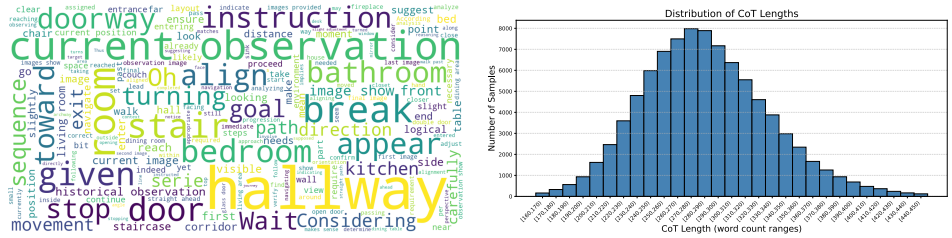
# B Technical Appendices and Supplementary Material




(a) (b)

Figure 5: (a) Word cloud of Chain-of-Thought in the R2R-CoT-320k dataset, highlighting frequent visual and spatial reasoning patterns. (b) Distribution of CoT lengths (in word count), showing a wide and diverse range of reasoning complexity.

## B.1 More Details of the R2R-CoT-320k Dataset

To provide a deeper quantitative understanding of our proposed R2R-CoT-320k dataset, we present statistics on CoT content and complexity. As shown in Fig. 5a, the word cloud reveals frequent reasoning patterns grounded in spatial semantics, such as "doorway," "current observation," "hallway," "turning," and "goal." These tokens suggest that the dataset captures rich, step-by-step reasoning tightly aligned with embodied navigation semantics.

Fig. 5b shows the distribution of CoT lengths, where most reasoning chains fall within the 200–300 word range, but with a long tail reaching beyond 450 words. This indicates that the dataset covers both concise and highly detailed reasoning processes, posing a greater challenge than typical short-form CoT datasets used in static tasks.

Overall, R2R-CoT-320k represents the first large-scale reasoning-augmented dataset for VLN with diverse, high-coverage CoT annotations. It offers a valuable benchmark to study the role of language-based reasoning in long-horizon, partially observable navigation tasks.