Journal of the Linguistic Society of Papua New Guinea

ISSN: 0023-1959

Vol. 30 No.1, 2012

Language in Papua New Guinea: the Value of Census Data

Adam Blaxter Paliwala
Department of Linguistics
University of Sydney
adam.paliwala@gmail.com

Abstract:

One of the most linguistically diverse nations on earth, Papua New Guinea, has consistently included a question about language in censuses conducted in the past (Laycock 1985; National Statistical Office 1994, 2002b).

Laycock (1985) presents language data from the censuses conducted under the Australian Administration in 1966 and 1971 and draws some conclusions about the distributions of Indigenous languages, the two National Language pidgins (Tok Pisin and Hiri Motu), and of English, the Official language. He also raises some concerns about a change in the format and focus of the language question in the first census taken post-Independence, in 1980.

With reference to more recent scholarship and the data from the most recent reported censuses in 1990 and 2000, this paper considers the changing form of the language question and its impact on the value of data collected for longitudinal analyses of language in PNG.

In particular, we address the shift in the focus of the language question from competence in 1966 and 1971, to an assessment of domains of language use in 1980, and then to a singular focus on literacy in 1990 and in 2000.

This paper presents a discussion of this changing data set and an analysis of its value in assessing long-term language trends from 1966 to 2000, and into the future.

In addition, this paper presents basic figures drawn from the 2000 census on the languages used in Papua New Guinea, revealing the important reality of multi-language skills in the country. It considers the value of the data collected post-1985 for linguists' understanding of the linguistic makeup of the national community, and for pidgin and creole studies, in particular[1].

1   Introduction

The research and analysis that led to the production of this paper was begun during fieldwork in Papua New Guinea (PNG) in 2002. My PhD project was on the development of Tok Pisin (TP) post-Independence, and I had been focusing on adult Tok Pisin / English bilingualism as a key factor. The publication of the National Statistical Office (NSO) Reports on the 2000 National Census (National Statistical Office 2002a, 2002b) allowed me to consider the extent to which such bilingualism existed in the country.

In my work analyzing the data from that Census, I attempted a comparison between the information gathered on language in PNG through the 2000 National Census with the information gathered in censuses taken before 2000.

An important discussion by Laycock (1985) of trends in language use and literacy drawn from the pre-Independence censuses in 1966 and 1971 highlighted some key issues we face when attempting to use the census data on language to understand the changes in language use in PNG over the 20th Century.

This paper presents some of the questions raised by such a comparison, not only in regards to Tok Pisin and English but also into language in PNG more generally. It also provides some interpretations of the 2000 National Census data on language in PNG.

1.1   Data & Method

Data from the tables presented in the 2000 National Census Reports (National Statistical Office 2000a, 2000b) and other sources (see Table 1) was entered by hand into Microsoft Excel spreadsheets. The Excel software allowed calculations and transformations to aggregate the reported figures and to allow comparisons with earlier reported census data on language use, on literacy, on multilingualism, on differences between rural and urban communities, and between age groups within the national population. Through using this software I was also able to produce the raw data figures in graphical form, and this paper presents a range of graphical figures that visibly show the underlying patterns and relationships within the Census data.

Table 1 lists the various sources consulted for Census data:

Census Year	Data Source
1966-1971 :	Laycock, D. (1985). Tok Pisin and the census. *The Handbook of Tok Pisin*. S. A. Wurm and P. Mühlhäusler: 223-231.
	Sankoff, G. (1980). Multilingualism in Papua New Guinea. *Social Life of Language*. G. Sankoff: 95-132.
1980 :	Romaine, S. (1992). *Language, education, and development : urban and rural Tok Pisin in Papua New Guinea*. Oxford, Clarendon Press.
	Jenkins, R. S. (2000). *Language Contact and Composite Structures in New Ireland*, Papua New Guinea. PhD, University of South Carolina
	National Statistical Office, C. (1994). *Report on the 1990 National Population and Housing Census in Papua New Guinea*. Port Moresby, National Statistical Office: 397
1990 :	National Statistical Office, C. (1994). *Report on the 1990 National Population and Housing Census in Papua New Guinea*. Port Moresby, National Statistical Office
2000 :	National Statistical Office, C. (2002a). *2000 Census Basic Tables - National Level*. Port Moresby, National Statistical Office.
	Sankoff, D. (2008). How to predict the evolution of a bilingual community. *Social lives in language*. M. Meyerhoff and N. Nagy. Amsterdam, John Benjamins: 179-194.

Table 1. Sources of Census Data for this study

Where this paper presents charts drawn from this data they are the result of careful analysis of the underlying data. With many different combinations of languages reported in the 2000 National Census, both the original data tables (National Statistical Office 2000a) and a pie-chart (Figure 1) representing the data directly as reported are somewhat confusing:
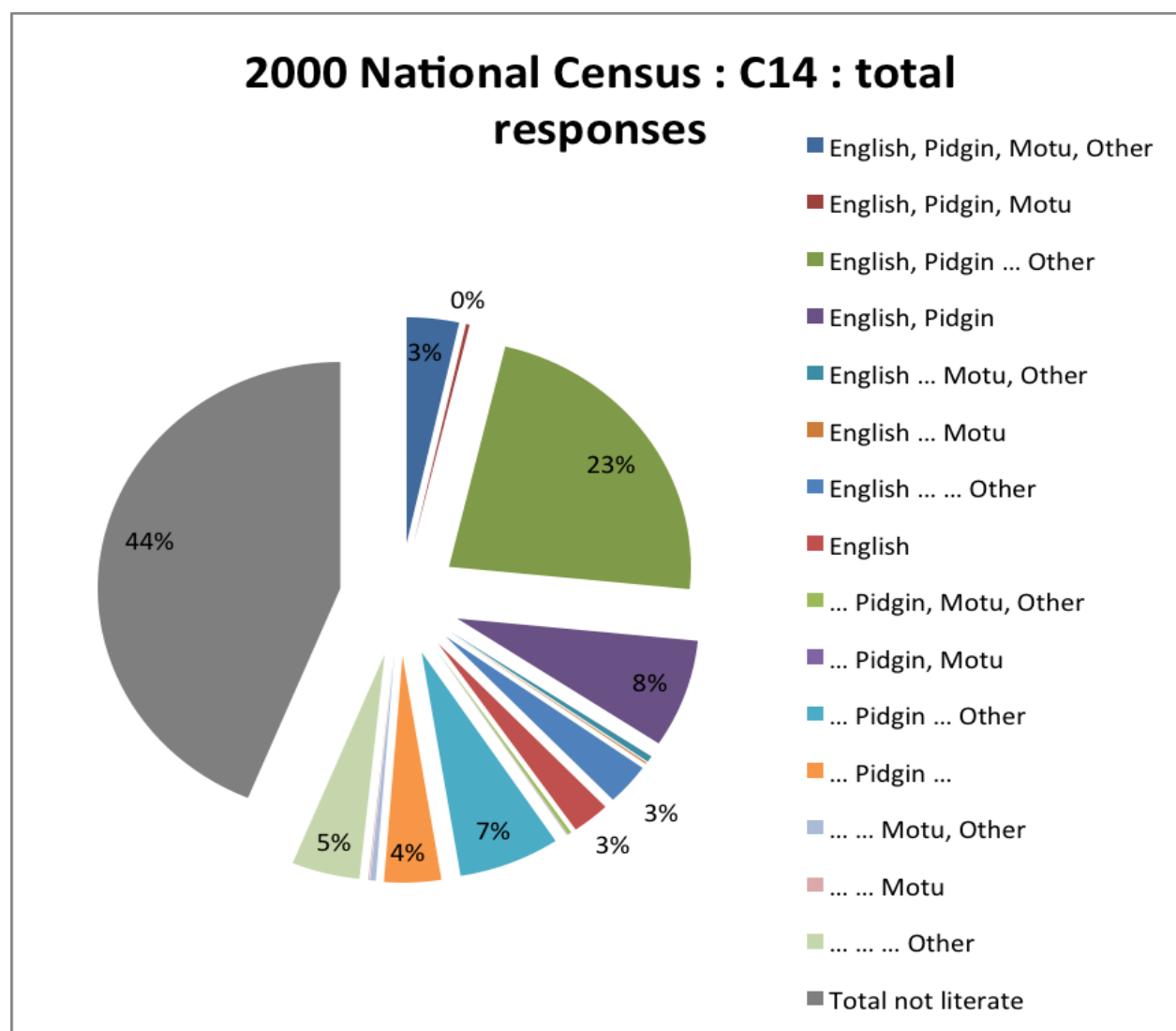
Figure 1. 2000 National Census: C14 : total responses

While such a representation, and indeed the raw data figures themselves, does allow us to see e.g. the largest groups of responses, we need to perform calculations in order to answer broader questions. Alongside its discussion of the patterns within, and value of, Census data, this paper provides charts as various Figures based on simple analyses of the raw data. These charts give a clearer picture of underlying patterns of language use reported in the 2000 National Census.

1.2   Outline

In the next section (s2) the changing nature of the language question on the census is itself addressed. The crucial association between language skills and education is introduced and discussed in the third section (s3). Questions are raised about using educational attainment to calculate national literacy rates and the question of literacy as a baseline for understanding language ability is addressed. In the final part of this paper (s4) we present some basic

findings from the existing census data, based on our answers to the questions posed in the previous sections.

2   The Language Question

Papua New Guinea is and remains one of the world's most linguistically diverse nations, and a question on language has been included in all of the long form questionnaires of each of the six censuses taken since 1966 (Laycock 1985; National Statistical Office 1994, 200b). The size and distribution of the census population, the questions asked, and the answers given, have, however, changed over the years.

This section presents a discussion of the changing national census questions on language (s2.1, s2.2) along with a brief discussion of other sources on language use in PNG (s2.3) and some later analyses of trends in the census data. It provides a background for the discussion and analysis that follows.

2.1   *Laycock*

In his study of the Census data available to him, Laycock (1985) outlined the changes in census-takers' approach to surveying language use in the community. Under the Australian Administration, censuses were taken of people living in urban and semi-urban areas in 1966 and in 1971. The first National Census was taken in 1980. Table 2 presents Laycock's key observations on each of the three censuses he discusses in table form:

	1966 (Laycock 1985: 223-224)	1971 (Laycock 1985: 228)	1980 (Laycock 1985: 230)
Census Question	Asked if people could speak national languages, and if they could read/write simple sentences in national languages, or any other language : a literacy test card was used	Asked for the home language, then: spoken national languages yes/no, with literacy question for 'yes' answers, and a literacy question for 'any other language'	Asked for the single language most spoken in house and also the single language most spoken when buying at market
Data Collection	Yes/no answers with test card for literacy	Yes/no answers with a yes/no question for literacy	List of languages including national language, major languages, with *tokples* category
Size of Census	Census was of urban population and 10% of village population		Language question was only on the 'long form' used in urban areas and some rural areas
Specific Observations	Prestige miscalculation, confusion over TP and English	No distinction of 'other' languages	Language USE vs. COMPETENCE

	"taking figures at face value" : TP with 3x English, and 4x Hiri Motu speakers	No assessment of skills or literacy "and it is likely that competence in more languages was claimed than was really the case"	Laycock refers to claim of maximum expansion for TP (1982) at 45% (p. 232)
	TP low in some regions, e.g. Highlands	Some errors in figures e.g. Decline in TP in New Ireland and West Sepik	
		Publication of population errors	
		No estimates of first language TP …	

Table 2. Laycock's (1985) observations on early census

Crucially, while the 1966 and 1971 Census specifically asked questions about people's ability to speak different languages and literacy in those languages, the 1980 Census took a different approach, asking respondents to list their primary 'home' and 'market' languages from a list of regional lingua franche. As Laycock observed:

> [I]t is difficult to see how they [the 1980 data] can be compared with the language data of previous censuses.

(Laycock 1985: 231)

Laycock also observed publicized errors in the population estimates on which 1980 census data was based which may make comparing the percentages of speakers in various languages problematic. Writing before the 1980 data was published, Laycock discussed the expansion of Tok Pisin as the fastest growing language in PNG 1966-1971, and, significantly, it's apparent acquisition alongside English in the school context.

The question of whether Tok Pisin had reached a level of maximum expansion was left open, though Laycock suggested that that point might already have been reached.

Laycock's discussion provides us with vital early summaries of the data on language collected by the Australian Administration, as well as alerting us to the problems with data that only represented a small proportion of the Papua New Guinean population. In addition he observes that the prestige associated with multilingualism in some communities may have led people to over-estimate their language abilities, and so highlights a key concern about the role of self-reporting of language ability in skewing the census data. Laycock also highlights the problem that this paper addresses at length: of comparing census data from the different surveys with each other.

While the earliest censuses were limited in the proportion of the population surveyed, they collected rich data on the range of languages people spoke, and also whether they could read or write those languages. The 1966 census actually tested literacy skills to substantiate people's statements of ability. The first census of the Independent State of Papua New

Guinea, in contrast, targeted an urban/rural subset of the national population for a 'long form' questionnaire which recorded the primary languages in use at home and in the marketplace.

Though the 1980 census collected very different data to the 1966 and 1971 censuses, the focus on the primary languages used in different domains provided data that was linguistically very valuable, enabling Mühlhäusler to observe that there were "tens of thousands of households where Tok Pisin is the principal language" (1985: 149).

2.2   1990 and 2000 National Census

The change made in the 1980 Census language question reflects a post-independence shift in interest from language competence, and literacy as an indicator of development, to language use, and the role of languages in the community.

In the 1990 Census, however, there was a re-focus on recording language literacy.

The pre-Independence question about whether a person could speak a particular language was gone, and with it an indicator of verbal vs. non-verbal competence in the community. The census simply asked if subjects were literate in English, Tok Pisin, Motu, or any other language. The test was if a subject could read and write simple messages or letters in a language, but no specific identification of language skills was made.

The manual issued to census Enumerators clarified the question by emphasizing that a person was considered as able to read and write a language only if he or she could both read and write with understanding "a short simple letter or message in that language" (National Statistical Office 1994) not simply their name or a memorized phrase.

This was in common with the 1966 and 1971 census questions.

However, unlike those previous census questions, there were no specific tests of ability:

> [I]t was assumed that as in most cases the enumerator was appointed from the locality, he also knew who was able to read and write with understanding.

(National Statistical Office 1994)

The, more recent, 2000 census was conducted using a single one-page form, printed entirely in English. The task of interpreting the questions for non-English-literate informants was left to English-speakers in individual households or the community at large, or to the census enumerators.

In 2000 the language question was again framed in terms of literacy. Language skills were only assessed for persons over 10 years of age, along with advanced educational attainments and details of employment.

Question 14 asked "Which languages can the person read and write with understanding?" (National Statistical Office 2000a). The four options given were: English, Pidgin, Motu, *tokples*. Yes/No answers were provided by placing a mark in the appropriate check-box.

As a result of the change to an exclusive focus on literacy post-1990, any attempt to understand the change in language use in PNG over time using census data raises a number of crucial questions.
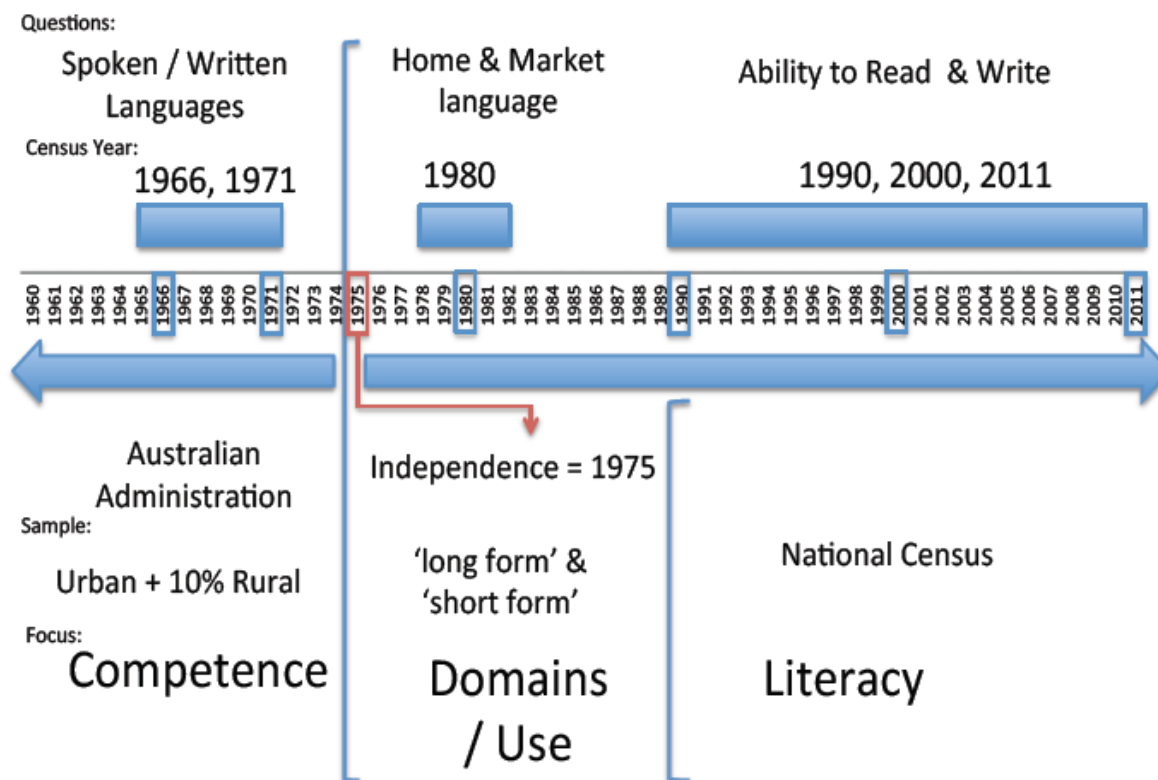
Figure 2. Papua New Guinea Census timeline

Figure 2 presents a timeline showing the 5 censuses we will discuss in this paper[2], and indicates the different phases in the evolution of the census 'language question' in PNG.

2.3   Other Work on Census and PNG

In addition to Laycock's (1985) work a number of other scholarly sources are useful in understanding the changing picture of language use in PNG in the late 20th century.

2.3.1   Noel

Noel (1975) considers the languages used by government employees during the period leading up to Independence in 1975. His estimates of the use of Tok Pisin in the House of Assembly and within the Government (reproduced here as Table 3) indicate the rapid growth in the use of that language during the period between the 1966 and 1980 censuses that Laycock discusses:

---

[2] The 2010 National Census was taken in June 2011 and no data has been reported at the time of writing.

House of Assembly		% Tok Pisin
1964-68		40%
1968-72		60%
1972-73		95%
A.E.C., to Cabinet and Ministry		
1964-68	Under secretary	25%
1968-72	A.E.C.	45%
1972-73	Cabinet	85%

Table 3 : Noel (1975) observations on Government use of Tok Pisin

Writing in 1975, he indicated the significance of Tok Pisin in the official roles of other Government employees, estimating the on-duty and off-duty use of this language, reproduced here in Table 4:

	On-Duty Tok Pisin use	**Off-Duty Tok Pisin use**
Labour Officers—12	80%	20%
Kiaps—45	95%	5%
Didman—30	95.5%	4.5%
Business Extension Officers—6	95.5%	4.5%
P.W.D. Officers—10	80.5%	19.5%
Local Government Council Advisers—8	90%	10%
D.I.E.S Extension Officers—10	90%	10%
Government Liaison Officers	80%	20%
Police	95%	5%

Table 4. Noel (1975) observations on use of Tok Pisin by Government Employees

Noel observes that the significant role of English in Government communications accounted for the difference between the percentages presented here as Table 4. He also noted an important difference between the home language use of people in urban and rural areas:

> Most Papua New Guineans in urban centres speak a great deal of Pidgin at home. The rural areas in the country vary the usage of Pidgin due to the fact that Motu, English, or local languages are used in those circumstances.

(Noel 1975: 80)

2.3.2   G. Sankoff

Gillian Sankoff (1980) also discussed changes in language use with reference to census figures, and made an important observation about the role of Tok Pisin in bilingualism with other languages:

> Tok Pisin is currently the most important language in Papua New Guinea as far as bilingualism is concerned.

(Sankoff 1980: 125)

She also observed that in 1980 there were twice as many speakers of Tok Pisin than in the earlier census, and that the language had a much greater national spread.

2.3.3   Jenkins

In her PhD dissertation, Jenkins (2000: 20) compares the 1971 census figures reported by G Sankoff with those noted by Romaine (1992: 87), and observes that there has been very little change between the two sets of figures. Romaine was reporting figures retrieved from existing statistical records (pers comm). These may have simply reproduced the 1971 figures in the absence of new data on literacy from the 1980 census question

On the basis of these figures and her own work in New Ireland and elsewhere, Jenkins predicted that English language use as a spoken language would stabilize at around 20-25% of the population. Jenkins also predicted that Tok Pisin would increase in use over time, while Hiri Motu would decline in use over time.

We will consider Jenkins' predictions with reference to the 2000 National Census data below (s4.2).

2.3.4   D. Sankoff

David Sankoff (2008) considered the use of census data for the application of his DMLX software to provide demolinguistic projections for the evolution of a bilingual community. Working as a statistician for the Bureau of Statistics of the Australian Administration in 1967 he produced a report of projections over the short and mid-term, which were later confirmed by his superior to be "bang on" (Sankoff 2008: 183). The method he developed in PNG provided the basis for the DMLX software, which provides projections for the impact of language-revival programs.

When D Sankoff considered the patterns present in the PNG 2000 National Census data he found that the focus on literacy and the absence of language-revival indices made his program inapplicable. This highlights the continuing importance of the questions raised by Laycock on the comparability of the different sets of census data: how to compare the figures on competence, domains/use, and literacy collected over the 5 reported censuses.

D Sankoff did make a few observations on the basis of the 2000 National Census data, including the higher literacy rates in urban populations and the fact that the age-spread of Tok Pisin literacy shows a build-up in ability up to about 40 years ago in males, with stability

thereafter, while female rates have continued to rise. He observed that English follows the same pattern, though with a lag, and now is at higher rates than Tok Pisin in the school-age population (Sankoff 2008: 191).

D Sankoff observed that Tok Pisin was a widespread feature of the male urban work environment and less so in rural areas and in women's workplaces. He also emphasized that social, educational and political evolution is key to understanding the levels of knowledge, use and literacy in Tok Pisin and English in PNG, noting in particular "a pattern of extensive usage of English hinted at by the 2000 literacy statistics" (Sankoff 2008: 192).

He also observes a trend in the 2000 National Census data on literacy that he explains with reference to non-formal learning:

> Acquisition continues later for Tok Pisin than English, suggesting that there is a greater tendency to acquire literacy in Tok Pisin as a young adult, presumably in the work place, whereas English acquisition is more a function of formal schooling.

(Sankoff 2008: 191-192)

We will return to this statement later, in our interpretation of some questions raised by the 2000 National Census data (s3.2.1).

2.4   Summary

This section has established that the concerns raised by Laycock about the changing nature of the language question are important for a consideration of the value of census data today.

As well as highlighting a question about the reliability of self-reported language ability, Laycock (1985) also draws our attention to the difference between ability to speak a language and the ability to read and write it, and the difference between questions about what different languages people may know and which are the main languages they use. While the census questions in 1990 and 2000 essentially collected the same data, as did the 2011 census, some work must be done not only to interpret that data but also if we are to make any reasonable comparisons between these and previous censuses. The next section will discuss this in detail.

3   Looking at Census Data: questions and answers

3.1   2000 National Census: Literacy

The basic figures in the report on the 2000 National Census (National Statistical Office 2000a) allow us to directly answer some essential questions about the literacy reported in different languages. The following charts and figures are compiled from the basic table data (C14) directly presented as Figure 1, above.

Of the recorded Census population of c. 5.14 million in 2000, those over 10 years old who were expected to answer the language question were c. 3.7 million people. Of these, 55% reported literacy in at least one language:
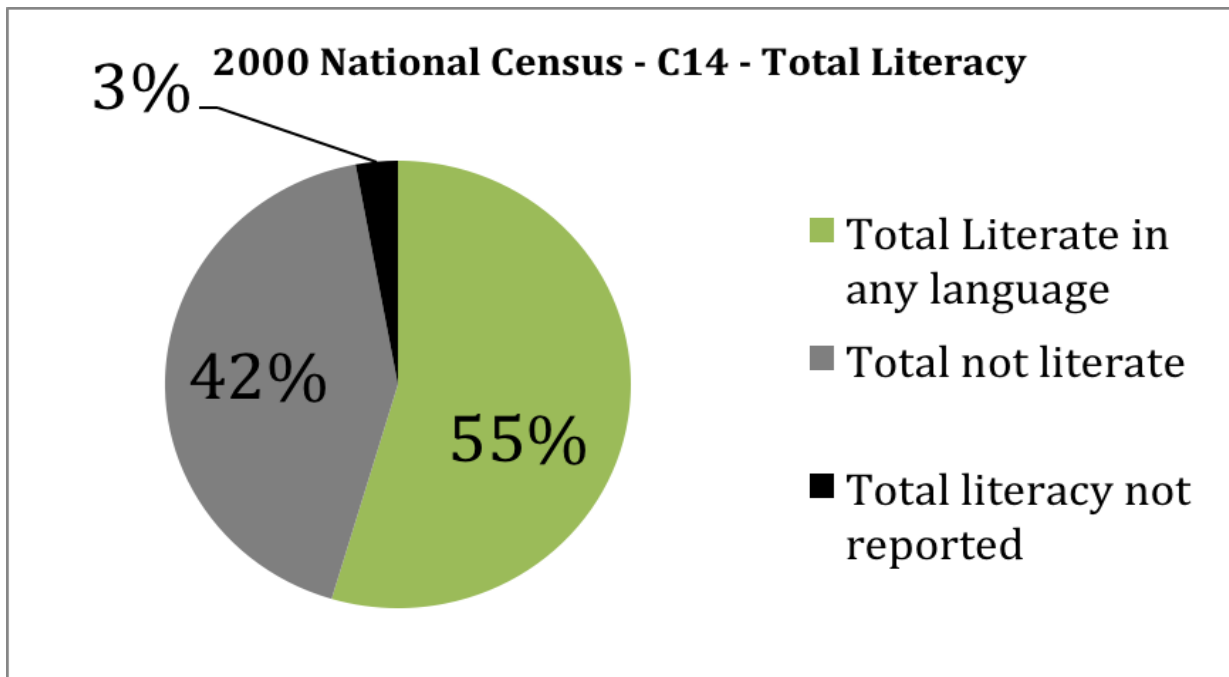
Figure 3. 2000 National Census: Question C14 : literacy in 10+ population

The c. 42-45% illiterate population according to the 2000 National Census compares very favorably with the pre-Independence 1971 figure of c. 71% discussed by Jenkins, and is a 10% improvement on the 55% illiteracy rate recorded from the 1990 census.

We can also use the basic figures recorded from responses to the 2000 National Census to look at literacy in the national languages separately:
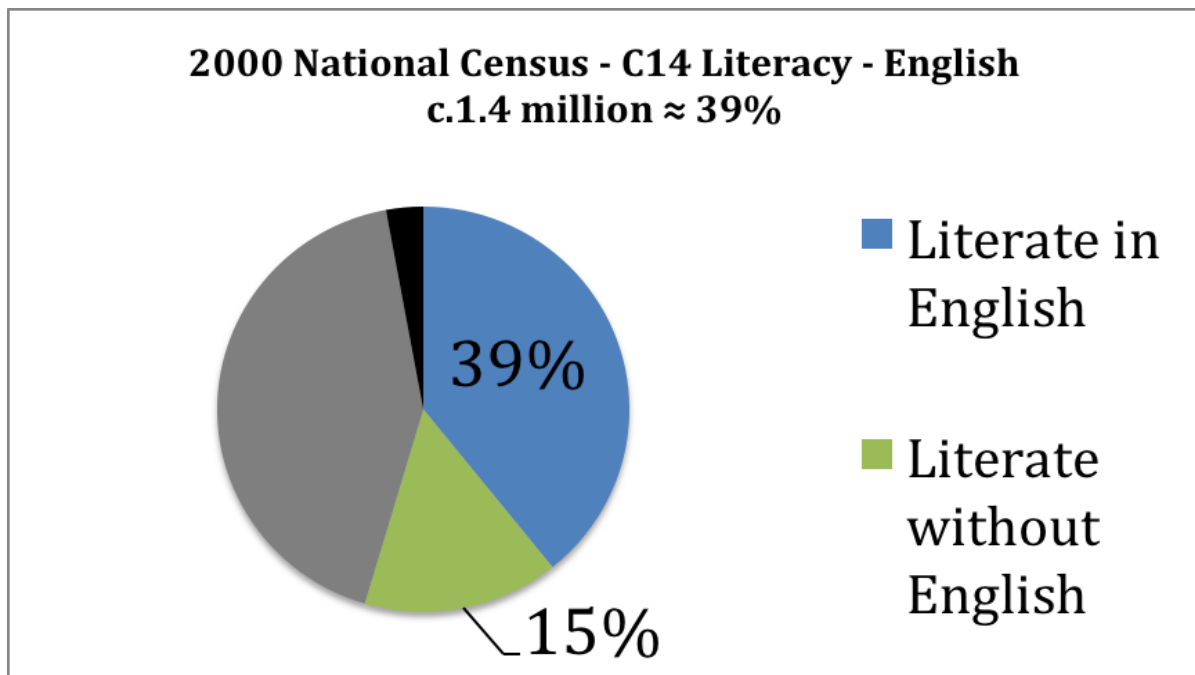
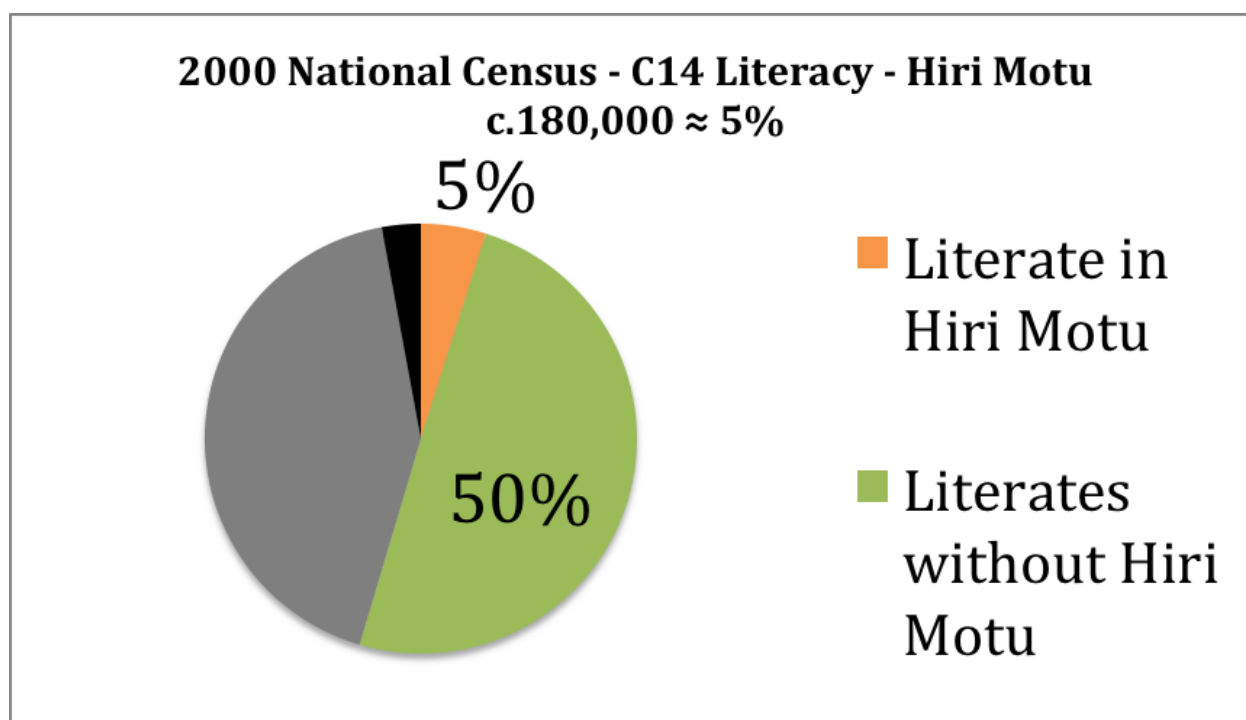**Figure 4. 2000 Literacy - English : c.1.4 million ≈ 39%**
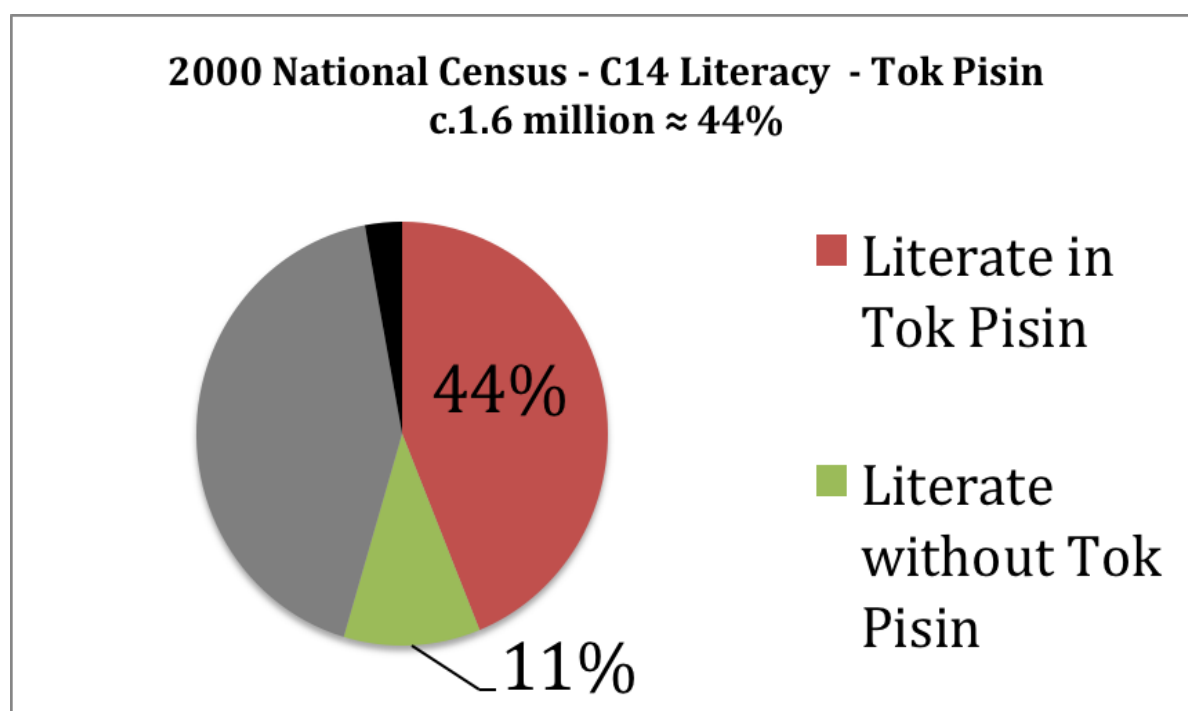
Figure 5. 2000 Literacy – Hiri Motu : c.180,000 ≈ 5%

Figure 6. 2000 Literacy – Tok Pisin : c.1.6 million ≈ 44%

These pie charts enable us to see clearly that the greatest proportion of people reporting literacy in the 2000 National Census reported literacy in Tok Pisin, and that this was slightly larger than the proportion of people reporting literacy in English. The majority of literate people in the country did not report being literate in Hiri Motu.

Where we are interested in long-term language trends in PNG, however, some crucial questions must first be addressed.

3.2   Education, Literacy, and Ability

Looking across the data reported by Laycock, Romaine, and the National Statistical Office, and taking the total percentages of the census populations reporting literacy for each year it was recorded we can see clear growth over time:

Literacy Rates :
1966 : 22% → 1990 : 45% → 2000 : 55%

Figure 7. Census Literacy 1966-200 - basic figures

However, we have gaps in this pattern. Laycock does not report literacy figures for 1971, and the 1980 census asked a very different kind of language question, and collected different data.

Laycock raised some questions about comparing data from 1980 with earlier censuses, and the report on the 1990 Census (National Statistical Office 1994) addressed those questions by providing a re-evaluation of the data from 1980 to estimate literacy based on educational attainment. This was a significant decision.

The national education policy at the time of the 1980 census specified English-medium education, and Grade 3 completion required reading and writing skills, so a decision was made to calculate the 1980 literacy rate as equivalent to the Grade 3 and above attainment. This gave an estimated literacy rate for the 1980 population at 27.7%[3].

The figure for 1980 as calculated in the 1994 report was lower that the 1971 census figure of around 30% literacy. However, this same calculated 1980 value allows for an estimated literacy growth rate per annum 1980-1990 of 6.4%, much higher than the population growth rate of 2.3%. While suggesting a drop in literacy between 1971 and 1980, the calculated figure indicted great success in the most recent decade for this educational index of development.

Using data now available from two subsequent Censuses we can derive additional insights into literacy figures and their relationship to educational attainment and language ability, as well as the distribution of Tok Pisin and its ongoing presence alongside *tokples* and English in the National Community.

Figure 8 shows a comparison between reported and estimated literacy rates. It includes the reported 1966 and 1971 figures alongside the estimated (National Statistical Office 1994) 1980 figure and reported figures from the 1990 and 2000 census. The dip in the line at 1980 where figures are estimated on basis of educational attainment (Grade 3+ completion), compared to the overall upward trend in the data, could be taken as indicating a decline in literacy post-Independence, and a subsequent steeper improvement in literacy 1980-1990.

---

[3] Romaine (1992: 87) includes a national 'illiteracy' percentage for 1980 that indicates national literacy is, at 32%, 5% higher than the estimated figure from the 1994 report. It is not clear from the nature of the 1980 census question, as Laycock (1985) describes it, how this 'illiteracy' figure would have been derived.
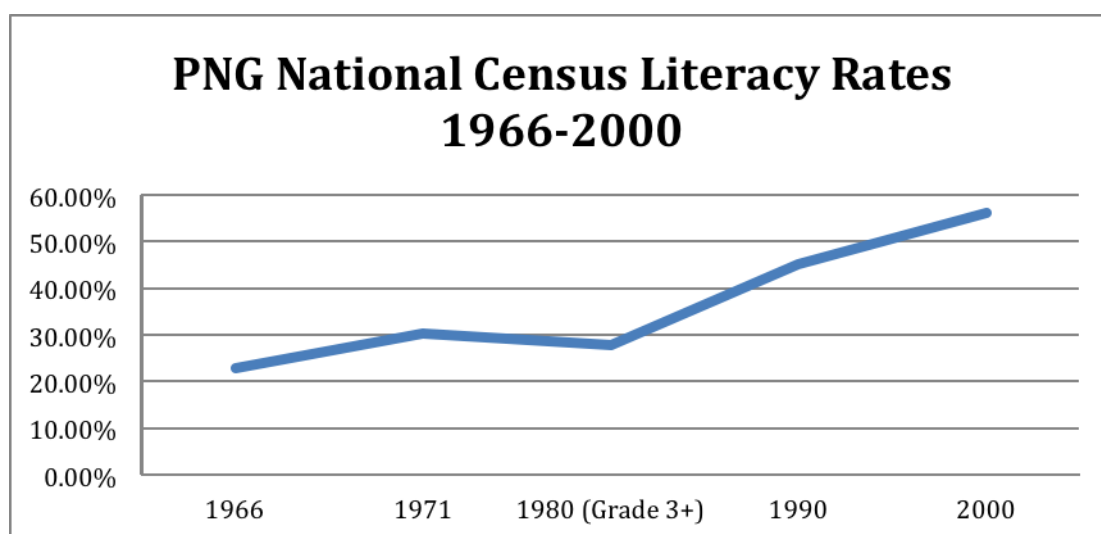
Figure 8. National Census Literacy (Nb. 1980 figure estimated on basis of educational attainment)

Though the association between education and literacy is a natural one, however, a consideration of the 2000 National Census data (National Statistical Office 2002a) reveals a 10% difference between the percentage of the sample population completing grade 3+ (c. 46%) and the percentage of over 10 year olds reporting literacy in one or more languages (c. 56%). This is equivalent to around 500,000 citizens, and the difference is present even though only citizens over 10 years of age answered the literacy question, whereas the education statistics could include children under 10.

The 10% discrepancy between a similarly estimated value for the 2000 census and the actual recorded figure suggests that the estimated 1980 literacy rate may be similarly low. Two possible explanations for this will be considered in the next section.

While observing the 'literacy gap' between literacy through educational attainment and the figures recording responses to question 14 in the 2000 National Census may help us to adjust our projections over time, the fact remains that the pre-Independence and 1980 censuses only recorded language data from the urban population and a percentage of people in rural communities (Laycock 1985).

As a consequence, even the earliest literacy figures may be questionable. As Crowley observed with reference to his analysis of language data from the census in Vanuatu:

> …care needs to be exercised in how the figures are to be interpreted.

(Crowley 1994: 2)

The following sections explore some different questions raised by the 2000 National Census data.
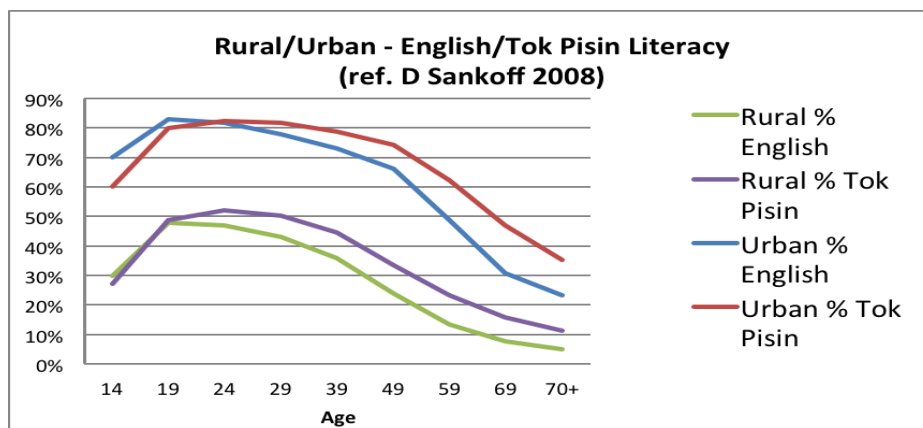
3.2.1   Education and Literacy

D Sankoff (2008) discusses the relationship between education and literacy skills in his analysis of the 2000 Census data.

As described above, statistically significant in the Census 2000 figures on literacy are approximately half a million citizens who reported reading and writing ability at question 14 who had not completed grade 3 education.

This represents approximately 10% of the population of Papua New Guinea.

As observed above, D Sankoff's interpretation of this difference is that there are large numbers of citizens who have gained literacy skills outside of formal education – perhaps through life-experience or community-based learning in the workplace. Figure 9 reproduces D Sankoff's (2008) bar chart as a line graph of literacy % in English and Tok Pisin by age. The overlapping lines for both Urban and Rural people illustrate an increase in Tok Pisin literacy for the 19-24 year age group as well as higher English literacy for those of school age:

Figure 9. 2000 National Census literacy by age: Rural/Urban English/Tok Pisin Literacy: ref. D Sankoff 2008

Before considering an alternative explanation for the difference between educationally-attained literacy and the self-reported figures in the 2000 National Census, we can use the 10% 'literacy gap' between Educational attainment and reported literacy figures in the 2000 National Census to adjust the 1980 Literacy figure calculated in the 1994 Report.

Figure 10 shows the calculated 1980 and 2000 figures alongside all the reported literacy figures, and marks the 1980 figure with the same difference as found in the 2000 census.
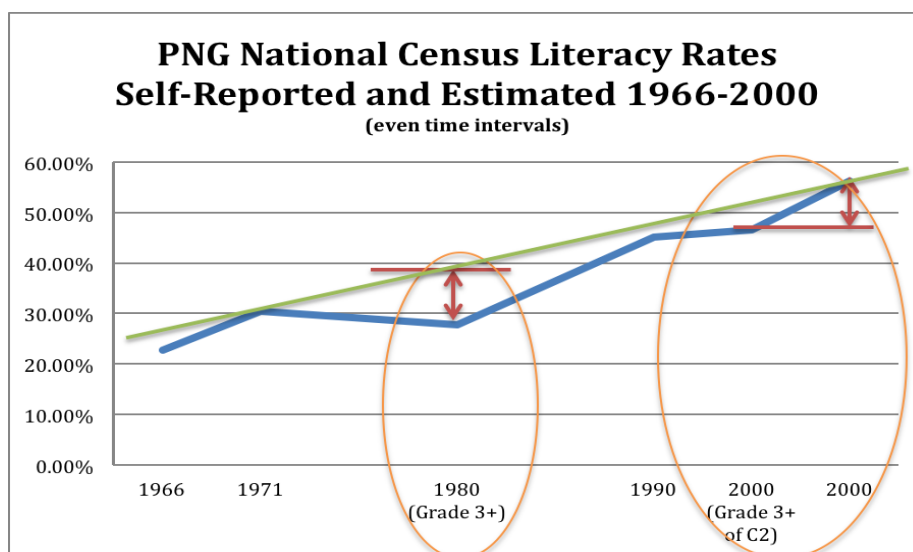
Figure 10. PNG Census Literacy Rates - Reported and Estimated

The adjusted (+10%) 1980 figure lies close to a linear increase (straight green line) in literacy over time. This suggests that a gap between educational attainment and self-reported literacy may be a consistent feature of PNG national census data.

3.2.2   Literacy in Indigenous Languages

In investigating the 'literacy gap' we may also compare this apparent finding of literacy skills in the absence of formal education with the 2000 National Census data on literacy in 'other' languages.

The NSO basic data indicates that 1,494,475 respondents (c.40%) reported that they could read and write an 'other' language, usually a *tokples*, traditional vernacular, with understanding.
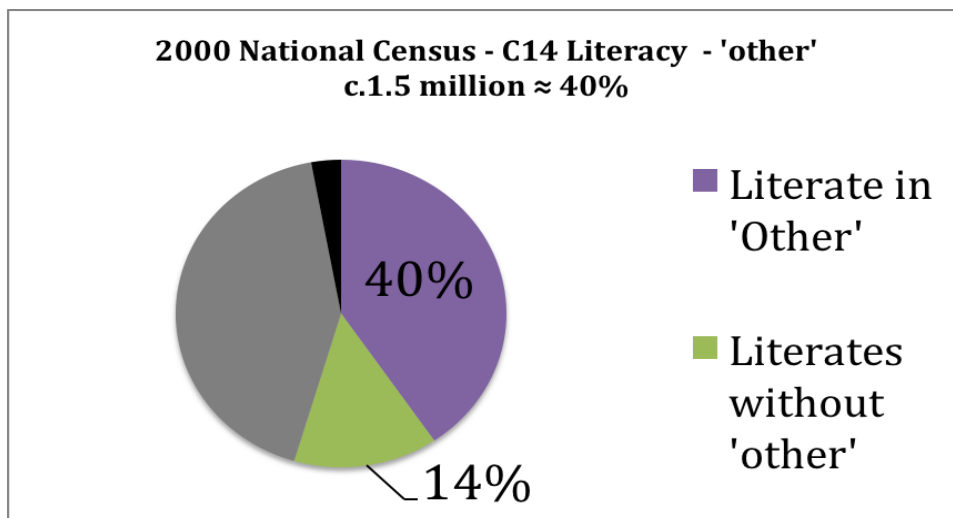

Figure 11. 2000 Literacy – 'Other': c.1.5 million ≈ 40%

This literacy in traditional languages is larger than literacy in English, the formal language of education. Significantly 162,914 respondents claimed literacy in *tokples* alone, indicating that they had learned to read and write a *tokples* without ever using English or a lingua franca as an intermediate or supporting language of literacy.
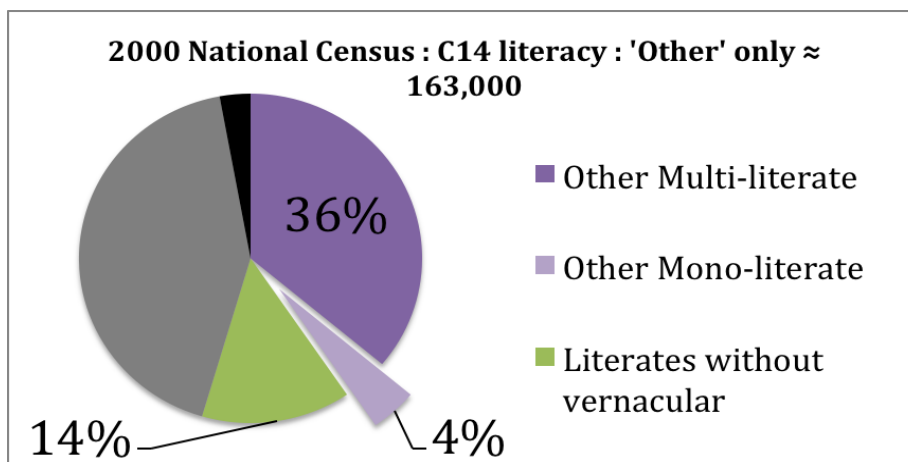

Figure 12. 2000 Literacy: 'other' mono-literate ≈ 4%

On one level, this looks like an indication of great success for vernacular maintenance programs, adult literacy, and the Tok Ples Pri Skul (TPSS) Grade 1&2 vernacular education schemes which have grown in popularity since 1980 (Siegel 1998).

However, a closer look at the numbers suggests that even with 1994 enrolments of around 8,000 adults in literacy programs, which may include *tokples* literacy, and with 80,000 Grade 1&2 students being taught basic literacy in *tokples*, a significant rate of expansion would have been required for these programs to produce 1.5 million 'other' literates in time for the 2000 census. If we also consider that some TPSS programs were in Tok Pisin, the reported figure remains hard to reconcile. The success of indigenous literacy is not called into question, but the massive success indicated by the 2000 National Census figures calls for some consideration.

Without digging into regional data for an answer to this question posed by this substantial community of mono-literate vernacular responses, let alone the very substantial multi-literate responses, as to how such literacy skills might have been acquired, I gained insight from discussing the 2000 census language question with people who had participated in the census.

3.2.3   Answering the Language Question

In the 2000 Census, enumerators were relied on as the gatekeepers of information from respondents. They were considered to be local authorities able to correct false evaluations by respondents. On the language question, it was assumed that they would know to a reasonable degree who could read and write particular languages.

However, our interviews revealed a patterned assumption among respondents that the ability to read and write English or Tok Pisin implied a similar ability to read and write in a *tokples* that was also known to an individual, even if there had never ben an attempt to read or write in that language.

This interpretation is supported by the 2000 Report table C4 (National Statistical Office 2002a), where we find significantly higher numbers of respondents ticked 'yes' for literacy in English, Pidgin, and *tokples* together (818,864) than for literacy in English and Pidgin alone (277,776).

Similarly for other combinations of languages: people tended to include the ability to read and write a *tokples* with literacy in English, Tok Pisin, Motu, or any combination of them, consistently more frequently than they recorded an ability in the national languages without being able to read and write *tokples*. This may have encouraged the diversity of responses illustrated by Figure 1, above. Furthermore, of householders who had participated in the 2000 census and interviewed by this researcher two years later about their experience, less than 50% could recall if a language question was included on the enumeration form at all. Among those who could recall the presence of a language question, only c. 35% specifically identified it as a question on reading and writing, that is literacy, skills.

The majority of my informants appear to have interpreted the 2000 National Census Question 14 on which languages they could "read and write with understanding" as a more general question on language competence: on what languages they knew or used.

This was true even for informants entirely comfortable with written and spoken English, for whom the monolingual census enumeration forms should have presented no obstacle to interpretation and understanding.

The 'literacy gap' could, perhaps, be a result of a conflation of reading and writing skills with language skills more generally, a consequence of the literacy question being the only question on language on the census, or of misinterpretations, or mistranslations, of the precise meaning of Question 14.

It is possible, therefore, that over-estimations of literacy may provide an alternative explanation for the gap between Grade 3+ educational attainment and literacy reported in the census.

Laycock has various reasons for remaining somewhat skeptical about the figures he discusses (Laycock 1985: 224, 228). Among these he includes concerns about the impacts of self-reporting and the prestige associated with multilingualism on census data on language in PNG.

Responses to the census 'language question' in 2000, and possibly before, may not be entirely accurate indicators of literacy in particular languages. We might expect, however, that those reporting multi-literacy would certainly be practiced in reading and writing at least one of those languages. Yet, the 2000 National Census data can still be shown to have high value as a baseline for understanding language distributions and multilingualism in PNG.
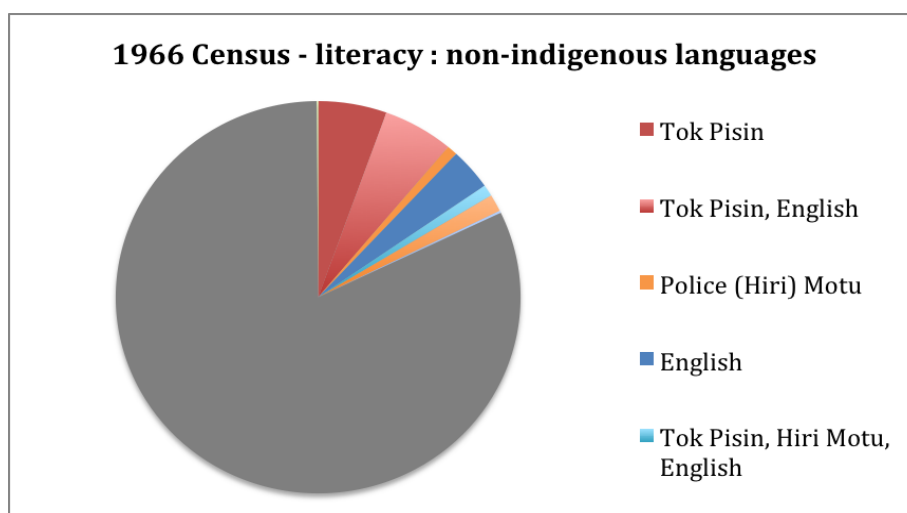
3.3   Literacy as a baseline

Even with skepticism about census data we can use the reported numbers to draw some conclusions about the patterns of language use across the national community. Though people may, as Laycock noted in regards to the earlier surveys, choose to over-emphasize multi-lingual skills as a symbol of prestige, we can use the literacy figures as a baseline indicator.
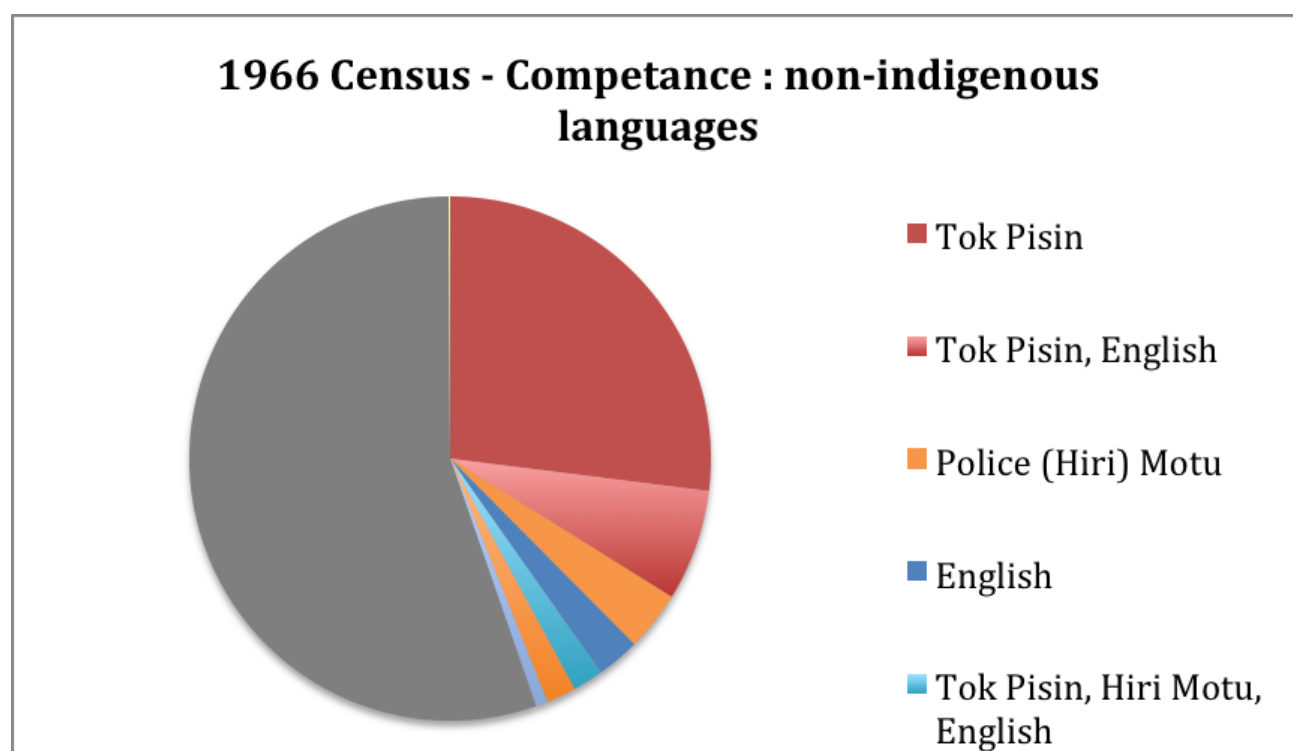
Some general patterns do emerge (s4).

Focusing on non-indigenous languages, we can also consider the value of the 1990-2000 question on 'literacy' as an indicator of general 'ability' in language.

Comparing the proportions of the 1966 survey reporting literacy in non-indigenous languages with those reporting competency more generally we can clearly see that those people reporting literacy are only a portion of those actually able to use a language:

Figure 13. 1966 Literacy: non-indigenous languages (from Laycock 1985b)

1966 Census - Competance : non-indigenous languages

- Tok Pisin
- Tok Pisin, English
- Police (Hiri) Motu
- English
- Tok Pisin, Hiri Motu, English

Figure 14. 1966 Competence : non-indigenous languages (from Laycock 1985b)

We can clearly see that more people in 1966 were reporting spoken ability (competence: Figure 14) than written ability (literacy: Figure 13).

On this basis we may presume that literacy rates are generally lower than competence. Certainly, the 45% of Papua New Guinea's population who are reporting non-literate[4] in the 2000 Census are speaking some language.

In 1966 the number of people reporting an ability to speak a non-indigenous language was roughly three times the number of people recorded with an ability to read and write it. Though this is not necessarily the case today, the implication is that any indicator of literacy in the community can be assumed to underpin an equivalent, if not much larger, community of speakers of those languages, and possibly other languages as well.

While Laycock does not provide figures for literacy from the 1971 census, in comparing the spoken language figures for the 1966 and 1971 census surveys there is clear and significant growth in reported ability in Tok Pisin and English together. Compare Figure 14 (above) with Figure 15:

---

[4] All sections of the pie charts here reflecting non-literate people are shaded grey.
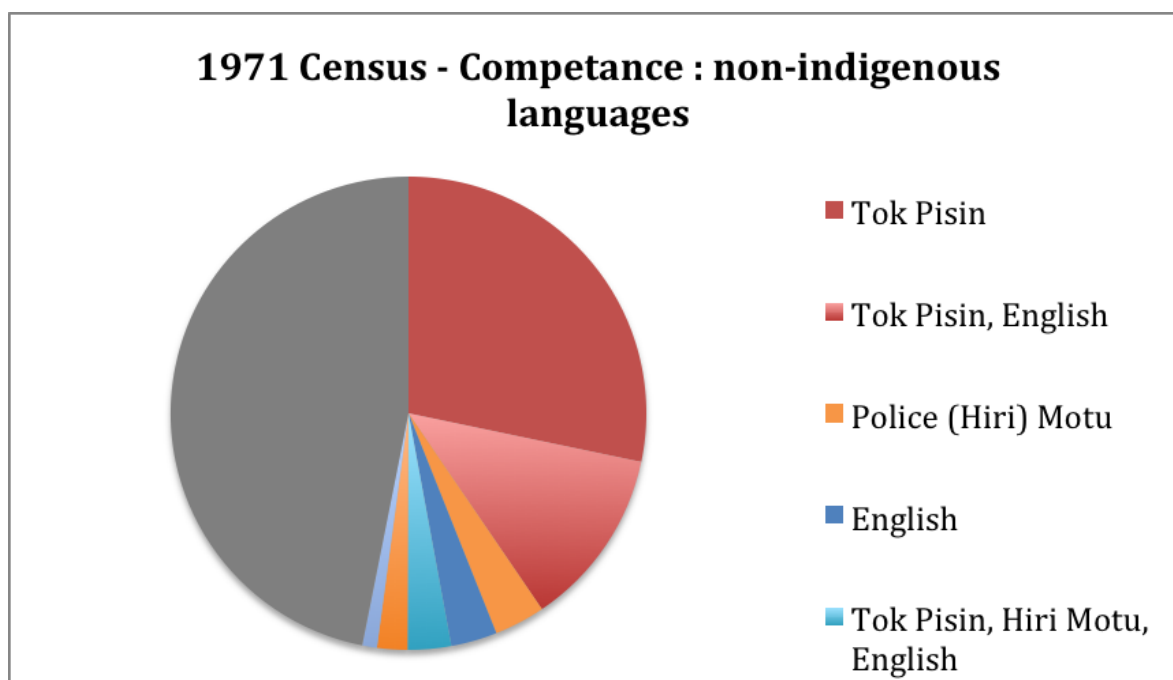
Figure 15. 1971 Competence: non-indigenous languages (ref. Laycock 1985b)

The growth in English/Tok Pisin bilingualism was something I was particularly interested in for my PhD research, but a number of other findings emerged from my analysis.

Holding in mind the appreciation of the significant questions about census data on literacy as a true indicator of literacy in any particular language, about the underlying relationship between education and literacy and between spoken language ability and literacy developed in this section, the final section of this paper discusses some findings from the 2000 National Census language question, and looks at overall trends from 1966 to 2000.

4   Patterns in the Census Data

This section discusses findings from the 2000 National Census data in terms of basic figures on rural and urban literacy (s4.1), the spread of the national languages (s4.2), and in terms of the information they reveal about multi-language skills (s4.3). Some particular insights about Tok Pisin are highlighted (s4.4).

4.1   Rural and Urban Literacy

Urban literacy rates, and use of Tok Pisin and English generally, are regularly higher than those of rural communities, and this is borne out by the 2000 census data.
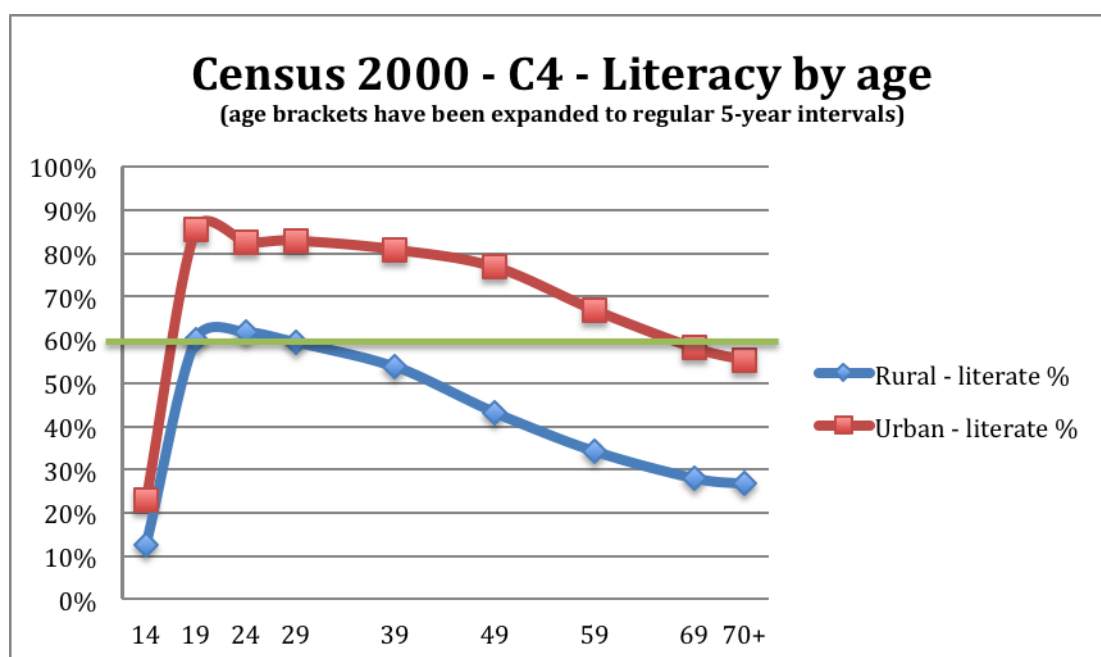
Figure 16. 2000 Literacy - Rural/Urban by age

One interesting observation that can be drawn from the 2000 National Census data is that the percentage of the rural population under 24-29 years of age acquiring literacy in 2000 is equivalent to the percentage of under 59-69 year-old urban literates (60% : the green line on Figure 16).

4.2 *Spread of National Languages*

To track the changes of language ability over time I compiled all the language data from the different censuses in our sources (listed in Table1) into a single resource.

Table 5 represents summary comparisons of the findings of the six National Censuses.

Source	Laycock 1985	Laycock 1985 Table 3	Laycock 1985	Romaine 1992	Romaine 1992	Report 1994	Report 2002
	1966	**(Literacy) 1966**	**1971**	**1980**	**(Literacy) 1980**	**(Literacy) 1990**	**(Literacy) 2000**
English	13.27%	11.41%	20.37%	22.34%	20.17%	29.1%	39.2%
Tok Pisin	36.5%	12.23%	44.49%	45%	20.2%	30.7%	43.9%
Motu	8.14%	3.38%	9.48%	9.4%	4.5%	6%	4.9%
Other		15.93%			14.23%	38.5%	40.5%

Table 5. Comparing reported census speaking/literacy percentages for different languages

21

A line chart showing the change in literacy rates using these figures is below (Figure 17), and clearly indicates the rise of Tok Pisin and English literacy, as well as the more equivocal status of Hiri Motu on the national scale over the period.
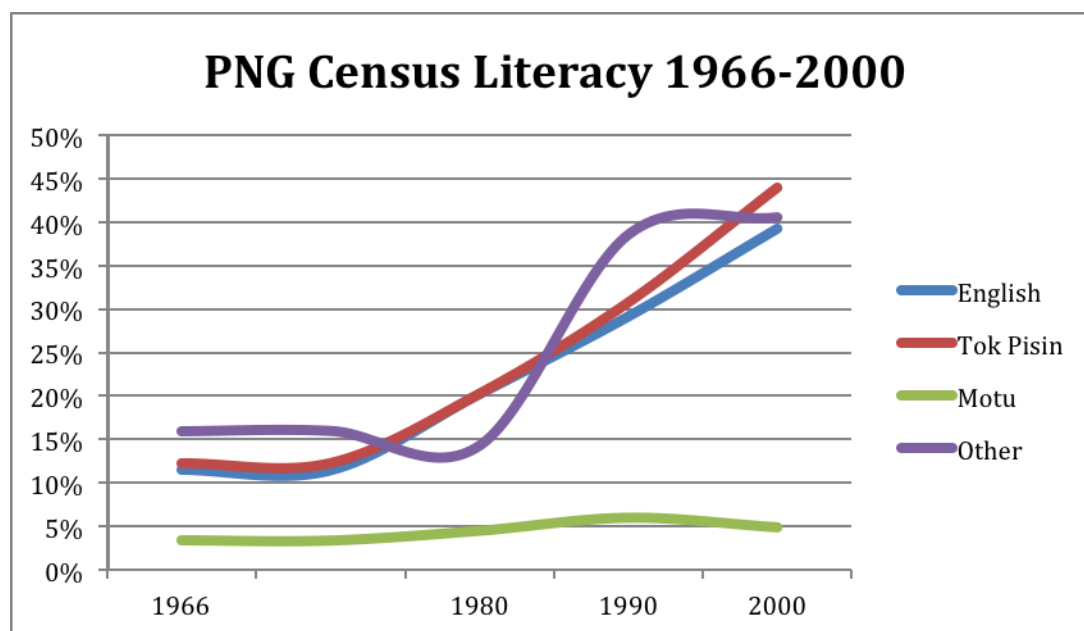

Figure 17. PNG Census literacy 1966-2000 - by language

The steep increase in 'Other', presumably indigenous language *tokples,* literacy from 1980 to 1990 may be best explained as a result of the common assumptions discussed by my informants in regards to the 2000 census above: as a correlate of literacy in the national languages and as a consequence of competence/literacy confusion.

This line chart clearly illustrates that recorded literacy in Hiri Motu 1990-2000 has declined slightly, while in all other languages it has increased.

If we compare this with Jenkins (2000: 20) speculations (s2.3.3 above) we may find that, if literacy figures are any indication, her prediction regarding Hiri Motu is borne out. However, in terms of English language ability remaining at 20-25% she may not be correct. Though, as we discussed above, the 2000 National Census figures do not aim to record actual frequency of use or general ability in the language, just reading and writing "with understanding", 40% literacy in English is substantially higher than Jenkins' estimate.

In terms of Tok Pisin both Jenkins' (2000) speculation that it would increase slightly and Laycock's (1985) suggestion that it's use might have reached maximum extensions only hold up if we assume the 2000 National Census question was interpreted universally as regarding speaking ability, rather than literacy.

However, even if a metric based on 1966 data is only impressionistic and flawed, we can assume that reported literacy levels will be lower than similarly reported general ability levels. As a result the percentage of Tok Pisin speakers in the National community is higher than the reported literacy figure of c. 44%: it is unlikely that Tok Pisin use plateaued at c. 45% (the 1971 figure for speakers that was the latest available to Laycock (1985)), if 44% of people over the age of 10 claim to be able to read and write it.

Many people who speak a language may not be able to read or write: the 45% of people not reporting literacy or not responding to the language question, the 'grey area' of language and

literacy data (labeled in Figure 3, and shaded grey throughout the charts here), are certainly speaking something.

4.3   Impacts of Multilingualism with Indigenous Languages on National Languages

We have already considered (Figure 4 – Figure 6) the relative proportions of the population reporting literacy in each of the three national languages separately, and discussed some of the implications and interpretations of the high percentage of people reporting literacy in an 'other' *tokples* language (Figure 11). As discussed above, the figures on literacy from the 2000 National Census may only provide a benchmark for actual spoken multilingual skills, a superficial impression of the true languages spoken by Papua New Guineans which bundles together some 800+ indigenous languages under a single 'other' category and provides no information on the language skills of the 45% of the population who do not report literacy in any languages. However, the census data is of value in comparing the proportion of multi-literacy with 'other' languages for those reporting literate in each of the national languages.

The fundamental issue here is in terms of understanding the ongoing impacts indigenous languages may have on the national languages. For the pidgins Tok Pisin and Hiri Motu these may be considered 'substrate' effects. More significantly, however, they also indicate something about the overlap between the national languages and traditional languages in the lives of Papua New Guineans.

Figure 18 presents an image of all (c. 180,000) respondents who indicated literacy in Hiri Motu, subdivided into sections in terms of the additional languages they reported literacy in. It may be thought of as a pie made up of the single slice (5%) of the national population shown in Figure 4. 2000 Literacy - English : c.1.4 million ≈ 39%
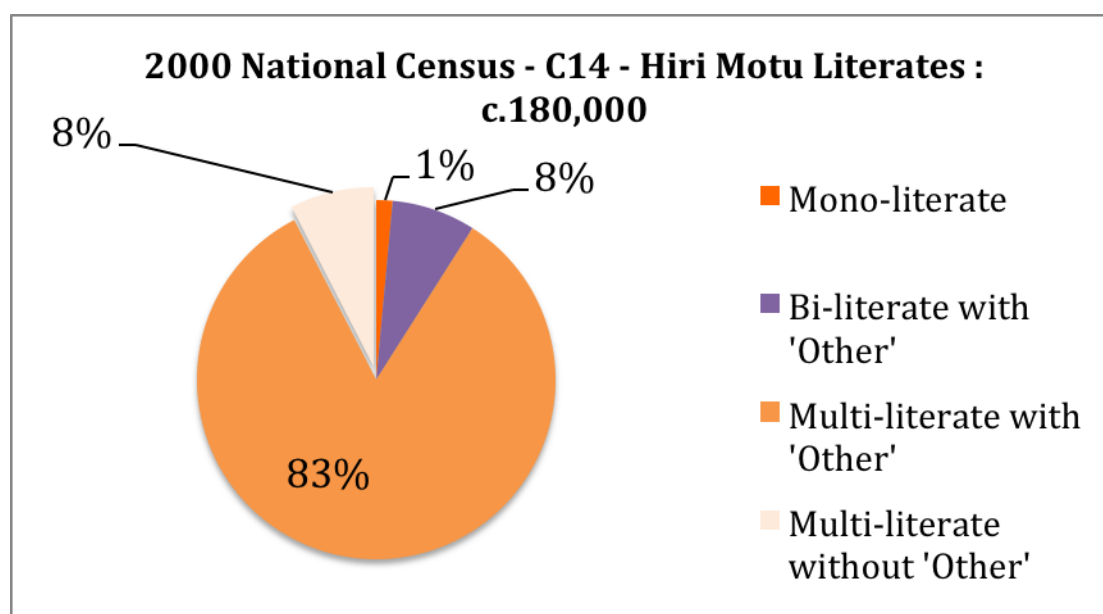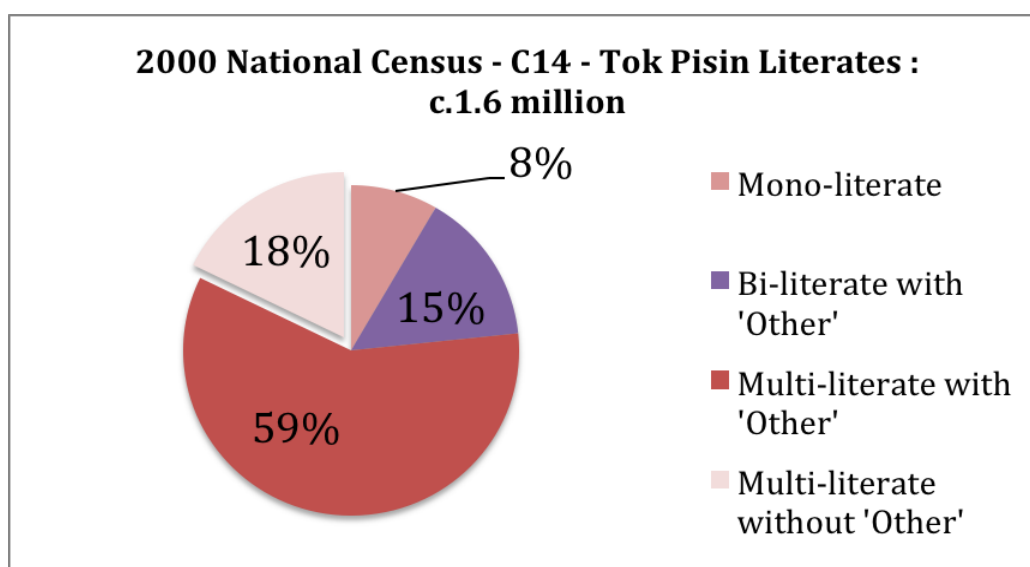

Figure 18. Substrate Effects I – Hiri Motu Literacy with 'other'

From this we can see that only a very small proportion of the Hiri Motu literates (c. 1%) reported that they could only read and write Hiri Motu and no additional languages (mono-

literate). Roughly the same proportion of people reported that they were literate in Hiri Motu and an 'other' language (c. 8%) as reported that they were literate in Hiri Motu and either Tok Pisin or English (or both) without being literate in an 'other', *tokples*, language (c. 8%). The majority of people who were literate in Hiri Motu reported literacy in an 'other', *tokples*, language as well as either English, or Tok Pisin, or both (multi-literate with 'other').

This suggests that most (99%) Hiri Motu literates also know another language. Hiri Motu literates were as likely to rely only on a *tokples* in addition as they were to rely only on the other national languages.

We may compare this with Figure 19, an image of all (c. 1.6 million, c. 44% of total respondents, or 80% of literate) respondents who indicated literacy in Tok Pisin, again subdivided into sections by additional languages. It is an image of the c. 44% 'slice' indicated in Figure 6.
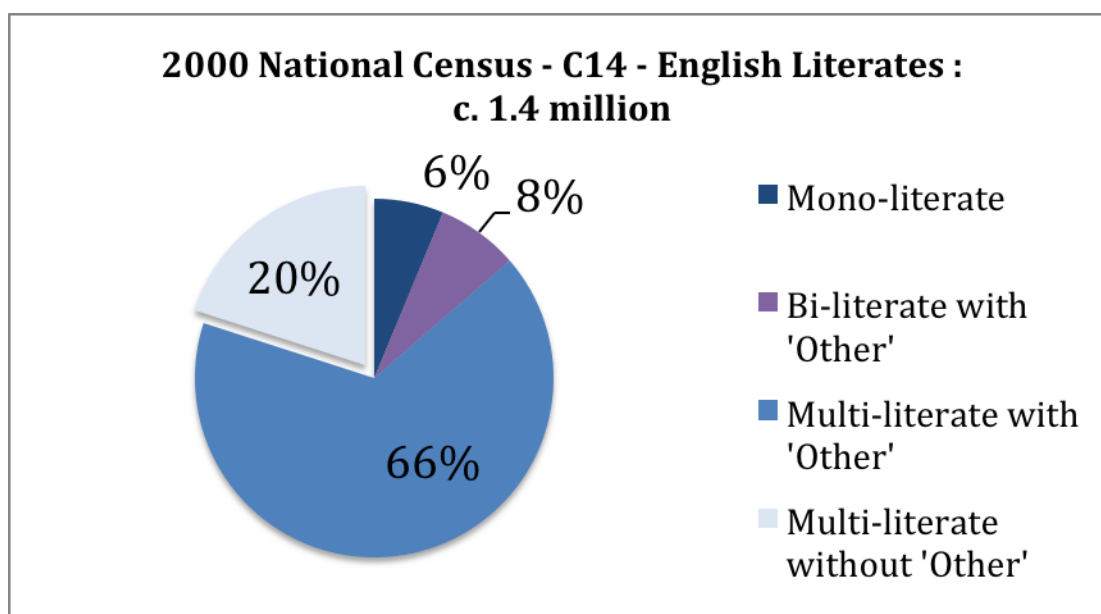

Figure 19. Substrate Effects II – Tok Pisin Literacy with 'other'

Bearing in mind that the real number of people represented by Figure 19 (c. 1.6 million) is slightly less than 9 times that of Figure 18 (c. 186,000) we can also see that the subsections for mono-literacy, bi-literacy with a *tokples* 'other' language (meaning people reporting literacy in Tok Pisin alongside an indigenous language) and multi-literacy without 'other' (meaning that people reported literacy in Tok Pisin and other national languages, but not in a traditional language) are more substantial.

For Tok Pisin literates, proportionally more people reported literacy only in Tok Pisin and a *tokples* (bi-literate with 'other') perhaps reflecting communities where Tok Pisin and *tokples* are used without an additional national language. Again, this was a similar proportion to those people who reported literacy in Tok Pisin and other national languages, without *tokples* literacy (multi-literate without 'other'). These would be people without literacy in *tokples* but with either Hiri Motu or English alongside Tok Pisin. Here there is also a significant (c. 64,000, c. 4% of all respondents or c. 8% of Tok Pisin literates) group of people who reported they were literate in Tok Pisin alone. These people could be considered as first-language 'creole' Tok Pisin speakers. This leads to a potential impact of multi-language skills on 92% of the Tok Pisin literate population.

For the third, official, non-indigenous language, English a slightly different pattern appears. Again, Figure 20 represents the 'slice' of English literates (c. 1.4 million or c. 39% of respondents) as a group divided by multi-literacy.

2000 National Census - C14 - English Literates : c. 1.4 million

- Mono-literate
- Bi-literate with 'Other'
- Multi-literate with 'Other'
- Multi-literate without 'Other'

6%   8%   20%   66%

Figure 20. Substrate Effects III – English Literacy with 'other'

Compared with Hiri Motu and Tok Pisin we can see that it is the group of people reporting literacy in English alone (mono-literate) and in English and an 'other' *tokples* language alone (bi-literate with 'other') that are in roughly equally proportion. This contrasts with the other national languages, where bi-literacy with an 'other' language was as commonly reported as multi-literacy without an 'other' language. For English too the proportion of people reporting multi-literate skills without a *tokples* was the highest of the three national languages.

This suggests that English literacy is more commonly learned in the absence of *tokples* literacy. Additionally, though the proportion of people bi-literate with English and an 'other' language, and not in Tok Pisin or Hiri Motu, is slightly higher than mono-literates, it is not as substantially different as for Hiri Motu or Tok Pisin illiterates: English literacy overlaps with literacy in other languages.

These findings indicate that while Hiri Motu literacy is largely supplemented by literacy in other languages, both Tok Pisin and English literacy skills are reported alone for significant proportions of the population. The 2000 National Census data also shows that English has a place alongside Tok Pisin as a significant language of bilingualism, a clear development from the picture described by G Sankoff of Tok Pisin (1980: 125) (s2.3.2).
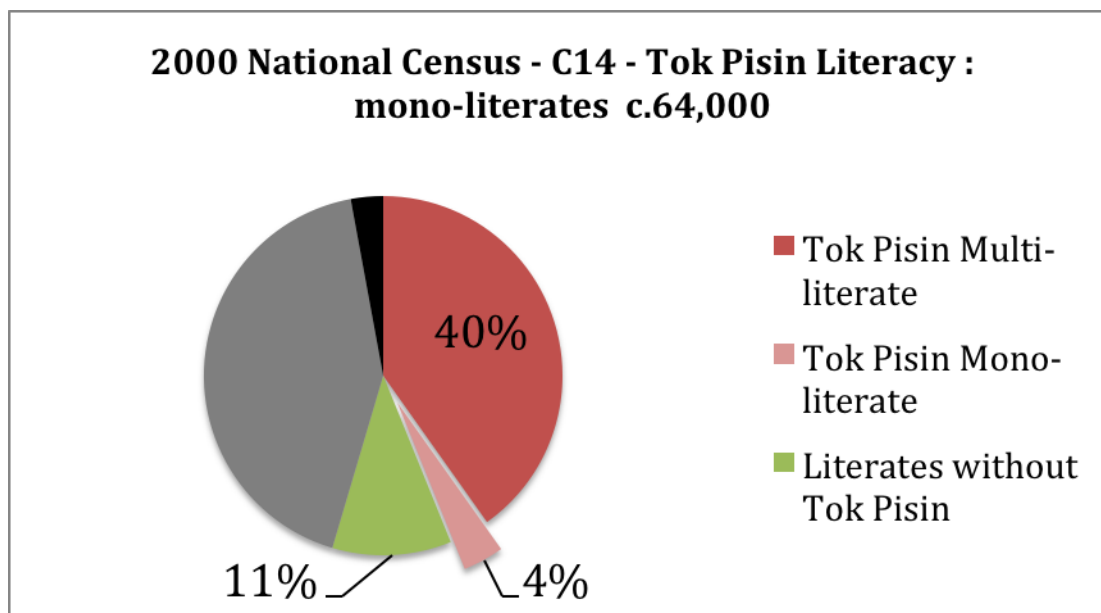
Overall, it is clear that the reported literacy in 'other' languages alongside the national languages indicates a vibrant presence of *tokples* in any picture of national multilingualism.

4.4   Tok Pisin and English

We can refer to the 2000 National Census literacy figures to draw some conclusions about the use of Tok Pisin in PNG.

4.4.1   Mono-literacy and 'creole' Tok Pisin

As discussed above (s4.3 ref. Figure 19), the c. 4% of the population who reported mono-literacy in Tok Pisin may be considered a baseline for a categorical creole (first language) speaking community (Figure 21):

2000 National Census - C14 - Tok Pisin Literacy : mono-literates c.64,000

40%

11%   4%

- Tok Pisin Multi-literate
- Tok Pisin Mono-literate
- Literates without Tok Pisin

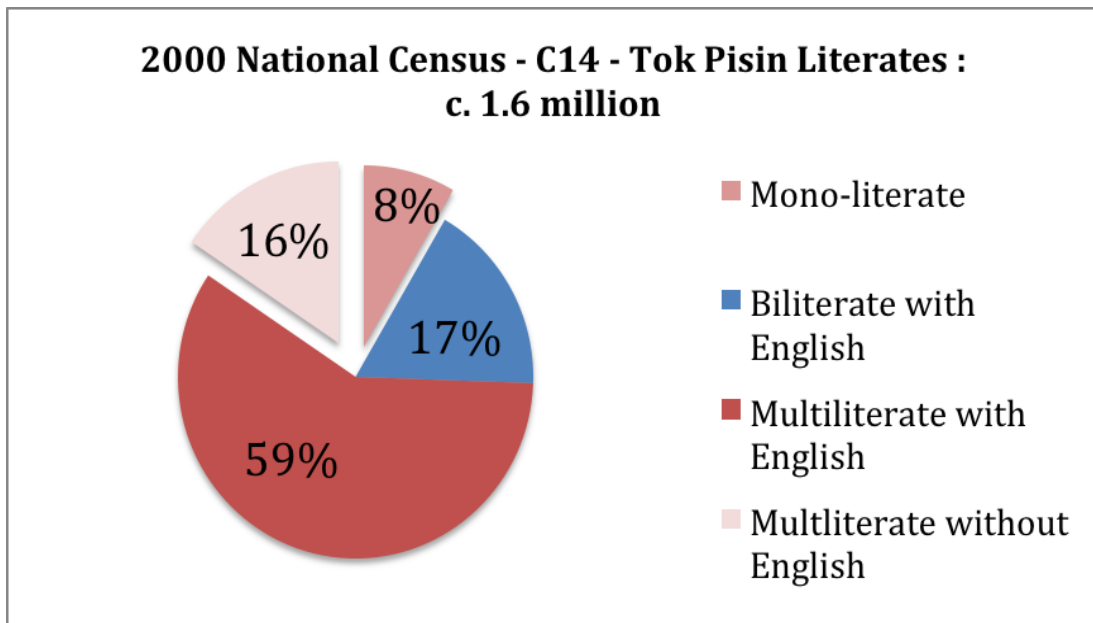**Figure 21. 2000 Literacy - Tok Pisin mono-literacy c.64,000 ≈ 4%**

This represents 8% of all Tok Pisin literates. The remaining 92% of Tok Pisin literates who reported multi-literacy with additional languages may be considered formal 'pidgin' (multilingual) language users. However, these literacy figures tell us nothing about which languages people use most in their day to day lives, or anything about their fluency in those languages.

Mühlhäusler (1979) reports from the 1971 Census c. 91,000 people (c 5.79% of the reporting population) using Tok Pisin as their home language, yet, like the 1980 Census data, this figure cannot be compared to the more recent National Censuses.

Without reference to census data, the Summer Institute of Linguistics (SIL) reports 50,000 Tok Pisin monolinguals (Lewis 2009), which is c. 36% of the 2000 National Census figure (c. 136,000) for Tok Pisin mono-literates. Assuming the SIL figure is correct, this difference indicates that multi-lingualism lies behind much of the mono-literacy reported in the National Census.

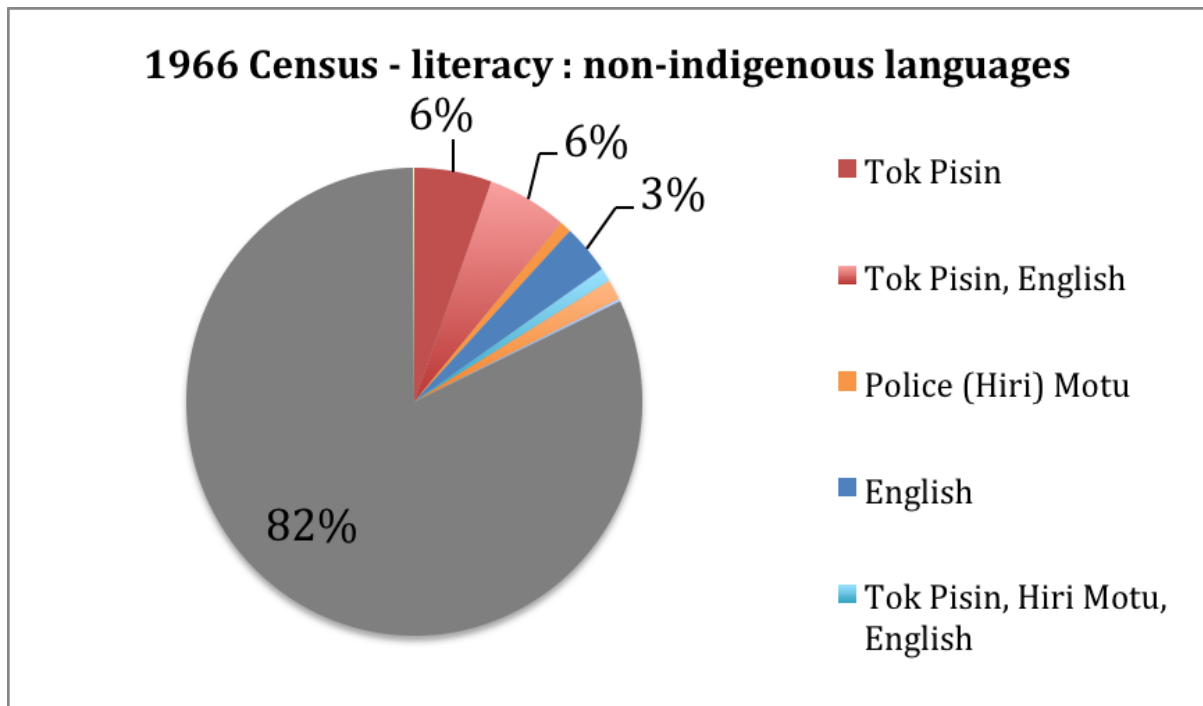### 4.4.2   Multi-literacy: the rise of Tok Pisin + English

In terms of multi-literacy in Papua New Guinea, the 2000 National Census reveals the dominance of combined literacy in both English and Tok Pisin. If we consider the overlap between Tok Pisin literacy and different languages with a focus on English (Figure 22) we can see that three-quarters (c. 76%) of Tok Pisin literates are literate with English.

2000 National Census - C14 - Tok Pisin Literates : c. 1.6 million

- Mono-literate — 8%
- Biliterate with English — 17%
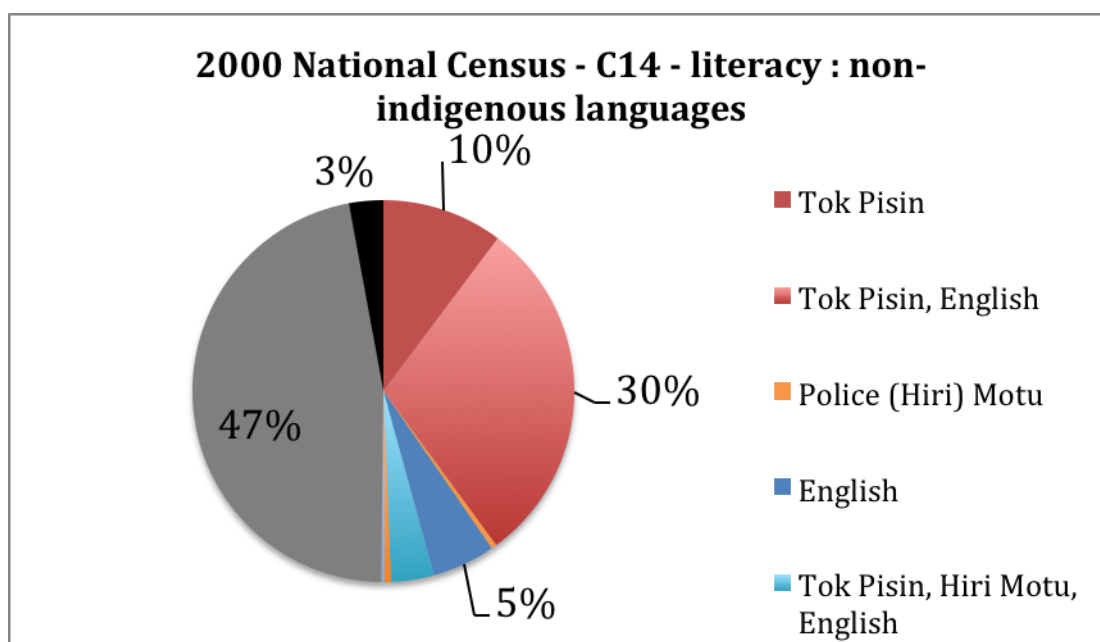- Multiliterate with English — 59%
- Multliterate without English — 16%

Figure 22. Superstrate Effects - Tok Pisin Literacy with English

Over time, census data reveals that both Tok Pisin and English skills have grown significantly. This change in the population of Tok Pisin/English multi-linguals is very well illustrated by a comparison of non-indigenous language literacy from 1966 (Figure 23, see also Figure 13, above) and those same figures from the 2000 National Census (Figure 24):

1966 Census - literacy : non-indigenous languages

- Tok Pisin — 6%
- Tok Pisin, English — 6%
- Police (Hiri) Motu
- English — 3%
- Tok Pisin, Hiri Motu, English
- (82%)

Figure 23 (also Figure 16). 1966 Literacy - non-indigenous languages

Figure 24. 2000 Literacy - non-indigenous languages

Viewed alongside each other, these pie charts perfectly illustrate not only the significant achievement of national literacy education programs in PNG over the 2nd half of the 20th Century, but also the great expansion in a single class of multi-language skills: of people who can use both English and Tok Pisin.

The 2000 National Census data clearly indicates multi-literacy for the majority (c. 80%) of the literate population. In a discussion of creole biliteracy around the world, Siegel (2010) rightly poses a question that we may relate to our discussion (s3.2.2) about the use of literacy skills in all the languages people indicate on the census forms:

> In other cases where there is clearly widespread creole literacy, as in Haiti and Papua New Guinea, it is not known whether there is actual biliteracy – for example, whether readers of the *Wantok Niuspepa* in Tok Pisin actually read materials in English as well.
>
> (Siegel 2010: 391)

However, with the census data reporting high rates of multi-literacy (c. 35% of the entire reporting population) across Papua New Guinea, an appreciation of multi-language skills, and of the Tok Pisin / English interaction in particular, must be kept central to our understanding of language in the nation.

Mühlhäusler characterized Urban Pidgin as resulting from bilingualism with English:

> The principal force shaping Urban Pidgin is the renewed contact with English resulting from English education and widespread English-Tok Pisin bilingualism.
>
> (Mühlhäusler 1985b: 248)

The data from the 2000 National Census reveals this kind of contact in the 40% of the Rural population under the age of 30 reporting literacy in English. This indicates that any influence of English on Tok Pisin previously thought of as 'Urban' may also be a significant factor in Tok Pisin on the national level. This in turn supports Smith's (2003) assertion that the Rural/Urban distinction is not as significant as bilingualism with English.

5   Conclusions

As Laycock discussed in the mid-1980s, and as we have discussed with reference to more recent censuses (s2, above), the fundamental differences between the questions asked, methods of data collection, and reliability of responses over the 44-year period covered by the reported censuses make strict statistical comparisons impossible. None-the-less this paper has outlined some of the ways in which we may use census data to understand the changing language situation.

The value of using educational attainment of literacy skills (Grade 3+) as an indicator of national literacy (s3.2) has been discussed, as have different explanations for the 'literacy gap' between formal education and self-reported figures (s3.2.1). This discussion revealed the important value of an appreciation of both informal education and the role of self-reporting in patterning census data.

The modern-day National Census sets out to collect information only on the languages a person may "read and write with understanding" (Question 14). However people completing the census may actually answer in terms of the languages they know and understand, without being specific about those they read and write (s3.2.3). As suggested here, this fact does not entirely undermine the value of Census data for linguists, though it does require us to take care when giving it our consideration.

The Census data from 1966 recorded both spoken languages and literacy skills in those languages, and we may wish to use that as the basis for an interpretation of modern 'literacy' figures (s3.3). While English language and literacy levels are likely to be similar (due to the educational context for English language learning), Tok Pisin may be spoken by **three times as many people** as report being able to read and write the language. Such a generalization from the 1966 census would put Tok Pisin verbal ability at over 100% of the 2000 National Census population. It certainly would appear that the true figure is considerably higher than Laycock's (1985) suggestion that maximum extension of Tok Pisin may have been reached by 1982. Without reference to National Census data, the Summer Institute of Linguistics reports that 4 million people speak Tok Pisin in Papua New Guinea (Lewis 2009). Written less than a decade after the 2000 National Census this figure is over 100% of the reporting (age over 10+) population for the census language question (Question 14), and indicates that Laycock's proposal of maximum extension has clearly been surpassed in the following generation.

With the great linguistic and cultural wealth of Papua New Guinea largely obscured by the simplicity of the current question on language on the National Census, we might hope that future censuses address a question about which languages people speak and understand alongside the current question on literacy. Including two questions, a question on language competence/proficiency as well as a question on literacy, on the census would provide much richer data as well as removing any confusion between census 'literacy' and general language skills.

Moving beyond these concerns, we can use the Census data in order to gain insight into the spread of language skills in different languages in urban and rural areas over time (s4.1), on the spread of the national languages (s4.2), on the overlap between vernaculars and the national languages and the growing role of English as a language of multi-lingualism (s4.3), and on the status of Tok Pisin as a categorical pidgin (a language which is additional to other languages) (s.4.4.1) and also into the levels of bilingualism with English in the national community (s4.4.2).

In terms of my PhD project, census data established a baseline for understanding of Tok Pisin / English bilingualism in Papua New Guinea. Though comparisons with earlier censuses are problematic, and we may prefer a question on spoken language skills to be re-included on the census to allow a better understanding of language in the country today, it is clear that around 40% of people responding to the 2000 National Census felt able to indicate they had skills in both Tok Pisin and English.

The 2000 National Census data does not allow us to determine which language is primary for respondents. In addition, there are problems, as discussed above, with generalizing from patterns of literacy to patterns of actual language skills or language use. This paper has shown, however, that general trends can be observed by linguists using census data, and the reports of subsequent National Censuses should be carefully studied in the years to come.

References

Crowley, Terry. 1994. Linguistic Demography: Interpreting the 1989 Census Results in Vanuatu. *Journal of Multilingual and Multicultural Development* 15.1. 1-16.

Jenkins, Rebecca Sue. 2000. Language Contact and Composite Structures in New Ireland, Papua New Guinea. PhD. Dissertation. University of South Carolina.

Laycock, Don. 1985. Tok Pisin and the Census. *The Handbook of Tok Pisin*. (eds.) Wurm, S A and Peter Mühlhäusler. 223-31.

Lewis, M. Paul (ed.) 2009. *Ethnologue: Languages of the World*, Sixteenth edition. Dallas, Tex.: SIL International. Online version, Tok Pisin entry: http://www.ethnologue.com/show_language.asp?code=tpi, accessed March 2012.

Mühlhäusler, Peter. 1985. Internal Development of Tok Pisin. *The Handbook of Tok Pisin*. (eds.) Wurm, S A and Peter Mühlhäusler. 75-166.

Mühlhäusler, Peter. 1985. Variation in Tok Pisin. *The Handbook of Tok Pisin*. (eds.) Wurm, S A and Peter Mühlhäusler. 233-273.

National Statistical Office, Census. 1994. *Report on the 1990 National Population and Housing Census in Papua New Guinea*. Port Moresby: National Statistical Office.

National Statistical Office, Census. 2002a. *2000 Census Basic Tables - National Level*. Port Moresby: National Statistical Office.

National Statistical Office, Census. 2002b. *Papua New Guinea 2000 Census National Report*. Port Moresby: National Statistical Office.

Noel, John. 1975. Legitimacy of Pidgin in the Development of Papua New Guinea toward Nationhood. *Tok Pisin I Go We?* (ed.) McElhanon, K. A. Port Moresby: Linguistic Society of Papua New Guinea. 76-84.

Romaine, Suzanne. 1992. *Language, Education, and Development : Urban and Rural Tok Pisin in Papua New Guinea*. Oxford: Clarendon Press.

Sankoff, David. 2008. How to Predict the Evolution of a Bilingual Community. *Social Lives in Language*. (eds.) Meyerhoff, Miriam and Naomi Nagy. Amsterdam: John Benjamins. 179-94.

Sankoff, Gillian. 1980. Multilingualism in Papua New Guinea. *Social Life of Language*. (ed.) Sankoff, Gillian. 95-132.

Siegel, Jeff. 1998. Literacy in Melanesian and Australian Pidgins and Creoles. *English World-Wide* 19.1. 104-33.

Siegel, Jeff. 2010. Bilingual literacy in creole contexts. *Journal of Multilingual and Multicultural Development* 31.4. 383-402.

Smith, Geoff. 2002. *Growing up in Tok Pisin*. London: Battlebridge.

Wurm, S A and Peter Mühlhäusler. 1985. *The Handbook of Tok Pisin*. Pacific Linguistics C-70. Canberra: Research School of Pacific Studies, Australian National University.