Indigenous vocabulary in Singapore and Indian Englishes

Lotta Fors
University of Tampere
Faculty of Communication Sciences
Master's Program in English Language and Literature
MA Thesis
May 2017

Tämä pro gradu –tutkielma tutkii alueiden paikalliskielistä tulevia lainasanoja Singaporen englannissa ja Intian englannissa ja vertaa varieteettien korpustutkimuksen tuloksia keskenään sekä aikaisempiin tuloksiin ja teorioihin. Tutkimuksen tavoitteena on paremmin kartoittaa lainasanojen käyttöä maailmanenglanneissa sekä selvittää minkälaisia sanoja lainataan enemmän kuin toisia. Tutkimuksessa tarkastellaan myös, esiintyvätkö lainasanat enemmän puhutussa vai kirjoitetussa kielessä ja perustuvatko tulosten eroavuudet maiden erilaisiin kolonisaatiokokemuksiin ja kielipoliittisiin päätöksiin.

Tutkielmassa käsitellään myös englanninkielen historiaa molemmissa maissa, Singaporen englannin ja Intian englannin pääpiirteitä ja luokitteluja, sekä aikaisempaa teoriaa maailmanenglanneista ja lainasanastoista. Tutkimus- ja pohdintaosioiden kannalta tärkeimpiä teoksia ovat Balasubramanianin *Register Variation in Indian English* (2009) ja Schneiderin *Postcolonial Englishes* (2007). Balasubramanianin teosta käytettiin erityisesti hyväksi tutkimustulosten rajauksessa ja tulosten jäsentelyssä, kun taas tutkimuksen perustelussa ja tulosten pohdinnassa tukeuduttiin useammin Schneiderin teoriamalliin maailmanenglantien kehityskaaresta.

Korpustutkimuksen aineistona toimivat ICE-korpuksen (*International Corpus of English*) Singaporen ja Intian englantien komponentit, joiden samankaltainen rakenne takaa tutkimustulosten kiistämättömän vertailukelpoisuuden. ICE-korpuksien materiaali tulee niin puhutusta kuin kirjoitetusta kielestä, joka on kerätty eri lähteistä. Puhutun ja kirjoitetun materiaalin erottelu pienemmiksi tarkoin määritellyiksi osioiksi korpuksissa mahdollisti myös tulosten tarkkojen esiintymiskontekstien analysoinnin.

Tutkimuksessa selvisi, että Singaporen englannissa esiintyy huomattavasti vähemmän paikalliskielistä lainattuja sanoja kuin Intian englannissa. Vaikka lainasanoja määrässä oli korpusten välillä eroja, tulosten kategorisoinnin avulla voitiin kuitenkin huomata, että molempien varieteettien lainasanat ovat keskenään toistensa kaltaisia. Suurin osa paikalliskielistä lainatusta sanastosta koski ruokaa, ihmisiä, uskontoa ja kulttuurin muita aspekteja, näin ollen tutkimustulokset vastasivat jokseenkin hyvin aiheesta jo kirjoitettua teoriaa. Sanastotutkimuksessa kävi ilmi, että vaikka Singaporen englannissa esiintyy muita lainasanoja suhteellisen harvoin, ovat keskustelupartikkelit olennainen osa maassa käytettyä englanninkieltä. Tutkimus osoitti myös, että lainasanoja esiintyy huomattavasti enemmän puhutussa kuin kirjoitetussa kielessä, ja että todennäköisyys sille, että paikallissanoja esiintyy on suurempi vapaamuotoisessa kielessä kuin muodollisessa tai virallisessa kielessä.

Avainsanat: Singaporen englanti, Intian englanti, lainasanat, korpuslingvistiikka

Table of Contents

1. Introduction

Due to colonization and later globalization, English has become to be used widely around the world. In many of the former British colonies, regional varieties have arisen from contact between various indigenous languages and the language of the colonizers. *World Englishes,* such as Singapore English and Indian English, are nowadays considered to be independent varieties of English with their own standards ways of using the language rather than faulty versions of British or American English. Although many World Englishes have been quite extensively studied over the past few decades, there have been a lot more studies concentrating on the grammatical and phonological features of these varieties than there have been studies on their vocabulary. The interest of this thesis, therefore, lies in the lexis of such World Englishes since that area of research is less studied and more unexplored than the World English studies on grammar and phonology. The thesis will focus on two separate varieties, Singapore and Indian Englishes, and explore their indigenous vocabulary. I am interested to know how extensively people speaking Singapore and Indian English bring words from their language backgrounds into the English they use. I will focus only on *loanwords*, meaning words that are incorporated to other languages without translation. Singapore and Indian English were chosen for the study because I want to compare indigenous vocabulary in two varieties of Englishes coming from roughly the same area (here South and Southeast Asia) but which have had rather different colonization experiences and evolution phases following Schneider's *Dynamic Model of Postcolonial Englishes* (2009).

The first part of the thesis will concentrate on providing a historical background of the linguistic situation and use of English in the regions, as well as a cross-section of the functions of English in the countries today before presenting previous studies on World Englishes and indigenous vocabulary. In the sections concentrating on sociolinguistic history of the varieties I will make use of earlier studies conducted on Singapore and Indian Englishes but also rely on more general studies

and theories concerning World Englishes. A few specific studies and theories, such as Chandrika Balasubramanian's vocabulary study on Indian English (2009) and Edgar W. Schneider's theory and *the Dynamnic Model of Postcolonial Englishes* (2007)*,* which deal with the emergence of indigenous words into regional varieties of English, will be used as points of reference throughout the thesis to create coherence between the theory and the research parts of the thesis.

The material studied comes from the Singapore and India components of *the International Corpus of English* (ICE) - each corpus consists of 300 spoken and 200 hundred written texts about 2,000 words in length. Since the corpora have been comprised the same way and are roughly the same size, the comparison between the results should be valid and telling. The assumption before commencing on the research is that indigenous words will be found and that the spoken sections will contain more indigenous loanwords than the written sections of the ICE corpora. Although the corpora are quite small, they should still provide the study plenty of results to analyze since results from both corpora will first be looked at separately and then compared with each other and analyzed using previous studies and theory on indigenous vocabulary on World Englishes. In processing and categorizing results gathered from the corpora, I will loosely follow Balasubramanian's previous study conducted on Indian English using *Corpus of Contemporary Indian English* (CCIE) (2009).

The aim of this thesis is to tap into the less studied field of vocabulary in connection to World Englishes and to make further conclusions about loanwords in Singapore and Indian Englishes based on the corpus study conducted on the Singapore and Indian components of ICE. The specific goals of the thesis are the following:

1) To provide thorough background information concerning the sociolinguistic history of Singapore and India in order to explain the linguistic complexity of the regions today.

2) To investigate how extensively Singapore English and Indian English use words from indigenous languages. Also, to categorize the indigenous words used into lexical categories (e.g. food, clothing, people) to see in what context they are used the most.

3) To compare the findings of the varieties to see if differences arise and if they can be explained by the different sociolinguistic backgrounds of the countries.

2. English in Singapore

This chapter will first give a brief introduction into English in Singapore before providing a more detailed look into the many stages it took for the language to develop from a foreign language used by the colonizers to Singapore English being considered a variety of English spoken by majority of Singaporeans. Lastly, Singapore English and its division into Standard Singapore English and Colloquial Singapore English will be discussed in order provide a better understanding of the actual uses of English in Singapore.

Singapore is a city-state located in Southeast Asia at the southernmost tip of continental Asia between Malaysia and Indonesia. Because of its location and the multicultural background of the population, Singapore is nowadays a center of international commerce and transportation. English language has almost a 200-years-long history and a significant role in the Singaporean society. Since the colonialization of Singapore by the British in 1819, the role of English has increased, first slowly and then more rapidly through government strategies, especially during the last few decades. Singapore English is nowadays considered an English variety of its own. It is a World English variety influenced by the indigenous languages and mother tongues in the country – Mandarin Chinese, Malay and Tamil.

At present, there are four official languages in Singapore; English, Mandarin Chinese, Malaysian language Malay and Indian language Tamil. English functions as the language of the government, administration, education and finance. Chinese, Malay and Tamil are the appointed mother tongues of the three biggest ethnic groups in the country but do not have many functions in the society outside being the reported mother tongue of the equivalent ethnicities and being the language used at home and with family. Based on a policy issued by the Singaporean government, if you are of Chinese, Indian or Malaysian origin you must report your mother tongue as being the corresponding language of that ethnic group (Tan 2014, 319-20). Although a great number of people from these ethnic groups use English as their mother tongue, it cannot be considered their native

language because of the government regulation. What follows is that the assigned mother tongues do not necessarily represent actual mother tongues or proficiencies of the speakers. Based on data gathered by the government, English is actually replacing other languages as the language used at home and indigenous languages are starting to lose the little function they still had in the society, especially among the young speakers (Tan 2014).

2.1. History of English in Singapore

The history of English in Singapore is shorter than in many other countries colonized by the British. The English language originally came to the region with British colonizers and traders in 1819 when the British East Asia company founded a free port in Singapore, which at the time was only a primitive country with only a small population (Schneider 2007, 153-4). The ethnic segregation system which divides people of different ethnicities into groups of Malays, Indians and Chinese already existed (Lim 2004, 2-3). The population growth of Singapore started with the establishment of the free port and further accelerated when Singapore officially became a British colony in 1867 (Schneider 207, 153-4). Despite the massive influx of population into the country, the biggest ethnic groups remained the same (Lim 2004, 2-3) and the number of European settlers stayed relatively low compared to many other colonized regions (Schneider 2007).

The first colonizers who arrived in Singapore were traders and the promotion of English language in schools was not a priority at the time (Lim 2004, 3). Still, a few English-medium schools were established by 1830 and the number of people enrolled in them grew from 722 students in 1850's to 27,000 in 1937 before Japanese settlement and second world war, which momentarily halted the advancements in the education system (ibid., 3-4). At this time, however, English was mostly a language of the elite (Wee 2008, 259-60) and not spoken or used by the larger mass. Singapore officially gained its independence in 1965 and rather than getting rid of the language of the colonizers the government chose to keep English as a national language. After independence, the government

opted for reformations which further strengthened the status of English in the society "with the intention of keeping Singaporean society open to global and regional forces" (Lim et al. 2010, 3).

Singapore is no stranger to strict regulations and the government of Singapore has manipulated the language use of the population by launching several campaigns aimed at using specific languages for specific purposes while promoting the status of English. The variety of languages in used in Singapore used to be much broader than it currently is. In the 1950's, the population of Singapore was reported to speak 33 different mother tongues in contrast to the three they use today (Kuo 1980, 40-2). In 1950s and 60s, the government issued *The Official Languages policy* which determined the four official languages which are still in use today. One language was chosen to represent one ethnic group and the languages chosen as official languages (Tamil, Mandarin, Malay and English) were not based on the real number of users - for example, the majority of the Chinese spoke Hokkien not Mandarin and the other ethnicities also used various languages within their ethnic groups (Kuo 1980, 41-3). In 1966, the state continued increasing the importance of English when it started a *Bilingualism policy* aiming at English-ethnic language bilingualism (Bolton and Ng 2014) and appointed English as the language of workplaces due to "political and economic reasons" (Lee 2012, 59-60). Although the government had appointed Mandarin Chinese an official language status and this had resulted in decreased use of other Chinese languages, the state was still not happy (ibid., 158). In 1979, the state yet again launched a new language policy, called *Speak Mandarin Campaign* to urge the Chinese to use Mandarin in all aspects of life and to further unify the Chinese ethnic community and by the late 80s, most of the Chinese Singaporeans spoke Mandarin (ibid.). Initially, the four official languages were all used as mediums of instruction in schools (Bolton and Ng 2014, 309). In 1987, the state closed all the ethnic language schools due to decrease in attendance and chose English as the only medium of education while downgrading the ethnic languages to be taught only as subjects (Tan quoted in Bolton and Ng 2014, 309). The latest government language campaign called *Speak Good English* movement, launched in 2000, was set to

urge Singaporeans to use formal Standard Singapore English variety instead of informal Colloquial Singapore English not only in official situations and contexts but also in their everyday lives (Wee 2008, 264).

It is very likely that these radical language policies have affected the languages used in the country. A recent study by Tan (2014) confirms that the language planning has had consequences in unifying the ethnic groups but also in strengthening the role of English in Singapore. The young members of each ethnic group seem to nowadays prefer to use English in all domains of life much more than older generations (ibid.)

2.2. Singapore English

As mentioned earlier, Singapore English is a World English but terminology may slightly differ depending on the scholar, Schneider (2007, 2010), for example, uses the term *Postcolonial English* to refer to new English varieties arising in regions formerly colonized by the British. The term employed in this thesis is for newer varieties of English is 'World English' because it is neutral and widely used among scholars. One of the earlies theories and descriptions on World English varieties as compared to Standard Englishes is Braj B. Kachru's *Circles of English model* (1988). Kachru's Circle model makes a distinction between three different kinds of regions depending on the type of English used in the countries – *Inner Circle, Outer Circle* and *Expanding Circle*. The *Inner Circle* is constructed of countries where English is spoken as a native language and specifically those countries where an older 'standard' version of English is used, such as the United States and the United Kingdom. *Outer Circle* consists of countries and regions where English is often used as a Second Language and where a newer variety of English has developed into a standardized variety of its own, as in many Asian and African countries. Singapore and India both fall into the category of Outer Circle varieties. Countries on the *Expanding Circle*, on the other hand, use English as a Foreign Language. Figure 1 below is a modified representation of Kachru's original model illustrating the circles with a few example countries and varieties.

Figure 1*. A modified version of the Circle Model*
of World Englishes (Kachru 1988)
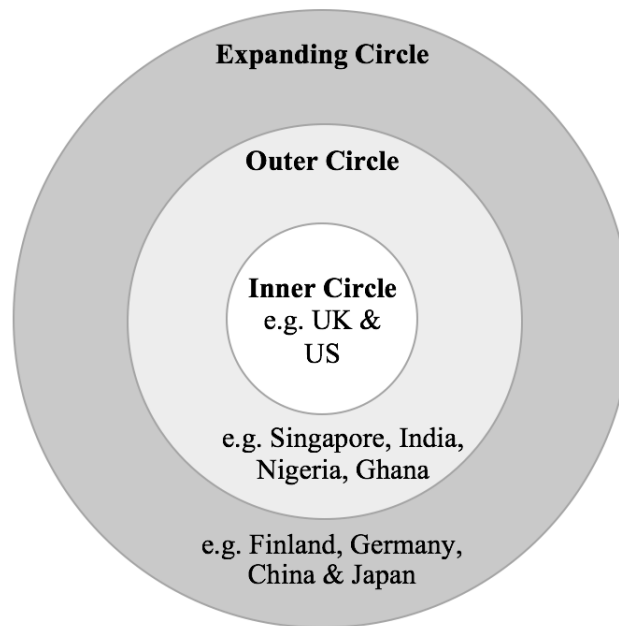
Before being accepted as a variety of its own, scholars often referred to Singapore English as a *creole* language (Lim 2004, 7-11). Following Mesthrie and Bhatt's (2008, 5) description, creoles arise in language contact situations where English is mixed with indigenous languages and are "fully developed speech forms, which show so much restructuring as to bear little resemblance grammatically to their lexifiers". Nevertheless, creoles may develop into their own varieties over time, as in the case of Singapore English due to high usage and position in the society (Schneider 2007).

Previously, it was thought that Singaporeans used three types of English - *acrolectal*, *mesolectal* and *basilectal* – based on their level of education (Deterding 2007, 87-8). An acrolect is a form that most closely resembles Standard English, mesolect contains some grammatical forms and features of indigenous languages and basilect is the creole form furthest away on a scale from Standard English (Schneider 2011, 97). Nowadays, since the majority of the population is capable of using English fluently, a more common theory of Singapore English often divides Singapore English into *Standard Singapore English* (SSE) and *Colloquial Singapore English* (CSE), also known as

Singlish (Lim 2004; Bao 2015; Schneider 2007). These two varieties are used for different purposes in different domains of life, in other words they occur in *diglossia* (Gupta 1989). The *high* variety, in this case SSE, is used in formal and official situations and the *low* variety, CSE, in more informal and spoken contexts (Leimgruber 2011, 2). While SSE does not differ much from Inner Circle Standard English, CSE is more nativized with influences from the other languages in the region (ibid.). According to Alsagoff (2010) in Singapore English, everything global is communicated in Standard Singapore English and everything local in Colloquial Singapore English and the speaker's stance is therefore also represented in the choice of variety used in a situation. Diglossia of Englishes in Singapore has also been called *register variation* between formal and informal registers (Bao and Huang 2006). In this thesis, a distinction between Standard Singapore English and Colloquial Singapore English will be made but as Figure 2 in the next subsection shows, they are not thought of as separate varieties or dialects but as different points of reference on a scale where the speakers may move depending on the speech situation.

2.2.1. Standard Singapore English

As said earlier, SSE is the high variety of Singapore English used in official situations. Gupta (2010) states that there is not much variation between Standard forms of World Englishes and Standard English and that grammatical "differences [between World Englishes and Standard English] are often exaggerated". According to her, variation between different regional varieties of English occur mostly in their vocabulary. Therefore, when describing features of different World Englishes, we are mostly concentrating on lower and more informal dialects of the varieties. For example, most of the features described as being characteristic of Singapore English occur only in Colloquial Singapore English and very rarely in Standard Singapore English. Because of this, there is no point in attempting to describe SSE more thoroughly since it can be thought of as a 'refined' version of Colloquial Singapore English containing only some of the regional features used more frequently in CSE.

Based on this, Figure 2 below was drawn to illustrate the aforementioned takes on Singapore English and linguistic variation between Colloquial Singapore English and Standard Singapore English. In the continuum, the space between CSE and SSE is the range employed by English-knowing Singaporeans. The continuum represents the assumption that a speaker does not have to choose between using one variety (or register) over the other but can move up and down the continuum in order to adjust the formality of the language depending on the speech situation. Based on the theory and the scale, speakers may end up using language which is somewhere between the varieties since there is no clear boundary between the two varieties or registers.

Figure 2. Standard Singapore English on a continuum from CSE to SE
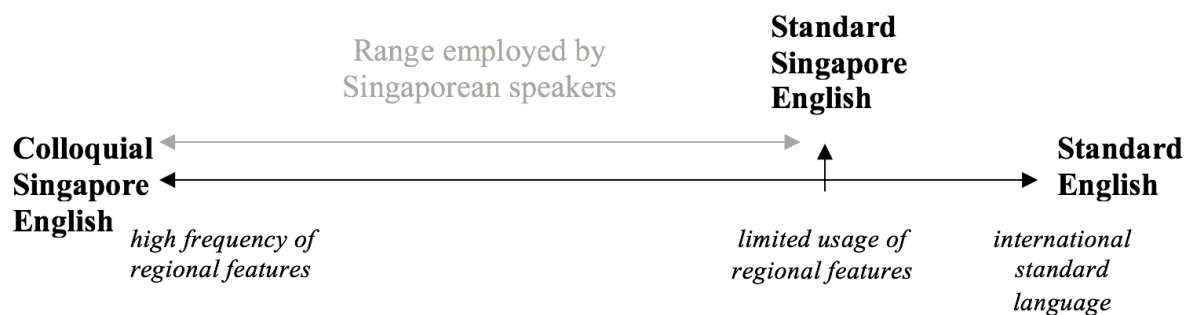
2.2.2. Colloquial Singapore English

Most of the Singaporeans use Colloquial Singapore English as opposed to Standard Singapore English as their language of communication in their daily lives. CSE is more distinguishable from Standard English than SSE because its pronunciation, grammar and lexicon have been more influenced by language contact with indigenous languages (Lim 2004).

What is common in Colloquial Singapore English and other World Englishes is that there can be significant differences in the word order and verb construction as compared to Standard English (Mesthrie and Bhatt 2008). Verbs are often left uninflected for number and tense, which are then only discoverable from the context of the sentence (Wee 2008b; Lim 2004). Especially in spoken

language, subject and object noun phrases as well as copular verb *be* might be left out altogether (Lim 2004). Although the SVO (subject-verb-object) word order seems to generally be applied in Singapore English, the object topic of the sentence can sometimes be preposed to the beginning of the sentence, especially in colloquial speech (Leimgruber 2011, Wee 2008b, Lim 2004). According to Tan (2003) this particular feature exists also in Chinese and Malay. The optionality in using articles is a feature shared by CSE and many other World Englishes (Mesthrie and Bhatt 2008). Characteristics which are often mentioned as typical for Colloquial Singapore English include features such as *already* and other words as markers of perfect tense, postponed *one* as a relativizer in relative phrases, non-concord tags such as *is it* and *can or not* added to the end of the sentence to form a question and ways of passivizing sentences that differ from Standard English (Wee 2008b; Lim 2004). One of the most significant features of Colloquial Singapore English is the use of various indigenous discourse particles (Wee 2008b; Lim 2004; Leimgruber 2011). Discourse particles *lah, ma, wat, meh, leh, lor, hor* and *hah* are frequently used to convey attitude and mood of the speaker (Wee 2008b; 604).

It has been argued that the government issued Speak Good English movement could lead not to the demise of Colloquial Singapore English but to the reduction of some of the most essentially Singaporean features in CSE (Wee 2008b, 606-7). On the other hand, based on Schneider's (2007) model, Singapore English is in the *Endonormative Stabilization* phase of development during which variation within the variety is often radically reduced so some features may naturally disappear from use even without government intervention. The next and last phase in the model again allows more variation and differences in using the language (Schneider 2007) but since Singapore English has not reached the last phase, it remains to be seen whether it will develop to be more heterogeneous than it currently is.

3. English in India

The structure of this chapter will closely follow the structure used in the previous chapter. It will start by briefly introducing English in India before giving a more thorough description of the history of English in India and functions of English in the country. The latter subsections will then aim at describing the Indian English variety in its complexity to create better understanding of the differences between the two varieties studied in this thesis.

The difference in the sizes of the regions studied is enormous, since India is the seventh largest country by size in the world (United Nations 2013). It is located in South Asia and shares a border with multiple countries including Sri Lanka, Pakistan, China, Nepal, Bhutan, Myanmar and Bangladesh. It is not only large in size but also in terms of population – India is the world's second most populous country just after China (United Nations 2015). It consists of 28 states and seven Union territories, which have been formed on the basis of linguistic background. Each of the states and territories also have their assigned official languages.

India is the third largest country by the number of English-users (Bhatt 2008, 548). It serves as a subsidiary official language with Hindi, which is the most used language in India. English has been used in India since the early 17th century, when the British East Asia Company arrived in India. The role of English in the society has remained strong even after India gained its independence from the British Empire in 1947, although there have been various attempts at diminishing the "controversial" status it has gained (Sedlatschek 2009, 1). Due to its colonial background and low number of mother tongue speakers, it has even been said that English enjoys a "disproportionate role" in the country where it is mainly used by people in larger cities and their immediate surroundings (Sailaja 2009, 2).

Nevertheless, Indian English is nowadays considered to be a variety of its own despite the fact that its classification as an English variety has been questioned from time to time (Sedlatschek 2009, 2). The challenging views highlight the variation in English used by people in different regions

and ethnic backgrounds, and also the wide range in the level of proficiency of its users (Balasubramanian 2009, 7-11).

3.1. History of English in India

As mentioned above, English language first arrived in India with the British East India Company in the 17th century. At the time, Portuguese was the main lingua franca between the European traders and Indian natives and English was used almost solely by the British traders (Sailaja 2009, 96-7). English slowly gained some ground but for some time it was merely used as a lingua franca between the British traders and the few Indians capable of using it (ibid.). The early traders and missionaries arriving in India were not interested in establishing English-medium schools but were there just for business (Sailaja 2009, 98) so the language did not spread as quickly as it did in some of the other countries colonized by the British. Indian-English bilingualism started growing more rapidly only during the 18th century through the continuing trade and conversion of Indians to British and western values (Bhatt 2008, 547).

Although the East India Company and English language had roles in the Indian society since their arrival in 17th century, it was not until 1765 that the British established political control in the country (Bhatt 2008, 547). In spite of the British gaining control over India and smaller attempts at teaching English at schools, the spread of English language only truly began in 1835 after Macaulay published his *Minute* (1835) on Indian education, where he argued for the necessity of Western education and English medium instruction in schools in India. Shortly after the publication of *Minute*, English became to be used as a medium of instruction in schools and additionally as lingua franca in the country (Balasubramanian 2009, 7) due to different government issued policies which concentrated on building an education system which emphasized the importance of English and opening lower governmental jobs to Indians (Sailaja 2009, 107-8), thus increasing the number of English-speaking Indians. Later, after the Magna Carta of Indian Education recommendations in 1854, education aims of the country shifted so that education in general was now offered to a wider

population (ibid.). Universities were established and English was chosen as the language of higher education, whereas vernacular languages were encouraged in lower education (ibid., 108). The Independence of India in 1947 was preceded by the birth of nationalism in India, which also included opposition against the status of English in the country. Nationalist leaders, such as Gandhi, promoted the use of Hindustani instead of English. The controversial English education and language actually helped in unifying the population and creating ideas about liberty among the nationalist movement (Sailaja 2009, 108). India is not alone in using the language of colonizers to their benefit since it has been claimed that in many former colonies the English-knowing locals used English as a tool to separate themselves from under the crown (Kachru 1988). English, therefore, often served as a unifier language among people with different mother tongues and strengthened the bond between the population, which in turned helped create national uprisings and start the process of anticolonization (ibid.)

English gained an official language status, alongside Hindi, in 1950 when it was appointed an official language of the Union in the Constitution of India. Although English has a significant nationwide role, the states are free to choose their own official languages for state internal purposes and most of the states have opted for local languages (Sailaja 2009, 4-6).

It should not be surprising then, that in 1991, when the population of India was around 889 million (The World Bank), only 0,02% of the people reported speaking English as a mother tongue (Census of India 2001) but the numbers are quite different when looking at those who speak English as a second or a third language. As a second language, English is the second most spoken language after Hindi with 86 million speakers and it is also the language Indians most frequently learn as their third language with 39 million users (Rukmini 2014).

According to Sailaja (2009, 5), "[t]he primary areas of English use are […] education, administration, law, mass media, science and technology" and it is also the preferred language in the field of business. Mehrotra (qtd. in Sedlatchek 2009, 21) adds that English is also widely used in

bigger gatherings with family and friends, in the field of sports, in politics and even for religious purposes. Although English is most frequently used among what can be called the "educated elite" it must be noted that English has become to be used also in less formal situations and everyday life (Sedlatchek 2009, 22) so the original elitist nature of the language has somewhat diminished and English serves as "an intranational link" among the middle and upper classes in the Indian society (ibid., 23-4), although mostly only in bigger cities.

3.2. Indian English

As mentioned earlier, Indian English is an Outer Circle variety of English based on Kachru's Circle Model model (see Figure 1), meaning that in general, people in India speak English as a Second Language and Indian English is considered a variety of its own. Its placement on the Circle Model, however, does not mean that it functions as a second language for all Indians because for some (especially those who have had to relocate from their home regions) it may be a first language and secondly because there are a lot of Indians who do not speak the language at all (Sailaja 2012, 360).

While Singapore English is often almost unanimously divided into Standard Singapore English and Colloquial Singapore English (terminology might vary depending on the scholar), no clear classification of different registers or dialects in Indian English seem to exist. Bhatt (2008) makes a distinction between Standard Indian English and Vernacular Indian English, which resembles that of Singapore English. McArthur's (2002) distinction is roughly the same but he uses the terms Standard and Non-Standard Indian English whereas Sailaja (2009) recognizes standard, non-standard and informal uses of Indian English. Some scholars do not even agree that Indian English exists while others argue that we should be talking about Indian *Englishes* instead of uniform Indian English (Sailaja 2009, 13). In addition to standardized Indian English, Kachru (qtd. in Sailaja 2009, 112) also mentions regional and occupational Indian English Pidgins – Boxwallah English, Babu English and Butler English which are pidgins or "broken Englishes" with roots in colonial India under the British Empire. Whatever the division or terminology, it seems to be evident that there is no uniform Indian

English. In this thesis, a division will be made between formal and informal registers and terms *Standard Indian English* and *Vernacular Indian English* will be employed while keeping in mind that the registers are only attempts at simplifying the complex and heterogenous nature of the Indian English variety as a whole.

3.2.1. Standard Indian English

It is hard to define Standard Indian English, because the written form of educated Standard Indian English greatly resembles Standard English and studies exploring spoken Standard Indian English are scant (Lange 2012). According to Sedlatchek (2009, 27), "educated IndE is determined by the high proficiency of its users, its wide range of functions in Indian society, and supraregional intelligibility". In that case, educated Indian English can be thought of as an equivalent to Standard Indian English, since terms regarding Indian Englishes often differ depending on the scholar. Many of the descriptions of Indian English only contain lists of features appearing in connection to Indian English, and not many extensive descriptive studies have been conducted (Sedlatschek 2009, 28). There is also no agreement in India whether the target standard for grammar should be Standard English or the *standard* Indian English used in newspapers and by the educated people in India (Trudgill and Hannah 2002, 129). Pingali (2012) states that what is considered standard in Indian English is the language used by the "proficient" speaker, which sounds like a tempting and easy description but is rather complicated since it involves defining who constitutes as a proficient speaker. The features used by the proficient speakers are said to be the standard in India but since there is no clear homogeneity in Indian English (even in standard use), simplifications and illusion of homogenous language use are often found in descriptions of Indian English (ibid.)

In sum, there hardly exists a unified standard version of Indian English but what is often referred to as Standard Indian English is the language used by a minority of Indians, specifically educated Indians in highly formal situations and by the written media. Figure 3 is my rough attempt at representing the types of Englishes used in India, based on the aforementioned and following pieces

of theory, when the categorizations of Standard Indian English and Vernacular Indian English are used.

Figure 3. Representation of Englishes used in India
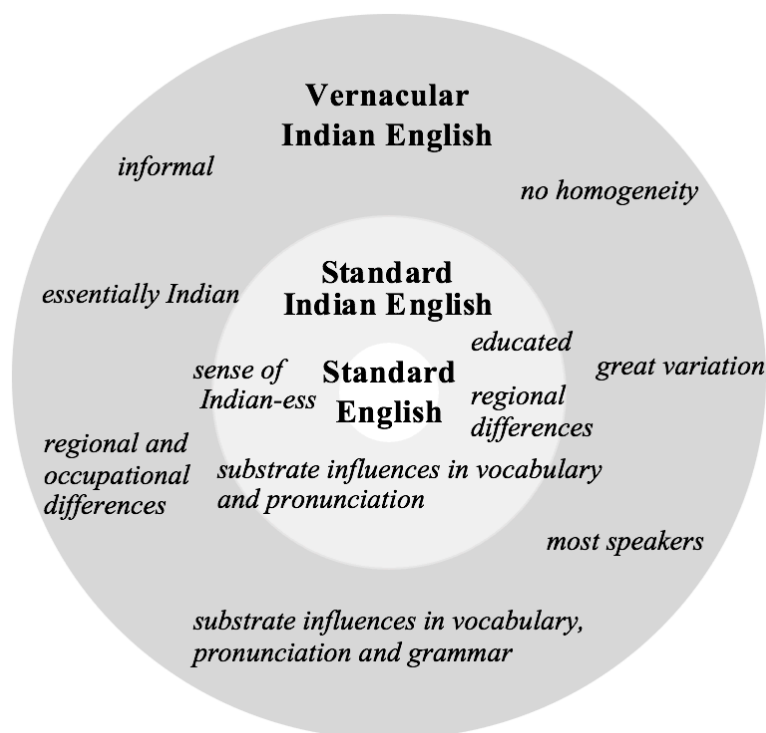
3.2.2. Vernacular Indian English

Since spoken Indian English greatly varies depending on the person speaking it, it can almost only be said that in general, Vernacular Indian English manifests the true "Indian-ess" of English used in India and is much more affected by Indian culture and indigenous languages of the region than Standard Indian English (Bhatt 2008). There is no uniform Vernacular Indian English and it is predominantly used in informal "low" register situations (ibid., 546). Contrary to the situation in Singapore, the majority of Indians do not possess the ability to move between high and low registers of Indian English, more clearly between Standard and Vernacular Indian Englishes, but only use a vernacular variety in all domains where English language is needed and used.

4. Previous Studies on World Englishes and variations in question

This chapter concentrates on providing information about previous studies on World Englishes and indigenous loanwords in World English varieties. In addition to that, it will also give a closer look at two works relevant for the research and discussion parts of this thesis.

4.1. Studies on World Englishes

The study of World Englishes is relatively new field of study, mostly because new varieties of English started developing around the world and standing out from Standard English only after the decolonization of the British Empire in 1940's when many of the former colonies gained their independence. It was not until much later, in 1980's, when the varieties of English arising in postcolonial regions started to be more comprehensively studied. Görlach, Kachru and McArthur created similar models of World Englishes which included older standard versions of English, such as American and British English, and regional varieties in Africa, Asia, South America and the Caribbean (Jenkins 2009, 17-20). Kachru's Circle Model was already briefly presented in Section 2.2. of the thesis, and out of the three models it is the one that is still often referred to in textbooks and studies of World Englishes today. It still being used today might be due to its "vagueness" compared to the other two models and the fact that it includes a place for the regions where English is spoken as a Foreign Language as well as English as a First and Second Language. What has been brought up regarding these types of models and studies on World Englishes is that they are not as relevant today as they previously were due to their monocentricity, meaning they are models with one center, some type of Standard English which is often British and/or American English. It has been suggested by Kachru himself (2005, 18-9), that a pluricentric model might better represent the situation today as he makes a distinction between "norm-providing" and "norm-dependent" varieties. Norm-providing countries or regions are those that have a somewhat established variety of English which most often influence the use of English in norm-dependent countries and especially those

geographically close to them (ibid.). By this definition, Singapore English and Indian English are both considered norm-providing varieties (ibid.).

Another model that is widely referred to today is Schneinder's Dynamic Model of Postcolonial Englishes (2007) which places World English varieties on a scale according to the phase they are in regarding the process of standardization. Schnenider's model will be more thoroughly explained in the next chapter. After the 80's, a growing number of studies have been conducted on different varieties of English and their features and characteristics. Many of the studies frequently focus on phonological or grammatical features and extensive studies on vocabulary alone are scant.

4.2. Schneider's Dynamic Model of Postcolonial Englishes

Schneider's (2007) model of *Postcolonial Englishes*, which is the term he uses for World English varieties, is a model illustrating development and standardization within World Englishes arising in language contact situations. Schneider mentions that previous studies have more or less concentrated on language contact and World English development on separate varieties and have not focused on finding out what is common about to all of them (ibid., 29). The model gives emphasis to the sociocultural aspects and events leading up to the changes as well as effects the events have in the language use of the people (Schneider 20017). Additionally, "social identities" of the indigenous people of the region and the British colonials are essential in the model and many of the phases of World English standardization are in his opinion explainable by these two social group identities coming closer together and starting to see themselves as sharing a "social identity" rather than being two separate groups (ibid.). The model consists of five phases or rather "sequences[s] of characteristic stages" which World Englishes go through in the process of becoming their own varieties (ibid., 29-33). Schneider acknowledges that the model is a simplification of a complex process and that in reality phases might often overlap or sometimes be even skipped completely.

Most of the World Englishes today originate from British colonial times (hence Schneider's term *Postcolonial English*) and Schneider uses Mufwene's (qtd. in ibid. 24-5) definition regarding

the types of colonialism as he distinguishes between three different patterns of colonialization. These colonization types are *trade colonization*, *exploitation colonization* and *settlement colonization* (ibid.). According to Schneider, trade colonies are set up to conquer central trading posts and activities, and to secure trade routes (ibid., 66). There is hardly any influx of colonizer population into trade colonies and the indigenous people often learn a pidginized version of English at first to collaborate with the colonizers but in general inter-ethnic communication remains low (ibid.). Later, English is often chosen as the lingua franca of the inter-ethnic population of the region but the lower version of the language remains in use but is frowned upon by government authorities (ibid., 67). In exploitation colonies, on the other hand, the colonizer population is somewhat bigger than in trade colonies but at the same time, not as big as in settlement colonies (ibid., 65-67). However, political and economic power of the colonizers is relatively strong in most exploitation colonies and the goal of the colonization is to spread the economic and political power of the colonizers (ibid. 65). Larger scale spread of English language is often not wanted by the colonizers in fear of rebellion and disobedience by the locals, and English remains an elitist language in the region (ibid. 65-6). After colonization, vast majority of the colonizer population leaves the region but English still remains as a language in education and of wider communication (ibid. 66). Based on the theory, Singapore is a trade colony and India a case of exploitation colonization. Different types of colonization histories of Singapore and India is also the reason Singapore and Indian English varieties were chosen for this thesis because I wanted to study two varieties from roughly the same region but with two distinct colonization experiences.

The main focus of the Dynamic Model, however, is the five phases English varieties go through in their process of becoming fully functional varieties of their own (Schneider 2007). The phases of the model are (1) *foundation*, (2) *exonormative stabilization*, (3) *nativization*, (4) *endonormative stabilization* and (5) *differentiation* (ibid.). I will now briefly explain the phases with focus on vocabulary and other things central to this thesis.

The first phase starts when English language arrives in the region with the earliest traders or colonizers (Schneider 2007, 33-6). During this phase there is very little communication between the locals and the colonizers since they do not have languages in common, but it is also at this time that heavily pidginized versions English often arise among a minority of the locals to meet the demands of communicating with the colonizers (ibid.). In the course of foundation phase, indigenous place names are the first local language lexical items to be adopted into English spoken by the colonizers (ibid.). The second phase begins after the colonies have been founded and English has come to be used more commonly and in many cases it is even the administrative language of the region (ibid. 36-40). During this phase, there is more inter-ethnic communication between the two groups and their social identities start to slowly merge together (ibid.). Vocabulary vice, a great number of indigenous loanwords are being adopted into English (ibid.). The first loanwords are often connected to the nature of the region and are, for example, words for animals and plants (ibid.). According to Schneider, the third stage, nativization, is the most important phase in the model (ibid., 40-8). At this time, the colonized nation is often heading towards independence and the identities of the colonizers and locals are more bound together so that there is no longer a "us vs. them" positioning between them (ibid.). As the identities of the people are more intertwined, so are the languages (ibid.). At this point, the English language used by the colonizers and the locals resembles each other a lot, and so the English in the region becomes nativized as it absorbs more features from the indigenous languages (ibid.). An influx of cultural loanwords, as well as other words, enter English and make it harder for foreigners to understand (ibid.). As nativization occurs, talks of "correctedness of language" arise by scholars and people in authority, although as time goes by, the localized versions of English become more commonly accepted (ibid.). The fourth phase of the model is often connected to the independence of the formerly colonized nation (ibid., 48-52). The identity of the people in the independent nation is nationalistic and one, and the division of people between the colonizers and locals is no longer relevant (ibid.). The English variety used in the nation is standardized but it does

not mean that variation is non-existent (ibid.). Standardization often leads to the emergence of official dictionaries of the variety (ibid.). The final phase of the dynamic model is reached when the stabilized and uniform language results in different dialects emerging (ibid., 52-5).

As an example, American English and British English are varieties that have reached the final phase in Schneider's Dynamic Model. The varieties studied in this thesis, however, are not as clear cut and a little bit harder to place on the model. According to Schneider, Singapore English manifests many of the characteristics of and can be considered to be going through endonormative stabilization (ibid., 153-61). Pointing out the phase of Indian English on the model is far more complex and boundaries between phases are fuzzier. Schneider places Indian English in phase 3 but mentions that the variety manifests some characteristics of phase 4 as well (ibid., 161-73). What has halted Indian English in the stage of nativization is that English in India is only used by a small minority and it is not connected to the population's national or linguistic identity (ibid., 171). On the contrary, Singapore English, which has reached phase 4, is used by the majority of Singaporeans and is considered to be a part of the population's national identity.

In the discussion section of the thesis, the results will be viewed giving consideration to differences between Singapore English and Indian English regarding the colonization type and place on the dynamic model.

4.3. Balasubramanian's Register Variation in Indian English

Chandrika Balasubramanian's (2009) corpus study on the vocabulary of Indian English conducted on the *Corpus of Contemporary Indian English* (CCIE) is the main work relevant for my own corpus research since I plan on roughly following her categorization of the results, and also because the results of her study provide an excellent point of reference for this thesis.

The vocabulary study Balasubramanian conducted was part of a larger scale study on Indian English but since the rest of the study is not relevant for this thesis, it will not be further discussed. What is of interest to the thesis are only the results from the vocabulary study. Balasubramanian

searched the CCIE for loanwords from Indian languages and categorized them based on their semantic meanings. The categories she used to organize the result tokens are 'Food', 'Clothing', 'Music (Arts)', 'Religion', 'People', 'Discourse markers', 'Greetings', 'Politics', 'Larger chunks' and 'Other'. Balasubramanian does not give the concrete number of results in the corpus but uses normalized frequencies and presents the results a little differently than I plan to in this thesis so there is no point in focusing on the numbers but rather on which of the categories provided the most results and in what part of the section loanwords occurred the most in.

Based on the results of her study, the most used loanwords are words relating to food, people, religion, and music and arts, and politics. Loanwords for types of clothing also occur in CCIE but not as frequently as words in the previously mentioned categories. Noticeable about Balasubramanian's results is that indigenous discourse markers seem to be used rather infrequently, but their low occurrence is easily explainable by the large proportion of written material in the corpus.

In her research on Indian English in CCIE, Balasubramanian found that indigenous loanwords are used more often in spoken than in written language. A staggering 85% of the indigenous words occurred in the spoken language sections of the corpus.

Balasubramanian's (2009) study is one of the few larger scale studies conducted on the vocabulary of a World English variety and therefore relevant to this thesis. Balasubramanian's methods and categorization will be explained more thoroughly in the next chapter of this thesis and the results of her study will be looked back at later in Chapter 7.

4.4. Indigenous vocabulary in World Englishes

The methods of creating new words in general in languages apply to varieties of World Englishes as well. New words are formed by borrowing from indigenous languages, blending different lexemes, clipping or shortening words, creating acronyms, compounding, creating hybrid compounds of words from different languages, new collocations, via semantic shifts (Schneider 2007, 78-82) or by loan translations (Kachru and Smith 2008, 103-10). In many Outer Circle Englishes, new words are created

or adopted from indigenous languages because no existing word in Standard English can convey an idea, word or an object usually connected to the culture of the region (ibid., 106). According to Kachru and Smith (ibid., 110), nativization of the vocabulary is one of the key points in acculturation and nativization of the whole regional variety of any World English. In their opinion, "the borrowings from local languages make English more relevant to local contexts [and] they also serve ideological purposes" (ibid., 190). In this thesis, the concentration is solely on loanwords, words that borrowed as such from indigenous languages, such as, *padi*, which is a Malay word for 'rice' and sometimes used in Singapore English, or *lakh,* a word used in Indian languages meaning 'one hundred thousand' adopted into Indian English.

Kachru and Nelson write that in South Asian Englishes (under which categorization Indian English belongs) the indigenous vocabulary often consists of words related to nature and places, government and legal systems, as well as transaction (2006, 158). As Schneider's (2007) opinion is that most indigenous loanwords in World Englishes are culture and nature words, and Balasubramanian's (2009) results point that in Indian English, loanwords are mostly used for food, people, religion, music and arts, as well as politics, it can be said that comments and research about loanwords are somewhat conflicting. Matras (2009) also adds that content words are not the only words borrowed. In fact, the most easily borrowed items in language contact situations, especially in bilingual and diglossic situations, are discourse markers (ibid., 193). It will be interesting to see what differences or similarities arise when conducting a vocabulary study on two distinct varieties based on corpora which have been comprised exactly the same way.

5. Data and Methods

In this chapter I will present the corpora where the studied material and words will be gathered from and go through the processes and methods employed in conducting the research for the thesis.

5.1. ICE Corpora

The corpora used in the study are the Singapore and India components of *the International Corpus of English* (ICE). The ICE project was first initiated in the 80's and launched in early 1990's by Sydney Greenbaum to help the research on Englishes in countries and regions where English was not the first language. The project now includes research and corpora of various countries and varieties such as Australia, Cameroon, Canada, East Africa (Kenya, Malawi, Tanzania), Fiji, Great Britain, Hong Kong, India, Ireland, Jamaica, Kenya, Malta, Malaysia, New Zealand, Nigeria, Pakistan, Philippines, Sierra Leone, Singapore, South Africa, Sri Lanka, Trinidad and Tobago, and USA.

Each corpus component in the International Corpus of English is comprised the same way. They consist of a spoken and a written section which are further divided into more specific subsections regarding their style, formality, situation and subject content. The sections and subsections are represented more thoroughly in *Table 1* below. Each ICE corpora is comprised of 500 texts with 300 texts in the spoken section and 200 in the written. The texts are all approximately 2,000 words in length so the total word count of each component should be around 1,000,000 words. Comprising a word list with AntConc and checking the number of word tokens in the corpora reveals that ICE-SIN has 1,106,186 words while contains ICE-IND 1,120, 340 words.

Table 1. Construction of International Corpus of English corpora

SPOKEN (number of texts 300)	DIALOGUE (180)	**Private** (100)	• Direct conversations • Telephone calls
		Public (80)	• Class lessons • Broadcast discussions • Broadcast interviews • Parliamentary debates • Legal Cross-examinations • Business transactions
	MONOLOGUE (120)	**Unscripted** (70)	• Spontaneous commentaries • Unscripted speeches • Demonstrations • Legal presentations
		Scripted (50)	• Broadcast news • Broadcast talks • Non-broadcast talks
WRITTEN (200)	NON-PRINTED (50)	**Non-professional writing** (40)	• Student essays • Examinations scripts
		Correspondence (30)	• Social letters • Business letters
	PRINTED (150)	**Academic writing** (40)	• Humanities • Social and Natural sciences • Technology
		Non-academic writing (40)	• Humanities • Social and Natural sciences • Technology
		Reportage (20)	• Press news reports
		Instructional writing (20)	• Administrative writing • Skills and hobbies
		Persuasive Writing (10)	• Press editorials
		Creative Writing (20)	• Novels and stories

Although the corpora in the ICE project are not very comprehensive regarding their size, their design is what makes them valuable. Because each corpus is comprised the same way, it is possible to conduct relatively accurate comparisons between different varieties of English. This is also the reason why ICE corpora were chosen for this thesis although there are multiple corpora of both Singapore and Indian Englishes that are larger in size but that would be harder to compare. Comparing ICE-SIN and ICE-IND together will guarantee results that are comparable as they are and even without using normalized frequencies. The division of the corpora into sections and subsections will also be useful since in addition to finding indigenous words in the corpora I also plan on finding out and comparing in which section of the corpora they occur in. Comparison of this kind would not be possible using any other corpora.

ICE corpora material is gathered largely from bigger city areas with educated people, therefore in conducting a study on them, it must be kept in mind that the language used is probably more standard than the language used by people in those countries on average. This, however, does not mean that the corpora are representative of the standard variety of the counties or regions but that they are just more formal in nature. Also, that is not to say that informal language in the corpora is not used, and certain sub-sections, such as 'private dialogue' and 'unscripted monologue' in the spoken section may at times contain language that is extremely non-standard.

5.2. Research methods

In this thesis, it is my aim to find the indigenous loanwords used in Singapore English and Indian English in ICE-SIN and ICE-IND and to compare the results to see if one variety uses more words from local languages than the other. In addition to that, I also want to know what are the words used and to which semantic category they belong to, as well as in which section or subsection of the corpora they occur in.

The data and words from the corpora was gathered using AntConc software by creating a word list of all the texts in the corpus. A wordlist creates a list of all the words used in the corpus

based on the number of times they are used in the texts, and based on that frequency the words are put in order with higher frequency words, such as articles, at the start of the list and words that occur in the texts only once at the end. After creating the list, I went through all the words in the list and further examined the ones that seemed indigenous. Indigenous words were relatively easy to spot from the word list but each word needed to be inspected more closely in order to decide whether the word was a proper noun, part of a longer indigenous phrase or sentence or a sole indigenous word used in otherwise English phrase or sentence. Although indigenous proper nouns, especially place names, were plentiful, they were not taken into consideration in this thesis because the aim of the thesis was to focus on indigenous words that more spontaneously occur in the middle of Singapore and Indian English phrases and sentences. For that same reason, longer phrases and sentences which occurred in the texts were also disregarded. Indigenous words in ICE-IND are tagged with <INDIG> marker but the same method of gathering the words as in ICE-SIN was still used because the words were tagged using slightly different criteria than is employed in the thesis.

If a word was concluded to be an indigenous word stemming from one of the languages used in the region, the meaning of the word was further examined. This was done using various unofficial and official dictionaries on Singapore and Indian English but also on the indigenous languages, as well as "googleing" the word and examining the results if nothing conclusive was found on the dictionaries. Although using Google may not be the most formal way of conducting research, it proved to be very helpful regarding the aim of this thesis since some of the words used were rather colloquial and therefore could not be found in the dictionaries. Googleing proved to be useful also because many of the words were not spelled the way they occur in dictionaries and in such cases Google often suggested the right spelling after trying to perform a search with the form found in the corpora. Also, many words for food were not found in any dictionaries but could easily be discovered otherwise searching for them on the Internet.

Some words were used in the same text multiple times in a way that would skew the research. In that case, I used Balasubramanian's (2009, 72-3) approach to them – if a word was repeated (by the speaker themselves or another speaker) in the immediate context of the first occurrence, the word was disregarded but if the word was later used in the text at a different time, it was picked up again. Another criterion on occurrences of the same token within a text was also employed – if a word was used within the same text more than five times, only five tokens were recorded. The criterion was employed because some of the texts repeated certain words for tens of times but at the same time, large majority of the tokens occurred within a text a maximum of five times. However, this criterion was not used with discourse markers since their use is more spontaneous and multiple occurrences of the same discourse marker in the same text is not considered as repetitive.

Uses of indigenous words in parenthesis or apostrophes were not picked up from the corpus. Particularly, some of the texts featured explanations of Singapore and Indian Englishes and their characteristics, and included lists of words frequently used in the varieties. This type of explanatory uses of the words were ignored because the indigenous words did not occur in them "naturally".

When indigenous words were identified, they were taken down for further analyzation and placed in different categories depending on their lexical meaning. Simplified versions of the excel worksheets used in processing and categorizing the results are attached to the thesis as appendices. Initially, I used the semantic categories Balasubramanian (2009) employed in her study of Indian words in the Corpus of Contemporary Indian English but took out and added some categories to better represent the results of my own study. This was done because some categories, such as 'Greetings' only had a few tokens, and many words in 'Other' category could be better placed in a category of their own. For example, there were various instances of words for different animals, plants and trees, and terms inherently part of the culture of the regions. Balasubramanian's categorization also included a category called 'Larger chunks of language' for indigenous utterings that were between two and 10 words in length. I did not find this category relevant for my thesis so I left it out from the

study altogether. Balasubramanian's original categories and the modified categories used in this thesis are represented in *Table 2*. What can be seen is from it is that, as mentioned, 'Larger chunks' and 'Greetings were taken out and categories 'Animals/Nature' and 'Culture' were added.

Table 2. *Balasubramanian's categorization (top) vs. my categorization (bottom)*

Food	Clothing	Music/Arts	Religion	People
Discourse markers	Greetings	Politics	Larger chunks	Other

Food	Clothing	Music/Arts	Religion	People
Discourse markers	Animals/ Nature	Politics	Culture	Other

At the same time as going through the word lists and examining indigenous words and categorizing them, additional information about them was also gathered. As mentioned earlier, in addition to finding out which words were used, the thesis set out to see in which section of the corpora the words occurred in. The information about the text was picked up from AntConc and put aside for further analyzation in order to draw tables for the results section of the thesis. The type of information picked up regarding the indigenous words in the corpora and the indigenous words occurring in ICE-SIN and ICE-IND are found in Appendix 1 and Appendix 2 attached to the thesis. The analysis of the registers the loanwords occur in involves using parts of the corpora to be representative of smaller subsections of ICE-SIN and ICE-IND. The types of registers, subsections and texts in the corpora can be seen in Table already presented above in Section 5.1., but *Table 3* below is a representation of the subsections of the corpora used in analyzing the results of this thesis. The choice in using the subsection labels in question is due to not wanting to go into too specified details about the occurrence

of the loanwords but also needing them to be specific enough so that accurate conclusions about registers can be made.

Table 3. *Subsections of ICE corpora used for categorizing results*

SPOKEN	Private dialogue	Public dialogue	Unscripted monologue	Scripted monologue
WRITTEN	Non-professional writing	Academic writing	Reportage	Persuasive writing
	Correspondence	Non-academic writing	Instructional writing	Creative writing

It should be noted that the word list examination process was conducted completely manually so it is possible that some words might have been overlooked. This may have led to a few words being left out from the study but since the word lists were examined carefully, the number of accidentally disregarded words should not be high. Some words resembling English words in their spelling might have also been left out since the word tokens for only indigenous looking words were further examined. Balasubramanian (2009, 72) also recognized this problem but mentions that words like that are rare. Categorization and choosing to include words and disregard others can also be considered quite subjective. Because of this, same criteria were tried to be employed throughout the research but having said that, some words might also be categorized differently and some words may not have been included or disregarded if slightly different criteria would have been used.

Lastly, a few other noteworthy comments about the categorization should be made. Following Balasubramanian's example, words relating to yoga were included in 'Religion' loanwords since yoga is closely connected to spirituality, especially in India. The category of 'Politics' includes names for local government systems, not found outside the regions, as well as words relating to the caste system in India which could have as well been categorized in 'People'. Additionally, 'Politics' does not include names for political parties since they are not considered as part of indigenous

vocabulary in this thesis. Also, completely excluded from the study are names of religions and terms for people partaking in the religions, such as *Hindu*, languages, for instance as *Marathi*, and names for peoples connected to the geographical area they originate from because such words were not considered to correspond with the aim of the thesis.

6. Corpus study

After reviewing the theoretical framework essential for the thesis, I will now move onto presenting the results gathered from the ICE corpora using AntConc and manually going through the word lists created. The results and tokens from both corpora will be first looked at separately in this chapter and the differences between the results will be further discussed in the next chapter of the thesis. At first, findings in the entire corpus will be examined with focus on occurrence of tokens in the previously mentioned semantic categories and corpus sections. Then, results in each semantic category will be reviewed more thoroughly while also presenting examples of the indigenous words occurring in the categories.

6.1. Findings in Singapore English

The overall number of indigenous loanword occurrences in the Singapore English component of ICE was relatively low as compared to Balasubramanian's results in CCIE and the results in the Indian English component of ICE. Based on the findings, indigenous discourse markers seem to be used quite frequently in Singapore English while other types of loanwords are far more infrequent, as can be seen in Table 4, which presents the results based on the semantic category of the tokens found in the corpus. Apart from 'Discourse markers', indigenous 'Food' words category is the most fruitful semantic category in regards to the number of tokens, with approximately triple the number of tokens as compared to the next biggest specified categories, which are 'People' and 'Culture'. 'Other' category consist of miscellaneous words that cannot be placed in the specified categories and due to its vague definition, it has a high quantity of tokens ranging from verbs to indigenous expressions. An additional 28 tokens retrieved from the corpus were not placed in any of the categories since the meaning of the words could not be deciphered. The results in Table 4 also indicate that while the number of tokens in categories are somewhat high, variations of same words occur frequently and the number of different words, or rather *lexemes*, is much lower. In table 4, the darker gray columns represent the number of tokens in total in the categories, and the adjacent lighter gray columns the

number of different lexemes among the tokens. Based on the Table, while the number of tokens in, for example 'Discourse markers' is high, only 13 different discourse markers are found within the corpus. The results in each of the categories will be discussed in detail later in the following subsections.

Table 4. ICE-Singapore results organized by the semantic category of the tokens
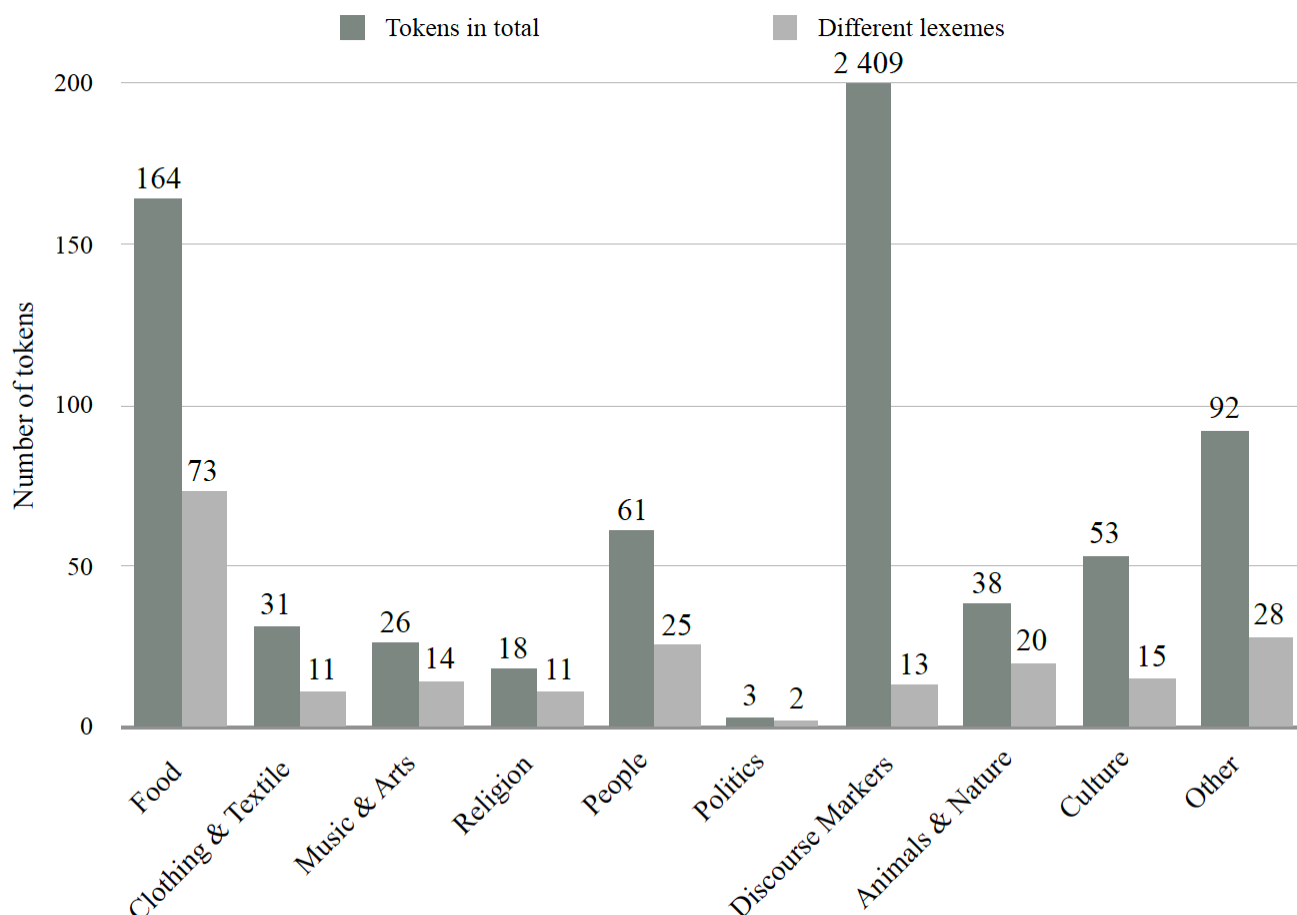
In Table 5 below, the results in ICE-SIN are organized by the subsection of the corpus the tokens occur in. The lighter gray columns represent the spoken sections of the corpus while the darker gray columns reflect the written subsections in ICE-SIN. As can be seen, the majority of the indigenous loanword in ICE-SIN occurred in the spoken sections of the corpus where 'Private dialogue' undisputedly contains the most tokens. The number of tokens in this category is largely due to the high percentage of discourse markers, but leaving them out of the equation would not change

the fact that most of the loanwords still occur within direct conversations and telephone calls. Most of the indigenous words in the written section occurred in 'Academic writing' where the theme of text often directly corresponded with the used loanwords. The second biggest subsection in the written section of the corpus is 'Creative writing' since borrowings from indigenous languages were sometimes used in fictional novels. The number of indigenous words in the written section is otherwise rather low and in 'Persuasive writing' subsection loanwords were not used at all.

Table 5. *ICE-Singapore results organized by corpus section the tokens occurred in*
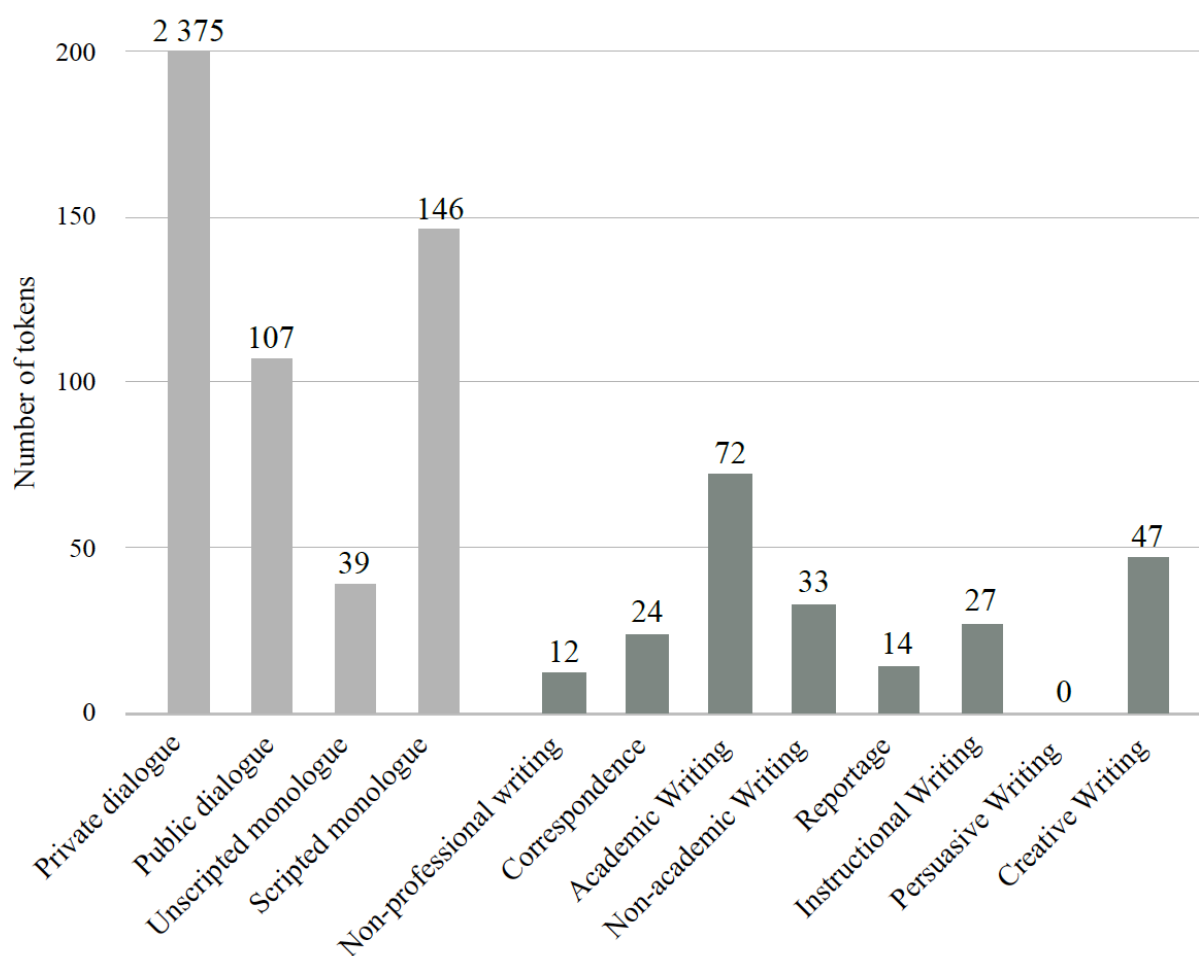
6.1.1. Food

The majority of the indigenous food tokens in ICE-SIN were words for different types of traditional dishes (especially noodle and rice dishes) peculiar to the region. The category also includes a number of indigenous words for fruits, vegetables and spices used in the Singaporean cuisine. The origin of

the loanwords spread quite evenly between the languages traditionally spoken in Singapore with most of the tokens coming from Chinese, Malaysian and Indian languages. The high rate of indigenous food word occurrences in not surprising, since many of tokens do not have an equivalent in Standard English and therefore an indigenous word must be used in order to convey the intented sense and meaning of the subject in question. Some of the indigenous words in this category have also been borrowed into Standard English and can be found in dictionaries such as *the Oxford English Dictionary* (OED). A case in point is the Malay-Indonesian word *durian*, which is a fruit from Southeast Asia, and is known by its indigenous name only in the English language as well. Below are three examples of indigenous words occurring in Singapore English sentences. The meanings of the words are included in brackets after the bolded loanwords.

(1) After much sweating and swearing he managed to open a **durian** (a fruit)
<ICE-SIN:S2B-029#30:1:A>
(2) Ya lah **bee hoon** (a noodle dish) or fried rice lor   <ICE-SIN:S1A-007#255:1:B>
(3) Bring your chilli **padi** (rice) lah   <ICE-SIN:S1A-040#305:1:B>

A large number of the food tokens occur in the spoken section of the corpus but they are also present in written registers. Especially in 'Instructional writing' subsection, which includes multiple recipes containing many indigenous food words also making it the written subsection with the most tokens in 'Food' category.

6.1.2. Clothing and textile

Clothing and textile words, especially in the spoken section of the ICE-SIN corpus, are not very common. Almost ¾ of the tokens in the corpus were found in the written section and many times in contexts where the indigenous words were used in writings describing traditional clothing and textiles, such as in example (4). Clothing and textile words which occurred in the texts more "naturally" in different contexts were few. Most of the words in the category are loanwords for different types of traditional men and women's clothing, and surprisingly many of them can also be found in OED.

(4) For example, the width of the **kepala** (part of a sarung) of a **sarung** (a traditional garment) is measured to occupy either front or the back of the wearer, while the **badan**'s (part of a sarung) length is calculated so that after folding, it forms the counterpart of the **kepala** on the wearer's body.   <ICE-SIN:W2A-009#81:1>

(5) Yes can get can get **batik** (hand dyed fabric) quite cheap <ICE-SIN:S1A-057#95:1:A>

6.1.3. Music and arts

The category of 'Music and arts' does not contain a lot of tokens. The few indigenous words that were used are mostly terms for different forms of art and traditional instruments, as in examples (6) and (7). Most of the occurrences of indigenous loanwords were found in 'Private dialogue' and 'Academic writings'. When the words were used in 'Private dialogue' they occurred quite naturally within casual chit-chat but in 'Academic writings' they were often used in contexts where traditional art and music culture of the region was being described in greater detail.

(6) Unlike performers of the traditional **wayang kulit** (form of puppet theater) or **makyong** (a form of Malay dance-drama) who usually had other sources of income such as farming, for the **bangsawan** (Malay opera) performer, particularly the stars viewed, acting was a full-time occupation.   <ICE-SIN:W2A-008#8:1>

(7) B: Maybe I take **sitar** or **veena** (string instruments)   <ICE-SIN:S1A-085#78:1:B>
    A: Take **sitar**   <ICE-SIN:S1A-085#79:1:A>

6.1.4. Religion

The number of indigenous religious words in ICE-SIN is rather low, other types of cultural loanwords are used more frequently in the corpus. The tokens found are miscellaneous words relating to different religions present in Singapore. According to the Department of Statistics' survey (2015), there are five major religions in Singapore, the biggest of them Buddhism, and over 80% of the population report having a religion. It is therefore somewhat surprising that religious loanwords do not occur in the spoken section of the corpus almost at all, most of the tokens in the 'Religion' category come from the written subsections. However, it needs to be noted that names of religions and derivations of such words, such as 'Hinduism' and 'Hindu', were not picked up from the corpus.

(8) Sorry I didn't write to your last destination, the **ashram** (a spiritual hermitage or a monastery) in Tamil Nadu.   <ICE-SIN:W1B-007#43:2>

(9) You you go there and then beg and beg and and and **kowtow** (an act of showing respect or worshipping) uh   <ICE-SIN:S1A-094#291:1:B>

6.1.5. People

'People' is one of the biggest categories in the Singapore component of ICE with 61 tokens. The category consists of various words for different ethnicities and words describing family ties and relations. Many of the indigenous loanwords regarding ethnicity involve a "us vs. them" mental separation of the "local" Asian ethnicities and "foreign" European people. Examples of words for local and foreign people can be seen below in sentences (10) and (11). Different variants of *Peranakan* and *ang mo* were the most used types of tokens in the 'People' category. It was interesting to notice that the variants of *ang mo* found in ICE-SIN often carried a negative connotation which could allude to a conclusion that although Singapore is westernizing at a rapid pace, the original ethnic groups are still holding on to the colonial past and view European people as 'others'. A great example of this is found in a conversation in example (12). Also meaningful is that the literal translations of *ang mo* and *kwai* lo are 'redhead' and 'foreign devil'. In addition to ethnicity words, other tokens, such as *kanna* in example (12), occur in the texts only a few times per type.

(10) I got it from an old **Peranakan** (a local born of Chinese/Malay descent) family at a steal.   <ICE-SIN:W2F-006#30:1>

(11) I don't want to tell that stupid **ang mo** (a white/European person) you know
                                       <ICE-SIN:S1A-096#X365:2:E>

(12) A: Got some **ang mos**
B: Got a lot of **kwai lo** (a white/European person) uh
A: No there are more there are more locals than there are **ang mos**
                                  <ICE-SIN:S1A-025#345-7>

(13) That's Steven Tan's **kanna** (a kid)   <ICE-SIN:S1A-058#130:1:B>

A little over 60% of the tokens in 'People' category were used in the spoken section of ICE-SIN, but taking into consideration that the spoken section consists of 300 texts whereas the written

section is comprised of 200 texts, it is safe to say that the tokens were spread quite evenly throughout the corpus.

6.1.6. Politics

Loanwords from indigenous languages in ICE-SIN consisted of only a couple of tokens. Since English functions as the language of all governmental activities in Singapore, it is not surprising that indigenous politics vocabulary barely exists in Singapore English. Based on Schneider's theory, it is vocabulary related to flora, fauna and culture that is most likely to be borrowed into language contact varieties. However, in the 'westernized' political system of Singapore, there is no need for indigenous vocabulary since the terms and words needed already exist in English language and can be easily modified if needed to better represent politics in Singapore. (14) is an example of one of the two types of tokens found. The token is a Malay term, which in this context is used in 'grass roots' level politics among the Malay community, so no general assumptions of its usage can be made.

> (14) One example are the **jawatan kuasa** (a committee) blok or block committees which emerge among the Malays who reside in HDB estates.   <ICE-SIN:W2A-014#56:1>

6.1.7. Discourse markers

Discourse markers are said to be one of the most distinctive features of Singapore English, and the findings in the Singapore component of the ICE corpora attest to that. With 2409 tokens found in the corpus, discourse markers are the most used vocabulary items loaned from indigenous languages into Singapore English. With regard to the number of tokens, it must be stated again that tokens in 'Discourse markers' category were treated differently compared to other categories in so that there was no limitation to the number of same type of tokens occurring within a text. This restriction was not employed with the tokens in the category because the use of discourse markers is not repetitive but impulsive even if the same discourse marker is used multiple times within a text or a conversation.

Table 6. Top 5 discourse markers in ICE-SIN based on frequency of occurrences

Discourse marker	Tokens
lah	1756
lor	145
aiyo	137
wah	75
hor	68

The top 5 tokens based on the frequency of their occurrences are listed in Table 6. *Lah,* which is used to "convey the mood and attitude of the speaker" (Wee 2008b, 604-6) is by far the most used discourse marker in the corpus. *Lor* and *hor*, which are traditionally said to be the most used discourse particles after *lah* place second and fifth in ICE-SIN. *Lor* marks something as obvious information (ibid.) and *hor* presents a proposition and tries to gather support for it (Lim 2004, 123-4). *Wah* and variants of *aiyo*, which are Indian expressions, are not usually listed as characteristic discourse markers of Singapore English but in ICE-SIN they occurred rather frequently in a natural way so they were included in the study. According to OED, *wah* can be used "to express admiration, encouragement, delight, [or] surprise" while *aiyo* conveys "distress, regret, [and] grief". Examples (15-19) include contexts in which the top five discourse particles have been used in the corpus. Discourse particles *yah, ya, ha* and *aye* were excluded from the findings since they are typically not considered discourse particles indigenous to Singapore English or Southeast Asian Englishes. Other types of tokens in the category not yet mentioned include *leh, loh, mah, meh, suh, dah, hah* and *orh*.

(15) C: Did he invite all the branch people
    D: No I'm not so sure who he invited
    C: I think he should **lah**   \<ICE-SIN:S1A-026#129-31>
(16) Uh that's provided you don't buy things **lor**   \<ICE-SIN:S1A-011#129:1:B>
(17) **Aiyah** what a waste   \<ICE-SIN:S1A-016#16:1:A>
(18) **Wah** he got hit first you know then he jumped out you know
                                        \<ICE-SIN:S1A-082#129:1:B>
(19) I mean sword fighting movies are always like that **hor** \<ICE-SIN:S1A-041#280:1:B>

Since discourse markers are most often used in spoken situations, it is no wonder that most of the tokens in the category occurred in the spoken section and more specifically in the 'Private dialogue' subsection which includes direct conversations and telephone calls. Some instances of discourse particles were also found in the written section with most of them in 'creative writing' where they were used in fictional characters' speech.

6.1.8. Animals and nature

According to Schneider (2007), vocabulary items relating to animals and nature are the first items of indigenous languages borrowed into English in World English language contact situations. Therefore, the number of results in the category is surprisingly low, but taking into consideration the number of indigenous loanwords in the whole corpus (except for discourse markers), 'Animals and nature' does not differ from the other categories. Most of the loanwords in the category are words for different types of plants, bushes and trees, but loanwords for birds, fish and other kinds of animals also occur. It has to be noted that the corpus included a lot of place names which involved indigenous nature vocabulary such as '*Bukit* (hill) Gombak', a neighborhood in Singapore, but were not picked up because names of places are not featured in the corpus study of this thesis. Sentences (20) and (21) illustrate the occurrence of loanwords belonging to the category in the corpus.

> (20) Export of the popular song bird, **Mata-Puteh** (a Southeast Asian bird), is banned in Malaysia.   <ICE-SIN:W2C-007#77:3>
> (21) Many of the new towns and factories were built upon the land once covered with **lallang** (a species of grass in Southeast Asia) or by secondary forests.
>   <ICE-SIN:W2A-025#31:1>

The tokens in 'Animals and nature' category have mostly been picked up from the written section of ICE-SIN and only a few tokens occurred in the spoken registers. The subsections where most of the tokens occur in are 'Academic' and 'Non-academic writings', therefore the loanwords in the corpus appear in scientific and semi-scientific contexts.

6.1.9. Culture

The 'Culture' category was not one of the original categories used by Balasubramanian (2009) but was added to the thesis since many of the tokens found had cultural value but were not semantically similar to the words in the pre-existing categories. The existence of indigenous cultural words is an essential part of World English varieties (Schneider 2007) and the addition of 'Culture' category ensures the suitable categorization of the tokens found in the corpora.

The added category is one of the biggest categories of Singapore English results with 52 tokens. Out of the 52 tokens, a few types were used more than others, the word *ang pow* (22) was used 9 times and variants of *kampong* (23) occurred in the texts 19 separate times. Other types of tokens were used only a few times per token and included words for sports and games (24), specific types of gardens and parks, and different terms for currency relevant for the region.

(22) Once people give you oranges you must give **ang pows** (a traditional monetary gift in a red envelope) right   \<ICE-SIN:S2B-035#140:4:D\>

(23) Where the petrol kiosk now stand was a **kampong** (a traditional village or an urban slum) area and behind it a private garden   \<ICE-SIN:S2B-027#64:1:A\>

(24) Activities like weight-lifting, **sepak takraw** (a Southeast Asian game of kick-volleyball) and table tennis which are still popular today \<ICE-SIN:S2B-015#113:2:D\>

Indigenous culture vocabulary was used in spoken language slightly more than in written language. Most of the tokens found in the spoken section of the corpus occurred in 'Scripted monologue' which includes broadcast news and talks as well as other types of scripted speeches, in fact only discourse markers and food terms were used more in 'Scripted monologue'. 'Scripted monologue' subsection is probably the most formal subsection of the spoken part of the ICE corpora and the occurrence of tokens in there should be a clear indication that the indigenous words are very much in use in Singapore English.

6.1.10. Other

Tokens in the 'Other' category are words which could not be placed in any of the other semantic categories. The most reoccurring words in the category are *kiasu* (see example 25) and *shiok* (26),

which are characteristic Singapore English words also found in most of English language dictionaries. In addition to *kiasu* and *shiok,* the findings include a number of verbs and adjectives with English equivalents as well as a few greetings and unplaceable nouns.

(25) People can say I'm **kiasu** (a grasping, selfish attitude), but as a parent I only want the best for my children   <ICE-SIN:W2C-019#90:2>
(26) A: It's quite **shiok** right (a feeling of pleasure and happiness)
      B: It's quite good   <ICE-SIN:S1A-062#84-5>
(27) So I must **kau tim** (finish) my laundry first   <ICE-SIN:S1A-091#104:1:A>

Almost 90% of the tokens in 'Other' category occurred in the spoken section of the corpus. The spoken texts are in general more informal than the written texts in the ICE corpora, so the existence of indigenous words in spoken registers is understandable because there is more variation in the language used in informal settings.

6.2. Findings in Indian English

The number of indigenous loanword tokens found in the Indian English component of ICE proved to be much higher than in the Singapore English component. The overall number of indigenous loanwords in ICE-IND is 3,918 while in ICE-SIN the corresponding figure is 2,895. Furthermore, leaving out the category of 'Discourse markers' and only taking into consideration the tokens in other categories makes the difference even more remarkable since discourse markers in ICE-SIN make up 83% of the overall number of tokens meanwhile only 28% of the tokens in ICE-IND are discourse markers.

Table 7. *ICE-India results organized by the semantic category of the tokens*
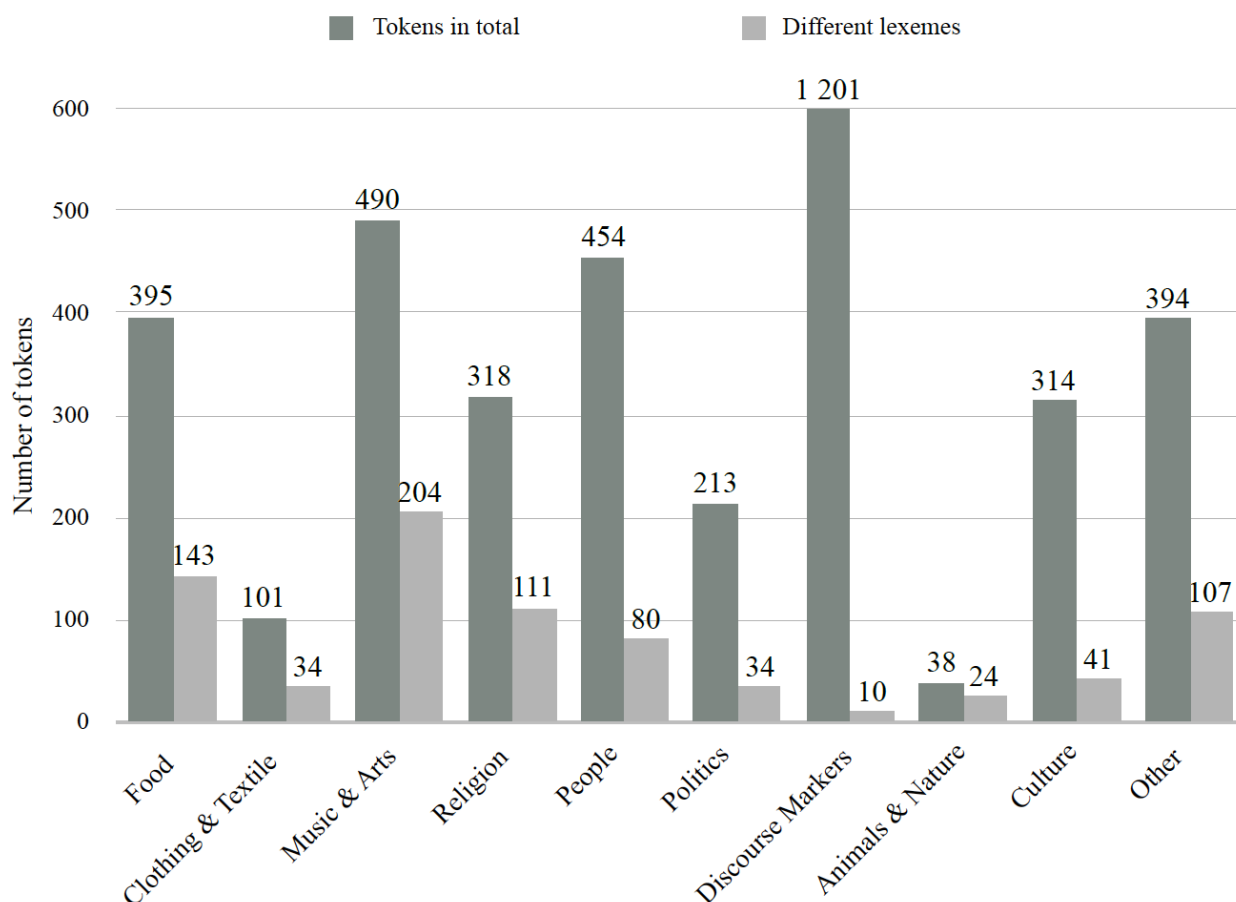
As evidenced by ICE-IND results presented in Table 7, 'Discourse markers' is the biggest semantic category of indigenous loanwords in Indian English but loanwords in other categories are also very well represented. 'Food', 'Music and arts', 'Religion', 'People' and 'culture' categories are the largest specified categories after 'Discourse markers'. In addition to the semantically specified categories, the category of 'Other' contains 396 miscellaneous indigenous words which make it the a rather big category on its own regarding the number of tokens. An additional 61 tokens were not included in any of the categories since their meaning was left unclear. As with findings in ICE-SIN, the number of different lexemes among the tokens is much lower compared to the number of tokens in total. Because the number of tokens in ICE-IND is much higher than in ICE-SIN, it is natural that in additon to havin more tokens, different lexemes (or loanwords) were also employed more in ICE-

IND. Each of the categories and occurrences of tokens within them will be discusses more thoroughly later in the following subsections.

Table 8. ICE-India results organized by corpus section the tokens occurred in
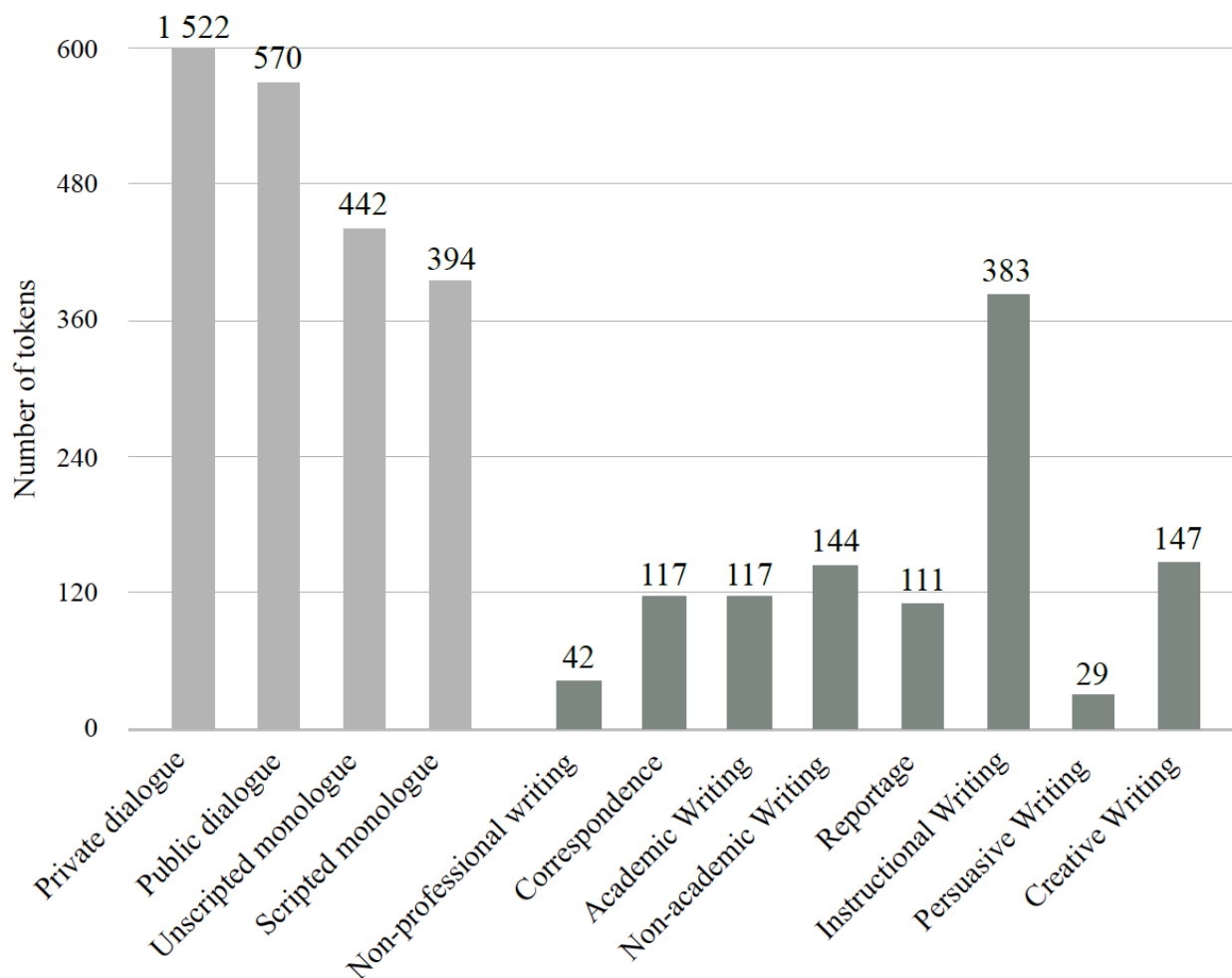
In Table 8 above, the results are categorized under subsections of the corpus based on their place of occurrence. 73% of all the indigenous words occur in the spoken section of the corpus. Taking into consideration that the spoken section texts comprise 60% of the whole corpus, the spoken section still contains more tokens, thus, loanwords tend to more generally be used in spoken Indian English. Again, this is not surprising since spoken contexts are often more informal in nature and therefore allow more variation in the language used, also in vocabulary. Most of the tokens in the spoken section were found in 'Private dialogue' and the tokens in other spoken subsections spread out quite evenly. The distribution of the results in the written subsections was more varied. 35% of

the loanwords occurring in written contexts were used in 'Instructional writing' texts, which include various recipes and detailed descriptions of traditional arts in India. On the other hand, 'Non-professional' writing and 'Persuasive writing' included far less tokens than other subsections of the corpus.

6.2.1. Food

Food words are one of the most frequently used indigenous loanwords in ICE-IND, and out of the 395 words found in the corpus, no type of token occurred significantly more than others although some tokens were used more times than others. The vocabulary in the category includes a varied mix of words for different types of traditional bread, dishes, snacks, restaurants, but also Indian terms for words such as nuts, beans, vegetables and wheats which have English equivalents and whose uses are not triggered by non-existence of corresponding English language vocabulary. A lot of the words, specifically dishes and snacks, in the category have found their way into OED. Examples (28-31) below present sentences with some of the most common food vocabulary tokens found in ICE-IND.

> (28) But this is best eaten with the steamed rice or **chapati** (Indian flatbread)
> <ICE-IND:S2A-051#74:1:B>
> (29) My husband must be just keeping the cooker and they must be eating just **daal** (a term used for beans, peas and lentils) and rice   <ICE-IND:S1A-030#41:1:B>
> (30) It is the winter season in Delhi and the **dhabas** (a roadside restaurant in India) are overflowing.   <ICE-IND:W2B-017#73:2>
> (31) 250 g. paneer; 150 g. curd; 50 g. cashewnuts; 2 tomatoes; 1/2 cup fresh cream; 3 tbsp. **ghee** (clarified butter) for deep frying.   <ICE-IND:W2D-016#199:1>

Clear majority of the Indian loanwords occurred in the spoken section of the corpus, especially in 'Private dialogue' and 'Unscripted monologue' subsections. A great number of tokens was also used in 'Instructional writing' in the written section, but as mentioned earlier, the subsection includes several recipes in which the occurrence of loaned food vocabulary is not particularly astonishing.

6.2.2. Clothing and textile

With 101 tokens, 'Clothing and textile' is one of the smallest categories of ICE-IND results, only 'Animals and nature' vocabulary was used more infrequently. However, being one of the smallest categories does not mean that clothing and textile words were barely used, it just means that in comparison to words in other categories they were used less. Note that, were a category to get 101 tokens in the Singapore component, it would be proportionally bigger compared to most Singapore English categories.

The most common types of tokens in 'Clothing and textile' category are variations of the word *saree* (32) which occur in the corpus 35 times. A large part of the tokens in the category are words for different types of garments, especially pieces of female clothing, but words men's clothing, types of accessories and terms for specific type of textiles, fabrics and cloths are also found. Instances of different types of words in the category can be seen in examples (32-34). While many of the food words have entries in OED and other dictionaries, only a handful of the loanwords in 'Clothing and textile' are discoverable in any Standard English dictionary.

> (32) Recently, we had seen her wearing a **saree** (a female garment) for granny's birthday
> <ICE-IND:S1A-040#194:1:B>
> (33) But the **Baandha** (a type of textile cloth) is an art that does not need extra yarn for its design. <ICE-IND:W2D-020#146:1>
> (34) A white-haired gentleman in a spotless ivory **khadi kurta** (a loose hand-woven shirt) and **churidar** (tight-fitting Indian trousers) outfit was regarding No. 24 with studied casualness, but his eyes gave him away. <ICE-IND:W2F-008#49:1>

Notable regarding the 'Clothing and textile' results is that 60% of the tokens found in the corpus occur in the written section which is contrary to other categories of results. Most of the tokens in the written section are again found in 'Instructional writing', which includes texts with thorough descriptions of Indian textile industry and therefore indigenous clothing and textile words are common in that specific subsection of the corpus.

6.2.3. Music and arts

'Music and arts' is the biggest specified semantic category after 'Discourse markers' with 490 tokens. Out of the occurring loanwords in the category, terms relating to music are used most frequently. Indigenous music vocabulary includes terminology for Indian music notes, scales, forms and melodic modes, as well as musical concepts, names of genres, singing styles, parts of composition and traditional instruments. The most used lexeme in the category is *raga* (35) with 47 occurrences while other lexemes are not used to the same extent. Other art styles, such as theater, dance, poetry and literature are also represented in the results. Theater vocabulary includes terms for different types of folk theater, performance arts and expressions, and names of different steps, hand gestures and traditional dances mostly comprise the indigenous dance vocabulary of the category. Vocabulary relating to other types of art forms is extremely scant. Looking at the results, it is clear that indigenous vocabulary relating to music and arts is used so extensively because Indian art forms differ from Western forms of art and therefore, no equivalent terminology exists. All of this leads to the use of Indian loanwords in the field. In addition, Indian forms of music, theater and dance have a long history and are deeply rooted in the society, and terms relating to them tend to be used quite frequently when discussing traditional arts.

(35) The combinations he presented during **jor** (a formal section of a composition) were many and varied, clear and imaginative, and with these he developed the **raga**'s (a melodic mode in Indian classical music) form which has reposeful grandeur, best suited for **alaap** (an opening section in traditional music) and **jor**. <ICE-IND:W2D-014#71:4>

(36) The ceremony was followed by **Kathak** (an Indian form of dance mixed with miming) dance recital by Pandit Birju Maharaj   <ICE-IND:S2B-001#70:1:A>

(37) Ponani and Sengadi with uh instrumental music cymbals, **chengala** (a Keralan instrument) or **eluttalam** (a Keralan instrument), are being used by the singers <ICE-IND:S2B-021#52:1:A>

'Music and arts' is one of the two categories, along 'Clothing and textile' in which the written subsection tokens exceed the occurrences found in the spoken section of the corpus. 66% of the overall results in the category occur in the written section, majority of them in 'Instructional

writing' but 'Academic writing' and 'Non-academic writing' also have a great number of occurrences. Most of the tokens found in the spoken section were found in 'Scripted monologue'. Consequently, it would seem that indigenous vocabulary relating to 'Music and arts' is generally used more in formal contexts.

6.2.4. Religion

A respectable number of tokens were placed in the category of 'Religion' when carrying out the corpus study. Despite having 318 tokens, the category is still placed in the middle ground when comparing the number of tokens with other categories. Religious indigenous vocabulary found in the corpus contains words relating to religious festivals and ceremonies, rituals, customs and practices, places of worship, and terms and concepts in relation to Hinduism, Buddhism, Sikhism and Islamism. In addition to indigenous words from "conventional" religions, yoga vocabulary was listed under the 'Religion' category following Balasubramanian's categorization. Among the tokens found were a few types that were used more than others, particularly variations of *Diwali* with 35 tokens and *pooja* which was used 21 times, but none of the rest of the type of words in the category rose clearly above others in the number of times they were used. Examples of the most used words and a couple of other typical indigenous loanwords in the category are presented in sentences (38-41).

(38) We are planning to go in **Diwali** (a Hindu festival/holiday) vacation to Mahabaleshwar
<ICE-IND:S1A-086#116:1:B>
(39) And then you start with **pooja** (a Hindu prayer ritual) before sleeping, also when you get up in the morning, that time also <ICE-IND:S1A-043#194:1:A>
(40) So this **asana** (a place and posture in yoga) trains the brain to get acclimatise to keep the head to the ground, then easily the confidence come and they can practise **shirsana** (a headstand asana) <ICE-IND:S2A-055#128:2>
(41) People are visiting different **Gurudwaras** (a place of worship for Sikhs) in the capital and other places since early this morning <ICE-IND:S2B-016#122:2:A>

A remarkable 88% of the tokens in 'Religion' category were used in different spoken contexts. Since indigenous religious vocabulary was used so frequently in informal spoken texts in the corpus, it leads to the assumption that religion is an inherent part of the Indian society and

everyday lives of the people. In addition, loanwords in religious contexts are used because oftentimes no English word can represent the complex concepts at the core of Indian religions. Regardless of how used they are in Indian English, it is still only the most basic indigenous words in the 'Religion' category which can be found in OED and other dictionaries.

6.2.5. People

Indigenous words used for people are among the most commonly used Indian loanwords found in ICE-IND with 454 tokens in total. The category includes a wide range of indigenous words used as honorifics and forms of address (42), terms for leaders and teachers (43), relatives (44), and ethnic groups of people (46), as well as various occupational words for traditional Indian jobs (45-46). Loanwords for honorifics, leaders and teachers are used more often than other types of words for people but each of the above mentioned semantic groups is still well represented in the results. Some tokens, such as *shri* (143 tokens), *guru* (38 tokens) and *pandit* (35 tokens) occurred in the texts more frequently than other words in the category, but the results in 'People' are still some of the most varied and substantial findings in the corpora. It must also be noted, regarding the results in the category, that names of peoples deriving from the geographical area they originate from are excluded from the study. Also, terminology referring to the Indian caste system is listed under 'Politics' category.

> (42) Children's day was celebrated on fourteenth November, sweets were distributed to students in memory of late **Shri** (a honorific) Baburao Tol <ICE-IND:S1B-079#97:1:F>
> (43) So nobody should take any field without becoming a thorough **pandit** (a scholar or a teacher) in that particular field […]   <ICE-IND:S2A-024#50:1:A>
> (44) How much are you charging **didi** (an older sister/female cousin) for the yoga
> <ICE-IND:S1A-043#198:1:B>
> (45) They would send their other **gundas** (a hired thug) there in that locality to settle everything   <ICE-IND:S2A-061#61:1:A>
> (46) To the servants of the locality, especially the **chokras** (a boy servant), she was neither **desi** (a person from the Indian subcontinent) nor **angrez** (a British/white person), man nor woman.   <ICE-IND:W2F-002#16:1>

The 'People' tokens are spread evenly throughout different sections and subsections of ICE-IND, which cannot be said for the results in any other category of findings from the two corpora. As a result, it can be assumed that Indian loanwords in relation to people or groups of people are used as frequently in informal and formal situations as well as spoken and written contexts.

6.2.6. Politics

Indian English 'Politics' category has less tokens than most of the other categories of results regarding the variety, but with 212 tokens it is still much bigger in comparison to the same category of Singapore English results which was virtually non-existent. The indigenous loanwords in the category could be further organized into five sub-categories based on their semantic meanings – caste system (see example 47), titles for royals and rulers (48), strikes and protests (49), government systems (50), and other political words (51).

(47) Countless number of **Harijan** (people at the bottom of or outside the caste system) women have been victims of atrocities by caste Hindus.   <ICE-IND:W2D-003#115:1>
(48) It's also a great privilege that later in the evening respected **Rashtrapatiji** (a president) Dr Shankar Dayal Sharma will be here to present award   <ICE-IND:S2A-017#53:1:C>
(49) At the same time they participated in a **morcha** (a march/rally) and **dharna** (a non-violent sit-in protest) organized by the [unclear]   <ICE-IND:S1B-077#49:1:A>
(50) We would rather have more and more of **panchayat** (an old system of local government), district and **zilla parishad** (a disctrict council in the panchayat system)-sponsored schemes than Centrally-sponsored schemes   <ICE-IND:W2C-011#12:1>
(51) Mr. Singh gave a patient hearing to the **Samiti** (a committee/association/society) delegation […]   <ICE-IND:W2C-009#75:12>

Caste system vocabulary included terms for different castes and caste members. What is notable is that the results include more different types of words for the lower casts and people outside the caste system than for people in higher castes. Indigenous title words, at the same time, include various types of titles mostly used in association with different types of male rulers. Non-violent strike and protests have been an essential part of Indian culture since Gandhi and protests against British colonizers. Therefore, it is interesting to see that even in such a small corpus, there are three

different words used for a 'non-violent protest' as well as a few other words connected to protesting. Indigenous vocabulary used for government systems were the most plentiful tokens placed in the category, largely due to *Lok Sabha* and *Raya Sabha*, which are Indian terms used for the lower and upper houses of the parliament and were mentioned dozens of times in the texts. Other political words in the category are words with English (and Western) equivalents, such as Indian terms for assembly and committee.

The majority of the loanwords used in the category, however, do not have a corresponding word in English since many of the concepts and meanings do not match the 'Western' political system. The political and government systems in India use multiple languages depending on the state and only few of them use English as an official language, as stated earlier in section 3.1., so the use of Indian words in political contexts is only natural. Most of the words in the category have their separate entries in OED but only a few of them are listed in other English language dictionaries.

6.2.7. Discourse markers

Whereas other categories of Indian English results have significantly more tokens than Singapore English categories, 'Discourse markers' is the only category that has less tokens in Indian English than in Singapore English. Based on the corpus study, indigenous discourse markers are used over twice as much in Singapore English compared to Indian English. Out of the 1201 tokens occurring in ICE-IND, four types were used recurringly in the corpus. The most used indigenous discourse markers are presented below in Table 9.

Table 9. *Top 4 discourse markers in ICE-IND based on frequency of occurrences*

Discourse marker	Tokens
haan	449
ahn	267
accha	247
yaar	130

Use of discourse particles has not been studied as much as other kinds of vocabulary items in Indian English, therefore, much cannot be said about them in connection to previous studies. It should be noted that all kinds of function words, including loanwords for *yes* and *no,* are placed in 'Discourse particles' category following Balasubramanian's (2009) categorization. As a result of this, the most used type of token in the category is *haan*, which means yes. The second most recurring particle in ICE-IND is *ahn*, whose meaning could not be deciphered but which could be a variation of the word *hahn* listed as a discourse particle in Balasubramanian's (2009) study. The third most used item is *accha*, which according to Cambridge Dictionary is used for showing agreement, understanding and ashtonishment. The last frequently reoccurring word listed is *yaar*, probably the most common Indian English discourse particle, which as per OED was originally used as a term of address meaning friend but which has been listed as a discourse particle by many scholars, such as Balasubramanian (2009). All of the discourse markers, except for a single occurrence of *yaar* in 'Creative writing' were used in spoken contexts which is typical for discourse markers as they are part of spoken communication.

6.2.8. Animals and nature

The number of tokens in 'Animals and nature' category was the lowest in all of the categories of Indian English results, loanwords in the category only amounted to 38 uses. Words related to 'Animals and nature' were used in ICE-SIN exactly the same number of times, which is proportionally higher taking into consideration the overall number of tokens in the categories. Most of the tokens were words for different kinds of trees common in the Indian peninsula and Southeast Asia. Bird types were the most used items of animal vocabulary which also includes a few tokens for other animals. Other words in the category include a couple of words for Indian grains and a few for types of terrain. Some of the words occur in OED but even so, they are mostly known by their English language names. Below are instances of loanwords related to 'Animals and nature' found in ICE-IND.

(52) Here if you view by the side or the back of the temple, there is…we have a **Ashoka** (a rain-forest tree)… **peepal** (a sacred fig tree) tree…   <ICE-IND:S2A-043#111:1:A>

(53) And at the same time son and father-in-law are sleeping in deep slumber as if two **azgars** (an Indian rock python) are sleeping   <ICE-IND:S1A-006#118:1:A>

(54) Because the **Hilsa** (a herring fish) itself is a very I mean easily cooked fish
     <ICE-IND:S1A-007#165:1:C>

6.2.9. Culture

Over two thirds of the 314 tokens in 'Culture' category are words related to money and transaction. As with the Singapore English results, names of currencies, money and coins were categorized as 'Culture' words. Therefore, the most used term in the category is *rupee*, which was used in the corpus almost 200 hundred times. Other vocabulary concerning money and trading include indigenous loanwords for different types of coins (ancient and current), marketplaces, as well as trading and loan systems. Also regarded as cultural vocabulary are words related to different ceremonies and customs (which are more cultural than religious), education system, traditional buildings, sports and games, philosophy, Indian medicine, and a few other words.

(55) But if black money which is being created in this country is being siphoned out and uh the **hawala** (a popular and informal money transfer system) market they are having their own fare   <ICE-IND:S1B-060#3:1:B>

(56) As if **swayamvaras** (the custom of choosing the husband for your daughter) nowadays or the people are simply making their daughters obey the orders
     <ICE-IND:S1A-074#108:1:A>

(57) The record book will note that the only gold India won was in **kabaddi** (a contact team sport), which South Asians have played for generations and others are barely familiar with.   <ICE-IND:W2E-003#47:1>

(58) We would like to discuss indigenous health services such as **Ayurveda** (a traditional Indian system of medicine), Acupuncture, **Unani** system (a traditional Persian-Arabic system of medicine)   <ICE-IND:S1B-035#156:1:B>

Large majority of the tokens in 'Culture' category were used in the spoken section of the corpus, especially in 'Private' and 'Public dialogue' as well as 'Unscripted monologue'. Based on this, it would seem that indigenous cultural loanwords are more frequently used in informal spoken situation than more formal and written contexts. The scant use of indigenous cultural loanwords in

formal contexts can be due to the fact that many of the inherently Indian words do not have entries in OED or other English dictionaries.

6.2.10. Other

In addition to the tokens in the previous specified categories, 394 tokens with mixed meanings and functions were picked up and grouped under the label 'Other'. A few tokens, particularly words *crore* and *lakh*, were used considerably more than other. *Crore*, which means "ten million", was used in 90 different occasions while *lakh*, a term that stands for "one hundred thousand", occurred in the corpus 111 times. Other indigenous loanwords in the corpus were used only a handful of times each, and include numerals, verbs, adjectives and adverbs, buildings, spaces, and tools, plus other miscellaneous words. Examples of 'Other' indigenous words occurring in ICE-IND can be seen in examples (59-62) picked up from the corpus.

> (59) **Dekhye** (to look/see), we're confusing two issues   <ICE-IND:S1B-041#71:1:B>
> (60) So what we'll do is make a gravy of chicken, ah, thickish, ah, or as we usually do the Kolhapuri style, uhm, and uh we'll make that mutton **ekdam** (completely) dry, uh, like a roast
> (61) Hereafter, he along with the accused and uh the knife and his staff came to Shastri Nagar Police **chowki** (a police station) and reported the matter to Vakola Police Station
>   <ICE-IND:S2A-067#20:1:A>
> (62) Whether the police used to visit the the liquor den during this period for drinking or for uh **hafta** (protection money collected by criminals or corrupt police) or something?
>   <ICE-IND:S1B-064#22:1:B>

Many of the indigenous words in 'Other' category have equivalents in Standard English, hence their uses are not caused by actual need for the word. 70% of the words occured in the spoken section of the corpus so miscellaneous indigenous words are used only slightly more frequently in spoken than in written texts in ICE-IND.

7. Discussion

In this chapter, the results in the Singapore and Indian components of ICE will be compared with each other to see what are the biggest differences regarding the number of tokens, the category the tokens occur in, and types of tokens used. The results in both corpora will then also be compared to the vocabulary study conducted on Indian English by Balasubramanian (2009) to see whether the results of the corpus study of the thesis are different or similar to her study and results which were already presented in section 4.3. of this thesis. In addition, the results will also be analyzed by looking at the history and different colonization experiences of the countries, as well as discussing whether the language policies (or lack thereof) the countries have employed have had an effect on the indigenous vocabulary used in variations of English spoken in the regions. Lastly, the results will be viewed taking into account Schneider's (2007) Model of Postcolonial Englishes and other previously mentioned approaches to indigenous vocabulary in World Englishes.

As briefly mentioned earlier and presented by the data on results, the overall number of indigenous loanwords tokens in ICE-SIN and ICE-IND somewhat differs since the number of used loanwords in ICE-SIN is 2,895 while in ICE-IND it is 3,918 so the Indian component of ICE has 35% more occurrences of indigenous loanwords than the Singapore component has. The difference becomes even more significant when discarding 'Discourse markers' and focusing on content words in other semantic categories. The number of tokens in ICE-SIN excluding tokens in 'Discourse markers is 485 whereas in ICE-IND it is 2,717. Approximately 83% of the indigenous loanword tokens in ICE-SIN are different types of discourse markers and other vocabulary is used rather infrequently, while at the same time, discourse markers in ICE-IND represent only 31% of the overall number of tokens in the corpus and other types of indigenous vocabulary is used substantially more frequently than in ICE-SIN. The overall numbers with and without 'Discourse markers' are presented in Table 10 below.

Table 10. ICE-SIN and ICE-IND overall results, number of
discourse markers and tokens excluding discourse markers
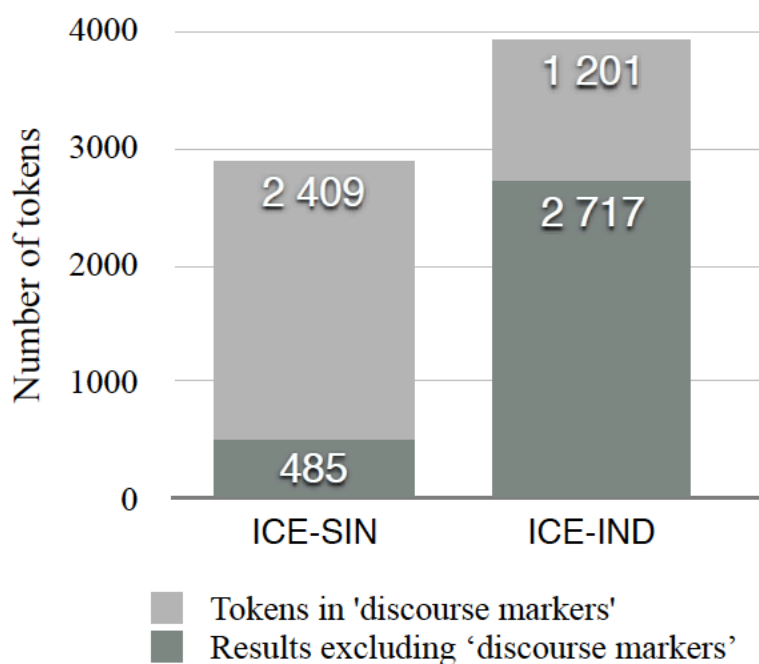
The results not only differed in the number of tokens found in the corpora but also in the distribution of results between categories. No category, except for 'Discourse markers' rose notably above others in the number of tokens found in the corpus in ICE-IND but 'Food' vocabulary in ICE-SIN was used far more frequently than any other types of loanwords. After 'Food', all the rest of the categories in ICE-SIN, not including 'Other', have under 60 tokens in the category with no big differences in numbers, except for 'Politics' which only has 3 tokens listen in the category. While 'Food' vocabulary in ICE-IND also has a high number of tokens, it is not the biggest category of ICE-IND results since based on the tokens found in the corpus since 'Music and arts' and 'People' categories have more listed in them than 'Food' does. Only indigenous vocabulary concerning 'Animals and nature' in ICE-IND has less than 100 tokens whereas all the specified categories except for 'Discourse markers' and 'Food' in ICE-SIN have significantly less than 100 tokens. The top five specified categories in ICE-SIN are 'Discourse markers' (with 2409 tokens), 'Food' (164), 'People' (58), 'Culture' (52) and 'Animals and nature' (38). Meanwhile, the five categories with the most tokens in ICE-IND are 'Discourse markers' (with 1201 tokens), 'Music and arts' (490), 'People'

(454), 'Food' (395) and 'Religion' (318). Based on the results and the found loanwords, it would seem that Indian society in general is much more focused on traditional culture including food, music and arts as well as religion than Singaporean society. This is not surprising since Indian culture has a much longer history than Singaporean culture does and many religions, philosophical ideas and arts in Asia have their roots in India. Singapore as a nation is also more concentrated on the future and 'Westernization' of the country, as they become more and more multicultural by the year which might also influence the indigenous vocabulary and traditional cultural words in Singapore English and count for the lack of loanwords from languages of the region.

Differences between the results in ICE-SIN and ICE-IND also arise when looking at the parts of the corpora the tokens occurred in. While in both of the corpora, most of the tokens were found in the spoken section of the corpus, the overall percentage of tokens in spoken contexts differ, since in ICE-IND a little over 70% of the tokens occur in spoken subsections while in ICE-SIN more than 90% tokens are used in the spoken section. Therefore, written informal Singapore English does not differ much from Standard English on part of vocabulary but written Indian English might be harder to understand for outsiders, since more indigenous loanwords are used even in formal contexts. Indigenous vocabulary in Singapore English, excluding discourse markers, was mostly used when talking about food or clearly describing or pointing out older traditional culture. Spoken Singapore English vocabulary, on the other hand, is clearly marked by the use of discourse markers which have been deemed characteristic of Colloquial Singapore English by scholars on the subject (eg. Lim 2004; Leimgruber 2011; Wee 2008b). Indian English is a completely different matter and indigenous loanwords are far more common in both spoken and written contexts as they are not only used in connection to food and older traditional culture but also to everyday life, customs and culture of today. The only category with a higher frequency of token occurrences in ICE-SIN, excluding 'Discourse markers', is 'Animals and nature'. Otherwise, more indigenous loanwords were used in each category of results in ICE-IND.

When comparing the corpus study results (mostly ICE-IND) of this thesis to the results in Balasubramanian's (2009) similar study conducted on CCIE, a few observations arise. First, the number of indigenous loanwords in CCIE in general seems to be slightly lower than in ICE-IND. However, it has to be noted that Balasubramanian does not include any numbers for tokens in 'Other' category so the ICE-IND and CCIE could be closer in their number of indigenous tokens if numbers for 'Other' tokens were also given. The difference in the number of tokens may result from differences in the ways the corpora are compiled since the written register in CCIE covers approximately 70% of the whole corpus (ibid., 41-2) whereas the written section of ICE corpora is only 40% of the corpus. As a result, CCIE is more formal in nature because written language in general is more formal than spoken language and therefore also contains language that is less influenced by indigenous languages, thus less loanwords from regional languages. If someone were to conduct a study on an indigenous grammatical feature in CCIE and ICE, the number of results in CCIE as compared to ICE would probably be even more significant. The second thing that greatly differs in Balasubramanian's CCIE results and the ICE results of the thesis is the number of tokens listed under 'Discourse markers', discourse markers seem to occur far less frequently in CCIE than in ICE-SIN and even in ICE-IND. In ICE-IND (as well as ICE-SIN), 'Discourse markers' is the biggest category of results but in CCIE it is the second smallest, the only category smaller than it is 'Greetings' which was completely cut out from the categories in this thesis since it provided only a handful or results and was deemed altogether unnecessary. However, the low number of discourse markers in CCIE is also explainable by the large size of the written register. Since discourse markers are mostly a characteristic of informal spoken language, it is only natural that they occur less in CCIE than in ICE.

Although the results of Balasubramanian's study and the study on ICE-SIN and ICE-IND differ to some extent, similar observations based on the results have arisen. Balasubramanian (ibid., 146-147) makes conclusions that match the ones that have been made earlier in this thesis with regard

to the corpus study on ICE-SIN and ICE-IND. Some of the observations are that spoken contexts tend to include more indigenous loanwords than written contexts do and when they are used in discourse, they occur more naturally as part of the speech but in written texts, they are often specifically talked about and most of the time "appear as clearly borrowed" (ibid. 146). According to Balasubramanian, in many cases when and indigenous loanword is used there is no equivalent English word (ibid., 147), which is something that has also been noted earlier regarding the results of the my corpus study. Additionally, I noticed that Indian words especially have found their way into OED and are acceptable uses of English nowadays. In addition, Balasubramanian acknowledges that while words in some categories are clearly more frequent in certain registers, words belonging to 'Food' are most evenly spread out throughout the corpus (ibid.) which holds true for the results from ICE-SIN and ICE-IND as well.

Differences in the results from different corpora can in general be due to many factors, such as the background of the speakers and the compilation of the corpora (i.e. whether some corpus includes more written or spoken material). This, however, is not the reason between differences in results regarding the Singapore and Indian components of ICE since their structure is the same. The differences between ICE-SIN and ICE-IND results have already been discussed earlier in the thesis, but as of yet, no reason for the differences have been suggested. To answer one of the research questions of thesis, it must be investigated whether these differences can be explained by the history of English and different onsets for English in the countries or the language policies and planning governments of Singapore and Indian have implemented and followed. My argument is that the seeming lack of indigenous vocabulary in Singapore English (not counting discourse markers) as compared to the multitude of loanwords in Indian English can be explained by differences in the colonization histories and sociolinguistic backgrounds of the countries. Both of the nations have had very different experiences of colonization and therefore onsets for emergence of English language. As discussed earlier in Sections 2 and 3 about the history of English in the countries and subsection

4.2., Singapore was merely a trading colony for British merchants whereas India is often regarded as more of an exploitation colony which was ruled by British and exploited in many ways (Schneider 2007, 65-7). The kind of colonization experienced by India where the original population of the country was exploited and ruled by outsiders using a foreign language (English) could easily result in negative or mixed attitudes towards English language by the local people, and raise the appreciation of their indigenous languages which might be one of the reasons why Indian people like to favor indigenous loanwords in Indian English even today. According to Hohenthal (2003), while many Indians recognize the advantages of English and admit to using it to varying degrees, they still value their mother tongues more and prefer using Hindi or any of the other Indian languages whenever possible. Therefore, it would not be surprising that their appreciation of mother tongue carries over to Indian English in the sense of vocabulary loans. The attitude towards English by the Singaporean government and society at large has been completely different. Ever since Singapore became independent, English has gained more and more ground and is widely used in the country today (Tan 2014). English is used so widely that the majority of Singaporeans, especially the younger ones, identify English as the language which makes them Singaporeans (ibid. 335-7). This kind of language identification may certainly be connected to the low number of indigenous loanwords in Singapore English as compared to Indian English or most of the other World English varieties. As mentioned earlier, Singapore has conducted extensive language planning which has strengthened the role of English in Singapore while at the same time diminishing the sphere of indigenous languages. The most important language policy in regard to indigenous vocabulary in Singapore English would probably be the *Speak Good English* movement launched in 2000 but because plain texts versions of ICE-SIN were published only in 2002, it is safe to assume that the texts included in the corpus predate the language policy have not been influenced by the policy. However, it would be interesting to study whether the policy has had an effect on Singapore English used after 2000. Indian government, on the other hand, has not much enforced the use of English in India and although it has a co-official

status in India, it remains a language spoken by a minority of the population while Hindi plays the role of lingua franca language in the country. The different roles and extents to which English is used in the two countries today might also have more or less influenced the differences in the number of indigenous loanwords found in the corpora. In addition to the previously mentioned factors affecting indigenous vocabulary in the varieties, it must be added that the long history of Indian culture, the vast size of the country and its population, as well as the hundreds of languages spoken in India compared to small and much younger Singapore most probably contribute to the results as well.

The results of the corpus studies on ICE-SIN and ICE-IND confirm Schneider's (2007) claims on the kind of vocabulary that transmitted from indigenous languages into regional varieties of English. As previously stated in Chapter 4.2., loanwords from indigenous language are usually place names, words for animals and nature, and different kinds of cultural words quintessential for the region and while most of the words may remain just in local used, some may be transferred into international Standard English as well (Schneider 2007). According to Schneider, later during the nativization phase of the regional variety other types of vocabulary items (not only culture-specific words) are also borrowed from indigenous language, although not as extensively as cultural vocabulary (ibid., 44). As can be seen from the types of tokens gathered from the corpora, the results quite well correspond with Schneider's statements. Different types of words related to the culture of the region, whether that be food, music, arts, politics or something else, are the most borrowed items of vocabulary in addition to discourse makers in Singapore and Indian Englishes based on data found in ICE-SIN and ICE-IND. Indigenous words for animals and plants are also found but they are far more infrequent than vocabulary tied to culture. Other types of vocabulary loans from indigenous language picked up from the corpora include discourse markers (mostly in spoken contexts), words denoting people and groups of people, function words, verbs, adjectives and various types of miscellaneous vocabulary. The results seem to confirm Matras's (2009) claim about discourse markers being the most prone items to be carried over in language contact situations. Furthermore,

although not relevant for the aim of the corpus study, it was also observed that both of the varieties included a large number of place names borrowed from indigenous languages, as well as a fair amount of code-mixing, which Schneider (2009) claims are both essential parts of the formation processes of new varieties of English. It is difficult to further speculate on the development of the varieties based on Schneider's model, since the majority of the developments regarding vocabulary occur during the first three phases of the model. Therefore, it can only be said that my corpus study seems to confirm that both varieties have at least developed well into the third phase of the Dynamic Model. But as the theory presented earlier in this thesis and as Schneider suggests, Singapore English shows more signs of already being in the Endonormative Stabilization phase because the language is tied to the national identity of the people and appears largely homogenized, as compared to Indian English, which is not in any shape or form a stabilized and unified language nor does it act as an identity carrier among Indians since it is still only mostly spoken by the educated minority.

8. Conclusion

This thesis has examined the extent to which indigenous vocabulary is used in Singapore and Indian Englishes and further studied which types of loanwords occur more frequently than others. In addition, it has also examined in what registers the loanwords found occur in and whether their occurrence is triggered by lack of equivalent vocabulary in English or something else. The findings in ICE-Singapore and ICE-India were compared with each other and analyzed according to previously conducted research on World Englishes and vocabulary. The results clearly show that indigenous loanwords are much more frequently used in Indian English than they are in Singapore English.

Singapore English makes use of indigenous vocabulary mostly by borrowing food words and using an extensive amount of discourse markers originating from regional languages. Majority of the tokens gathered from the corpus presented different types of discourse markers, said to be characteristic features of Singapore English, a claim which is affirmed by the great number of discourse markers in the Singapore component of the ICE corpora. Apart from discourse markers and food words, other types of indigenous vocabulary in Singapore English were also found, but used rather infrequently in more specific contexts and not very often in natural spontaneous speech. A large percentage of the occurring tokens in Singapore English were used in the spoken section of ICE-SIN but that was mostly due to discourse markers and their use in spoken registers. When discourse markers were taken out of the equation, it became evident that the rest of the tokens were spread quite evenly between the spoken and the written section of the corpus. Some of the corpus subsections had more tokens than others, but with such a small number of tokens, it is hard to say anything more specific.

As mentioned, the Indian component of ICE contains a much greater number of indigenous loanwords than the Singapore component. The borrowed loanwords in Indian English cover all types of words associated with different cultural spheres, such as food, music, arts, religion, politics and trade, as well as terms for people and discourse markers. Results based on my corpus study indicate

that use of indigenous loanwords in Indian English is a common phenomenon in spoken and written registers but that they occur more frequently in spoken language. Leaving out discourse markers, which almost without an exception occur in 'Private dialogue', shows that other types of indigenous loanwords are used evenly in different subsections of the spoken register. The number of tokens in the written subsections of ICE-IND, on the other hand, are more varied since large majority of the tokens are used in 'Instructional writing' while the rest of the tokens are more evenly spread in the rest of the subsections of the written register.

Examining the context of the loanword occurrences showed that indigenous loanwords were used in the corpora for different purposes. Most of the time words from regional languages were borrowed into English because English language equivalents for the ideas, objects or entities just did not exist. Sometimes, the indigenous words were not just loanwords in the Singapore and Indian varieties of English but have come to be generally used in all English varieties. This is the case with many ideas, objects or entities which originate from the regions and have come to be known or used all over the world, such as *yoga*, as well as with many animals, plants, and fruits, like *durian.* Many of the concepts or entities only exist in Southeast Asia and are therefore known by their indigenous names which are borrowed into English. Some uses of the indigenous loanwords were not triggered by absence of corresponding English language words, but were more conscious choices of code-switching by the speaker or the writer, and at times some loanwords seemed to occur quite spontaneously in spoken contexts, especially in conversations. Surprisingly many of the loanwords found in the corpora have entries in the online version of *Oxford English Dictionary* but at the same time, not many of the words found in OED are listed in other English language dictionaries.

Comparing the results of the corpus study of this thesis with Balasubramanian's (2009) study and research on indigenous vocabulary on World Englishes showed that while most of the indigenous words borrowed into regional varieties are words associated with animals, plants and mostly different spheres of culture, the types of words borrowed can greatly vary depending on the studied corpus,

and its compilation, as well as the historical and present-day culture in the region where the variety is spoken in. Other factors which might contribute to the number of loanwords used in varieties are colonization experience and attitude of the people of the formerly colonized nation toward English, as well as language policies and planning exercised by the government.

Although Schneider's Dynamic model of Postcolonial Englishes (2007) provided a lot of useful historical information and framework for the study of the varieties, it was acknowledged that research on the development of Singapore English and Indian English would require more sociolinguistic and grammatical studies. This is because indigenous vocabulary is a crucial part of only the first three phases out of the total five phases varieties of Englishes go through in their process of nativization and standardization.

As for further research on vocabulary in World Englishes and the particular varieties, it would be interesting to conduct a corpus study on a bigger corpus to see whether the proportions of results in different categories remain the same. Also, it would be fascinating to know how extensively loanwords from different varieties occur in OED and other English dictionaries. These are only few of the questions that left me wondering while conducting research for the thesis, for indigenous vocabulary in World Englishes has not been studied that thoroughly and there is still plenty of room for further research.

Bibliography

Corpora

Nelson, Gerald. 2002. *The Singapore Corpus Manual*.
    <http://www-bcf.usc.edu/~billmann/diversity/ICE-SIN_Manual.PDF>
    Corpora available online at <http://ice-corpora.net/ice/>

Nelson, Gerald. 2002. *The ICE India Corpus User Manual*.
    Corpora available online at <http://ice-corpora.net/ice/>

Software

Anthony, Lawrence. 2014. AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda
    University. Available at < http://www.laurenceanthony.net/>

Works Cited

Alsagolff, Lubna. 2010. "English in Singapore: Culture, Capital and Identity in Linguistic
    Variation." *World Englishes* 29.3: 336-48.

Balasubramanian, Chandrika. 2009. *Register variation in Indian English.* Amsterdam/Philadelphia:
    John Benjamins Publishing Company.

Bao, Zhiming and Hong, Huaqing. 2006. "Diglossia and Register Variation in Singapore English."
    World Englishes, 25: 105-14.

Bhatt, Rakesh M. 2008. "Indian English: syntax" in *Varieties of English 4: Africa, Sout and
    Southeast Asia*, edited by Rajend Mesthrie, 546-62. Berlin/New York: Mouton de Gruyter.

Bolton, Kingsley and Ng, Bee Chin. 2014. "The Dynamics of Multilingualism in Contemporary
    Singapore." *World Englishes,* 33.3: 307-18.

Cambridge Dictionary Online. Cambridge University Press. < http://dictionary.cambridge.org>
    [Accessed 30 April 2017]

Census of India. 2001. "Data on Language"
    <http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/data_on
    _language.aspx> [Accessed 24 January 2017]

Department of Statistics Singapore. "General household survey 2015: Religion".
    <http://www.singstat.gov.sg/publications/publications-and-papers/GHS/ghs2015content>
    [Accessed 24 January 2017]

Deterding, David. 2007. *Singapore English*. Edinburgh: Edinburgh University Press.

Gupta, Anthea Fraser. 1989. "Singapore Colloquial English and Standard English". *Singapore Journal of Education*, 18: 31-57.

Gupta, Anthea Fraser. 2010. "Singapore Standard English Revisited" in *English in Singapore*, edited by Wee et al., 57-89. Hong Kong: Hong Kong University Press.

Hohenthal, Annika. 2003. "English in India: Loyalty and attitudes". *Language in India,* Vol 3. <http://www.languageinindia.com/may2003/annika.html> [Accessed 3 April 2017]

Jenkins, Jennifer. 2009. *World Englishes: A resource book for students*. 2nd ed. Oxon: Routledge

Kachru, Braj B. 2005. *Asian Englishes beyond the canon*. Hong Kong: Hong Kong University Press.

Kachru, Braj B. 1988. "The sacred cows of English". *English Today,* 16.4: 3-8.

Kachru, Yamuna and Cecil L. Nelson. 2006. *World Englishes in Asian Contexts*. Hong Kong: Hong Kong University Press.

Kachru, Yamuna and Larry E. Smith. 2008. *Cultures, Contexts, and World Englishes*. New York: Routledge.

Kuo, Eddie C.Y. 1980. "The sociolinguistic situation in Singapore: Unity in Diversity" in Alfredas, Evangelos A. (ed.) *Language and Society in Singapore*, 39-62. Singapore: NUS Press.

Lange, Claudia. 2012. *The syntax of spoken Indian English*. Philadelphia: John Benjamins Publishing Company.

Lee, Kuan Yew. 2012. *My lifelong challenge: Singapore's bilingual journey*. Singapore: Straits Time Press.

Leimgruber, Jakob R. E. 2011. "Singapore English." *Language and Linguistics Compass* 5.1: 47-62.

Lim, Lisa (eds). 2004. *Singapore English: A grammatical description*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lim, Lisa, Anne Pakir, and Lionel Wee. 2010. *English in Singapore: Modernity and management.* Hong Kong: Hong Kong University Press.

Macaulay, Thomas B. 1835. *Minute by the Hon'ble.* <http://www.columbia.edu/itc/mealac/pritchett/00generallinks/macaulay/txt_minute_education_1835.html> [Accessed 3 January 2017]

Matras, Yaron. 2009. *Language contact.* New York: Cambridge University Press.

McArthur, Tom. 2002. *The Oxford guide to world English*. Oxford: Oxford University Press. .

Mesthrie, Rajend and Bhatt, Rakesh M. 2008. *World Englishes: The study of new linguistic varieties.* Cambridge: Cambridge University Press.

Oxford English Dictionary Online. Oxford University Press. <http://www.oed.com>

Rukmini, S. 2014. "Sanskrit and English: there's no competition". *The Hindu*.
    <http://www.thehindu.com/data/sanskrit-and-english-theres-no-
    competition/article6630269.ece> [Accessed 3 January 2017]

Sailaja, Pingali. 2009. *Indian English*. Edinburgh: Edinburgh University Press.

Sailaja, Pingali. 2012. "Indian English: Features and sociolinguistic aspect". *Language and
    Linguistics Compass,* 6.6: 359-70.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world.* Cambridge:
    Cambridge University Press.

Schneider, Edgar W. 2011. *English around the World: An introduction*. Cambridge: Cambridge
    University Press

Sedlatschek, Andreas. 2009. *Contemporary Indian English: Variation and change*.
    Amsterdam/Philadelphia: John Benjamins Publishing Company.

Tan, Ludwig. 2003. "Topic Prominence and null arguments in Singapore Colloquial English", in
    Deterding et al. (eds), *English in Singapore: Research on* Grammar, 1–10. Singapore:
    McGraw-Hill.

Tan, Ying-Ying. 2014. "English as a 'Mother Tongue' in Singapore". *World Englishes* 33.3: 319-39.

Trudgill, Peter and Jean Hannah. 2002. *International English: A guide to the varieties of Standard
    English*. 4th ed. London: Arnold.

United Nations, Department of Economic and Social Affairs, Population Division. 2013. *World
    Population Prospects: The 2012 Revision.*
    <https://esa.un.org/unpd/wpp/publications/Files/WPP2012_HIGHLIGHTS.pdf> [Accessed 23
    January 2017]

United Nations, Department of Economic and Social Affairs, Population Division. 2015. *World
    Population Prospects: The 2015 Revision.*
    <https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf> [Accessed 23
    January 2017]

Wee, Lionel. 2008a. "Singapore English: Phonology" in *Varieties of English 4: Africa, South and
    Southeast Asia,* edited by Rajend Mesthrie, 259-77. Berlin/New York: Mouton de Gruyter.

Wee, Lioel. 2008b. "Singapore English: Morphology and syntax" in *Varieties of English 4: Africa,
    Sout and Southeast Asia*, edited by Rajend Mesthrie, 593-609. Berlin/New York: Mouton de
    Gruyter.

The World Bank website. "Data on Languages".
    <http://data.worldbank.org/indicator/SP.POP.TOTL?locations=IN> [Accessed 3 January
    2017]

Appendix 1

Singapore English results tables[1]

1) Food

WORD	MEANING	COMPONENT	TOKENS
asam	tea	S2B	1
assam	tea	W2D	2
ayam pakang	a chicken dish	S2B	1
bak-pau	a chinese salty bun	W2F	1
bee hoon	noodle dish	S1A	8
beehoon	noodle dish	W2C	4
blachan	a spicy sauce/paste	W2D	1
buah keras	candlenut	W2D	1
chai	tea	W1B	1
chapati	indian flatbread	S1A	1
chappati	indian flatbread	W2D	1
char siew	a cantonese meat dish	S1A	4
char siew	a cantonese meat dish	W2F	2
char siew	a cantonese meat dish	S2B	1
chee cheong fun	type of sauce	S1A	1
chin chye	a vegetable	S1A	2
chutney	indian sauce	W2D	1
chye sim	vegetable	S1A	1
congee	Asian rice porridge	W2D	1
currypok	a malaysian snack	W2B	1
dhall	an Indian dish	W2D	1
dhall	an Indian dish	S1A	1
dosai	a type of pancake	S1A	2
durian	a fruit	S2B	10
durian	a fruit	S1A	3
durian	a fruit	W2A	2
durian	a fruit	W2D	1
gado-gado	an indonesian dish	S1A	1
garam masala	indian spice mix	W2D	1
goreng	a rice dish	S2B	1
Hokkien mee	noodle dish	S1A	1
hong hong bao	asian type burger	S1B	1
ice-kacang	bowl of shaved ice (literally ice peanut)	S2B	1
ikan asam	sour fish	S2B	1

---

[1] The meanings and translations for the indigenous words were gathered solely for my own purpose with the help of online dictionaries and Wikipedia.

ikan bilis	driend anchovis	S1A	1
kai lan	a vegetable	S2B	1
kaiya	malaysian coconut jam	S2B	2
kalamanci	calamondin	S2B	1
kang kong	a vegetable plant	S2B	1
kangkong	water spinach	W2B	1
kaoliang	sorghum wine	W1A	1
kati	a street-food dish	W2D	3
kati taugeh	a street food dish	W2D	1
kofta curry	a food dish	W2D	1
kong piah	a pastry	S1A	1
koo chye	chive	W2D	1
kopi-o	a traditional black coffee	W2B	1
kua-chi	pumpkin seeds	W2F	1
kway teow	noodle dish	S1A	5
kway teow	noodle dish	S2B	4
kway teow	noodle dish	S1B	3
kway teow	noodle dish	W2B	1
makan	to eat/a meal	S2A	1
mee	noodles	W2A	1
mee siam	siamese noodles	S1A	1
mee soto	spicy noodle soup dish	S1A	1
mochai kottai	a type of bean	S1A	1
nanten	some food	S1A	1
nasi lemak	a rice dish	S1A	1
nasi padang	an indonesian rice dish	S2B	1
nasi-lemak	a rice dish	W2B	1
nonya	peranakan cuisine	S1A	1
nonya	peranakan cuisine	S2B	1
nonya kweh	a bit sized peranakan sweet snack	S1A	1
otah-otah	steamed fish	W2B	1
otar otar	a fish dish	S2B	1
pa choi	type of spinach	S2B	1
padi	rice	S1A	3
padi	rice	W2A	1
paiyasum	a South Asian dessert dish	S1A	3
pandan	eadible leaves	S1A	1
pandan	eadible leaves	W2F	1
pang teh	tea	S1A	1
pao	asian type burger	S1B	2
paos	noodles	S1B	1
poh piah	a fresh spring roll	W2B	1
popiah	a fresh spring roll	S1A	2

popiah rojak	(see popiah & rojak)	S1A	1
rambutan	a fruit growing in Southeast Asia	W2A	1
rambutan	a fruit growing in Southeast Asia	W2A	1
rendang	a spicy Indonesian meat dish	S1A	1
rojak	a traditional fruit and vegetable salad dish	S1A	1
rojak	fruit	S1B	1
roti prata	a flat bread	S1A	1
sago	starch	S2B	1
sago	starch	W2A	1
sago	starch	W2B	1
samba blachan	a spicy sauce/paste	S2B	1
sambal	a spicy sauce	S1A	1
sambal balachan	a spycy sauce/paste	S1A	1
Satay	a grilled meat dish with sauce	S1A	5
satay	a grilled meat dish with sauce	W2D	3
satay	a grilled meat dish with sauce	W2B	2
satay	a grilled meat dish with sauce	S2B	1
satay	a grilled meat dish with sauce	W1B	1
satay	a grilled meat dish with sauce	W2A	1
satay	a grilled meat dish with sauce	W2F	1
siew-mai	dumplings	W2F	2
ta bao	something to do with food and eating	S1A	1
tandoori	Indian style oven baked food	S2B	1
tandoori	Indian style oven baked food	W1B	1
tau cheo	soybean paste	W2D	1
tau kuah	some type of vegetable	W2D	1
taucheo	soybean paste	W2D	2
taugeh	brussel sprout	W2D	1
taugeh	brussel sprout	W2D	1
tempe kedele	indonesian food	W2A	1
tiao kai	noodle dish	S1A	1

164

2) Clothing and textile

WORD	MEANING	COMPONENT	TOKENS
badan	"body" but in this case part of sarung	W2A	1
barong	an embroided formal traditional shirt	W2F	1
batik	hand dyed fabric	S1A	3
batik	hand dyed fabric	W2A	5
batik	hand dyed fabric	W2C	1
batik	hand dyed fabric	W2F	1
batik	hand dyed fabric	W2A	1
cheongsam	a traditional chinese dress	W2B	1

cheongsam	a traditional chinese dress	W2F	1
cheongsam	a traditional chinese dress	W2C	1
dhotis	a traditional mens garment	S1A	1
kain panjang	a wrap-around skirt	W2A	1
kain songet	textile material	W2B	2
kepala	"head" but in this case part of sarung	W2A	1
parang rusak	a traditional batik pattern	W2A	1
saree	a traditional dress	W2F	1
sarong	a traditional blouse/dress	W2B	1
sarong	a traditional blouse/dress	S1A	1
sarong kebaya	a traditional blouse/dress	S2A	1
sarong kebaya	a traditional blouse/dress	W1A	1
sarongs	a traditional blouse/dress	S1A	2
sarung	a traditional blouse/dress	W2A	2
			31

3) Music and arts

WORD	MEANING	COMPONENT	TOKENS
bangsawan	malay opera	W2A	5
dalang	a shadow puppeteer	W2A	2
gamalan	traditional ensamble music of Indonesia	S1B	1
gamallon	traditional ensamble music of Indonesia	S1A	1
kompang	a traditional Malay drum instrument	W2C	1
kronchong	a ukulele-like instrument	W2A	1
kroncong	a ukulele-like instrument	W2A	1
makyong	a traditional dance-drama from Malaysia	W2A	1
orang	a theater term	W2A	1
orang muda	the leading male in bangsawan	W2A	3
sitar	an Indian traditional instrument	S1A	2
veena	a traditional Indian string instrument	S1A	3
wayang	traditional puppet theatre	S1A	1
wayang kulit	traditional puppet theatre	W2A	1
wayang wong	traditional dance	W2A	1
xiyang hua	something to do with paintings	S2A	1
			26

4) Religion

WORD	MEANING	COMPONENT	TOKENS
ashram	a spiritual hermitage or a monastery in Indian religions	W1B	5
ashrams	a spiritual hermitage or a monastery in Indian religions	W1B	1
chou-li	rites	W1B	1
dukkha	a buddhist concept of "suffering" and "pain"	W1B	1
hari raya	a religious festival celebrated by the muslims in Singapore (and Malaysia)	W1B	1

karma	in Buddhism, the sum of a person's actions in one life	W2F	1
keramat	a shrine	W2B	1
koan	a paradox to be meditated upon that is used to train Zen Buddhist monks	W1B	1
kowtow	an act of deep respect shown by prostration/worshipping in Buddhism	S1A	1
kowtow	an act of deep respect shown by prostration/worshipping in Buddhism	W2F	1
madrasah	private schools which are overseen by Islamic Religious Council of Singapore	W2C	1
qi kong	a holistic system	S2B	2
sedekah	alms	W2A	1
			18

5) People

WORD	MEANING	COMPONENT	TOKENS
amah	a female domestic servant	S2B	1
anak	a child	W2A	1
Ang Mo	"white people"	S1A	1
Ang Mo Lang	"red haired people"/ white people	S1A	1
ang moh	"white person/people"	S1A	1
ang mos	"white people"	S1A	2
ang-mo	"white person"	S1A	1
angat	an adopted person	W2A	1
angmohis	"whiteish"	S1A	1
baba	"local born" "local descendant" (chinese)	W2A	2
bo chap	an indifferent, not bothered, not caring person	S1A	1
bo-chap	an indifferent, not bothered, not caring person	W2B	1
buaya	crocodile (used for bad boys / not so good men)	W1B	1
bumi	the Malay race and other indigenous peoples of Southeast Asia	S1A	3
cucu	a grandchild	W2A	1
gundu	a fat person	S1A	1
kampung	a villager	W2D	1
kanna	a kid	S1A	1
kay poh	"a busybody person"	S1A	7
kwai lo	a white/european person	S1A	4
macik	an aunt	S2B	1
mahjong kaki	kaki= a person or a group of people who you do things with	W2B	1
makandan	a person who does something in a traditional malay wedding	S2B	1
mamasan	a woman in a position of authority	S2B	2
mata-mata	a police	W2F	1
peranakan	"local born" "local descendant" (chinese)	S1A	1
peranakan	"local born" "local descendant" (chinese)	S2B	2
peranakan	"local born" "local descendant" (chinese)	W1A	7
peranakan	"local born" "local descendant" (chinese)	W2F	3
pernakans	"local born" "local descendant" (chinese)	S2B	1
pernakans	"local born" "local descendant" (chinese)	W2A	1

samsui	chinese women who immigrated to Singapore between 1920-1940	S2B	3
sayang	an endearment	S2B	1
sinkheh	a newcomer	W2A	1
yogi	a yoga practitioner	S1A	1
yogi	a yoga practitioner	W2A	1
			61

6) Politics

WORD	MEANING	COMPONENT	TOKENS
jawatan kuasa blok	block committee	W2A	2
khairat kematian	welfare fund	W2A	1
			3

7) Discourse markers

WORD	MEANING	COMPONENT	TOKENS
aiya		S1A	3
aiya		W2F	3
aiyah		S1B	1
aiyah		S1A	87
aiyo		S1A	2
aiyoh		S1A	33
aiyuh		S1A	6
ayah		S1A	2
dah		S1A	2
hah		S1A	45
hah		S2A	2
hah		S2B	1
hah		W2F	1
hor		S1A	62
hor		S1B	1
hor		S2B	5
lah		S1A	1312
lah		S1A	293
lah		S1B	6
lah		S1B	4
lah		S1B	58
lah		S2A	15
lah		S2A	1
lah		S2B	3
lah		S2B	42
lah		W1B	2
lah		W2F	16

lah	S2A	4
leh	S1A	37
leh	S2B	2
leh	S1A	38
leh	S2B	5
loh	S1A	9
lor	S1A	116
lor	S1A	23
lor	S2B	4
lor	W1B	2
mah	S1A	14
meh	S1A	15
orh	S1A	46
suh	S1A	10
suh	S2A	1
wah	S1A	68
wah	S1B	5
wah	W1B	2
		2409

8) Animals and nature

WORD	MEANING	COMPONENT	TOKENS
belukar	a bush	W2F	1
gambien	a species of tropical bush	W2A	3
gambien	a species of tropical bush	W2B	1
gamier	an extract derived from the leaves of a climbing shrub	W2B	1
gingko	a maidenhair tree	W2B	1
gourami	an Asian fish	W2B	1
kapu	a tropical tree	W2F	1
keng wah	a plant/flower	W2D	1
lalang	a species of grass in Malaysia	W2A	1
lallang	a species of grass in Malaysia	W2A	1
lotong kelabu	silvery lutung (a monkey species)	W2B	1
maingayi	a tree/plant in Malaysia/Singapore	W2A	1
mata-puteh	a southeast Asian bird	W2C	3
medaka	a Japanese rice fish	W2A	5
merbok	a spotted dove	W2C	1
merbok	a spotted dove	W2B	1
myna	a bird	W2B	5
nipah	a type of palm tree	W2A	3
pariah	a stray dog	S1A	1
pitta	a bird	W2A	2
poonai	a cat	S1A	1

sharma	a bird	W2A	1
tembusu	a Southeast Asia tree	W2B	1
			38

9) Culture

WORD	MEANING	COMPONENT	TOKENS
ang pow	"red envelope"/"red packet"	S1A	1
ang pow	"red envelope"/"red packet"	S1B	1
ang pow	"red envelope"/"red packet"	S2B	5
ang pow	"red envelope"/"red packet"	W2B	1
angpow	"red envelope"/"red packet"	S2B	1
cha-tek	malaysian game/feathered ball	S2B	1
chap gee kee	a chinese card game	S2B	1
garang guni	a practition conducted by modern "rag-and-bone" men	S2B	1
kampong	a traditional village	S1B	3
kampong	a traditional village	S2B	5
kampong	a traditional village	W2A	4
kampong	a traditional village	W2F	1
kampongs	plural of kampong	S2B	1
kampung	a traditional village	W2F	4
kampungs	plural of kampong	W2F	1
kopi tiam	a traditional coffee shop	S2B	1
kowtow	showing respect by kneeling and bowing down	S1A	1
pasar malam	a night market	W1B	1
pasar malam	a night market	S1A	1
pencak	an indonesian martial art	S1A	1
renminbi	chinese currency	W2C	1
ringgit	malaysian currency	S1A	1
ringgit	malaysian currency	S2B	4
ringgit	malaysian currency	W1B	1
rupiah	indonesian currency	S2B	6
sepak takraw	Southeast Asian game of kick-volleyball	S2B	1
taichi	a chinese martial art	S2B	1
taiji	a chinese martial art	S2B	1
taman	a park/garden	W2B	1
			53

10) Other

WORD	MEANING	COMPONENT	TOKENS
alamak	exclamation, "oh no!"	S1A	6
alamak	exclamation, "oh no!"	S1B	1
alamak	exclamation, "oh no!"	S2A	1

ang kau	familiar singular "you"	W2B	1
bukalau	bukalau	S1A	1
fa lao sao	to whine	W1B	1
gelat	expression to denote extreme weariness and irritation at repetition	S1A	1
habis	a verb-like adjective that means both exhausted and finished	S1A	1
hantam	to beat or to hit	S2A	2
kana	mis-spelling of 'kena' passive	S1A	4
kau tim	to finish	S1A	1
kena	passive	S1A	2
kiasu	a grasping, selfish attitude	S1B	5
kiasu	a grasping, selfish attitude	S2A	2
kiasu	a grasping, selfish attitude	S2B	5
kiasu	a grasping, selfish attitude	W1B	1
kiasu	a grasping, selfish attitude	W2C	1
kiasuism	a grasping, selfish attitude	S1A	1
kiasuism	a grasping, selfish attitude	S1B	4
kiasuism	a grasping, selfish attitude	S2B	1
kiasus	a plural of kiasu	S1B	2
kosong	number zero	S2A	1
kuching kurak	something or someone is insignificant and half-baked, or incompetent	S1A	1
lao sai	diarrhea	S1A	1
lay chay	irritating, troublesome	S1A	1
makan	to eat/ to consume	S1A	4
makan	to eat/ to consume	S2A	2
makan	to eat/ to consume	W2A	1
maloo	(qualified/ashamed/embarrased?)	S1A	1
manja	to pamper/to take care	W2A	2
pak-tor	go on a date	S1A	1
parang	a malay knife	S1B	5
parang	a malay knife	S2A	1
parang	a malay knife	W2A	1
sama	same	W2F	1
sayang	to pity	S1B	1
shiok	a feeling of sheer pleasure and happiness	S1A	5
shiok	a feeling of sheer pleasure and happiness	S2A	4
shiok	a feeling of sheer pleasure and happiness	S2B	2
suka suka	to do as one likes, regardless of whether someone else approves	S1A	2
terok	irritating, troublesome	S1A	3
teruk	irritating, troublesome	S1A	2
ting xia lai	to stop	S1B	1
vanakam	hello!	W2B	1
xiao	small	S1A	3
zhen de	really!	S1A	1

Appendix 2

Indian English results tables[2]

1) Food

WORD	MEANING	COMPONENT	TOKENS
ajawain	bishop's weed	S2A	3
ajwain	bishop's weed	S2A	2
aloo	a potato	W2B	1
aloo ki bhaji	a type of bhaji	W2D	1
alu ka parontha	bread	S2A	3
alu ke paratha	a flatbread	S1A	1
alu ke parathe	a flatbread	S1A	1
amritsari machali	a fish dish	S2A	1
baati	a hard wheat roll	S1A	1
badam	almond	W2C	1
bajras	pearl millets	S1A	1
bajre ki roti	a type of bread	S1A	1
bajri	pearl millet	S1A	1
bansi	indian wheat seed	W1A	1
basmati	indian rice	W2F	1
batis	hard wheat rolls	S1A	1
besan	gram flour	W2D	1
bhaji	a spicy Indian snack or entree dish similar to a fritter	S1A	3
bhaji	a spicy Indian snack or entree dish similar to a fritter	W1B	1
bhakri	a round flat unleavened bread	S1A	7
bhakris	a round flat unleavened bread	S1A	1
bhalla	a North Indian and Pakistani snack food	S2A	2
bhallas	a popular fast food snack in Northern India	S2A	2
bhaturas	a fluffy deep-fried leavened bread	S2A	2
bhujiya	a popular crispy snack	S1A	1
biryani	Indian mixed rice dish	S1A	6
bonta	same as bati	S1A	1
brinjal	an aubergine	W2A	1
brinjals	aubergines	W2D	1
chaat	a savory snack in India, Bangladesh, Pakistan etc	W2B	3
chaat	a savory snack in India, Bangladesh, Pakistan etc	W2D	2
chaats	a savory snacks, typically served at road-side tracks from stalls or food carts	W2B	1
chaibandi	tea (something)	S1A	1
chaipatti bandi	tea (something)	S1A	1

---

[2] The meanings and translations for the indigenous words were gathered solely for my own purpose with the help of online dictionaries and Wikipedia.

chakli	a savoury snack from India	S1A	1
chakli	a savoury snack from India	W2D	1
chana	chick pea	W2D	5
chana dal	chick pea	S2A	1
chanas	chickpeas	S2A	4
chandrakala	an Indian Diwali sweet recipe with maida and khoya as main ingredients	W2C	1
chapati	Indian flatbread	S1A	14
chapati	Indian flatbread	S2A	5
chapaties	an unleavened flatbread	W2D	1
chapatis	Indian flatbreads	S1A	3
chapatis	Indian flatbreads	W2D	1
chappattis	an unleavened flatbread	W1B	1
chatpata khana	Zesty Chickpeas	S1A	1
chawal	rice	S1A	2
chhole parathe	a type of flatbread	S2B	1
chikki	a traditional Indian peanut sweet	W2D	5
chole bature	a combination of spicy chick peas and fried bread	S2A	2
chota hazari	a meal served in households and barracks shortly after dawn	W2F	1
churmur	crunchy street food snack	S1A	1
chutney	a sauce in the cuisines of the Indian subcontinent	S2A	1
chutney	a sauce in the cuisines of the Indian subcontinent	W1B	1
chutney	a sauce in the cuisines of the Indian subcontinent	W2B	1
chutney	a sauce in the cuisines of the Indian subcontinent	W2D	2
chutneys	a sauce in the cuisines of the Indian subcontinent	W2B	1
chutneys	a sauce in the cuisines of the Indian subcontinent	W2D	1
daal	lentil/pea/bean	S1A	5
daal baati	an Indian dish comprising dal (lentils) and baati (hard wheat rolls)	S1A	1
daal batti	an Indian dish comprising dal (lentils) and baati (hard wheat rolls)	S1A	1
daba	a roadside restaurant in India	S1B	1
dahi	yogurt	W2B	4
dahi balle	a popular fast food snack in Northern India	S2A	1
dahi bhalla	a popular fast food snack in Northern India	S2A	1
dahi bhallas	a popular fast food snack in Northern India	S2A	1
dahi wada	a popular fast food snack in Northern India	S2A	4
dahiwadas	a popular snack in India	W2B	1
dal	lentil/pea/bean	S1A	1
dal	lentil/pea/bean	S2A	2
dals	lentil/pea/bean	S1B	1
dals	lentil/pea/bean	W2D	2
dawat	a feast	S2A	3
desi ghee	"country made" clarified butter	S1A	6
desi ghee	"country made" clarified butter	S2A	1
dhaba	a roadside restaurant in India	S1A	1
dhabas	roadside restaurants in India	W2B	1
dhokla	a vegetarian food item that originates from the Indian state of Gujarat	W2D	2
dosa	a savoury cake	S1A	1

elaichi	cardanom	W2D	2
garam masala	a spice mix	S2A	1
garam masala	a spice mix	W2D	5
gatta	a besan dumpling	S1A	1
gatte	a besan dumpling	S1A	1
ghee	clarified butter or vegetable oil	S1A	4
ghee	clarified butter or vegetable oil	S2A	2
ghee	clarified butter or vegetable oil	W2D	5
ghee bhat	a Bengali rice dish	S1A	1
gobi ke parathe	a flatbread	S1A	1
golguppas	Indian street food	W2B	1
goshtaba	a dish of minced mutton balls cooked in curd and spices	W2B	1
gulab jamun	a milk-solid-based South Asian sweet	S1A	1
gulab jamun	a milk-solid-based South Asian sweet	W2D	1
gurdani	some type of food	W2C	1
haldi	turmeric	S1A	3
idli	a savoury cake	S1A	1
idli	a savoury cake	S1B	1
idli	a savoury cake	W1B	1
idli dosa	a savoury cake	S1A	2
jaiphal	nutmeg	W2B	1
jalebis	a sweet popular in countries of South Asia	W2C	1
jilabi	a crunchy snack	S1A	1
jowari	something edible	S1A	2
kabuli chana	chickpea	W2D	1
kachoris	a sweet dish dipped in sugar syrup	S2A	1
kala jamun	an exotic sweet dish served on special occasions	W2C	1
kaliya	a fish dish	S1A	1
kalmi kabab	a delicious Indian recipe served as a Side Dish	W2B	1
kalyan sona	a wheat type	W1A	1
kanak	wheat	W2F	3
karam ka saag	a leaf-based dish	W2B	1
keema	a traditional meat dish	S1A	2
keema pulav	minced meat and rice dish	S1A	1
kesharbadam	a milk drink made from saffron, almonds and milk	W2C	1
khajur	a date	S2A	1
khatti mithi chutney	sweet & sour chutney	S2A	2
kheema matter	Minced mutton and green peas turned into a spicy dish	W2D	1
kheer	a rice pudding	W2B	1
khus-khus	poppy seeds	W2D	1
kismish ka raita	an Indian vegetable dish	S2A	1
kurmura	puffed rice	W2C	1
lal maas	a Rajasthan curry dish	S2A	4
lonavala chikki	a type of chikki	W2D	1
maida	a wheat flour from India	S2A	1
maida	a wheat flour from India	W2D	3

makke ki roti	a type of bread	S1A	2
malai	a South Asian cooking ingredient	W2C	1
mamra-badam	almond	W2D	1
masala	spice or spice mix	S1A	1
masala	spice or spice mix	S2A	1
masala	spice or spice mix	W2D	5
masalas	spices	S1A	1
methi	fenugreek	S2A	1
methi	fenugreek	W2D	1
moongfali	peanut	S1A	1
mung dal	mung bean	S2A	1
murgh musallam	Masala roasted chicken	W2D	1
murghi badami	curried chicken in almond sauce	W2D	1
murkagoli	a chicken dish	S2A	1
navratna farsan	a kind of Chivda that uses lentils like chick peas and moong in its fried mix	W2C	1
paan	a leaf snack	S1A	5
paan dhabba	roadside restaurants (which in this case serve paan)	S1A	2
palak	spinach	S2A	2
palak	spinach	W2B	3
palak sev	a spinach dish	W2C	1
paneer	a type of fresh cheece	W2D	6
paneer ka tikka	an Indian vegetable dish	S2A	1
pani	water?	W2B	1
pani-puri	a common street snack	S1A	1
papad	a tortilla like preparation that can be fried or roasted	S1A	1
papad	a tortilla like preparation that can be fried or roasted	W1A	1
papadams	a thin, crisp, disc-shaped food from the Indian subcontinent	S2A	1
papadum	a thin, crisp, disc-shaped food from the Indian subcontinent	S2A	3
pappad	a thin, crisp, disc-shaped food from the Indian subcontinent	S1A	5
papri	chip/nacho-like thin crispy snacks	W2B	2
papri chaat	a type of chaat	W2B	1
parathas	a flatbread that originated in the Indian subcontinent	S1A	1
parathas	a flatbread that originated in the Indian subcontinent	W2D	1
parathas	a flatbread that originated in the Indian subcontinent	W2F	2
parontha	a flatbread that originated in the Indian subcontinent	S2A	1
paronthas	a flatbread	S2A	2
pau bhaji	an Indian street food dish	S1A	1
payas	a traditional South Asian food	S2A	3
payasam	kheer	S1B	1
pedas	a sweet from the Indian subcontinent	S1A	1
phirni	a classic indian sweet pudding	W2B	1
pista chikki	a type of chikki	W2D	1
pudiana	mint	W2D	1
puran poli	a maharashtrian sweet flat bread	S1A	3
puran poli	a maharashtrian recipe of sweet flat bread	S1A	3
puratha	a flatbread	S2A	1

putthi	something edible	S1A	1
raita	a condiment from the Indian subcontinent	S2A	3
raita	a condiment from the Indian subcontinent	W2D	1
raitas	a condiment from the Indian subcontinent	S2A	3
rasagullas	a syrupy dessert popular in the Indian subcontinent	S1A	1
rasgulla	a syrupy dessert popular in the Indian subcontinent	S1A	1
rasmalai	a dessert originating from the Indian subcontinent	W2C	1
roomali rotis	flatbread	W2B	1
rosgullas	a syrupy dessert popular in the Indian subcontinent	W2C	1
roti	a flatbread originating from the Indian subcontinent	S1A	3
rotis	flatbread	W2D	1
saag	a leaf-based dish	S1A	2
saagwala dal	leaf & lentil dish	W2D	1
sabji	cauliflower	S1A	3
samari	some type of food	S1A	1
sambar	a lentil-based vegetable stew or chowder	W2F	1
sambar dabba	a roadside restaurant that sells sambar	S1B	1
sambar daba	a roadside restaurant that sells sambar	S1B	1
samosas	a fried or baked dish with a savoury filling	S1A	2
samosas	a fried or baked dish with a savoury filling	S2A	2
samosas	a fried or baked dish with a savoury filling	W2D	5
samosas	a fried or baked dish with a savoury filling	W2F	1
sarson ka saag	Indian sauce/dip	S1A	4
sarson ka sag	Indian sauce/dip	S1A	1
satvik	Sattvic diet is a diet based on foods in Ayurveda and Yoga literature	S1A	2
satvik rajsik	Sattvic diet is a diet based on foods in Ayurveda and Yoga literature	S1A	1
satvik rajsik	Sattvic diet is a diet based on foods in Ayurveda and Yoga literature	S2B	1
shahi	indicates its status as a prestige dish	W2D	1
shahi korma	an Indian dish	W2D	1
shanjahani	an Indian dish	W2D	1
sona	indian wheat seed	W1A	1
sowar	sorghum	S1A	3
tabakmaaz	lamb ribs	W2B	1
tali	an Indian-style meal, made up of a selection of various dishes, served on a platter	S1A	1
tamsik	There are three categories of food in yoga; Rajsik, tamsik and Satvik	S2B	1
tamsik	There are three categories of food in yoga; Rajsik, tamsik and Satvik	S1A	1
tandoor	a type of oven / food cooked in that oven	S2A	4
tandoori	a type of oven / food cooked in that oven	S2A	3
tandoori	a type of oven / food cooked in that oven	W2B	5
tangadi	leg (a chicken leg)	W2D	1
tikka	a type of Indian food, usually served as an appetizer	S2A	2
tikka	a type of Indian food, usually served as an appetizer	W2B	1
tikka masala	an Indian dish	W2D	1
udad dal	black lentil	S2A	1
udad dal	black lentil	S2B	2
urad dal	black lentil	W2D	2

vada pau	a vegetarian fast food dish	S2B	6
wazwaan	a multi-course meal in Kashmiri cuisine	W2B	2
			395

2) Clothing and textile

WORD	MEANING	COMPONENT	TOKENS
addakam	a type of hand-painted or block-printed cotton textile	W2D	1
amru	woven from silk and zari on silk to produce variegated designs	W2D	1
baandha	a type of textile cloth	W2D	5
Baluchar sarees	plural, a type of saree from the Bengal area	W2D	1
baluchari butidar	a type of saree	W2D	2
butta	a female garment from the Indian subcontinent	W2D	1
chanderi saree	a traditional sari made in Madhya Pradesh, India	W2D	1
chappals	footwear	S1A	2
chappals	footwear	S2A	1
chikan	a style of embroidery, common in India	S2B	5
churidar	tightly fitting trousers worn by both men and women in South Asia	W2C	1
churidar	tightly fitting trousers worn by both men and women in South Asia	W2F	1
dupatta	a long, multi-purpose scarf that is essential to many South Asian women's suits	S1A	1
durri	a thick flat-woven rug or carpet] used traditionally in India as floor-coverings	W1A	1
ghagra	a traditional dress	S1A	1
himru	a fabric made of silk and cotton	W2D	5
himrus	a fabric made of silk and cotton, which is grown locally in Aurangabad	W2D	1
jala	a weaving pattern	W2D	1
jamdanee	one of the finest muslin textiles of Bengal	W2D	5
jari	an even thread traditionally made of fine gold or silver	W2D	2
jari bhutta	a type of jari	W2D	1
kalamkaari	a type of saree	W2D	1
kalamkari	a type of hand-painted or block-printed cotton textile	W2D	5
kalamkaries	a type of saree	W2D	1
kameez	a traditional outfit originating in the Indian subcontinent	W2F	1
khadi	a term for handspun and hand-woven cloth	S2A	1
khadi kurta	a type of khadi	W2F	1
kuccha	Short trousers ending above the knee	W2C	1
mundu	a garment worn around the waist in Kerala	S1A	1
namdahs	a felted wool carpet, woven in very few states of India	W2F	1
paithan saree	a variety of sari	W2D	1
palampores	a type of chintz cloth used for bedspreads, wall hangings, etc.	W2D	2
pallu	The loose end of a sari	W2F	1
panchha	a traditional garment for men	S2B	2
punjaa	Crotch of the pants, part of the cloth that is between the legs	W2D	1
salwar	a traditional outfit originating in the Indian subcontinent	S1A	1
salwar	a traditional outfit originating in the Indian subcontinent	W2F	1

salwar-kameez	a traditional outfit originating in the Indian subcontinent	W2F	1
sarees	plural of saree, a female garment	S1A	15
sarees	plural of saree, a female garment	S2A	10
sarees	plural of saree, a female garment	W2C	1
sarees	plural of saree, a female garment	W2D	1
sari	a female garment from the Indian subcontinent	W2F	5
saris	a female garment from the Indian subcontinent	W2F	3
seborag jamdanee	a type of jamdanee	W2D	1
seborag jamdanee	one of the finest muslin textiles of Bengal	W2D	1
sherwani	a long coat-like garment worn in the Indian subcontinent	W2D	1
			101

3) Music and arts

WORD	MEANING	COMPONENT	TOKENS
a laap	The opening, improvised section of a Classical raga performance	W2D	1
aahariya	a music term	S2B	1
aahariya abhinaya	a music term	S2B	1
aahariya abhinaya	expression of costume and scene	W2D	1
aangika abhinaya	expression of the limbs	W2D	2
aatakathas	a literary genre in Malayalam language	S2B	1
abhinaya	the art of expression in Indian aesthetic	S2B	2
abhinaya	the art of expression in Indian aesthetic	W2B	5
abhinaya	the art of expression in Indian aesthetic	W2D	4
abhinayam	the art of expression	S2B	1
adavu	basic step in nritta dance	W2D	1
aharyabhiava	the expression of costume and scene in Indian theatres	S2B	1
alaap	The opening, improvised section of a Classical raga performance	W2D	7
alaap-jor	when a steady pulse is introduced into the alap, it is called jor	W2D	4
alankara	a concept in Indian classical music	S2A	1
alapa	the art musics of South Asia	W2A	2
alapadma	mudra hand gesture	W2D	1
alapana	a form of melodic improvisation that introduces and develops a raga	W2D	1
alapanas	mudra hand gesture	W2D	1
alarippu	traditionally the first dance piece that dancers learn and perform	W2B	1
anudaatta	a musical note	S2B	1
anupallavi	In Carnatic music, the anupallavi comes after the pallavi	W2D	1
asammita mudra	a type of hand gesture	S2B	1
ashtapadi	Indian hymns where the music has eight lines	S2B	2
ashtapadi	Indian hymns where the music has eight lines	W2B	1
ashtapadi	Indian hymns where the music has eight lines	W2D	1
ata tala	a common tala	W2D	1
atakatha	Aattakatha (performance), story for dancing and acting	S2B	1
atam	a music/theater term	S2B	1
avaroha	in the context of Indian classical music, is the descending scale of any raga	W2D	1
badhat	an Indian music term	W2D	1
bandha	an art term	W2D	1

bandish	a fixed, melodic composition in Hindustani vocal or instrumental music	S2A	1
bandishes	a fixed, melodic composition in Hindustani vocal or instrumental music	W2C	1
banis	a term in carnatic music	W2B	1
bharat natyam	form of Indian classical dance from Tamil Nadu	W2D	1
bharata natyam	form of Indian classical dance from Tamil Nadu	W2B	5
bhavai	a popular folk theatre form of western India	W2D	1
bhavas	the emotion or mood conveyed by a performer	S2B	1
bhavatmaga	a theater term	S2B	2
bhavatmaka	a theater term	S2B	1
bhibhatsa	one of the eight rasas in the dramatic theory of Indian arts; disgust	W2A	1
bhinna shadja	a hindustani music term	W2D	1
bhoopali	a Hindustani classical raga	W2D	1
bibhatsa	one of the eight rasas in the dramatic theory of Indian arts; disgust	S2A	1
bol-taans	a type of taan that can be sung by utilizing the words of the bandish	W2D	1
chakyarkut	a performance art from Kerala, South India	S2B	1
chakyarkuta	a performance art from Kerala, South India	S2B	1
chanda	a cylindrical percussion instrument	S2B	3
charukesi alapana	a form of melodic improvisation that introduces and develops a raga	W2D	1
chengala	a metal circular Keralan instrument	S2B	1
churnika	a music form	W2B	3
dandamu	a music term	W2D	1
dha komal	a note in Indian music	S2B	1
dhaivat	a musical note	S2B	1
dhaivat	a musical note	W2D	1
dhenuka	Dhenuka (pronounced dhēnukā) is a rāgam (musical scale) in Carnatic music	W2D	1
dholki	a South Asian two-headed hand-drum	W2D	1
dhrupad	a genre in Hindustani classical music	W2A	1
dhrupad	a genre in Hindustani classical music	W2D	5
dhwani	melody/music	S2A	2
drut	the concluding section, in fast tempo	W2D	2
drut ek-taal	a music term	W2D	3
drut et-taal	a type of taal	W2D	3
drut khayal	faster tempo khayal	W2D	2
dwitiya	a musical note	S2B	1
ek-taal khayal	a type of khayal	W2D	1
eluttalam	an instrument / a term relating to music	S2B	1
esraj	a string instrument	W2D	1
gamaka	refers to ornamentation that is used in the performance of Indian classical music	W2A	1
gana	something relating to Indian music	W2A	2
gandhar	a musical note	S2B	2
gandhara	a musical note	S2B	1
gandharva	a term for skilled singers in Indian classical music	W2A	3
ganpati	a dance	S1A	1
gatkari	idiomatic instrumental elements	W2D	1
gats	gat = a fixed, melodic composition in Hindustani vocal or instrumental music	W2D	4
gattodas	a music term	W2D	1

gayaki	Indian vocal music	W2C	1
gharana	a system of social organization linking musicians or dancers by lineage or apprenticeship	S2B	1
gharana	a system of social organization linking musicians or dancers by lineage or apprenticeship	W2C	5
gharana	a system of social organization linking musicians or dancers by lineage or apprenticeship	W2D	1
gharanas	a system of social organization linking musicians or dancers by lineage or apprenticeship	S2B	1
gharanas	a system of social organization linking musicians or dancers by lineage or apprenticeship	W2C	3
ghazal	a poetic form with rhyming couplets and a refrain, each line sharing the same meter	W2C	2
ghazals	a poetic form with rhyming couplets and a refrain, each line sharing the same meter	W2F	1
gouri	a raga	W2D	2
guna alankara	a concept in Indian classical music	S2A	2
hindolam	a rāgam in Carnatic music	W2D	1
idakya	an instrument	S2B	2
iluttalam	a musical term	S2B	1
janya raga	a Carnatic music raga	W2D	1
jathi	Jathi of a tālam specifies beat count of the rhythm cycle	W2D	1
javalis	a music term	W2B	2
jhap-taal	one of the most famous talas of Hindustani music	W2D	2
jhinjhoti	a Hindustani classical music raga	W2D	1
jor	a music term	W2D	2
jugupsa	one of the eight bhavas; disgust	S2A	1
kaata	an Indian style of religious storytelling, performances of Hinduism	S2A	3
kahalya	a tribal dance	W2D	1
kaimani	a Keralan instrument	S2B	1
kalam	a paint brush	W2D	2
kalyan	a type of raga	W2D	1
kanjira	a South Indian frame drum	S2B	1
kathak	one of the ten major forms of Indian classical dance	S2B	1
kathak	one of the ten major forms of Indian classical dance	W2B	1
katputhli	a string puppet theatre, native to Rajasthan, India	W2B	1
katputhlis	a string puppet theatre, native to Rajasthan, India	W2B	1
kavyangas	limbs of poetry	S2A	1
keeravani	a rāgam in Carnatic music	W2D	1
kharaharapriya	a rāga in Carnatic music	W2D	2
khayal	modern genre of classical singing	W2D	5
khayalas	modern genre of classical singing	W2D	1
khayals	the modern genre of classical singing in North India	W2D	5
killekyatra	shadow puppetry	W2B	2
kirwani	a musical scale in Hindustani classical music	W2C	1
kirwani	a musical scale in Hindustani classical music	W2D	2
kodiattam	a form of Sanskrit theatre traditionally performed in the state of Kerala, India	S2B	1
komal	music notes in indian music	S2B	4
komal nishad	a type of music note?	W2D	1
kriti	a form of Carnatic music	W2B	1
kriti	a form of Carnatic music	W2D	4
kudiattam	a form of Sanskrit theatre traditionally performed in the state of Kerala, India	S2B	1
kuttu	a theater term	S2B	1

laghu	a music term	W2D	1
laghus	a music term	W2D	1
layakari	Laya means tempo and 'kari' means doing; a musical term	W2C	1
lokadharmi	the mode of representation in Indian arts that deals with the worldly activity of people	W2B	1
maand	a famous style of singing in Rajasthan	W2D	1
madanikas	Bracket figures	W2B	1
maddalam	a drum made out of the wood of the jackfruit tree	S2B	2
madhya saptak	middle octave	S2B	1
madhyam	a note on the shuddha scale	S2B	3
madhyam	a note on the shuddha scale	W2D	2
malhar	an old raga in Indian classical music	W2F	1
Mandala Sthana	a dance position	W2D	1
mandra saptak	lower octave	S2B	1
manjutala	the first portion of pada	S2B	1
manodharmatam	an art term	S2B	1
mehndi	a form of body art from Ancient India	S1A	5
melappada	a music term	S2B	1
mohini attam	a semi-classical dance form of Kerala	W2B	1
mohiniyattam	one of two classical dances of India that developed in the state of Kerala	S2B	1
moorchana	change of scale	W2D	1
mrigandam	an Indian drum	W2D	4
mudra	a hand gesture in Indian classical dance	S2B	1
mudra kadagamugham	a hand gesture in Indian classical dance	S2B	1
mudra kadagamugram	a type of hand gesture	S2B	1
muyav	an instrument	S2B	2
natya	a dance/music term	S2B	3
natya	a dance/music term	W1B	1
navarasabhinaya	a dance term	S2B	1
nayika	In the performing arts (especially dance) of India: a lead female role	W2D	1
nishada	a twelve-note system in Carnatic music	W2D	1
nishadam	a music term	W2D	1
nrita	classical dance tradition	S2B	5
nritta	classical dance tradition	S2B	3
nritta	classical dance tradition	W2D	2
nritya	classical dance tradition	S2B	3
nritya	classical dance tradition	W2B	2
odissi	a major ancient Indian classical dance	W2B	1
odissi	a major ancient Indian classical dance	W2D	2
oppana	a popular form of social entertainment among the Mappila community of Kerala	W2D	1
pada	a part of music	S2B	3
padam	a form of musical composition in Carnatic Music	W2D	1
paiccal adavu	a certain type of step in nritta dance	W2D	1
PaikaAkhada	a dance term	W2D	1
paikakhadas	a dance term	W2D	1
pakhawaj	a Percussion instrument	W2D	1
pallavi	In Carnatic music pallavi' is the thematic line of a song	W2D	3

pallavis	the thematic line of a song	W2D	1
pancham	a note on the shuddha scale	S2B	3
pantuvarali	a rāgam in Carnatic music	W2D	1
pataka	a mudra hand gesture	S2B	1
pataka tritakam	a mudra hand gesture	S2B	1
pilu	a raga (melodic mode) of Indian classical music	W2D	1
prabandhas	a literary genre of medieval Indian Sanskrit literature	W2A	1
pramana	refers to epistemology in Indian philosophies	W2A	1
prathama	a musical note	S2B	1
puriya danashri	a rāga in Hindustani classical music	W2D	1
puriya kalyan	a type of raga	W2D	1
purvaranga	Purvaranga or the preliminaries of a play is an important chapter in Indian theatre	W2A	1
qawwalis	a form of Sufi devotional music popular in South Asia	W2F	2
raag	raga	W2C	2
raga	a melodic mode in Indian classical music	S2B	4
raga	a melodic mode in Indian classical music	W1B	1
raga	a melodic mode in Indian classical music	W2A	5
raga	a melodic mode in Indian classical music	W2B	1
raga	a melodic mode in Indian classical music	W2D	10
raga Anand Shri	a type of raga	W2D	1
raga Gouri	an Indian musical saga	W2D	1
raga Maand	Rajastani folk music	W2D	1
raga Shankara	Hindustani classical music	W2D	1
raga Shree	a rare but popular concert raga	W2D	1
raga-dhyana	a type of raga	W2A	1
raga-dhyanas	a type of raga	W2A	1
raga-hyuana	a type of raga	W2A	1
raga-hyuanas	a type of raga	W2A	1
ragam	raga	W2D	2
ragas	melodic modes in Indian classical music	S2B	6
ragas	melodic modes in Indian classical music	W2A	3
ragas	melodic modes in Indian classical music	W2D	4
ramkali	a sweet early morning raga in Hindustani classical music	W2D	1
rasa	an ancient concept in Indian arts about the aesthetic flavor	S2A	5
rasa	an ancient concept in Indian arts about the aesthetic flavor	W2A	2
rasabhinaya	the essence of emotive acting	S2B	1
rasas	an ancient concept in Indian arts about the aesthetic flavor	S2B	1
riti	The mode of expression or the style of a literary composition	S2A	2
roopak	a popular tala in Hindustani music	W2D	2
sadir katcheri	the dance known as sadir katcheri (artistic dance) served as a form of worship	W2B	1
salagabhairavi	a rāgam in Carnatic music	W2D	1
samyuta mudra	a type of hand gesture	S2B	1
sanchari rasas	elements of ancient Sanskrit text of dramatic theory	W2A	1
santoor	a trapezoid-shaped hammered dulcimer or string musical instrument	W2D	5
sarangi	a bowed, short-necked string instrument from India	W2D	1
sargam taans	If the vocalist sings the notes of the tan it is called a "sargam tan"	W2D	1

sarod	a stringed instrument of India, used mainly in Indian classical music	W2C	1
sarod	a stringed instrument of India, used mainly in Indian classical music	W2D	2
satyakabhinaya	a dance term/music/theater	S2B	1
shadaja	a musical note	S2B	4
shahanai	a musical instrument similar to the oboe, common in India	S1B	1
shiravanjani	a musical scale used in Indian classical music	W2D	1
shoka	a rasa	S2B	1
shringara	one of the nine rasas	S2B	3
shringara	one of the nine rasas	W2D	2
shrutis	the smallest interval of pitch that the human ear can detect	S2B	5
shuddha	a musical scale	S2B	1
Shuddha kalyan	a type of raga	W2D	5
shuddha madhyam	a note on the shuddha scale	W2D	1
shudha swaras	a Sanskrit word that connotes a note in the successive steps of the octave	S2B	1
shudhha	a musical note	S2B	1
sitar	a plucked stringed instrument used mainly in Hindustani music and Indian classical music	S2B	1
sitar	a plucked stringed instrument used mainly in Hindustani music and Indian classical music	W2C	2
sitar	a plucked stringed instrument used mainly in Hindustani music and Indian classical music	W2D	5
sloka	a category of verse line developed from the Vedic Anustubh poetic meter	W2B	1
slokas	a category of verse line developed from the Vedic Anustubh poetic meter	W2B	2
sopan	a form of Indian classical music that developed in the temples of Kerala	S2B	1
sopan sangeet	a style of singing	S2B	1
sopan sangeeta	a form of Indian classical music that developed in the temples of Kerala	S2B	1
sopana	a method of singing	S2B	4
sovaraksharam	a music term	W2D	1
sthayi	an initial phrase or line of a fixed, melodic composition in Hindustani music	W2A	1
svara	a note in the successive steps of the octave	W2A	1
svaras	a Sanskrit word that connotes a note in the successive steps of the octave	W2A	1
swara	a single note, which defines a relative (higher or lower) position of a note	W2D	1
swarakalpana	raga improvisation within a specific tala	W2D	2
swarams	a Sanskrit word that connotes a note in the successive steps of the octave	W2D	1
swaritta	a musical note	S2B	1
taal	a music term	W2D	1
taals	a music term	W2D	1
taankari	a music term	W2D	1
taans	taan = a virtuoso technique used in the vocal performance of a raga	W2C	1
taans	a music term	W2D	1
taans	taan = a virtuoso technique used in the vocal performance of a raga	W2D	3
tala	the term used in Indian classical music to refer to musical meter	S2B	5
tala	the term used in Indian classical music to refer to musical meter	W2A	4
talas	a music term	S2B	1
tamas	a music term	W2B	2
tanam	one of the methods of raga improvisation (manodharma)	W2D	1
tara shadja	a hindustani music term	W2D	1
tarana	a type of composition in Hindustani classical vocal music	W2D	3
teen-taal	a music term	W2D	6

teernaman	a dance term	W2D	1
theka	a musical composition in classical Indian music for percussion instruments	W2D	1
thumir-ang	a music term	W2D	1
thumri	a common genre of semi-classical Indian music	S2B	1
thumri	a common genre of semi-classical Indian music	W2A	1
thumri	a common genre of semi-classical Indian music	W2C	1
thumri	a common genre of semi-classical Indian music	W2D	2
tivra	a note	S2B	1
tivra saptak	higher octave	S2B	1
todi	a Gurmat classical raga	W1B	1
todi	a Gurmat classical raga	W2D	1
Vamsaka Brinda	flute-led orchestra	W2D	1
varnam	a form of song in the Carnatic music repertoire	W2D	4
veena	a multistringed chordophone of the Indian subcontinent	W2A	1
veena	a multistringed chordophone of the Indian subcontinent	W2D	1
vidusaka	a comic stock character in classical Sanskrit drama.	W2D	2
vidushaka	a comic stock character in classical Sanskrit drama.	S2B	1
wachikabhinaya	Expression of speech in Indian music and drama	S2B	2
yakshagana	a theatre form that combines dance, music, dialogue, costume etc.	S2B	5
yakshagana	a theatre form that combines dance, music, dialogue, costume etc.	W2B	4
yaman kalyan	a type of raga	W2D	1
			490

4) Religion

WORD	MEANING	COMPONENT	TOKENS
abhayasa	a spiritual practice which is regularly and constantly practised over a long period of time	W2A	5
acharya	a preceptor or instructor in religious matters	S1A	1
ahimsa	not to injure "compassion", virtue in Indian religions	S1A	1
ahimsa	not to injure "compassion", virtue in Indian religions	S1B	5
ahimsa	not to injure "compassion", virtue in Indian religions	W1A	1
akal purakh	a Sikh name used for God	W2F	1
anadi vasana	"beginnengless impression" in Indian religions	W2A	1
anumana	in Indian philosophy, the second of the pramanas, or the five means of knowledge	S2A	1
apsara	a female spirit of the clouds and waters in Hindu and Buddhist mythology	W2F	2
apsaras	a female spirit of the clouds and waters in Hindu and Buddhist mythology	W2F	1
ardhmasendra asan	a yoga pose	S2A	1
ardhmasendra asan	a yoga pose	S2A	1
artha	one of the four aims of human life in Indian philosophy	W2A	1
arti	a Hindu religious ritual of worship	W2F	3
arya samajists	an Indian religious movement	W1A	1
asana	a place and posture in yoga	S2A	4
asana	a place and posture in yoga	W2D	1
asans	a yoga term	W2F	1
ashram	a spiritual hermitage or a monastery in Indian religions	W1B	3
ashrama	a spiritual hermitage or a monastery in Indian religions	W2F	2
ashramas	a spiritual hermitage or a monastery in Indian religions	S2B	1

ashwa sanchalak	a yoga pose	S2A	1
ashwa sanchalak	a yoga term	S2A	1
bhajan	any song with a religious theme	S2B	1
bhajan	any song with a religious theme	S2B	4
bhajan	any song with a religious theme	W2D	5
bhakti	Bhakti in Indian religions is "emotional devotionalism"	S2B	1
bharadwaja asana	a yoga pose	S2A	1
bhava	in Hinduism "becoming, being, existing, occurring, appearance"	W2D	1
bindi	a red dot worn on the centre of the forehead, commonly by Hindu and Jain women	S1A	1
bodhi	sacred trees	S2A	3
brahma nirvan	the state of release or liberation; the union with the Brahman in Hinduism	S2A	1
brahmo samajists	the societal component of Brahmoism	W1A	1
chakra	an energy point or node in the subtle body in Hinduism and Buddhism	S2B	1
chaturthi	the fourth day (Tithi) of any lunar month in the Hindu calendar	S1A	2
danda asana	a yoga pose	S2A	1
dasara	the tenth day of the festival, Vijayadashami	S1A	1
daseera	a Hindu festival symbolizing the triumph of good over evil	S1A	1
dasera	a Hindu festival symbolizing the triumph of good over evil	S1A	1
daserra	a Hindu festival symbolizing the triumph of good over evil	S1A	1
dassara	a Hindu festival symbolizing the triumph of good over evil	W2B	1
dassera	an important Hindu festival celebrated in a variety of ways in India and Nepal	S1A	6
deepawali	hindu festival of lights / a holiday	S1A	3
devi	a goddess	S1A	1
devi	a goddess	S1B	3
devi	a goddess	W1A	1
dharma	a key concept with multiple meanings in the Indian religions	S1A	1
dharma	a key concept with multiple meanings in the Indian religions	W2B	2
dipavali	hindu festival of lights / a holiday	S1A	3
dipawali	hindu festival of lights / a holiday	S1A	1
diwali	hindu festival of lights / a holiday	S1A	28
diwali	hindu festival of lights / a holiday	S1B	2
diwali	hindu festival of lights / a holiday	W1B	4
dukha	an important Buddhist concept, commonly translated as "suffering", "pain" or "unsatisfactoriness"	S2A	1
dussera	an important Hindu festival celebrated in a variety of ways in India and Nepal	S1A	1
grihastha	the second phase of an individual's life in a four age-based stages of the Hindu ashram system	W2B	1
gudhi	A pole wrapped with a cloth erected on the first day of the Shalivahan Hindu year	W2A	1
gudhyas	A pole wrapped with a cloth erected on the first day of the Shalivahan Hindu year	W2A	1
gudipadva	a hindu festival	S1A	1
guna	a key concept in various schools of Hindu philosophy	S2A	4
guna alankara	a guna	S2A	2
guna manasya	a guna	S1A	1
gunas	a key concept in various schools of Hindu philosophy	S2A	1
gurudwara	a place of worship for Sikhs	W2F	1
gurudwaras	a place of worship for Sikhs	S2B	1
hastamudras	a hand gesture used in yoga and Indian religions	S2B	1
havan	a ritual, wherein an oblation or any religious offering is made into fire	W2F	1

jathre	a religious day or a festival	W2B	1
kalpa	a relatively long period of time (by human calculation) in Hindu and Buddhist cosmology	W2F	1
kama	one of the four goals of Hindu life	W2A	1
kar seva	one of the main teachings of Sikhism	S2B	1
kar sevak	a volunteer for a religious cause	S1B	1
karma	the spiritual principle of cause and effect	W2B	1
karseva	one of the main teachings of Sikhism which means selfless service of others	S1B	5
karsevak	a volunteer for a religious cause	S1B	2
karuna	compassion that is a fundamental quality in the bodhisattva ideal of Mahayana Buddhism	S2B	1
karuna	compassion that is a fundamental quality in the bodhisattva ideal of Mahayana Buddhism	W2A	1
kshatriya	one of the four varna (social orders) of the Hindu society	S2A	1
kumkum tilaka	In Hinduism, the Tilaka (Hindi: तिलक) is a mark worn usually on the forehead	S1A	1
makar sankranti	a Hindu festival of harvest	W1B	1
makarsankranti	a Hindu festival of harvest	S1A	1
mandir	a hindu temple	S1A	2
mantras	a sacred utterance, a numinous sound, a syllable, word or phonemes, or group of words	S2A	1
masjid	a place of worship for muslims	S1A	1
moksa	one of the four aims of human life in Indian philosophy	W2A	1
mudra	a symbolic or ritual gesture in Hinduism and Buddhism	W2D	1
naats	poetry in praise of Allah and his Prophet Muhammad	W2F	1
nandipatha	an ancient Indian symbol	S2B	1
navaratri	a multi-day Hindu festival celebrated in the autumn every year	S1A	1
navmi	a Hindu festival	S1A	1
navratra	a multi-day Hindu festival celebrated in the autumn every year	S1A	1
nivritti	the mind's thoughts	W2B	1
niyoga	an ancient Hindu tradition, in which a woman would requests a person for helping her bear a child	S2B	3
nukta	a diacritic mark introduced in Devanāgari (and some other Indian scripts)	S1B	1
om	a sacred sound and a spiritual icon in Hindu religion	S1A	5
om	a sacred sound and a spiritual icon in Hindu religion	S2B	5
padva	Hindu lunar new year's Day	S1A	1
pardah	seclusion of women from public observation	S2B	1
parshwathana asana	a yoga pose	S2A	4
parshwauttana asana parshwa	a yoga pose	S2A	2
pawammukta asan	a yoga pose	S2A	1
pooja	a Hindu prayer ritual	S1A	17
pooja	a Hindu prayer ritual	W1B	1
prasadita padasthana asana	a yoga pose	S2A	1
prasadita padasthana asana	a yoga pose	S2A	1
prasavika padastha asana	a yoga pose	S2A	2
pratipada	the first day in the lunar fortnight (Paksha) of the Hindu calendar	S1A	1
pravritti	a Sanskrit meaning "different," and vrtti representing citta vritti, or the mind's thoughts	W2B	1
puja	a prayer ritual performed by Hindus to host, honour and worship one or more deities	W2E	1
pujas	the act of showing reverence to a god or a spirit	S1A	1
pujas	the act of showing reverence to a god or a spirit	W1B	1
purusarthas	a key concept in Hinduism, and refers to the four proper goals or aims of a human life	W2A	1

rajas	one of three "tendencies, qualities": sattva, rajas and tamas	W2B	2
rakhi	sacred thread worn on wrist	S2B	1
raksha bandhan	a Hindu religious and secular festival	S2B	5
rakshasa	a mythological being in Hindu mythology	S2B	3
rath yatra	a Hindu festival	S1B	1
rishi tarpan	in Hinduism celebrated on a full moon day in the month of August	S2B	1
rishi tarpan	an older name for shravani festival	S2B	1
sadachakra	a popular yantra or mandala (mystical diagram) used for worship in Jainism	S2B	1
sakti	the primordial cosmic energy in Hinduism	W2A	1
samadhi	a state of meditative consciousness	S1A	2
samadhi	a state of meditative consciousness	S2A	3
samadhi	a state of meditative consciousness	S2B	1
samadhi	a state of meditative consciousness	W1B	1
samadhi	a state of meditative consciousness	W2F	1
samsara	a fundamental assumption of all Indian religions	W2B	1
sanjeevan samadhi	a practice to sum up the life after entering into a deep meditative state	S1A	1
sanyasi	the life stage of renunciation within the Hindu philosophy	S1B	2
sanyasis	people at the stage of renunciation	S1A	1
sati	an obsolete Hindu funeral custom	S2A	1
sati	an obsolete Hindu funeral custom	S2B	2
sati	an obsolete Hindu funeral custom	W1A	5
sati devdasi	a girl's dedication to worship and service a deity or a temple for the rest of her life	S2B	1
satvaguna	one of the three Guṇas (tendencies, qualities, attributes)	S2B	1
seva	"selfless sacrifice"	W2B	3
seva	"selfless sacrifice"	W2D	1
shabad	hymn, a piece of Holy Text in Indian culture	W2F	1
shilanya	a sacred brick in Hinduism	S2B	1
shilanyas	an elaborate ritual during the initial phase of temple construction	S2B	2
shilanyas	an elaborate ritual during the initial phase of temple construction	W2E	2
shirsana	a headstand asana	S2A	2
shirshana	a headstand asana; yoga pose	S2A	1
shravani	holy month	S2B	1
skanda chalan	a yoga term	S2A	1
somvari	a day on which sea water is believed to turn holy	S1A	1
suptahastapadangosth asan	a yoga pose	S2A	1
surya namaskar	a Yoga warm up routine	S2A	1
swarg lok	Swarga Loka, is any of the eight loka or planes in Hindu cosmology	S2B	1
tada asana	a yoga pose	S2A	6
tamoguna	one of the three Gunas (tendencies, qualities, attributes)	S2B	1
tamoguna	one of the three Gunas (tendencies, qualities, attributes)	W2B	1
tapascharya	penance	S2B	1
tatva	in Kashmir Shaivism, the 36 tattvas are the 36 elements or principles of reality	S2A	2
tilak	a mark worn usually on the forehead in Hinduism	S2B	2
trikon asan	a yoga pose	S2A	1
tyaga	renunciation; gradual disociation of the mind from the seed of desire for worldly objects	W2B	4
vaisakha purnima	the birthday of Gautama Buddha amongst southern Buddhists	S2A	2

vakra asan	a yoga pose	S2A	1
vedanta	one of the six orthodox (āstika) schools of Indian philosophy	W2B	3
vichar	the Sanskrit and Pali term for a Buddhist monastery	S1A	1
vihars	the Sanskrit and Pali term for a Buddhist monastery	S1A	1
yagas shileniyas	some type of ceremonies	S2A	1
yoga	a group of physical, mental, and spiritual practices or disciplines	S1A	5
yoga	a group of physical, mental, and spiritual practices or disciplines	W2B	3
yogasan	a yoga term	S2A	2
			318

5) People

WORD	MEANING	COMPONENT	TOKENS
aaya	a nanny	S2A	1
acharya	a preceptor or instructor in religious matters; founder, or leader of a sect	S1A	1
acharya	a preceptor or instructor in religious matters; founder, or leader of a sect	S1B	5
acharya	a preceptor or instructor in religious matters; founder, or leader of a sect	S2B	4
acharya	a preceptor or instructor in religious matters; founder, or leader of a sect	W2A	1
acharya	a preceptor or instructor in religious matters; founder, or leader of a sect	W2C	1
adivasis	the tribal groups population of South Asia	S1B	3
ajja	grandfather	W1B	1
akka	sister	S1A	1
amma	One's mother (often used as a familiar form of address)	S1A	2
amma	One's mother (often used as a familiar form of address)	W1B	9
amma	One's mother (often used as a familiar form of address)	W2F	5
ammijan	"dearest mother"	W2F	4
amritdhari	a sikh title	W2F	1
angrez	a British person	W2F	1
asthana vidushi	an honorary title bestowed to a court musician or dancer in India	W2B	3
avidasis	the tribal groups population of South Asia	S2B	2
ayah	a nursemaid or nanny employed by Europeans in India	W2F	1
baba	guru/saint/teacher or father/grandfather	S1A	6
baba	guru/saint/teacher or father/grandfather	S1B	2
baba	guru/saint/teacher or father/grandfather	S2A	1
bai	a maid	S1A	1
bandobast	a group responsible for protecting a person, building, organization	S1A	1
beti	a daughter/woman	S1A	1
bhai	brother/friend	W2F	2
bhaisaab	borther	W2F	1
bhaiyya	a derogatory word used to denote hindi speakers of India	S1A	2
bhakt	a person who believes or faith on somebody he follows	S1A	1
bhakta	a religious devotee	S1A	1
brown sahibs	a term used to refer to natives of South Asia who imitate Western	S1A	1
chokras	a boy employed as a servant	W2F	1
dadas	elder brothers	S1A	1
dai	a traditional midwife in India	S1B	1
dai	a traditional midwife in India	W2F	4

dais	dai = a traditional midwife in India	W2F	2
desi	someone who is from the Indian subcontinent.	W2F	1
devadasis	a girl "dedicated" to worship and service of a deity or a temple for the rest of her life	W2B	3
dhiralalita	a person who is naturally very funny, always in full youthfulness, expert in joking	S2B	1
dhirodatta	a very grave, gentle, forgiving, merciful, determined, humble, and bodily attractive person	S2B	1
didi	An older sister or older female cousin	S1A	2
goonda	a hired thug	S1A	5
goondas	a hired thug	S1A	1
gundas	a hired thug	S2A	1
guru	a "teacher, guide, expert, or master" of certain knowledge or field	S1A	3
guru	a "teacher, guide, expert, or master" of certain knowledge or field	S1B	10
guru	a "teacher, guide, expert, or master" of certain knowledge or field	S2A	3
guru	a "teacher, guide, expert, or master" of certain knowledge or field	S2B	6
guru	a "teacher, guide, expert, or master" of certain knowledge or field	W1B	1
guru	a "teacher, guide, expert, or master" of certain knowledge or field	W2A	1
guru	a "teacher, guide, expert, or master" of certain knowledge or field	W2B	1
guru	a "teacher, guide, expert, or master" of certain knowledge or field	W2D	3
guru	a "teacher, guide, expert, or master" of certain knowledge or field	W2F	1
gurukulites	people who attend gurukula schools	W2D	1
gurus	a "teacher, guide, expert, or master" of certain knowledge or field	S1A	1
gurus	a "teacher, guide, expert, or master" of certain knowledge or field	S2B	2
gurus	a "teacher, guide, expert, or master" of certain knowledge or field	W2B	3
gurus	a "teacher, guide, expert, or master" of certain knowledge or field	W2F	2
jamedars	a title used for various military and other official in the Indian subcontinent	W2F	1
jawans	jawan = junior soldier (especially an infantryman) in South Asia	S2B	2
jawans	jawan = junior soldier (especially an infantryman) in South Asia	W2E	1
jawans	jawan = junior soldier (especially an infantryman) in South Asia	W2F	1
jee	respectable, honorable for a person (adjective)	S2A	1
jee saheb	a form of address for a man	S1B	5
jija	a term for girls who are elder than the speaker	S1A	1
kaka	a paternal uncle in Marathi	S1A	1
kavi	a poet	S2A	1
kawwalis	A group of qawwali musicians	S2B	1
khala	a mother's sister	S1A	1
kukis	people from a nomad tribe in India	S1A	2
kukis	people from a nomad tribe in India	S1A	5
mahasabhas	members of an orthodox Hindu nationalist movement	S1B	1
mahatma	Sanskrit for "Great Soul", similar in to Western term saint	S1A	1
mahatma	Sanskrit for "Great Soul", similar in to Western term saint	S1B	8
mahatma	Sanskrit for "Great Soul", similar in to Western term saint	S2B	1
mahatma	Sanskrit for "Great Soul", similar in to Western term saint	W1A	4
mahatma	Sanskrit for "Great Soul", similar in to Western term saint	W2A	1
mahatma	Sanskrit for "Great Soul", similar in to Western term saint	W2B	3
mahatma	Sanskrit for "Great Soul", similar in to Western term saint	W2F	1
mamaji	a maternal uncle	S1A	1
mummela	a dance and dialog group	S2B	1

nagas	an ethnic group of several tribes native to the North Eastern part of India	S1A	6
nagas	an ethnic group of several tribes native to the North Eastern part of India	S2B	1
nagas	an ethnic group of several tribes native to the North Eastern part of India	W2D	1
narad	a Vedic sage	S1B	1
nayaka	a lead male role in dance or a romantic hero in drama	S2B	1
needaya	some type of singer	W2D	1
pahadi	people belonging to hill tribes in India	S1B	1
pahadi	people belonging to hill tribes in India	S2B	1
pahadi	people belonging to hill tribes in India	W2D	2
pahari	people belonging to hill tribes in India	S1B	1
pahari	people belonging to hill tribes in India	S2B	1
panchas	a clan of the Bharwad people of India	S1B	2
panchas	a clan of the Bharwad people of India	S2A	5
pandit	a scholar or a teacher of any field of knowledge in Hinduism	S1B	4
pandit	a scholar or a teacher of any field of knowledge in Hinduism	S2A	4
pandit	a scholar or a teacher of any field of knowledge in Hinduism	S2B	10
pandit	a scholar or a teacher of any field of knowledge in Hinduism	W1B	5
pandit	a scholar or a teacher of any field of knowledge in Hinduism	W2D	10
pandit	a scholar or a teacher of any field of knowledge in Hinduism	W2E	1
pandit	a scholar or a teacher of any field of knowledge in Hinduism	W2F	1
parnivatkas	those who spent a lot of time and energy in performing Yagas	S2A	1
praudha	adult	S2A	1
puratchi	a leader of some type	S1B	1
puratchithalaivi	revolutionary leader	S1B	1
purohit	in the Indian religious context, means family priest	S2B	1
rajah	a word for king and meaning "monarch"	W2A	2
rickshawallah	a rickshaw driver	S1A	2
rishi	in Hinduism "seers" or "sages"	W2F	2
rishi	in Hinduism "seers" or "sages"	S2A	1
rishi	in Hinduism "seers" or "sages"	W2F	2
saab	"Saab" means sir and it is meant for respect	S1A	1
saab	"Saab" means sir and it is meant for respect	W2F	1
saab	"Saab" means sir and it is meant for respect	S2A	1
saheb	a form of address for a man	S1B	2
saheb	a form of address for a man	W2F	1
sakhi	a friend	W2D	2
sardarji	a leader	S1A	2
sardarji	a leader	S1B	1
sardarji	a leader	W2F	2
sareewalah	a person selling sarees	S1A	1
sarpanch	an elected head of a village-level statutory institution of local self-government	S2B	1
sepoy	an Indian soldier	W2F	1
sharnarthis	a refugee	S1B	1
Shehenaiwallah	a type of wallah	S2A	1
shisyas	a pupil of a guru	S2A	1
shree	an honorific commonly used in India	S1B	2

shree	an honorific commonly used in India	S2A	8
shree	an honorific commonly used in India	W1B	3
shri	an honorific commonly used in India	S1B	34
shri	an honorific commonly used in India	S2A	28
shri	an honorific commonly used in India	S2B	11
shri	an honorific commonly used in India	W1A	1
shri	an honorific commonly used in India	W1B	50
shri	an honorific commonly used in India	W2F	6
shrimati	an honorific commonly used in India (for women)	S1B	1
shrimati	an honorific commonly used in India (for women)	S2A	9
sootradhara	the stage manager and director of Sanskrit drama	W2B	1
sutradhara	the stage manager and director of Sanskrit drama	W2D	1
swami	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	S1A	1
swami	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	S1B	3
swami	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	S2A	1
swami	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	S2B	2
swami	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	W1B	5
swami	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	W2A	1
swami	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	W2B	4
swamiji	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	S1A	1
swamiji	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	W1B	1
swamiji	(1)a religious teacher, (2) a mystic; a yogi, (3) a form of address for such a person	W2F	4
tamkuls	people belonging to hill tribes in India	S1A	5
Vimukhta Jati	the tribes listed under the Criminal Tribes Act of 1871	S1A	1
wallah	person in charge of or employed at a particular thing	S1A	3
zamindar	a landowner, especially one who leases his land to tenant farmers	S1A	2
			454

6) Politics

WORD	MEANING	COMPONENT	TOKENS
bandh	a form of protest used by political activists in South Asian countries	S1A	2
bandh	a form of protest used by political activists in South Asian countries	W1B	1
bandh	a form of protest used by political activists in South Asian countries	W2C	2
begum	a female royal an aristrocratic title	S2B	2
begum	a female royal an aristrocratic title	W2F	5
brahmin	is a caste group of priests	S1A	3
brahmin	is a caste group of priests	S1B	2
brahmin	is a caste group of priests	W2F	4
brahmins	is a caste group of priests	S1A	3
brahmins	is a caste group of priests	S2B	2
brahmins	is a caste group of priests	W2B	4
brahmins	is a caste group of priests	W2F	1
bundh	A general strike, shutdown, or other form of protest	W1B	1
chatrapati	an Indian royal title equivalent to an King used by the Marathas	W1B	1
chhatrapati	Indian royal title used by the Marathas	S2B	1
chhatrapati	Indian royal title used by the Marathas	W1B	5

chhatrapati	Indian royal title used by the Marathas	W2A	7
dalits	the self-chosen name of castes in India which are "untouchable"	W2B	2
dalits	the self-chosen name of castes in India which are "untouchable"	W2C	1
dharana	a non-violent sit-in protest	S2B	2
dharna	a non-violent sit-in protest	S1B	2
dharna	a non-violent sit-in protest	W2C	2
gheraoed	A protest in which a group of people surrounds a politician, building, etc.	W2C	1
harijan	an individual who is born at the bottom of or outside the Hindu caste system	S1B	1
harijan	an individual who is born at the bottom of or outside the Hindu caste system	W2D	1
harijans	individuals who are at the bottom of or outside the Hindu caste system	S1B	2
harijans	individuals who are at the bottom of or outside the Hindu caste system	S2B	1
kendriya vidyalaya	a system of central government schools in India	S1A	1
khstriya	one of the four varna (social orders) of the Hindu society	S2B	1
Lok Sabha	the Lower house of India's bicameral Parliament	S1B	9
Lok Sabha	the Lower house of India's bicameral Parliament	S2A	1
Lok Sabha	the Lower house of India's bicameral Parliament	S2B	21
Lok Sabha	the Lower house of India's bicameral Parliament	W2C	4
Lok Sabha	the Lower house of India's bicameral Parliament	W2E	1
maharaj	a title for a ruler or a king	S1B	1
maharaj	a title for a ruler or a king	S2B	2
maharaj	a title for a ruler or a king	W2A	8
maharaja	a title for a ruler or a king	S1A	3
maharaja	a title for a ruler or a king	S1B	2
maharaja	a title for a ruler or a king	S2A	1
maharaja	a title for a ruler or a king	S2B	3
maharaja	a title for a ruler or a king	W2A	4
maharajas	a title for a ruler or a king	W2B	2
majaraja	a Sanskrit title for a "great ruler", "great king" or "high king"	W2A	1
panchayat	the oldest system of local government in the Indian subcontinent	W2C	2
panchayat	the oldest system of local government in the Indian subcontinent	W2E	1
panchayati raj	a South Asian political system	S1B	1
panchayati raj	a South Asian political system	S2B	1
panchayati raj	a South Asian political system	W2E	1
panchayats	a South Asian political system	W2E	1
panchnama	a member of the lowest caste in India	S1B	9
panchnama	a member of the lowest caste in India	S2A	5
parishad	(in India) an assembly	W1B	1
parishats	(in India) an assembly	W2B	1
patti	a traditional caste of Herdsmen from Sri Lanka's feudal past	S2A	1
pradhans	a ministerial title used in regions of Hindu cultural tradition	S2B	1
raja	a title for a monarch or a princely ruler in Southt Asia	W2F	5
rajasaheb	a title for a ruler or a king	W2A	1
rajmata	"queen mother" a title	S1A	2
Rajya Sabha	the Upper house of India's bicameral Parliament	S1B	1
Rajya Sabha	the Upper house of India's bicameral Parliament	S2B	20
Rajya Sabha	the Upper house of India's bicameral Parliament	W2E	4

ranisaheb	a title	W2A	2
Rashtrapatiji	a president	S2A	1
samiti	A committee, society, or association	S1B	1
samiti	A committee, society, or association	W1B	1
samiti	A committee, society, or association	W2C	3
sammelan	a summit/a conference	S2B	1
sammelan	a summit/a conference	W2D	1
shimpi	a person belonging to Meru Kshatriya Shimpi caste	S1A	1
shrimant	title used by Maratha royals and nobles	W2A	1
shudra	the fourth and lowest of the traditional varnas, or social classes, of India	W1A	1
tahsil	an administrative divisions of India denoting a sub-district	S1A	1
taluk	an administrative divisions of India denoting a sub-district	S2A	1
taluk	an administrative divisions of India denoting a sub-district	W2B	1
taluka	an area of land with a city or town that serves as its administrative centre	S1A	2
taluka	an area of land with a city or town that serves as its administrative centre	W1A	1
taluka	an area of land with a city or town that serves as its administrative centre	W2E	1
taluks	an administrative divisions of India denoting a sub-district	W2A	2
zilla parishad	the third tier of the Panchayati Raj system	S2B	1
zilla parishad	the third tier of the Panchayati Raj system	W2C	1
zilla parishads	District Council	S2B	1
morcha	An organized march or rally	S1B	2
morcha	An organized march or rally	W2C	1
satyagraha	a particular form of nonviolent resistance or civil resistance	S1B	1
			213

7) Discourse markers

WORD	MEANING	COMPONENT	TOKENS
accha		S1A	202
accha		S1B	40
accha		S2A	1
achha		S1A	4
ahn		S1A	239
ahn		S1B	17
ahn		S1B	11
aiyyo		S1A	2
ayyah		S1A	1
ayyo		S1A	1
haan		S1A	367
haan		S1B	65
haan		S2A	17
ki		S1A	8
ki		S1B	5
ki		S2A	1
lekin		S1A	1
lekin		S1A	1
lekin		S2A	1

nahi		S1A	1
nahi		S1A	18
nahi		S1B	5
nahi		S2A	2
nahi nahi		S1A	19
nahi nahi		S1B	2
nahi nahi nahi		S1B	1
nahi nahi nahi		S1A	1
nai		S1A	3
nai		S1B	1
nai nai		S1A	3
nai nai nai		S1B	1
nei		S1A	28
wah		S1A	2
yaar		S1A	127
yaar		W2F	1
yar		S1A	2
			1201

8) Animals and nature

WORD	MEANING	COMPONENT	TOKENS
ashoka	a type of tree	S2A	1
ashwattha	a type of tree	W2B	1
azgars	an Indian rock python	S1A	1
banyan	a type of tree	W2F	5
bodhi vriksha	a type of tree	S2A	1
bulbul	a passerine songbird	S2B	1
bulbuls	a passerine songbird	W2F	1
champaka	a large evergreen tree	W2B	1
chatak	Jacobin cuckoo, pied cuckoo, or pied crested cuckoo	S2B	1
chikoo	a long-lived, evergreen tree	W1A	1
hilsa	a type of fish	S1A	2
jawar	sorghum	S1A	2
jawar	sorghum	W1A	1
jawar bajri	sorghum	S1A	1
jawar bajri	a type of millet	S1A	1
kaudis	seashells	S2A	1
krauca	a bird "crane", "demoiselle crane"	W2A	1
kraunch	a bird "crane", "demoiselle crane"	S2B	1
neem	Indian Lilac is a tree in the mahogany family	W1A	1
pallam	a low lying area	W2A	1
peepal	a type of tree	S2A	1
peepal	a type of tree	W2B	1
peepul	an Indian moraceous tree	W2F	3
pipiha	a colloquial term for a type of bird	W2F	1
sarkanda	elephant grass	W2F	2

tamla	a type of tree	W2D	1
tendu	a species of flowering tree	S2B	1
thungundram	slope of a hill	W2A	1
vriksha	a type of tree	S2A	1
			38

9) Culture

WORD	MEANING	COMPONENT	TOKENS
akshata	Rice Sprinkled During Hindu Puja and Weddings	S1A	1
anganwadis	"courtyard shelter" for homeless children	W1B	1
asaab	a greeting often associated with the courtly culture of the Muslim Nawabs	W2F	1
ashtapada	an Indian board game which predates chess	W2B	1
ayurveda	a system of medicine with historical roots in the Indian subcontinent	S1B	1
ayurveda	a system of medicine with historical roots in the Indian subcontinent	S2B	2
bada din	Christmas	S1B	1
balwadi	an Indian pre-school run for economically weaker sections of the society	S1A	1
barse	the naming ceremony for a newborn	W2A	1
bazar	a permanently enclosed marketplace or street	S1A	2
bazar	a permanently enclosed marketplace or street	S2A	1
bazar	a permanently enclosed marketplace or street	W2F	1
bissi	a slang term for a 20-paisa coin, which is worth 1/5 of one Indian rupee	S1A	4
bissis	a slang term for a 20-paisa coin, which is worth 1/5 of one Indian rupee	S1A	2
challa	a game of some sort	S2A	3
dosha	one of three bodily Bioelements that make up one's constitution	S2A	1
doshas	your unique physical and mental constitution, which influence your personal well-being	S2A	2
goddu mandu	ancient medicine for sterility	W2C	1
gurkula	a type of residential school in India with pupils living near the guru	W2D	1
gurukulavasam	a term relating to gurkula	W2D	1
gurunkula	a type of residential school in India	S2B	1
gurunkula	a type of residential school in India	W2D	2
haldi kunku	a social gathering in India	S1A	1
havala	a popular and informal value transfer system	W2C	2
haveli	generic term used for a traditional townhouse and mansions	S1A	1
hawala	a popular and informal value transfer system	S1B	5
hawala	a popular and informal value transfer system	W1A	4
janapadas	the realms, republics and kingdoms of the Indian Vedic period	S2B	1
kaata	prearranged form of exercise in any martial arts	S2A	5
kabaddi	a contact team sport that originated in ancient India	W1B	1
kabaddi	a contact team sport that originated in ancient India	W2E	1
maidan	an open area or space often used as a marketplace	S1B	6
mangalsutra	a necklace that a Indian groom ties around the bride's neck in the ceremony	S1A	3
namaskar	is a respectful form of greeting in Hindu custom	S1A	2
namaskar	is a respectful form of greeting in Hindu custom	S2A	4
namaskaras	is a respectful form of greeting in Hindu custom	W1B	1
namaskars	is a respectful form of greeting in Hindu custom	W1B	1
namaste	is a respectful form of greeting in Hindu custom	W2F	1

nayapaisa	Indian coin, was a unit of currency equaling 1/100) of the Indian rupee	S2A	1
nishka	an ancient Indian coin	S2B	1
pada kachari	a courthouse, a hall, an office of administration, a place of business transactions	S2B	1
paisa	a rupee is divided into 100 paise (sing. Paisa)	S1A	1
paisa	a rupee is divided into 100 paise (sing. Paisa)	S1B	7
paise	a rupee is divided into 100 paise (sing. Paisa)	S2B	4
paise	a rupee is divided into 100 paise (sing. Paisa)	W1A	1
paise	a rupee is divided into 100 paise (sing. Paisa)	W2B	1
paise	a rupee is divided into 100 paise (sing. Paisa)	W2D	2
paise	a rupee is divided into 100 paise (sing. Paisa)	W2F	4
pana	a punch-marked coin in the early coinage of India	S2B	2
parda nasheen	covered in hijab or any other clothes	S2A	1
purdah	the practice of screening women from men or strangers	S1A	1
purdah	the practice of screening women from men or strangers	W2F	1
raksha sutra	a bracelet that protects from the evil	S2B	5
Rashtrabhasa	national language	S1A	1
rupee	a common name for currencies in South and Southeast Asia	S1B	12
rupee	a common name for currencies in South and Southeast Asia	S2B	4
rupee	a common name for currencies in South and Southeast Asia	W1A	1
rupee	a common name for currencies in South and Southeast Asia	W2A	1
rupee	a common name for currencies in South and Southeast Asia	W2B	2
rupee	a common name for currencies in South and Southeast Asia	W2C	5
rupee	a common name for currencies in South and Southeast Asia	W2E	1
rupee	a common name for currencies in South and Southeast Asia	W2F	4
rupees	Indian currency	S1A	52
rupees	Indian currency	S1B	45
rupees	Indian currency	S2A	51
rupees	Indian currency	W1B	8
rupees	Indian currency	W2A	2
rupees	Indian currency	W2C	3
rupees	Indian currency	W2D	2
rupees	Indian currency	W2F	5
satamana	a coin in ancient India	S2B	1
satmana	a coin in ancient India	S2B	1
sattva	one of the three Guṇas (tendencies, qualities, attributes)	W2B	2
satva	one of the three Guṇas (tendencies, qualities, attributes)	S2A	1
sutra	the string of the raksha sutra bracelet	S2B	1
swabhava	"own-being" or "own-becoming", the essential nature or essence of living beings	S2A	1
unani	the Perso-Arabic traditional medicine as practiced in Mughal India	S1B	1
swayamvaras	a practice of choosing a husband by a girl of marriageable age	S1A	1
swayamwara	a practice of choosing a husband by a girl of marriageable age	S1A	1
tilak	one of the initial wedding ceremonies in India	S2B	1
			314

10) Others

WORD	MEANING	COMPONENT	TOKENS

aashirwad	a blessing	S2A	1
abhidana	vocabulary	W2A	1
abhipraya	intention	S2A	1
adeshraj	when something comes from above	S1A	1
afeem	opium	W2F	1
ahe	"is"	S1B	3
ahe	"is"	S1A	2
akhada	training hall for professional fighters	W2D	1
almirahs	A free-standing cupboard or wardrobe	W2F	1
angeethi	a traditional brazier used for space-heating and cooking	S1A	1
ankusha	a tool employed in the handling and training of elephants	S2B	1
anumiti	consent	S2A	2
apna	oneself,own,mine,one's,my/our,and Indian,our guy,very	S1A	1
aram se	"easy, take your own time, do it slowly, softly, be cool...etc."	S1A	3
arre	hey	S1A	2
auchitya		S2A	5
bahut acchi	"pretty good"	S2A	1
balavastha	infancy	S2A	1
bandobast	protection of a person, building, orrorganization against crime or attack	W2C	1
benami	any transaction in which property is transferred to one person paid by another person	W2E	1
bhava	"emotion, sentiment, state of body or mind, disposition"	S2B	2
bhava	"emotion, sentiment, state of body or mind, disposition"	W2A	2
bidi	a thin, Indian cigarette filled with tobacco flake and wrapped in a tendu	W1B	1
bighas	a traditional unit of measurement of area of a land	S1B	1
chaadar	a sheet/veil/waterfall	S2A	1
chakki	a grainmill	S1A	1
chalega	a question or a statement as to whether something will work	S1A	2
chalegaa	a question or a statement as to whether something will work	W2A	1
challan	An official form or document, such as a receipt, invoice, or summons	S1B	2
challaned	fine, penalty, tax, levy	S2B	1
chalo	"let's go!"	S1A	1
chalta hai	"it's ok" attitude	S2A	1
chandas	prosody	W2A	1
charkha	a type of Indian spinning wheel	S2A	5
chawl	a name for a type of residential building found in India	S1B	2
chawl	a name for a type of residential building found in India	W2C	2
chhota	small/little	S1A	1
chowki	a police station	S2A	4
crores	ten million	S2A	3
crores	ten million	S1B	43
crores	ten million	S2A	1
crores	ten million	S2B	7
crores	ten million	W1A	7
crores	ten million	W2A	4
crores	ten million	W2B	1
crores	ten million	W2C	13

crores	ten million	W2D	1
crores	ten million	W2E	10
dawayee	drug/medicine	S2B	1
dekchi	a cooking pot	W2D	2
dekhye	look, see, watch	S1B	1
desi	something that is from the Indian subcontinent	S1A	1
desi	something that is from the Indian subcontinent	S1B	1
desi	something that is from the Indian subcontinent	W2A	1
dhairya	patience	S2A	3
durlabh	rare/extinct	S2A	1
ek	one	S1A	1
ek	one	W2A	2
ekdam	completely, totally	S1A	1
ekka	A small vehicle with two wheels that is pulled by a horse	S1A	1
faltu	useless, idle	S1A	1
gadbad	confusion, trouble	S1A	1
ghar	a house/home/residence	W2C	1
ghat	a series of steps leading down to a body of water	W2A	1
gobar	bio	S1B	1
godowns	godown = a warehouse	S2B	1
godowns	godown = a warehouse	W1A	3
godowns	godown = a warehouse	W2C	1
gupti	a traditional dagger from India	W2C	1
guptis	a traditional dagger from India	W2C	1
hafta	slang for the protection money collected by gangsters and corrupt policemen	S1B	2
hai	can mean different things	S1A	2
hasya	happy/laughter/joy	S2B	1
hasya	happy/laughter/joy	W2A	1
hota hai	"happens" "is"	S1B	1
hundi	a type of cooking pot or pan	S2A	1
jaisa	as/such as	S1A	1
kada	harsh	S1B	1
kadai	an iron cooking pan	S1A	3
kadhai	an iron cooking pan	S2A	2
kadhai	an iron cooking pan	W2B	1
kalakriti	"work of art"	S2A	1
karke	do or get done	S1A	1
karman	action	W2A	1
khaki	soil-coloured	S2A	1
khamma gani	a form of greeting in Marwari, similar to 'hello',	S1A	1
kharara	some kind of tool, a brush (?)	W2F	1
kolsa	coal	S1B	1
kuch kuch	"very little"	S1A	1
kula	race	S2A	1
kya	"what?"	S1A	2
lakh	one hundred thousand	S1A	1

lakh	one hundred thousand	S1B	10
lakh	one hundred thousand	S2A	4
lakh	one hundred thousand	S2B	26
lakh	one hundred thousand	W1A	1
lakh	one hundred thousand	W2A	3
lakh	one hundred thousand	W2C	16
lakh	one hundred thousand	W2E	2
lakhs	one hundred thousand	S1A	11
lakhs	one hundred thousand	S1B	17
lakhs	one hundred thousand	S2A	2
lakhs	one hundred thousand	S2B	5
lakhs	one hundred thousand	W1A	1
lakhs	one hundred thousand	W2A	3
lakhs	one hundred thousand	W2C	8
lakhs	one hundred thousand	W2F	1
lath	a heavy stick often used in India as a weapon especially by police	W2D	1
lathi	a heavy stick often used in India as a weapon especially by police	S2A	1
lathi	a heavy stick often used in India as a weapon especially by police	W2B	1
lathi	a heavy stick often used in India as a weapon especially by police	W2C	1
lathis	a police tactic used to disperse crowds	S2B	1
loktantra	democracy	S2A	2
mali	garden?	W2F	1
matlab	"meaning"	S1A	2
mela	a fair	S1B	2
mohalla	a country subdivision or neighbourhood	W2F	1
nazakat	"neatness, elegance, politeness"	S1A	1
nipunata	proficiency	W2A	1
nirjeev	"inanimate, lifeless"	S2A	1
nivas	a place of residence; a house, block of flats, etc.	W1B	1
pada	a leg	S2A	1
pakka	(1) matured , (2) confirmed (1) finalised etc	S1A	1
panch	number five	S2A	4
panchanama	a very important document, a law document (?)	S2A	16
pankha	a type of fan used since the early 500 B.C	S1A	1
prameya	knowledge	S2A	2
pucca	refers to dwellings that are designed to be solid and permanent	S1A	1
pucca	refers to dwellings that are designed to be solid and permanent	S1B	1
rampuri	an Indian gravity knife of formidable reputation	S2A	3
sandarbha	context	S2A	2
sastra	"precept, rules, manual, compendium, book or treatise"	W2A	1
shastras	"precept, rules, manual, compendium, book or treatise"	S2B	2
shegadi	a pan of live-coals or embers	W2A	2
siddhant	Indian slang term meaning awesome	S2A	1
sthavarajangama	Fixed and moving	W2A	1
tali	a round platter used to serve food	S1A	1
tambaku	tobacco	S1A	1

tawa	a cooking implement used in South Asia	S2A	1
teepoy	an item of furniture	S1B	1
thali	an Indian style platter	W2D	1
thana	to stretch	S2A	1
thoda	little	S1A	2
thoda thoda	very little	S1B	1
trami	a huge plate	W2B	1
uttana	extension	S2A	1
verandah	A veranda or verandah is a porch or long, open room with a roof over it	S1A	1
verandah	A veranda or verandah is a porch or long, open room with a roof over it	S2A	2
verandah	A veranda or verandah is a porch or long, open room with a roof over it	W2F	3
vrata	"vow, resolve, devotion"	S2A	3

394