

Target Article

**Linguistic and social typology:  
The Austronesian migrations and  
phoneme inventories**

PETER TRUDGILL

*Abstract*

*There is a challenging issue for linguistic typology which involves the relationships which might exist between societal type and aspects of linguistic structure. Linguistic-typological studies have provided us with insights into the range of structures available in human languages, but we do not yet have explanations for why, of all the possible structures available, particular languages select particular structures and not others. A legitimate sociolinguistic viewpoint would be that some social explanations may be available. The sociolinguistic factors suggested as being relevant are language contact versus isolation, and community size and network structure. This paper deals with this thesis from the point of view of Austronesian phonology, with particular reference to Polynesian phoneme inventories.*

*Keywords:* areal linguistics, Austronesian, community size, language contact, phoneme inventories, Polynesian, social structure

**1. Linguistic and social typology**

In earlier papers (Trudgill 1989, 1996, 1998) I have suggested that there is a challenging issue for linguistic typology which involves the relationships which might exist between societal type and aspects of linguistic structure. Linguistic-typological studies have provided us with insights into the range of structures available in human languages, into what the constraints on these structures might be, and into relationships between different typological characteristics. But we do not yet have explanations for why, of all the possible structures available to human languages, particular languages select particular structures and not others. A legitimate sociolinguistic viewpoint would be that some such explanations may be available, and that some of these might be social in nature; that is, the distribution of linguistic features over languages may

not be totally random when viewed from a social perspective. The sociolinguistic factors I have suggested as possibly being relevant here revolve around language contact versus isolation, and community size and network structure. They include:

- (i) Small, isolated, low-contact communities with tight social network structures are more likely to be able to maintain linguistic norms and ensure the transmission of linguistic complexity from one generation to another. Such communities are thus likely to be more linguistically conservative, i.e., to show a slower rate of linguistic change, and more likely to demonstrate complexities and irregularities. Though changes are less likely to occur in such communities, these changes may also tend to be of a more marked type, because of the ability of tightly-networked societies to, as it were, force such changes through; and the languages of such communities may therefore be more likely to have more marked forms and structures. Thus, Faroese is a language which is more conservative and more morphologically complex than the closely related Norwegian, but one which has experienced more marked phonological changes (Trudgill 1989).
- (ii) Small, isolated, low-contact communities with tight social network structures will have large amounts of shared information in common and will therefore be able to tolerate lower degrees of linguistic redundancy of certain types. It is possible, for example, that such communities may therefore show a higher incidence of fast speech phenomena, and higher proportions of phonological changes due to the institutionalisation of such fast speech phenomena.
- (iii) On the other hand, communities involved in large amounts of language contact, to the extent that this is contact between adolescents and adults who are beyond the critical threshold for language acquisition, are likely to demonstrate linguistic pidginisation, including simplification, as a result of imperfect language learning.

Note that it is important to distinguish between pidginisation, as a process, and a pidgin language, as only one of a number of possible outcomes of the process. Most often, language contact involving adults does not lead to the development of new (pidgin or creole) languages. In those cases where new varieties do form, however, there appear to be two fundamental mechanisms which are instrumental in their formation, namely the inability of post-adolescent humans to learn new languages perfectly, and the process of focussing (see LePage & Tabouret-Keller 1985), which may occur in certain social and linguistic circumstances.

Whenever adults and post-adolescents learn a new language, pidginisation takes place (Trudgill 1989). Pidginisation consists of three related but distinct processes: reduction, admixture, and simplification. Reduction refers to the fact

that, in pidginised form, there is simply LESS OF a language as compared to the form in which it is spoken by native speakers: the vocabulary is smaller, and there are fewer syntactic structures, a narrower range of styles, and so on. Admixture refers to interference – the transfer of features of pronunciation and grammatical and semantic structure from the native language to the new language, an obvious feature of adult second-language acquisition. Simplification, as is well known (see Mühlhäusler 1977), is a rather complex phenomenon, but it refers crucially to regularisation of irregularities, to loss of redundancy (such as grammatical gender), and to an increase in TRANSPARENCY, by which is meant an increase in forms such as *eye-doctor* as opposed to *optician*, and *did go* as opposed to *went*. Imperfect learning, that is, leads to the removal of irregular and non-transparent forms which naturally cause problems of memory load for adult learners, and to loss of redundant features. This can in turn lead to an often dramatic increase in analytic over synthetic structures. Reduction can be considered as being due to incomplete learning and restriction in sociolinguistic function, while admixture and simplification are the result of imperfect learning.

In some cases, where exposure to the new language is minimal, such pidginisation may be extreme and remain extreme. In certain cases, moreover (see Whinnom 1971), such extremely pidginised forms of language may, in the absence of native speakers of the original language, become important as a lingua franca, a means of communication between two or more groups who have acquired the pidginised forms and who have no native or other language in common. In these cases, focussing may well occur: the pidginised forms of the original language may acquire stability, with widely shared norms of usage, and a new language variety, a pidgin, will have come into being.

Typically, then, a pidgin is a stable language, without native speakers, which is the outcome of reduction, admixture, and simplification of some source language, and where, also typically, pidginisation has occurred to such a degree that mutual intelligibility with the source language is no longer possible.

The important thing for our purposes, however, is that the simplification that occurs as part of pidginisation will be found in many contact varieties other than genuine pidgins and their creole developments; note that I am following here the pidgin-creole scenario favoured by McWhorter (2000) and Hancock (1988), rather than Mufwene (2001). The most extreme forms of such native-speaker languages are of course creoles (McWhorter 2001), but simplification is by no means confined to them.

## 2. The Austronesian expansion

In this paper I consider this thesis concerning the link between societal type and linguistic type from the point of view of Austronesian phonology, with

Table 1. *Approximate chronology and possible homelands*

Proto-language	Chronology	Possible homeland
Proto-Austronesian	4000–2500 BC	Taiwan
Proto-Malayo-Polynesian		Philippines
Proto-Central-Eastern Malayo-Polynesian		Cenderawasih Bay, Irian Jaya
Proto-Eastern Malayo-Polynesian	1500–1300 BC	Moluccas (Maluku)
Proto-Oceanic		Bismarck Archipelago, Papua New Guinea
Proto-Central Pacific	1000–800 BC	Fiji
Proto-Polynesian	500 BC–200 AD	Tonga-Samoa-Uvea- Futuna

particular reference to Polynesian phoneme inventories. As is well known, the Austronesian language family includes more than 1,200 members (Pawley & Ross 1993: 429); and it is found over a bigger area of the globe than any other family, stretching from Madagascar, in the western Indian Ocean, to Easter Island, in the eastern Pacific, and covering 70 degrees of latitude, from Hawai'i in the north to New Zealand in the south. Of particular relevance to this paper are the migrations which led to this enormous spread, particularly the remarkable settlement of the islands of the remote Pacific Ocean by the Polynesians.

Current thinking (Blust 1999, Pawley 1999) suggests that the pedigree of the Polynesian languages is as follows. Proto-Austronesian split into several first-order subgroups (Blust 1999 thinks as many as ten). All but one of these groups are located in Taiwan. The exception is Malayo-Polynesian, which thus contains all the Austronesian languages spoken outside of Taiwan. Malayo-Polynesian then split into a number of groups, including Central-Eastern Malayo-Polynesian. Next, Central and Eastern Malayo-Polynesian split from one another, and subsequently Eastern Malayo-Polynesian subdivided into a number of groups, one of them being Oceanic. The Oceanic languages include all the languages that were formerly labelled as Micronesian, Melanesian, and Polynesian. One of the sub-branches of Proto-Oceanic is Central Pacific (Geraghty 1983, 1986; Pawley 1996a, 1996b, 1999). Proto-Central Pacific was an extensive dialect chain. Polynesian is thought to have diverged from the eastern end of this chain (Geraghty 1983) and Rotuman from the northwestern end of the chain (Pawley 1996a).

The approximate chronology of these developments (Pawley 1996a, 1999; Pawley & Ross 1993; Kirch & Green 2001) and possible homelands for these proto-languages (Pawley & Ross 1992, Kirch & Green 2001) are given in Table 1.

The well-known although not entirely accepted (see Marck 1999, 2000) outline of the structure of the Polynesian family can be portrayed, in a somewhat simplified form, as follows. Proto-Polynesian split initially into Tongic and Nuclear Polynesian. Nuclear Polynesian then subdivided into Samoic-Outlier and Eastern Polynesian. The main Samoic languages are Samoan, Uvean, Futunan, Tokelauan, Pukapuka (Cook Islands), and Tuvaluan. The Outliers are languages which are the result of, as it were, backwards migration of Polynesian speakers into Micronesia, Vanuatu, the Loyalty Islands, and the Solomons (see further below). The Eastern group divided into Central Eastern and Rapanui (Easter Island). The main Central Eastern Polynesian languages are Marquesan, Mangarevan, and Tahitian (French Polynesia), Rarotongan and Penrhyn (Cook Islands), Hawai'ian, and New Zealand Maori. Tahitian, Penrhyn, Rarotongan, and Maori form the Tahitic sub-branch; Marquesan, Mangarevan, and Hawai'ian the Marquesic sub-branch.

It is probable that the split into Tongic and Nuclear Polynesian occurred 100 BC–200 AD, with a probable homeland for Nuclear Polynesian being Samoa. The migration eastwards that led to the establishment of eastern Polynesian, perhaps in the Tahiti area, could have been around 500 AD (Kirch & Green 2001). The chronological end point of the major colonisation of the Pacific by Austronesians lies with the settlement of the North Island of New Zealand, perhaps around 1000 AD, probably from Rarotonga, by the ancestors of the Maori, and the subsequent settlement of parts of the South Island. The final, minor acts of Polynesian expansion probably involved migration from Mangareva to Rapa in the southern Australs, around 1300 (Fischer 2001), and from New Zealand to the Chatham Islands, probably around 1400. The Moriori language of the Chathams is now extinct.

The chronology for the development of New Zealand Maori might thus be something like this: Proto-Polynesian 500 BC, Nuclear Polynesian 200 BC, Eastern Polynesian 0–500 AD, Central Eastern Polynesian c. 500 AD, Tahitic 500–1000 AD, Maori 1000 AD. All of the Central Eastern languages retained a good degree of mutual intelligibility at least until the beginning of the 19th century.

### **3. Phoneme inventories**

This dispersal over a period of more than 5,000 years of the Austronesian language family into the Pacific, culminating in the settlement of New Zealand, was accompanied by a remarkable series of phonological developments involving phoneme inventories. Two of the Polynesian languages at the extreme geographical end points of this dispersal, Hawai'ian in the far north and Rurutu in the far southern Australs, are remarkable in that they have very small phoneme inventories indeed (Maddieson 1984). In particular, both of them have only

eight consonants. South Island Maori (now extinct), which also lay at the extreme end point of a migration route, is also sometimes claimed to have had eight consonants (Biggs 1978: 708–709), but nine seems more probable (Harlow 1987). Importantly, moreover, these unusually small inventories are simply the phonological end point of a millennia-long reduction in the number of consonants as languages spread further and further into the Pacific: Austronesian 23 consonants, Proto-Malayo-Polynesian 20, Proto-Oceanic 23, Proto-Central Pacific 21, Proto-Polynesian 13, Nuclear Polynesian 11, Central Eastern Polynesian 10, Hawai'ian 8. The reduction was effected as follows.

Proto-Austronesian (see Tryon 1993), which was spoken, as we have seen, around 4000 BC, had a phoneme inventory including 23 consonants (Table 2).

Ross (1992) posits a reduction to 20 consonants for Proto-Malayo-Polynesian.

Proto-Oceanic (Lynch et al. 2002: 63) had a slightly larger system, although a number of the additions had low functional load (Table 3).

Proto-Central Pacific is believed to have had about 21 consonants (Andy Pawley, personal communication; Geraghty 1986 argues for a higher figure).

Proto-Polynesian, whose separate identity has to postdate the settlement of Tonga and Samoa, demonstrates a rather dramatic loss of consonants, with a system considerably reduced as compared to that of Proto-Austronesian and Proto-Oceanic. It had (see Clark 1976, Krupa 1982) a consonant inventory of only 13 consonants (Table 4).

This was somewhat reduced in Nuclear Polynesian (the ancestor of all modern Polynesian language groups except Tongic) by the loss of /h/, and the merger of /t/ with /l/, giving a system of eleven consonants. In Central Eastern Polynesian, which postdates the eastward expansion of the Polynesian peoples into the more remote areas of the Pacific, this was further reduced to ten consonants as a result of the loss of /ʔ/. This is already a very minimal consonant system, especially bearing in mind that there were only five vowels. Then, however, and remarkably, Hawai'ian, whose separation from the other eastern Polynesian languages obviously postdates the settlement of Hawai'i from the Marquesas (see Sutton (ed.) 1994), reduced the consonant system even further to eight, as mentioned above, by merging /f/ and /s/ as /h/, and merging /ŋ/ with /n/. In addition to this, /k/ became /ʔ/ and /t/ changed to /k/, see Table 5 (left side).

The Rurutu language of the Austral Islands (Tubuai), situated on the extreme southern fringes of French Polynesia, also developed an extremely attenuated consonant system in which the glottal stop had three different historical sources (Table 5, right side).

And, while New Zealand North Island Maori has ten consonants, the same number as in Central Eastern Polynesian, in South Island Maori (Harlow 1987), now extinct, this was also reduced, as we mentioned above, to nine, as /ŋ/

Table 2. *Proto-Austronesian consonant phonemes (Tryon 1993)*

<i>m</i>	<i>n</i>			ŋ		
<i>p</i>	<i>t</i>			<i>k</i>	<i>q</i>	ʔ
<i>b</i>	<i>d</i>	ɖ	ʃ	<i>g</i>		
	<i>ts</i>					
	<i>dʒ</i>					
	<i>s</i>		ʃ			<i>h</i>
	<i>z</i>					
	<i>l</i>	ɭ				
	<i>r</i>				ʁ	

Table 3. *Proto-Oceanic consonant phonemes (Lynch et al. 2002)*

<i>mw</i>	<i>m</i>	<i>n</i>	ɲ	ŋ	
<i>pw</i>	<i>p</i>	<i>t</i>	<i>c</i>	<i>k</i>	<i>q</i>
<i>bw</i>	<i>b</i>	<i>d</i>	ʃ	<i>g</i>	
		<i>s</i>			
		<i>r</i>			<i>R</i>
		<i>dr</i>			
		<i>l</i>			
<i>w</i>			<i>j</i>		

Table 4. *Proto-Polynesian consonant phonemes (Clark 1976, Krupa 1982)*

<i>m</i>	<i>n</i>	ŋ	
<i>p</i>	<i>t</i>	<i>k</i>	ʔ
<i>f</i>	<i>s</i>		<i>h</i>
<i>v</i>			
	<i>l</i>		
	<i>r</i>		

Table 5. *Consonant phonemes in two Polynesian languages, Hawai'ian and Rurutu*

Hawai'ian				Rurutu		
<i>m</i>	<i>n</i>			<i>m</i>	<i>n</i>	
<i>p</i>		<i>k</i>	ʔ	<i>p</i>	<i>t</i>	ʔ
			<i>h</i>	<i>f</i>		
<i>v</i>				<i>v</i>		
	<i>l</i>				<i>r</i>	

Table 6. *Consonant phonemes in New Zealand North Island Maori (left) and South Island Maori (right)*

<i>m</i>	<i>n</i>	<i>ŋ</i>		<i>m</i>	<i>n</i>	
<i>p</i>	<i>t</i>	<i>k</i>		<i>p</i>	<i>t</i>	<i>k</i>
<i>wh</i>			<i>h</i>	<i>f</i>		<i>h</i>
<i>w</i>	<i>r</i>			<i>w</i>	<i>l</i>	

merged with /k/ (/r/ also became /l/); see Table 6. There is also a claim (Biggs 1978) that /f/ merged with /h/, giving an eight-consonant system: /m n p t k w l h/.

Similar reductions of consonant inventories can be found in some of the Polynesian Outlier languages spoken on small isolated atolls on the fringes of Micronesia and Melanesia: Kapingamarangi, for example, has nine consonants. Nine-consonant systems are also found in Mangarevan, spoken in the Gambier Islands about 1,600 kms southeast of Tahiti, and some forms of Marquesan (Clark 1976: 20, Krupa 1982: 26).

#### 4. Possible sociolinguistic explanations

The question I now want to pose is this: is there any connection between this geographical penetration deeper and deeper into the formerly uninhabited Pacific and the loss of consonants? Andy Pawley (personal communication) paints a very nice picture of Polynesians setting off in their canoes, throwing consonants overboard as they go. But is there anything which linguists can actually say about this? Is it just a coincidence that the gradual centuries-long but dramatic and pioneering dispersal of the ancestors of the modern Polynesians from mainland Asia into more and more remote areas of the hitherto uninhabited Pacific Ocean was accompanied by an equally gradual but no less dramatic reduction in the size of the phonological inventories of the languages spoken by these people? In particular, is an explanation available that might link this aspect of linguistic structure to aspects of societal structure? Recall that the possible explanatory societal factors introduced at the beginning of this paper concerned language contact versus isolation, and community size and network structure. One encouraging pointer that such an explanation for the loss of consonants in Austronesian might be at hand is that we can suppose that the two factors of isolation and small community size would have become increasingly relevant as Polynesian groups advanced further and further into the remote Pacific.

##### 4.1. *Contact and isolation*

The issue of the relationship between societal type and size of phonological inventories (see Trudgill 1998) has been addressed by Haudricourt (1961), who



cites the Caucasian language Ubykh, which had a very large phoneme inventory including 78 consonants. He points out that this large-inventory (and now dead) language was spoken by a smaller population in a smaller geographical area than other, related Caucasian languages which had smaller inventories. He also points to North America, where East Coast languages have fewer than 20 consonants (e.g., Oneida with ten), while the further west one goes, the bigger the inventories get AND the more languages there are, or at least were, per square mile. Is this, he asks, just a coincidence? Given that languages generally both lose and develop new phonemes, how do we explain this relationship between geographical language density and phoneme inventories?

Nichols has one answer. She writes (Nichols 1992: 193): "It can be concluded that contact among languages fosters complexity, or, put differently, diversity among neighbouring languages fosters complexity in each of the languages." (I would point out that such contact will of course have to be of a very particular type, namely long-term contact situations involving childhood – and therefore proficient – bilingualism. As indicated above, adult language contact tends to lead, on the contrary, to simplification.) Large inventories will be favoured by stable contact situations because the long-term presence of many neighbouring languages means that segments can readily be borrowed from one language to another, thus leading to increased inventories, such as the well-known borrowing of velaric-ingressive stops by some Bantu languages from the Khoisan languages of southern Africa. This process can also be seen in the case of certain Polynesian languages. Some Polynesian Outlier languages which came into contact with phonologically more complex languages, as in parts of Melanesia, have added consonants. For example, Rennellese has 13 consonants, Emae 15, Mele-Fila 16, and West Uvean 26 (Clark 1994).

Large phonological inventories, then, may be the result of borrowing. But what of small inventories? Why do some varieties have very small numbers of vowels and consonants?

An obvious suggestion is that this may result from relatively mechanical factors associated with language contact of the type involving adult language contact and acquisition. The point is that simplification, both in language contact and in dialect contact situations (see Chambers 1995: 160), is brought about by the imperfect language learning of adults and post-adolescents (McWhorter 1997). Simplification may very well lead to loss of phonological contrasts: the smaller the inventory, the easier it is to learn, which is why the most extreme products of pidginisation, pidgin languages themselves, tend to have small phoneme inventories. (This is not necessarily true of pidgin-like languages which have grown up in situations involving contact between closely related languages – some variants of Fanakalo preserve clicks, for instance.) Labov has also maintained (1994) that in dialect contact situations, mergers tend to spread at the expense of contrasts. Isolated dialects are thus those which

are likely to resist mergers most strongly (and thus have larger phonological inventories). Many examples of this could be given. For example, Middle English *o:* and *ou* which have for centuries been merged in most varieties of English, including Received Pronunciation and the central dialects of England, remain distinct, as in *moan* and *mown*, in peripheral areas of Great Britain including East Anglia and South Wales. Obviously, other things being equal, mergers also lead to smaller inventories.

So, long-term contact involving child bilingualism may produce large inventories through borrowing, and adult language contact may produce smaller inventories through imperfect learning, pidginisation, and simplification. Unfortunately, however, this latter cannot be the only explanation for small inventories. The Polynesian languages which we have been focussing on in particular, namely those which have only eight consonants, can be supposed to have been relatively isolated, LOW-contact languages.

Haudricourt (1961) attempts an explanation for this. Small inventories, he says, are the result of the impoverishment which occurs in situations characterised by monolingualism and isolation (the opposite of the situation obtaining in the Caucasus) – and/or by non-egalitarian bilingualism. Haudricourt suggests that in certain situations the superiority of a dominant group in a diglossic bilingual environment may be

so obvious they no longer have any need to articulate well to be understood – they may confuse two different phonemes or no longer pronounce one – no one will dare to mock them. This is why we find fewer consonants in the language of the Iroquois who terrorised their neighbours, or in the languages of the people of Tahiti and Hawai'i who combine island isolation with significant demographic development as compared to other less favoured archipelagos. [my translation; PT]

#### 4.2. *Community size and network structure*

Haudricourt's is not an especially happy thesis, but it does perhaps contain the germs of an explanation. Maddieson (1984) argues that there is no actual evidence that languages such as Hawai'ian show signs that they suffer from problems due to lack of contrastive possibilities. Let us suppose, however, that a small number of available syllables, and therefore a relatively small amount of redundancy may, other things being equal, lead to greater communicative difficulty. If this is so, then we should probably turn away in this case from high contact versus low contact as an explanatory factor. We should turn instead to the other major factor mentioned above, community size and network structure, as being the most important. My argument here is as follows: initial small community size (the number of people who could arrive on a relatively small number of relatively small boats) would have led to tight social networks, which would have implied large amounts of shared background information –

a situation in which communication with a relatively low level of phonological redundancy would have been relatively tolerable.

There is, however, a problem here. One group of languages which, like the Caucasian languages, is known to have very large phoneme inventories are the San languages of southern Africa: on one analysis, !Xū has 95 consonants (Maddieson 1984), for example. It seems unlikely that we can explain this, in Nichols' terms, as being due to a high degree of contact, especially since very many of the consonants are clicks which, although they have been borrowed from San into a few southern African Bantu languages, are otherwise unknown in the area outside the Khoisan grouping. Yele, an isolated Papuan language spoken on Rossel Island, also has a very complex phoneme inventory, with 38 vowels and 56 consonants.

Why then do we find this extreme contrast between two different sets of languages traditionally spoken in small, isolated communities, the San and the Yele, with very large inventories, on the one hand, and (some of the) the Polynesians, with very small inventories, on the other?

#### 4.3. *Contact again*

The answer to this conundrum of very large versus very small inventories appears to lie in psycholinguistic issues to do with memory. One of the biggest difficulties encountered by adult non-native learners in contact situations has to do with learning and remembering, which is one reason why irregularities tend to disappear in such situations, most dramatically in the case of pidgin languages. At the phonological level, learning and remembering are probably less likely to cause particular difficulties than at other linguistic levels, but they will nevertheless be relevant in one of three ways.

4.3.1. First, Khoisan-type consonant systems are unlikely to be present in high-contact language situations because of the inherent difficulty for adults and adolescents of remembering and mastering such a rich series of articulations.

4.3.2. On the other hand, languages such as South Island Maori and Hawai'ian will cause problems of memory load of a very different type. According to Maddieson (1984), Hawai'ian has only 162 possible syllables. This inevitably leads to a situation no doubt even more extreme than that in North Island Maori, of which Harlow (2001) says a very high proportion of all possible words consisting of two morae actually occur. This is of course what leads to the lack of contrastive possibilities mentioned above. I suggest that while this lack of contrastive possibilities is entirely unproblematical for native speakers, it does cause problems for non-natives. The problem lies in the relative lack

of distinctiveness between one vocabulary item and another, due to the necessarily high level of usage of all possible syllables. The problem is one of CONFUSABILITY OF CONSTITUENTS. Consider the case of Maori, which also has, as we have seen, a small phoneme inventory. I submit that learning and remembering, as an adult, which of the following 23 words is which, is no easy task: *pakake* 'minke whale'; *paakaakaa* 'brown'; *pakeke* 'hard'; *paakeeke* 'grate'; *pakikau* 'wing'; *paakiki* 'inquisitive'; *pakoke* 'random'; *pakoki* 'distort'; *pakoko* 'dried up'; *paakoko* 'childless'; *paakoukou* 'shoulder blade'; *piikaokao* 'rooster'; *pookaakaa* 'stormy'; *pookeka* 'chant'; *pookeeke* 'gloomy'; *poukoki* 'stilts'; *puukaakaa* 'burning fiercely'; *puukaki* 'source of river'; *puukeeke* 'armpit'; *puukeko* 'swamp hen'; *puukiki* 'stunted'; *pukoko* 'lichen'; *puukukai* 'greedy'.

It is no easy task because, according to Lively et al. (1994: 274), there are NEIGHBOURHOOD EFFECTS (also called LEXICAL SIMILARITY EFFECTS) having to do with what other words a given word has to be differentiated from (Luce & Pisoni 1998). Neighbours are taken to be words that differ from a given target word by only one phoneme (see also Luce 1986, Goldinger et al. 1989). Lively et al. (1994: 275) show that words of equal frequency are identified less accurately if they come from dense neighbourhoods than if they are from sparse neighbourhoods. A word that has neighbours higher in frequency tends to have its recognition speed and accuracy depressed. Therefore, if we can generalise from recognition to memory, then my point is made – and it seems clear that we can generalise in this way, since recognition obviously depends on memory.

4.3.3. There is also the issue of word length. Nettle (1995, 1999) has shown that there is a statistical connection between inventory size and word length. It is rather obvious that this should be so on purely mathematical grounds (although it has been disputed by Maddieson 1984): the smaller the phoneme inventory, the longer words, on average, have to be, although Nettle's work shows that this is not guaranteed in any given case. The length of lexical items, moreover, in terms of syllables and segments, will have an effect on memory load. The less there is to remember, the easier language acquisition is (on the LENGTH EFFECT, see Garman 1990: 251–253). The longer a word is, in terms of syllables and/or segments, the more difficult it will be, other things being equal, to remember.

## 5. Conclusion

In the absence of a large-scale database of evidence on this topic, taken from different language families in different parts of the world, any conclusions to be produced here can be only suggestive and tentative. However, it seems that

we have reached the following preliminary conclusions concerning contact, isolation, community size, and tendencies involving phoneme inventories:

- (i) In cases where there is long-term language contact involving child-language acquisition, high degrees of language contact may lead to larger phoneme inventories, as a result of borrowing, as suggested by Nichols.
- (ii) Situations involving adult language contact, on the other hand, are likely to favour medium-sized phoneme inventories, i.e., inventories which are not so large as to be difficult for adolescents and adults to remember and acquire, but not so small as to cause confusability of constituents and high word length.
- (iii) Low degrees of language contact may lead to languages with small inventories, because the memory load difficulties caused by confusability and word length will not be relevant, since post-critical threshold learning is not involved. They may also just as well lead, however, to large inventories, because, equally, the memory load difficulties caused by the acquisition of large numbers of phonemes will not be relevant either. We have as yet no explanation for why one of these factors may come to dominate the other.
- (iv) Large community size will favour medium-sized phoneme inventories, i.e., inventories which are not so small as to cause communicative difficulties as a result of a low degree of redundancy.
- (v) Languages spoken in small communities can develop very small inventories since lower degrees of redundancy can be tolerated because of the large amounts of shared information present. Small communities can also, equally, develop very large inventories because of the ability of such communities to encourage continued adherence to norms from one generation to another, however complex they may be. We have as yet no explanation for why one of these factors may come to dominate the other. One possibility is that we should distinguish between isolated communities which do, however, have neighbours, such as the San and the Yele, and isolated communities which do not, such as the Hawai'ians. As has been suggested by Thurston (1994), small communities which have neighbours, but which wish to remain as isolated as possible, may elaborate systems in order to make them more opaque to these neighbours. However, it is not at all clear that this can occur in the case of phoneme inventories – can speakers deliberately expand phoneme inventories at will?

In the case of the Polynesian languages with very small inventories, we can point to a process whereby increasing isolation and diminution in community size, as migrations carried settlers to ever more remote parts of the Pacific, was accompanied by ever smaller phoneme inventories, as per points (iii) and (v). As points (iii) and (v) show, however, the correct generalisation is not that such

languages will necessarily have very small inventories but that they will be more likely to have EITHER very small inventories OR very large ones. (Non-isolated languages spoken in larger communities will, on the other hand, tend to have medium-sized inventories.) The factors of isolation and small community size can quite simply lead to the development of UNUSUAL phonological systems, as has also been suggested by Nettle (1999: 147): these systems may be either unusually small, as in the case of South Island Maori and Hawai'ian, or unusually large, as in the case of !Xū and Yele.

Received: 5 November 2002

Université de Fribourg

Revised: 4 August 2003

Correspondence address: Département d'anglais et slavistique, Université de Fribourg, Miséricorde, 1700 Fribourg, Switzerland; e-mail: peter.trudgill@unifr.ch

Acknowledgements: I am very grateful to the anonymous *LT* reviewers and to the following people who have provided very helpful information and/or commented on earlier drafts of this paper: Kurt Braunmüller, Michael Garman, Patrick Griffiths, Jean Hannah, Ray Harlow, John McWhorter, Daniel Nettle, Frans Plank, Malcolm Ross, and, especially, Andy Pawley.

## References

- Biggs, Bruce (1978). The history of Polynesian phonology. In Stephen A. Wurm & Lois Carrington (eds.), *Second International Conference on Austronesian Linguistics: Proceedings. Fasc. 2: Eastern Austronesian* (Pacific Linguistics, C-61), 691–716. Canberra: Australian National University.
- Blust, Robert (1999). Subgrouping, circularity and extinction: Some issues in Austronesian comparative Linguistics. In Zeitoun & Li (eds.) 1999, 31–94.
- Chambers, Jack K. (1995). *Sociolinguistic Theory: Linguistic Variation and its Social Significance*. Oxford: Blackwell.
- Clark, Ross (1976). *Aspects of Proto-Polynesian Syntax*. Auckland: Linguistic Society of New Zealand.
- (1994). The Polynesian Outliers as a locus of language contact. In Dutton & Tryon (eds.) 1994, 109–139.
- Dutton, Tom & Darrell Tryon (eds.) (1994). *Language Contact and Change in the Austronesian World*. Berlin: Mouton de Gruyter.
- Fischer, Steven Roger (2001). Mangarevan doublets: Preliminary evidence for Proto-Southeastern Polynesian. *Oceanic Linguistics* 40: 112–124.
- Garman, Michael (1990). *Psycholinguistics*. Cambridge: Cambridge University Press.
- Geraghty, Paul (1983). *The History of the Fijian Languages*. Honolulu: University of Hawaii Press.
- (1986). The sound system of Proto-Central-Pacific. In Paul Geraghty, Lois Carrington, & Stephen A. Wurm (eds.), *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics* (Pacific Linguistics, C-94), 289–312. Canberra: Australian National University.
- Goldinger, Stephen D., Paul A. Luce, & David B. Pisoni (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28: 501–518.

- Hancock, Ian (1988). Componentiality and the origins of Gullah. In James L. Peacock & James C. Sabella (eds.), *Sea and Land: Cultural and Biological Adaptation in the Southern Coastal Plain*, 13–24. Athens, Ga.: University of Georgia Press.
- Harlow, Ray (1987). *A Word-list of South Island Maori*. 2nd edition. Auckland: Te Reo.
- (2001). *A Maori Reference Grammar*. Auckland: Longman.
- Haudricourt, André (1961). Richesse en phonèmes et richesse en locuteurs. *L'Homme* 1: 5–10.
- Kirch, Patrick Vinton & Roger C. Green (2001). *Hawaiki, Ancestral Polynesia: An Essay in Historical Anthropology*. Cambridge: Cambridge University Press.
- Krupa, Victor (1982). *The Polynesian Languages*. London: Routledge.
- Labov, William (1994). *Principles of Linguistic Change*. Oxford: Blackwell.
- Le Page, Robert & Tabouret-Keller, Andrée (1985). *Acts of Identity: Creole-based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.
- Lively, Scott E., David B. Pisoni, & Stephen D. Goldinger (1994). Spoken word recognition: Research and theory. In Morton A. Gernsbacher (ed.), *Handbook of Psycholinguistics*, 265–301. San Diego: Academic Press.
- Luce, Paul A. (1986). Neighbourhoods in the mental lexicon. (Research on speech perception technical report, 6.) Bloomington: Dept. of Psychology, Speech Research Laboratory, Indiana University.
- Luce, Paul A. & David B. Pisoni (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19: 1–36.
- Lynch, John, Malcolm Ross, & Terry Crowley (2002). *The Oceanic Languages*. Richmond: Curzon.
- Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.
- Marck, Jeff (1999). Revising Polynesian linguistic subgrouping and its culture history implications. In Roger Blench & Matthew Spriggs (eds.) *Archaeology and Language IV: Language Change and Cultural Transformation*, 95–122. London: Routledge.
- (2000). *Topics in Polynesian Language and Culture History*. (Pacific Linguistics, 504.) Canberra: Australian National University.
- McWhorter, John (1997). *Towards a New Model of Creole Genesis*. Frankfurt: Lang.
- (2000). *The Missing Spanish Creoles*. Berkeley: University of California Press.
- (2001). The world's simplest grammars are creole grammars. *Linguistic Typology* 5: 125–166.
- Mufwene, Salikoko (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- Mühlhäusler, Peter (1977). *Pidginization and Simplification of Language*. (Pacific Linguistics, B-26.) Canberra: Australian National University.
- Nettle, Daniel (1995). Segmental inventory size, word length, and communicative efficiency. *Linguistics* 33: 359–367.
- (1999). *Linguistic Diversity*. Oxford: Oxford University Press.
- Nichols, Johanna (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Pawley, Andrew (1996a). On the position of Rotuman. In Bernd Nothofer (ed.), *Reconstruction, Classification, Description: Festschrift in Honour of Isidore Dyen*, 387–410. Hamburg: Abera.
- (1996b). On the Polynesian subgroup as a problem for Irwin's continuous settlement hypothesis. In Janet Davidson, Geoffrey Irwin, Foss Leach, Andrew Pawley, & Dorothy Brown (eds.), *Oceanic Culture History: Essays in Honour of Roger Green*, 387–410. Dunedin: New Zealand Journal of Archaeology.
- (1999). Chasing rainbows: Implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In Zeitoun & Li (eds.) 1999, 95–138.
- Pawley, Andrew & Malcolm Ross (1993). Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22: 425–459.



- Ross, Malcolm (1992). The sound of Proto-Austronesian: An outsider's view of the Formosan evidence. *Oceanic Linguistics* 31: 23–64.
- Sutton, Douglas G. (ed.) (1994). *The Origins of the First New Zealanders*. Auckland: Auckland University Press.
- Thurston, William (1994). Renovation and innovation in the languages of north-western New Britain. In Dutton & Tryon (eds.) 1994, 573–609.
- Trudgill, Peter (1989). Contact and isolation in linguistic change. In Leiv Egil Breivik & Ernst Håkon Jahr (eds.), *Language Change: Contributions to the Study of its Causes*, 227–237. Berlin: Mouton de Gruyter.
- (1996). Dialect typology: Isolation, social network and phonological structure. In Gregory R. Guy, Crawford Feagin, Deborah Schiffrin, & John Baugh (eds.), *Towards a Social Science of Language: Papers in Honour of William Labov*, Volume 1: *Variation and Change in Language and Society*, 3–21. Amsterdam: Benjamins.
- (1998). Typology and sociolinguistics: Linguistic structure, social structure and explanatory comparative dialectology. *Folia Linguistica* 31: 349–360.
- Tryon, Darrell (1993). *Comparative Austronesian Dictionary*. Berlin: Mouton de Gruyter.
- Whinnom, Keith (1971). Linguistic hybridization and the 'special case' of pidgins and creoles. In Dell Hymes (ed.), *Pidginization and Creolization of Languages*, 91–115. Cambridge: Cambridge University Press.
- Zeitoun, Elizabeth & Paul Li (eds.) (1999). *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Academia Sinica.