

Multiword Expressions and the Low-Resource Scenario from the Perspective of a Local Oral Culture

Steven Bird

Charles Darwin University
steven.bird@cdu.edu.au

Abstract

Research on multiword expressions and on under-resourced languages often begins with problematisation. The existence of non-compositional meaning, or the paucity of conventional language resources, are treated as problems to be solved. This perspective is associated with the view of Language as a lexico-grammatical code, and of NLP as a conventional sequence of computational tasks. In this talk, I share from my experience in an Australian Aboriginal community, where people tend to see language as an expression of identity and of ‘connection to country’. Here, my early attempts to collect language data were thwarted. There was no obvious role for tasks like speech recognition, parsing, or translation. Instead, working under the authority of local elders, I pivoted to language processing tasks that were more in keeping with local interests and aspirations. I describe these tasks and suggest some new ways of framing the work of NLP, and I explore implications for work on multiword expressions and on under-resourced languages.