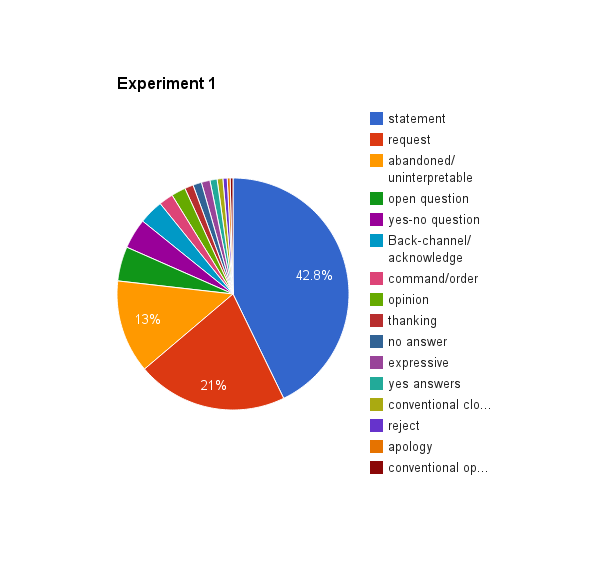
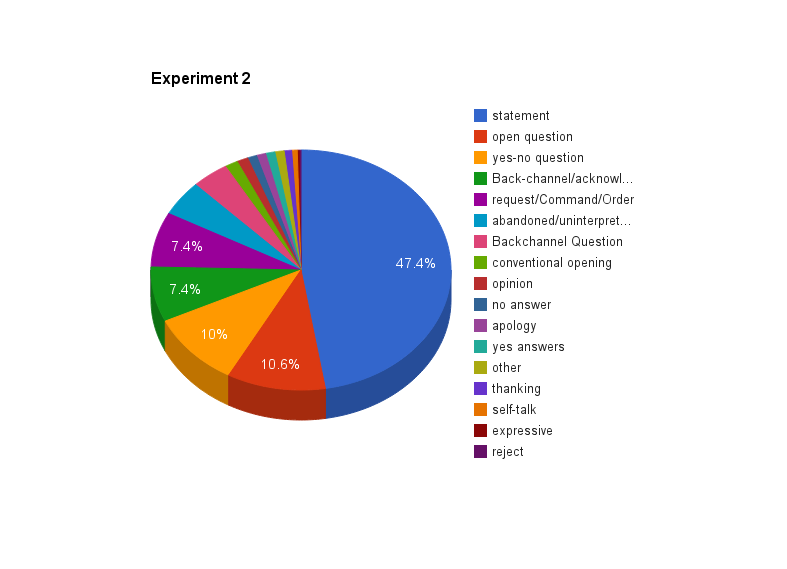
**Tag Set Selection**

In order to DA recognition for Sinhala, we need to have pre identified set of tags which suits the nature of Sinhala. Selecting suitable dialogue act tag set is a vital step in Dialogue Act Recognition and it’s heavily depends on the language, culture and the context of the target application/task. So, we went through popular tag sets like DAMSL, Meeting recorder DA tag set and VERMOBIL tag set and identified which DAs has frequently used in related studies by doing a horizontal analysis on tag set used in related studies .We did this tag selection process in an iterative manner by feeding the results of the experiments conducted. The following are results of some experiments we conducted. We paid significant attention to nature of Sinhala language while selecting the DA tag set.

**Tag Set Selection Iteration 01**

* Decided to merge request and command /order tag into one. It’s hard to differentiate them in Sinhala compared to English.
* Decided to introduce new tags such as back channel question and self-talks. Back channel question is often used in Sinhala.(EX: එහෙමද? -ehemada?)

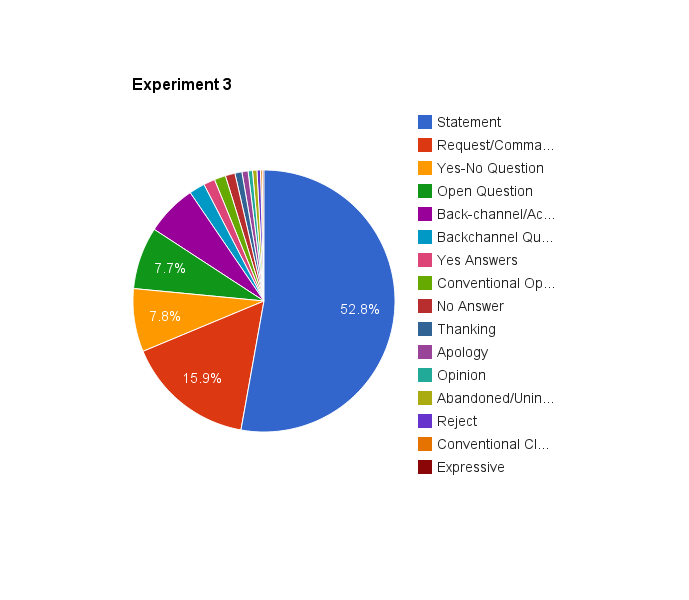


**Tag Set Selection Iteration 02**

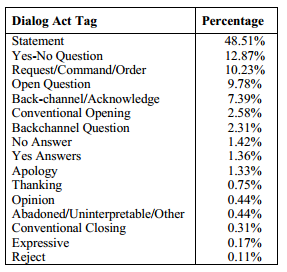
* Decided to merge ‘Other’ tag and ‘Abandoned/Uninterruptable’ into a single tag.
* Decided to remove newly introduced self-talk(0.6 %) due to its infrequent nature in Sinhala language conversations.

**Some Common observations**

* Wh- questions like what’, ‘when’, ‘why’, ‘which’ is not identifiable in Sinhala due its lexical characteristics. So, we have a wider tag call open questions which contain Wh-question and other forms of questions except yes-No questions and backchannel questions.

****

**Finalized tag set with percentage**

****

* The percentage of each DA type approximately matches with observations of related studies.