Turn-taking Labeling Guidelines

Developed by Julia Hirschberg, Stefan Benus, Agustín Gravano, Héctor Chávez, and Enrique Henestroza, and based on the labeling scheme proposed in [Beattie 1982].

1) Turn exchanges, ‘turns’ tier

Label only the turn intervals inside tasks (tasks are marked by intervals that start with “Images:” in the ‘tasks’ tier).

For each turn interval by S2, where S1 is the other speaker, label S2’s turn interval as follows:

Is S2’s utterance in response to S1’s utterance and indicates only “I’m still here / I hear you and please continue”?\(^{(1)}\)

Simultaneous speech present?

S2 is successful?

S1’s utterance complete?\(^{(2)}\)

Overlap (O)

Butting-in (BI)

Smooth switch (S)

Pause interruption (PI)

Backchannel with overlap (BC_O)

Backchannel (BC)

(1) Backchannels were identified by three annotators for the affirmative cue words project, who were provided with the following definition:

*Backchannel: The function of ‘okay’ [or ‘alright’, ‘mm-hm’, ‘yeah’, etc.] in response to another speaker’s utterance that indicates only “I’m still here / I hear you and please continue”.*

When a simple majority of annotators (i.e., at least two out of three) considered an utterance to be a backchannel, it was labeled ‘BC’ or ‘BC_O’.

(2) We use Beattie’s informal definition of utterance completeness: “Completeness was judged intuitively, taking into account the intonation, syntax, and meaning of the utterance” (Beattie 1982, page 100, second note).

Special cases:

We identified three common cases in which no turn exchange occurs, and the corresponding turn interval receives a special label ‘X[1-3]’.
• **Task beginnings:** If a turn interval begins a new task, then label it ‘X1’.

• **Continuation after a backchannel:** If a turn interval $t$ is a continuation from the previous turn by the same speaker after a ‘BC’ or ‘BC_O’, then label it ‘X2_O’ if $t$ overlaps the backchannel, or ‘X2’ if not.

• **Simultaneous start:** If two turn intervals begin almost simultaneously—formally, within 210 ms of each other [Fry 1975]—then the speakers are most probably reacting to the preceding turn interval:

$$A_1 \quad \frac{x}{y} \quad B_1 \quad A_2 \quad 0 < |y - x| < 210 \text{ms}$$

In the figure, $A_2$ and $B_1$ occur most likely in response to $A_1$. Thus, $B_1$ should be labeled with respect to $A_1$ (not $A_2$); $A_2$ should be labeled ‘X3’.

**Notes:**

• The figure below shows a frequent pattern consisting of a complete short utterance ($B_1$) fully contained within a longer utterance ($A_1$) by the other speaker, such that the floor is briefly *shared* by both speakers, and $A_1$ is not disrupted by $B_1$. In such cases, the most appropriate label for $B_1$, according to our labeling scheme, is ‘O’; it is neither ‘I’ nor ‘BI’ because both utterances are complete.

2) **Miscellaneous tier**

**Collaborative contributions:** If a speaker completes, or attempts to complete, an utterance from their interlocutor, as if trying to help them, add a ‘H’ label in the misc tier.

**Other:** Mark in the misc tier any other situation not contemplated in these guidelines.

**References**
