The first LingDy Forum in 2016
April 5, 2016

Building an open-source computational grammar for Indonesian "INDRA"

David Moeljadi
Nanyang Technological University, Singapore

Abstract:
In this short presentation, I would like to introduce my ongoing project to build an open-source computational grammar for Indonesian "INDRA" (Indonesian Resource Grammar) (Moeljadi et al. 2015), modeled in Head-Driven Phrase Structure Grammar (HPSG) (Sag et al. 2003), which can parse and generate text. INDRA is being created and developed using tools from the Deep Linguistic Processing with HPSG-Initiative (DELPH-IN) (http://www.delph-in.net/) and employs Wordnet Bahasa (Nurril Hirfana et al. 2011, Bond et al. 2014) as lexical source. INDRA aims to parse and treebank Indonesian text in Nanyang Technological University Multilingual Corpus (NTU-MC) (Tan and Bond 2012, Bond et al. 2013) and will be applied to machine translation.

Bond, Francis, Shan Wang, Eshley Huini Gao, Hazel Shuwen Mok and Jeanette Yiwen Tan. 2013. Developing parallel sense-tagged corpora with wordnets. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse (LAW 2013), 149–158. Sofia.
Bond, Francis, Lian Tze Lim, Enya Kong Tang and Hammam Riza. 2014. The combined wordnet bahasa. In NUSA: Linguistic studies of languages in and around Indonesia 57, 83–100.
Moeljadi, David, Francis Bond and Sanghoun Song (2015) Building an HPSG-based Indonesian Resource Grammar (INDRA). In Proceedings of the Grammar Engineering Across Frameworks (GEAF) Workshop, 53rd Annual Meeting of the ACL and 7th IJCNLP, 9–16, Beijing, China, July 26-31, 2015.
Nurril Hirfana Mohamed Noor, Suerya Sapuan and Francis Bond. 2011. Creating the open Wordnet Bahasa. In Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation (PACLIC 25), 258–267. Singapore.
Sag, Ivan A., Thomas Wasow and Emily M. Bender. 2003. Syntactic Theory: A Formal Introduction. Stanford: CSLI Publications 2nd edn.
Tan, Liling and Francis Bond. 2012. Building and annotating the linguistically diverse NTU-MC (NTU-multilingual corpus). In International Journal of Asian Language Processing 22(4), 161–174.

Keywords:
Indonesian grammar, computational grammar, grammar engineering, HPSG, Wordnet Bahasa, parallel corpus, NTU-MC, machine translation