THE ORIGINS OF EMPTY CATEGORIES IN SINGAPORE ENGLISH

Bao Zhiming

National University of Singapore

The system of empty categories in Singapore English, a contact language with an endogenous ecology, arises through the interaction of three parameters: [topic-prominence], [pro-drop], and [wh-movement]. These parameters are re-set under the pressure of the languages in the contact ecology, mainly the substrate Chinese dialects, and the lexifier English. The paper adopts a holistic approach to creole genesis, in which substrate and superstrate influence is expressed in terms of parametric re-structuring constrained by principles of Universal Grammar. Surface-true substrate, superstrate, or novel features are exponents of this parametric re-structuring.

KEYWORDS: pidgins/creoles, creole genesis, universals, substratum, superstratum, Singapore English, Chinese

The Issue

For the past two decades or so, the literature on creole genesis has been dominated by the debate between the universalists and the substratists. The universalists argue that the process of pidginization and creolization is driven by linguistic universals that are shared by many of the world’s extant languages. The strongest proponent among them is Bickerton, who, influenced by Chomsky’s mentalistic conception of language and grammar (Chomsky,
1965, 1986), has argued since the 1970s that creoles emerge among first-
generation speakers under extreme linguistic conditions — severe paucity
of input data for the child. Hence, they reflect the unmarked choices of the
1999). The substratists, while not denying the importance of linguistic uni-
versals, place the substrate languages of the contact community at the core
of their explanation of creole genesis. There are scholars, notably Mufwene
(1986), who argue that the universalist and substratist approaches are not
incompatible. Indeed, they complement each other.2

There are two important components of Bickerton’s universalist position:
the principles-and-parameters theory of UG, and the assumption that the
UG parameters assume default settings in the absence of a sufficiently rich
acquisitional environment. It is the latter component that proves controversial.
While the universalists typically see creolization as first-language acquisition
by children, the substratists see it as an instance of imperfect second-language
learning by adults. There is, however, no a priori reason to support this
conceptual and explanatory alignment. One can adopt the parametric theory
of UG without theoretical and empirical commitment to the exact mode of
acquisition. This is especially pertinent for pidgins and creoles (henceforth,
PCs) that arose in non-plantation, non-insulate contact communities. Under
such socioeconomic and linguistic conditions, PCs are dynamic languages to
which both native-born children and immigrant adults contribute grammatical
systems or sub-systems, and reinforce them at the same time. The grammar
of an emerging PC draws features from the languages in the contact situation
in accordance with its parametric settings. These settings arise under the
influence of UG, the superstratum, or the substratum of the contact ecology

2 There is a large body of literature on the roles of UG, linguistic universals, substratum,
and superstratum in creole genesis, and on the parallel between pidginization and creolization
on the one hand, and child or adult language acquisition on the other. For a glance at these
issues, see Schuchardt (Gilbert, 1980, Alleyne 1971, 1979), Kay and Sankoff (1974), Le Page
Roberts (1998), Ham (1999), and DeGraff (1999).
THE ORIGINS OF EMPTY CATEGORIES IN SINGAPORE ENGLISH 277


In this paper, I consider the empty categories of Singapore English, a PC with a constant linguistic ecology throughout its short history. The system of empty categories in Singapore English is determined by three factors: pro-drop, topic-prominence, and optional wh-movement. Unlike typical pro-drop languages such as Spanish and Italian, where pro is licensed by agreement markers on the verb, Singapore English is a pro-drop language by virtue of its non-inflectional morphology, which is a universal feature of PCs.3 This affects the interpretation of subject null pronouns, as we shall see. Although topic prominence is also a universal PC feature, at least statistically (cf. Bickerton, 1981; Siegel, 1997), I will show that the topic structure of Singapore English is derived from Chinese, the main substrate language.4 As for wh-question formation, wh-movement is obligatory in English, optional in Singapore English, and not allowed in Chinese. (It is, however, optional in Malay, a minor substrate language.) In Singapore English empty categories, we see the effect of the interplay of the substratum, the superstratum, and universals. It is inconceivable that such an intricate grammatical and interpretative system can be the result of a loose assemblage of null elements — accidental mixing, for example, of English variable with Chinese pro. The empty categories of Singapore English arise in a principled manner. The so-called “cafeteria principle,” if unconstrained by principles of UG, needs

---

3 As a universal feature of PCs (cf. Kay & Sankoff, 1974; Givón, 1979; McWhorter, 1998), lack of inflection could only be interpreted as a statistically significant and unmarked tendency. In any case, it is true only of surface observations. Theoretically, one may postulate AGR as part of INFL, with agreement features on par with surface true inflectional languages; see Lefebvre (1996) and DeGraff (1996) for analyses of Haitian Creole AGR, which, like other PCs, is not realized morphologically.

4 A brief note on the terms “Singapore English,” “English,” and “Chinese” is in order. I use the term “Singapore English” to refer to the basilectal and mesolectal varieties of vernacular English spoken in Singapore, see Section 2. The term “English” or “native English” refers to native varieties of English associated with the inner-circle countries of Kachru (1985). I shall make no distinction between standard English and colloquial English. The term “Chinese” is a cover term for the mutually unintelligible dialects which share a common written language. The linguistic difference among the dialects lies mostly in phonology, with negligible differences in syntax and core vocabulary. The major Chinese dialects represented in Singapore originate from the southern provinces of Fujian and Guangdong; see the sections entitled “Languages of the Immigrants” and “Lingua Franca of the Straits Settlements.”
to be rejected as a viable explanation of the origin of Singapore English empty categories (cf. Bickerton, 1981; Mufwene, 1996; Siegel, 1997). I argue that the system of empty categories results from the typological shift that Singapore English has undergone — Singapore English and Chinese are both pro-drop and topic-prominent. English, by contrast, does not allow pro-drop and is subject-prominent. Under the assumption that typological distinctions are expressed by UG parameters, we can view substrate influence (or superstrate influence) in this way. The substratum or superstratum does not directly affect surface grammatical features of an emerging PC; instead, it affects parametric settings, and surface-true grammatical features are selected from the languages in the contact environment under the pressure of the new parametric arrangement, subject to the constraint of UG principles (cf. Mufwene, 1990a, 1990b, 1991, 1996). Here, we will focus on three parameters: [wh-movement], [topic prominence], and [pro-drop]. The settings of these parameters give rise to the system of empty categories in Singapore English.

The rest of the paper is organized as follows. In the following, I discuss the sociohistorical background of Singapore English and argue that continuous English dominance and immigration of Chinese-speaking people to Singapore makes it imperative to view Singapore English as a dynamic system, in the sense of Mühlhäusler (1986), with convergent contributions from (imperfect) second-language learning by new adult immigrants and first language learning by native-born children. In the third section, I outline the system of empty categories proposed by Chomsky (1981, 1995), without the technical jargon of the principles-and-parameters theories. I classify the empty categories in terms of their positions in the clause, and the positions of their antecedents or operators. Readers who are familiar with Chomsky’s theory of empty categories may wish to skip this section. In the fourth section, I discuss the salient properties of topic structure in Chinese, highlighting important similarities and differences in topic structure between topic-prominent languages and their subject-prominent siblings. The clustering of properties is linked to the parameter [topic-prominence] (cf. Li & Thompson, 1976; Huang, 1984). In the fifth section, I examine the properties of empty categories in Chinese, English, and Singapore English, and topic structure in Singapore English. The final section is the conclusion.
The typological status of Singapore English, or Singapore Colloquial English, is controversial. Broadly speaking, the controversy divides into two camps: those who think Singapore English is a PC, and those who do not. Platt (1975) labels it a creoloid, which has creole-like features but lacks the usual pidgin as its predecessor. In the same general direction, Arends, Muysken, and Smith (1994) classify it as an extended pidgin. The difference between creoles and extended pidgins is a matter of definition. Structurally, the two are indistinguishable (cf. Mühlhäusler, 1986). The opposing camp includes Kachru (1985), and Pakir (1991), among others, who consider Singapore English as a New English, a non-native variety which has been indigenized, or institutionalized, in its adoptive community. Whatever the label — Singapore Colloquial English, creoloid, New English, indigenized English, or institutionalized English — it is an undisputed fact that the grammar of non-native varieties in general, and Singapore English in particular, exhibits the profound effect of contact-induced change. The taxonomic controversy underscores the fluidity of the state of the English language in Singapore. It is plausible that the opposing camps are concerned with different “lects” of the English continuum — the PC camp with the basilectal or mesolectal variety, and the New English camp with the acrolectal variety. I use the term “Singapore English” to refer to the basilectal or mesolectal variety one hears spoken informally in Singapore. This variety, often derisively labeled Singlish, is among the languages that children acquire (cf. Kwan-Terry, 1986; Gupta, 1991 1994). In this section I describe the historical and sociolinguistic characteristics of the contact ecology that give rise to (and encourage the development of) Singapore English.

Chaudenson (1977) distinguishes between exogenous PCs and endogenous PCs in terms of two socioeconomic factors: the economic activity of the lexifier-speaking group, and the geographical location of the contact community. The exogenous PCs arise in so-called plantocratic communities, which are re-constituted for the purpose of agriculture, with linguistically diverse groups of slaves or indentured laborers, far removed from the geographical origin of any of the constituent groups. The endogenous PCs, by contrast, arise in contact communities in which the language of the indigenous population is used, and the dominant economic activity of the lexifier-speaking group was commerce, rather than agriculture. Linguistically, however, the
commercial or agricultural activities of the contact groups are not as important as the presence of the main substrate language(s) in the contact environment. Endogenous PCs arise in communities in which the main substrate language(s) are spoken alongside the emerging PCs, while exogenous PCs arise in communities that do not have a common, numerically significant, substrate language. For ease of reference I will call the ecology that gives rise to endogenous PCs the endogenous ecology, and that which gives rise to exogenous PCs, the exogenous ecology.

The linguistic ecology in which Singapore English is developing is very different from plantation ecologies that give rise to exogenous creoles that have long occupied the attention of creolists. I will consider three factors which make this ecology endogenous: population change in historical context, the local languages of the main segments of the population, and the lingua francas of Singapore’s multi-lingual and multi-ethnic community.

An Immigrant Community

When Stamford Raffles annexed it in 1819 as a trading post from the Sultanate of Johore, Singapore was a small island at the southern tip of the Malayan peninsula with about 150 fishermen and pirates living in make-shift huts (Newbold, 1839, p. 279). With the encouragement of the British colonial government, immigrants started pouring in rapidly from the region, that is, from peninsular Malaya such as Malacca and Penang, from the Indonesian archipelago, and from the southern Chinese coastal provinces of Fujian and Guangdong. Many of the early settlers of Chinese descent who migrated from the region were already familiar with European commercial practices and with local indigenous cultures. The lure of a new trading post was too strong to resist. Table 1 contains the census figures of the migrants who came to Singapore between 1824 and 1836, obtained from Newbold (1839).

5 Among the three main linguistic groups, Chinese, Malay, and Indic languages, the latter two made a negligible contribution to the grammar of Singapore English (cf. Platt & Weber, 1980). For this reason we will focus on the ethnographic profile of the Chinese community. Linguistically, the Malay community is quite homogeneous; if there is any dialect variation, the variation does not cause problems in intelligibility. The Indian immigrant community is by no means linguistically homogeneous. Some speak Hindi, an Indo-European language; others speak the Dravidian languages Tamil and Malayalam. Tamil speakers form the largest segment of the Indian community.
In 1836, as Table 1 shows, the numbers of Chinese and Malay were about equal. Nevertheless, since the 1840s the Chinese have constituted the majority of Singapore’s population. Table 2 gives the proportion of Chinese, Malays, and Indians between 1840 and 1980 in Singapore (Pan, 1998, p. 200).

The 1980 distribution still holds today. Given the numerical dominance of Chinese-speaking people, it is not surprising that the grammar of Singapore English has recognizable Chinese influence.

Singapore has never had a plantation economy of the type that characterized the relatively insular island communities in the Caribbean and Hawaii. In the first 20 years after the annexation of Singapore, the ethnic groups were engaged in various economic activities: the Europeans in commerce, the Chinese in various trades, and Malays in fishing and agriculture (cf. Newbold, 1839). The nascent economy of Singapore was not dependent on plantations, but on trade involving tropical spices such as nutmeg, cloves, and pepper — and subsequently on entrepôt trade.
In 19th century and early 20th century Singapore, the formative period of Singapore English, the population was very fluid, especially among the Chinese community (Xu, 1950). Not only was there constant movement of people, there was also an enormous imbalance between the sexes among Europeans, Indians, and especially Chinese. The 1849 census, for example, recorded only one Chinese woman for every 17 Chinese men in Singapore (Jackson, 1850). This problem was alleviated towards the end of the 19th century, with relaxation of emigration in China, and the improved financial means of the bachelor migrants (Turnbull, 1977, p. 58). Economic reality in Singapore (and the region) and political upheaval in China at the turn of the century forced many of the early sojourners to settle permanently. As more and more sojourners settled down and started families, the Straits-born Chinese community grew in number. In the 1840s, Siah (1848) counted 1,000 Chinese who were born in the Straits Settlements, about 2.5% of the total Chinese population at the time. By the 1880s, the percentage of Straits-born Chinese grew to 9.5%. Immigration started to decline after the Second World War, due to tight government control. By 1990, immigrant Chinese constituted a “mere seven percent of the resident Chinese population (Pan, 1998, pp. 201–202).” Although most Straits-born Chinese continued to speak the dialects of their parents, many picked up some Malay, and English. Exposure to the multilingual environment of Singapore and Malaya tends to dilute the effect of the languages spoken by newly arrived immigrants.

The Languages of the Immigrants

Linguistically, the Chinese in pre-modern Singapore spoke a plethora of mutually unintelligible dialects. The two largest dialects are both Southern Min dialects: Hokkien, spoken in the region around the city of Xiamen (Amoy in the vernacular) in Fujian Province, and the closely-related Teochew, which is spoken in the border area of southern Fujian Province and northern

---

6 In the literature on Chinese linguistics, seven major dialects are recognized. They are: Mandarin, Wu, Southern Min, Northern Min, Cantonese, Hakka (Kejia), and Gan (Jiangxi Province) (cf. Norman, 1989; Yuan et al., 1989). Most dialects spoken in Singapore belong to Southern Min (Hokkien and Teochew) and Cantonese.
Table 3. Distribution of Major Chinese Dialects

<table>
<thead>
<tr>
<th>Dialect</th>
<th>1840s</th>
<th>1881</th>
<th>1891</th>
<th>1901</th>
<th>1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hokkien</td>
<td>22.7</td>
<td>28.9</td>
<td>37.6</td>
<td>39.2</td>
<td>42.2</td>
</tr>
<tr>
<td>Teochew</td>
<td>47.9</td>
<td>26.2</td>
<td>19.5</td>
<td>18.3</td>
<td>21.9</td>
</tr>
<tr>
<td>Cantonese</td>
<td>15.0</td>
<td>17.2</td>
<td>19.2</td>
<td>20.3</td>
<td>15.3</td>
</tr>
<tr>
<td>Hakka (Kejia)</td>
<td>10.1</td>
<td>7.1</td>
<td>6.1</td>
<td>5.6</td>
<td>7.3</td>
</tr>
<tr>
<td>Hainanese</td>
<td>1.8</td>
<td>9.6</td>
<td>7.1</td>
<td>6.3</td>
<td>7.0</td>
</tr>
<tr>
<td>Malacca-born</td>
<td>2.5</td>
<td>11.0</td>
<td>10.5</td>
<td>10.3</td>
<td>–</td>
</tr>
</tbody>
</table>

Guangdong (Canton) Province. Table 3 summarizes the distribution of the major dialects spoken in Singapore between the 1840s and 1990. The Malacca-born Chinese are the descendants of early Hokkien (Fujian) immigrants who migrated to Malacca centuries before the establishment of the British Straits Settlements (Vaughan, 1879; Purcell, 1948). Except for a few words of Hokkien origin, they did not speak the language of their forebears. Their mother tongue was Baba Malay, a creole with a Hokkien substrate and a Malay vocabulary base (Shellabear, 1913; Pakir, 1986). They were among the first group of people in Singapore to switch to English not only as a business language, but also as a home language (cf. Tan, 1988).

Few Chinese emigrated from Mandarin-speaking areas of China to Singapore, and to the region known collectively as Nanyang, or South Sea. Mandarin therefore was not a dialect spoken by early immigrants. It was, nevertheless, the medium of Chinese education. Mandarin-speaking teachers were specially brought to the Straits Settlements to teach in Chinese-medium schools, but their number was not significant. At the turn of the century some English-educated Chinese intellectuals in Singapore, alarmed at the level of ignorance of Chinese culture among Straits-born Chinese, started to advocate Mandarin learning in the Chinese community (cf. Kiong, 1907). As Mandarin-medium education spread in Malaya, especially after the Chinese Revolution of 1911 that overthrew the Qing Dynasty (1644–1911), more and more people were able to speak Mandarin (Purcell, 1948). Although it

---

7 The 1840s figures are obtained from Siah (1848), the 1990 figures from Pan (1998), and the rest from Lee (1978). The 1990 figures do not have the category “Malacca-born.” The remaining 6.3% of resident Chinese population speak a dialect other than those listed in the table.
was not a “mother” tongue of the early immigrants, it was, and still is, an important component of the contact ecology of Singapore English.

After Singapore became an independent nation in 1967, Mandarin-medium education, after years of decline, was abandoned in favor of English-medium education, and in the 1970s the government of Singapore started the annual “Speak Mandarin” Campaign among Chinese Singaporeans. Today, Mandarin is one of the four official languages of Singapore, along with Malay, Tamil, and English, and it is the most common dialect of Chinese spoken in Singapore, especially among the younger Singaporeans. The indigenous dialects are rapidly dying out.

Although the major Chinese dialects are mutually unintelligible, they differ mainly in phonology. In addition to a common syntax, they share a common written language and a common core lexicon (cf. Yuan et al., 1989). As far as the genesis and development of Singapore English are concerned, the dialects form a homogeneous substratum, and they collectively exert their influence.

**Lingua Franca of the Straits Settlements**

The multi-ethnic, multi-lingual community of Singapore, and the Straits Settlements, provided an excellent environment for the emergence of pidgins and creoles. Unfortunately we do not have written documents about the state of English use among the local population in 19th-century and early 20th-century Singapore. The extant written documents — government records, newspapers, and magazines — are written in standard native English. Consequently, they do not reflect the actual state of the language among the local population. However, from personal reminiscences and accounts written by government officials, travelers, and explorers at the time, we can piece together a likely scenario of the linguistic ecology that gave rise to early Singapore English.

Early Singapore was divided into geographically separate enclaves based on ethnic and dialectal affiliation. Communication within each enclave was based on the language of that enclave — Malay, Tamil, Hokkien, Cantonese, and so on. Outside the enclaves, the lingua franca was Bazaar Malay, a Malay-based pidgin, of which Dennys (1878b) is a valuable description. Although Bazaar Malay was the “universal” language in Malaya, English-based pidgin was emerging to replace it as the lingua franca of the Straits Settlements, as is evident from descriptions by contemporary travelers and residents such as
Thomson (1864), Dennys (1878a), and Vaughan (1879). Dennys (1878a) was a brief description of Pidgin English that was the lingua franca of foreign trade spoken mostly in the southern coastal areas of China. The few grammatical features mentioned in that work are identical with those of the Chinese Pidgin English described by Hall (1944). According to Dennys, this Pidgin English spread to Malaya, including Singapore, especially after the Taiping Rebellion of 1850. Undoubtedly there were Pidgin English-speaking traders among the immigrants to the Straits Settlements — after all, one of the functions of the Settlements was to protect the China trade (Lee, 1978). It is equally likely that their number was small, since most immigrants were illiterate laborers trying to carve out a living in the new land for themselves and their families.

Even without external input, one would have expected contact languages to emerge from the linguistic and cultural mosaic of the Straits Settlements. Malacca, a port about 200 kilometers north of Singapore, had been a Portuguese and Dutch colony before the British took control and incorporated it into the Straits Settlements in 1824. The existence of a Portuguese-based pidgin did not escape the notice of 19th-century travelers, and Pidgin English was also making its appearance in Malacca, by now part of the British Straits Settlements (Thomson, 1864; Wallace, 1869). Malacca-born Chinese were the first immigrant settlers in Singapore, bringing, in the words of Thomson, their “broken-English-Dutch-Portuguese” along with them.

From the historical accounts, it is certain that a pidginized form of English, either local or foreign in origin, preceded the present-day Singapore English. English-medium schools were also instrumental in the spread of English, a point much emphasized by Platt and Weber (1980) and other scholars of Singapore English. But mass education in the English language in Singapore did not start until after the Second World War, and started in earnest only after independence in 1965. In the 19th century and the first half of the 20th century, English education was a luxury. Only the elite of the mercantile class had the financial means to send their children to English-medium schools such as Raffles Institution, which was established in 1823, and to England and America for higher education. Nevertheless, the political and commercial dominance of English in Singapore, the capital of the Straits Settlements, was a strong motivating force for people to acquire some knowledge of English. Most people, if they knew English, acquired it without the benefit of formal instruction. Whether acquired in school or on the street, Chinese influence on the English language being acquired did not escape the
notice of government officials. One official by the name G. T. Hare gave a speech in 1897 to a civic society in Singapore, in which he lamented the poor mastery of English of the “youthful citizen”: He “knows neither Chinese nor English well. He has a smattering of both, but an intelligent understanding of neither” (Hare, 1897, p. 6). It is fair to assume that Hare was referring to the contact features in the English of the citizens of the Straits Settlements. One hundred years later, the situation has hardly improved. The government and the local news media are as concerned with the widespread use of “Singlish” in today’s Singapore as Hare was with the local English of his time. A “Speak Good English” campaign is now under way.

Over the years Singapore has received a continuous stream of Chinese-speaking immigrants in varying numbers, and has experienced continuous pidginization and creolization as successive generations of immigrants and their children settled down and adapted to the new cultural and linguistic environment. The endogenous ecology of the 19th century has changed little in the 20th century: a homogeneous Chinese substratum for the majority of the population, with English as the prestige language performing the full load of high-level communicative functions of modern Singaporean society. Certainly, due to English-medium education, made mandatory in the 1980s, more and more people are able to speak English, albeit at various levels of proficiency (cf. Pakir, 1991). Most Singaporeans are effectively bilingual: English and one of the “ethnic” languages: Mandarin for Chinese, Malay for Malays, and Tamil for Indians. Many are able to speak a smattering of Malay, Tamil, and the Chinese dialects represented in Singapore. One may be tempted to classify Singapore English as a mixed language (cf. Thomason, 1997). But this is hardly characteristic of the language situation in Singapore as a Straits Settlement. There are important changes in the linguistic ecology of Singapore English, especially wider access to English than was available to the founding population of 19th century Singapore. But the linguistic substratum remains unchanged. If the founding population exerted influence out of proportion to its numbers (Mufwene, 1996), a constant linguistic substratum (and superstratum) reinforced the effect of the founding population on the grammar of Singapore English. Today, Singapore English is a robust PC that children acquire as their mother tongue (Kwan-Terry, 1986, 1989; Gupta, 1991, 1994), even though it continues to be stigmatized in Singaporean society.
The Typology of Empty Categories

Chomsky (1981, 1995) recognizes four types of empty category, specified in terms of the two binary features [pronominal] and [anaphor]. These are shown in (1):

(1) a. [+Anaphor, –Pronominal] NP-trace
   b. [+Anaphor, +Pronominal] PRO
   c. [–Anaphor, –Pronominal] variable (wh-trace)
   d. [–Anaphor, +Pronominal] pro

By virtue of their feature specifications, empty categories are pronouns or anaphors with no phonological material. They are subject to the same binding principles as their overt counterparts. The two features determine the syntactic position in which a given empty category may occur and the binding principle that it must obey. An NP-trace, like overt anaphors such as himself or herself, occurs in an argument position, and obeys Principle A of the binding theory. The antecedent of an NP-trace occupies the subject position. An NP-trace typically is found in a passive sentence (a) or in a sentence with a so-called raising verb, such as seem (2b).

(2) a. The glass is broken e (cf. . . . break the glass)
   b. Mei Mei seems [e to enjoy the show]

Notice that an NP-trace and its antecedent must be in the same minimal clause, which is the matrix clause in both sentences in (2). In generative analysis, NP-traces result from raising. In the passive, the NP the glass is raised from the object position to the subject position. In constructions with raising verbs, the NP (Mei Mei in (2b)) is raised from the embedded subject.

---

8 Binding theory deals with the referential potentials of nominal expressions, which include empty categories. The three principles which comprise binding theory are as follows (Chomsky, 1981 p. 188):

Principle A: An anaphor is bound in its governing category.
Principle B: A pronominal is free in its governing category.
Principle C: An R-expression is free.

[+Pronominal] elements, whether phonologically null or overt, are subject to Principle B, and [+Anaphor] elements to Principle A. Simply put, a governing category of an element is the minimal NP or IP that contains that element, its governor, and an accessible subject or a finite AGR (agreement). The governing category is also called a complete functional complex. See Chomsky (1981, 1986 1995) and Haegeman (1994) for technical details of binding theory.
position to the matrix subject position. In both cases, the raised NP forms an argument chain with its trace, [the glass, e] in (a) and [Mei Mei, e] in (2b).

Unlike an NP-trace, PRO occurs in the subject position of an untensed clause, which in English is always embedded in a larger constituent, a clause, or a phrase. Depending on the main verb, the antecedent of PRO can be the subject or the object of the matrix clause. In (3a) the antecedent of PRO is the matrix subject Mei Mei, indicated by the subscript index i; in (3b) it is the matrix object Barney. PRO can also be arbitrary in reference, as the example (3c) shows.

(3) a. Mei Mei_i wants [PRO_i to see Barney]
b. Mei Mei told Barney_i [PRO_i to eat the carrot]c. [PRO to error] is human

The antecedent of a variable, which is commonly called the operator, occurs in COMP, a non-argument position. In generative linguistics, variables are derived through wh-movement, which is why they are also called wh-traces.9 The sentences in (4) show the typical positions of the wh-trace.

(4) a. Who [e saw Barney]? b. What [did Mei Mei like e]? c. Why [did Mei Mei cry e]? d. How much did you say [Mei Mei likes Barney e]?

The three types of empty category that we have discussed so far have one thing in common, namely, they are all licensed by their antecedents. The fourth type, pro, is a bit tricky in that it does not need to be licensed by an antecedent in the same sentence, and yet its reference is not arbitrary, like the arbitrary PRO in [PRO to err] is human. Since it is specified as [−Anaphor, +Pronominal], pro ought to behave like an overt pronoun. As

---

9 There is another type of variable, bound to quantified phrases such as every man. In generative linguistics, this type of variable is derived through movement at the level of Logical Form. Sentences containing quantified expressions often exhibit scopal ambiguity. In English, for example, the sentence Every man loves some woman is ambiguous. If every man has wide scope, the sentence means that each and every man loves his own woman; if some woman has wide scope, the sentence means that all men love the same woman. These two readings have the LF representations (i) and (ii)

(i) [for every x, x a man] [there exists y, y a woman] [x loves y]
(ii) [there exists y, y a woman] [for every x, x a man] [x loves y]
In this paper, we will not be concerned with quantifier-variable binding.
a pronoun, pro cannot have an antecedent within the minimal clause that contains it, but may have an antecedent outside its minimal clause. In other words, it is subject to Principle B of the binding theory (cf. footnote 8). The occurrence of pro is subject to the principle of recoverability on the basis of inflection (cf. Chomsky, 1981, 1995; Haegeman, 1994). Typically pro is found in languages with rich agreement inflection, such as Italian and Spanish. In English, inflection is not rich enough to allow unambiguous recovery of the null pronoun. We can see this contrast in the Spanish sentence (5a), which may contain a null pronoun in the subject position of a tensed, embedded clause (Huang, 1984, p. 533):

(5) a. José sabe [que él/e ha sido visto por María].
    José know that he/e has been seen by María
    'José knows that [he] has been seen by María.'

b. José sabe [que María lo/*e ha visto]
    José know that María him/e has seen
    'José knows that María has seen him/*e'

Note that the Spanish sentence is grammatical, with or without the overt pronoun él. Its English counterpart is grammatical only with the overt pronoun he — pro renders the sentence ungrammatical. However, in Spanish, as in English, pro does not occur in the object position: compare (5b). Spanish verbs do not exhibit object agreement. In languages with even richer inflection than Spanish-type languages, such as Pashto, a language spoken in Afghanistan, object pro is possible. The object agreement morphology ensures recovery of the null pronoun (cf. Chomsky, 1981, 1986; Huang, 1984, 1989).

Thus, the occurrence of pro is correlated with inflection, especially agreement. In Pashto-type languages, pro occurs in both subject and object positions; in Spanish-type languages, it occurs in subject position only; and in English-type languages, it does not occur at all. Huang (1984) suggests that in languages with no inflection, such as Chinese, Korean, and Japanese, pro is also possible in subject position, although not in object position. Huang’s proposal is motivated by his analysis of empty pronouns in Chinese, as we will see in the following section.

We summarize the typology of empty categories in terms of their syntactic positions and those of their antecedents or operators in (6).
It has been widely accepted in Chinese linguistics that Chinese is a topic-prominent language, in contrast with English, which is a subject-prominent language (Chao, 1968; Li & Thompson, 1976; Huang, 1984, 1989; Xu & Langendoen, 1985). According to Li and Thompson (1976), one major typological difference between topic-prominence and subject-prominence has to do with the grammatical relation encoded in the structure of basic sentences: in topic-prominent languages, it is topic-comment; in subject-prominent languages it is subject-predicate. This is illustrated in (7).

(7) a. John hit Mary.
   Subject Predicate
b. As for education, John prefers Bertrand Russell’s ideas.
   Topic Comment
c. Zhangsan xihuan pingguo. ‘Zhangsan likes apples’
   Zhangsan like apples
   Subject Predicate
d. Pingguo Zhangsan xihuan e1 ‘Apples, Zhangsan likes.’
   apple Zhangsan very like
   Topic Comment

The two English sentences are cited from Li and Thompson (1976, p. 459). The formal difference between the two clausal types can be expressed by the two phrase structure rules, or schemata, given in (8) (TP: topic phrase).\(^\text{10}\)

\(^{10}\)The two rules in (8) are adapted from (i) and (ii), proposed by Chomsky (1977, 1981) and Xu and Langendoen (1985c).
In (8b), TOP is the position for the topic, and IP provides the comment on the topic. Both COMP and TOP are non-argument positions.

Chinese topic structure is well studied in the literature. An important issue in the analysis of topic structure is the relationship between TOP and IP (or a constituent of IP). Consider the topic structure of (7d). At first glance, (7d) looks as if it is derived from (7c), whereby the object pingguo ‘apple’ moves to the TOP position, leaving behind the trace e. In this view, topicalization is analogous with wh-movement, with the null element, a variable, bound to TOP instead of COMP. Although this works for the simple topic structure exemplified in (7c, 7d), the movement analysis fails to account for the whole range of topic structures in Chinese. Xu and Langendoen (1985) argue convincingly that Chinese topic structure exhibits robust properties that violate the usual constraints on movement, such as subjacency and bijection. The topic structure is a basic construction type and (8b) is part of the phrase structure grammar of topic-prominent languages.

The topic-comment relationship is established through an interpretative strategy, stated below (Xu & Langendoen, 1985, p. 20).

(i) \[ \text{COMP } S \]
(ii) \[ \text{TOP } S \]

Given the phrase-structure schema of X-bar theory (cf. Haegeman, 1994, p. 105; Chomsky, 1995, p. 53), the rules in (8) can be recast as follows:

(iii) \[
\text{CP}
\]
\[
\text{Spec} \quad \text{C'}
\]
\[
\text{C} \quad \text{IP}
\]

(iv) \[
\text{TP}
\]
\[
\text{Spec} \quad \text{T'}
\]
\[
\text{T} \quad \text{IP}
\]

For ease of exposition I will continue to use the simpler structure of (8), with COMP and TOP occupying the specifier position of C and T, respectively.
(9) \([\text{TP} \alpha [\varepsilon \ldots \beta \ldots]]\), where \(\alpha\) is a major category and \(\beta\), possibly empty, is related to \(\alpha\).

In Xu and Langendoen’s analysis, which is adopted here, (7d) is base-generated, and the null object \(e_i\) is “related” to the topic pingguo; ‘apple’ through co-indexation, in accordance with (9). Since TOP is \([\text{Spec}, \text{TP}]\), a non-argument position, we will treat the null element in topic structure as a variable, and the relationship as TOP-variable binding. The terminology is for expository convenience. It does not imply that TOP-variable binding is established through movement, as is the case for COMP-variable binding.

In what follows, we introduce five salient properties of Chinese topic structure, which will provide sufficient evidence for the structural congruity between Chinese and Singapore English in TOP-variable binding.

First, topics in Chinese involve all major phrasal types and all grammatical functions, as shown in (10) (Xu & Langendoen, 1985, pp. 3–5).

11 The data in (10), and elsewhere in the paper, are quoted verbatim from the source. Since most accessible literature on Chinese topic structure is Mandarin-based, the data used in this paper are from Mandarin, transcribed in pinyin, the official romanization scheme widely used in Chinese linguistics. Works on other Chinese dialects in English include Bodman (1955) on Hokkien, and Matthews and Yip (1994) on Cantonese. Syntactically the dialects are the same in relevant respects. For comparison I give the Hokkien (H) and Cantonese (C) equivalent of the Mandarin (M) data in (10) below.

(i) M: [NP Wu xiansheng], [IP e renshi wo]  
H: [NP gə xiaam], [IP e linsuk gua]  
C: [NP NS in NS an], [IP e jinjëk NS]  
Wu Mr know I

(ii) M: [NP Zhexie hua] [IP wo bu xiangxin e]  
H: [NP tsia-e hue] [IP gua put jënjën e]  
C: [NP jit-ti wa] [IP NS i NS sin e]  
these words I not believe

(iii) M: [IP Ta hui shuo zhexie hua] [IP wo bu xiangxin e]  
H: [IP i hue suat tsia-e hue] [IP gua put jënjën e]  
C: [IP k’øy wui jët jit-ti wa] [IP NS m siuNSin e]  
he can say these words I not believe

(iv) M: [TP [NP Zhexie hua] [IP ta hui shuo e]] [IP wo bu xiangxin e]  
H: [TP [NP tsia-e hue] [IP i hue suat e]] [IP gua put jënjën e]  
C: [TP [NP jit-ti wa] [IP k’øy wui jët e]] [IP NS m siuNSin e]  
these words he can say I not believe
As the above data show, the topic position can be occupied by a noun phrase (10a, b), a clause (10b, d), and a prepositional phrase (10e). Moreover, the clausal topic may itself contain a topic: cf. (10d). In terms of grammatical function, the empty category could be an argument (10a–d) or an adjunct (10e). In all these examples, the empty category is a variable bound to the topic.

There is some trivial difference in lexical choice between Mandarin and Hokkien or Cantonese. For example, my Hokkien informants prefer to use the colloquial bat ‘know’, instead of the more formal linsik (cf. [i]), and my Cantonese informants prefer to use kong ‘say’ instead of jyt (cf. [iii] and [iv]). Syntactically, as the data show, Hokkien, Cantonese, and Mandarin share the same topic structure. The same is true of the data in the rest of the paper. Henceforth only Mandarin data will be cited.

I am grateful to two JCPL reviewers for raising this point.

I am grateful to two JCPL reviewers’ comment that led to this clarification.
The type of topic construction illustrated in (10) is not unique to topic-prominent languages. It is readily found in English, as indicated by the glosses. Chafe (1976, p. 49) calls them “English-style” topic structure, where the fronted expression is the “focus of contrast.” One important structural feature of the English-style topic construction is that the comment contains a pronominal form, which need not be phonologically null. The two sentences below contain the so-called resumptive pronoun \( t\.\) ‘he’:

(11) a. \([\text{NP } \text{Wu xiansheng}] , [\text{IP } t\. \text{renshi wo}]\)
   \begin{align*}
   &\text{Wu Mr he know I} \\
   &\text{‘Mr Wu, he knows me.’}
   \end{align*}

b. \([\text{NP } \text{Wu xiansheng}] , [\text{IP } \text{wo renshi ta}]\)
   \begin{align*}
   &\text{Wu Mr I know he} \\
   &\text{‘Mr Wu, I know him.’}
   \end{align*}

Referentially, the null pronominal form or the resumptive pronoun is bound to the topic.

Second, let us consider the so-called “Chinese-style” topic construction, which lacks a pronominal form in the comment. This is illustrated in (12) (Li & Thompson, 1976. p. 462).

(12) a. \([\text{NP } \text{nei-chang huo}] [\text{IP } \text{xingkui xiaofang-dui lai}]\)
   \begin{align*}
   &\text{that fire fortunate fire-brigade come} \\
   &\text{de kuai} \text{adv. particle quick} \\
   &\text{‘That fire, fortunately the fire-brigade came quickly.’}
   \end{align*}

b. \([\text{NP } \text{nei-xie shumu}] [\text{IP } \text{shushen da}]\)
   \begin{align*}
   &\text{those tree tree-trunk big} \\
   &\text{‘Those trees, the trunks are big.’}
   \end{align*}

In these sentences, the topic specifies a reference frame for the comment, or some constituent within the comment. In (a), \( \text{nei-chang huo} \) ‘that fire’ is the context in which the entire comment is to be interpreted. In (b), \( \text{shushen} \) ‘tree trunk’, the subject of the comment, refers to the trees specified by the topic. The Chinese-style topic structure is the primary empirical evidence against a movement analysis of Chinese topicalization.
Third, clauses may contain multiple topics, with each topic binding its own variable. Two such structures are shown below (Xu & Langendoen, 1985, p. 17).

(13) a. [zuotian]₁ [Li xiansheng]₂ [wo e₁ kanjian le e₂]
   yesterday Li Mr I see ASP
   ‘Yesterday, Mr Li, I saw.’

b. [zhejian shi]₁ [Li xiansheng]₂ [wo gaosu guo e₁ e₂]
   this event Li Mr I tell ASP
   ‘This event, Mr Li, I have told about.’

In (13a), the topic ‘yesterday’ binds the adjunct variable e₁, and the topic ‘Mr Li’ binds the null object e₂. In (13b), the direct and indirect objects of the verb ‘tell’ appear as topics, each binding its own variable. While multiple-topic structures are grammatical, and indeed common, in Chinese, their English counterparts do not fare as well: The English gloss in (13a) may be acceptable, it is definitely questionable in (13b).

---

13 The phrase-structure rule (8b) needs to be revised to accommodate multiple topics. Xu and Langendoen (1985) propose the schema shown in (i):

(i) TP → TOP {IP, TP}

which expands into the two rules in (ii):

(ii) a. TP → TOP IP
    b. TP → TOP TP

(ii-b) is recursive. Applying (ii-b) twice generates the hierarchical structure (iii):

(iii)

For the sake of simplicity, we assume, without argument, a “flat” structure for multiple topics, that is, (iv):

(iv)

Nothing in our discussion hinges on the internal configuration of multiple-topic structures.
Fourth, one topic may bind more than one variable, or serve as the point of reference for multiple comments. This violates the bijection principle, which stipulates a one-to-one relation between the operator and the variable (Xu & Langendoen, 1985). Some relevant examples are shown in (14).

(14) a. [liulian], [Zhangsan xihuan chi \(e_1\)], [keshi bu yuan yi mai \(e_2\)]
   durian Zhangsan like eat but not willing buy
   ‘Durians, Zhangsan likes to eat, but is not willing to buy.’

b. [zhe-ke shu], [ye duo], [hua shao]
   this-CL tree leaf many flower few
   ‘This tree, the leaves are many, the flowers are few.’

c. [nei-kuai tian], [daozi zhangde hen da], [suoyi e hen zhiqian]
   that-CL land rice grow very big so very valuable
   ‘That piece of land, rice grows very big, so (it) is very valuable.’

In (14a), the topic binds two variables, \(e_1\) and \(e_2\), the object of their respective comment clauses. In (14b), which is a Chinese-style topic structure, the topic provides the point of reference for two comment clauses. In (14c), cited from Li and Thompson (1976), the topic provides the context for the interpretation of the first comment, and it binds the null subject in the second.

Finally, the topic itself may be phonologically null. This is especially common when the topic is understood in the context of the utterance (15a), or as answers to questions (15b):

(15) a. Zhangsan shuo Lisi bu renshi e
   Zhangsan said Lisi not know
   ‘Zhangsan said Lisi did not know [him].’

b. Q: Zhangsan xihuan pingguo ma?
   Zhangsan like apple Particle
   ‘Does Zhangsan like apples?’
   A: e xihuan e
      like
      ‘Yes (literally, like)’
In (15a), the null object $e$ refers to someone understood in the context, not to the matrix subject Zhangsan. In (15b), the references of the null subject and object in the answer are derivable from the question. Huang (1984) proposes that these null elements are variables bound to topics that are themselves null. The representation for (15a, b) is as follows:

$$
(16) \begin{array}{l}
\text{a.}\ [TP [TOP i] [IP Zhangsan shuo Lisi bu renshi } e_i]\] \\
\text{b.}\ [TP [TOP i [TOP j] [IP } e_i \text{ xihuan } e_j]\]\n\end{array}
$$

The order of the two topics is inconsequential. The null topic is only possible in topic-prominent languages, where topic-comment structure, formally captured by phrase-structure rule (8b), is a basic sentence type.

The formal properties of topicalization in Chinese are summarized below.

$$
(17) \begin{array}{l}
\text{a.}\ \text{“English-style” topic structure} \\
[TP [TOP } \alpha_i\] [IP ...$\beta_i$ ... ]], ($\beta_i$ is phonologically null). \\
\text{b.}\ \text{“Chinese-style” topic structure} \\
[TP [TOP } \alpha_i\] [IP ...$\beta_i$ ... ]], ($\beta_i$ is the comment, or a non-null constituent of the comment). \\
\text{c.}\ \text{Multiple-topic structure} \\
[TP [TOP } \alpha_i\] [TOP $\alpha_j$] ... [IP ...$\beta_i$ ...$\beta_j$ ... ]], ($\alpha_{ij}$, $\beta_{ij}$ as defined in (a) or (b)). \\
\text{d.}\ \text{Multiple-comment structure} \\
[TP [TOP } \alpha_i\] [IP ...$\beta_i$ ... ] [IP ...$\beta_i$ ... ] ... ], ($\alpha_i$, $\beta_i$ as defined in (a) or (b)). \\
\text{e.}\ \text{Null topic} \\
[TP [TOP } \alpha_i\] [IP ...$\beta_i$ ... ]], $\alpha_i$ is phonologically null, ($\beta_i$ as defined in (a) or (b)).
\end{array}
$$

These topic structures are not exhaustive, nor are they mutually exclusive. The null topic, for example, may bind variables in multiple comments (cf. (17d, e)), and multiple topics and multiple comments (cf. [17c, d]) may be found in the same sufficiently complex topic structure. They provide a strong diagnostic of the typological status of Singapore English. We will discuss Singapore English topic structure, in connection with TOP-variable binding, in the following section.
We now proceed to examine empty categories in Chinese and Singapore English. As we have mentioned earlier, of the four empty categories, we will exclude NP-trace from consideration, since its existence is dependent on theoretical assumptions about raising verbs and passivization (cf. Haegeman, 1994). We will focus on PRO, pro, and variable, and show that these empty categories in Chinese and Singapore English have the same grammatical properties.

The existence of empty categories in Singapore English has been noted in the literature, and it has been attributed to Chinese influence (cf. Tay, 1979; Platt & Weber, 1980; Gupta, 1991; Platt & Ho, 1993). But so far there has been little work on the systematic analysis of Singapore English empty categories. The analysis of empty categories in Chinese, however, is controversial. Two positions have been put forth in the literature. One position holds that empty categories in Chinese obey the same binding principles as their counterparts in English, and indeed in other natural languages. This position is most eloquently argued in Huang (1984, 1989). The opposing camp rejects this analysis, arguing instead that Chinese empty categories are not subject to the binding principles, and are therefore “free” (Xu, 1986; Huang, 1992). Here I will not attempt to examine the theoretical and empirical merits of the opposing arguments. For our purpose, it is sufficient to demonstrate the structural and interpretive parallel between the empty categories in Chinese and those in Singapore English. For expository convenience, I will adopt the argumentative structure of Huang’s (1984) analysis. It should be pointed out that whatever the theoretical framework, the analysis of Chinese empty categories can be easily modified to accommodate Singapore English empty categories.

**PRO**

A PRO is a null pronoun that occurs in the subject position of an un-tensed clause, which is often embedded; its antecedent may be arbitrary, or an element in the matrix clause. These two types of PRO can be found in Chinese as well as Singapore English. Relevant data are shown in (18) and (19).  

---

14 The Singapore English data are culled from various sources, including the on-going Singaporean section of the International Corpus of English (cited as ICE), and from the intuition
(18) a. [PRO xiyan] you hai
   smoke have harm
   ‘Smoking/To smoke is harmful.’

   b. *[PRO xiyan-le] you hai
      smoke-ASP have harm
      ‘Smoked is harmful.’

(19) a. [PRO walk in Pulau Ubin] also can

   b. [PRO walking in Pulau Ubin] also can

   c. [PRO leave one stroke only], [e wrong already] (Platt, Weber, & Ho, 1983, p. 25)
      ‘Leave out one stroke, and (the character) is wrong’

In (18a), PRO occurs in the subject position of an untensed clause, and has arbitrary reference. (18b) shows that PRO can not occur as the subject of a tensed clause, as marked by the perfective marker le. This is the expected behavior of PRO (cf. [6]). In Singapore English, the infinitival clauses need not be introduced by to (Platt, Weber, & Ho, 1983, Tay, 1979). As (19) shows, PRO is found in the same constructions as it is in English and Chinese. In (19c), already in the second clause is the perfective marker (Kwan-Terry, 1989; Bao, 1995), and the gap e is not a PRO, but a variable bound to the discourse topic, which is a Chinese character not mentioned in the clause, but understood in the context. We will discuss variables in the section on variable and COMP.

Besides arbitrary PRO, controlled PRO is also found in Chinese and Singapore English. The data in (20) are from Chinese.

(20) a. Zhangsan_i daying Lisi_j [PRO_i/*j kan zhe-bu dianying]
      Zhangsan promise Lisi see this-CL movie
      ‘Zhangsan promised Lisi to see the movie.’

   b. Zhangsan_i jueding [PRO_i bu lai] le
      Zhangsan decide not come ASP
      ‘Zhangsan decided not to come.’

   c. Zhangsan_i quan Lisi_j [PRO_i/*j bie kan zhe-bu dianying]
      Zhangsan persuade Lisi not see this-CL movie
      ‘Zhangsan persuaded Lisi not to see the movie.’
d. Zhangsan, bi Lisi [PROi kan zhe-bu dianyin]
   Zhangsan force Lisi see this-CL movie
   ‘Zhangsan forced Lisi to see the movie.’

In (20a, b), the null subject of the embedded clause is a PRO controlled by
the matrix subject Zhangsan. (20c, d) the null subject is a PRO controlled
by the matrix object Lisi.

Consider now the data from Singapore English, shown in (21).

(21)  
   a. So you decided [PROi not to quit] already hah (ICE)
   b. I should think that he would want [PROi to go back to Eu-
      rope] (ICE)
   c. You call her [PROi walk there], e very far le (Platt & Ho,
      1989, p. 219)
   d. Older generation, they mind child [PROi to get married to
      their own dialect] (Platt, Weber, & Ho, 1983, p. 60)
   e. Bernard ask me [PROi bring you along with me] (Tay 1993,
      p. 100)

(Hah in (21a) and le in (21c) are sentence-final particles.) The PROs in
(21a, b) are controlled by the matrix subject, and those in (21c, d, e) by the
matrix object. These are the expected uses of the verbs decide, want, call,
and ask. In (21c), e, the null subject of very far le, is not a PRO. It refers to
the destination there mentioned in the previous clause. In (21d), mind is used
as an “object control” verb, very much like call, so the sentence means that
parents prefer their children to marry people who speak the same (Chinese)
dialect.

Pro

Pro is an empty pronominal that occurs in an argument position of a
tensed clause, and its antecedent occurs in an argument position as well.
Like overt pronouns such as him, pro and its antecedent cannot occur in
the same minimal clause. Languages which allow pro are often called pro-
drop languages, of which Italian and Spanish are often-cited examples. Since
English is not a pro-drop language, sentences with pro’s are ungrammatical,
as shown in (22).

(22)  
   a. *John said [proi will like Mary]
   b. *John said [Mary will like proi]
Consider now the distribution of pro in Chinese. Huang (1984) shows that pro occurs in the subject position of a tensed clause, but not in the object position. The subject-object asymmetry is illustrated in (23) ([23a,c] are cited from Huang, 1989, p. 187; the subscript $k$ refers to discourse topics understood in context of use).

\[(23)\]
\[
\begin{align*}
\text{a. } & \text{ Zhangsan shuo } [e_{ik} \text{ hen xihuan Lisi}] & \text{Zhangsan said very like Lisi} \\
& \text{‘Zhangsan said (he/someone else) liked Lisi.’} \\
\text{b. } & \text{ Zhangsan shuo } [ta_{ik} \text{ hen xihuan Lisi}] & \text{Zhangsan said he very like Lisi} \\
& \text{‘Zhangsan said (he/someone else) liked Lisi.’} \\
\text{c. } & \text{ Zhangsan shuo } [Lisi \text{ hen xihuan } e_{ik}] & \text{Zhangsan say Lisi very like} \\
& \text{‘Zhangsan said Lisi liked (*him/someone else).’} \\
\text{d. } & \text{ Zhangsan shuo } [Lisi \text{ hen xihuan } ta_{ik}] & \text{Zhangsan say Lisi very like he} \\
& \text{‘Zhangsan said Lisi liked (him/someone else).’} \\
\end{align*}
\]

Huang (1984) observes that the null subject in (23a) may refer to the matrix subject Zhangsan, or someone understood in the discourse. This is exactly the behavior of the overt pronoun ta ‘he, she’: compare (23b). In (23c), the null object can only refer to the discourse topic, but not to the matrix subject Zhangsan. This behavior differs from that of the overt pronoun ta, which is free to take the matrix subject as its antecedent: compare (23d). The data in (23) suggest that the null subject, but not the null object, gives rise to ambiguity. Huang (1984) accounts for the subject-object asymmetry by allowing only the null subject as pro, and both the null subject and null object as possible variables. As a pro ($e_i$), the null subject takes the matrix subject as its antecedent, as sanctioned by Principle B of the binding theory; as a variable ($e_k$), it is bound to TOP, a non-argument position. (23a) is therefore ambiguous. (23c) is not ambiguous, since the null object can only be interpreted as a variable bound to TOP. The syntactic representations of (23a,c) are as shown in (24) (var, variable).

\[(24)\]
\[
\begin{align*}
\text{a. } & \text{ TP [TOP $k$ [IP Zhangsan shuo [pro$_i$ hen xihuan Lisi]]]} \\
\text{a’. } & \text{ TP [TOP $k$ [IP Zhangsan shuo [var$_k$ hen xihuan Lisi]]} \\
\end{align*}
\]
One consequence of this explanation is that pro and PRO are indistinguishable in terms of the grammatical position that they occupy. The only syntactic difference between the two empty categories is that pro occurs in the subject position of a tensed clause, and PRO in the subject position of an untensed clause. Neither pro nor PRO can occur in the object position.¹⁵

When we turn to Singapore English, we find the same distribution of pro. The relevant data are shown in (25):

(25) a. [...]k [Sar Che and Sar Ee] i said [eₖ go to the airport to fetch you and Li Sa] (Platt, Weber, & Ho, 1983, p. 123)
   b. [...]k then [my mum] i said [eₖ must call her sister] (ICE)
   c. [Mei Mei] i said [eₖ finished lunch already]
   d. [...]k they said [they will send eₖ to you] (Tay, 1979, p. 104)

(25a) is cited from a dialogue in a Singaporean play. Within the context of the play, the antecedent of e could be the matrix subject Sar Che and Sar Ee, or someone else. In other words, the null subject of the embedded clause could be interpreted as a pro with Sar Che and Sar Ee as its antecedent, or as a variable, bound to the empty TOP. The same is true of (25b, c), where e could be a pro or a variable. The null object in (25d) is a variable; it can only refer to something understood in the context, not to the matrix subject they.

¹⁵The existence of pro and PRO in Chinese is by no means an accepted conclusion in Chinese linguistics. The difference between the two empty categories crucially depends on the finiteness of a clause. Since there is no morphological marking of finiteness in Chinese, it is often difficult to identify the null subject as belonging to one category, rather than the other. Xu and Langendoen (1985), Xu (1986), and Huang (1992) use this type of empirical evidence to argue against Huang’s (1984 1989) analysis of Chinese empty pronominals. As we shall see, the same problem exists in Singapore English: if there is no consistent inflectional marking, what is the surface-observable evidence of finiteness? In this paper, we will not attempt a theoretical analysis of finiteness in Chinese and Singapore English. For our purpose, it is sufficient to establish structural parallels between the two languages in the distribution and interpretation of empty categories. Whatever theoretical approach one adopts to the finiteness issue, and consequently to the identification of pro/PRO in one language can be readily extended to the other. We will therefore not be concerned with the theoretical details of the opposing arguments.
In Chinese and Singapore English, the null subject of a tensed clause can be identified as a pro or a variable, but the null object can only be identified as a variable. The subject-object asymmetry is readily observable in pragmatically neutral cases. However, extraneous pragmatic or discoursal factors may create difficulties in the identification of the null object. Consider the Chinese sentences in (26).

(26) a. Zhangsan shuo [Lisi bu renshi e] 
   Zhangsan say Lisi not know 
   i. ‘Zhangsan said Lisi did not know (others).’
   ii. ‘Zhangsan said Lisi did not know (Zhangsan).’

b. Zhangsan danxin [mama yao ma e le] 
   Zhangsan worry Mom will scold ASP 
   i. ‘Zhangsan is worried that Mom will scold (others).’
   ii. ‘Zhangsan is worried that Mom will scold (Zhangsan).’

Under normal circumstances, the null object in (26a) can only be interpreted as a variable, referring to someone understood in the context (reading [26a-i]). But as an answer to the question Who did not know Zhangsan?, the null object could refer to the matrix subject Zhangsan, making it look as if it is a pro (reading [26a-ii]). As for (26b), the two readings are equally plausible in some pragmatically appropriate context. It appears that the null object, like the null subject, can be a pro as well as a variable.16

The same is true of the null object in Singapore English. Consider the discourse fragment below, uttered by the father of a one year old boy:

(27) My son wants to carry all the time. 
   ‘My son wants to be carried/(someone) to carry him all the time.’

From the context, it is clear that the utterance means that the child wants someone, preferably the father, to carry him all the time. This interpretation forces us to assign the following syntactic representation to the clause:

---

16 From the theoretical point of view, the behavior of the null object exemplified in (26) is problematic. If we allow it to be a pro as well as a variable, we cannot account for the subject-object asymmetry in such familiar cases as (23). On the other hand, if we identify it only as a variable, we cannot give a straightforward account of the pro-like readings. For theoretical solutions that have been proposed in the literature, see Huang (1984, 1989), Xu (1986), and the discourse-oriented work of Huang (1992).
(28) \[ \text{TP} \ [\text{TOP} \ [\text{IP} \ \text{my son}, \text{wants} \ [\text{PRO to carry} \ e_1 \ \text{all the time}]]] \]

PRO has arbitrary reference and the null object \(e_1\) is a \(\text{pro}\) taking the matrix subject \(\text{my son}\) as its antecedent. Furthermore, the verb \(\text{want}\) is not a "subject-control" verb, as it is in native English. In fact \(\text{My son wants to carry all the time}\) is the exact rendering of the Chinese sentence, with the exception of the position of the temporal adverbial \(\text{all the time}:\)

(29) \(\text{wo er zi zongshi yao } [e_1 \ bao \ e_2]\)

\(\text{I son all-the-time want carry}\)

In Chinese, temporal adverbials are placed before the main verb; in English, they are placed after the main verb. Temporal adverbials in Singapore English retain the position of the lexifier language. The Chinese verb \(\text{yao}\) is not a subject-control verb either. In the sentence, \(e_1\) may refer to the matrix subject, or someone else understood in the context (arbitrary \(\text{PRO}\)), and \(e_2\) refers to the matrix subject (\(\text{pro}\)). The null objects in Singapore English and Chinese exhibit the same behavior.\(^{17}\)

**Variables and COMP**

Variables occur freely in argument as well as non-argument positions within a clause, and are bound to operators in non-argument position, which in English is COMP and in Chinese, TOP. In Singapore English, a variable could be bound to both COMP and TOP, depending on the nature of the variable, and the type of construction in which it is found. We will discuss TOP-variable binding in the following section. Here, we discuss COMP-variable binding in Singapore English.

In generative linguistics, variables are treated as traces left behind when \(\text{wh-elements}\) are moved to COMP to form \(\text{wh-questions}\) (cf. Haegeman, 1994; Chomsky, 1995). \(\text{Wh-movement}\) is optional in Singapore English, betraying

\(^{17}\) One might argue that, since the use of the copula \(\text{be}\) and the verbal inflection \(-\text{ed}\) is optional in Singapore English, the embedded clause \(\text{to carry}\) could be interpreted as a passive.

In this view, the null object is an NP-trace. In generative linguistics, the passive is represented as follows:

\[[\text{be carried my son}] \rightarrow [\text{my son}, \text{is carried} \ e_1]\]

The null element is the NP-trace "left" behind when \(\text{my son}\) is moved to the subject position of the sentence (Chomsky 1981, 1995). Whatever the analysis, the structural congruity between the Singapore English clause and its Chinese counterpart is still unmistakable.
a compromise under the influence of the two main contact languages: the obligatory movement in English, and no movement in Chinese.

In English, variables are typically found in wh-questions:

\[(30)\]
\begin{enumerate}
  \item Who, \(e_i\) took the apple?
  \item What, \(e_i\) did you take?
  \item Why/where/how, \(e_i\) did you take the apple?
  \item How, \(e_i\) did you take the apple?
  \item Where, \(e_i\) did you hide the apple?
\end{enumerate}

In (30a, b), the gaps occur in argument positions (subject and object); in (30c, d, e), they occur in non-argument position. The wh-words occupy the Comp position in all cases — they are the operators that bind the variables.

COMP-variable binding is not found in Chinese, where wh-elements do not move to form wh-questions.\(^{18}\) In Singapore English, wh-movement is optional. In addition to (30), questions in which the wh-elements remain in situ are equally grammatical, and common. A few examples follow.

\[(31)\]
\begin{enumerate}
  \item John must buy what? (Chow, 1995, p. 25)
  \item John must have dinner with who? (Chow, 1995, p. 32)
  \item Have you decided roughly what amount? (ICE)
  \item After that, you went to what? (ICE)
\end{enumerate}

The positioning of wh-phrases may be motivated by discoursal factors. Take (31d) for example. It is taken from the following extract:

\[(32)\] How’s your job? You work there only not long right? Your other job . . . Where are you now uh? Previously you’re at Kentucky right? Then after that you went to what? NCB, is it?

In the extract, the emphasis is on the temporal sequence of job-hopping; moving what to COMP breaks the temporal flow of the information intended in the conversation (cf. Ho, 2000).

\(^{18}\) In generative linguistics, Chinese wh-question formation is typically analyzed as involving wh-movement at the level of Logical Form, even though wh-phrases remain in situ at the surface; see Huang (1982). “COMP-variable” is a descriptive term that has no theoretical status — it is a variable that is bound by a wh-operator at the surface level. The same is true for “TOP-variable.”
Variables and TOP

The five topic structures in Chinese enumerated in (17) are readily attested in Singapore English. We will consider them in turn.

The “English-style” topic structure is exemplified in (33).

(33) a. [Cantigas], [I think $e_1$ is Portuguese music] (ICE)
b. [certain medicine], [we don’t stock $e_1$ in our dispensary] (Platt & Weber, 1980, p. 73)
c. [three months], [we have to wait for $e_1$] (Platt & Weber, 1980, p. 73)
d. [inside the schoolbag], [they have got food $e_1$] (Gupta, 1991, p. 130)
e. [is very interesting], [I find geography $e_1$] (Platt & Weber, 1980, p. 73)
f. [you can investigate some of the case], $e_1$ actually will be good] (ICE)

All major phrasal categories occur in TOP: NP in (33a,b,c), PP in (33d), VP (33e), and IP (or TP) in (33f). In terms of grammatical function, they are the subject (33a,f), the object (33b), the prepositional object (33c), the adverbial (33d), and the object complement (33e). In native English, (33a,b) are readily acceptable, but not the other cases in (33). Singapore English topicalization is as robust as Chinese topicalization, even in “English-style” topic structure.

We turn now to Chinese-style topic sentences. These are given in (34).

(34) a. [my family], [everybody is educated in English] (Platt, Weber, & Ho, 1983, p. 47)
b. [a pick-up taxi], [you have to share the trip with others] (Platt, Weber, & Ho, 1983, p. 48)

---

19 This sentence is extracted from the following dialogue:

A: So what do they do there? Quite busy uh?
B: No lah. Not too too bad. But you can investigate some of the case actually will be good. They probably will not allow you to investigate the whole case, but you can tag along you see.

The ICE corpus is transcribed without detailed prosodic information such as pauses. A few native speakers I consulted put a pause between case and actually, indicating that the sentence you can investigate some of the case actually will be good functions as a topic. For some scholars, notably Platt and Weber (1980) and Gupta (1991), this fragment is analyzed as lacking the conjunction if — the sentence is paraphrased as If you can investigate some of the case, actually it will be good.
c. [stay longer] [they have to overcharge] (Platt & Weber, 1980, p. 78)
d. [take paper] [I also want to tell mummy] (Gupta, 1991, p. 136)

In each of these sentences the comment lacks a null form. The topics provide a context for the interpretation of the comments. Platt and Weber (1980) and Gupta (1991) consider (34c, d) as examples of the missing conjunctions. Their native English counterparts are as follows:

(35) a. If you stay longer, they have to overcharge.
b. If you take paper, I also want to tell mummy.

However, (34c, d) have little conditional meaning. In view of our analysis of topicalization, it is better to treat (34c, d) as topic structures, with VPs in the TOP position. Stay longer and take paper provide the context, rather than the condition, for their respective comments.

Consider now the multiple-topic structures shown in (36).

(36) a. [One time], [the flats], nobody want $e_j e_i$. (Platt, Weber, & Ho, 1983, p. 48)
b. [to my sister], [sometime], I $e_j$ speak English $e_i$. (Platt & Weber, 1980, p. 73)
c. [another time], [same MRT], [this man], $h_j$ sat opposite me $e_i e_h$ (ICE)
d. [ ], [everything], $e_k$ also don’t know $e_i$
e. [about six-thirty], [my friend and I], $w_je e_i$ left $e_i$ hah] (ICE)

In relevant respects these structures are identical to the corresponding structures in Chinese, with each TOP binding one variable in the comment. In (36d), we see a case with a null TOP, referring to someone understood in the discourse. The utterance means that that person does not know anything. In (36e), the topic my friend and I binds the resumptive pronoun we.

The multiple-comment structure of Singapore English is exemplified in (37).

(37) a. [that fish], [I bought $e_i$ last week], [$e_i$ spoiled already]
b. [((Mandarin)), [but if they speak $e_k$ to me, I will speak $e_k$].
(Platt, Weber, & Ho, 1983, p. 51)
c. [(LY)\textsubscript{k}] [one day] [e\textsubscript{k} must go to the beach] and [e\textsubscript{k} collect more shells] and [e\textsubscript{k} make my shell collage] (ICE\textsuperscript{20})

d. [that boy\textsubscript{i}] [I know e\textsubscript{i} very naughty] [e\textsubscript{i} pinch my sister]

In all the data in (37), a single TOP binds more than one variable in a series of comments (37a, c, d), or in one complex comment (37b). The words in parentheses in (37b, c) are not expressed, but understood from the discourse context.

Finally, we consider the null-topic structure. We have seen topic sentences with null topics (cf. [36d] and [37b, c]). Here are a few more examples.

(38) a. [(about flowers)\textsubscript{k}] [I never ever draw e\textsubscript{k} what] (Gupta, 1991 p. 135)

b. [(on what is just said)\textsubscript{k}] [I think e\textsubscript{k} is prayer] (Platt, Weber, & Ho, 1983, p. 87)

c. [(about an order)\textsubscript{k}] [I don’t think I can send e\textsubscript{k} to you today] (Platt, Weber, & Ho, 1983, p. 90)

d. [(about black bean fish)\textsubscript{k}] [I mean e can’t eat it all the time but occasionally I like e\textsubscript{k}] (ICE\textsuperscript{21})

\textsuperscript{20}This sentence is taken from a lengthy conversation between two individuals, JY and LY, about, among other topics, kite-flying at Changi Beach. The immediate context of the fragment is as follows:

JY: It’s quite nice because the area so big.
LY: Ya ya quite busy.
JY: And there’ll all these little stores you know.
LY: That sell right?
JY: Sell the kites and all.
LY: Remember Changi beach we couldn’t even fly the kite. So no wind. Actually one day must go to the beach and collect more shells and make my shell collage.

It is clear that LY is the subject of the fragment (37c).

\textsuperscript{21}The sentence is cited from the following conversational fragment:

A: Actually I don’t like that black bean fish you know.
B: You don’t ah?
A: Very stiff one the fish.
B: I like ah. I, I, I mean can’t eat it all the time but occasionally I like. That fish has been selling very well.
Based on the data discussed above, we conclude that Singapore English is a topic-prominent language, like the substrate. Since wh-movement is optional in Singapore English (cf. Gupta, 1994; Chow, 1995; Ho, 2000), we propose (39) as part of the phrase-structure grammar of Singapore English:

(39) \[ TP \ TOP \ [CP \ COMP \ IP] \]

The structure (39) encodes structural elements from two sources: TOP from Chinese and COMP from English. For ease of comparison we expand the three phrase-structure rules, (8a, b) and (39), into trees, as follows:

(40)

a. English

```
        CP
       /   \   
COMP  IP
```

b. Chinese

```
        TP
       /   \   
TOP  IP
```

c. Singapore English

```
        TP
       /   \   
TOP  CP  
    /   \  
COMP IP
```

---

22 It is possible that the fronted wh-phrase in Singapore English does not appear in COMP, but in TOP. So the question what did you take? has the structure in (i), rather than (ii):

(i) \[ TOP \ [TOP whati] \ [IP did you take ei] \]

(ii) \[ COMP \ [COMP whati] \ [IP did you take ei] \]

Under this analysis, COMP-variable does not exist in Singapore English, at least not at the level of surface structure (see footnote 18). The difference between (i) and (ii), though theoretically important, is inconsequential for our purpose: either analysis establishes the substrate source of TOP/COMP-variable binding in Singapore English. We will therefore adopt the schema (39) without further argument.

I am grateful to one reviewer for raising this issue.
Note that TOP is higher on the tree than COMP, that is, topics precede wh-phrases in wh-questions. Reversing the order of TOP and COMP produces odd-sounding questions, as illustrated by the following pairs of questions:

(41) a. [TOP Changi beach], [COMP what bus], do you take e_i?
    a'. ![COMP what bus], [TOP Changi beach], do you take e_i?
    b. [TOP That car over there], [COMP why] you don’t polish e_i?
    b'. ?[COMP Why], [TOP that car over there], you don’t polish e_i?

The primed examples are either ungrammatical, or extremely awkward.

Thus, wh-question formation in Singapore English is a curious mixture of the substrate and superstrate. TOP-variable binding in Singapore English has its origin in Chinese, the main substrate language, whereas COMP-variable binding is obviously derived from English, the lexifier. Although wh-in-situ, a substrate feature, is a grammatical option, the robust pattern of wh-question formation involves wh-movement, with or without subject-verb inversion. This is evident not only in the speech of pre-school children (Kwan-Terry, 1986; Gupta, 1994), but also in the data culled from the ICE corpus. The contribution of the superstrate to the grammar of Singapore English is not limited to the vocabulary. The superstrate contributes grammatical features as well.

Conclusion

In the preceding sections we have established the structural and interpretative congruity between the empty categories of Chinese and those of Singapore English. We summarize the main results of our discussion below.

(42) a. Basic phrase-structure:
    English: [CP COMP IP]
    Chinese: [TP TOP IP]
    Singapore English: [TP TOP [CP COMP IP]]
    b. Empty categories
    | PRO     | English | SingEng | Chinese |
    |---------|---------|---------|---------|
    | pro, subject | no     | yes     | yes     |
    | pro, object   | no     | no      | no      |
    | variable, COMP| yes    | yes     | no      |
variable, TOP

<table>
<thead>
<tr>
<th></th>
<th>English-style</th>
<th>Chinese-style</th>
<th>Multiple-topic</th>
<th>Multiple-comment</th>
<th>Null topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

From the above table, we can see clearly the source of the empty categories in Singapore English. Wh-movement is derived from English, even though it is not as robust in Singapore English as it is in the source language. Variables bound to COMP are traces of wh-movement. Chinese influence is most obvious in pro, topic structure, and variables that are bound to TOP.

Huang’s (1984, 1989) analysis of empty categories crucially depends on two theoretical considerations: the nature of AGR and the existence of TOP, which may itself be phonologically null. As we have seen, Chinese and Singapore English converge on these two subsystems: both lack inflection, and both have the same topic structure. However, one needs to be cautious when attempting to explain the convergence. Lack of inflection in Singapore English need not be the result of substrate influence, even though both Chinese and Malay are non-inflectional languages. (Standard Malay has a rather productive derivational morphology, cf. Omar, 1989.) Since PCs lack inflection, it could be due to universal tendencies in pidginization and creolization, or to some obscure UG processes that give rise to those tendencies. The existence of subject pro is diagnostic of a pro-drop language. Singapore English is a pro-drop language in the sense that Chinese is a pro-drop language: the identification of pro does not depend on inflection, of which Singapore English and Chinese have none; it depends on control, if Huang’s analysis is correct.

The topic structure of Singapore English is clearly derived from Chinese. It is true that most, if not all, languages allow topic structure. In subject-prominent languages it is typically viewed as an example of movement, as is implied in the term topicalization. However, Singapore English topic structure does not allow a movement analysis — Xu and Langendoen’s (1985) argument against such an analysis of Chinese topic structure can be easily extended to Singapore English topic structure. Other PCs do not share the full range of syntactic properties of Singapore English topic structure. Bickerton (1981) and Veenstra and Besten (1994) show that topicalization, focusing, and clefting in Caribbean PCs, which I group together under the general
heading of topic structure, are syntactic and discoursal devices derivable through movement, subject to the usual constraints on movement. This is not the case in Singapore English. Furthermore, Singapore English imposes no categorial restriction on what can appear in TOP — all major phrasal categories can serve as topics. This is not true in other PCs. In Guyanese Creole, for example, clauses introduced by se cannot be fronted (Bickerton, 1981, p. 107), as illustrated by the sentence pair in (43).

(43) a. mi no aredi se dem gaan
   I know already that they gone
   ‘I knew already that they’d left.’

b. *se dem gaan mi no aredi
   that they left  I know already
   ‘That they’d left I knew already.’

By comparison, Singapore English topic structure is much richer than equivalent constructions in other PCs. We conclude that it is substratum-derived, made possible under typological convergence between Chinese and Singapore English. We express this convergence in terms of three parameters: [wh-movement], [topic-prominence], and [pro-drop]. The settings are shown in (44).

(44) Parametric Setting SingEng Chinese English
    [wh-movement]  +/−  −  +
    [topic-prominence]  +  +  −
    [pro-drop]  +  +  −

The debate between the universalists and substratists — whom Hall (1966) calls substratophobes and substratomaniacs — diverts attention exclusively to one constituent aspect of linguistic ecology, to the neglect of contributions from all other constituents. Since PCs do not arise in a social or linguistic vacuum, a holistic approach is needed to explain the emergence of PC grammar, which, like the grammar of all human languages, must meet the requirements of UG. The system of empty categories in Singapore English results from the re-structuring of the relevant parametric settings under the pressure of UG and the languages that constitute the linguistic ecology of Singapore English.
References


