

An English-in-Asia Database

Susan Butler¹

Macquarie University, Australia

The dictionary project began—and I began dictionary work with it—in 1970. At that stage we worked on 6" x 4" filing cards—roughly 150,000 of them in the main file, with subsidiary files of specialist collections and work in progress. We now work on a Unix based system called Titan—and the change is profound. Processes which took person-years on filing cards, take a fraction of that time on computer.

In the old days any sense that you were writing the dictionary was diluted by the sheer number of hours you had to labour to effect any change in it, and then later by the fact that the computer in which the dictionary was housed was owned by the typesetter, so editing was a remote control operation. Now we are face-to-face with the dictionary, and the processes for editing are under our control.

The other wonderful thing about this system is that I can switch from dictionary to corpus for background to a dictionary entry. When the dictionary first appeared in 1981 our citation collecting was minimal—more notes to remind ourselves of things that should be in the dictionary than a complete set of citations. We now have a fifteen million word corpus supplemented by accessing the various newspaper corpuses around Australia which have been steadily growing since the mid '80s. Our focus in our own corpus has been on Australian books—fiction and non-fiction—and now that that is more or less accomplished, on special areas of language not covered by the available corpuses.

For example, we are interested in amassing self-edited texts—texts which are correct in the eyes of the writer but which have not passed under the eye of editorial gatekeepers, machine or human. We are scanning a number of student essays sent in each year for a competition run by a newspaper. From this we hope to find what Australians really think on matters of style and usage, as opposed to what editors think they ought to think. 'Color' and 'colour' are both observable spellings in Australian texts, but which one would Australians, left to their own devices, offer.

So, the corpus has been and is continuing to be of great benefit to us. Moreover, these days a corpus can be put together with remarkable ease. Once a month we access the newspapers and siphon off a sample. Books can be fed in via the scanner at the rate of one a day, provided they present no particular difficulties in terms of reproduced type or difficult typeface. When I remember the difficulty with which we compiled our citation cards in 1970, it all seems quite miraculously simple.

By using the Macquarie Spellchecker (or a modified version of it) we can produce lists of items which are not already covered by the dictionary. These were something of a surprise at first. It is remarkable how many lexical items we use in our written texts which do not fit tidily into the lexicon. Some are adventurous words, some are adventurous spellings, some are efforts to capture speech in writing.

¹ Susan Butler's background is in commercial lexicography. She is an editor of the *Macquarie Dictionary of Australian English*.

Some are perfectly regular but one-off formations for existing words, for example:

Items Not Found By Spellchecker In A Corpus Sample

Anglo-Australian	addition
Edna Mays	rhyming slang
Golden Joeys	rhyming slang
Jappy Dicks	rhyming slang
Hawksian	one-off
Mozambique Rose	?
Proustian	addition
U P	marginal
Wilderian	Marginal
Woolloomooloo spit	obscure
Yellow Boys	obscure
Zapata moustache	dictionary entry?
abdominals	addition
abear	obscure
aberrate	obscure
abouts	one-off
absooty	speech
acacian	one-off
accessive	one-off
acne-ed	addition
agamemnonian	one-off
Airfix	trademark
airletters	solid
airmiles	solid
alevena	one-off
allfired	spelling
allie	odd spelling

But from all of these the key dictionary items emerge.

Our interest in English has extended from Australian English to the English that is spoken in neighbouring countries, particularly in South East Asia. I spent three months of last year in Singapore as a Visiting Fellow at the University of Singapore. I experienced an odd sense of

deja vu—that the issues in English in Singapore were ones that exercised Australians twenty years ago. What was standard English? Was there more than the then allowable two? Who was to set the standard? Were levels of formality and informality, allowability and non-allowability to be decided by the people who lived there, or others. These issues are more-or-less resolved in Australia although there would still be a conservative element in the community that regarded British English as the only ‘real’ correct English, and *Oxford* as the only possible dictionary. The rest of us are grappling with the problems of standard English, particularly with the fact that standard English is an ever-shifting quantity. We are writing our own dictionary, our own style manuals, conducting our own style conferences on contemporary Australian English.

In Singapore we held a lexicography workshop which started out tentatively asking was there any such thing as Singapore Standard English, and moved ever more confidently and agreeably to asking how Singapore English in all its aspects, Standard Singapore English, colloquial Singapore English, Singlish, pidgin English, might be recorded and mapped. I think that shortly there will be a boom of Singaporean books focused on language—just as twenty years ago there was a book of similarly motivated Australian books on Australian English.

There are some collections of Singaporean English already—which I think are all exemplified by the one published by RELC. They are one-person collections and there is always a limit to what one person can do in this field. For the lexicographer they present a problem because they tend to mix up standard items of high frequency indiscriminately with non-standard items of great colour but low frequency.

Examples:

hwd	aquafarm
cite	Another big flaw in the law is classifying poultry and livestock raising as well as aquafarms , fishponds and prawn raising, within the scope of the CARP when these are entirely different enterprises.
source	<i>Manila Bulletin</i>
hwd	atin-atin
cite	Most of the diners are charmed by the homey atin-atin entertainment...
source	<i>Daily Inquirer</i>
hwd	balagtasan
cite	Give strategically-located stores radio and TV sets and start a <i>rural distance education literacy program</i> on the air. Folk wisdom, balagtasan , soap opera, basketball games, can all be repackaged...
source	<i>Business Star</i>
hwd	bangketa
cite	Diarrhea and parasitism, for instance, are not always caused by the failure to eat the right kind of food or food bought less expensive from sidewalk vendors or in bangketas .
source	<i>Daily Inquirer</i>
hwd	barangay
cite	A seminar-workshop for barangay development councils in Tondo on the composition, role, functions, and responsibilities of the council under the

- Local Government Code will be conducted by the Cabangis-Genaro Youth Club (CYGC) on Sunday, Aug. 9, at the Tondo Sports Complex.
 source *Manila Bulletin*
- hwd **barangay**
 cite Four army troopers were killed and others were seriously wounded when the service jeepney they were riding was ambushed by the New People's Army (NPA) at **barangay** San Roque, Camaraines Sur yesterday afternoon.
 source *Manila Bulletin*
- hwd **barrio**
 cite Dr Flavier, our new Secretary of Health and coined a Tagalog phrase for family planning that made him world famous; "pagpigi sa panggigil" when he pioneered family planning in the **barrio**. It means "to control your lust".
 source *Manila Bulletin*
- hwd **calamansi**
 cite This fiery liquid concoction is a combination of tropical fruit juices such as pineapple, **calamansi** and orange that's loaded with Vitamin C (to help guard against the showers of the season!), while the dash of rum provides the tasty fire to flavor.
 source *Business Star*
- hwd **churrascarias**
 cite Many **churrascarias** (steak or barbecue houses) have adopted the *rodizio* system under which all of these delicious meats are brought to your table on skewers, directly from the grill...
 source *Daily Inquirer*
- hwd **churrasco**
 cite Another typical dish that both Eli and Aurora like to serve is **churrasco**, or barbecue, which consists of various types of meat grilled over an open flame and served in large, succulent pieces.
 source *Daily Inquirer*
- hwd **cocolumber**
 cite It was explained that the decline in wood production has put much more pressure on coconut trees as "**cocolumber**"—now widely used as scaffolds for building construction—has gained popularity.
 source *Business Star*
- hwd **demonstration sport**
 cite The Philippines, reeling from two bitter losses in boxing quarterfinals, consoled itself with a bronze medal in the **demonstration sport** of takewondo when Stephen Fernandez reached the semifinals of the bantamweight division in the 25th Olympics here.
 source *Manila Bulletin*

- hwd **doy pack**
cite There will include paper pouches, plastic laminates, aluminum foil pouches, sachets, **doy packs**, and tetra brik cartons.
source *Business Star*
- hwd **ecotourism**
cite He also reported that the Department of Environment and Natural Resources (DENR) agreed to support proposals from the DOT and the Baguio community to package the camp into an "**ecotourism**" area by preserving the 284 reservation area which was under the DENR's management.
source *Business Star*
- hwd **flying voter**
cite Comelec will field policemen, soldiers and ROTC cadets to watch out for **flying voters**.
source *Daily Inquirer*
- hwd **jeepney**
cite Senior Police Officer Casimiro Soriano of the Muntinlupa police said that the victims on board Gloria's **jeepney** with license plate number DHV 868, were cruising along the National Highway in barangay Bayanan at past 10 p.m. when the incident occurred.
source *Manila Bulletin*
- hwd **jueteng**
cite Sari-sari stores are **jueteng** and gambling joints for PBA games and pintakasis as well as sweepstakes selling centers.
source *Business Star*
- hwd **kraft board**
cite While any ordinary **kraft board** would have done the same job, she said, Japanese buyers prefer white over brown because it looks cleaner.
source *Business Star*
- hwd **kundiman**
cite At periodic intervals, they take a break from attending to the restaurant patrons and converge around the piano at the center of the dining room to sing **kundiman**, art songs, arias.
source *Daily Inquirer*
- hwd **midnight decree**
cite Harbor pilots groups are currently engaged in a legal war with the port agency over the validity of Executive Order (EO) 1088, which was issued in the mid-80s by then President Marcos as a **midnight decree** raising pilotage rates by as much as 5,000%
source *Business Star*
- hwd **mini-mini store**
cite Train owner, wife, children, kins in the basics of entrepreneurship—costing, bookkeeping, **mini-mini store** management...

- source *Business Star*
- hwd **miting de avance**
cite Debating clubs and pros-cons are spawned here to parallel **mitings de avance**, to get a pulse of voters' preferences, and to check out political balimbings and loyalists.
- source *Business Star*
- hwd **ozeki**
cite Another Hawaiian, Akebono, or Chad Rowan, won the tournament with a 133-2 record and looked set to join Konishiki at summo's second-highest rank—**ozeki** (champion).
- source *Business Star*
- hwd **PBA game**
cite Sari-sari stores are jueteng and gambling joints for **PBA games** and pintakasis as well as sweepstakes selling centers.
- source *Business Star*
- hwd **pintakasis**
cite Sari-sari stores are jueteng and gambling joints for PBA games and **pintakasis** as well as sweepstakes selling centers.
- source *Business Star*
- hwd **prawn raising**
cite Another big flaw in the law is classifying poultry and livestock raising as well as aquafarms, fishponds and **prawn raising**, within the scope of the CARP when there are entirely different enterprises.
- source *Manila Bulletin*
- hwd **pros-cons**
cite Debating clubs and **pros-cons** are spawned here to parallel mitings de avance, to get a pulse of voters' preferences, and to check out political balimings and loyalists.
- source *Business Star*
- hwd **Pugutan**
cite The Lenten tradition highlighted by the "**Pugutan**", a depiction of the miracle to Longinus, the one-eyed centurion who was made to see the blood that trickled when he lanced Christ's side...
- source *Daily Inquirer*
- hwd **quedan**
cite The Department of Agriculture (DA) in a bid to pump prime agricultural credit through the **quedan** guarantee system, has signed up as the biggest single government investor in the newly expanded **Quedan** and Rural Credit Guarantee Corporatio (Quedancor).
- source *Business Star*

- hwd **sari-sari store**
cite The **sari-sari store**. There is absolutely no equivalent in the sprawling Indian subcontinent and the South Asian societies. Nor in East and the Southeast Asian nations, although variations of the **sari-sari store** can be found in Thailand and Indonesia.
source *Business Star*
- hwd **solon**
cite One **solon** we had discussed with on the matter had the impression that one-and-a-half hectares would be enough for certain crops.
source *Manila Bulletin*
- hwd **solon**
cite **3 solons** join NPC
source *Manila Bulletin*
- hwd **stand-by**
cite The sari-sari store is the neighborhood socio-politico center...It is the **stand-by** center for children and adults, for bullies and good neighbors.
source *Business Star*
- hwd **talipapa**
cite They (Sari-sari stores) represent unharnessed power equivalent to the neighborhood church and its pulpit, and the neighborhood **talipapa**.
source *Business Star*
- hwd **tetra brik**
cite There will include paper pouches, plastic laminates, aluminum foil pouches, sachets, doy packs, and **tetra brik** cartons.
source *Business Star*
- hw **thrift bank**
cite Presently, DBP has accredited a total of 60 commercial banks, private development banks, **thrift banks**, savings and loan associations, leasing and finance companies and investment houses under its wholesale lending program.
source *Business Star*
- shwd **tokwa**
cite When before all veggie food had to be drab, boring, tasteless and had **tokwa** coming out of your ears, there is now a certain sophistication that it offers...
source *Malaya*
- hwd **yaya**
cite A less amusing sight: there's this lovely celebrity who's here with her husband, their little daughter, and the girl's **ya-ya**.
source *Daily Inquirer*

hwd **yokozuna**
 cite Konishiki...denied newspaper reports that he blamed racism for gaining promotion to **yokozuna** or grand champion.

I did a little reading myself of Philippine newspapers just to get a feel for English here. I picked up a number of items which are not in the Minidictionary but which seem in context to be comfortable and familiar items. Of course, on a few newspapers I have no hope of assessing currency. *Bangketa* and *barrio*, *doy pack* and *jueteng* seem established. But is a *flying voter* a real entry? How well-known are *kundiman*? Are there lots of *mini-mini stores*? Is *pros-cons* a real word?

Usage matters are even more difficult to assess.

Usage

hwd **chic**
 cite If there is no sign to be found in a “**chic**” hotel, for example, to advise the uninitiated diners which, if any, of its eateries has a dress code.
 source *Manila Bulletin*
 year 1992
 date 5 Aug 7
 cmnt
 type usage usu. applied to a person or an article of clothing

hwd **discuss with**
 cite One solon we had **discussed with** on the matter has the impression that one-and-a-half hectares would be enough for certain crops.
 source *Manila Bulletin*
 year 1992
 date 7 Aug 6
 cmnt
 type usage

hwd **fisherfolk**
 cite ...40% is allotted to small farmers, **fisherfolk**, cooperatives and other private investors.
 source *Business Star*
 year 1992
 date 20 July 2
 cmnt seems quaint, archaic
 type usage

hwd **jibe with**
 cite If it does not **jibe with** the teaching of the chu
 source *Business Star*
 year 1992
 date 20 July 4
 cmnt in Random House US English
 type usage international

hwd	public transports
cite	...; where they blow cigarette fumes in every face, whether in public transports , elevators, movie theaters, or restaurants...
source	<i>Manila Bulletin</i>
year	1992
date	5 Aug 7
cmnt	plural form
type	usage
hwd	request for
cite	This refers to your letter... requesting for a clarification as to whether you shall refund the withheld and remitted in the 1991 Personnel Economic Relief Allowance (PERA) of your personnel...
source	<i>Business Star</i>
year	1992
date	20 July 4
cmnt	
type	usage
hwd	said
cite	Villahermosa expects said packages to become even more popular as the number of single households increases
source	<i>Business Star</i>
year	1992
date	20 July 5
cmnt	use of said occurs a number of times
type	usage

Discuss with is odd to me but seems natural here. *Request for* likewise. This turns up in the small Singapore corpus too.

Fisherfolk seems quaint to me—is it standard here?

Likewise *said*—and in the letter I received—*the undersigned*.

What I as a commercial lexicographer am looking for is a collection of standard lexical items of high frequency in each of the varieties of English in South East Asia. I am leaving South Asia out of it for the moment—one thing at a time.

To do this I wish to amass a corpus of English in each variety. For example, for Singapore English we have begun with a representative sampling of Singapore fiction and non-fiction to be scanned into the computer. In addition to this I hope to persuade *Straits Times* to hand over samples of machine-readable text. It is possible to amass reasonably quickly a corpus that will not tell everything there is to be known about Singapore English in all its aspects, but which will produce the lexical items which are thoroughly accepted or culturally significant. In two books *kiasuism* has emerged a number of times, for example.

This is the list of items which has emerged from an analysis of two books.

Figure 1: Words in Asiacorp not in Spellchecker

Group 1	September 1992	
airwell	kiddiepix	sarabat
aiya	koay	sayang
allamah	kueh	seashells
amokry	lalang	seelo
ang pows	meaningfully	singaporeans
angsana	nanyang	starfruit
bathmats	nonya	sunkist
bearingscraper	oi	supervisee
boringly	phonegossippy	susah
breadman	phui	taximen
chandeliered	plotless	troup
clippedly	quiveringly	uncompleted
genteely	ratlan	unfilial
glasspanes	realer	unlegislated
hati	remisier	vroom
kebaya	samfoo	vroomed
kebun	samfoos	

Words in Asiacorp not in Spellchecker

Group 2	September 1992	
above the back collar	coopulate	girlchild
asianness	detac	goodnesses
beehoon	ditmov	grandmudder
bonzo	eat all you can	granduncle
busstops	eeee	greatgrandfather
cagful	everything	hokkien
calibans	filipina	hometown
centrestage	fincombed	humsub
ceray	fishhead	humsubism
clusterings	fishseller	humsubs

kiasium	pinyin	shophouse
kiasuer	plentitude	sociocultural
kiasuism	plepare	sociohistoricocultural
latenight	preelection	suay
lomanic	profusive	surefire
meet your member of parliament	relevant	talcumed
melonsced	reruns	tightnesses
mrt	revitalisation	twochild
mudder	ricesacks	ultraviolated
multicolours	rojak	urmt
multiracial	roughnesses	vermincaused
nononsense	sarongclad	viceconsul
patriach	schoolgoing	whateveritis
pileup	shaveless	wordplays

The importance to us of the corpus is that it can claim to be representative of standard English in each of the targeted countries so that we can with some confidence produce a set of corpus-based dictionaries for South-East Asia which cover the key lexical items of contemporary standard English in the region in a systematic way.

For this project I need from each country quantities of text—machine-readable or scannable, and a lexicography consultant. In Singapore we have established a working relationship with the Department of English Language and Literature which involves transferring our dictionary and Australian corpus to them for research purposes. They are interested in the totality of Singaporean English for which a corpus-backed dictionary of another variety is a useful reference point. There is no reason why the Asian corpus should not develop on a cooperative research basis. It is in our interests to promote greater research into the individual varieties of English which will lead to a broader range of information to use as a basis for a general market dictionary in South-East Asia.

I can imagine that some comparative studies on the differences and similarities between the various Englishes of the region would be of interest. This kind of information would lead to more informed discussions of what standard English is in each community of English speakers and would have significance in formulating educational policies.