bls Berkeley Linguistics Society

Quantitative Analysis of Variation: i in Tok Pisin
Author(s): Christopher Smeall
Proceedings of the First Annual Meeting of the Berkeley Linguistics Society (1975), pp. 403-409

Please see "How to cite" in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via http://linguistics.berkeley.edu/bls/.

The Annual Proceedings of the Berkeley Linguistics Society is published online via eLanguage, the Linguistic Society of America's digital publishing platform.

A Quantitative Analysis of Variation: i in Tok Pisin

Christopher Smeall

University of California at Berkeley

The principal goal of this paper is to describe the variation in the distribution of the verbal particle i in Tok Pisin and to present grammatical evidence that, in the course of its assimilation from substrate Austronesian languages, i has changed from a subject agreement particle to something more akin to a complementizer, and that part of the surface variation in its occurence is due to phonological conditioning of a non-deterministic nature.[1]

The corpus for this study is primarily based on tape-recorded data gathered by Gillian Sankoff and Suzanne Laberge in New Guinea during the summer of 1971, and consists of roughly 1,350 sentence tokens drawn from the speech of four fluent speakers of Tok Pisin.

Tok Pisin, or Neo-Melanesian, is a pidgin language spoken in the Trust Territory of New Guinea, parts of Papua New Guinea, the British Solomon Islands, and the New Hebrides. According to Wolfers (1968), its speakers number about 530,000, and the roughly 10,000 native speakers among these testify to its current creolization.

The principal lexical source languages of Tok Pisin are English, which accounts for the bulk of the lexicon, and Kuanua, an Austronesian language spoken in the Gazelle peninsula of New Britain.

The verbal particle in question is exemplified in (1)-(3):

(1) Em i no giaman. "He wasn't lying."
(2) Olotaim mitupela i stap long kunai. "All the time we've been living in the grasses."
(3) Man i klostu i laik i lapun. "The man was getting old."

As these examples illustrate, i occurs in pre-verbal position. (1) exemplifies the negative no, one of the few elements which can intervene between i and the verb, and (3) shows that i marks embedded verbs as well, as both klostu and laik are auxiliaries. Turning to (4), we can see that i has an analogue in Kuanua, the Tok Pisin substrate language mentioned above.[2] It is quite clear that i is not, as was previously thought by some early grammarians, derived from the third singular subject pronoun in English.

(4) KUANUA: a tuntana i takpa ra vudu. "The man takes the banana."

The example from Motu in (5), an SOV language with the analogue e, exemplifies the particle's transcendence of the SVO/SOV factionalism that exists among New Guinea Austronesian languages. Note that it still occurs pre-verbally, however.

(5) MOTU: mero ese hudi e vaia "The boy takes the banana."

Some other Austronesian languages with i-analogues are illustrated in (6) and (7).

(6) WEDAU: wei numa-na i gaegae "This house is large."
(7) KIRIWINAN: bua bogwa i yagi "The betel has now ripened."

The distribution of the i-analogues in those Austronesian languages in which it occurs seems to be controlled by the subject. Although these languages vary in morphological complexity, and some of them incorporate tense markers into their i-analogues, subject agreement, with different forms of the particle occuring depending on the person and number of the subject, is the norm. An example of a conjugation from Petats is given in (8) for the verb la 'go'.

(8) PETATS:
elia a la nauk "I am going."
elo o la nom "You are going."
eiau e la no "He/she is going."
eri e la ro "We (incl.) are going."
elam e la nam "We (excl.) are going."
emi e la nomi "You (pl.) are going."
eru e la er "They are going."

In Petats, as (8) shows, only the first and second singular have special forms, with e marking the remainder of the pronouns. In general, the e or i form is associated with majority representation and usually includes third person, but the subgrouping of forms among remaining pronouns varies from language to language.

What distinguishes these languages from Tok Pisin, first of all, is the categorical nature of the subject agreement constraints. In Tok Pisin, like in other pidgin languages, there is a great deal of variation, and the distribution of i is no exception. Consequently, the constraints are more difficult to determine. Table I lists the percentage occurence of i in simple sentences in my corpus, correlated with the various pronominal forms, in rank order.

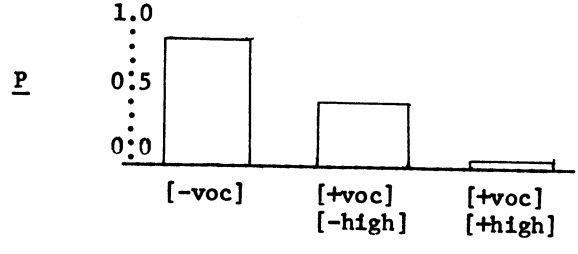
TABLE I Percentage occurence of i before the verb as a function of pronominal subject in simple sentences.

Person/number	Percentage	Tok Pisin form
3rd/sing.	83	em
3rd/plur.	80	ol
3rd/plur.	60	ologeta
1st/plur.(excl.)	55	mipela
2nd/plur.	48	yupela
2nd/sing.	15	yu
1st/plur.(incl.)	6	yumi
1st/sing.	5	mi

This table indicates, among other things, that for no pronoun is the presence of i correlated in a truly categorical way with its person and number. Moreover, if one examines more closely the distribution of these percentages, the fact emerges that they form a descending hierarchy with respect to the final segment of the pronoun. Those forms with which i occurs most frequently end in a consonant, those of intermediate frequency end with a low vowel, and those of lowest frequency end with a high vowel. In addition, the figure for ologeta, which is also a third plural pronoun, but ends in a low vowel rather than a consonant, shows a considerably lower frequency of occurence with i than ol does.

To further test the hypothesis that the variation in occurence of i was at least partially determined by phonological constraints, all of the data in the corpus were lumped together, that is to say both sentences with pronominal subjects and sentences with NP subjects, and the occurence/non-occurence of i was correlated with the nature of the segment immediately preceding the space where i occured or should have occured, regardless of the nature of the word in which the segment was located. The Sankoff-Cedergren program was used to determine multiplicative probabilities for the factors preceding high vowel, preceding low vowel, and preceding consonant, among others. In coding the data, reduced occurences were counted as full occurences, and dubious cases were thrown out. The results, shown in Table II, are similar enough to the percentages in Table I to suggest that phonological conditioning may have something to do with the pattern of variation associated with person and number of subject. Under this hypothesis, i would be deleted by a phonological rule with roughly the form given in (9):

TABLE II Probability of non-application of i-deletion as a function of preceding environment.



(9) $$i \longrightarrow \emptyset \; / \; \left\langle \begin{array}{l} a([-voc]) \\ b\left(\begin{bmatrix} +voc \\ -high \end{bmatrix}\right) \\ c\left(\begin{bmatrix} +voc \\ +high \end{bmatrix}\right) \end{array} \right\rangle \; \_\_\_\_$$

If i is not functioning as a subject agreement particle, then what is its function in Tok Pisin? I would suggest that i has become a kind of complementizer or sentence marker, that it is still sensitive to the presence, but not the nature, of subjects, and marks S-nodes in constituent structure.

Various patterns of variability in i-occurence with respect to selected syntactic processes of Tok Pisin suggest that a cyclic rule of i-insertion operates to place the particle pre-verbally, and distributional facts suggest that this rule, like the phonological deletion rule, must be equipped with variable constraints.

What little data there is concerning syntactic processes in Austronesian languages with i-analogues does not appear to display any of the characteristics to be discussed below, but a detailed comparison is unfortunately impeded by the lack of reliable grammatical descriptions of these languages.[4]

First, i-insertion is variably verb class governed. The rule is controlled by both the matrix and embedded verbs, bearing in mind that since i can complementize root sentences, the term 'embedded verb' has been extended to include the verb in simple clauses, embedded in a deleted performative if you will.

The verb class restrictions are of three types. Embedded verbs of saying and thinking, and perhaps all epistemic non-factives, exemplified in (10) and (11), highly favor non-application of the rule:

(10) em tok) em i tok "He talks"
(11) man i larim ol tingting "The man let them think"

Motion verbs like go, kam, and stap, which are used aspectually in Tok Pisin, have the opposite effect: they strongly favor insertion, and are exemplified in (12) and (13)

(12) man i mekim dok i go outsait "The man made the dog go outside."
(13) man i mekim yu i kam kisim "The man made you come get

The third type of constraint is proposed tentatively, due to under

exemplification. According to Wurm(1971), the auxiliaries and modals can be divided into two groups, depending on whether i insertion occurs in the cycle below them or not (this is my interpretation of Wurm's data). In other words, this is a question of matrix, rather than embedded, verb conditioning. (14) and (15) exemplify inhibitory aux's, (16) and (17) excitatory ones.

(14) mipela i mas kaikai "We must eat."
(15) em i traiim lukluk long mipela "He tried to look at us."
(16) man i klostu i laik i lapun "The man is getting old."
(17) em i laik i brukim paiawut "He wants to chop firewood."

Woolford(1975) distinguishes another verb-class, the set of verbs which potentially take the transitive suffix -im, and presents figures which indicate that this class also has an inhibitory effect. It may be that this constraint can accomodate the variation found in the auxiliaries. In Table III there is some data summarizing the effect of the above classes, with other factors held as constant as possible.

TABLE III Occurence of i as a function of verb class, third person subjects only.

Verb Class	Percentage
Aspectuals	90
Aux's	i.d.
Saying or thinking	10

A second feature of the variation involves the presence or absence of subjects. According to Wurm, when an embedded NP has adjuncts it may optionally be postposed, and when it has an embedded relative, postposition is obligatory. Furthermore, the operation of this process of 'Heavy NP Postposition', again according to Wurm, has a categorical blocking effect on the insertion of i. So one finds the normal ordering as in (18), the optional variants (19) and (20), and the obligatory postposing of (21):

(18) larim dok i kam insait "Let the the dog come inside."
(19) larim tispels tupela man i kaikai "Let these two men eat."
(20) larim kaikai tispela tupela man "Let these two men eat."
(21) larim kaikai man i sindaun i stap lohap "Let the man who is sitting over there eat."

Note that i is not found when postposing has occured. Although there are no examples of this type in my corpus, it may be that it is the subject NP's presence or absence which is crucial here, and that i-insertion is inhibited when subject NP's are ripped out or deleted. Some evidence that does bear on this question

is the patterning of i with respect to conjunction reduction. When the second of two identical NP's in subject position is deleted in conjoined sentences, there is a strong tendency for i to occur less often in the second conjunct, approximately 30% of the time in my data, compared with simple sentences restricted to third person pronominal subjects.

A summary of these constraints is given in the formulation of i-insertion in (22). Although I have presented some order of magnitude values for the constraints in the body of this paper, because of the relatively small size of the corpus, speakerwise and tokenwise, because of the fact that all variables must be analyzed simultaneously for correct interpretation, and because of the fact that the 'correct' choice of independent variables is largely a matter of judgement, exact specification of coefficients must await a more comprehensive analysis. In particular, when the correct set of determining variables is isolated, the patterning of i may turn out to be more categorical than here indicated.

(22) Cyclic i-insertion

$$\emptyset \longrightarrow \langle i \rangle \;/\; \langle aS_1 \rangle [\, \langle bV_1 \rangle_{S_2} [\, \langle cNP \rangle \_\_ + \langle dV_2 \rangle \,]_{S_2} \,] \langle aS_1 \rangle$$

where a,b,c,d are coefficients representing contributions towards probability of rule application, and a depends on presence or absence of S_1, b depends on verb class of V_1, c depends on presence or absence of NP, and d depends on verb class of V_2.

To summarize then, I have tried to show that while i-analogues in many Austronesian languages are deterministically constrained subject agreement particles, the patterns of variation in the data tentatively suggest a more attenuated role for Tok Pisin i as a stochastically constrained sentence marker dependent on such things as verb class, presence or absence of subject NP, occurence in subordinate or superordinate position, and phonological environment.

1. I would like to thank Gillian Sankoff and Willian Labov for much valuable assistance, and the participants in Gillian Sankoff's Tok Pisin seminar at the 1973 Linguistics Institute for transcriptions. Responsibility for the above ideas rests with me.

2. Examples (4)-(8) are from A. Capell.
3. The assumption of non-interaction of the independent variables is made, but possibly not warranted.
4. See for example Codrington(1885).

REFERENCES

Capell, A.
1969 A survey of New Guinea languages, Sydney, University Press.
1970 The Austronesian languages of Australian New Guinea, in Linguistics in Oceania, ed. by T.A. Sebeok et al., Current Trends in Linguistics 8, The Hague, Mouton.

Chowning, A.
1968 Languages of New Britain, PL, A21.

Codrington, R.H.
1885 The Melanesian Languages

Laycock, D.
1970 Materials in New Guinea Pidgin (Coastal and Lowlands), Pacific Linguistics, Series D, No. 5.

Mihalic, Rev. F.
1971 The Jacaranda Dictionary and Grammar of Melanesian Pidgin, Jacaranda Press, Port Moresby.

Sankoff, G.
1973 Above and beyond phonology in variable rules, in New Ways of Analysing Variation in English, ed. by C-J. N. Bailey and R. W. Shuy, Washington, D.C., Georgetown University Press.

Sankoff, G and S. Laberge
1973 On the acquisition of native speakers by a language, Kivung 6:32-47.

Wolfers, E.
1971 A report on Neo-Melanesian, in Pidginization and Creolization of Languages, ed. by D. Hymes,Cambridge, University Press.

Woolford, E.
1975 Variation and change in the i predicate marker of New Guinea Tok Pisin, paper prepared for the 1975 International Conference on Pidgins and Creoles.

Wurm, S.A.
1971 New Guinea Highlands Pidgin; course materials, Pacific Linguistics, Series D, No. 3.