Information structure and discourse markers in Tok Pisin: differences in genres

Masahiko NOSE (Shiga University)
mnosema@gmail.com

Abstract:
This study describes the information structure of Tok Pisin, an English-based creole language, spoken in Papua New Guinea. It has a rigid word order SVO, and does not have a grammatical voice. This study illustrates several cases of grammatical behavior related to the information structure and clarifies grammatical means of Tok Pisin to deal with information flow. This study utilizes the written texts of the New Testament and translated folk tales, as well as spoken data from radio news and chats in the field. Topic and comment in Tok Pisin are affected by emphasizing, topicalization (or fronting), and by using several discourse markers. Finally, I claim that Tok Pisin has an extremely limited means of dealing with information structure in its grammar, although we observed slight differences between the spoken and written data.

Keywords: information structure, topic, focus, discourse marker, Tok Pisin

1. Introduction
There are approximately 1000 native languages in the Melanesia area. They are classified mainly as Austronesian and New-Guinea (or Papuan) languages. Moreover, some lingua franca languages are spoken for communicating with each other (see Figure 1). The three main languages (Tok Pisin, Solomon Pijin, and Bislama) are English-based creole languages and their lexifier is English.

This study describes the information structure in Tok Pisin and will pay special attention to both spoken and written data. Crowley (2004) described the grammar of Bislama and wrote about some aspects of its information flow. For this study, I conducted fieldwork in the Amele-speaking area in Madang Province (Nose 2014) and gathered Tok Pisin data (Amele in Figure 1).

Section 2 presents basic information about Tok Pisin, and explores several previous studies on creole languages. Section 3 presents examples of spoken and written data of Tok Pisin; in particular, we observe changing word orders and usage of discourse markers. Section 4 presents a discussion on managing information flow and the differences in genres, and Section 5 is the conclusion of the study.
2. The grammar of Tok Pisin and preliminary studies on information structure

This section introduces basic information about Tok Pisin and, subsequently, indicates how previous studies have described the information structure of creole languages. This generally concerns creole languages (McWhorter 2011, Veenstra & Besten 1994), Melanesian pidgins (Crowley 2004, Dutton 1985, Mihalic 1971), and information structure (Andersen 1983, Foley 2007).

2.1. Introduction of Tok Pisin grammar

Tok Pisin is an English-based creole language and one of the official languages of Papua New Guinea. Grammatically, Tok Pisin has an isolating tendency and little inflection on nouns and verbs. It has a rigid SVO order and an adjective-noun and demonstrative-noun order.

(1) Mi bin rid-im dispela nicepela buk long haus bilong mi.
1sg past read this nice book prep house prep 1sg
“I read this nice book in my house”

In (1), the subject mi appears at the beginning of the sentence and, in contrast, the object buk appears later, while the verb ridim occupies the middle position between the subject and the object. The form bin indicates the past tense marker, and the -im of the verb ridim is a transitive marker. There are two prepositions in Tok Pisin; one is the multifunctional preposition long, and the other is the possessive preposition bilong (Mihalic 1971: 38). Possessive formation is expressed in “object bilong owner,” as haus bilong mi “my house” in (1). Tok Pisin has only one active voice and, therefore, the sentence (1) cannot be passivized.

2.2. Discourse markers of Tok Pisin

Mihalic (1971:40-41) described the usage of conjunctions in Tok Pisin as fulfilling discourse markers. The most frequent marker is na, which means “and” and “nor.” Other significant markers are olsem “in order to,” “and,” “then” and “as,” and tasol “but,” “however.” Moreover, we found several kinds of discourse markers in (2) in the process of our translation (Nose and Tamo (2015 forthcoming)).

(2) Representative discourse markers in Tok Pisin:
    na “then”; tasol “but”; mi tok olsem “in fact, that is to say”; bikos “because, that’s why”;
    long mi “for me, as for me”; kain olsem “for example”; long sampla kain tok “in other words”;
    nambawan tain “firstly”; bihain orait “next, secondly”; bihain pasten “finally”;
    kain samting olsem “how could I say, something like”

In (2), the forms long, olsem, and bihain are preferable and they function as connectives, in terms of text organization and speaker-hearer interaction, and they influence the information structure, as well (Schiffrin 1988, Heine 2013). In section 3 and 4 we discuss how several discourse markers influence the information structure.
2.2 Some previous studies

This section describes the characteristics of creole languages in terms of several previous studies (Crowley 2004, McWhorter 2011), and the definition and properties of information structure, based on Kiss (2002), Veenstra & Besten (1994) and Sankoff (1983). As McWhorter (2011) argued, creole languages tend to have simple grammar. According to Dutton (1985) and Mihalic (1971), Tok Pisin does not have the complicated characteristics of grammar, and has no means to specify topic and focus. However, Sankoff (1993) presented another perspective.

2.2.1 Creole has simple grammar

McWhorter (2011) indicated that creole languages are generally simple in phonology, morphology, and syntax, and he explained the reasons for creole languages being at the infancy stage and why they have not experienced much grammaticalization. Actually, the grammar of Tok Pisin is simpler than that of the other native New-Guinea and Austronesian languages, and therefore, Tok Pisin is easier to learn. Thus, it spread as a lingua franca throughout Papua New Guinea. For example, verbs in Tok Pisin have few inflections and the words are rather isolated; moreover, word orders are fixed in SVO. Crowley (2004) described the grammar of Bislama, but he did not point out whether there is a topic/focus system in Bislama, although he enumerated several discourse markers relevant to the information structure. Moreover, Sankoff (1993) discussed topic and focus in Tok Pisin, and Meyerhoff (2011) described some discourse markers in Bislama.

2.2.2 Information structure from a typological perspective

Word order and information structure are related to each other, and they have been a popular topic for studies since the Prague school and classical studies in word order typology (cf. Andersen 1983). This study reviews the information structure and word order of Tok Pisin using Hungarian sentence examples (Kiss 2002).

Hungarian is a language with free word order, but its orders are influenced by topic and focus. In (3), the transitive sentence in Hungarian can be changed in (3a-d), in terms of information structure.

(3) Hungarian:

a. Péter el-olvassa a köny-et a bolt-ban.
   Peter prev-read.3sg.pres the book-acc the store-in
   “Peter read through the book in the store.”

b. A könyvet Péter elolvassa a boltban.
   “As for the book, Peter reads it through in the store.”

c. Péter a könyv-et olvassa el a boltban.
   “It is the book that Peter reads through in the store.”

d. A boltban a könyvet olvassa el Péter.
   “In the store, it is the book that Peter reads it through.”
In (3), (3a) has a neutral order, SVO order, and the first element Péter is the subject and also occupies the topic position. In (3b), the sentence-initial topic becomes the direct object a könyv, and in (3c), the topic is Péter, but a könyv occupies the focus at the preverbal position. Finally, in (3d), a boltban occupies the topic and a könyv occupies the focus, respectively. Generally, topic comes at the beginning of the sentence, and the focus position is near the verb, as shown in Hungarian (Andersen 1983:70-71, 74-75). However, unlike Hungarian, creole languages do not have a developed information structure in their grammar.

Veenstra & Besten (1994:303) suggested the concept of fronting instead of the topic and focus system similar to Hungarian. Veenstra & Besten claimed that fronting is functioning as topicalization, left-dislocation, and focusing. Fronting indicates moving an element to the initial position in the sentence. Fronting occurs in the following three conditions: in declarative contexts, interrogative contexts, and relative clauses. Topicalization is applied in declarative contexts and this movement also includes cleft, and focusing. Wh-movement is applied in interrogative contexts, and the third one in relative clauses.

(4) Tok Pisin:

a. **Long ples** mi lainim tok ples Amele.
   prep village 1sg learn native tongue Amele
   “At village, I will learn a native language Amele.”

b. **Husait** em i nicepela strait?
   who 3sg cop nice very
   “Who is the fairest of us all?” (FT)

c. Em kamap **gutpela hap** we i gat haus i stap.
   3sg come up good place where cop get house cop be located
   She came to a clearing where there was a small cottage. (FT)

In (4a), the word long ples takes the initial position in the sentence. This fronting is topicalization. Next (4b) is an example of wh-movement, and the interrogative pronoun husait (who) occupies the initial position. The relative clause in (4c) is also a fronting movement. However, focusing in Tok Pisin has another view (Sankoff 1993). Therefore, we observe another creole, Saramaccan (Veenstra & Besten 1994:306), as shown in (5).

(5) Saramaccan:

Di buku **we** mi bi lesi.
the book focus 1sg past read
“I read THE book.” we: focus marker

There is a focus marker, we, in Saramaccan. Sankoff (1993) claimed that there are postponed focus particles in Tok Pisin; yet, tru, moa, tasol, and wampela. In the following paragraphs, this study provides spoken and written data of Tok Pisin, and we examine how information structure works and
we consider whether there is focus marker in Tok Pisin.

3. Information structure in different genres

This section investigates the information structure in Tok Pisin by observing several different kinds of texts. Mainly, we examine fronting and other means of topicalization/focusing; moreover, we observe the usages of discourse markers. This study classifies the texts of Table 1 according to genres; namely, spoken/written and city/ rural texts.

<table>
<thead>
<tr>
<th>Spoken discourses</th>
<th>City Tok Pisin</th>
<th>Rural Tok Pisin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written texts</td>
<td>YM</td>
<td>FTC, Bel</td>
</tr>
<tr>
<td></td>
<td>RA, NT</td>
<td>FT</td>
</tr>
</tbody>
</table>

Table 1. Spoken and written Tok Pisin data in this study

3.1. Spoken data

The spoken data are YM, FTC, and Bel, and there are several differences between city and rural genres. Lexically speaking, YM has frequent usages of English words, and in contrast, FTC and Bel maintain Tok Pisin vocabulary. Grammatically, not all discourses keep a strict word order, and we can observe many discourse markers that are relevant to the information structure.

First, in (6), the fronting element puttim tan long is emphasized by stress. Therefore, adding stress phonologically is a means of focusing in spoken Tok Pisin.

(6) FTC/ rural

\[
\text{puttim tan long, maus bilong em... em em yet... em pinis a? em i go ... long brik haus put tongue prep mouth prep 3sg 3sg 3sg yet 3sg perf intj 3sg cop go prep block house}\]

“put tongue on his mouth, (he licked his lips, as he went to the brick house)”

(7) Bel/ rural

\[
\text{wankain olsem}\]

same like (DM)

“It is the same like this.”

Another means of focusing is to say only the necessary word, as in (7). In (7), the fragment words wankain olsem are foregrounded in the discourse and other information is not specified, but they are supplemented by the context. In (8), it is an example of fronting, but the word “Snow White” is repeated in the following sentence and the element is emphasized by this repetition. Example (9) is peculiar in that the 3rd person singular pronoun he (instead of em) is used. The speaker (newscaster on the radio) accidentally used the English word he. Needless to say, the specific name (HK) of the governor is operated by fronting.

(8) FTC/ rural

\[
\text{Snow White .... Snow white em pilim tait, Snow White Snow White 3sg feel tired}\]

“Snow White feels tired”

(9) YM/ city
**Gabanaa Haveila Kaabo, he** givim bikpela tenkyu bilong em i go long travel foundation
Governor 3sg give big thanks prep 3sg cop go prep travel foundation
“Governor HK gives big thanks to the travel foundation.”

(10) FTC/ rural
i no ken writim pinis mi askim yu, yu writim pinis oo
cop no can write perf 1sg ask 2sg 2sg write perf intj
“You cannot write it up, and I am asking you, have you already written it?”

In (10), the sentence does not have a subject and the verb comes first. It seems to be fragment information, but this type is frequently observed in both spoken and written texts. In this case, focus is considered to be on the negative action or verb phrase.

(11) YM/ city
**na ting blong ol na i tok olsem**
DM think prep 3pl DM cop talk like (DM)
“And their idea, they say like this”

In (11), the discourse marker *na* is used. This *na* means “and” and “then,” and this marker is apt to appear at the beginning of the sentence. The discourse marker *na* can connect the previous and present sentences; moreover, speakers can give additional information through the second *na* in (11). In (11) and (12), the form *olsem* is used, which means “like, that, in this way.” This is also a discourse marker and it connects the discourse or omits background information.

(12) FTC/ rural
em mas meri mi ting wanpla tok **olsem**
3sg must girl 1sg think one talk like (DM)
“She must be a girl, and I think one says like this.”

(13) FTC/ rural
**Nau** mipla go long 42, 41 pinis
now (DM) 1pl go prep page 42, 41 finish
“Now we go to the page 42, the page 41 is finished.”

Finally, in (13), the form *nau* is naturally derived from the English “now”. It is also a discourse marker (as pointed out by Sankoff 1993) and this marker can introduce a new topic at the beginning of the sentence.

### 3.2. Written data

In this section, we examine written texts of Tok Pisin and examples of their behavior are observed through fronting, discourse markers, and other means. Moreover, we try to find differences
in genres. RA and NT are regarded as city Tok Pisin, and FT is a translation work of Nose and Tamo (2015 forthcoming).

(14) RA/ city

Long Madang Provinces blong PNG, ol keis blong measles i wok long go daun.
prep Madang Province prep PNG 3pl case prep measles cop work prep go down
“In Madang Province, PNG, all cases of measles are going down.”

In (14), the locative phrase with preposition long is moved to the initial position by fronting. The fronting of locative phrases is frequently observed in Tok Pisin.

(15) NT/ city

Na em i salim sampela man i go long Betlehem na ol ples klostu long en,
DM 3sg cop send some man cop go prep Bethlehem DM 3pl place near prep it
na ol i kilim i dai olgeta pikinini man i no winim tupela yia yet.
DM 3pl cop kill cop die all child man cop no suppress two year yet
“and he sent forth and put to death all the male children who were in Bethlehem and in all its districts, from two years old and under.”

The discourse marker na is used frequently in written texts as well as in spoken texts. In particular, in (15), the three na markers connect discourses, each meaning “and.”

(16) FT 5/ rural

SnowWhite i no save olsem em haus blong ol sevenpla dwarf.
Snow White cop no know DM 3sg house prep 3pl seven dwarf
“What SnowWhite did not know was that the cottage belonged to seven dwarfs.”

(17) FT 7/ rural

Olsem na em i go long tower na wokim apul long outsait
DM DM 3sg cop go prep tower DM work apple prep outside
look olsem apul tasol em pulap poison.
“So she went to her tower where she made an apple.”

The discourse marker olsem is frequently observed, as shown in (16) and (17). The form olsem has several usages. First, in (16), this marker functions as a complementizer or for quoting (Meyerhoff 2011: 256). In contrast, in (17), olsem and na appear in sequence and their meaning is “so, then,” functioning as a conjunction.

(18) FT 4/ rural

so em i sleep insait long wanpla blong ol dispela bed.
DM 3sg cop sleep inside prep one prep 3pl this bed
“and so she yawned and stretched and lay down to sleep in one of the beds.”

On the other hand, in (18), the English word “so” is used as a conjunction. The meaning of so is almost the same as olsem or na.

(19)NT/ city
Nau Herot i save, ol saveman bilong hap sankamap
now Heod cop know 3pl wise man prep side of sun rise
ol i giamanim em, orait bel bilong em i hat moa.
3pl cop deceive 3sg DM belly prep 3sg cop hard more
“Then, Herod, when he saw that he was deceived by the wise men, was exceedingly angry”

In (19), there are two discourse markers, nau and orait. Nau functions by presenting a new topic or story, and it is used preferably in spoken discourse, like in (13). In contrast, orait means “so” in English. Both nau and orait play a role in expressing the speaker’s (or storyteller’s) confirmation (cf. Schiffrin 1988: 230, “speaker progression”).

4. Discussion
This section summarizes the observations of the previous sections and discusses several usages. The present study examined spoken and written texts and tried to find their differences and to summarize several sentence types related to their information structure. Overall, the observed grammatical means are summarized in (20). There are limited means expressing topic and focus in Tok Pisin; moreover, only a few discourse markers are preferred for usage. Additionally, this section discusses the differences in genres that were observed.

(20) Grammatical options regarding information structure in Tok Pisin
a. Fronting: topic and focus: repeating, cleft
b. Emphasizing (focusing): necessary word only, emphatic words (wanpla, tru, tasol: Sankoff (1993)’s focus particles)
c. Using discourse markers: connecting discourse, introducing, omission
c-1: conjunction: na, tasol, so, yet, olsem (like this, thus)
c-2: demonstrative: ya, olsem (introducing quotation),
c-3: confirmation: nau, orait; confirmation, new topic
d. others (verb-initial, yu tok): verb-focus, turn-taking

Mainly, Tok Pisin utilizes emphasizing in spoken discourse, and prefers fronting in both genres. Although we did not find a proper example in this study, cleft construction (see footnote 8) is also possible as a fronting movement. Discourse markers have a function of introducing topics, connecting sentences, and implying speaker’s confirmations, but these kinds are limited and only na and olsem are commonly observed. Other usages are considered, but this study does not discuss them.

8
Next, we try to clarify the differences in genres. Spoken discourse can depend on phonological clues (stress and pause); moreover, the speaker can choose to relay the important information only. Instead, written data prefers using discourse markers; in particular, the discourse markers *na* and *olsem* are frequently used for coherence. Next, we consider the differences between city and rural areas, but there is no significant difference in the information structure between them. Fronting is observed in both genres, and some English words, *so, he*, etc. are preferred in the spoken/city texts.

Moreover, we discuss that discourse markers can become topic/focus markers. For example, McWhorter underscored that the discourse marker *nɔˈɔ*, “then” in Saramaccan, has been grammaticalized to a new information marker, as shown in (21).

(21) Saramaccan (McWhorter 2011:127-128)

A bu’nú.  **nɔˈɔ** mi ʼ o ʼ ta’ ha’i ka i.

it good DM I future imperfective listen you “Good. So I’ll be listening for you (waiting for your answer)”

Sankoff (1993) indicated that Tok Pisin has the focus marker *yet*, originally meaning “yet, still.” Sankoff claimed that the form *yet* in (22) has been grammaticalized from intensifier to focus marker. However, this study did not find such a usage, and the usage of *yet* is extremely limited in texts. This study considers that *yet* is not a fully grammaticalized marker in Tok Pisin, and that it only has fronting or emphasizing functions for focusing (the usage *yet* is included in (20b)).

(22) Tok Pisin: Sankoff (1993: 131)

Tok “Orait yu **yet** kilim pikinini bilong mi.”

say alright 2sg focus kill child prep 1sg

“(She) said, “Alright, **you’re the one** who killed my child””

We subsequently discuss the usages of discourse markers. The discourse marker *na* is frequently observed in connecting sentences. In (23), these sentences are taken from FT and the parallel texts in Tok Pisin, Amele, and English. Foley (2000:387) claimed that the New Guinea languages have common discourse characteristics. First, given and presupposed information is normally omitted, and second, only one piece of new information is introduced per clause. By contrasting the parallel texts in (23), these characteristics are applicable to Tok Pisin and also visible through language contact. That is, the discourse markers *na, olsem* are used to form consistent discourse structures, as observed in New Guinea languages.

(23) Tok Pisin/ Amele: FT (cf. Nose 2014)

**Na ol i kilim em Snow White.**

DM they cop call 3sg Snow White

**Odocob uqa ayan snow-white boin.**

And then (DM) 3s name-poss snow white call-3s.past

“and then she called snow white.”
Finally, this paper considers answers to why Tok Pisin has a limited number of managing information structures, and explains the decisive differences between spoken and written texts. First, Tok Pisin is a creole language and it still has the characteristics of a simple grammar in its construction, which cannot yet fully grammaticalize topic/focus markers. As a result, it depends on fronting, emphasizing, and several discourse markers, as shown in (20). In spoken discourse, speakers utilize stress, pause, emphasizing, and repeating, while, in contrast, discourse markers are preferred in written texts. This study claims that discourse markers such as nau, oraip, na, and olsem function to introduce new information or focus to sentences (cf. Foley 2000: 386-387, Nose 2014).

5. Conclusion

We summarize the characteristics of topic and focus in Tok Pisin that this study has found.

First, fronting is the most frequent method of indicating topic/focus. Tok Pisin has fronting movement through the indication of topicalization and partly through focusing; but focusing is realized through other options, namely adding stress, emphasizing, and cleft constructions. In particular, spoken discourses prefer adding stress, using emphasis, and repeating, while, in contrast, several discourse markers are frequently observed in both genres, and they are effectively used to organize foreground and background information in written texts. However, these discourse markers have not yet been grammaticalized into topic or focus markers.

Notes:
1. I would like to thank Neret Tamo, Nelau Lagia and the villagers in Sein, Madang Province, Papua New Guinea for their data and kindness. I claim sole responsibility for any errors.
2. The lingua franca in Melanesia are Indonesian in the Western part of New Guinea Island, Hiri Motu in the Western Province and around Port Moresby in Papua New Guinea, Tok Pisin in the other area of Papua New Guinea, Solomon Pijin in Solomon Island, and Bislama in Vanuatu. Tok Pisin, Solomon Pijin and Bislama are called Melanesian Pidgin.
3. Abbreviations: acc, accusative; cop, copula; DM, discourse marker; future, future tense; intj, interjection; sg, singular; past, past tense; perf, perfective; pl, plural; prep, preposition; prev, verbal prefix, pres, present tense; 1, 2, 3, first person, second person, third person respectively.
4. Thus, word order position and information structure in Hungarian are summarized in (i).

   (i) (Topic) X (Focus)-VERB X

   Topic: Sentence initial position, Focus: Preverbal position and X: neutral position or supplementary element

5. There is no description of information structure in Tok Pisin and Bislama, as far as we checked Dutton (1985), Crowley (2004), and Mihalic (1971).
6. Tok Pisin is mainly a spoken language and there are few materials of written texts (mainly Christian texts). This study collected both spoken and written discourses and they are shown in Texts section underneath.
8. Tok Pisin has cleft constructions, although the texts that this study used were not found. These examples (ii) and (iii) are from my recording, September 6, 2014. Cleft construction has a focusing effect and it is included in fronting movement.

(ii) food asde mpla kaikai

food yesterday 1pl eat

“The food we ate yesterday”

(iii) dispela haus hap youngpela boi bin wokin

this housethere young boy past build

“This house there young boy built”

9. There are emphatic words for expressing importance in information: *tru* (very, really), *tasol* (only, just)

10. Nose (2014) claimed that Tok Pisin has a kind of turn-taking expression, *yu tok/*yu toktok “you say” and *mi tok* “I tell you.”

**Texts:**

1. Bel: Bel speaker interview, elicitation of Bel grammar, spoken in Tok Pisin (Recorded on September 1, 2012)


3. FTC: Recoded discourse in translating FT (Recorded on August 25, 29, 2014)


5. RA: Web news in Radio Australia; *Ol toktok long sik measles long Solomon Islands na PNG* (23, October, 2014), Accessed on October 27, 2014 (http://shar.es/1HOaJ6)

6. YM: Yumi FM radio news (Recoded on September 3, 2013)

**References**


Mihalic, Francis. 1971. *The Jacaranda dictionary and grammar of Melanesian Pidgin.* Hong Kong:
Jacaranda.