The present volume of *NUSA* features Malay. Among Malay’s numerous regional and sociolinguistic varieties, the volume is primarily concerned with the varieties in the Malay Peninsula, though other varieties are also discussed. The papers in this volume represent what we think are the “current trends in Malay linguistics”: the use of digitized data such as corpora and web data, theory-informed language description and attention to language contact.

The first two papers make extensive use of digitized data and discuss the same theoretical issue, namely passive agents. The currently most popular view of the voice system of Malay is that of a “symmetric voice system” akin to those of the so-called Philippine-type languages, whereby in addition to the traditional active and passive, there exists an additional voice referred to as the ‘objective voice’, ‘bare passive’, etc. (e.g. Arka & Manning 1998; Cole, Hermon & Yanti 2008).

Paul Kroeger accepts the symmetric voice hypothesis, and discusses the proper classification of clauses with *di-* verbs whose agents are expressed by the third person enclitic pronoun \(^{=}nya\) without the preposition *oleh* ‘by’ (*di-V=nya*). The grammatical functional status of \(^{=}nya\) has been controversial. While I Wayan Arka initially analyzed it as “core” (Arka & Manning 1998), later he reanalyzed it as “semi-core” (Arka 2005). In his paper, Kroeger claims that it is “oblique,” pointing out that the binding fact on which the aforementioned studies rely should be accounted for in terms of discourse functions of \(^{=}nya\) rather than its syntactic status. The syntactic status of \(^{=}nya\) is important in frameworks that consider grammatical functions as primitive (e.g. LFG), as it determines the voice category of *di-V=nya* clauses: objective if \(^{=}nya\) is core, neither passive nor objective if it is semi-core, and passive if it is oblique. Kroeger’s conclusion is that *di-* always marks the passive voice.

Hiroki Nomoto and Kartini Abd. Wahab are skeptical about the symmetric voice hypothesis. For them, the verbal morphology (i.e. *di-* or bare) does not demarcate two distinct voices (i.e. passive and objective), but it is one of the distinctive characteristics of different passive subtypes. This view is based on their analysis of the following facts concerning *di-* passive agents: first and second person agents are infrequent but possible, and are usually expressed implicitly. They account for these facts in terms of the (in)compatibility between the informational salience/givenness required for the *di-* passive agent position and that of the noun phrase which fills this position. Nomoto and Kartini claim that the salience status of the former is determined by that of the eventuality described by the *di-* verbal phrase, which in turn is linked to the syntactic verb movement involved in *di-* passives. The presence of this verb movement distinguishes *di-* passives from bare passives syntactically and informationally. However, this does not mean that the verbal morphology marks different voices.

The next two papers deal with lexical aspects of the Malay language. Corpus research...
on Malay has been mostly corpus-based, that is, corpora have been used as databases of exemplars of pre-determined patterns. The corpus-driven approach, which discovers patterns from generated sequences, is not easily executed for Malay data. This is due to the fact that most Malay corpora (Korpus DBP\textsuperscript{1}; MalayWac,\textsuperscript{2} a collection of Malay websites in the Sketch Engine\textsuperscript{3}; Malay Concordance Project,\textsuperscript{4} a collection of historical manuscripts with advanced search functions, etc.) do not provide word-for-word tags. Even though there have been some tagged Malay corpora, these are not made available to the public. Therefore, due to the shortage of resources, corpus research on Malay is largely dependent on running texts without part-of-speech tagging. Most authors find their own ways of retrieving linguistic patterns in their work.

The paper by Siaw-Fong Chung demonstrates how Malay research could be carried out using running texts from a large number of collected texts. She utilizes a self-collected corpus from a Malaysian newspaper Utusan Malaysia for the purpose of observing the use of Malay antonyms bawah:atas and dalam:luar. She makes use of a concordance tool, AntConc,\textsuperscript{5} to retrieve the patterns she would like to observe—antonyms bawah:atas and dalam:luar, and their co-occurrences with the locative marker di. She analyzes whether the same collocates that appear with a preposition are also found in its antonymous counterpart. Similarly, she also compares the collocates when di is present or absent for both antonyms. Her results show that antonym pairs are not always symmetrical, and when they are, they often collocate with concrete nouns such as tingkat atas/bawah, atas/bawah meja, etc. As for the co-appearance with di, unlike what many have found about the specificity of di, some collocates of di luar are not as concrete and specific as those of luar. On the other hand, bawah and di bawah do not seem to be very different in their list of collocates, indicating the specificity feature of di is not always present with many prepositions.

As mentioned previously, the application of technology in Malay research is in high demand. While Chung’s study demonstrates how a corpus with running texts could be utilized, a tagged corpus could have made it more sophisticated. Quality, manually annotated lexical resources (e.g. sense categorization, part-of-speech tagging, semantic relation links, etc.) need to be developed to serve as the standard resources in the linguistic and natural language processing research of Malay.

The paper by Francis Bond, Lian Tze Lim, Enya Kong Tang and Hammam Riza provides the details of the creation of a Malay lexical resource, the Open WordNet Bahasa. The English WordNet has been used in multiple ways by linguists, especially computational linguists. With the presence of a lexical resource, mapping to other resources will be possible. For example, the English WordNet has mapping to the Suggested Upper Merged Ontology (SUMO) that allows concept information to be added to an existing lexical resource. For Malay researchers, and for Malay lexical semanticists and compu-

\textsuperscript{1} http://sbmb.dbp.gov.my/korpusdbp/SelectUserCat.aspx

\textsuperscript{2} MalayWac consists of websites in Malay regardless of its regional differences. Thus, Indonesian, Malaysian Malay, and other standard forms of Malay are possibly found in this web corpus.

\textsuperscript{3} http://www.sketchengine.co.uk

\textsuperscript{4} http://mcp.anu.edu.au

\textsuperscript{5} http://www.laurenceanthony.net/software.html
tational linguists, a better Malay lexical resource will definitely provide new directions of research. The paper by Bond et al. clearly answers this need. It shows how a lexical resource with broad coverage is needed for Malaysian Malay and Indonesian. Their Open WordNet Bahasa is an open combined semantic resource for Malaysian Malay and Indonesian, combining and extending several existing Malay wordnets. This resource replicates the lexical relations that are found in the English WordNet by providing relations such as hypernym, hyponym, meronym-part, etc. The resource also has a mapping to the SUMO and dictionary meanings. The Open WordNet Bahasa has a portion of sense tagged data for Indonesian.

Jakrabhop Iamdanush and Pittayawat Pittayaporn describe the causative constructions in Patani Malay, a Malay variety spoken in Thailand’s Deep South. Although language contact is a favourite topic of the linguists working on the languages in the Malay Archipelago, the contact situation in this northernmost part of the Malay-speaking world has not caught much attention. One obvious reason for this is that Patani Malay is spoken in Thailand, a country whose national/official language is not Austronesian, let alone Malayic. A quick inspection of Iamdanush and Pittayaporn’s references reveals that it is not the case that linguists in Thailand downplay Patani Malay, but some research results are not easy to access for non-Thai linguists, as they are published in Thailand in the Thai language. Iamdanush and Pittayaporn’s paper thus counts as a valuable source of knowledge to many of us who do not read Thai. They describe three periphrastic causative constructions in Patani Malay, which involve waʔ ‘to do’ (cf. Standard Malay buat [buwah’]), wi ‘to give’ (cf. Standard Malay beri [bər̥i]) and the combination of these two morphemes waʔwi. Iamdanush and Pittayaporn compare these constructions with causative constructions in other Malay varieties (Colloquial Malaysian Malay, Standard Indonesian and Eastern Malay varieties in Indonesia) and Thai, and show that the Patani Malay causative constructions resemble the latter. Based on socio-historical and linguistic facts, they conclude that the resemblance is due to long-term asymmetrical language contact with Thai, the language of prestige in the area.

As we conclude this introduction, it must be noted that there are two recent trends in general linguistics that have yet to be explored well in Malay linguistics: studies with rigorous statistical analysis, especially in phonology and morphology, and experimental studies. We hope that a similar volume in the future will include contributions in these areas.

References

