Within and Between Speaker Transitions in Multiparty Casual Conversation

Emer Gilmartin
ADAPT Centre, Trinity College Dublin
gilmare@tcd.ie

Carl Vogel
Trinity College Duboin
vogel@tcd.ie

1 Introduction

Casual conversation, 'talk for the sake of talking', has been observed to occur in two main phases or sub-genres – interactive *chat* where most or all participants contribute, and more monologic *chunk* phases, where one speaker dominates the conversation, often telling a story or giving an extended opinion (Eggins and Slade, 2004). Previous work has shown differences in the length, composition in terms of speech, silence and overlap, and in the relative frequencies of chat and chunk phases in casual conversation (Gilmartin et al., 2019). In this work we use the timing of speech and silence in chat and chunk phases to explore transitions between single party speech by a speaker and the next stretch of single party speech by the same speaker (*within speaker transition*) or another speaker (*between speaker transition*). We define ***1Sp*** as an interval of single party speech and ***1Sp1*** as a ***1Sp*** of duration one second or more. We also adapt the terminology used in (Heldner and Edlund, 2010) for dyadic interaction. For speakers A and B, within speaker silence (WSS) is defined as **A_GX_A** and between speaker silence (BSS) is defined as **A_GX_B** where GX denotes global silence, while within and between speaker overlap are **A_AB_A** and **A_AB_B**. Thus, ***1Sp*** can transition back to ***1Sp*** with one intervening interval of silence or overlap, e.g. **1_0_1** or **1_2_1**. For multiparty interaction, more possibilities emerge. As multiparty transitions can involve a combination of overlap and silence, we define only two transition types – *within speaker transitions* (WST) beginning and ending with the same speaker, and *between speaker transitions* (BST), which start with one single speaker and transition to another single speaker.

2 Data and Annotation

The CasualTalk dataset is a collection of six 3 to 5 party casual conversations of around one hour each, drawn from the d64, DANS, and TableTalk corpora (Oertel et al., 2010; Hennig et al., 2014; Campbell, 2008).

The data were segmented and transcribed manually and a total of 213 chat and 358 chunk phases were identified and annotated, as described in (Gilmartin and Campbell, 2016). The data were also segmented into 30688 floor state intervals reflecting the participants speaking or silent at any time.

3 Transitions between Single Speakers

For each ***1Sp1***, we searched forward in the dataset to locate the next ***1Sp1*** and extracted the sequence of intervals (in terms of speaker numbers) from the initial ***1Sp1*** to the next ***1Sp1***. As an example, **1_2_3_2_1_0_1** contains 5 intervening intervals between the two stretches of ***1Sp1***.

Distributions of ***1Sp1–1Sp1*** transitions are shown in Figure 1, where it can be seen that the vast majority of intervening intervals are in stretches of odd numbers of intervals, with the number of cases dropping with increasing intervals. Overall, 95.53% of all ***1Sp1-*** intervals are closed by a later ***1Sp1*** in fewer than 16 intervening intervals. Even-number cases accounted for only 112 (2.1%) of the 5382 transitions between 1 and 15 intervals long. The most frequent class of transitions are those with one intervening interval which account for 41.13% of cases. 21.74% were WSTs while BSTs accounted for 73.78%. For the remaining 4.47% of ***1Sp1-*** intervals, labelled 16+, at least 16 intervals occurred before a second ***1Sp1*** interval was encountered.

In both chat and chunk, disregarding the even-number cases, the number of transitions

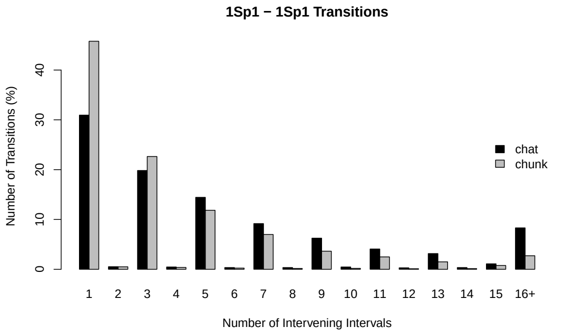

Figure 1: Number of floor state intervals between single-speaker intervals of 1 second or more in duration

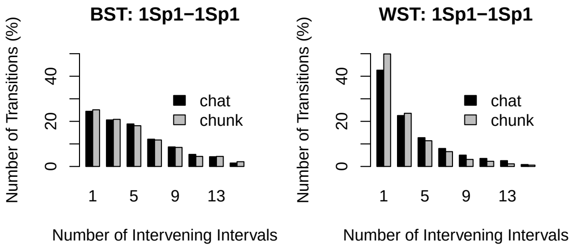

Figure 2: Number of floor state intervals between (***1Sp1***) intervals in Between Speaker Transitions (BST, left) and Within Speaker Transitions (WST, right) in chat and chunk phases.

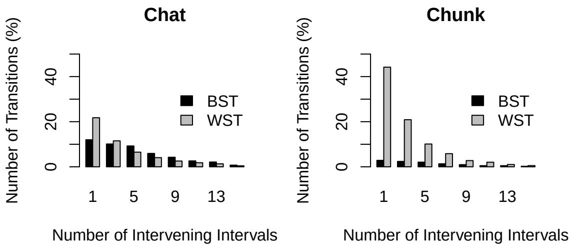

Figure 3: Percentage of Between and Within Speaker Transitions per number of floor state intervals in (***1Sp1***-***1Sp1***) in chat and chunk phases.

declines monotonically with the number of intervening intervals between ***1Sp1*** intervals. The chat condition starts with a smaller percentage of 1-interval transitions and declines at a lower rate than the chunk condition. In both conditions, it is likely that numbers continue to decline with increasing intervals in a long tail. The 16+ category, a bucket category, is more than three times as large proportionally in chat (8.31%) as in chunk(2.71%).

The odd numbered cases and the 16+ interval bucket class were excluded from the ***1Sp1***-***1Sp1*** transition data, leaving 5270 transitions, comprising 77.24% WST and 22.76% BST with intervening intervals ranging from 1 to 15. Figure 2 shows these BST and WST transitions by number of participants, while Figure 3 shows interval types in chat and chunk phases, and the proportion of transitions per interval total. One-interval transitions were the largest group for BST and WST for both chat and chunk, with the proportion of 1-interval transitions particularly high for WST, and very much so in the case of chunk

4 Discussion and Conclusions

The results on transition n-grams between intervals of one speaker speaking in the clear for at least one second (***1Sp1***) show that chat and chunk differ in that between speaker transitions in chat interaction are spread over more intervening intervals than in chunk, thus increasing the frequency of more complex transitions. This could reflect more turn competition, or indeed more backchannels and acknowledgement tokens being contributed by more participants. Within speaker transitions are predominantly one-interval, perhaps reflecting breathing pauses. One-interval transitions comprise the largest class, with a higher proportion of one-interval transitions in chunk than chat, and higher proportions of within speaker than between speaker one-interval transitions in both, but particularly in monologic chunk. However, one-interval transitions only account for 41.03% of transitions overall, reflecting the need to consider more complex transitions around turn change and retention. It would be very interesting to separate within speaker breathing pauses from other transitions in order to better understand transitions around silence. Future work involves further classification of transitions depending on the number of distinct speakers involved, and investigation of the duration of transitions. It is hoped that this study, and similar studies of other corpora, will allow us to inventory transition types in multiparty spoken interaction, and then analyse examples of the statistically more likely transitions in detail to better understand speaker transitions.

Acknowledgments

This work is supported by the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

N. Campbell. 2008. Multimodal processing of discourse information; the effect of synchrony. In *Universal Communication, 2008. ISUC'08. Second International Symposium on*, pages 12–15.

S. Eggins and D. Slade. 2004. *Analysing casual conversation*. Equinox Publishing Ltd.

Emer Gilmartin and Nick Campbell. 2016. Capturing Chat: Annotation and Tools for Multiparty Casual Conversation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Emer Gilmartin, Benjamin R Cowan, Carl Vogel, and Nick Campbell. 2019. Chunks in multiparty conversationâĂŤbuilding blocks for extended social talk. In *Advanced Social Interaction with Agents*, pages 37–44. Springer.

Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568.

Shannon Hennig, Ryad Chellali, and Nick Campbell. 2014. The D-ANS corpus: the Dublin-Autonomous Nervous System corpus of biosignal and multimodal recordings of conversational speech. Reykjavik, Iceland.

Catharine Oertel, Fred Cummins, Jens Edlund, Petra Wagner, and Nick Campbell. 2010. D64: A corpus of richly recorded conversational interaction. *Journal on Multimodal User Interfaces*, pages 1–10.