

# **Knowledge Gaps in Contemporary Javanese Comprehensive Description: Towards the Making of Javanese Annotated Corpus 1945-2015 periods**

Atin Fitriana  
Dwi Puspitorini  
Totok Suhardijanto

University of Indonesia

## **Abstract**

Since its beginning, the descriptive study of the Javanese is actually done by many foreign researchers. They are Uhlenbeck (1978), Robson (1992), and Oglobin (2001). One of the obstacles in descriptive studies on the Javanese is the absence of a database that can be used by the researcher. Meanwhile, the database of Javanese, either in the form of newspaper or magazine articles, novels, text books, manuscripts show, lyrics, song, and etc are very abundant compared to other local languages in Indonesia. This paper will discuss the matters in the preparation of Javanese annotated corpus making in Javanese Study Program, University of Indonesia. The annotations that will be developed include the word class labeling (POS tagging), semantic labeling that consist of NER (name and entity recognition), up to syntactic labeling for the parser needs. The making of contemporary Javanese corpus is also mind the aspect of equilibrium (balance) and representativeness. The type of data that is inserted into the corpus consists of various types of text, but with the period limitation. Temporarily, Old Javanese text and Middle Javanese are not into the goal in the preparation of this corpus because of the consideration that the Old Javanese corpus is being developed by foreign researchers. The annotated Javanese corpus is expected to be utilized as a research database for language and culture of Java researchers, not only in Indonesia, but also outside of Indonesia.

Keywords: Javanese, corpus, language database, labeling (tagging), text and text types

## References

- Balai Bahasa Yogyakarta. (2006). *Pedoman Umum Ejaan Bahasa Jawa yang Disempurnakan*. Edisi Revisi. Yogyakarta: Penerbit Kanisius.
- Chafe, W. (1985). Linguistic Differences Produced by Differences between Speaking and Writing. In David R. Olson, Andrea Hildyard, and Nancy-Torrance (eds.), *Literacy, Language, and Learning*. Cambridge: Cambridge University Press. hlm. 105-123.
- Jonsson, E. (2015). *Conversational Writing: A Multidimensional Study of Synchronous and Supersynchronous Computer-Mediated Communication*. Frankfurt am Main: Peter Lang Publishing Group.
- Robson, Stuart. (1992). *Javanese Grammar For Students*. Australia: Monash University.
- Subroto, D Edi. (1991). *Tata Bahasa Deskriptif Bahasa Jawa*. Jakarta: Departemen Pendidikan dan Kebudayaan.
- Sudaryanto. (1991). *Tata Bahasa Baku Bahasa Jawa*. Yogyakarta: Duta Wacana University Press.
- Ogloblin, Alexander K. (2001). Imperative sentences in Javanese. In Xrakovskij, Victor S. (ed.), *Typology of Imperative Constructions*, 221-242. München: Lincom Europa
- Uhlenbeck, E.M.. (1978). *Studies in Javanese Morphology*. Netherlands: Springer Netherlands.
- Wedhawati dkk.. (2001). *Tata Bahasa Jawa Mutakhir*. Jakarta: Departemen Pendidikan Nasional.