

# The *Ausbau* Issue in the Dravidian Languages: the Case of Tamil and the Problem of Purism.

Harold Schiffman  
South Asia Studies  
University of Pennsylvania

**Introduction:** Tamil is a Dravidian language that began to try to develop itself, i.e. undergo *Ausbau*, when the perception developed that there was a threat from other languages to its very existence.<sup>1</sup> But Tamil linguistic culture's *Ausbau* project has had limited success, I would claim, principally because of an excessive concern with and devotion to **purism**. That is, rather than apply a number of different strategies for vocabulary (i.e. 'word-formation') development that are available to all languages, such as abbreviation, blending, borrowing and nativization, Tamil linguistic culture has opted primarily for the strategy of loan translation, using 'native' roots (or what are thought to be 'pure' native roots) or neologizing *ex nihilo*, making up new terms but using only what are thought to be 'pure' native resources. To understand this, we must look back in the history of the language to a number of moments when Tamil (and the other Dravidian languages) were confronted with linguistic 'threats' of various sorts, i.e. the influx of loan words from Sanskrit, and later pressures from both English during the British colonial period, and from Hindi, after Independence<sup>2</sup>. Finally, the fact that higher education has always been in English in India, and continues to resist any switch to other languages, has severely hampered the development of scientific and technological registers in *any* South Asian language.

**Review of Literature:** Purism, especially the movement for Tamil purism, has been extensively studied by Annamalai (1979)<sup>3</sup>. The notion of being able to influence and 'develop' the corpus of a language was pioneered by Haugen (1966). The issue of diglossia, which figures strongly in any discussion of Tamil *Ausbau* relies on the pioneering work of Ferguson (1959) and the extension of diglossia to bilingual situations by Fishman (1967 [1981]); the topic as applied to Tamil has extensively been covered by Britto (1986). For a discussion of register and repertoire, see Schiffman 1996. The terms *Ausbau* and *Abstand*, mainly as a diagnostic for determining whether a speech form is a 'language' or a 'dialect' were first proposed by Kloss (1967). For background on the issue of language of science, and the problems that have plagued Tamil as it has attempted to modernize and claim registers dominated by English, see Andronov 1975. Reversing language shift has been extensively dealt with by Fishman (1991).

---

<sup>1</sup> The perception was that Sanskrit, and later Hindi, were a threat to the continued existence of the language, such that if steps were not taken, Tamil would become completely *corrupted* by these languages, if not annihilated by them.

<sup>2</sup> Much of what I am claiming here has been presented in other research (Schiffman 1987, the chapter on Tamil linguistic culture in Schiffman 1996, Schiffman 1997, 1999, and 2003), and material from a forthcoming article on language policy implementation in Singapore.

<sup>3</sup> For an extensive review of Annamalai 1979 and others see Schiffman 1996.

**What is *Sprachausbau*?** The notion of *Ausbau* as a term in the sociology of language was proposed by Heinz Kloss (Kloss 1967), and, when used in conjunction with another German term, *Abstand*, are applied together to use as a diagnostic in deciding whether some speech form is a language or perhaps ‘only’ a dialect. *Abstand* languages are languages by linguistic ‘distance’, such that even if the language in question is not ‘developed’ it is still a separate language by virtue of its structure and vocabulary being divergent enough from other languages that it cannot be understood by speakers of them, and thus subsumed under another language. A classical example of this would be Icelandic, which has only a few hundred thousand speakers, but is isolated enough from its Scandinavian sister languages that it is not mutually intelligible with them<sup>4</sup>. The major Dravidian ‘literary’ languages are all *Abstand* languages, and even some of the minor ones, such as Tuḷu and Brāhui, though not ‘developed’ enough to be used for schooling and other purposes, are still not mutually intelligible with any other language, Dravidian or otherwise<sup>5</sup>.

*Ausbau* languages, on the other hand, are languages that have been ‘outfitted’ or developed<sup>6</sup> so that they have the lexical resources (vocabulary and technical terminologies) that enable them to be used for science and technology. *Ausbau* languages can be close (i.e. lacking in *Abstand*) enough to be mutually intelligible, and in some cases may even share similar vocabularies; Czech and Slovak, Norwegian and Swedish, Serbian and Croatian, and a number of other similar pairs exist in the world; they have been ‘*Ausbau*-ed’ as independent languages because their speakers resisted the notion that they were mere ‘dialects’ of each other. Over time, of course, languages that are genetically close to each other can choose a path that leads to more rather than less ‘*Abstand*’, so that eventually, the two may differ in their literary or other registers, while still being mutually intelligible at a day-to-day level<sup>7</sup>. Thus Serbian and Croatian, Hindi and Urdu etc. can be easily understood by their respective populations at a simple (spoken) level, but vocabulary for ‘higher’ registers is borrowed or developed separately, to make these registers less mutually unintelligible. Hindi borrows from Sanskrit, while Urdu borrows from Perso-Arabic sources, rendering the higher registers mutually unintelligible<sup>8</sup>. After the collapse of the former Yugoslavia, Croatian, Serbian and Bosnian lects have gone separate ways in order to assert their linguistic and political

---

<sup>4</sup> Other obvious linguistic isolates are Basque, Burushaski, and languages spoken on islands far from other languages.

<sup>5</sup> Though Tuḷu has 1.5 million speakers and possesses a written grammar and dictionary, Tuḷu speakers have always accepted the use of Kannada for education; when Tuḷu is written, it is written with Kannada script.

<sup>6</sup> A concrete use of the term *Ausbau* in German is applied to the task of finishing-off or outfitting a space in a building that is unfinished—an attic storeroom or a basement can be *ausgebaut*, i.e. ‘refurbished’ by finishing the walls, floors and ceiling in a manner similar to other rooms in the house, and equipping this new ‘room’ with appliances, furniture, etc. that make it useable as a regular ‘room’.

<sup>7</sup> The Scandinavian languages have famously cooperated through the *Nordisk Sprakrad* (Nordic Language Council) to make sure that new terminologies are shared in the major languages, rather than developed independently.

<sup>8</sup> But at the conversational level, speakers of Hindi and Urdu cannot say for sure which ‘language’ others are speaking. One may in fact know Hindi, the other Urdu, but there is no way to distinguish them at this level.

independence, whereas in the 19<sup>th</sup> and 20<sup>th</sup> centuries, attempts were made to bring them closer together, i.e. lessen the *Abstand*.<sup>9</sup>

If a lect has neither *Ausbau* nor *Abstand* features, it is probable that it will be relegated to the status of being a dialect of some other language, and will get subsumed under the umbrella of a neighboring language which has been *Ausbau*-ed. Thus *Lëtzebuergesch*, a Mosel-Fränkisch Germanic dialect spoken in Luxembourg (and in neighboring parts of Germany), though its speakers insist on their own linguistic identity, can not be used for higher education in that country because it has not undergone *Ausbau*, so standard German is employed from about third grade onward (Hofmann 1981, Schiffman 1992). In many places in the world we find examples of struggles by various lects to assert their linguistic independence from other neighboring, more developed ‘languages’ and the strategy often used is to increase the *Abstand* while also working on the *Ausbau*. In the Dravidian area, Malayalam has struggled for almost a millennium to distance itself from Tamil, of which it was originally considered a dialect; one of the strategies used was to borrow more extensively from Sanskrit, which Tamil has eschewed. Another way to increase the *Abstand* (at least of the written form) is to develop a different writing system, a strategy we see in force with Malayalam (vs. Tamil), Serbian vs. Croatian, Thai vs. Lao, and Hindi vs. Urdu.

Tamil got its start in the *Ausbau* project when the missionary grammarians such as Caldwell (1856) helped establish the notion that Tamil and the Dravidian languages were genetically separate from Sanskrit. But it was Tamil ‘orientalists’ such as U. Ve. Cuvaminataiyar who ‘rediscovered’ ancient Tamil literature and helped make it available to a wider audience. Before 1881, when U. Ve. Cu. first laid eyes on manuscripts (written in a kind of Tamil he had never seen before), few in the Tamil tradition knew of any pre-Aryan Tamil culture or literature. The subsequent development of the Pure Tamil Movement (*tanit-tamiR iyakkam*), the renaissance of Tamil literature, and the development of a Dravidian political movement (Justice Party, DK, DMK) must be seen both as stimulated by and reactions against the rise of English education and the work of ‘orientalist’ missionary-grammarians. As the Dravidian languages began to assert themselves as independent of Sanskrit, it became clear that the main literary languages (and even some of the tribal languages), were clearly already separate languages by *Abstand*, and did not constitute a chain of mutually-intelligible dialects such as we find in the Indo-Aryan area.<sup>10</sup>

**Corpus Management:** A concern for developing a language so as to allow it to function in many if not all linguistic domains falls under what sociologists of language call ‘corpus management’, that is, the care, feeding, and grooming of the body of the language—its vocabulary, grammar, and literature. As was the case for most of the literary languages in the subcontinent, registers were already in existence for poetry and

---

<sup>9</sup> It may be noted that in the ethnic cleansing that went on in Yugoslavia during the 1990’s, Croats, Serbs and Bosnians were identified by their adversaries, not by any linguistic features, but by their given names, or by denunciation.

<sup>10</sup> The Indo-Aryan dialect continuum stretches across the subcontinent from Pakistan to Bangladesh, with no clear break between one dialect and another (Shapiro and Schiffman, 1982).

writing about religion, music, and dance, but registers for science and technology were sorely lacking. The *Ausbau* project for Tamil, then, was to develop vocabulary for science, technology, and other ‘modern’ registers. Even though higher education was the almost exclusive domain of English, many language groups thought that English could be ejected from these domains after independence was won .

Because there has long been a perception that Tamil is ‘threatened’ by other languages, and therefore needs to develop its *own* linguistic resources, there has resulted an excessive attention to linguistic purism as a strategy for corpus management. Tamils hold that the reason for the decline and fall of Tamil in the past was the invasion of Sanskrit vocabulary, but now the problem is *angilak kalappu*, or mixing English words in with pure Tamil; if foreign loan words could be eliminated, and Tamil equivalents for them could be developed, Tamil could then ‘hold its head up’ and be known as a ‘real’ language, equal to other world languages. Corpus management and *Ausbau*, then, are rarely undertaken in a politically neutral situation, but are more likely to happen when the perception exists that the existence of a language or its people is ‘threatened’ by some other language, or when social conditions are in a state of flux (Annamalai 1979). The issue then becomes emotional; someone may fan the flame of nationalistic or ethnic fervor, and purism may be resorted to as a way to thwart the linguistic ‘corruption’ that will not only ‘ruin the language’ but undermine the very existence of the culture and people themselves. Corpus management then may be couched in terms of being a matter of life and death, and rational language planning may be consigned to the waste basket. But as Fishman notes, it is status planning, not corpus planning, that is the engine of all language planning success. (Fishman 1991: 349) If the corpus planners of Language X decide to exaggerate the dangers posed by putative enemies, or create the perception that some other language group is planning the extermination of Xish, this may rally the troops internally, but it makes status planning all the more difficult. An all-or-nothing stance may result in nothing, rather than ‘all.’ Linguistic purism unfortunately blames corrupting elements from outside as the reason why Xish is not succeeding, rather than making compromises of various sorts, including compromises on strategies of word-formation. As Andronov has pointed out, this ought also to have involved looking at the resources of the modern *spoken* language, rather than attempting to revive classical Tamil.<sup>11</sup>

“All efforts of Tamil linguists were not directed at studying modern Tamil, at determining its norms and preparing it for new functions in society, but at artificially galvanizing the dead literary language.” (Andronov 1975:188). Andronov goes on to point out that, as far as developing scientific and technological vocabulary, that “translating into or writing in Tamil original purely scientific works on narrowly specialized complicated problems in contemporary technology, [...] encounter almost unsurmountable difficulties. (op.cit. 188) such that introducing Tamil as the medium of instruction in higher education can only succeed if terminology is borrowed from languages that already possess it, alongside the creation of indigenous loan-translations. Purism, Andronov notes, in eschewing loanwords entirely, can only lead to “an unjustified impoverishment of the language.” (op. cit. 189). Given the abundance of terminology already extant in just one field of inquiry, which may have a vocabulary of

---

<sup>11</sup> It should be noted that Bengali and Telugu, at least, have abandoned the erstwhile H-varieties of their languages in favor of focusing on developing the L-varieties and extending their domains

over 100,000 words, the impossibility of translating millions of terms from other growing fields is obvious.

**Care and Feeding, but of which Variety?** In general in language maintenance movements, it is difficult to convince people that language maintenance for ‘threatened languages’ cannot be restricted to the literary or educated variety, but must also be geared to efforts on behalf of the *spoken* language, the language of ‘hearth and home’.<sup>12</sup> Many linguistic changes that have occurred in the spoken language (ST) since its 13<sup>th</sup> century codification have not been reflected in LT, so the situation is comparable to what would have happened in English if we still wrote English as Chaucer did, and maintained a formal pronunciation style of Middle English separate from the way we pronounce modern English.<sup>13</sup>

In diglossic languages (Ferguson, 1959) as extreme as Tamil, the linguistic culture almost always values the older (literary, formal) variety more, viewing it as the “real” language, while the spoken variety is seen as corrupted and “ungrammatical.” So diglossic linguistic cultures are most concerned with the corpus management of the literary language; in the late 19th and early 20th century this became a battleground in Tamil Nadu, and as we have noted, led to the rise of a number of political parties, one of which, the *Dravida Munnetra KaRaham*, gained power in Madras State in 1967. The revolutionary zeal, the evangelical, rhetorical and oratorical skills of the DMK mesmerized a generation of diaspora Tamils as well, but the power of that rhetoric has now waned in India, as people have seen successive DMK governments fail to deliver the promised goods, and lapse into internecine struggles over which offshoot of the DMK is the most devoted to Tamil.

Ironically, many meetings held to discuss Tamil and its ‘problems’ that I have attended were conducted in three varieties: LT, ST, and English, and not just for my benefit (though I, like some others, actually used two of the three). What Tamil corpus managers fail to see is that Tamil speakers, wherever they happen to be (i.e. in Tamil Nadu, Sri Lanka, or in Malaysia/Singapore), actually have a *repertoire* of languages available to them; they use LT for certain purposes, ST for many purposes, and English for yet others. The three are in a kind of symbiotic relationship; trying to deny any of them their place is detrimental to the others, or is seen as artificial and unnatural. To some, listening to a speech in hypercorrect Tamil is an enjoyable experience, akin to attending a cultural performance of some sort; but the *communication of ideas* is not one of its goals.<sup>14</sup>

---

<sup>12</sup> Tamil, being extremely diglossic (Britto 1986), has a literary variety (LT) that is not comprehensible to the uneducated, and vice versa. LT is based on a thirteenth century norm, the *nannuul* codified by the grammarian Pavanandi. Some sources date the *nannuul* from the 14<sup>th</sup> century, but this is not crucial to our argument.

<sup>13</sup> Thus, e.g. these lines from the *General Prologue* of Chaucer’s *Canterbury Tales*: “Thanne longen folk to goon on pilgrimages And palmeres for to seken straunge strondes To ferne halwes, kowthe in sondry londes.”

<sup>14</sup> I once was invited to attend a speech by a DMK orator at a university in Tamilnadu. After listening to the oration for some time, I confessed to the person who had invited me that I didn’t understand a word of it. “Neither do I,” he said, “but isn’t it *beautiful*?”

Evidence for the need to maintain the spoken language as well as the written I have presented in other research, most notably in my 1987 study of language shift among German-Americans. (Schiffman 1987). The languages involved are different, but the outcomes and conditions have similar results—the exclusive maintenance of the written language tends to result in language shift in spoken domains to another language, in both cases English; further shift to the exclusive use of English in all domains then may result.

**Some issues that are often ignored:**

1. *Literacy*. In Tamil Nadu, it is not expected that all adults will master LT; the latest census figures indicate that on average about 73% of Tamilnadu residents are literate, or at least functionally literate, for example up to 4th standard level only.<sup>15</sup>
2. *Registers, varieties, levels*. Tamil speakers fail to recognize that Tamil is, because of its ancient traditions, actually a language with many more levels or styles than, for example, English. In English, we usually expect all high-school graduates to be able to read Shakespeare, but we do not expect them (or even all college graduates), to demonstrate *active mastery* of Shakespeare's kind of English. That is, we do not expect Americans or British or Australian students to be able to *write sonnets or plays* using Shakespeare's language. We only require a passive understanding of his writings, and even attaining that is not easy. Certainly most literate Americans and British speakers cannot read Middle English (e.g., Chaucer) without a dictionary or a grammar, but Tamil linguistic culture expects literate Tamils to do the equivalent of that. And Old Tamil (dating from the 2nd-3rd century CE) is also treated as if it is part of the expected competence of A-level students. As for Old English, though we in the English-speaking world do have high-school seniors read Beowulf<sup>16</sup>, they read it in a *modern translation*, and we would never expect anyone but historians of English to learn Old English (which only goes back to the eighth century, not the second century, like *Sangat tamiR*).

In effect, Tamil consists of at least five different styles or varieties of language; we can distinguish at least the following:

1. Spoken Tamil (*peeccu tamiR*), spoken and understood by all speakers;
2. Modern LT, the language of social novels and short-stories; educated speakers can read this and write it.
3. Older modern LT, the language of Pavanandi (thirteenth century); educated people understand it, and some people can generate accurate specimens of it.
4. Medieval Tamil (the Alvars, TirukkuraL), and
5. Old Tamil (*sangat-tamiR*), which cannot be understood actively by anyone without the commentaries and dictionaries.

---

<sup>15</sup> Figures obtained from <http://www.tn.gov.in/schooleducation/statistics/table7and8.htm>

<sup>16</sup> At least this was true when I was in high school, but that was 50 years ago.

Some people would add another level, that of the epic literature, that is, *silappatikaaram* and *maNimeekalai*, which comes between Medieval and Old Tamil, and since there exists a hyperpuristic style incorporating material from a number of stages of the language, cultivated by pundits and taught in secondary-schools by Tamil teachers, we must also take “Pundit Tamil” into consideration even though it is artificial and hypercorrect; it typically mixes material (lexicon, morphology) from various stages of the language. One must also note that most literary norms of the past are *poetic* norms, not norms intended for writing prose. The only prose writing that existed before the modern novel came into existence was that used to comment on poetry; it was in itself not thought of as literature. Scholars and linguists therefore differentiate these different stages and forms of the language, but as noted, pundits do not, mixing various levels at will, and ignoring only the spoken styles, which they consider corrupt and degenerate.

It could actually be argued that the kind of Tamil being taught is not in fact hyper-archaic, because it contains many words that did not exist in the past. This claim is in fact to some extent true, as many new words have been coined to serve as replacements for English words. I refer to this kind of “fake” Old Tamil as “*archaistic*,” that is, it “looks archaic” because it uses old roots, but it is not really old, i.e. attested in ancient texts. As already noted, this hyperpuristic style was a strategy developed early in this century in Tamil Nadu, beginning first as a hyperarchaic style (which uses only words from older sources, such as *illam* for “house,” instead of *viiDu* or some other more modern form), but then was adapted for coining new words, that is, neologisms, for things that never existed in the past. In some ways this might be termed ‘hyper-*pseudo*-archaic’, since there is a certain falseness to this supposed ancient style; another might be simply epicurean, that is, concerned primarily with a certain style. I prefer the term “archaistic” because it encapsulates best the notions found in the term “hyper-pseudo-archaic.”

For example, in the building I lived in in Singapore, the English word “lift” (“elevator”) is referred to by one term *lipTu* on a sign on the outside, telling people in four languages what to do if there is a breakdown, and as a *mintuukki* on the inside, telling people, again in four languages, that the lift is under “closed-circuit television surveillance” (*uL amaippu kaamira* plus a verb meaning “surveille.”)<sup>17</sup> Notice that the ‘lift’ is neologized as *mintuukki*, which I take to mean “electric-lifter,” whereas the word *kaamira*, borrowed straight from English “camera” is used without impunity, for reasons I am unable to fathom.<sup>18</sup>

But this tendency to constantly tinker with terminology is something that is typical of hyperpuristic language planning, even if it is probably not conducive to convincing people who care about Tamil to adopt the terminology—one should probably settle on one form and eschew all others. The attitude is also paternalistic, that is, “the terminology

---

<sup>17</sup> I am unaware of what the official term for “lift” might be in Tamil Nadu, since I have never seen a Tamil sign on any lift there—the assumption there seems to be that if one is using the lift, one knows what to do with it, and what it is called

<sup>18</sup> If one travels around Singapore and reads such signs, one finds great inconsistency in the terminology, as well as in the ‘reasoning’ behind it, which tends to lead Tamils to distrust the credibility of the language planners.

planners know best” but this is also typical of the way Singapore is governed—it is a kind of intellectual arrogance that works *against* language maintenance because it conveys the message that the planners have not ‘gotten their act together, and are therefore not to be taken seriously.’<sup>19</sup>

### **An Example of This From Tamil Nadu.**

The English word “television” has five (six, if you count the Singapore word) equivalents in Tamil. That is, Tamil speakers may use any one or several of the following words to refer to “television”:

1. *duurdarshan* is the term used, even in Tamil, for government television broadcasting originating in New Delhi. When announcers say *itu duurdarshan* one knows that they are saying “This is the Indian government television broadcasting system.”
2. *tolaikaaTci* is the term used for television broadcasts originating in Madras. Another term, *tolainookki*, may also be used, and is usually the term used for television in Singapore.
3. “Television,” pronounced *TelaviSan*, is used by many educated people, often to refer to the concept of television, even when speaking Tamil, for example, “*TeleviSion vandadukku munnaale, reeDiyoo taan irundadu*” (“Before there was television, there was only radio”).
4. “TV,” pronounced *Tiivii*, is also used widely in spoken Tamil. Example: “*naan ungaLe Tiiviile paatteen*” (“I saw you on TV”).<sup>20</sup>
5. *paDapoTTi*, a folk term used by uneducated people, is derived from *paDam* “picture” and *poTTi* “box,” and is probably modeled on another older similar item, *paaDa poTTi* “song box,” that is, “radio.”<sup>21</sup>

Thus Tamil has at its disposal at least five words. One, from Sanskrit, is never used except to refer specifically to All-India government television programs, those that originate in Delhi. The second item is only used in the speech of television broadcasters, and would be the written form as well. Three and four are of course borrowed, and are widely used. But the last item comes from the spoken resources of the language, and though somewhat puristic-sounding (because it uses indigenous resources), its meaning is obvious to any speaker of Tamil, whereas the others would not be clear to people without education. In any event both one and two are loan-translations or calques from the international terms: *tele-* is the Greek word for “distance” and both *duur-* and *tolai*

---

<sup>19</sup> It is interesting to note that other languages that have attempted to modernize, such as Turkish in the early 20<sup>th</sup> century, have also adopted a hyperpuristic approach, rejecting loans from Arabic and Ottoman, and then claiming that certain forms that are of dubious origin are actually originally Turkish, in order to maintain the notion that 100% purity has been obtained. (Tachau 1964:198)

<sup>20</sup> This was said to me by a man passing in the street after I had appeared on Singapore Television.

<sup>21</sup> I am grateful to Vasu Renganathan for this example.



convey this; ‘vision’ is from Latin, and *darshan* and *kaaTci/nookki* are the Sanskrit and Tamil equivalents; the Tamil also conveys, by the ending *-i* on *nookk-i* (and *kaaTci*) that an instrument or thing is involved. It is not a “look-*ing*” but a “look *er*” or a “thing for looking.” This is also present in *mintuukk-i* “electric thing that lifts,” or “electric lift-er.”

My point here is that Tamil does not need five words for television. If it is going to try to displace English or Sanskrit, it needs at most one, and if the language is going to be useful to people, it needs the one that is most obvious and natural and acceptable to the most people in its meaning, even if it does not come from the most ancient source. *paDapoTTi* is a “pure” Tamil word but it is from a source that is neither archaic nor deliberately puristic, the natural resource of the living spoken language.

**Vocabulary Development Processes.** One of the aspects of the evolution of new vocabulary in any language that I know of is that the words that succeed and prevail tend to eventually be short and easy to use. Much of new English vocabulary, for example, starts out as a longish term of some sort, for example, “television” and is then reduced to “telly” (in Britain) or “TV” in North America. The word “fan” meaning person avidly interested in sports or entertainment, comes from *fanatic* and is a reduction or shortening of that term.<sup>22</sup> But Tamil purists do not allow Literary Tamil to use resources such as shortening, or acronyms, despite the prevalence of English acronyms in general use in the spoken language: MGR, HDB, MOE, COE, etc. In South India there is a college known by its initials OCPM. This has in fact evolved into a “word” *oociipiyam* and even obeys the morphophonemic *caariyai* rules for words ending in *am*, when case markers are added, that is, *oociipiyattule* “in/at OCPM (college).”

**Blending:** In English (and other languages) the taking of pieces from several words (not just letters) to make new words is known as “blending.” Such forms as “smog” (from **s**moke and **f**og), “modem” (from **m**odulator-**d**emodulator) are known as blends, and they are very common in Japanese, Russian, Spanish, Indonesian, and many other languages.<sup>23</sup> Computer terminology, consumer electronics, and the media abound with these things: mobisode, premercial, infomercial, infoganda, humint, Fortran, cyborg, quackupuncture, mockumentary, Hi-Fi, GIGAFLOPS, etc. In India the word “kilometre” has been borrowed in various languages; in Tamil it is *kiiloomiiTTar*, often abbreviated on signs as *kii.mii*. In some areas this has been taken as an actual word, and the blended form *kiimii* now stands for *kiiloomiiTTar* at least in the speech of uneducated people. This seems to me a laudable and natural development, and should be encouraged, because shortening and abbreviating is what seems natural in all the other languages I know, and because it is a folk resource, rather than something handed down by a higher authority. If this process is not allowed to happen in Literary Tamil, how will *minsaravacadi* (for “electricity”) manage to replace *karaNTu*? How will *mintuukki* (“elevator”) ever be expected to replace *lipTu*? How will *tuunduvaNDi* (“bus”) succeed in replacing *bassu*? Language planning edicts issued from on high will not make the longer terms succeed.

---

<sup>22</sup> Tellingly, the Tamil word for “fan” is a mis-translation of [*electric*] fan, i.e., *viciri*.

<sup>23</sup> German: *Gestapo* (from *Geheime Staats-Polizei*), and *Stasi* (*Staats Sicherheitsdienst*); Russian: *Kolkhoz*, *Komsomol*, *Gulag*, and many others..

There are those that argue that Tamil should be open to borrowings from other languages, and that it will not survive if it does not borrow, but these are mostly the voices of foreigners, such as Andronov (1979). In fact Tamil does borrow, but borrowings are only common in *peccut tamiR*. There are other resources, such as those already mentioned, that is, the natural resources, such as those that produced *paDapoTTi*. Literary Tamil needs to open itself to these resources—acronyms, blends, shortenings, initials—all of which are natural sources for new words in many languages of the world.

**Another example from the language of science.** I have recently been examining the development of terminology in the scientific discourse of what is known popularly as ‘mad cow’ disease. The discourse about ‘mad cow’ has again brought home to me that scientific discourse, and its terminology, is best developed by the *users*, and not by language planning bodies. And it also reminds us that ‘late modernizers’ (especially in the language planning field) are hampered by a tremendous backlog of vocabulary they must develop, in order to catch up even minimally with the language of science that is already established (Andronov 1979).

**Mad Cow disease.** Consider the terminological issues for those researchers who have recently studied the phenomenon of ‘mad cow disease,’ or, as it is known scientifically, *(Bovine) Transmissible Spongiform Encephalopathy*. The terminology for this in English is, of course, based on Greco-Latin roots, so that scientists working in other western European languages can easily recognize what it means.<sup>24</sup> But how would TSE (note the use of abbreviations) be *easily* rendered in another, non-western language? Scientific terminology in the west uses various sources, such as Greco-Latin vocabulary, abbreviations, blending (taking parts of various words), and other innovative re-uses of older vocabulary.<sup>25</sup> Stanley Prusiner, who was awarded the Nobel Prize in Medicine on October 7, 1997 for his work on TSE’s, claimed that ‘mad cow disease’ (or TSE) was caused by **proteinaceous infectious** particles, a term which he shortened by taking the first syllable of the first two words and combining them in a ‘blend’; but he rearranged the vowels slightly, to get ‘pri-on’ instead of \*pro-in. Thus the term ‘prion’ was coined to provide a shorter, catchier term than ‘*proteinaceous infectious particles*’ which, like ‘*Transmissible Spongiform Encephalopathy*’ is too much of a mouthful to use easily.

This very aptly illustrates how the people *developing the science* also *develop the register*, and then use various devices (borrowing, abbreviation, blending, etc.) to make the register even easier to use. Someone working in another language who wanted to do research on TSE’s cannot afford to expend time translating these terms and finding suitable ‘native’ terminology for something as complicated as ‘transmissible spongiform encephalopathy’. Their time would be better spent doing research on the *disease*, not on the terminology. It should be emphasized, however, that besides the scientific ‘register’ there are also often ‘folk’ or lay terminologies for the same phenomena. Thus the folk

---

<sup>24</sup> In French, the term is ‘l’encéphalopathie spongiforme bovine’ and in other western languages, the Greco-Latin roots are also used, with some tweaking of spelling and morphology to fit the individual languages’ patterns.

<sup>25</sup> In high-energy gaseous physics, the somewhat quaint English verb ‘quench’ has been taken to refer to ‘rapid cooling of a hot gas’, which is of course not its original meaning.

terms ‘mad cow’ (Spanish *vaca loca*, French *vache folle*, German *Rinderwahnsinn*, etc.) are used by lay people, journalists, etc. and even appear on some of the ‘scientific’ websites<sup>26</sup>. So here we have a kind of diglossia that not only keeps the scientific and lay discussions separate, but there are also H-variety terms (TSE and its expanded form) that have equivalent L-variety terms (‘mad cow’). As is perhaps obvious, any language can come up with an equivalent for ‘mad cow’ but finding an equivalent for ‘Transmissible Spongiform Encephalopathy’ is another issue. If the language in question uses Greco-Latin vocabulary as a source for scientific terminology, there is no problem. If other resources have to be resorted to, the problems are more complex. Tamil, it should be noted, can easily produce the lay term *maaTTu payttiyam*, (‘bovine madness’), but I doubt if there is a scientific H-variety version of this in use anywhere.

To summarize: the scientific discourse about this disease uses several devices or strategies to develop the terminology:

- The **Greco-Latin** ‘full’ form: *Transmissible Spongiform Encephalopathy*;
- **Abbreviation**: the abbreviated form ‘TSE’ is widely used also in the scientific register, and not only in English.
- **Blending**: the term ‘*proteinaceous infectious particle*’ is shortened by taking the first syllable of the first two words, and then manipulating them slightly to give it the pronounceable form ‘prion’.
- **Lay discourse**: Even in the scientific discourse, but mostly in the lay or folk discourse, the term ‘mad cow’ is also used, even though it only refers to bovine TSE’s.

This can perhaps explain why in a nation like India, (and a language like Tamil) which originally intended to abandon English in favor of indigenous resources, has still, almost six decades after independence, not managed to do so—Indian doctors and technologists prefer to be ‘in the loop’ of international work in all these fields, so higher education in India continues to be almost exclusively conducted in English. This also means that Indian doctors and engineers can easily obtain jobs in other countries, and send home remittances that help sustain the economy of the subcontinent. Keeping up with the terminology in ‘cutting-edge’ fields like computer science and medicine means knowing what is going on in a language of wider circulation (LOWC), and using the terminology already developed, rather than relying on ‘indigenous’ or ‘pure Tamil’ resources.

But of course the reliance on English for science and technology means that diglossia of the Fishman type is maintained and reinforced, and will probably continue to be

---

<sup>26</sup> <http://www.sciam.com/article.cfm?chanID=sa006&colID=1&articleID=000E3363-C45B-10CF-BCE683414B7F0000>. The German website <http://www.rinderwahnsinn.org> gives the scientific term for ‘Rinderwahnsinn’ as BSE, i.e. Bovine Spongiforme Encephalopathie, a term clearly borrowed from English, and also refers to ‘Prionen’ i.e. ‘prions’ as the agent that transmits the BSE. The Spanish term *vaca loca* follows the folk naming pattern, while the scientific term is *Encefalopatías Espongiformes Trasmisibles*, but is referred to as one of a kind of ‘TSE’s, using the English term as the basis, but also as *Encefalopatía Espongiforme Bovina*, or EEB. (<http://www.revistaoccidente>) In any case, all the western languages use the Greco-roman vocabulary as the basis for their own scientific terminology.

maintained, and not just in India—English is also important in other former British colonies as well as elsewhere in the world of science.

**The problem of ‘late modernization.’** As is perhaps obvious by now, it seems clear to me that linguistic cultures that attempt to modernize their languages late in the game, i.e. centuries after other languages have done so, are going to have problems. As various writers on the subject of corpus planning have pointed out, modernization of a language involves a number of features. Clyne, Baldauf and Ingram’s statement on this can serve as a useful example:

**“Modernization** If a language is to function for science, technology, commerce, politics, and scholarship, then new lexical items must be introduced. This has occurred in decolonization situations, among others. Alternatives include borrowed terms (in Papua New Guinea, Tok Pisin *dimokretik*), loan-translation (Afrikaans *draadloos* < Eng. *wireless*), formations based on existing words (Tok Pisin *wok dokta* ‘health service’), and semantic extensions (Icelandic *sima* ‘rope’ > ‘telephone’). As corpus treatment systems become established, the ideology of indigenization as a driving force becomes less motivated (Jernudd 1983).

During much of the 20th century, even in Western societies, attempts have been made to standardize terminology, including definitions and abbreviations in specialized fields such as electrical engineering and navigation. This has only been partly achieved. Terminological planning has occurred at three levels: national, supranational (within one language), and international (across languages).(Clyne, Baldauf and Ingram 1992, 2003)”

As my ‘mad cow’ example shows, however, there are limits to how much an erstwhile pre-modern language can modernize, for the reasons that I have already adduced. The amount of effort involved, the willingness of the citizenry to accept the new vocabulary, the enormous effort it takes to assimilate the new terminology (even if there is willingness to do so), and the time delay between the beginning of terminology development and its actual availability, are all factors that indigenous scientists have to consider. In other words, if speakers of Language X want X-ish<sup>27</sup> to be a viable alternative to other LOWC’s that are already in use for science and technology, they have to decide whether they wish to stand back and wait for a language academy or some other body to come up with the terminology for a given field, such as ‘mad cow’ disease, or whether they will be better off economically and professionally if they opt to continue to use English or some other LOWC for this purpose. For those who already know English or another LOWC, the rewards of professional isolation while waiting for lists of dubious terminology are not worth it. Add to this the freedom to participate in the larger discourse of the scientific field in question, and also have ones own work *read* in it, plus the concomitant opportunities for employment in a larger job market, and the answer for most scientists will be obvious.

---

<sup>27</sup> This is the terminology Fishman proposes for talking about a hypothetical language engaged in this process.

Language groups such as the Tamils typically cite Japanese as an example of a language that has modernized successfully, but this ignores the fact that Japan started this process 150 years ago, and was also willing to *borrow* large amounts of vocabulary from Chinese, German, English, and even Dutch.

As Fishman has pointed out (Fishman 1989: 395), such groups often seem to want modernization to happen all at once, in all domains of life, rather than concentrating on crucial areas such as inter-generational transfer. These language ‘utopians’ often have the idea that their language cannot hold up its head as a real and vital language unless it does so in all domains, simultaneously.

“However, the huge amount of time, effort and funds necessary in order to reach that goal [of teaching Astrophysics in Irish] (both in connection with perfecting the necessary terminologies and texts, and in connection with getting teachers themselves to learn and use them) and the small number of individuals who will ever use Irish or Xish for that purpose, seriously raises the question of whether *that* is a wise policy or not. [...]

But, as Fishman goes on to point out, if this goal is questioned (especially when the language is simultaneously losing ground on the home front) then the bugaboo of ‘folklorization’ raises its head. Relegating Xish to vernacular and hearth-and-home uses, and denying them meaningful roles in the modern world is taken (by the would-be modernizers) to mean that they are nothing more than anachronistic ‘spinning wheels’ in a world dominated by computers and other kinds of 21<sup>st</sup> century technology. It is perhaps not accidental that analogies to simpler, homespun tools have often been used when referring to pre-modern languages and their utility. The renowned French scholar Meillet once said that:

"In comparison with French the Breton language is an instrument of such crudity and such little real use, that no intelligent Breton speaker could even dream of using it in preference to French. One might as well claim that the electric lamp oppresses the candle or the torch, or complain that the reaper oppresses the scythe." (Meillet 1928:178)<sup>28</sup>

Fishman’s focus here is on language maintenance, but the point is the same:

“The most important lesson to remember [...] is the following one: *the intimate domains are also the most sheltered*; they nurture inter-generational continuity, even if these domains are not the major movers and shakers of the modern world. On the other hand, *the most powerful modern domains are also the*

---

<sup>28</sup> Meillet then goes on to say that "They [regional languages] are not really in competition with major languages. They are nothing but poor relations who have fallen on hard times." (op.cit. p. 185; my translation—HS)

*most exposed* to the vicissitudes of power confrontation. [...] Endangered languages must assume control of the former—the intimate spheres of family and community.” (Fishman op.cit. 399)

The attempt to modernize a pre-modern language for use in all modern domains, therefore, is, in this day and age, probably *counterproductive*—it fails to achieve the results it sets out to achieve, while simultaneously ignoring the maintenance of the language in the hearth-and-home domains. It locks the barn door, but leaves the windows open. In the case of German-Americans cited above (Schiffman 1987), the policy of puristic devotion to the literary language was covertly and unwittingly a policy that encouraged language *shift* to English; Gopinathan and Mani (1983) refer to it as ‘subtractive bilingualism’ in Singapore, and it is not far-fetched to claim that the same can be said for puristic devotion to Tamil in Tamil Nadu—if scientists don’t wish to accept the Tamil terminology proposed for them, they not only abandon the Tamil language, they may also abandon Tamil Nadu, since emigration is an easy out, and their knowledge of English allows them that option. This is surely not what advocates of Tamil want, but it is what seems to be not only a dangerous possibility, but the inevitable (and regrettable) outcome.

## References

- Annamalai, E. (1979) “Movement for Linguistic Purism: the Case of Tamil”. Central Institute of Indian Languages, Mysore.
- Andronov, M. S. (1975) Problems of the national language in Tamilnad. *Anthropos*, 70: 180-93.
- Britto, F. (1986). *Diglossia: A study of the theory with application to Tamil*. Washington, DC: Georgetown University Press.
- Caldwell, Robert (1961) [1856] *A Comparative Grammar of the Dravidian or South-Indian Family of Languages*. University of Madras, Madras.
- Clyne, Michael, R. B. Baldauf, Jr., D. E. Ingram. (1992, 2003) "Language Planning" *International Encyclopedia of Linguistics*. William J. Frawley. ©1992, 2003 Oxford University Press. International Encyclopedia of Linguistics: (e-reference edition). Oxford University Press. University of Pennsylvania. 22 August 2006 <http://www.oxford-linguistics.com/entry?entry=t202.e0581.s0001>
- Ferguson, C. (1959). “Diglossia.” *Word*, 15, 325-337.
- Fishman, J. (1981). Societal Bilingualism: Stable and Transitional. In J. Fishman

- Sociolinguistics: a Brief Introduction*. (pp. 78-89). (Originally Published in 1967 as Bilingualism with and without diglossia; Diglossia with and without bilingualism. *Journal of Social Issues*, 23(2), 29-38).
- Fishman, J. (1981). *Sociolinguistics: A brief introduction*. Rowley, MA: Newbury House.
- Fishman, J. (1989), *Language and Ethnicity in Minority Sociolinguistic Perspective*. Multilingual Matters 45. Clevedon and Philadelphia: Multilingual Matters, Ltd.
- Fishman, Joshua. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Clevedon, England: Multilingual Matters.
- Gopinathan, S. & Mani, A. (1983). Changes in Tamil language acquisition and usage in Singapore: A case of subtractive bilingualism. *Southeast Asian Journal of Social Science*, 11(1), 104-116.
- Gopinathan, S. (1994). Language policy changes 1978-1992: Politics and pedagogy. In S. Gopinathan, A. Pakir, H. W. Kam, & V. Saravanan (Eds.) *Language, society, and education in Singapore: Issues and trends*. Singapore: Times Academic Press.
- Gopinathan, S., Pakir, A., Kam, H. W., & Saravanan, V. (Eds.) (1998). *Language, society, and education in Singapore: Issues and trends*. Singapore: Times Academic Press.
- Haugen, Einar I. (1966) *Language conflict and language planning; the case of modern Norwegian*. Cambridge: Harvard University Press.
- Hoffmann, F. (1981) 'Triglossia in Luxemburg'. In Einar Haugen, J. D. McClure and D. Thomson (eds) *Minority Languages Today*. Edinburgh University Press, Edinburgh.
- Kloss, Heinz (1967) *Abstand languages and Ausbau languages*. *Anthropological Linguistics*, 9(7): 29-41.
- Meillet, Antoine. 1928. *Les langues de l'Europe nouvelle*. Paris: Payot.
- Schiffman, H. F. (1987). Losing the battle for balanced bilingualism: The German-American case. *Language Problems and Language Planning*, 11(1), 66-81.
- Schiffman, Harold F. (1992) 'Resisting arrest' in status planning: structural and covert impediments to status change implementation'. *Language and Communication*, 12(1): 1-15.
- Schiffman, H. F. (1996). *Linguistic culture and language policy*. New York: Routledge.
- Schiffman, H. F. (1997). Diglossia as a sociolinguistic situation. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics*. London: Basil Blackwell.
- Schiffman, H. F. (1999). *A reference grammar of spoken Tamil*. New York: Cambridge University Press.
- Schiffman, H. F. (2003) "Tongue-Tied in Singapore: A Language Policy for Tamil?" *Journal of Language, Identity & Education*, Vol. 2, No. 2: pages 105-125.
- Shapiro, Michael and H. Schiffman. 1982. *Language and Society in South Asia*. Delhi: Matilal Banarsidass.
- Tachau, Frank. (1964) "Language and Politics: Turkish Language Reform." *The Review of Politics*, Vol. 26, No. 2, pp. 191-204.