

Building a semi-parallel corpus of Malay varieties: some preliminary findings

Asako SHIOHARA

ILCAA, Tokyo University of Foreign Studies

YANTI

Atma Jaya Catholic University of Indonesia

This article describes our attempts to build a semi-parallel corpus of Malay varieties spoken in Indonesia and Malaysia based on data collected through ‘the Jackal and Crow picture task’. The corpus enables us to directly compare the use of syntactic structures or deictic items under similar pragmatic conditions in elicited narratives as well as to provide an outline of the linguistic features of varieties whose affiliations are not well known. Our preliminary findings from the corpus are that (i) voice selection in narratives varies across varieties, and (ii) narratives elicited from some speakers in Makassar exhibit a distinctive structure of transitive clauses influenced by Makassarese.

1. Introduction¹

The aim of this paper is to introduce our attempts to build a semi-parallel corpus of Malay varieties spoken in Indonesia and Malaysia and report our preliminary findings based on the corpus. This is one of the research activities conducted as part of the joint research project titled ‘A Research on Varieties of Malayic Languages’ at the Research Institute for Languages and Cultures of Asia and Africa (AY 2017–2019).

The linguistic area of Malay covers a large area, including Indonesia, Malaysia, Brunei, Singapore, and South Thailand (Adelaar 2018:202). Malay belongs to the Malayic subgroup, which includes Malay proper and a large variety of Malay dialects and Malayic languages. Adelaar (2018:571–573) provides a socio-linguistic classification of Malay varieties, grouping them into (i) standard varieties such as *Bahasa Indonesia* (Standard Indonesian) and *Bahasa Malaysia* (Standard Malaysian); (ii) vernacular varieties, which are spoken in traditional Malay regions; and (iii) regional lingua franca varieties, which are spoken mainly in and around urban areas. The third category includes several types of varieties, which can be further sub-categorised as follows:

- (a) Established varieties: Varieties used before the spread of Indonesian as a national language. This category covers the varieties Paauw (2008) deals with under the

¹ The research based on this paper was supported by JSPS KAKENHI Grant Number 15K02472 and LingDy3 project of the ILCAA, TUFs. This paper was developed based on our presentations at the International Workshop on Malay Varieties held at the Tokyo University of Foreign Studies in 2018 and 2019, which was organised by the ILCAA joint research project titled ‘Research on Malayic Varieties’ (AY2017–2019). We would like to express our thanks to the participants of the workshops for their comments and suggestions during the workshops. We also thank two anonymous reviewers whose comments have helped improve this article. All errors are ours.

category of ‘the contact varieties of Eastern Indonesia’ or those that Adelaar and Prentice (1996) call ‘trade Malay’.

Its speakers often call it by a name combining the word *Bahasa* ‘language’ or *Melayu* ‘Malay’ and the name of the location, e.g. *Bahasa Manado* ‘Manado Malay’, *Bahasa Kupang* ‘Kupang Malay’, and *Melayu Papua* ‘Papuan Malay’.

(b) Emerging varieties, or regional inflected Indonesian: Varieties that have emerged after the spread of Indonesian as a national language. To some extent, these varieties have developed distinct structural features varying among them. The status of a variety may not be sufficiently established as independent, and reflecting its status, its speakers often recognise it as a colloquial or informal version of standard Indonesian, which is the view that some researchers adopt for languages in the ‘central’ area of Indonesia near Jakarta. Djenar et al. (2018), in their study of youth language observed in Indonesian urban areas, hold the position of ‘colloquial and Standard Indonesian as a continuum of registers in the same language, Indonesian’, and discuss how young people use the registers as a resource for style construction and move between styles during an interaction, rather than remaining within a certain defined register for particular speech activities (Djenar et al. 2018:141–142). Ewing (2020), in his study of the linguistic features of Indonesian spoken in Bandung, adopts a similar view. Based on his observation that ‘different speakers display different patterns, frequencies and interactional practices with regards to Sundanese elements that occur in their speech’, he states that ‘there is not a fixed Sundanese variety of Indonesian. Rather, speakers deploy Sundanese elements as part of their repertoire of semiotic resources’. In this paper, we will take a similar position regarding the varieties belonging to the category of emerging varieties. We will eschew giving the specific data elicited from individual speakers a label of the variety, such as Makassar Indonesian, Yogyakarta Indonesian, or Sumbawa Indonesian, because they have yet to be well-attested as established regional varieties. Rather, we will simply refer to the data as ‘Indonesian spoken in Makassar/Yogyakarta’, and so on.

Several survey works have been made of Malay varieties, such as Collins (1987) on Malay dialect research in Malaysia, Adelaar and Prentice (1996), Adelaar (2005), and Adelaar (2018), which give an exhaustive survey of Malay varieties from the historical and socio-linguistic points of view. Also, a number of descriptive works focusing on a specific variety or varieties of Malay have been written.² However, many Malay varieties remain understudied. Various vernacular varieties, which show great typological diversity (Adelaar 2018:573), have not been adequately described. Also, little work has been done on emerging varieties, especially outside the island of Java, although, as mentioned above, some recent works have clarified the status of urban Indonesians in the central part of Indonesia (Djenar et al. 2018, Manns 2019, Conners & Brugman 2020, and Ewing 2020). For some understudied or less-studied varieties, data collected employing this method may be the first step in describing their distinctive lexical and structural features and identifying the affiliation of each variety, while for other more established varieties for which substantial descriptions have been made, the

² For vernacular Malay, see Collins (1989), Nothofer (1997), Yanti (2010), and McDonnell (2016), and for established regional lingua-franca varieties, see Paauw (2008), a description and typological comparison of seven varieties that he calls ‘the contact varieties of Eastern Indonesia’ – Manado Malay, North Moluccan Malay, Ambon Malay, Banda Malay, Kupang Malay, Larantuka Malay, and Papua Malay – as well as Litamahuputty (2012) and Kluge (2017).

data collected may be used to compare the lexical or structural features of these varieties under similar discourse conditions.

The structure of this article is as follows. In Section 2, we present the methodology we adopted in building the corpus. In Section 3, we roughly describe the nature of the data that have been collected. In Section 4, we present our preliminary findings. In Section 5, we provide a summary.

2. The collection methodology in building the semi-parallel corpus

In building the semi-parallel corpus, we adopted the data elicitation method proposed by Carroll et al. (2011) of using the Jackal and Crow picture task with a set of nine pictures depicting scenes from the story of Jackal and Crow.

This type of non-linguistic stimulus for eliciting data has been adopted in many recent studies (Majid 2012). Two of the first widely known attempts were the ‘pear story’ project (Chafe 1980), which used a video as stimulus to seek various aspects of narrative production, and the ‘frog story’ project (Mayer 1969, Berman and Slobin 1994), which used a picture book to examine the development of linguistic ability in building a narrative.

The Jackal and Crow picture task was originally designed to record ‘data about a wide range of categories relevant to psycho-social cognition’, more precisely, data about the points ‘where the protagonists are motivated by their own thoughts and the suggestions and actions of others to do things, or feel things’ (Carroll et al. 2011); studies employing this task to investigate social cognition include Kratochvíl et al. (2018)³, dealing with Singapore Malay.

Our primary aim in employing this method, however, was to collect data as a first step to establish a comparative grammar of Malay that cover larger linguistic features rather than specific domains of the grammar.

The outline of the story shown by the nine pictures is as follows.⁴ First, a crow takes a fish from a basket. Second, a jackal comes and sees the crow fly to a tree. Third, the jackal thinks of the fish and salivates. Fourth, the jackal looks up at the crow with the fish. Fifth, the jackal calls to the crow that he should sing. Sixth, the crow is flattered and drops the fish. Seventh, the jackal eats the fish. Eighth, the jackal, satisfied, licks his lips. Ninth, the crow, having lost the fish, looks sad. Figure 1 shows all nine images in the task.



Figure 1: Images of the Jackal and Crow picture task

³ In Kratochvíl et al. (2018), narratives elicited using this task as well as other stimuli such as the Pear Story film (Chafe 1980) or the Frog story (Berman and Slobin 1994) are used to demonstrate (i) how the speakers’ stance is coded, (ii) how referents are categorized, and (iii) how information structure is marked in the variety.

⁴ This story is a traditional Sherpa tale. Similar versions of the story have been recorded in many places, including Botswana (Knappert 1985) and France (Carroll et al. 2011).

The data collector requests that the participants tell a story based on the cards. The cards are given to the participants one at a time, and the participants are asked first to tell the story while looking at each picture, and second to tell the story without looking at the pictures. In some sessions, the participants are asked to tell the story from the viewpoint of one of the protagonists (i.e. the crow or the jackal).

3. The data collected

We conducted fifteen sessions as shown in Tables 1 and 2.⁵ We chose the target varieties to cover all three categories mentioned in the Introduction. The selection of the varieties within the categories is rather opportunistic, that is, it depends on, to a great extent, the availability of the speakers and data collectors. Thus, in some places we have just collected data from only one participant.

The first two (1–2) include standard varieties (Category (i) in the Introduction), and the next four (3–6) belong to the category of vernacular varieties (Category (ii)); however, we realize that the data only cover representative varieties spoken in Borneo (Kalimantan) and Belitung, and have not included those spoken in Sumatra. Varieties 7–9 are ‘established’ lingua-franca varieties (Category (iii)a). The following two (10 Sabah and 11 Makassar) belong to the emerging varieties (Category (iii)b). In the remaining four locations (12–15) in Indonesia, namely Sumbawa Besar, Yogyakarta, Tarakan, and Nurabelen (Flores), we did not aim at any clear target varieties. They are all locations in which non-Malayic indigenous languages are predominant; sessions were conducted to capture a rough picture of the Indonesian as the second language spoken in those areas.

Table 1. List of recorded sessions (Part1)

No.	Location and data collector	Variety told to use in the recording	Number of participants
1	Jakarta, Indonesia (Yanti)	Standard Indonesian (<i>Bahasa Baku</i>) and colloquial Indonesian spoken in Jakarta (Sneddon 2006)	37
2	Kuala Lumpur, Malaysia (Hiroki Nomoto and Asako Shiohara)	Standard Malaysian	5
3	Weston, Sabah (Malaysia) (Asako Shiohara)	Brunei Malay (Bahasa Barunei)	10
4	Asajaya, Sarawak, Malaysia ⁶ (Hiroki Nomoto and Asako Shiohara)	Sarawak Malay (Bahasa Sarawak) (Collins 1987)	5

⁵ All the sessions were conducted in the area where the given variety is spoken, except the session with the Papuan Malay speaker, which was conducted in Tokyo, where the speaker was visiting. Most of the sessions were conducted by the present authors, but some were conducted by local collaborators who have been working with other members of the ILCAA joint research project mentioned in the Introduction.

⁶ https://github.com/matbahasa/Melayu_Sarawak/blob/master/README.md

Table 2. List of recorded sessions (Part2)

No.	Location and data collector	Variety told to use in the recording	Number of participants
5	Pontianak, Indonesia (Kazuya Inagaki)	Pontianak Malay (Kamal 1986, Mecer 1983)	1
6	Belitung, Indonesia (Yoshimi Miyake)	Belitung Malay (Bahasa Belitung) (Hoogstad and Tjik 2007)	1
7	Savanajaya, Buru, Maluku, Indonesia (Yoshimi Miyake)	North Maluku Malay (<i>Bahasa Maluku</i>) (Bowden 2012)	1
8	Kupang, Indonesia (Asako Shiohara)	Kupang Malay (Bahasa Kupang) (Jacob et al. 2000, Jacob 2020)	1
9	Tokyo (Asako Shiohara)	Papuan Malay (<i>Bahasa Papua</i>) (Kluge 2017)	1
10	Kota Kinabalu, Sabah, Malaysia (Hiroki Nomoto)	Sabah Malay (Bahasa Sabah) (Hoogervorst 2011, Wong 2000)	3
11	Makassar, Indonesia (Yanti and Asako Shiohara)	Standard Indonesian and colloquial Indonesian spoken in Makassar (Jukes 2013)	10
12	Sumbawa Besar, Indonesia (Asako Shiohara)	Indonesian	6
13	Yogyakarta, Indonesia (Yoshimi Miyake)	Indonesian	2
14	Tarakan, Indonesia (Antonia Soriente)	Indonesian	3
15	Nurabelen, Eastern Flores, Indonesia (Naonori Nagaya)	Indonesian	4

Figure 2 shows the locations where the data were collected.

**Figure 2. Locations where the data were collected**

All the sessions in Jakarta and Kuala Lumpur and some sessions in Makassar were conducted at universities and the participants are university students or lecturers. In other locations, the participants are in general educated people. In two locations, Jakarta and Makassar, we asked the participants to tell a story both in the standard variety (*Bahasa Indonesia baku*) and the colloquial variety (*Bahasa Indonesia sehari-hari*) in order to clearly establish differences between the two varieties. Despite these attempts, eliciting data exhibiting the features of the target variety is not very easy, though the emergence of distinctive features is well recognised by speakers and attested in previous studies (i.e. Sneddon (2006) for Colloquial Indonesian spoken in Jakarta and Jukes (2013) for Indonesian spoken in Makassar). This is presumably at least in part due to the fact that the method was predominantly designed to collect narratives, which speakers tend to associate with a formal register. The location and the consultant's profession may influence the variety recorded. We conducted two sessions in Makassar, the first with senior university lecturers as participants (six people), and then with younger people (four people), three of whom were employees, the other a graduate student. The first session was conducted on campus and the second outside the campus setting. The lecturers tended to use similar registers throughout the sessions, whereas the other four participants used quite distinct registers as per the data collectors' request. A similar result was observed in the sessions conducted in Jakarta (See Yanti and Shiohara 2019).

Presumably because the story is not native to the target societies, a large diversity was observed among the participants in their interpretations of the species of animals, as well as the events depicted in the pictures; only a small number of participants told the story according to the plot we expected. This diversity is most salient in the description of the scenes in which the jackal and crow exchange the fish; for the expected scenario of the jackal's getting the crow's fish through flattery, some participants described the crow as sharing the fish with the jackal because he felt pity for him. Others thought that the jackal asked the crow to share the fish and the crow agreed to give the fish to the jackal. Thus, the plots of the elicited stories vary among participants. In Section 4, we compare the introductory part of the collected stories, where variations of the plot are observed to a lesser extent.

Notwithstanding the above-mentioned limitations, the corpus has allowed us to make direct lexical and structural comparison among the varieties; the collected data may be used to compare the usage of clause structure type and NP type (e.g., lexical NPs or pronouns) only observable under specific pragmatic conditions, for example, the use of voice-related clauses and agreement between the predicate and the argument.⁷ We present our preliminary findings in this direction in 4.1. In addition, when we do not have a clear idea about the affiliation of the variety spoken in one area, the data collected may help us to identify distinctive features of less studied or unstudied varieties. In 4.2, we present our preliminary findings on such features found in the Indonesian spoken in Makassar.

⁷ Another point that could exhibit considerable difference among varieties is how 'given' referents (i.e., referents that are already introduced to the discourse) are marked (e.g., by demonstratives or special anaphoric markers, such as *tersebut* 'PFV.PASS-mention' in Standard Indonesian). However, it is beyond the scope of this paper to pursue the point any further.

4. Preliminary findings

4.1 Voice selection in narratives

The data collected enable us to compare the usage of structures and lexical items observable under specific pragmatic conditions. One of the points exhibiting considerable diversity among the varieties collected is the usage of transitive clause structures. In this section, we will present our preliminary findings on this point by comparing the clauses describing the first two pictures shown in Figure 3 in certain varieties.

Standard Indonesian and Malaysian have two voices, active and passive voice⁸ (Sneddon et al. 2012:255–264).

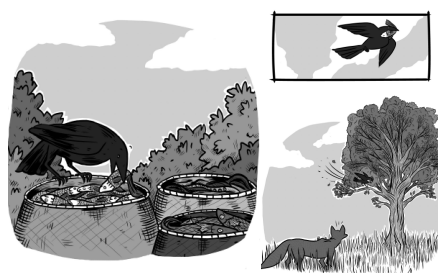


Figure 3. Pictures 1 and 2 from the ‘Jackal Crow Picture Task’ (Carroll et al. 2011)

Standard Indonesian and Standard Malaysian appear to be similar in that active voice is used in all the clauses. As shown in (1), an example from Standard Indonesian, and (2), an example from Standard Malaysian, all the participants told the story using active clauses that mark the verb with the standard nasal prefix *meng-*, such as *mengambil* ‘to take’, *membawanya* ‘to take it’ (1), *mencuri* ‘to steal’, and *membawa* ‘to take’ (2).

(1) Standard Indonesian

jadi ter-dapat... ikan dan burung. burung ter-sebut...
 so PFV.PASS-exist fish and bird bird PFV.PASS-mention

meng-ambil... ikan dan meng-ambil=nya
 ACT-take fish and ACT-take=3

mem-bawa=nya ke... atas pohon.
 ACT-bring=3 to above tree

‘So, there was a fish and a bird. The above-mentioned bird took a fish and took and brought it to (the top of) a tree.’

⁸ In Indonesian, there is another voice type which has been discussed quite robustly in the literature and is referred to as ‘object voice’ or ‘passive type two’ (among others: Chung 1976, Arka & Manning 1998, Cole et al. 2008). However, we currently only focus on voice on the active and passive voice.

(2) Standard Malaysian spoken in KL

pade *suatu* *hari* *si* *gagak* *telah* ***men-curi*** *se-ekor* *ikan*
 on one day ART crow PFV ACT-steal one-CLF fish
daripade *nelayan,* *se-terus=nye,* *si* *gagak* *telah*
 from fisherman one-continue=3 ART crow PFV
mem-bawa *ikan* *ter-sebot* *ke* *se-buah* *pokok.*
 ACT-bring fish PFV.PASS-mention to one-CLF tree

‘One day a crow stole a fish from a fisherman, then the crow brought the above-mentioned fish to a tree.’

The Indonesian spoken in Tarakan and in Flores appears very similar to Standard Indonesian, as shown in (3) and (4).

(3) Indonesian spoken in Tarakan

di *sini* *ada* *se-ekor* *burung* *yang* *coba* ***meng-ambil*** *ikan*
 at here exist one-CLF bird REL try ACT-take fish
di *wadah* *atau...* *tempayan* *tempat* *ikan,* *keranjang* *ikan.*
 at container or crock place fish basket fish
kemudian *burung* *itu* ***meng-ambil*** *ikan.*
 then bird that ACT-take fish

‘Here, there was a bird which was trying to take a fish in a container or a crock for fish, a basket for fish. Then the bird took the fish.’

(4) Indonesian spoken in Flores

ini *ada* *se-buah* *kantong* *ikan* *atau* *tiga* *bakul*
 this exist one-CLF bag fish or three basket
yang *ter-isi* *penuh* *dengan* *ikan.*
 REL PFV.PASS-contents full with fish
lalu *ada* *burung* *hitam* *yang* *makan* *ikan* *ter-sebut*
 then exist bird black REL eat fish PFV.PASS-mention
lalu *burung* *hitam* *itu* ***mem-bawa*** *ikan* *itu.*
 then bird black that ACT-bring fish that

‘There was a fish bag, or three baskets filled with fish. Then there was a black bird which ate the fish. Then the black bird took the fish.’

Pontianak Malay, one of the vernacular varieties spoken in Indonesia, and the Indonesian spoken in Sumbawa Besar and in Yogyakarta demonstrate slightly different structural features, as exemplified in (5), (6), and (7), respectively. Sentence (5) is the example of Pontianak Malay. Here, the speaker tells the story employing an active clause in the first clause and then a passive clause with the *di*-prefixed verb in subsequent clauses.

(5) Pontianak Malay⁹

jadi, buyong gagak men-cuyi iwak dalam kyanjang.
 become bird crow ACT-steal fish inside basket

kemudian iwak tu di-bawa' teybang ke atas pohon,
 then fish that PASS-bring fly to above tree,
tempat sayang=a. di-tengok lah sama musang.
 place nest=3 PASS-see PTC by weasel

‘So, a crow stole a fish from a basket. Then (the crow) flew bringing the fish to the tree, the place of his nest. (The crow) was seen by a weasel.’

Voice selection varies among speakers in Yogyakarta and Sumbawa Besar. The narratives told by some elderly speakers exhibit a similar pattern to that observed in Pontianak Malay, as shown in Sentences (6) and (7), respectively. In Yogyakarta, where we collected data from only two speakers, another speaker gives a narrative exhibiting a similar pattern to Standard Indonesian shown in Sentence (1). Also, in Sumbawa Besar, it is only two relatively elderly speakers who used passive voice clauses; the remaining four speakers give narratives exhibiting a similar pattern to Standard Indonesian. At this stage of our research, it is still uncertain that the differences in the voice forms are due to generational differences, as the number of samples obtained for each variation is small.

(6) Indonesian spoken in Yogyakarta

burung gagak me-lihat banyak ikan yang ada
 bird crow ACT-see a.lot fish REL exist

di keranjang, milik nelayan, dia langsung matuk ikan,
 at basket owned fisherman 3 direct ACT.peck fish

ambil¹⁰ satu ikan, kemudian di-bawa=nya terbang.
 take one fish then PASS-bring=3 fly

‘A crow saw a lot of fish in a basket owned by a fisherman. He pecked it right away, taking one fish and then he flew bringing (it).’

(7) Indonesian spoken in Sumbawa Besar (a relatively elderly speaker)

tiba-tiba datang se-ikor burung gagak lalu meng-ambil ikan itu.
 suddenly come one-CLF bird crow then ACT-take fish that

dan ikan itu di-bawa terbang ke jauh.
 and fish that PASS-bring fly to far

‘Suddenly a crow came and took the fish. And (the crow) brought the fish to (a) far (place).’

⁹ The example of Pontianak Malay includes phonetic, morphological, and lexical elements mentioned in the description of this variety in Kamal (1986:2), cited in Adelaar and Prentice (1996:680). For example the voiced velar fricative γ is the reflex of Standard Indonesian r .

¹⁰ The bare stem form *ambil* ‘take’ shows that this clause is neither an active nor a passive voice clause. The status of this clause in this variety is not clear.

In the Brunei Malay data, one of the vernacular Malay varieties spoken in Malaysia, speakers in general use passive structures, which are marked by prefix *di-*, to a greater extent, as shown in Sentences (8) and (9).

(8) Brunei Malay (I)

gagak ani ka-lapar-an, di-ambil=nya sa-ikung lauk atu
crow this PASS-hungry-CIRC PASS-take=3 one-CLF fish that

‘The crow was hungry; he took that fish.’

(9) Brunei Malay (II)

di siring utan ada sa-ikung gagak ka-lapar-an tarabang
at side forest exist one-CLF crow PASS-hungry-CIRC fly

punya tarabang ba-jumpa lauk, lauk dalam bayung
have fly INTR-meet fish fish inside bucket

di-ampir-i lauk ani di-rungkup=nya sa-ikung
PASS-approach-APPL fish that PASS-take=3 one-CLF

di-bawa=nya tarabang sampai ka s-puhun kayu.
PASS-bring=3 fly to to one-CLF tree

‘Near the side of forest, there was a crow who was hungry, flying and found a fish, a fish in the basket. (The crow) approached the fish, he took one, (then) he brought it, flying to one tree.’

Some Brunei Malay speakers also use a so-called zero-passive clause, in which the verb stem with an enclitic pronoun indicating the agent is used,¹¹ confirming the description of Clynes (2001:27).

(10) Brunei Malay (III)

pada masa atu ada sa-ikung burung gagak
one time that exist one-CLF bird crow

sadang ah anu kan man-curi lauk-lah
PROG ITJ what’s.it.called will ACT-steal fish-PTC

jadi curi=nya lauk, lauk atu di satu bakul.
then steal=3 fish fish that in one bucket

‘One time, there was a crow, (he) was going to steal a fish. Then he stole the fish, the fish in one bucket.’

It is widely known that *di-*prefixed passive clauses have the ‘event focus’ function (Hopper 1984 and Kaswanti 1988, among others), that is, the function of expressing ‘sequenced events which pertain to the main line of the discourse’ (Hopper 1984:84). As claimed by Cumming (1991) regarding the written Indonesian of the time, and as can be seen in the Standard Indonesian data collected in this study (e.g., sentence (1)),

¹¹ A similar type of ‘zero-passive’ structure is attested in Sarawak Malay. See Nomoto (2019).

this function has become less common in recent Indonesian.¹² We can see that narratives obtained from one of the speakers in Yogyakarta and speakers of vernacular varieties retain this traditional use of the passive clause.

As for the Indonesian spoken in Sumbawa Besar, we should consider another factor, the influence of the speaker's first language. In Sumbawa Besar, the participants in the task were all native speakers of Sumbawa. Though it is not a Malayic language, it has a transitive structure quite similar to the passive voice of Standard Indonesian. Sentence (11) is a narrative elicited through the same task, told in the Sumbawa language.

(11) Sumbawa (ISO 639-3: smw)¹³

tiba-tiba datang pio gagak datang ètè jangan nan
suddenly come bird crow come take fish that

lalu ya=bawa ngibar ko do' benar
then 3=bring fly to far really

nyampè mo ko sópó puin-kayu
arrive PTC to one tree

‘Suddenly a crow came and took the fish. Then it (the crow) flew bringing (it) to a really far (place), (and it) arrived at a tree (branch).’

Sumbawa does not have a voice opposition of active voice and passive voice, but only uses one type of transitive clause with a bare stem verb. The transitive verb may be marked by a pronominal proclitic indicating the transitive agent, and the conditions under which the pro-cliticised verb occurs overlap with the event focus *di*-passive clauses in Indonesian to a great extent. The pro-cliticised verb occurs only under the condition that (i) the agent referent is given in discourse and (ii) the agent NP does not occur in the form of a pre-predicate NP. In addition, the third person proclitic occurs only with transitive verbs.

In Sentence (11), the pro-cliticised verb occurs only in the second clause, not the first, which matches the occurrence of the *di*-passive clause in the corresponding Indonesian example shown in Sentence (1). Thus, one possibility is that the Sumbawa speakers use *di*-passive clauses in narratives on analogy with the corresponding Sumbawa sentence.

As mentioned above, younger speakers, however, do not use this type of ‘event focus’ passive clause. Sentence (12) is an example obtained from a high school student. We could see this as reflecting the use of active clauses in Standard Indonesian as a default structure in narratives through education; her frequent use of a formal expression *tersebut* ‘PFV.PASS-mention’ might support this view.

¹² For details of the ‘event focus’ function, see Kroeger (2014:21ff.), who gives a survey of this issue.

¹³ Note that in Sumbawa, the reflex of so-called nasal prefix which marks active voice in many Malay varieties, derives intransitive verbs. Thus, the verbs *ngibar* ‘fly’ and *nyampè* ‘arrive’ in sentence (11) are intransitive verbs derived from the base *-kibar* and *sampè*, respectively, both of which are not transitive bases; *-kibar* is a morpheme which cannot be used as an independent word by itself, and *sampè* is used as preposition ‘until, to’ when used alone.

(12) Indonesian spoken in Sumbawa Besar (a relatively young speaker)

burung ter-sebut mem-bawa se-ekor ikan, terbang ke atas
 bird above-mentioned ACT-bring one-CLF fish fly to above
pohon. lalu di bawah pohon ter-sebut ternyata ada
 tree then at under tree above-mentioned it.turned.out exist
se-ekor anjing. anjing ter-sebut memandang si burung.
 one-CLF dog dog above-mentioned ACT.gaze ART bird.

‘The above-mentioned bird brought a fish, (and) flew onto a tree. Then, under the above-mentioned tree, it turned out that there was a dog. The above-mentioned dog gazed at the bird.’

Thus far, we have seen differences in the voice selection observed among varieties. The situation is different in regional lingua franca Malays spoken in eastern Indonesia. As Paauw (2008) and Adelaar (2005) suggest, lingua franca Malay varieties spoken in eastern Indonesia, such as Kupang Malay, Papuan Malay, and Manado Malay, do not exhibit an opposition of active and passive voice. The Kupang Malay and Papuan Malay data collected in our research confirm this point. Sentence (13) is an example of Kupang Malay. Here, the verb occurs in the bare stem form, which does not show voice.

(13) Kupang Malay

burung gagak hitam dia pi curi ikan di pinggir laut
 bird crow black 3 go steal fish in side sea
di sana ada bakul, dia pi toto ama ini ikan.
 at there exist bucket 3 go pick take this fish
jadi ia terbang
 so 3 fly

‘The black crow, he went stealing a fish in the seaside. There was a basket there. He went picking and taking this fish, then he flew.’

4.2 Indonesian spoken in Makassar

Jukes (2013:132), in his research on the TAM clitics of Makassarese, mentions that the Indonesian spoken in Makassar heavily employs Makassarese aspectual clitics. Sentence (14) is an example given by Jukes (2013:132) in which the perfective clitic =*mi* of Makassarese appears in Indonesian.

(14) Indonesian spoken in Makassar

*bikin apa=*mi* itu anak-anak?*
 make what=PFV that child.RED

‘What did those kids do?’ (Jukes 2013:132)

In Makassarese, the clitic *mi* is a combination of the TAM clitic *mo* and the third person absolutive enclitic =*i*, but in the Indonesian spoken in Makassar *mi* is a default perfective clitic used irrespective of the person of the referent of the S or P (Jukes 2014). Thus, we can see that the Makassar features are incorporated into Indonesian grammar in a simplified way.

Our data collected in Makassar show a more pervasive use of a transitive clause structure similar to Makassarese; it is observed in the use of bare stem verbs and the development of pronominal clitics.

Sentence (15) provides the first four clauses of the narrative elicited from one of the speakers. Each clause includes one or more transitive verbs, and they all occur in a bare stem form without any of the prefixes marking grammatical voice in Standard Indonesian. In addition, all the transitive verbs occur with the pronominal clitics indicating the agent or the patient of the situation; in (a), the third person absolutive enclitic in *ambil=ki* ‘take it’ indicates the patient, in (b), the third person ergative proclitic in *na=lia* indicates the agent, and in (c) and (d), the predicate is marked with both ergative and absolutive markings.

(15) Indonesian spoken in Makassar (I)

(a) *ada tadi burung, burung ambil=ki itu ikan=ka*
 exist before bird bird take=3ABS that fish=DEF

(b) *ada ikan tiga bakul di situ na=lia tuh*
 exist fish three bucket at there 3ERG=see that

(c) *na=ambil=mi satu itu ikan=ka*
 3ERG=take=PFV one that fish=DEF

(d) *baru na=makang=ki, na=bawa=ki*
 after 3ERG=eat=3ABS 3ERG=bring=3ABS

pergi di anu di pohon=ka
 go at what's.it.called at tree=the

‘There was a bird, the bird was taking the fish. There were three buckets of fish there, he saw that. He took one piece of the fish. Just after he ate it, he went to what’s-it-called, to the tree.’

In Makassarese, the unmarked transitive clause has a bare stem form with the ergative clitic for the agent and the absolutive clitic for the patient. The predicates observed in clauses in Sentence (15) to a large extent follow this pattern.

Sentence (15) shows other types of distinctive morpho-syntactic features that can also be seen as influence from Makassarese. First, the form *=ka*, which functions as a definite clitic, is used as a definite marker in the NPs *ikan=ka* ‘the fish’ in (a) and (c) and *pohon=ka* ‘the tree’ in (d). Second, the relative order of the demonstrative *itu* and the head noun is opposite the canonical order in Standard Indonesian, in which the head noun precedes the demonstrative. In the NP *itu ikan=ka* ‘that fish’ observed in (15a) and (15c), the demonstrative *itu* occurs before the head NP. The constituent order within the NP is observed in Makassarese (Jukes 2006:188).

Though the ergative pronominal clitic *na=* can be claimed as a borrowing of the corresponding Makassarese form *na=* ‘3ERG’ (Jukes 2013:124), the absolutive third person enclitic *=ki* and the definite clitic *=ka* cannot be seen as direct borrowings from Makassarese; the Makassarese counterpart of each element is *=a* and *=i*, respectively (Jukes 2013:124). In Makassarese, the sound *k* is inserted as a rule between a vowel-initial clitic, such as *=a* and *=i* and a glottal-final word (e.g., *juku?* ‘fish’). Thus, the combination of *juku?* ‘fish’ and the definite clitic *=a* becomes *juku-k=a* ‘the fish’ (Jukes

2006:97). The sound *k* observed in *=ki* ‘3ABS’ and *=ka* ‘DEF’ in these Indonesian narratives might have arisen on analogy with the Makassarese inserted sound *k*, after which the whole forms *=ki* ‘3ABS’ and *=ka* ‘DEF’ came to be recognised as the clitic in the contemporary Indonesian spoken in Makassar (Jukes 2013).

Sentence (16) is from a narrative elicited from another speaker. In this text, features of Makassarese origin are used to a lesser extent: although the TAM clitic *=mi* ‘PFV’ is extensively used, the pronominal clitic is observed with three predicates, *liat-liat-i=ki* ‘see=3ABS’ in Clause (d), *menyanyi=ko* ‘sing=2ABS’, and *na=makan=mi* in Clause (i). In addition, the third person possessive form *=nya* occurs instead of the clitic *=ka* indicating definiteness, in the same way that the form is used in colloquial Indonesian (Sneddon et al. 2012:154–156). In clause (b), the *di*-prefixed passive is used, although the use of the passive verb in this context is not typical in Standard Indonesian, as mentioned in 4.1.

(16) Indonesian spoken in Makassar (II)

- (a) *ada tuh cerita. ini burung gagak*
 exist PTC story this bird crow
dia mo pi cari makan
 3 want go look.for eat
- (b) *nah dia datang=mi di-ambil=mi ikan di dalam bak*
 ITJ 3 come=PFV PASS-take=PFV fish at inside basin
- (c) *terus toh, sementara dia bawa=mi itu ikan*
 then PTC while 3 bring=PFV.3 that fish
- (d) *ternyata ada serigala yang lihat-liat-i=ki dari jauh*
 unexpectedly exist jackal REL see-INS-APPL=3 from far
- (e) *nah itu serigala datang=mi, pergi bertanya di burung gagak*
 ITJ that jackal come=PFV go ask at bird crow
- (f) *gagak, eh... menyanyi=ko dulu eh*
 crow ITJ ACT.sing=2SG just ITJ
- (g) *terus itu si burung gagak, menyanyi=mi*
 then that ART bird crow ACT.sing=PFV
- (h) *dan secara ber-sama-an, jatuh=mi juga itu ikan=nya ...*
 and way INTR.together.NMLZ fall=PFV also that fish=3
- (i) *jadi na=makan=mi Itu serigala itu ikan...*
 then 3ERG=eat=PFV that jackal that fish

‘In the story, there was a crow, which went looking for food. Well, it came and took fish in the basin. Then while it (the crow) brought the fish, there happened to be a jackal who saw it (the crow) from a far place. Well, the jackal came and asked the crow. “Crow, just sing to me.” Then the crow sang and, at the same time, the fish fell. Then it (the jackal) ate the fish.’

In sum, the narratives elicited from some speakers in Makassar exhibit morpho-syntactic features quite distinctive from those of any other variety due to the heavy influence of Makassarese. However, the extent to which the distinctive Makassar-like features are observed in the narratives elicited varies among speakers.

5. Summary

In this paper, we have presented our attempts to build a semi-parallel corpus of Malay varieties spoken in Indonesia and Malaysia through the Jackal and Crow picture task. We conducted fifteen sessions to collect sixteen Malay varieties. Four sessions were conducted in locations where non-Malay indigenous languages are predominant, and in two locations, Tarakan and Nurabelen, the data exhibit features very similar to Standard Indonesian. In two other places, Sumbawa Besar and Yogyakarta, the data collected from older speakers exhibit slightly different features, as mentioned below.

The Jackal and Crow task we adopted for this study is predominantly designed to collect narratives, and as the features observable in interactions rarely appear in the data, the corpus data may thus miss some distinctive regional features that would be observable in conversations. The nature of the task also has the disadvantage of inducing speakers to choose a formal and standard variety, which is easily associated with recounting narratives, especially in locations such as Jakarta and Makassar where the regional varieties are considered colloquial or informal versions of Standard Indonesian.

Notwithstanding the above-mentioned limitations, the corpus has allowed us to make direct lexical and structural comparison among the varieties; the collected data may be used to compare the usage of structures only observable under specific pragmatic conditions, for example, the use of voice-related clauses and agreement between the predicate and the argument. Our preliminary findings can be summarised as follows:

(i) The pragmatic condition for voice selection varies among the varieties that exhibit an opposition of active and passive voice in the same way that standard varieties do; in the standard varieties and Indonesian spoken in Tarakan and Nurabelen, the active voice is extensively used, while in vernacular varieties like Pontianak Malay and Brunei Malay and in the Indonesian spoken by the older speakers of Sumbawa Besar and Yogyakarta, the ‘event focus’ usage of passive clauses (Hopper 1984, Kaswanti 1988), which has become less common in recent Indonesian (Cumming 1991), is retained.¹⁴

(ii) In the Indonesian spoken in Makassar, we could observe the pervasive use of a transitive clause structure similar to that of Makassarese, with bare stem verbs and the development of pronominal clitics, as well as the heavy use of the TAM clitic *mi* mentioned in Jukes (2013), which is also borrowed from Makassarese.

Further research based on other discourse types is needed to reveal more of these features: narrative data spontaneously told should be collected to verify the point (i), and collecting conversational data is necessary to investigate further details of Indonesian spoken in Makassar shown in the point (ii), in order to see how the first and second person pronominal clitics occur, and how TAM clitics, including the form *mi*, function in interactions. Our experience collecting the data in Makassar shown in Section 3 suggests that Indonesian exhibiting Makassarese influence is spoken only in limited registers. More socio-linguistic research through interviews with the speakers is needed to gain a better understanding of the linguistic situation in the area.

¹⁴ For details of the ‘event focus’ function, see Kroeger (2014:21ff.) for a survey of this issue.

Abbreviations

2 second person, 3 third person, ABS absolutive, ACT active voice, APPL applicative, ART article, CIRC circumfix, CLF classifier, DEF definite, ERG ergative, ITJ interjection, INTR intransitive, IST inserted vowel, PASS passive voice, PFV perfective, PROG progressive, PTC particle, REL relativiser, SG singular

References

- Adelaar, Alexander. 2005. Structural diversity in the Malayic Subgroup. Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian Languages of Asia and Madagascar*. 202-226.
- Adelaar, Alexander. 2018. Dialects of Malay/Indonesian. In Charles Boberg & John Nerbonne (eds.), *The handbook of dialectology*, New Jersey: Wiley-Blackwell. 571–580.
- Adelaar, Alexander K. & D. J. Prentice. 1996. Malay: its history, role, and spread. In S. A. Wurm, P. Mühlhäusler & D. Tryon (eds.), *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. 673–693. Berlin/New York: Mouton de Gruyter.
- Arka, I Wayan & Christopher D. Manning. 1998. Voice and grammatical relations in Indonesian: A new perspective. In M. Butt, T. H. King (ed.), *The proceedings of the LFG '98 Conference*. Stanford: CSLI Publications.
- Berman, Ruth & Dan Slobin. 1994. *Relating events in narrative: A crosslinguistic developmental study*. New York/London: Psychology Press.
- Bowden, John. 2012. Local languages, local Malay, and Bahasa Indonesia: A case study from North Maluku. *Wacana* 14(2). 313–332.
- Carroll, Alice, Barbara Kelly & Lauren Gawne. 2011. The jackal and crow picture task. Designed for use by the Social Cognition and Language Project. A collaboration of The Australian National University, Griffith University, University of Melbourne and the Max Planck Institute for Psycholinguistics.
- Chafe, Wallace. 1980. *The pear stories*. Norwood, NJ: Ablex.
- Clynes, Adrian. 2001. Brunei Malay: An overview. *Occasional Papers in Language Studies, Department of English Language and Applied Linguistics, Universiti Brunei Darussalam* 8. 11–43.
- Cole, Peter, Gabriella Hermon & Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118. 1500–1553.
- Collins, James T. 1987. *Dialek Melayu Serawak*. Kuala Lumpur: Dewan Melayu dan Pustaka.
- Collins, James T. 1989. Malay dialect research in Malaysia: The issue of perspective. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch Indië* 145. 235–264.
- Connors, Thomas J. & Claudia M. Brugman. 2020. Javanese influenced Indonesian: features from two conversations. In Atsuko Utsumi and Thomas J. Connors (eds.), *Various aspects of Malay varieties*. NUSA 68, 67-84. [Permanent URL: <http://repository.tufs.ac.jp/handle/10108/94894>]

- Chung, Sandra 1976. An object-creating rule in Bahasa Indonesia. *Linguistic Inquiry* 7. 41–87.
- Cumming, Susanna. 1991. *Functional change: The case of Malay constituent order*. Berlin: Mouton de Gruyter.
- Djenar, Dwi Noverini, Michael Ewing & Howard Manns. 2018. *Style and intersubjectivity in youth interaction*. Berlin/Boston: De Gruyter Mouton.
- Ewing, Michael. 2020. Features of Indonesian in Bandung. In Atsuko Utsumi and Thomas J. Connors (eds.), *Various aspects of Malay varieties*. *NUSA* 68, 51–66. [Permanent URL: <http://repository.tufs.ac.jp/handle/10108/94895>].
- Hoogervorst, Tom G. 2011. Some introductory notes on the development and characteristics of Sabah Malay. *Wacana* 13(1). 50–77.
- Hoogstad, Salim Yan & Tjik, Erna. 2007. *Kamus kecil: bahasa Melayu Belitong*. Belitung, Indonesia: Yayasan PAUN Belitung.
- Hopper, Paul. 1984. Ergative, passive, and active in Malay narrative. In Flora Klein-Andreeu (ed.), *Discourse perspectives in syntax*. 67–88. New York: Academic Press.
- Jacob, June & Charles E. Grimes, compilers. 2000. *Kamus pengantar Bahasa Kupang: Malayu Kupang–Bahasa Indonesia, Bahasa Indonesia–Bahasa Kupang, Kupang Malay–English, English–Kupang Malay*. *Paradigma B–10*. Kupang: Artha Wacana Press.
- Jacob, June. 2020. *A sociolinguistic profile of Kupang Malay, a creole spoken in West Timor, Eastern Indonesia*. MS.
- Jukes, Anthony. 2006. Makassarese (basa Mangkasara'): A description of an Austronesian language of South Sulawesi. Ph.D. Thesis, University of Melbourne. Department of Linguistics and Applied Linguistics
- Jukes, Anthony. 2013. Aspectual and modal clitics in Makassarese. In John Bowden (ed.), *Tense, aspect, mood and modality in languages of Indonesia*. *NUSA* 55, 123–133. [Permanent URL: <http://hdl.handle.net/10108/74329>].
- Jukes, Anthony. 2014. Belumpi: Makasar clitics in Makassar Indonesian. Paper presented at ISMIL18 on 13–15 June 2014, Procida, Naples, Italy.
- Kamal, Mustafa. 1986. *Morfologi dan sintaksis bahasa Melayu Pontianak* (Morphology and syntax of Pontianak Malay), Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
- Kluge, Angela. 2017. *A grammar of Papuan Malay*. Utrecht: LOT.
- Kaswanti Purwo, Bambang. 1988. Voice in Indonesian: A discourse study. In Masayoshi Shibatani (ed.), *Passive and voice*. 195–241. Amsterdam: John Benjamins.
- Knappert, J. 1985. *Myths and legends of Botswana, Lesotho, and Swaziland*. Leiden: E. J. Brill.
- Kratochvíl, František, Nur Izdihar Binte Ismail & Diyana Hamzah. 2018. Stance, categorisation, and information structure in Malay. In Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds.), *Perspectives on information structure in Austronesian languages*. 41–80. Berlin: Language Science Press. DOI: 10.5281/zenodo.1402537

- Kroeger, Paul. 2014. Passive agents in Malay: The binding properties and discourse functions of agentive =nya. In Siaw-Fong Chung & Hiroki Nomoto (eds.), *Current trends in Malay linguistics*. NUSA 57, 5–29. [Permanent URL: <http://hdl.handle.net/10108/79283>]
- Litamahuputty, Betty. 2012. *Ternate Malay: Grammar and texts*. Utrecht: LOT.
- Majid, Asifa. 2012. A guide to stimulus-based elicitation for semantic categories. In Nicholas Thieberger (ed.), *The Oxford handbook of linguistic fieldwork*. 54–71. Oxford: Oxford University Press.
- Manns, Howard. 2019. Gaul, conversation and youth genre(s) in Java. NUSA 67, 3–18. [Permanent URL: <http://repository.tufs.ac.jp/handle/10108/93962>].
- Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Press.
- McDonnell, Bradley. 2016. *Symmetrical voice constructions in Besemah: A usage-based account*. Ph.D. thesis. University of California, Santa Barbara, USA.
- Mecer, A. R. 1983. *Kedudukan dan fungsi bahasa Melayu Pontianak* (Status and function of Pontianak Malay). Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Nomoto, Hiroki. 2020. In Atsuko Utsumi and Thomas J. Connors (eds.), *Various aspects of Malay varieties*. NUSA 68, 141–159. [Permanent URL: <http://repository.tufs.ac.jp/handle/10108/94897>]
- Nothofer, Bernd. 1997. *Dialek Melayu Bangka*. Bangi: Penerbit University Kebangsaan Malaysia.
- Paauw, Scott H. 2008. *The Malay varieties of Eastern Indonesia: A typological comparison*. Ph.D. thesis. State University of New York: Buffalo, USA.
- Sneddon, James N. 2006. *Colloquial Jakartan Indonesian*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Sneddon, James N., K. Alexander Adelaar, Dwi N Djenar & Michael Ewing. 2012. *Indonesian: A comprehensive grammar*, 2nd ed. London/New York: Routledge.
- Wong, Jane Kon Ling. 2000. *The Sabah Malay dialect: Phonological structures and social functions*. Kota Kinabalu: University Malaysia Sabah.
- Yanti. 2010. *A reference grammar of Jambi Malay*. Ph.D. thesis. University of Delaware, Newark, USA.
- Yanti & Asako Shiohara. 2019. Some observations from Jackal and Crow data collected in Jakarta. Paper presented at the Third International workshop on Malay varieties, 31 Nov–1 Dec at the Research Institute for Languages and Cultures, Asia and Africa, Tokyo University of Foreign Studies.