ICE Age 2: ICE corpora of New Englishes in the making

'How' a Fiji corpus?
Challenges in the compilation of an ESL ICE component[1]

Carolin Biewer, Marianne Hundt and Lena Zipp,[2] University of Zurich

1   Introduction

In 1996, Tent and Mugler wrote an article arguing for the compilation of a Fiji English component of the ICE family of corpora (Tent and Mugler 1996). As they themselves could not pursue the project further, it was taken over by the current research team in 2005. We started with the written component and have recently begun to sample spoken data, as well. This article continues their argument, adding a 'how' to their previous 'why'.

The challenge that we face is to compile a Fiji English component which is at once representative of the current language use in Fiji, and at the same time comparable to all other ICE corpora.[3] This corpus should not only enable us to get a clear idea of the actual usage of acrolectal English in Fiji but also to compare this variety with other standard varieties of English worldwide. This task of creating a corpus that matches the general ICE framework, but which is still representative of the universe of the local and culture-specific production of standard English texts in Fiji, quickly started to look as difficult as squaring the circle. As Leech points out:

> While it makes sense to achieve success in both representativeness and comparability, there is a sense in which these two goals conflict: an attempt to achieve greater comparability may actually impede representativity and vice versa. (Leech 2007: 142)

There is a danger that, the more we follow the original ICE guidelines, which represent a western and native speaker perspective of text production, the less

likely it is that the corpus will represent the actual use of acrolectal English in Fiji (see Schmied 1996: 183 on ESL[4] corpora). Careful navigation is required in the compilation process as both comparability and representativeness are equally important features, and "a gradual approximation to these goals" (Leech 2007: 143) should be attempted. That also means that we need to be both rigorous and pragmatic at the same time in our decisions – following the general ICE framework but allowing for an adjustment to the local reality.

A very basic decision concerns the suitability of informants and texts to be included in the sample. In compiling ICE Fiji, we have had to readjust our decisions to the availability of data in Fiji and find a way to categorize the collected texts so that they matched the data included in the other ICE corpora. In the following, we will give a brief description of major issues concerning the compilation of ICE-Fiji, highlighting in particular areas of corpus design (Section 2), problems of data collection (Section 3) and text categorisation (Section 4). In doing so, the article documents the most important decisions that were taken in the compilation process. This information will also provide future scholars with a better idea on the potential and limits of this corpus-linguistic resource. Because English is the second language for most speakers in Fiji, the particularities of the ESL context will feature prominently.

2   Corpus design: Localising the framework

The two main criteria in the general ICE design for the inclusion of a text are age and education of the author or speaker and not the "evaluation of the language in a text" (Nelson 1996: 28). The definition of eligibility of authors or speakers is therefore crucial. The other important decision is how to limit the time span for the texts to be sampled in the corpus.

2.1   Who to include in the corpus?

For a text to be included in an ICE corpus the author must be at least 18 years old and must have benefited from formal education in English up to the secondary or tertiary level (Nelson 1996: 28; Greenbaum 1991: 3). An additional requirement is that the author was either born in the country or came to live there early in his or her life (Nelson 1996: 28). For the compilation of ICE-NZ, Janet Holmes had to ask the question "At what point does an immigrant become a New Zealander?" (Holmes 1996: 164). In a country like New Zealand, with its colonial background, large sections of the population are immigrants. The related challenge in the Fiji setting, where it is customary to leave the country for secondary or tertiary education, is to give an answer to the question 'When

does a Fijian become a Fijian emigrant?' Adapting the general ICE design and Holmes' (1996) definition of who counts as a New Zealander (1996: 165), we initially decided to define a speaker of Fiji English who is eligible for inclusion in for ICE-Fiji as someone who

- is currently a resident of Fiji
- is an adult of the age of 18 or older with a formal education in English up to tertiary level
- has ideally been living in Fiji since the age of ten and spent less than 10 years or less than half of their life-time abroad (whichever is shorter), and has not spent more than one year overseas in the last three years.

These rather strict criteria quickly proved to be impracticable. The inclusion of data from people who were educated abroad for some time is simply unavoidable. In thus adapting our sampling criteria, we are in good company: the ICE Hong Kong also includes people who were educated overseas (see Bolt 1994: 18). The restriction should rather be that at least not the entire higher education should have been acquired abroad, and that the authors should not have gained several degrees while staying overseas. We also allow for authors spending some time of the year abroad and writing from abroad, as long as they did not leave the country in the formative years before the age of 18. This more relaxed perspective is closer to the reality of the educated speaker of acrolectal Fiji English.

We also diverge slightly from the original sampling frame in that we decided to include class lessons recorded in secondary schools, although the pupils recorded are partly under 18 but not younger than 16 years old. It was agreed that these, too, are 'aspiring' for higher education, a term that Greenbaum uses in the general ICE guideline to justify the inclusion of students (Greenbaum 1991: 3f.); over-16-year-olds also belong to the target group in that they have already had a considerable exposure to English in their education. In Fiji, English becomes "the sole official medium of instruction after the first three years of primary school" (Tent and Mugler 2004: 751). Without these class lessons it would be extremely difficult to fill the class lesson section of the corpus because the lectures and tutorials from the University of the South Pacific (USP) also pose their problems for inclusion in the corpus. While the existence of a university in Fiji makes data collection easier, its rather special role as a regional institution has to be taken into account. USP is jointly funded by the governments of twelve island countries in the South Pacific, and this multi-national, multi-lingual and multi-cultural nature is reflected in its staff and students. It has altogether fifteen campuses in the twelve member states and also offers pro-

grams through Distance and Flexible Learning, which implies that students do not have to be present on campus.[5] All these factors prove difficult in the course of recording classroom discussions at the university, where several restrictions apply to the recording situations, such as the nationality/eligibility of teachers and students present.[6] Other ICE teams currently working on the compilation of an ESL/ESD corpus[7] too have relaxed the age restriction of the general guidelines for the recording of class lessons.

Expatriates from Fiji, i.e. people who are not residents of Fiji at the present or spend most of the year abroad, should not be part of the corpus, but it is not easy to find out who is an expatriate, as it is not easy to find out in general how many years somebody has spent abroad. We realized that for instance letters to the editor, which we included in the press editorial section to expand our number of authors, are often written from abroad, perhaps – so to speak – to strengthen the ties with the home country. Here it proved vital to check the content of the letters and the name of the town given underneath the letter. If the information clearly pointed towards an (Indo-)Fijian expatriate as the writer or other non-eligible authors, the letter was excluded.[8] But if such information cannot be retrieved easily, where do we draw the line in our efforts to discover it? This is a general issue to which we will return.

With 98 per cent of the population in Fiji speaking English as a second language (Tent 2001: 210), it seems straightforward that ICE-Fiji should be an ESL corpus. But does that mean we should exclude native speakers of English entirely? The two biggest ethnic groups in Fiji are the Fijians and Indo-Fijians[9] (who in 1996 comprised 95 per cent of the whole population),[10] but there are also the Part-Europeans, descendants of Fijians and European immigrants, who generally speak English as a first language (Tent 2001: 210). Their variety of English is close to Fiji English as it is spoken by Fijians and it is definitely distinct from British or American English (Tent and Mugler 2004: 753), which makes it not only interesting to investigate – it is also part of the local choice of English. In accordance with the real-life situation, in which the proportion of Part-Europeans is only 1.5 per cent of the total population (Tent 2001: 210), we decided to make sure that most of the data stems from Fijians and Indo-Fijians.

We have thus come to another important decision to be made, namely whether or not to base the proportion of texts written in Fiji by people of different ethnic backgrounds on a demographic reality. In the 1996 census, 51 per cent of the population were Fijian, 44 per cent were Indo-Fijians (Tent 2001: 210). We therefore attempted to include an equal proportion of texts written by Fijians and Indo-Fijians, including texts of authors of other ethnic background only when found by chance.[11] However, due to the political instability after the first

military coup in 1987, which also resulted in continuing hostility against the Indo-Fijian population, the number of Indo-Fijians in the country has been rapidly declining; in the 2007 census they constituted only 37 per cent of the whole population (http://www.statsfiji.gov.fj/).[12] The ratio of texts by the different population groups included in the corpus therefore does not reflect the current demographic proportion but a demographic reality of a time when compilation started in earnest. But what also needs to be considered is that Indo-Fijians actually produce more literary and, in some fields, academic texts than Fijians. It was decided that the reality of text production has to be taken into account as well, which goes some way towards our original aim of an equal proportion of texts produced by Fijians and Fiji Indians. Also, for a study of variation between different speech communities, it is important that we keep a 50:50 ratio.[13] Like the ICE-NZ team, we also try to attempt a gender balance whenever that information is given (cf. Holmes 1996: 165), but if data is scarce we cannot pursue this goal.[14]

Another important issue is that Fiji is a country with a population of roughly 840,000 people (http://www.statsfiji.gov.fj/). In other words, this is a fairly small community and the same people keep appearing in different text categories as there is only a limited number of people in Fiji who compose texts.[15] The L2 setting already limits our range of domains (see Schmied 1996: 185f. and Section 3.1 in this paper), but what is perhaps of more concern is that we may limit our focus to a small part of an already small community. However, this is in line with the general ICE guidelines, according to which only educated speakers should be included, and we would argue that ICE Fiji will still be a representative sample of educated Fiji English, even if some authors are sampled several times. Still, it is one of the details that have to be documented in the manual for future researchers to understand the nature of this corpus.

We do not always get all the sociolinguistic background information we would wish for; it is possible that we find no information on an author's ethnicity, education or whereabouts. In instructional writing e.g., authorship is seldom acknowledged, which is a significant problem because texts in this genre are often heavily plagiarised. How far do we try to trace that information? When do we stop and acknowledge we do not know much about the author, but decide to include him or her nevertheless? As Schmied notes "it is impossible to look behind every text's production history" (1996: 187). We therefore need to find a pragmatic solution to this problem. Our approach has been to research as much background information as possible on the texts to be included. We exclude texts whenever there is a suspicion or evidence that the text was plagiarised or the author does not meet the criteria detailed above. But in the end, there is

always a compromise involved as the amount of time that can be spent on a single text is limited. Most information on authors and plagiarised texts can be found by means of a simple Google search or via the plagiarism software *Docoloc* (which is also Google-based) for whole documents.

2.2   The timeframe

While the collection process for both written and spoken data is still in progress, data for ICE-Fiji may have been produced as early as 1990. This time span of well over 15 years might raise difficulties because ongoing language change might thus have a skewing effect on our data. However, this actually only really applies to the creative writing section and some direct conversations in the spoken component;[16] most other data stem from after 2000.[17] As Fiji is a society with a strong oral tradition not much literature has been published generally, and even less so in recent years, so older data had to be used for the creative writing section. In our attempts to collect novels and short stories for ICE-Fiji we even discovered that the most active phase of literary production in English in Fiji was before 1990 and that, therefore, a considerable number of texts fall outside our sampling frame.[18]

The time of text production ideally coincides with the actual time of compilation; to gain spoken data of an earlier date is nearly impossible. ICE-Fiji will therefore mainly use data from after 2000 and not from the time span of 1990 to 1994, which was the original timeframe envisaged for the ICE corpora (Nelson 1996: 28). However, as this is in line with the corpus design of all the other new ICE corpora (see this volume), comparability with all these corpora is provided. The recency can also be an advantage, especially with regards to spoken data, because they are more likely to include features like new quotatives (see Mair 2009).

3   Collecting the data: Problems and challenges

After having decided on major adjustments to the original corpus design outlined above, we set out to collect the suitable data. The Internet was used to download online editions from *The Fiji Times* and *The Fiji Sun*. It was also used as a research tool to gain information on authors and the availability of books in libraries worldwide. Jan Tent kindly allowed us to include some of the recordings he had made for his PhD on Fiji English. Contacts were set up in Fiji and several extensive field trips have been undertaken by the compilers since 2005 to collect both written and spoken material. But as soon as we embarked on the data collection we found that the broad range of text types of the ICE family,

envisaged by the founders of the ICE project as being available in "all ICE countries" (Nelson 1996: 29), was problematic. As noted by Schmied (1996: 185f.), second language environments tend to present difficulties with regard to the availability of certain text types. Also the quality of the accessible data can at times be questionable.

3.1   Restricted availability of data

Fiji is not only a small country with a traditionally oral culture and only a limited number of people actually producing written material. People in Fiji also do not necessarily use English extensively in their private conversations. Tent and Mugler, who undertook two surveys on language use and attitudes in Fiji in 1993 and 2005, found that at least 90 per cent of the parents use one of the local languages (Fijian and Fiji Hindi) in conversations with their children. They also found that English is not always used as a *lingua franca* between Fijians and Fiji Indians but that Indo-Fijians also switch to Fijian to converse with Fijians (Tent 2009). The sociolinguistic settings of the two main community languages in Fiji might also influence the collection of social letters. Fiji Hindi is not a written language. Indo-Fijians often have only limited competence in Standard Hindi, which is considered very formal. So the domain of social letters in the Indo-Fijian community is likely to be dominated by English. Fijians, on the other hand, are more likely to write private letters in Fijian.[19] These are only two examples of several which demonstrate that the impact of English on the socio-linguistic reality in Fiji is, albeit strong, nevertheless confined to certain domains. Owing to the changed reality of long-distance communication, it was decided that the new ICE generation (see this volume) should include e-mails in the letter category. We are also still optimistic that we will find enough situations of natural conversation in English to be recorded for ICE-Fiji.

Another restriction on the availability of texts has to do with the fact that local mass media production is limited. Until recently, there was only one local free-to-view television channel in Fiji, *Fiji One*, which started broadcasting in 1991.[20] It broadcasts news on the Internet (http://www.fijitv.com.fj/), and features discussion rounds and local sports commentaries. On the whole, however, locally produced material is scarce, with most of the broadcast material coming from Australia (ABC), New Zealand (e.g. the highly popular soap *Shortland Street*) and America (also mainly soaps).[21]

As a small developing country, Fiji only has limited access to publishing means, with the typical time lag in production facilities when compared to industrial countries. There are three English-medium national newspapers in Fiji, *The Fiji Times, The Fiji Sun* and *The Daily Post,* all of them produced in

Suva. Until 2004, there were only a few locally produced magazines or journals, such as *Islands Business* or *The Review*. We automatically ended up with data from only three different newspapers and magazines, and a "geographical spread by including national, regional and local newspapers" as has been done for ICE-GB (Nelson 1996: 32) would not have been possible at the time. We are fortunate, however, that since 2006, several lifestyle magazines have been launched and have since picked up in numbers (e.g. *Fiji Living, Marama, Turaga, Mai Life*). The last few years have also seen rapid changes in the spread of new technologies in Fiji: a second, private, free-to-air television channel was established in 2008, producing three local shows (http://www.tv.com.fj/); *Fiji Television* has also expanded its range and online accessibility of local productions, and there are radio live-streams and downloadable broadcasts on the web. Adding this new data to the corpus now enables us to gain a more diversified sample of Fiji English.

However, a note of caution is advisable at this point. Extensive use of the Internet for the acquisition of suitable corpus material may result in a slight skewing effect: obviously, not all texts produced in Fiji within a certain time span will be found on the WWW and the acquisition of data through this channel can lead to an overrepresentation of material considered 'worthy' of a global audience. But through our field trips we have been able to ensure the inclusion of more 'locally' produced and circulating texts.

Copyright issues placed another restriction on the availability of data. At the moment, data are included in the corpus, whether copyright clearance has been obtained or not. It therefore may happen that some data have to be excluded at a later stage if copyright clearance cannot be gained. Alternatively, we may be forced to restrict public access to parts of ICE-Fiji. An added difficulty may be that the notion of copyright cannot easily be translated into the Fijian cultural context. Therefore, oversampling of suitable texts would be an advisable safeguard.

3.2   Restricted quality of the data

Finding data that is eligible to be included in the corpus is one thing, but the quality of that data also has to be considered. At an early stage, we had agreed to focus on texts published in Fiji in order to reduce editorial influence from speakers with first language backgrounds. Publishing in Fiji is mainly undertaken by the newspaper publishing houses, the Government printer and the University of the South Pacific. The latter was our main source for publications of academic texts in the form of monographs and journal articles. A closer look revealed, though, that both the Institute of Education and the Faculty of Business and Eco-

nomics, which rank among the most prolific within the university, employ editors or 'publications officers' from England and New Zealand, respectively, with the additional twist that both had been living in Fiji for a very long time and had specialised in detecting "the commonest mistakes made by students in Fiji schools and tertiary institutions" (Pene 2003).[22] This shows that it is vital to have information on local editing procedures. Luckily, the editors of *The Fiji Times* and *The Fiji Sun*, for the years from which we collected data, seem to have been Fijians. Also, some of the publishers have less strict ideas about editing than others. In general, editing in the genre of academic writing and newspaper articles is common practice (see Nelson 1996: 32). The advantage of a broad range of text types in the ICE family is that the edited text types can be compared to spoken material and unedited written material. The problem with edited material was also one reason why we decided to include MA theses and doctoral dissertations as 'ready-to-be-published' material in the academic section of the corpus. Supervisors who are often native speakers of inner-circle varieties are likely sources of non-local language use for these texts, but at least the potential of such influence is easy enough to trace.

When collecting data we also have to ask ourselves how reliable our sources are. In instructional writing, for instance, plagiarised material from the Internet or documents from international organizations generally seem to be favoured over local authorship, but this (for obvious reasons) tends not to be documented. Similarly, student essays contained a fair amount of plagiarised material. This has to be declared as extra-corpus material, and if larger passages of the text are plagiarised, the text has to be excluded from the corpus altogether. That, of course, also means that oversampling during field trips is necessary as it is not clear beforehand whether a text can be used or has enough suitable passages to provide a sample of 2,000 words.

On a more critical note, press freedom, and thus access to Internet data, has been significantly influenced by the most recent political developments in Fiji: the 2006 military coup d'état and what has been termed 'the 2009 constitutional crisis'. Following April 10, 2009, public emergency regulations require military presence in every newsroom, and while most foreign journalists have left the country, local journalists are expected to adopt what is called "the journalism of hope" (http://www.radionz.co.nz/news/stories/2009/04/15/1245aa0f464). Radio stations, blogs and Internet cafes have apparently been shut down, and opinionated or uncontrollable journalistic text types such as editorials, news reports or live call-in programs are dwindling. Material taken from the Internet or other public sources after April 2009, therefore, is likely to have been censored (see Robie 2009). This actually makes it even more attractive to find new sources of non-

published or not yet published material, even though these texts have a smaller impact in the community, as they are only available to a limited readership.

Another problematic text category is that of parliamentary debates, because in the 2006 coup, Parliament was dissolved and has not been re-opened since. Hansards from earlier sessions of the Parliament may have to be used instead, although Hansards are usually not entirely accurate transcriptions (see Mollin 2007).

4   After the data collection: The issue of text categorization

Following data collection, the next step is to check which categories of the ICE framework can be filled with these texts.[23] It quickly becomes obvious that text categorisation is, to a certain extent, culture-specific and that the tension between representativeness and comparability results in a balancing act. The textual universe of Fiji English cannot easily be translated into western categories of text types. While some categories of the ICE family may not be prominent in Fiji, such as social letters, there will be other text categories, such as autobiographic writing, which are more predominant in Fiji than in, say, Europe. Moreover, text categories are interpreted differently in Fiji (see Section 4.2).

4.1   Non-existent text categories

The restricted availability of personal letters has already been mentioned. Apart from the restricted usage of English in letter writing, and the reluctance of potential informants to hand such a letter over to an unknown linguist, letter writing does not play an important role in Fiji society. This is predominantly an oral culture and it is very important to meet with family and friends in person to strengthen ties. With adding email communication to the data-base we can fill the category, but it has to be pointed out in the manual later that, on the whole, the category is not truly a representative one for text production in Fiji.

The political instability in Fiji effectively eradicates any spoken data from legal domains. This does not only refer to the parliamentary debates mentioned above. In 2009, the Fiji Court of Appeal ruled that the 2006 coup d'état was illegal, which resulted in the President's abrogation of the Constitution, in the course of which he also deposed the judiciary and nullified all judicial appointments (http://www.fiji.gov.fj/publish/page_14712.shtml). Data collection in the categories of legal cross-examination and legal presentation has to wait until new judicial officers are appointed. The text category is not non-existent *per se* but the political climate turns it into a genre that is either not productive or currently not accessible.

4.2   Different interpretation of text categories in Fiji

In Fiji, a novel can be a kind of patchwork creation including not only fictional prose but also poetry, autobiographical snippets and excursions into historical events. A student essay may include phrases from secondary literature that were learned by heart, class lessons may sound like a teacher's monologue. A public demonstration may turn out to be an actual demonstration without words. A discussion on TV may look more like an interview as only one person at a time answers the presenter's questions – moreover, the answers themselves strongly suggest that the questions were known in advance and the answers rehearsed. This makes it difficult to distinguish between scripted and unscripted speech or between interview and discussion.

The skills and hobbies section is another text category that is difficult to translate into the local culture if one does not know what typical leisure time activities there are in Fiji.[24] Technology is also a 'western' category that needs a new interpretation in a developing country. All these differences are likely to be culture-related, i.e. how creative writing is interpreted, how a student perceives the task of writing an essay etc. That a lesson may consist of a teacher's monologue is not surprising as in this culture pupils are ranked lower and, therefore, do not talk back (see Schmied 1996: 189). The participant in a discussion politely waits his turn; nothing is left to chance so that the discussion does not become face-threatening for any of the participants. A pastime event in the Fiji context would be something like setting up a kava session with story-telling or cooking for a family feast; an example of a text for the technology section would be the detailed description of how to build a canoe.

In general, we have to trust the local label; a discussion is a discussion if the community identifies and names it as such, even if the rules of discourse are different to ours. If someone is reading from a draft or has obviously memorized his or her statement, it is a scripted speech. In taking over these local interpretations of the text categories a certain level of representativeness, a mirroring of local usage, can actually be maintained.

5   *Conclusion and outlook*

Some problems we encountered in the compilation of the ICE-Fiji component are typical problems of compiling an ESL corpus, such as the restriction of the usage of English to certain domains, considerable exposure to ENL through television, native speakers as editors, new technologies only just becoming available, "political sensitivity" (Schmied 1996: 185), culture-specific interpretations of text types. These issues are also encountered and described by

Schmied (1996) in the case of ICE-East Africa. He concludes that "[d]espite intensive discussions [...] the ICE design is not geared towards ESL corpora primarily, but can be used as a convenient framework – provided that some special, sociolinguistic issues in corpus compilation are taken into consideration" (Schmied 1996: 185).

In addition to these more general ESL-related issues, we addressed the issue of balancing representativeness versus comparability in the specific context of Fiji. Firstly, Fiji provides a cultural setting that is different from the Western background which is implicit in the original ICE corpus design: it is a partly oral culture, so that certain text types are not routinely used (or English is not used for text production in some situations). Secondly, Fiji is a relatively small community with respect to all users of English, but even more so if the restriction to acrolectal speakers is taken into account: thus, the number of potential contributors is significantly limited. Thirdly, the political situation in the country is one that increasingly includes censorship.

"Absolute representativeness is an unattainable ideal", as Mukherjee (2004: 114) points out. The general motivation behind all our solutions was to carefully navigate towards the ideal, namely compile a corpus that is as representative as possible and at the same time ensure maximal comparability with existing and, especially, with emerging ESL/ESD corpora. Towards this end, we have also been collaborating closely with other teams that are currently compiling ICE corpora of New Englishes. One of the main aims of this article (as well as ongoing work on the manual that is to accompany the corpus) has been to document possible solutions and to make the process of compilation transparent to future users of the corpus.

To sum up, what we need in order to make the ICE project a success in providing an adequate database for the study of English worldwide is:

- a consensus among all ICE teams on fundamental issues
- a common agreement on changes in the general ICE framework, where necessary, to include the new generation of ESL/ESD ICE components
- a better coordination of all individual ICE projects
- a comprehensive documentation of decisions past and present of all ICE teams, in manuals that go beyond a mere listing of filenames and markup conventions

A difference between the old and the new ICE generation is certainly unavoidable because the new ICE corpora are all ESL corpora of New Englishes and the core of their data stems from after 2000. But if we all work together as closely as

possible, if we take the issue of representativeness versus comparability seriously and make our decisions in the compilation process as transparent as possible, we can make sure that the ICE corpora family becomes invaluable and indispensable for future research on English as a global language.

Trying to compile a corpus of Fiji English as a representation of the current usage of English in Fiji is like trying to catch a butterfly in midair. Whatever data you collect you are always one step behind the current usage. Fiji English is a newly developing variety, and thus in a way even more evasive than language use is generally, with the sociolinguistic situation in flux and the usage of English still restricted to certain domains.[25] With the current political situation there is not yet a pan-ethnic national identity that could be expressed with Fiji English (see Schneider 2007: 118). In his model on the development of new Englishes Schneider locates Fiji English in phase 2 (exonormative stabilisation) but slowly moving to phase 3 (nativisation), with first hints of structural nativisation (2007: 114–118). While we as the compilers of ICE-Fiji have to come to terms with a yet restricted availability of data for the moment, we can also already see these dynamics, the expansion of the usage of English in Fiji and we see some of our problems in the compilation of ICE-Fiji diminish over time. So on the one hand we are trying to keep up with ongoing sociolinguistic and structural changes, but on the other hand we can also say that we have the privilege of watching and documenting a variety on the move: this is Fiji English in the making.

Completion of the written part of ICE Fiji is envisaged for the end of 2009, with the exception of the category of letters (both personal and business) and instructional writing. The ICE-Fiji team is currently planning a student fieldwork trip to the University of the South Pacific for 2010, during which students will be set up with tandem partners to record face-to-face conversations. It is hoped that this will be a major step towards completion of the data collection stage of ICE Fiji.

Notes

1. We would like to thank Jan Tent for valuable comments on an earlier version of the article.
2. Lena Zipp would like to thank the German Academic Exchange Service (DAAD) for a four-month grant she was given to conduct data collection for her PhD project in Fiji.
3. For the general design of the ICE family, see Nelson (1996) and Greenbaum (1991). For the current state of ICE-Fiji, see Tables 2 and 4 in the appendix.
4. 'ESL' stands for 'English as a second language'.

5. See http://www.usp.ac.fj/index.php?id=usp_introduction, last accessed 14.07.2009.
6. Lynch and Mugler, however, claim that on the main campus in Suva 75 to 80 per cent of the students are Fijians or Indo-Fijians (Lynch and Mugler 2002: 79). Recordings of a classroom situation on the main campus for ICE-Fiji are therefore not entirely out of the question.
7. 'ESD' stands for 'English as a second dialect'.
8. It should also be pointed out that the notion of an 'expatriate status' is centred on a Western perspective. At least for the Fijian population we know that emigrants will always stay in close contact with their family in Fiji, i.e. participate in regular family reunions. First generation emigrants also often return to stay as soon as they retire (see Biewer, in preparation).
9. The label for the descendants of the indentured Indian labourers in Fiji varies: alongside *Indo-Fijian* (which stresses the Fijian part of the identity) people from this ethnic group also use the label *Fiji Indian*. We use both terms in this article without, however, making any political claims.
10. There are other minorities in Fiji who speak English as a second language: other Pacific Islanders, Rotumans, Chinese and Part-Chinese (see Tent 2001: 210).
11. The ICE-NZ team too made the decision to try for a representative sample of Maori people in the corpus but they did not try "to include representative samples from [other] particular ethnic groups" (see Holmes 1996: 165).
12. In 1987 Sitiveni Rabuka staged two military coups claiming that the newly elected government, that was perceived to be under Indo-Fijian dominance, would "undermine Fijian land rights, the integrity of Fijian society, and the sovereignty of its people" (Lal 1992: 298). Ethnic prejudices are stirred up again and again by politicians who are seeking power and two more military coups have taken place since 1987, one in 2000 and one in 2006. The politically highly unstable situation, among other things, causes more and more people to leave the country. The most recent development is the constitutional crisis of 2009.
13. For a first description of some linguistic differences between the speech communities, see Mugler and Tent (2004) and Tent and Mugler (2004).
14. It is not uncommon that the distribution in a corpus is skewed towards females, as women usually are more open to the idea of being recorded and more willing to contribute some of their writing. Also see Sigley and Holmes (2002: 154), who comment on the *Wellington Corpus of Written New Zealand English* having more female authors than male authors in the category of fiction.

15. For instance one of the teachers we recorded during a class lesson also contributed a short story to the corpus. It is also not unusual that authors write on a certain subject in an academic journal and then publish a feature on the same topic in a magazine or a newspaper.

16. These are the conversations that have been taken over from the original ICE-Fiji project.

17. It is a fact that the compilation of a corpus takes a long time, in particular when the compilation is not done locally. ICE-Jamaica, for instance, includes texts produced between 1991 and 2004 (see table accompanying the pre-release of ICE-Jamaica).

18. This even affects apparently new publications which at times turn out to be reprints from the 1980s.

19. A study conducted by Thor May in 1990 in Suva, the capital of Fiji, recorded that out of 834 respondents 76 per cent wrote at least one letter per year, of which 48 per cent were written in English, 21 per cent in English and Fijian, 16 per cent in Fijian, but only 3.7 per cent in Hindi and 6.3 per cent in English and Hindi (http://thormay.net/lxesl/tech7.html, last accessed 14.07.2009).

20. Since 1994, it has been produced by the local company *Fiji Television*.

21. As a consequence English is the dominating language on TV. But there are a few shows in Fijian and Fiji or Standard Hindi; also the main news in the evenings are broadcast three times, once in English and once in the two other major local languages (see Mangubhai and Mugler 2003: 415).

22. One of them (Frances Pene) even published a handbook called *Write it Right* (2003). The quotation was taken from the blurb on the back cover.

23. See Tables 1 and 3 in the appendix for the different text categories in the ICE family.

24. Mountain biking, yachting, and golf, for instance, are typical Western pastimes that are not widely practised in Fiji. Fishing and mat-weaving, on the other hand, can be seen as hobbies but are also practised as necessary, every-day skills.

25. Obviously, there are different sub-varieties of Fiji English (FE), e.g. Fijian FE, Indo-Fijian FE, Chinese FE, Part-European FE, Rotuman FE etc. And each of these has a continuum from basilectal to acrolectal levels.

References

Biewer, Carolin. In preparation. South Pacific Englishes. The dynamics of second language varieties of English in Fiji, Samoa and the Cook Islands. Post-doctoral thesis. University of Zürich.

Bolt, Philip. 1994. The International Corpus of English project – the Hong Kong experience. In U. Fries, G. Tottie and P. Schneider (eds.). *Creating and using English language corpora. Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora, Zürich 1993*, 15–24. Amsterdam: Rodopi.

Bolt, Philip and Kingsley Bolton. 1996. The International Corpus of English in Hong Kong. In S. Greenbaum (ed.), 197–214.

Greenbaum, Sidney. 1991. *International Corpus of English. The compilation of the International Corpus of English and its components.* London: Survey of English Usage.

Greenbaum, Sidney (ed.). 1996. *Comparing English worldwide. The International Corpus of English*. Oxford: Clarendon Press.

Holmes, Janet. 1996. The New Zealand spoken component of ICE: Some methodological challenges. In S. Greenbaum (ed.), 163–181.

Lal, Brij V. 1992. *Broken waves – a history of the Fiji Islands in the twentieth century.* Honolulu: University of Hawaii Press.

Leech, Geoffrey. 2007. New resources or just better old ones? The Holy Grail of representativeness. In M. Hundt, N. Nesselhauf and C. Biewer (eds.). *Corpus linguistics and the web*, 133–149. Amsterdam: Rodopi.

Lynch, John and France Mugler. 2002. Pacific languages at the University of the South Pacific. *Current Issues in Language Planning* 3 (1): 76–80.

Mair, Christian. 2009. Corpus linguistics meets sociolinguistics. Studying educated spoken usage in Jamaica on the basis of the International Corpus of English. In Th. Hoffmann and L. Siebers (eds.). *World Englishes – problems, properties and prospects*, 39–60. Amsterdam and Philadelphia: John Benjamins.

Mangubhai, Francis and France Mugler. 2003. The language situation in Fiji. *Current Issues in Language Planning* 4 (3/4): 367–458.

May, Thor. 1990. Language in Suva. Language use and literacy in an urban Pacific Community. Paper given at the Australian Linguistics Society Conference at Macquarie University in Sydney, NSW, 1990 http://thormay.net/lxesl/tech7.html (last accessed: 14.07.2009).

Mollin, Sandra. 2007. The Hansard hazard. Gauging the accuracy of British parliamentary transcripts. *Corpora* 2 (2): 187–210.

Mugler, France and Jan Tent. 2004. Fiji English: Morphology and syntax. In B. Kortmann, E.W. Schneider *et al.* (eds.). *A handbook of varieties of English. Volume 2: Morphology and syntax*, 770–788. Berlin and New York: Mouton de Gruyter.

Mukherjee, Joybrato. 2004. The state of the art in corpus linguistics: Three book-length perspectives. *English Language & Linguistics* 8 (1): 103–119.

Nelson, Gerald. 1996. The design of the corpus. In S. Greenbaum (ed.), 27–35.

Pene, Frances. 2003. *Write it right*. Suva: USP Institute of Education.

Robie, David. 2009. Free speech in Fiji. *Pacific Journalism Review* 15 (1): 5–8.

Schmied, Josef. 1996. Second-language corpora. In S. Greenbaum (ed.), 182–196.

Schneider, Edgar. 2007. *Postcolonial English – varieties around the World*. Cambridge: Cambridge University Press.

Sigley, Robert and Janet Holmes. 2002. Looking at *girls* in corpora of English. *Journal of English Linguistics* 30 (2): 138–157.

Tent, Jan. 2001. A profile of the Fiji English lexis. *English World-Wide* 22 (2): 209–245.

Tent, Jan. 2009. Language use and attitudes in Fiji 1993–2005. Paper given at the University of Zurich, Switzerland on 26/05/09 [handout].

Tent, Jan and France Mugler. 1996. Why a Fiji corpus? In S. Greenbaum (ed.), 249–261.

Tent, Jan and France Mugler. 2004. Fiji English: Phonology. In B. Kortmann, E.W. Schneider *et al.* (eds.). *A handbook of varieties of English. Volume 1: Phonology*, 750–779. Berlin and New York: Mouton de Gruyter.

Appendix

Table 1: Number of 2,000-word samples per category in the written compo-
nent of the ICE corpora family

Categories	no. of samples
student essays	20
correspondence	30
academic writing	40
non-acad. writing	40
instructional writing	20
press	30
creative writing	20
TOTAL	200

Table 2: 2,000-word samples collected per category in the written component
of ICE-Fiji (August 2009)

Categories	ICE-Fiji
student essays	20
correspondence	0
academic writing	33
non-acad. writing	27
instructional writing	11
press	30
creative writing	17
TOTAL	138

Table 3: Number of 2,000-word samples per category in the spoken compo-
nent of the ICE corpora family

Categories	no. of samples
dialogues private	100
dialogues public	80
monologues unscripted	70
monologues scripted	50
TOTAL	300

Table 4: 2,000-word samples collected per category in the spoken component of ICE-Fiji (August 2009)

Categories	ICE-Fiji
dialogues private	11
dialogues public	45
monologues unscripted	13
monologues scripted	27
TOTAL	96