

## Dominant Language Transfer in Minority Language Documentation Projects: Some Examples from Brunei

Adrian Clynes  
*Universiti Brunei Darussalam*

Language documentation often takes place in contexts of heavy language contact, where there is a shift in progress from a minority language to another culturally dominant language. For many younger speakers, the language of their parents is increasingly acquired as a second language, and their communication in this second language shows classic transfer effects, where transfer is “[a] general cover term for a number of different kinds of influence from languages other than the target language” on a learner’s acquisition of that target language (Ellis 1994:341). However, transfer can also be seen as a more pervasive phenomenon, “a constraint imposed by previous knowledge on a more general process, that of inferencing” (Schachter 1992:44). Considered in this light, transfer can influence far more than a given learner’s interlanguage. Assumptions, attitudes, and conceptual models associated with a culturally dominant language can all unconsciously influence assumptions made about minority languages. These can, in turn, affect various strategic decisions made in the documentation of such languages, including whether a given variety should be documented, which speakers should be recorded, which text types to collect, what orthography to use, even what constitutes a genuine feature of the lexis, phonology, morphology, and so on. This paper aims primarily to illustrate this phenomenon, and to explore ways of dealing with it. Dominant language influence needs to be taken into account at each stage of the documentation process, minimizing it where it is intrusive, and taking advantage of it where it can be of use.

**1. BACKGROUND.** Language shift from minority languages to a national language, or to another culturally dominant language, is a very real phenomenon in both Brunei (e.g., Martin 1995, Noor Azam OKMB Haji-Othman 2005) and Malaysia (e.g., Bibi Aminah Abdul Ghani & Abang Ahmad Ridzuan 1992; Bibi Aminah Abdul Ghani 2006), as elsewhere. Very often the shift is mediated by code-mixing, as well as by accommodation of older speakers to the mixed language of younger speakers. Code-mixing and accommodation increasingly become the norm (Martin & Poedjosoedarmo 1996; McLellan 2005), and communication in the “pure” minority language becomes less common. This has been the case in Brunei for minority languages like Tutong (Noor Azam OKMB Haji-Othman 2005), Belait (Martin 2005), and Dusun (Kershaw 1994).

With this shift comes a concomitant influence of the culturally dominant language or languages (in Brunei, these include Brunei Malay, Standard Malay, and increasingly, English) on the minority languages themselves. Contact with dominant languages has a range of structural influences on minority language learning, with dominant language elements transferring to varying degrees (Thomason & Kaufmann 1988; Thomason 2001). For many younger speakers, the language of their parents is increasingly acquired as a second language, and their communication in this second language shows what in the second language acquisition literature would be described as classic transfer effects, where trans-

fer is “[a] general cover term for a number of different kinds of influence from languages other than the target language” on a learner’s acquisition of that language (Ellis 1994:341).

Researchers documenting or describing minority languages in heavy contact situations must then address the challenge of what to do with mixed data—texts which contain significant amounts of code-mixing, or data which in other ways shows influence from other languages. Despite the widespread nature of this phenomenon, it appears that such issues have only recently begun to be directly addressed in the language documentation literature (e.g., Aikhenwald, *in press*; Bowern 2010).

The primary focus of this paper, however, is very different: it relates to influence from dominant languages at another level: that of the language documentation process itself, and the way in which such projects are conceptualized and implemented. This, too, can be thought of as a kind of transfer, if we follow those who view transfer as a more fundamental phenomenon: “a constraint imposed by previous knowledge on a more general process, that of inferencing” (Schachter 1992:44).

Considered in this light, transfer-like effects can influence far more than a given learner’s interlanguage. Assumptions, attitudes, beliefs, and conceptual models associated with the dominant language can all influence assumptions made about minority languages, and so they can affect various strategic decisions made in the documentation of such languages. This paper aims to illustrate this effect in a practical way, using examples from documentation work in Brunei on the Tutong, Belait, and Brunei Malay (BM) languages.<sup>1</sup> It shows how such “transfer” from the dominant language can influence answers to very basic questions like these:

- What is this language, actually?
- Which speakers should be recorded?
- Which text types should be collected?
- Who should undertake the documentation? (This includes deciding who should collect texts, and who should transcribe.)
- Should texts in the minority language be edited to remove dominant language content?
- Which items should be included in, or excluded from, a dictionary of the language, or marked as “intrusive” loanwords in transcribed texts?
- What should be included in, or excluded from, a grammar of the minority language?<sup>2</sup>

---

<sup>1</sup> This paper is the result of projects funded by Universiti Brunei Darussalam, and is a revised version of a paper first presented at the 10<sup>th</sup> Borneo Research Council Conference at Curtin University, Miri, July 2010. My thanks to two anonymous reviewers, as well to Cikgu Hj Ramlee Tunggal for helpful comments.

<sup>2</sup> The need to distinguish language documentation as a worthwhile enterprise in its own right, related to—but distinct in many ways from—much of language description, was rightly pointed out by Himmelmann (1998). At the same time, Himmelmann points out that language description issues arise in the transcription of primary data, and are therefore a vital part of that documentation. Deciding on an orthography, for example, requires a prior description of the phonology, while interlinear glossing of texts requires prior morphological analysis. Even accurate gist translations require a thorough understanding of the vocabulary and syntax. This paper therefore treats basic language description issues as falling within the domain of language documentation.

This paper does not aim to discuss general methodological and ethical issues in documentation that are already thoroughly covered by authors such as Mithun (2001) and Himmelmann (2008), among others. Nor does it focus on the phenomenon of influence from dominant languages on minority languages themselves; it is instead concerned with influence on documentation projects, and with how the “transfer” of attitudes and concepts associated with the dominant language can potentially distort the record of the minority language produced by such a project—either in the direction of the dominant language (making the variety recorded seem more like the dominant language than it actually is), or away from it (making it seem less like the dominant language than it actually is).

Such transfer of attitudes and concepts is, however, not always harmful: in some cases it can have a positive influence on documentation. If, for example, the orthography used in transcription is modelled on that of a dominant language, the task of transcribing minority language data is made simpler for workers already literate in the dominant language. Often, however, dominant language influence on documentation can be a negative: it could be, for example, that the same newly devised orthography, based on that of the dominant language, lacks symbols needed for phonemes present in the minority language. This could be because they are not found in the dominant language, or their use may be in some way inappropriate. The record produced would then be inaccurate, making the phonology of the language appear closer than it really is to that of the dominant language (see section 2.9). Alternatively, a zealous “purist” approach could lead a documentarian to consciously or unconsciously edit data, or to downplay similarities with the dominant language, again distorting the record away from the reality of the language (see, for example, section 2.6). Such influences can lead to misrepresentations, and so they can significantly affect the outcome of a language documentation project.

The remainder of the paper considers, in turn, questions like those posed in the list above and illustrates some of the possible ways of answering them—particularly with regard to dominant language influence—considering some of the potential consequences of such answers. All examples are taken from Bruneian data, in most cases from two documentation projects undertaken at Universiti Brunei Darussalam: one on the Tutong language (with Haji Ramlee Tunggal, and Noor Azam OKMB Haji-Othman), and the other on the Metteng Belait language and oral history (with Hajah Asiyah az-Zahra Haji Ahmad Kumpoh). Both languages are endangered, and both are severely under-documented.

For further information on the language situation in Brunei, see Martin (1995), Martin, Poedjosoedarmo & Ozog (1996), Haji-Othman (2005), and Coluzzi (2010).<sup>3</sup>

## 2. INFLUENCE OF “TRANSFER” FROM THE DOMINANT LANGUAGE.

**2.1. WHAT IS THIS LANGUAGE, ACTUALLY?** In the case of minority languages in Brunei, three quite distinct answers to this question are regularly heard, all framed in relation to the dominant language:

1. “A very different language, related to Malay, but not closely, and with very little in common with Malay.” This is the type of answer given by linguists for both Tutong and Belait. Neither language in its “pure” form is mutually comprehensible with Malay, and both share low percentages of cognate basic vocabulary (33% and 29%, respectively, in Nothofer’s (1991) 200-word list; compare English and German, with around 60% cognates).

2. “A dialect of Malay.” This answer is very commonly given in Brunei by many who do not speak the minority languages, and so do not realize just how different they can be. Thus one often hears references to *Melayu Tutong* (‘Tutong Malay’) and *Melayu Belait* (‘Belait Malay’). This is not surprising, given that members of the ethnic minorities are officially recognized as belonging to the “Malay race” (1961 Nationality Act of Brunei, cited in Coluzzi 2010), and that for census and other purposes, the government treats all the indigenous languages of Brunei as dialects of Malay (Martin 1995).

3. “Both of the above, plus a continuum of mixed varieties in between.” In this author’s experience, for many young ethnic Tutongs in Brunei, “basa’/bahasa

---

<sup>3</sup> Brunei Darussalam is a Malay sultanate on the north-central coast of Borneo, an enclave surrounded by Sarawak, Malaysia, with a population of approximately 400,000, and an abundance of languages. The constitution recognizes seven *puak jati* (‘original peoples’) of Brunei, each speaking endangered languages (Martin 1995, Coluzzi 2010): traditional local varieties of Brunei Malay (Clynes 2001), Kedayan, Tutong (Haji Ramlee Tunggal 2005), Belait (Clynes 2005), Dusun, Bisaya, and Murut (Lun Bawang). These are all Austronesian languages: Brunei Malay and Kedayan are dialects of Malay, while the others are only very distantly related to Malay. Martin’s survey work demonstrated that even in the early 1990s, fewer and fewer young people were speaking the *puak jati* minority languages (Martin 1995), which are all also severely under-documented. Speakers of the two languages mentioned in this study no doubt number less than 10,000 for Tutong, and in the hundreds for Metteng Belait. All the minority languages occupy the lower position in a diglossic relationship with Standard Malay and English. Standard Malay (close to that of Malaysia; see Clynes & Deterding 2011) is the language of government, the official media, and, along with English, education (education is bilingual from primary school onward). English is also used in high domains, is a *de facto* second national language, and is commonly code-mixed with other languages; it is also increasingly being used in low domains by younger speakers. While traditional local varieties of Brunei Malay are endangered, its situation as an urban variety is stable (Coluzzi 2010). Other languages spoken in the country include Iban (Coluzzi 2010), Penan, Mukah, various Chinese varieties, Nepali, and other languages spoken in expatriate communities.

Tutong” can mean a range of things, from the quaint, often incomprehensible language their grandparents might speak, through a continuum of code-mixed varieties, to what linguists would identify as clearly Malay sprinkled with a few Tutong lexical items as markers of Tutong identity.

Clearly these three answers involve very different notions of just what a given minority variety is, and so they could potentially lead to three very different types of language documentation projects, with very different outcomes:

1. The first answer will result in a “purist” documentation project, one which aims to focus only on the minority language as spoken in its purest form, with a minimum of dominant language (Malay) influence. Linguistic elements borrowed from Malay into the minority language would likely be considered “noise” in data collected in such a documentation project, and would be ignored where possible.

2. The second answer could lead to the documentation of “Tutong Malay” and “Belait Malay,” varieties minimally different from others spoken in the country, but such projects might ignore the less mutually comprehensible, “archaic” varieties spoken by some. Alternatively, it might well result in no project at all, as Malay is not viewed as being in danger.

3. The third answer would lead to a far more inclusive and ambitious “variationist” type of project, one which would attempt to document the full continuum of language use in the Tutong or Belait communities. (This is, of course, the ideal outcome in a documentation project, cf. Himmelmann 1998, 2008).

**2.2. WHICH SPEAKERS SHOULD BE RECORDED?** This question relates closely to the previous one, as do its possible answers. The answers are, again, in a sense dominant language-influenced, if not driven, since they reflect varying views about the degree to which dominant language influence should be shunned or accepted:

1. A purist answer would argue that only speakers of varieties least affected by Malay should be recorded. Generally these varieties would be considered those spoken by the oldest speakers, as well, perhaps, as varieties spoken in more isolated rural communities. The varieties spoken by younger members of the community, and by speakers living in less isolated areas, would be less likely to attract the interest of documentarians.

2. A variationist answer would argue that, ideally, all members of the speech community—or a representative sample—should be recorded, no matter what degree of dominant language influence is shown in their speech.

Each of these approaches, the purist and the variationist, has its own positive and negative aspects. The purist view results in a relatively uniform variety being recorded,

one which reveals the minority language in all its distinctiveness. Such a variety is of most interest to many social scientists, including comparative linguists and typologists. It is also potentially easier to document, since the pool of potential consultants is smaller than when using a variationist approach. A purist approach was used, for example, in Clynes 2005, which focused on the Metteng Belait of a very small number of elderly consultants. Problems with the purist approach, however, include the following:

1. The more restricted pool of potential consultants leads in turn to fewer—possibly *far* fewer—texts being recorded than in a project open to a wider range of consultants.
2. The exclusive approach of a purist project can unwittingly “send the wrong messages:” it can appear to devalue the varieties spoken by those excluded from the consultant pool, including those of younger speakers who should be targeted by maintenance or revitalization programs. This approach may also encourage negative purist attitudes among the most fluent speakers (cf. Florey 2004).
3. The problem of circularity: how does one identify the subset of speakers of a “purer” variety, in the absence of a study of a wider sample of speakers? One risks arbitrarily selecting individuals, say, on the basis of age, when their speech patterns may not necessarily be most representative of the “pure” variety being sought. This author ran into that problem working on Metteng Belait. After recording several texts, one speaker’s daughter expressed surprise that I had recorded so much of her father’s speech. From her and her siblings’ point of view, his Metteng was not “pure,” since as a youth he had lived for some time in a district to the west, where a different dialect is spoken.
4. A purist approach could be accused of attempting to describe something which has in fact never existed, a “monovariant” language.

The positive aspects of the variationist approach include the following:

1. It gives a truer picture of the overall state of the language.
2. It values the speech of all members of the speech community, and so it can be potentially motivating.
3. It can provide a variety of information on the usage of different generations or groups, which could then be valuable to revitalization programs.
4. It produces a record of most interest to many social scientists, such as sociolinguists and sociologists, while satisfying the comparativists and typologists.

Some negative, or more challenging, aspects of a variationist approach include that it is far more demanding of time and other resources, and that it may involve more difficult

descriptive and analytical tasks as potentially two languages—together with the code-mixing continuum between them—must be described.

**2.3. WHICH TEXT TYPES SHOULD BE DOCUMENTED?** A purist approach would collect only texts from genres recognized as truly native to the minority language culture. Texts in such genres are presumably less likely to be subject to linguistic and other influences from other cultures. Critics would argue that by rejecting innovatory genres—say blog posts or text messages—an artificial, “museum” view of the language is encouraged, which would be increasingly foreign to younger speakers.

The variationist response, allowing any genre, traditional or otherwise, will no doubt lead to more dominant language admixture in the data, particularly in texts from digital genres. It does, however, view the language as a living, evolving entity, and validates the varieties of younger speakers—those vital to maintenance or revitalization programs. From a practical point of view, this approach can be a valuable source of written texts for an otherwise unwritten language. Again, the weblog *Tutong Kita*’ (<http://tutongkita.blogspot.com>), started as an experiment by Noor Azam OKMB Hj-Othman, has become an extremely valuable and accessible source of texts in written Tutong (albeit in a “variationist” understanding of Tutong).

**2.4. WHO SHOULD UNDERTAKE THE DOCUMENTATION?** The observer’s paradox applies in documentation projects as elsewhere—the process of documentation, including the presence of a documentarian, influences the nature of what is documented (Labov 1972, Wertheim 2006). This applies at both the recording/text collection stage and the transcription stage. In both cases, unwanted dominant language influence can be the result. I will discuss each stage in turn in the following sections.

**2.5. WHO SHOULD COLLECT TEXTS?** One such fieldworker effect which has been encountered in recording sessions in the Tutong and Belait projects is consultants’ linguistic accommodation toward the dominant language because a documentation worker is present, someone who for various reasons consultants identify with a dominant language. As a result, they may use higher levels of Malay vocabulary or grammatical constructions in what is otherwise a Tutong or Belait text. They may even switch to the dominant language to offer comments or explanations. One unambiguous example in our data is the occasional use of English vocabulary items, but only when this writer—the only English speaker in the team—was present as sound recordist.

Ideally, of course, text collection will be done by native speakers of the minority language, using that language in all interactions when collecting texts. This can lead to the collection of texts with minimal code mixing, representing genres authentic to the speech community (e.g., conversations rather than interviews, a genre more likely to be associated with the national language as used on broadcast media, and so more likely to “pull” the data collected toward the lexis and structures of that language). Native speakers can of course also bring their own cultural knowledge to the collection context, thus enriching the final product in a variety of ways: they can, for example, help identify the best consultants, the most authentic genres, even particular topics and texts worthy of documentation; in record-

ing sessions they can prompt and guide consultants with relevant questions and comments. All these benefits have emerged in both the Tutong and the Belait projects.

Still, having native speakers collect texts does not guarantee a lack of dominant language influence if, for example, those documentarians are perceived of as representatives of national institutions associated with the dominant language culture, rather than as members of the local community. This seems to have occasionally been the case, particularly in the Tutong project, where the chief fieldworker was a retired schoolteacher who was widely known and respected in the community from his appearances on radio and television, and was now part of a university documentation team. This, together with an associated formal tone at the beginning of recording sessions, seems to be the reason that the initial few minutes of Tutong recordings in particular often contained a good deal of Malay code-mixing. In the case of the Metteng Belait project, both of these factors were less prominent—Metteng Belait is only spoken in one village, and the fieldworker, though a university lecturer, was recording people she had known and lived with since childhood—often her own relatives.

Of course, linguists are also needed at various stages in the documentation process, if only at the planning and data analysis stages. Where they are native speakers of the minority language, so much the better. Often, however, they will not speak the minority language well, only being fluent in the dominant language. When texts are being collected, and particularly in heavy contact situations, linguists should make every effort to minimize any accommodation effects their presence may engender. My own answer to this has been to restrict my role to that of sound recordist, and, like a stage hand in a *Noh* play, attempt to remain “invisible” during the linguistic interactions.

**2.6. WHO SHOULD TRANSCRIBE THE TEXTS?** Peter Austin (2008) points out that transcribers may make a variety of editorial amendments, including additions, deletions, and reorderings. Such amendments can alter the actual text, either making it appear “purer” than it is, or possibly adding dominant language elements; in either case they should be documented in detail. A complication not mentioned by Austin is that such amendments may be made unconsciously.

Where transcribers are also speakers of the dominant language, this can lead to unintentional influence creeping into transcriptions. Such transcribers will often be of a younger generation, literate, and speak a variety of the local language more influenced by the dominant language than that of the speaker recorded. They may, therefore, unconsciously transfer dominant language lexis and structures into their transcriptions. Examples in the Tutong project include Tutong lexical items incorrectly transcribed as their cognate dominant language equivalents. For example, Tutong has /ʔ/ where Malay regularly has final /h/; hence, inaccurate transcriptions like *tujuh* for [tudʒuʔ] ‘seven’ and *puluh* for /puluʔ/ ‘ten’. Similarly, where older speakers have homorganic plosive clusters such as /bp, dt, gk/, younger speakers often use the “cognate structure” from Malay, a homorganic nasal-plosive sequence. This also leads to incorrect transcriptions, such as [babpaʔ] ‘mouth’ being heard and transcribed as *bampa*, [dadtom] ‘fever’ as *dantom*, [magkol] ‘sleep’ as *mangkol*, and so on.

Other transcribers, such as comparative linguists or older native speakers, might unconsciously err in the other, more purist, direction. In the case of the homorganic clusters



just listed, some speakers vary between the two structures  $C_{\text{plosive}}C$  and  $C_{\text{nasal}}C$  such that either may surface in pairs like /babpa?/ ~ /bampa?/ ‘mouth’, /dadtom/ ~ /dantom/ ‘fever’, and /magkol/ ~ /maŋkol/ ‘sleep’. There is a danger that transcribers will give an inaccurate, overly prescriptive version of what “should have” been said, rather than an objective transcription. No matter who the transcriber is, or what their “linguistic orientation” may be, awareness of the issues and critical self-examination are needed to produce accurate transcriptions.

**2.7. SHOULD TEXTS IN THE MINORITY LANGUAGE BE EDITED FOR DOMINANT LANGUAGE CONTENT?** Presumably, in a truly descriptive documentation archive, few would condone the wholesale replacement of dominant language elements with minority language items in recorded audio files—say, by selectively re-recording and editing. When it comes to written transcriptions of recorded texts, however, such editing is not infrequent. In particular, published texts are “improved” by replacing intrusive dominant language content, with such edits not always signalled. In an interesting blog discussion of this issue at *Transient Languages and Cultures* ([http://blogs.usyd.edu.au/elac/2008/05/endangered\\_languages\\_outreach.html](http://blogs.usyd.edu.au/elac/2008/05/endangered_languages_outreach.html)), Peter Austin (2008) argues that both transcriptions and published texts should include a detailed record of editorial decisions. Dominant language content can be easily signalled in written transcripts, for example, with italics (see, e.g., Meakins 2008). This does, though, beg the question of what is a dominant language element and what is an authentic minority language element, which is discussed in the next section.

**2.8. WHAT IS “THE” VOCABULARY OF THE MINORITY LANGUAGE?** Signalling dominant language content in transcriptions requires an understanding of what is and is not an authentic minority language lexical item. There is a continuum from one-off intrusions of dominant language items into an otherwise unmixed minority language text, to fully assimilated loanwords, now felt to be truly minority language vocabulary items. And there is, of course, a large gray area in between. If loanwords are to be signalled in transcriptions, such “gray area” items can be a problem. A similar problem is met in compiling a dictionary of a minority language: how much, and which, dominant language content should be included? Recall that for many younger speakers, “speaking Tutong” includes speaking Malay with just a sprinkling of Tutong items. If we used that as a guide, our “Tutong” dictionary might contain the entire Malay lexicon as a subset of the Tutong language, not to mention those of Arabic and English. Clearly, a more selective approach is needed.

Another source of excessive dominant language influence is the compilation of a dictionary using a dominant language wordlist to elicit minority language items. It is likely that the final product will give a distorted picture of the minority language lexicon—one filtered through the vocabulary and worldview of the dominant language. The small Tutong-Malay dictionary (Dewan Bahasa dan Pustaka 1991a) appears to have been compiled using such a list: 72% of the first 551 headwords (entries beginning with the letters A–G) have Malay cognates. By contrast, a 200-word Swadesh list gives only a 33% cognate level with Malay (Nothofer 1991). There is clearly a high proportion of loanwords from Standard Malay (SM) in the Tutong dictionary. At the same time, it lacks entries for many

high-frequency native lexemes, as well as lexemes expressing distinctively Tutong cultural concepts.

A purist solution is to include, where possible, only “truly minority language” items, avoiding dominant language and other intrusive items, particularly where they appear to be relatively recent loanwords. The first edition of the *Brunei Malay Dictionary* (Dewan Bahasa dan Pustaka 1991b) tried a radical version of this solution (abandoned in the 2007 second edition), which omitted any item with a cognate form in Standard Malay. This approach, though, was too radical: high-frequency basic vocabulary items like *mata* ‘eye’, *aku* ‘1SG pronoun’, and other such items were not listed, since they have clear cognates in Standard Malay. The assumption seemed to be that these cognate words were “the same” in both dialects, and so no record was needed. This is not true, however: *aku* ‘1SG’, for example, has forms not found in Standard Malay, such as the oblique form *kadiaku* and an enclitic subject form *-ku*, each with its own unique morphology and syntactic behavior. Similarly, *mata* occurs in uniquely Brunei Malay compounds such as *mata garanjang* ‘lascivious’ and *kalat mata* ‘sleepy’. It is likely that almost any lexical item will have such unique features, which distinguish it from even “identical” cognates in related dialects and languages. In short, the problem of filtering out unwanted dominant language material is not solved by excluding all items with dominant language cognates.

So which borrowings from the dominant language should be included in a minority language dictionary? Tutong has a long history of contact with Brunei Malay. It would be reasonable to include older borrowings, those no longer seen as foreign, and even more recent borrowings which show evidence of having been assimilated into Tutong—say by semantic, phonological, or morphological adaptations. In Tutong, most older loanwords come from Brunei Malay and not Standard Malay (Standard Malay is, with English, the source of most recent loanwords.). Examples of Brunei Malay-influenced loanwords (from Arabic, via Brunei Malay) include *masgit* ‘mosque’ (compare SM *masjid*), *baang* ‘evening prayer’ (cf. SM *bang*), and *kain kina* ‘type of cloth: literally, Chinese cloth’ (cf. SM *kain cina*); all of these show phonological features now found only in the speech of older speakers of Brunei Malay. Similarly, older English loanwords found in Tutong and Belait are often no longer used in younger speakers’ contemporary Malay: *metuka* ‘car’ (cf. SM/contemporary BM *karita*), *libri* ‘library’ (cf. SM/contemporary BM *perpustakaan*), and *taim* ‘time’ (cf. SM/BM *masa*).

Ultimately, decisions will be made by individual editors about which dominant language items to include in dictionaries, or which to signal in transcriptions. No two editors will arrive at the same decisions, and even the same person will no doubt make different decisions about certain items on different occasions. Still, overall a parsimonious approach is clearly needed, rather than a “carte blanche” one, to produce a dictionary that will be practical and useful for others. A simple rule of thumb might be that most users will be new learners or semi-speakers, people who speak the dominant language, and those who generally rely on the dictionary to check the meanings of unfamiliar terms or lexemes not listed in other dictionaries: this would indicate little or no need to include recent dominant language loanwords.

**2.9. WHICH ORTHOGRAPHY SHOULD BE USED?** Again, both purist and dominant language-tolerant perspectives will answer this question differently. This is one area in which

dominant language influence may be useful, if minority language speakers have acquired literacy via that language. Clynes 2005, a sketch of Belait morphology and basic syntax, was written for readers trained in linguistics, and so it used a “purist,” non-dominant language-influenced spelling in Belait. This sometimes led to marked discrepancies between phonemic spellings and surface phonetics: thus /buik/ ‘flower’ spelled as *buik*, but usually pronounced [bujəq]; and /tuliw/ ‘body’ spelled as *tuliw*, but usually pronounced [tuljew]. This system was not acceptable to native speakers who, being accustomed to Malay orthography, expected a word’s spelling to be a reliable guide to its pronunciation.

Such a “phonetic” spelling system is indeed clearly needed if a language archive is to be usable by the proprietary community. And where that community is literate in the dominant language, it makes good sense to borrow relevant features of that language’s orthography into the minority language writing system. The result will in some cases indeed be a more phonetic than phonemic orthography (for example, Belait /buik/ ‘flower’ spelled as *buyek*). Such a spelling system is in fact valuable in revitalization programs: learners need to know how to pronounce the language, rather than negotiate the intricacies of an abstract (and possibly controversial) phonemic analysis (Seifart 2006:283–284).<sup>4</sup>

In the process, however, the dominant language-based orthography can unintentionally impose or imply phonological categories which do not apply in the minority language. Standard Malay has verbal prefixes *ber-* /bər/ ‘intransitive prefix’, *meng-* /mə(ŋ)/ ‘agentive’, and *ter-* /tər/ ‘non-volitional’, all of which contain the vowel /ə/. Brunei Malay has a three-vowel system with /i/, /u/, and /a/, but no schwa; the Brunei Malay equivalents of those affixes take the shapes /ba/, /ma(ŋ)/, and /ta/. The two editions of the Brunei Malay dictionary, *Kamus Bahasa Melayu Brunei* (DBP 1991b, 2007), list derived verb forms with these affixes; however, they are always spelled with <e> instead of <a>, following the Standard Malay spelling. They therefore implicitly and erroneously claim a four-vowel system for Brunei Malay. (For further discussion of these affixes, see the following section).

Orthographic choices between minority language- and dominant language-influenced forms are also encountered at the level of individual lexemes. Words may have more than one potential orthographic form, with one reflecting the older minority language phonology and the other a form influenced by the dominant language phonology: for example, Tutong *magkol* /magkol/ ‘sleep’ and the alternate (dominant language-influenced) *mangkol* /maŋkol/ ‘sleep’. In such cases, the question arises as to which form should become the headword in a dictionary: the older, more historically authentic form, or the form perhaps more widely used and recognized by younger speakers? Even if both forms are entered as headwords, one will presumably be given lower priority, redirecting the reader to the other (prioritized) entry, in which case a “purist” vs. “dominant language-influenced” choice must be made.

---

<sup>4</sup> On the other hand, as an anonymous reviewer points out, an orthography too similar to the one used for the dominant language may be problematic for some ethnic groups, who want to see their language clearly differentiated from the dominant one. A distinctive orthography may help lend prestige to a minority language as clearly distinct and autonomous, and so help avoid the perception that it is a mere “dialect” of the dominant language.

A further danger with a strongly dominant language-influenced orthography is that the dominant language writing system may not include symbols for phonemes found only in the minority language. In Malay the glottal stop [ʔ] is not phonemic, though it is predictably found in certain contexts, and it is not written in the Malay orthography. Tutong has contrastive /ʔ/, and it is common in the lexis. Trained in Malay conventions, writers of Tutong (including transcribers) normally do not mark the glottal stop, thus failing to distinguish pairs like /tuʔ/ ‘already’ and /tu/ ‘this’, /daun/ ‘sound’ and /daʔun/ ‘leaf’, and /tian/ ‘stomach’ and /tiʔan/ ‘see’. Clearly the dominant language orthography cannot simply be adopted uncritically: it needs to be checked and, if necessary, adapted for phonological fit with the minority language.

The inverse case is also found, in which the dominant language orthography encodes phonemes not found at the phonemic level in the minority language. For example, /h/ is phonemic in Malay, but not in Tutong, where [h] optionally occurs utterance-finally, and non-contrastively, on underlyingly vowel-final words: *nini* ‘quotative particle’ [nini(h)], *itu* ‘proximal demonstrative, this’ [itu(h)]. Such items are commonly transcribed with an *h*, due to the influence of Malay. This encoding of a purely phonetic element can sometimes be useful (e.g., see the discussion of Belait orthography above). However, in this case it leads to confusion. Many Tutong speakers add <h> to *all* occurrences of such words, even where they occur non-finally and are not pronounced with [h], no doubt due to the prestige of the dominant language orthography.

Finally, a quite different source of dominant-language influence on orthography comes in the informal conventions developed for composing text and short message service (SMS) messages. For speakers of many minority languages, practically the only context in which they write their language is when composing SMS messages. They devise informal orthographies for that purpose, and these are no doubt inevitably modelled on similar, dominant language-derived orthographies also used for texting. For example, conventions such “rebus spelling,” where letters of the alphabet are intended to be read as the phonetic value of their “names” (e.g., English *c u* for *see you*), are easily transferred to other languages. In Brunei, the letters of the alphabet are pronounced following the conventions of British English, so <a> [eɪ], <b> [bi], <c> [si], <d> [di], <e> [i], and so on. With the Tutong project, we have found that transcribers commonly use rebus spellings (even after training): particularly common is the use of <e> to represent [i] in spellings such as <ne> [ni] ‘3SG’ or <etu> [itu] ‘proximal demonstrative, this’. An unadvised reader could easily be misled as to the pronunciation of such words.

**2.10. WHICH MORPHOLOGY SHOULD BE USED?** As with phonology, dominant language-influenced analyses of the minority language’s morphology (and no doubt syntax) can affect transcription decisions in documentation, giving a distorted view of the minority language. A potential example is again evident with respect to the Brunei Malay verbal prefixes discussed above, as represented in the dictionaries (Dewan Bahasa dan Pustaka (DBP) 1991b, 2007). The Brunei Malay equivalents of Standard Malay *ber-* /bər/ ‘intransitive prefix’, *meng-* /məŋ/ ‘agentive’, and *ter-* /tər/ ‘non-volitional’ are /ba/, /ma(ŋ)/ and /ta/, respectively. These have quite different allomorphy from the Standard Malay forms. The dictionaries list derived verb forms; however, they list them as taking the Standard Malay affixes, not the authentic Brunei Malay forms. Thus there are entries for “hybrid” Brunei Malay forms such as *mengunjar* ‘look for’ (with the unlisted “pure” form being *maunjar*),

*mengisihi* ‘insert’ (the unlisted “pure” form being *maisih*) (DBP 1991b), *beringat-ingat* ‘remind’ (“pure” form: *baingat-ingat*), *berkurapak* ‘speak, say’ (“pure” form: *bakurapak*), and *terampai* ‘(be) put down’ (unlisted “pure” form: *tahampai*) (DBP 2007). It is easy to imagine transcribers similarly using standard Malay affixes when transcribing non-standard Malay dialects, unless they are aware of the issue. The prestige of the dominant variety can blind transcribers to the realities of the dialect they are recording.

**3. CONCLUSION.** Amery (2009) argues convincingly for an approach to language documentation that draws on insights gained from second language pedagogy, in particular the functional-notional curriculum. The present paper similarly draws on another influential concept from second language acquisition theory, that of transfer (Schachter 1992), arguing that an awareness of transfer effects is needed in the design and implementation of documentation projects. This paper has argued that in situations where there is a shift to a culturally dominant language, assumptions, attitudes, and conceptual models associated with that language can all be unconscious influences on key decisions made in minority language documentation—such as which speakers to record, which text types to collect, what orthographic system to use, and even whether the variety under consideration is worth documenting. Dominant language transfer therefore needs to be taken into account in a conscious way at each stage of the documentation process. If not accounted for, there is a danger that the output from the documentation process may be distorted, making the particular variety recorded seem either more or less like the dominant language than it actually is. Those involved need to be aware of such issues, as well as ways of either minimizing influence or using it where it is potentially helpful, such as in devising an orthography.

Documentation materials should address dominant language transfer issues explicitly, indicating the general approach that has been taken, e.g., purist or variationist, and why. They should give a clear idea of the degree of interspeaker variation in the speech community being studied. Further, they should indicate the various potential sources of dominant language influence on the project, giving the most detailed account possible of the circumstances of data collection, speakers’ biodata and backgrounds, and also those of recorders and transcribers. Transcriptions should be annotated to indicate all editorial decisions made in relation to such influences. Since documenting minority languages often occurs in heavy contact situations, the issues discussed here are no doubt relevant to many documentation projects.

#### REFERENCES

- Aikhenvald, Alexandra Y. To appear. Multilingual fieldwork and emergent grammars. In *Proceedings of the 33<sup>rd</sup> Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society.
- Amery, Rob. 2009. Phoenix or Relic? Documentation of Languages with Revitalization in Mind. *Language Documentation & Conservation* 3(2). 138–148.
- Austin, Peter K. 2008. Glossed texts—the fiddle factor. *Transient Languages & Cultures*, [http://blogs.usyd.edu.au/elac/2008/09/glossed\\_texts\\_the\\_fiddle\\_facto.html](http://blogs.usyd.edu.au/elac/2008/09/glossed_texts_the_fiddle_facto.html). (1 July 2010.)

- Bibi Aminah Abdul Ghani and Abang Ahmad Ridzuan. 1992. Language shift among the Orang Miriek of Miri, Sarawak. In Peter W. Martin (ed.), *Shifting patterns of language use in Borneo*, 131–145. Williamsburg, Va: Borneo Research Council.
- Bibi Aminah Abdul Ghani. 2006. Maintenance and shift in a language of Sarawak: A study of Bintulu. In Chong Shin, Karim Harun & Yabit Alas (eds.), *Reflections in Borneo Rivers*, 111–126. Pontianak: STAIN Pontianak Press.
- Bowern, Claire. 2010. Fieldwork in language contact situations. In R. Hickey (ed.), *Handbook of language contact*, 340–357. Malden, MA: Wiley-Blackwell.
- Clynes, Adrian. 2001. Brunei Malay: An overview. *Occasional Papers in Language Studies* (Department of English Language and Applied Linguistics, Universiti Brunei Darussalam) 7. 11–43. <http://www.ubd.edu.bn/academic/faculty/FASS/staff/docs/AC/Clynes-Brunei-Malay.pdf>.
- Clynes, Adrian. 2005. Belait. In Karl A. Adelaar & Nikolaus Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 429–455. London: Routledge Press.
- Clynes Adrian & David Deterding. 2011. Standard Malay (Brunei). *Journal of the International Phonetic Association* 41(2). 259–268.
- Coluzzi, Paolo. 2010. Endangered Languages in Borneo: A Survey among the Iban and Murut (Lun Bawang) in Temburong, Brunei. *Oceanic Linguistics* 49. 119–143.
- Dewan Bahasa dan Pustaka. 1991a. *Kamus Tutong-Melayu Melayu-Tutong*. Bandar Seri Begawan: Dewan Bahasa dan Pustaka Brunei.
- Dewan Bahasa dan Pustaka. 1991b. *Kamus Bahasa Melayu Brunei*. Bandar Seri Begawan: Dewan Bahasa dan Pustaka Brunei.
- Dewan Bahasa dan Pustaka. 2007. *Kamus Bahasa Melayu Brunei*, Edisi kedua [2nd edn]. Bandar Seri Begawan: Dewan Bahasa dan Pustaka Brunei.
- Ellis, Rod. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.
- Florey, Margaret. 2004. Countering purism: Confronting the emergence of new varieties in a training programme for community language workers. In P.K. Austin (ed.), *Language Documentation and Description*, vol. 2, 9–27. London: SOAS.
- Haji Ramlee Tunggal. 2005. *Struktur Bahasa Tutong*. Bandar Seri Begawan: Dewan Bahasa dan Pustaka Brunei.
- Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161–195.
- Himmelmann, Nikolaus P. 2008. Reproduction and preservation of linguistic knowledge: Linguistics' response to language endangerment. *Annual Review of Anthropology* 37. 337–350.
- Kershaw, Eva M. 1994. Final shifts: Some why's and how's of Brunei Dusun convergence on Malay. In Peter W. Martin (ed.), *Shifting patterns of language use in Borneo*, 179–194. Williamsburg, VA: Borneo Research Council.
- Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1. 97–120.
- Martin, Peter W. 1995. Whither the indigenous languages of Brunei Darussalam? *Oceanic Linguistics* 34. 44–60.
- Martin, Peter W. 2005. Language shift and code-mixing: A case study from northern Borneo. *Australian Journal of Linguistics* 25(1). 109–125.

- Martin, Peter W. & Gloria Poedjosoedarmo. 1996. Introduction: An overview of the language situation in Brunei Darussalam. In Peter W. Martin, Gloria Poedjosoedarmo & Conrad Ozog (eds.), *Language use and language change in Brunei Darussalam*, 1–26. Athens, OH: Ohio University Center for International Studies.
- Martin, Peter W., Gloria Poedjosoedarmo & Conrad Ozog (eds.). 1996. *Language use and language change in Brunei Darussalam*. Athens, OH: Ohio University Center for International Studies.
- Mataim Bakar. 2000. *Morfologi Dialek Brunei: Satu tinjauan kajian lampau*. Dewan Bahasa dan Pustaka Brunei.
- McLellan, James A.H. 2005. *Malay-English language alternation in two Brunei Darussalam on-line discussion forums*. Curtin University of Technology dissertation.
- Meakins, Felicity. 2008. Unravelling languages: Multilingualism and language contact in Kalkaringi. In Jane Simpson & Gillian Wigglesworth (eds.), *Children's language and multilingualism: Indigenous language use at home and school*, 247–64. New York: Continuum.
- Mithun, Marianne. 2001. Who shapes the record: The speaker and the linguist. In Paul Newman & Martha Ratliff (eds.), *Linguistic Fieldwork*, 34–54. Cambridge: Cambridge University Press.
- Noor Azam OKMB Haji-Othman. 2005. *Changes in the linguistic diversity of Negara Brunei Darussalam: An ecological perspective*. Leicester: University of Leicester dissertation.
- Nothofer, Bernd. 1991. The languages of Brunei Darussalam. In Hein Steinhauer (ed.), *Papers in Austronesian Linguistics* 1, 151–176. Pacific Linguistics A-81. Canberra: Australian National University.
- Schachter, Jacquelyn. 1992. A new account of language transfer. In Susan M. Gass & Larry Selinker (eds.), *Language transfer in language learning*, 2nd edn., 32–46. Amsterdam: Benjamins.
- Seifart, Frank. 2006. Orthography development. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation*, 275–299. Berlin: Mouton de Gruyter.
- Thomason, Sarah G. 2001. *Language contact*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Wertheim, Suzanne. 2006. Cleaning up for company: Using participant roles to understand fieldworker effect. *Language in Society* 35. 707–727.

Adrian Clynes  
adrian.clynes@ubd.edu.bn