and then expanded towards the coast. However, I have reservations about the phonological analysis in this work (see below). Moreover, the evidence presented for the spread of Malayic varieties from the inland to the coast is thin (see general discussion in §9.3). Mohd Hilmi et al. (2016, 2018) conducted excellent work on the acoustic aspects of word-initial geminates in KM, following earlier studies on initial geminates in Patani Malay (Abramson 1986, 1987, 2003). The phonological aspects of initial geminates in KM are discussed in §2.2.1. Noraien Mansor et al. (2013) wrote a short monograph on general features of CTM, but it does not offer much advancement compared to previous summaries such as Ismail (1973) and Asmah (1985).

Overall, it is regrettable that, with few exceptions, the morphosyntactic aspects of NEPMs have been generally overlooked, and basic descriptive studies are still lacking. It is also unfortunate that the quality of existing phonological studies is often disappointing, for two reasons.

First, a major issue with many existing phonological studies is the lack of systematic phonemic analysis and differentiation between phonetic realisations and phonological representations. Very often only the phonetic forms are transcribed, and no further attempts are made towards a phonological analysis. For instance, vowel length is sometimes marked in "phonological studies" (e.g., Mohd Tarmizi 2018a, b), yet there is no justification for considering it a distinct feature. Siti Nadiah's thesis (2020) on ITM monophthongs essentially takes all phonetically distinct vowels as phonemes, disregarding some clear allophonic alternations.

Second, a more common and serious problem arises from the lack of distinction between diachronic changes and synchronic derivations. SM or PM approximations have often been taken for granted as the underlying forms from which the phonetic realisations of NEPMs are derived using a set of convoluted rules. Adi Yasran (2005; 2010) analyses the consonant and vowel inventories of KM and formulates the derivation of KM surface forms within the framework of Optimality Theory. However, his analysis lacks justification for the underlying forms of the words being considered. For example, KM [ayɛ] 'chicken' is taken as having derived from the underlying form /ayam/ (cf. SM *ayam*) with the application of phonological rules including vowel raising and nasal deletion. Yet there is no reason to assume that the underlying form of KM [ayɛ] is /ayam/, and a more straightforward analysis would be to simply take /ayɛ/ as the underlying form. Similarly,

Sharifah Raihan (2018) discusses the realisation of consonant clusters consisting of a nasal + a voiceless obstruent in KM and other Malay dialects. She takes KM [gatoŋ] 'to hang' as the surface realisation of the underlying form /gantoŋ/, whereby /n/ in /-nt-/ is deleted in root-internal position. However, the foundation of such an analysis is also problematic: /gantoŋ/ is merely the earlier form from which KM [gatoŋ] developed *diachronically*, which happens to have survived in SM. In the synchronic sound system of KM, there is no evidence that nasal-voiceless obstruent clusters exist at all, even at the underlying level. Similar problems can be found in other publications on NEPMs (as well as other Peninsular Malay varieties), such as Zaharani (2006), Zaharani et al. (2011), Adi Yasran (2011, 2012), Nur Adibah & Sharifah Raihan (2017) and Mohd Tarmizi (2018a, c) and Nor Hashimah, Wan Athirah & Harishon (2021).

The manifestation of this issue culminates in the compilation of glossaries such as *Glosari dialek Terengganu* [A glossary of the Terengganu dialect] (1997). For each CTM word, a standardised spelling that resembles SM orthography is given, which mirrored earlier British linguists' practise of documenting Malay dialects. For instance, [igaʔ] 'to catch' and [tɛpɛʔ] 'to stick' are written as ⟨igat⟩ and ⟨tempek⟩ respectively, which seems to suggest that word-final /t/ and medial consonant sequence /-mp-/ are phonemic. This forced system is fortunately abandoned in *Glosari dialek Kelantan* [A glossary of the Kelantan dialect] (2016), another glossary in the same series that was published about two decades later. Here we can find [sɔʔmɔ] 'always' and [blɛ-blɛ] 'whilst' written as ⟨sokmo⟩ and ⟨ble-ble⟩ (instead of the potentially standardised spelling ⟨belan-belan⟩ which can be found in *Kamus Dewan* [The institute dictionary], Sheikh Othman 2007).

The issues discussed above highlight the need for more descriptive studies of NEPMs. Only then can we gain a better understanding of these languages, including their synchronic systems and diachronic development.

1.6 Present study

The present study has two goals. The first goal is to provide a description of NEPMs by gathering and analysing new data, with a focus on the phonology and morphology of NEPMs, while also providing a concise description of their syntactic structure. Based on synchronic descriptive facts, the second

goal is to establish the phonological and morphological changes that have taken place in NEPMs.

The data for this study were collected during two field trips to Kelantan and Terengganu, conducted from July to November 2018 and August to October 2022. These field trips were a part of the visiting studies at the University of Malaya, at the Academy of Malay Studies in 2018 and at the Faculty of Languages and Linguistics in 2022. A third field trip was originally planned for 2020 but was unfortunately cancelled due to the Covid-19 pandemic.

This section elaborates on the methodology in the present study, describing the fieldwork locations and native speaker consultants in §1.6.1 and explaining data collection and processing in §1.6.2. A summary of transcription conventions follows in §1.6.3.

1.6.1 Fieldwork locations and native speaker consultants

As NEPMs are spoken across an extensive area, the selection of field sites had several considerations. Firstly, the state capitals Kota Bharu and Kuala Terengganu were excluded as they serve as the first entry points for immigrants to these states, and thus are more likely to have received more external influences. For similar reasons, villages were preferred over local towns. Secondly, field sites should be preferably chosen to align with typical Malay settlement patterns. In Kelantan, the river systems play a crucial role in shaping the settlement pattern, while in Terengganu, both the river systems and coasts are important factors (see §1.4.1.1). Finally, feasibility and practicality were also taken into account, with preference given to locations where it was easier to establish relations through my contact network. With these considerations in mind, the following locations were chosen as primary field sites, as shown in Table 1.2 and Figure 1.6.

Table 1.2: Overview of field sites

Variety	Primary field site	Coordinates	District
KM	Kampung Kusial Bharu	5°45'N, 102°08'E	Tanah Merah
CTM	Kampung Gong Sentul	5°20'N, 103°06'E	Kuala Nerus
ITM	Kampung Dusun	5°04'N, 102°56'E	Hulu Terengganu

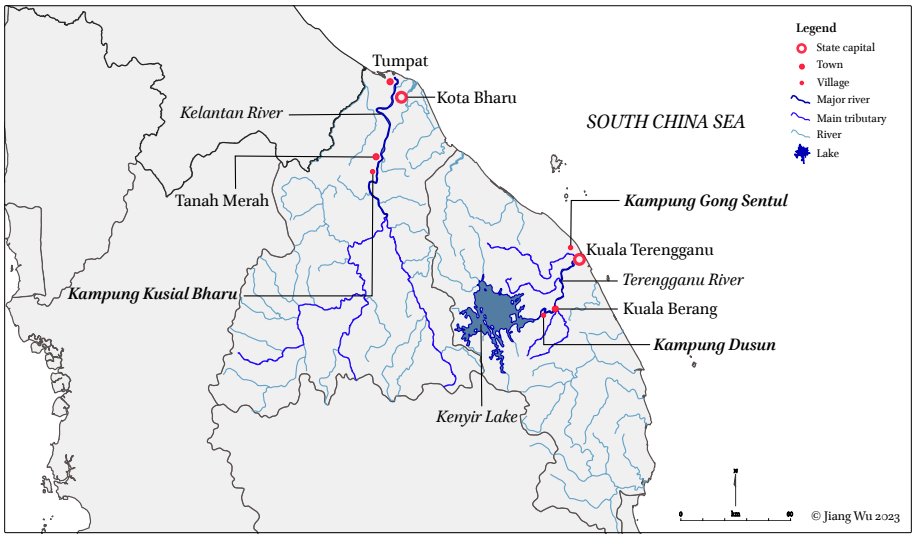

Figure 1.6: Locations of field sites

Data on KM were primarily collected in Kampung Kusial Bharu (*kampung* is Malay for 'village') in the district of Tanah Merah, which is situated in the Kelantan River basin, some fifty kilometres south of the state capital Kota Bharu. The village of Kusial Bharu is on the west bank of the Kelantan River. In addition, several visits were made to Kampung Palekbang in Tumpat.

CTM data were collected in Kampung Gong Sentul in the district of Kuala Nerus, which is located close to the confluence of the Nerus River and the Terengganu River. The town of Kuala Nerus is situated along the coastline of Terengganu, between the city of Kuala Terengganu and the state's airport, approximately seven kilometres from the city centre. Formerly a part of the Kuala Terengganu district, Kuala Nerus is now densely populated with many small villages in close proximity to each other, Kampung Gong Sentul being one of them.

ITM data were collected in several villages across the district of Hulu Terengganu given the vast intra-dialectal variation of ITM (see Collins 1983a). The primary field site is Kampung Dusun, and other villages visited include Kampung Tanjung Baru, Kampung Payang Kayu and Kampung Pasir Nering, where comparative lexical data were collected (see Figure 1.7).

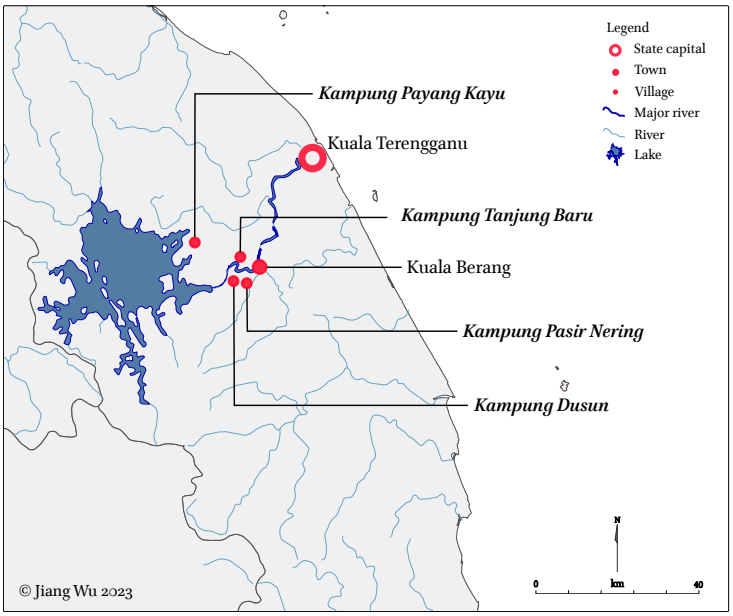

Figure 1.7: Locations of field sites in Hulu Terengganu

In comparison to Kuala Nerus, Hulu Terengganu is more sparsely populated, and the distribution of villages in this region clearly follows the riverine system. Kampung Dusun is a village on the bank of the Terengganu River in its upper reach, approximately ten kilometres from the town of Kuala Berang and fifty kilometres from the capital Kuala Terengganu. Hosting the oldest primary school and one of the oldest mosques in Hulu Terengganu, the village has apparently been of great importance.

In selecting the consultants, the rule of thumb was that non-mobile, older, rural females were preferred. This differs from the NORMs principle proposed by Chambers & Trudgill (1998: 29) for dialectology studies, which prioritises non-mobile, older, rural *males*. The preference for females over males was based on the consideration that males in traditional Malay villages tend to be more mobile than females, and thus are more likely to have received more external influence on their language. Many older men have worked in other states before returning to their hometown, leaving their families in the villages. Furthermore, men generally have higher social status and more interaction with outsiders. Younger generations also tend to move to bigger cities for study or work. Despite the preference for choosing older

consultants, data from some younger speakers were also collected, typically during elicitation sessions, which proved to be easier with younger speakers.

1.6.2 Data collection and processing

The data collection process followed the guidelines outlined by the European General Data Protection Regulation (GDPR, available at https://gdpr-info.eu/), whereby the consultants' consent was obtained before recording their speech. The linguistic data collected can be broadly classified into six types, as listed in Table 1.3. The table also presents an overview of the amount of raw data collected for each variety. Each type of data is briefly described below.

Table 1.3: Overview of raw data

Data type	KM	CTM	ITM
Word lists	≈4 hours	≈3,5 hours	≈5 hours
Narratives	≈20 minutes	≈45 minutes	≈70 minutes
Elicitation	≈6 hours	≈3 hours	≈6,5 hours
Conversations	≈2,5 hours	≈1 hour	≈6 hours
Discussions and interviews	≈1,5 hours	≈45 minutes	≈3 hours
Unrecorded field notes	N/A	N/A	N/A

The first type is word lists, which consist of a basic word list of 260 items, a modified version of the Swadesh 200 word list, and a supplementary list of 309 items. Both lists were recorded for all three varieties, and the basic word lists collected for each variety can be found in Appendix A. Additional word lists focusing on more specific phonological phenomena or words in particular categories were recorded as needed. These lists served as the basis on which phonological analyses were conducted.

The second type consists of narratives. Three standard stories were recorded with the aid of picture books and video clips: *The frog story* (Mayer 1969), a modified version of *The chicken thief story* (Rodriguez 2009) and *The pear story* (Erbaugh 2001, available at http://pearstories.org/). Some of these stories were recorded with multiple consultants. It is acknowledged that narratives obtained with visual stimuli may be less natural than free

narratives (see, e.g., Klamer & Moro 2020). Attempts were also made to record free-style storytelling, but they were unfortunately unsuccessful with most consultants, which could be attributed to two reasons. First, there is not a strong tradition of storytelling in the field sites I visited, and perhaps not in contemporary Malay villages on the peninsula in general. Second, the intermediate language used for data collection was SM, which is closely related to the vernaculars being investigated. The consultants often switched to or mixed their vernaculars with SM when telling free-form stories, which for them was not a natural setting.[9]

The third type of data is from elicitation sessions, which involved various tasks ranging from sentence translations, making sentences using the given words, to elicitation with video stimuli such as the Surrey clips (Fedden et al. 2010) and the *Give*-events clips recorded by Moro & Fricke (2020).

The fourth type of data comes from naturalistic conversations in the vernaculars between two or more interlocutors. Unlike the first three types of data, which were recorded in both video and audio formats (.mp4 and .wav respectively), conversations were only recorded using the audio recorder. Given the spontaneous nature of this type of recording, no videos were recorded so as to minimise potential influences of an artificial interview setup on the natural flow of the conversations.

The fifth type of data includes interviews of the speakers' sociolinguistic background and other types of discussions, such as those on the usage of certain words and constructions. These were typically recorded in SM.

Lastly, during the fieldwork, some observations and notes were made on the spot and written down in notebooks, but not recorded. These belong to the category of unrecorded field notes.

The duration of fieldwork in each site varies for various reasons, which results in differences in the amount of data collected for each NEPM variety. I spent approximately five weeks in Kelantan, four weeks in Coastal Terengganu, and seven weeks in Hulu Terengganu. The largest amount of data has therefore been collected for ITM, whereas the least amount is available for CTM (see Table 1.3).

[9] The close relatedness between the intermediate language and the target language posed a major challenge in data collection. Consultants were sometimes invited to listen to their own speech and point out parts that might have been influenced by SM. Other challenges and difficulties of data collection and processing have been explained in Klamer et al. (2021: 489–491).

Recorded data were transcribed by myself or with the assistance of consultants or student assistants. Word lists were transcribed in Microsoft Excel, some elicitation sessions were transcribed in Microsoft Word, while other recordings were transcribed using the linguistic annotation tool ELAN (The Language Archive, Max Planck Institute for Psycholinguistics, available at https://archive.mpi.nl/tla/elan). The Excel transcriptions (.xlsx) were exported as database files for Standard Format Lexicon (.db), while annotated ELAN files (.eaf) were exported as FLEXTEXT files. Both types of files were then imported into the software FieldWorks Language Explorer (FLEx, SIL International, available at https://software.sil.org/fieldworks/), which provides a corpus-building platform for interlinear glossing of texts and morphological analysis while expanding a lexical database. For each NEPM variety, a corpus was built in FLEx. Translations were done either in ELAN during annotation or in FLEx. Finally, the glossed and translated transcriptions were reimported into ELAN and linked with the recordings. Due to time limitations and the general difficulty in transcribing naturalistic conversations, as well as the varying quality of recordings, not all conversations have been transcribed. Appendix B provides three transcribed sample texts (one for each variety) from different types of recordings, which serve to illustrate the morphosyntactic structures of NEPMs.

In handling the recordings, file naming follows the convention outlined in (1). A file name starts with the language abbreviation, followed by the date of recording, the type of data and its number. The following abbreviations are used for each type of data: "wl" for word lists, "n" for narratives, "e" for elicitation, "cv" for conversations, "d" for discussions and interviews, and "fn" for unrecorded field notes. Additional information such as the content of the recording is sometimes provided, followed by the format of files if recordings were made in multiple formats ("a" for audio and "v" for video). For example, a file with the name "ITM_180907_n01_frogstory_a" shows that it is an audio recording of a frog story (which is a narrative, and the first narrative recorded on that day), recorded on 7th September 2018 in ITM.

(1) LANGUAGE ABBREVIATION_DATE OF RECORDING_TYPE OF DATA AND NUMBER_ADDITIONAL INFORMATION_FORMAT OF FILES

A metadata sheet was created to document the list of sessions, recordings and native speaker consultants. All types of raw data and processed data, as well as the metadata sheet, have been archived in the *Corpora of Kelantan*

Malay, Coastal Terengganu Malay and Inland Terengganu Malay (Wu 2023), available at https://doi.org/10.34894/HWUVLM.

1.6.3 Transcription conventions and citation codes

Linguistic examples in the present study are transcribed using the International Phonetic Alphabet (IPA) with slight modifications. For easier transcription and to align with the tradition in the field, I opted to use /y/ instead of the IPA symbol /j/ for the palatal glide, and /c, ɟ/ instead of IPA /c͡ç, ɟ͡ʝ/ to represent palatal affricates. For example, KM /ayɛ/ 'chicken' and /jaɣi/ 'finger' should be read as /ajɛ/ and /ɟ͡ʝaɣi/ respectively in standard IPA. During initial data processing, words were transcribed in broad IPA following their pronunciations at the phonetic level. After conducting phonological analyses, transcriptions were rendered in phonemic forms. As each NEPM variety has a different phonological system, no attempt has been made to create an orthography for each variety. Linguistic examples are primarily cited in their phonemic transcriptions to facilitate cross-dialectal comparisons, which are either enclosed in forward slash brackets "//" (as practised in the phonology chapters in order to differentiate phonemic forms from phonetic forms, which are given in square brackets "[]"), or simply in *italics* (as practised in other chapters).

When citing example sentences, I provide a slightly shortened name of the corresponding recording as the citation code, which follows the free translation. For instance, the example with the citation code "KM_180812_n01_12" indicates that it is a KM example from line 12 of a narrative recorded on 12th August 2018. The corresponding recording can be found in the corpora by searching for the file name starting with "KM_180812_n01". Morpheme-by-morpheme interlinear glosses are provided following the Leipzig Glossing Rules, with my own additions where necessary. An exception has been made in the transcription of reduplicated forms, where I use a hyphen "-" to link the two morphemes instead of the prescribed tilde "~", which is reserved for indicating free variation between two (or more) forms. A list of standard abbreviations for the glossing of grammatical categories, along with additional abbreviations and a symbol usage guide, can be found in the front matter on pages xvii–xx.

1.7 Structure of this dissertation

The present dissertation is structured into two main parts that align with the research goals.

Part one provides a synchronic description of NEPMs, comprising five chapters. Chapters 2 to 4 focus on the phonology of KM, CTM and ITM respectively, as this is the area where these varieties exhibit the most pronounced differences. Each chapter follows a parallel structure, starting with an examination of the segment inventory, and then building up towards syllable and word structure, as well as phonotactic constraints, before concluding with a brief discussion of stress patterns. Chapters 5 and 6 describe the morphology and syntax of NEPMs. Given the similarities across the morphosyntax of NEPMs, a comparative approach is taken in these two chapters.

Part two delves into the historical development of NEPMs. It consists of two chapters (Chapters 7 and 8), which explore the historical phonology and historical morphology respectively. Building on the synchronic analysis in Part one, these chapters draw comparisons between NEPMs and the preexisting reconstructions in the common ancestral language (in the present case PM) and establish sound changes and morphological changes that have taken place. In these chapters, I show that NEPMs have undergone remarkable sound changes and significant morphological reduction. I also discuss the potential factors contributing to these changes in the history of NEPMs, briefly examining the role of language contact. By analysing the chronological order of sound changes, I argue that NEPMs cannot form a lower-level subgroup within Malayic, despite superficial similarities in their sound patterns. Furthermore, there is no clear evidence of substratal influences, and the morphological reduction was primarily driven by phonological changes.

The main findings of this dissertation are synthesised and summarised in Chapter 9. In addition, I propose a hypothesis regarding the migration patterns of the speakers, combining data from linguistics and historical records. It is evident that ITM stands out as the most divergent and conservative variety among NEPMs, suggesting that its speakers may have a longer history when compared to the coastal population. Speakers of KM and CTM, on the other hand, likely represent a more recent migration, possibly occurring post-Malaccan times. Lastly, I discuss some limitations of the present study and suggest directions for future research.