

CONCORDANCE STUDIES OF THE LANGUAGE OF ADOLESCENT FIRST LANGUAGE TOK PISIN SPEAKERS

GEOFF P. SMITH

1. INTRODUCTION

Tok Pisin is one of the most extensively studied of all pidgin languages, and numerous detailed descriptions of its lexicon, grammar, historical development and sociolinguistic status are available. The raw material on which these analyses are based include written records of speech samples dating back to the mid-nineteenth century (Keesing 1988; Mühlhäusler 1978); corpora of recorded material based on research in a variety of regions (Romaine 1992; Mühlhäusler 1979; Sankoff 1980; Goulden 1990; Dutton 1973; Laycock 1970; Wurm 1971); written standards by expatriates based on close consultation with native speakers (Bible Society 1966; Bergmann 1982; Mihalic 1971, 1982); and native speaker insights (Thomas 1990; Nekitel 1994). Some of the data collected for specific studies have been subjected to further analysis by other writers (Smeall 1975; Woolford 1979).

While a wealth of data is available, one area which has not been adequately addressed is the degree of variation which currently exists and the extent to which such variation is 'acceptable' to the members of the language community. Some attempts have been made to establish norms of acceptability, for example the written standards referred to above, but detailed research into the parameters of acceptability is yet to be carried out. Similarly, in academic studies describing Tok Pisin grammatical structures, the criteria by which the correct is distinguished from the incorrect, or the acceptable from the unacceptable are not always clearly spelled out. A number of dimensions of variation can be identified: regional varieties (Laycock 1970; Wurm 1971; Mühlhäusler 1977, 1985), sociolects (Mühlhäusler 1975), distinctions based on age (Lynch 1979; Romaine 1984, 1990a; Sankoff & Laberge 1973), degrees of formality (Brash 1971; Mühlhäusler 1991; Smith 1990b), medium (Siegel 1983; Romaine 1990b) and relationships with other languages (Mosel 1980; Smith 1994).

This paper looks at evidence for regional variation between two roughly comparable groups of adolescent first-language Tok Pisin speakers from widely separated geographical areas. The evidence is based on an analysis of a small corpus of recorded speech, using the concordance programs Oxford *Microconcord* (Scott & Johns 1993) and Longman *Mini-concordancer* (Chandler 1989). Some comments on the usefulness of these concordance programs for this type of analysis will also be offered.

Jan Tent and France Mugler, eds *SICOL: proceedings of the Second International Conference on Oceanic Linguistics: vol.1, Language Contact*, 135–146.

Pacific Linguistics, C-141, 1998.

© Geoff P. Smith

2. THE STUDY SAMPLE

The texts used for this study were transcripts of recorded speech of adolescents in the Western Highlands (WH) and in New Ireland (NI) Provinces who spoke Tok Pisin as their first language. All the respondents were in secondary or upper primary education. The WH samples were recorded at Mt Hagen High School and Hagen Park High School in December 1990, while the NI samples were recorded at Utu Provincial High School and Carteret Community School, Kavieng, in August 1990. The respondents were asked to talk about personal experiences or tell a story based on either a modern or a traditional theme. The WH informants consisted of 12 males and 10 females ranging in age from 14 to 19 and the sample consists of just over 11,000 words. The NI sample is slightly larger, at almost 14,000 words, and was collected from 16 males and 8 females ranging in age from 12 to 16. These are very small samples by the standards of present-day corpora, but large enough to allow us to gain some interesting insights into some differences between the two.

3. CONCORDANCE PROCEDURE

Tape-recorded speech was transcribed using a broad phonetic transcription and a DOS text file was prepared for each of the WH and NI samples. These files were examined using the Oxford *Microconcord* and Longman *Mini-concordancer* (version 1.01) programs. Both these programs basically produce concordances of specified 'search words', that is, they identify all occurrences of the search word and display the line of text in which these words occur to show the context. Each line can be expanded to a larger context sample if desired. The Oxford program is useful for analysing extremely large amounts of text, and can simultaneously search through up to 500 different files from over 900 directories. However, it is limited to displaying a maximum of about 1,600 concordance lines, the exact number depending on other computer software used.

The concordance lines can be arranged according to alphabetical listing of the word immediately to the right or left of the search word, or two or three places from it. Searches can be conducted for groups of alternate words, which is useful when dealing with transcripts which contain various renderings of the same word, such as *dispela* or *disla*. Searches can also be limited to collocations with other context words at a specified distance or 'horizon' from the search word. Wild card characters can also be specified, for example *sa** would include *sa*, *save* and any other word beginning with 'sa'. Appropriate use of wild cards can make data analysis somewhat less arduous.

The Longman *Mini-concordancer* can in addition produce word frequency lists, which can not be produced by *Microconcord*. It can also produce larger numbers of samples than the 1,600 limit of *Microconcord*, which is occasionally useful in examining the contexts of frequently occurring items (e.g. *i*). However, the size of the files that the Longman *Mini-concordancer* can handle is limited to about 50,000 words. The following word frequency lists were produced using the Longman *Mini-concordancer*.

4. ANALYSIS

4.1 WORD FREQUENCY

In the sample from WH, 879 different words occurred, whereas the NI sample produced 918 different words. The 25 most frequently occurring words, including variations such as

phonological reductions, are shown in Table 1. The words are arranged in decreasing order of frequency, and the number of occurrences is recorded.

There were altogether 6,466 occurrences of the first 25 words in the WH sample, constituting 57.3% of the total. In the NI sample, the figures were 8358 and 60.2% respectively. Of the commonest words, function words such as *i*, *long*, *na*, *ia*, *olsem*, *bin*, and *sa* are well represented, as are the commoner pronouns (*mi*, *mipela*, *yu*, *em*, *ol* and *tupela*). The commonest verbs are *go*, *kam*, *tok*, *kisim* and *stap*. The only word used exclusively as a noun is *man*, and there are two adjectives, *dispela* and *wanpela*, although these function primarily as demonstrative pronoun and article respectively.

TABLE 1: THE 25 MOST FREQUENTLY OCCURRING WORDS IN WH AND NI

WH (11,292 words)					NI (13,888 words)				
Word	No.	Freq. (1/n)	Rank	NI Rank	Word	No.	Freq. (1/n)	Rank	WH Rank
<i>na*</i>	747	15	1	3	<i>i</i>	1157	12	1	6
<i>em</i>	652	17	2	4	<i>long*</i>	715	19	2	3
<i>long*</i>	617	18	3	2	<i>na*</i>	645	22	3	1
<i>go</i>	466	24	4	5	<i>em</i>	635	22	4	2
<i>ol</i>	463	24	5	7	<i>go</i>	623	22	5	4
<i>i</i>	311	36	6	1	<i>nau</i>	596	23	6	9
<i>mi</i>	311	36	7	8	<i>ol</i>	541	26	7	5
<i>bilong*</i>	310	36	8	9	<i>mi</i>	406	34	8	7
<i>nau</i>	245	46	9	6	<i>bilong*</i>	348	40	9	8
<i>tupela*</i>	230	49	10	19	<i>ia</i>	312	45	10	12
<i>stap*</i>	228	50	11	26	<i>bin</i>	229	61	11	19
<i>ia</i>	220	51	12	10	<i>kam</i>	225	62	12	13
<i>kam</i>	215	53	13	12	<i>wanpela*</i>	200	69	13	22
<i>mipela*</i>	199	57	14	16	<i>kirap*</i>	195	71	14	31
<i>dispela*</i>	189	60	15	18	<i>tok</i>	188	74	15	17
<i>olsem*</i>	168	67	16	24	<i>mipela*</i>	182	76	16	14
<i>tok*</i>	151	75	17	15	<i>taim</i>	163	85	17	20
<i>man</i>	150	75	18	21	<i>dispela*</i>	156	89	18	15
<i>bin</i>	124	91	19	11	<i>tupela*</i>	149	93	19	10
<i>taim</i>	97	116	20	17	<i>save/sa</i>	146	95	20	21
<i>save/sa</i>	92	123	21	20	<i>man</i>	132	105	21	18
<i>wanpela*</i>	90	125	22	13	<i>kisim</i>	121	115	22	24
<i>tasol*</i>	85	133	23	24	<i>yu</i>	116	120	23	26
<i>nogat*</i>	80	141	24	65	<i>olsem*</i>	113	123	24	16
<i>kisim*</i>	74	153	25	22	<i>tasol*</i>	113	123	24	23

* denotes the inclusion of phonological variants or reduced forms (e.g. *bilong* includes *bilong*, *blong*, *blo*, and *bl'*).

The first observation to be made from this list is an apparently remarkable uniformity between the two samples. The nine commonest words are identical in both samples, although the order of frequency is slightly different. The 25 commonest words in WH contain only two which are outside the commonest 25 in NI and vice versa.

In an earlier study (Smith 1994), the extensive borrowing or calquing of English expressions by young first language speakers of Tok Pisin in Manus was noted. Typically, English semantics were kept more or less intact, while the borrowings were fully integrated syntactically into Tok Pisin. This tendency is also seen to some extent in the present study, with integrated lexical items such as *spreadaut* 'spread out' and *kolektim* 'to collect' (WH), and *inteviu* 'interview' and *ektivitis* 'activities' (NI) well represented in the data.

4.2 DIFFERENCES IN WORD FREQUENCY

In spite of the apparent uniformity of the frequency lists, one or two differences are evident. In some cases, it may be that the subject matter of the discourse in a small sample gives a misleading impression of the relative frequency of words. For example, the word *keiv* 'cave' occurs seven times in the WH sample, not at all in the NI sample. This does not mean that the word does not occur in NI, but that the topic of one of the stories from WH happened to involve caves. To get a good idea of the frequency of such words with any degree of accuracy, a very large corpus would be needed. (Even extremely large corpora, such as the Bank of English's 200 million words used in the COBUILD English dictionary can give misleading impressions of the frequency of words as used in everyday speech.) However, in other cases it does appear that words are more commonly used in one location than the other. For example, the word *gel* 'girl' or *gels* occurs 30 times in the WH sample but only once in the NI text. This is not due to the nature of the topic, as the alternative *meri* is used with approximately equal frequency in the two samples. *Nogat* 'no' has a frequency of 1/113 in WH (78 occurrences), but only 1/350 (30 occurrences) in NI. As far as less frequent items are concerned, the word *gatim* does not occur in the NI sample, although one occurrence can be found in the WH data, which does indicate that this form may coexist with *gat* in some areas. *Stap* occurs more than twice as frequently in WH than NI and *i* more than three times as frequently in NI than WH. The significance of these differences will be discussed more fully below.

In the latter cases, and other examples, it appears that the difference in frequency reflects a fundamental difference in the function of the word in question. This calls for a more detailed look at contexts and collocations as revealed by concordances, which may reveal differences in the way the words are used.

4.3 ILLUSTRATIVE EXAMPLES OF COLLOCATIONS AND WORDS IN CONTEXT

4.3.1 *nabaut*

As an example, the occurrences of the word *nabaut* (including the phonological variant *nambaut*) will be analysed. *Nabaut*, generally glossed as the adverbs 'around' or 'about', occurs 40 times in the WH sample, giving it a frequency of 1/221. In the NI sample, the frequency is less than one third of this (1/657). However, the concordances show that the difference is not merely a question of number of occurrences, but that a differentiation of function is involved. A concordance highlighting the words preceding the target word was

examined to see how often it qualified nouns or noun phrases and how often verbs or verb phrases. In both the NI and WH samples, the use of *nabaut* as an adverb following verbs occurred with roughly comparable frequency. The usage is illustrated by the example *ol i go daiv nabaut* 'they went and dived around'. However, the WH data shows that the majority of the occurrences of *nabaut* followed nouns or noun phrases (e.g. *nogat ol kostiums nabaut* 'there were no costumes and things'). The overall frequency of this distinctive use in NI was only about one sixth of its frequency in WH.

TABLE 2: USE OF *nabaut* IN NI AND WH

	WH	NI
Total	40	16
following noun phrase	28 (70%)	6 (37%)
following verb phrase	12 (30%)	10 (63%)

Thus, the use of *nabaut* following nouns in phrases like *ol sotpela man nabaut*, meaning something like 'and that sort of thing, et cetera' appears to be well established in the WH sample, but far less so in the NI sample. However, it is not always clear if the *nabaut* refers to the verb or a noun in the verb phrase, for example in *ol go painim kaikai nambaut*, the *nambaut* could qualify either the verb *painim* or the noun *kaikai*. This gives a clue as to how the usage could have arisen, from an adverbial expression qualifying verbs reinterpreted as a qualification of nouns in verb phrases and then of nouns alone.

4.3.2 *kirap*

The frequency tables also indicate that the word *kirap* and its phonologically reduced variants *kira* and *kra* occur considerably more frequently in NI (1/71) than in WH (1/213). The word *kirap* normally means 'get up' or 'start up'. Again, it can be shown that the difference in frequency reflects different usage. The use of *kirap* in context shows a very frequent collocation with the word *tok*, illustrating its use in reporting speech. Standard phrases such as *em i kirap na tok olsem*, usually considerably reduced (e.g. *em kra to se*), are widely used when quotations are introduced. A collocation of *kirap* and its phonological variants with *tok* 'say', using a (0, 3) horizon, was examined (i.e. occurrences of *tok* within three words to the right of *kirap*). This picks up phrases such as *kra tok*, *kirap na tok* and *kirap na i tok*. The concordance showed 78 occurrences of this collocation in NI (1/134), but only 28 in WH (1/316), less than half the frequency in the total speech sample. However, collocation of *kirap* and *tok* accounted for a greater percentage of the uses of *kirap* in WH than in NI. The phrase usually indicates the initiation of an opening move in a dialogue as opposed to a response to what someone else has said.

TABLE 3: USE OF *kirap* WITH *tok* IN NI AND WH

	WH	NI
total occurrences of <i>kirap</i>	53 (1/166)	195 (1/71)
collocation with <i>tok</i> (0, 3)	28 (53%)	78 (40%)

Thus, the use of *kirap* in these samples is considerably more common in NI than in WH, but of those occurrences in WH a greater proportion are used to qualify *tok* to indicate an opening move in a dialogue.

4.3.3 ADVERBIAL CLAUSES WITH *taim*

Dutton and Bourke (1990) first brought to our attention a use of *taim* 'time' or 'when' from the Southern Highlands in clause-final position when used as a conjunction in 'when'-clauses. It has since been confirmed that this pattern occurs not only in the Southern Highlands, but elsewhere in the Highlands region and it has even been reported from NI (1990:259). A concordance of the word *taim* can quickly determine how common this syntactic pattern is in the WH and NI samples. In the NI samples, the word *taim* occurred 163 times, 55 of which were as conjunctions in 'when'-clause constructions such as *taim mi kam long hia* 'when I came here'. The remaining uses of *taim* were mainly as nouns (e.g. *dispela taim*, *wanem taim?*). None of the instances of *taim* used as a conjunction was in clause-final position. In the WH sample, *taim* occurred 97 times. Of the 31 uses as a conjunction, *taim* occurred in clause-initial position in 25 or 81% of instances, but in a significant minority of cases (6 or 11%) it was in clause-final position as in *mi stap long greid nain taim*, *mi fil olsem breiv liklik* 'when I was in grade 9 I felt a bit braver'.

TABLE 4: POSITION OF *taim* IN WHEN-CLAUSES

	WH		NI	
	No.	%	No.	%
<i>taim</i> – total	97	100	163	100
<i>taim</i> – when-clauses	31	32	55	34
clause-initial	25	81	55	100
clause-final	6	11	0	0

Thus, it can be seen that *taim* is used mostly as a noun both in WH and NI, but of the 30% or so of occurrences as a conjunction, it appeared clause-finally in 11% of cases in WH, not at all in NI.

4.3.4 THE PARTICLE *i*

Debate about the exact function of the particle *i*, variously described as a predicate marker, subject referencing pronoun or resumptive pronoun (Keesing 1988) continues unabated (e.g. Franklin 1980; Keesing 1988; Mühlhäusler 1987, 1990; Romaine 1993; Verhaar 1991; Woolford 1979). What is remarkable here is the occurrence of *i* as the commonest word in NI and the sixth commonest in WH, especially considering accounts of the decline of the use of the particle (Sankoff 1980:269) and even predictions of its possible eventual extinction

(Romaine 1993:259). Although the particle *i* is in the six most frequently occurring words in both lists, there is a significant difference between the two samples. To begin with, *i* occurs over three times more frequently in the NI sample with 8.3% of the total number of words compared with 2.8% in WH. But this is not merely an overall greater frequency in all contexts, and the pattern of use needs to be examined in more detail.

Concordances of the use of *i* in context were examined, displaying both the preceding word and following words listed alphabetically. A number of distinctive patterns became obvious from a cursory examination of these displays. When concordances were arranged with the preceding word sorted alphabetically, the number of occurrences of *ia i* was conspicuous in NI, 78 in all. The comparable figure for WH was only one. Similarly, there was a huge discrepancy in the frequency of occurrences of *na i*: 114 in NI compared with nine in WH. Comparison of occurrences of *nau i* was even more dramatic, with 104 in NI compared with zero in WH. When the word following *i* was sorted alphabetically, these concordances highlighted differences in the co-occurrence of *i* with *go* and *kam*, with 30% of occurrences in NI, but 42% in WH. More significantly, the collocation *i stap* accounted for 18% of the WH occurrences, but only 3% in NI.

TABLE 5: USE OF *i* IN WH AND NI

	WH		NI	
	No.	%	No.	%
<i>i</i> total	311 (1/36)	100	1,157 (1/12)	100
<i>ia i</i>	1	0.3	78	7
<i>na i</i>	9	3	114	10
<i>nau i</i>	0	0	104	9
<i>i sa</i>	2	0.6	25	2
<i>i go/i kam</i>	131	42	346	30
<i>i stap</i>	56	18	40	3
pronoun + <i>i</i>	102	33	373	32

These figures point to a striking difference in the use of *i* in discourse in the two samples. The preference for clause-initial use after *na* and *nau* represents a discourse style very rarely found in the WH sample. The very limited use of *i* in sequence with aspect and other particles such as *ia* is also noticeable. Nearly half of the WH use of *i* is in the constructions *i go* and *i kam*, including both use as directionals and also to indicate continuing action. In fact 60% of all occurrences of *i* in WH preceded only three words—*go*, *kam* and *stap*. A significant proportion of approximately one third of both samples consisted of *i* in its well-known role as marking predicates following subject pronouns *em*, *mipela*, *ol* and *tupela*.

4.3.5 RELATIVE CLAUSES WITH *longen*

In transcribing the texts, *longen* and its variants (*logen*, *loen*, *len*) were written as a single word, as *-en* does not have an independent existence. The usual meaning of *longen* is 'of it' or 'its' or 'to it' or 'to him' etc., as for example in the extract from NI: *mi kirap, mi lap longen nau...* 'I got up and laughed at him now...'. This is quite similar to the meaning of *long em*. Dutton (1973:39) draws the distinction between *em* and *-en* in terms of emphasis or focus. However, in the present data, *longen* appears to have an additional role in delimiting relative clauses in a significant number of cases. For example, in *wanpela ples i nogat man silip longen* 'a place where there was nobody sleeping', *longen* has anaphoric reference to *peles*, meaning 'there' or 'in it', but also has a role in delimiting the relative clause. This tendency is carried further in some instances, where the relative clause delimiting function appears primary, and *longen* does not appear to have any anaphoric reference:

- (1) *I man Wabag ia em poisinim em longen ia wokobaut kam i go.*
It was the Wabag man who poisoned him approaching.

In the above example, *ia* bracketing as described by Sankoff and Brown (1976) also contributes to the definition of the relative clause, although this is not always the case:

- (2) *...bihainim dispela ol liklik ston tupela bin tromei long rot longen.*
...followed these little stones which they had thrown on the road.

In this example, again there appears to be no anaphoric referent for *longen*. This clause delimiting function is rarely seen with *long em*, and it is much more common in WH than NI, as shown by the figures in Table 6.

TABLE 6: *longen* AND *long em* AS RELATIVE CLAUSE MARKERS IN NI AND WH

	WH		NI	
	No.	%	No.	%
total occurrences – <i>longen</i>	34 (1/332)	100	20 (1/694)	100
<i>longen</i> as relative clause boundary	23	68	5	25
total occurrences – <i>long em</i>	14 (1/806)	100	13 (1/1,068)	100
<i>long em</i> as relative clause boundary	0	0	2	14

4.3.6 *stap* AND OTHER FEATURES

Stap is used to mean 'is situated', as in *diwai stap hapsaid long rot* 'the tree was on the other side of the road', and more generally 'to live' or 'to be'. It also follows verbs to indicate continuing action as in *tupela wok long kaikai i stap* 'they two were eating'. *Stap* occurs with a frequency about three times greater in the WH sample than in NI. Again a number of collocations show a difference in co-occurrence of *stap* with other words.

Table 7 shows that roughly equal percentages of occurrences of *stap* follow *i*. There is a much greater tendency for *stap* to follow pronouns directly in the WH sample, approximately three times the percentage of occurrences in NI. The collocation *stap + go* or *i go* is again much commoner in the WH sample. *Stap* preceding *na* or *nau* at boundaries in discourse

occurs more frequently in WH, whereas the collocation *sa stap* or *save stap* is much more frequent in the NI sample.

TABLE 7: *stap* IN NI AND WH

	WH		NI	
	No.	%	No.	%
<i>stap</i> – all occurrences	209 (1/54)	100	98 (1/142)	100
<i>i stap</i>	47	22	20	20
pronoun + <i>stap</i>	59	28	9	9
<i>stap (i) go</i>	24	11	4	4
<i>stap na/nau</i>	52	25	17	17
<i>sa/save stap</i>	11	5	16	16

As pointed out by Lynch (pers. comm.), there may be substrate motivation for the greater frequency of *stap* in WH, as a variety of existential verbs have been described in languages of the Highlands region (Piau 1981).

Similar analyses can be carried out with other grammatical features such as the marker *bai* and other auxiliaries. *Bai* and its phonological reductions *ba* and *b'* (the form *baimbai* does not occur in either sample) occur with approximately equal frequency in both samples. There has been considerable discussion of the tendency for *bai* to shift to preverbal position in creolised forms (Romaine 1990a). In the present samples the particle occurs pre-verbally in WH in over half of the occurrences (52%), while preverbal use accounts for only 21% of occurrences in the NI sample. Combinations of *bai* with other modal, aspectual or negative particles such as *bai mas*, *bai no* and *bai sa* do occur, albeit infrequently, in the NI sample but are not found at all in the WH sample.

Preliminary analysis also shows interesting differences in the use of *bin* with other verbs, such as *gat*, and in collocations with other aspect markers, such as *sa*. For example, *bin gat* occurs nine times in the NI sample, *sa gat* twice and *bin sa* eight times. None of these collocations is found in the WH sample. The combination and ordering of modal and aspect particles could be one area where regional differences may become apparent in the future.

5. DISCUSSION: THE ORIGINS OF SYNTAX IN DISCOURSE

Sankoff and Brown (1976), in an important paper on the development of syntactic structures from discourse features, show how the particle *ia* became grammaticalised in the development of relative clause delimitation in the Tok Pisin spoken by young people in Buang. Some of the data presented here support the existence of an analogous route for the origin or elaboration of syntactic structures.

In some cases it appears that features are open to various interpretations. In the case of *nabaut*, described above, there may be some ambiguity about whether the verb or the noun in a verb phrase is being qualified, which could be the origin of innovative syntactic properties. Similarly with *longen*, the original role in anaphoric reference could be reinterpreted where the word occurred to indicate relative clause marking where the word occurred at the end of a clause. The role of *i*, in addition to marking the boundary between subject and predicate and other functions in indicating directionals and continuous action, appears to include discourse

marking, at least in the NI sample. And *kirap* in discourse appears to have a special role with the word *tok*, namely in indicating the opening move of a dialogue. Could *kirap* develop further as an auxiliary verb or even an aspect marker in future generations?

6. CONCLUSION

A comparison of the samples of speech from WH and NI shows both a remarkable uniformity and some interesting differences. Although the size of the samples is small, it appears that the extent of dialect differentiation is strictly limited, and manifested statistically rather than as a set of distinct phonological, lexical and syntactic features. There seems to be a differentiation of the function of some words such as *nabaut*, *i*, *stap*, *taim* and *longen*, which suggests significant regional variation. It remains to be seen whether such differentiation will become more marked with time or be held in check by social and regional mobility and the influence of more conservative second language speakers (Smith 1990a). Nevertheless, regional differences must be taken into account before making generalisations about the direction of change in creolised Tok Pisin in Papua New Guinea as a whole.

Concordance programs such as Oxford *Microconcord* and Longman *Mini-concordancer* have proved to be useful tools for an initial analysis of lexical and syntactic features. Patterns of use and differences in the frequency of occurrence can be observed much more easily than by examining text or recorded data alone. However, the value of these programs lies mainly in identifying features of potential interest, which can then be subjected to a fuller analysis.

The importance of large corpora of written and spoken language is increasingly being realised in the English-speaking world (Aarts et al. 1993). Dictionaries and grammars based on the evidence of use in the real world of the language community are taking the place of more idealised systems derived from pedagogical tradition or the intuitions of the writer. In Papua New Guinea, a large national corpus based on the spoken language of people throughout the country would be a most useful resource for language planning purposes.

REFERENCES

- Aarts, Jan, P. De Haan and N. Oostdijk, eds, 1993, *English language corpora: design, analysis and exploitation*. Amsterdam: Rodopi.
- Bergmann, Ulrich, ed., 1982, *Save na mekim: buk bilong kain kain wok na kain kain samting bilong helpim sindaun bilong yumi*. Lae: Liklik Buk Information Centre.
- Bible Society, 1966, *Niupela Testamen*. Port Moresby.
- Brash, Elton, 1971, Tok Pilai, Tok Piksa na Tok Bokis: imaginative dimensions in Melanesian Pidgin. *Kivung* 4/1:12–20.
- Chandler, Brian, 1989, *Longman Mini-concordancer version 1.01*. London: Longman.
- Dutton, T.E., 1973, *Conversational New Guinea Pidgin*. PL, D–12.
- Dutton Tom and R. Michael Bourke, 1990, *Taim* in Tok Pisin: an interesting variation in use from the Southern Highlands of Papua New Guinea. In Verhaar, ed. 1990:251–262.
- Franklin, Karl J., 1980, The particles *i* and *na* in Tok Pisin. *Kivung* 12/2:134–144.
- Goulden, Rick J., 1990, *The Melanesian content in Tok Pisin*. PL, B–104.
- Keesing, Roger M., 1988, *Melanesian Pidgin and the Oceanic substrate*. Stanford: Stanford University Press.
- Laycock, Don, 1970, *Materials in New Guinea Pidgin (Coastal and Lowlands)*. PL, D–5.

- Lynch, John, 1979, Changes in Tok Pisin morphology. Paper presented at the Thirteenth Papua New Guinea Linguistic Society Congress, Port Moresby.
- Mihalic, Francis, 1971, *The Jacaranda dictionary and grammar of Melanesian Pidgin*. Milton, Qld: Jacaranda Press.
- 1982, *Stail buk bilong Wantok Niuspepa*. Port Moresby: Word Publishing.
- Mosel, Ulrike, 1980, *Tolai and Tok Pisin: the influence of the substratum on the development of New Guinea Pidgin*. *PL*, B-73.
- Mühlhäusler, Peter, 1975, Sociolects in New Guinea Pidgin. In K.A. McElhanon, ed. *Tok Pisin i go we?*, 59-75. *Kivung* Special Publication No. 1. Port Moresby: Linguistic Society of Papua New Guinea.
- 1977, On regional dialects in New Guinea Pidgin. In S.A. Wurm, ed. *New Guinea area languages and language study*, vol. 3: *Language, culture, society, and the modern world*, 533-537. *PL*, C-40.
- 1978, Samoan Plantation Pidgin English and the origin of New Guinea Pidgin. *Papers in pidgin and creole linguistics* No. 1, 67-119. *PL*, A-54.
- 1979, *Growth and structure of the lexicon of New Guinea Pidgin*. *PL*, C-52.
- 1985, Variation in Tok Pisin. In S.A. Wurm and P. Mühlhäusler, eds *Handbook of Tok Pisin (New Guinea Pidgin)*, 233-273. *PL*, C-70.
- 1987, Tracing the predicate marker in the Pacific. *English World-Wide* 8/1:97-121.
- 1990, On the origins of the predicate marker in Tok Pisin. In Verhaar, ed. 1990:235-249.
- 1991, Watching girls pass by in Tok Pisin. In Jenny Cheshire, ed. *English around the world: sociolinguistic perspectives*, 637-646. Cambridge: Cambridge University Press.
- Nekitel, Otto M., 1994, Review of Francis Byrne and John Holm, eds *Atlantic meets Pacific: a global view of pidginization and creolization*. *Language and Linguistics in Melanesia* 25/2:194-196.
- Piau, Julie Ann, 1981, Kuman classificatory verbs. *Language and Linguistics in Melanesia* 13/1-2:3-31.
- Romaine, Suzanne, 1984, Relative clauses in child language, pidgins and creoles. *Australian Journal of Linguistics* 4/2:257-281.
- 1990a, Change and variation in the use of *bai* in young children's creolized Tok Pisin in Morobe Province. In Verhaar, ed. 1990:187-203.
- 1990b, Pidgin English advertising. In Christopher Ricks and Leonard Michaels, eds *The state of the language*, 195-203. Berkeley: University of California Press.
- 1992, *Language, education and development: urban and rural Tok Pisin in Papua New Guinea*. Oxford: Clarendon Press.
- 1993, The decline of predicate marking in Tok Pisin. In Francis Byrne and John Holm, eds *Atlantic meets Pacific: a global view of pidginization and creolization*, 251-260. Amsterdam: Benjamins.
- Sankoff, Gillian, 1980, *The social life of language*. Philadelphia: University of Pennsylvania Press.
- Sankoff, Gillian and Penelope Brown, 1976, The origins of syntax in discourse: a case study of Tok Pisin relatives. *Language* 52/3:631-666.
- Sankoff, Gillian and Suzanne Laberge, 1973, On the acquisition of native speakers by a language. *Kivung* 6/1:32-47.
- Scott, Mike and Tim Johns, 1993, *Microconcord*. Oxford: Oxford University Press.
- Siegel, Jeff, 1983, Media Tok Pisin. *Papers in pidgin and creole linguistics* No. 3, 81-92. *PL*, A-65.
- Smeall, Christopher, 1975, A quantitative analysis of variation: *i* in Tok Pisin. In Cogen, Cathey et al., eds *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society.

- Smith, Geoff P.**, 1990a, Creolised Tok Pisin: uniformity and variation. In R.B. Baldauf Jr and Allan Luke, eds *Language planning and education in Australasia and the South Pacific*, 197–209. Clevedon: Multilingual Matters.
- 1990b, Idiomatic Tok Pisin and referential adequacy. In Verhaar, ed. 1990:275–287.
- 1994, Husat i bin Othoraizim?: new verbs in Manus Tok Pisin. Paper presented to the Seventh International Conference of the Society for Pidgin and Creole Linguistics, Georgetown, Guyana, August 1994.
- Thomas, Dicks R.**, 1990, A course in practical Tok Pisin. In Verhaar, ed. 1990:399–409.
- Verharr, John W.M.**, ed., 1990, *Melanesian Pidgin and Tok Pisin*. Amsterdam: Benjamins.
- 1991, The function of i in Tok Pisin. *Journal of Pidgin and Creole Languages* 6/2:231–266.
- Woolford, Ellen**, 1979, Variation and change in the i 'predicate marker' of New Guinea Tok Pisin. *Papers in pidgin and creole linguistics* No. 2, 37–49. PL, A–57.
- Wurm, S.A.**, ed., 1971, *New Guinea Highlands Pidgin: course materials*. PL, D–3.